UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

Random Forest interpretability - explaining classification models and multivariate data through logic rules visualizations

Mário Popolin Neto

Tese de Doutorado do Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional (PPG-CCMC)

ICMC USP SÃO CARLOS

Mário Popolin Neto

Random Forest interpretability - explaining classification models and multivariate data through logic rules visualizations

Thesis submitted to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP – in accordance with the requirements of the Computer and Mathematical Sciences Graduate Program, for the degree of Doctor in Science. *FINAL VERSION*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Prof. Dr. Fernando Vieira Paulovich

USP – São Carlos
February 2022

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

P829r

Popolin Neto, Mário
Random Forest interpretability - explaining classification models and multivariate data through logic rules visualizations / Mário Popolin Neto; orientador Fernando Vieira Paulovich. -- São Carlos, 2022.
128 p.

Tese (Doutorado - Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional) -- Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2022.

1. Logic Rules Visualization. 2. Random Forest. 3. Classification Models Interpretability. 4. Models and Multivariate Data Explanations. I. Paulovich, Fernando Vieira, orient. II. Título.

Bibliotecários responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2:
Gláucia Maria Saia Cristianini - CRB - 8/4938
Juliana de Souza Moraes - CRB - 8/6176

Mário Popolin Neto

Intepretabilidade de Random Forest - explicando modelos de classificação e dados multivariados por meio de visualizações de regras lógicas

Tese apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Doutor em Ciências – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientador: Prof. Dr. Fernando Vieira Paulovich

USP – São Carlos
Fevereiro de 2022

Dedicated to Mercedes Sardinha Moretti,
who loved me as her son.

ACKNOWLEDGEMENTS

First, I thank God for my life.

I acknowledge my family, since, without their support, I would not be here. My mother, Maria Terezinha Sardinha Popolin, loves to learn and has passed on to me this gift; my father, Cláudio Popolin, helped me face great challenges in my academic career; my young brothers Guilherme Sardinha Popolin and Cláudio Popolin Júnior are the biggest enthusiasts of my work; my godfather, Antonio Humberto Moretti, is always cheering for me; and my grandfather Mário Popolin, grandmother Tereza de Oliveira Sardinha, and godmother Mercedes Sardinha Moretti (in memoriam) were responsible for many precious memories that inspired me.

In particular, I thank my beloved wife, Poliane Aparecida Segato Popolin, who deeply cared for me, listening countless times to the findings and obstacles I faced during this doctoral project - even in my sleep, I tried to explain what I was doing in the project (see the epigraph). Furthermore, she kindly took care of my right shoulder, which was (and still is) in extreme pain due to stress.

I want to express my gratitude to my advisor, Prof. Fernando Vieira Paulovich, who accepted me as his Ph.D. student, guiding me through the past four years and making me a better researcher. He led me to outstanding accomplishments, and I am very grateful for everything I have learned from him.

I wish to thank all professors and staff from the Instituto de Ciências Matemáticas e de Computação (ICMC) of the University of São Paulo (USP), and Prof. Osvaldo Novais de Oliveira Júnior (São Carlos Institute of Physics - USP) and Andrey Coatrini Soares - their promptness in helping me with numerous questions of analytical chemistry (sometimes over and over) was fundamental. I also express my thanks to Gabriel Dias Cantareira, Martha Dais Ferreira, and Mateus Malvessi Pereira for helping me directly and indirectly. I acknowledge the VGRS (Visiting Graduate Research Students) program from Dalhousie University, Nova Scotia, Canada.

I would like to mention my dear friends Prof. José Remo Ferreira Brega and Prof. Diego Roberto Colombo Dias, who encouraged me to pursue a Ph.D. degree, and also acknowledge the series of pictures from Matt Might in "The Illustrated Guide to a Ph.D." [1]. I saw these pictures for the first time in 2015, and such inspiring visual representations

[1] <http://matt.might.net/articles/phd-school-in-pictures>

frequently crossed my mind over the past four years.

I sincerely acknowledge all the teachers/professors that I had in my life as a student. Furthermore, I would like to thank all my friends. I thank Angela Cristina Pregnolato Giampedro and Dipankar Mazumdar for giving me feedback to improve my writing in this Ph.D. thesis. Last but not least, I thank the support from the Qualification Program of the Federal Institute of São Paulo (IFSP).

"Por essa a gente começa a entender como é que foi classificado."
(Sonilóquio do autor)

RESUMO

POPOLIN NETO, M. **Intepretabilidade de Random Forest - explicando modelos de classificação e dados multivariados por meio de visualizações de regras lógicas**. 2022. 128 p. Tese (Doutorado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2022.

Modelos de classificação possuem imenso potencial e futuro ubíquo, considerando o vasto número de tarefas preditivas em diferentes domínios onde estes modelos são aplicáveis. A interpretabilidade dos modelos pode ser tão importante quanto a performance, fornecendo explicações globais e locais para interpretar os conhecimentos adquiridos e auditar decisões. Além da capacidade preditiva, modelos de classificação podem ser aplicados como ferramentas descritivas, onde intepretabilidade envolve explicações de dados. Regras lógicas vêm sendo amplamente utilizadas em soluções para interpretabilidade e Decision Trees são reconhecidas pela geração de regras lógicas consistentes. A abordagem Random Forest – conjunto de Decision Trees – tem sido amplamente adotada devido a sua habilidade em produzir resultados precisos e manipular conjuntos de dados multivariados. Entretanto, a intepretabilidade de modelos Random Forest enfrenta o desafio de gerir um número considerável de regras. Baseado na visualização de regras lógicas em uma metáfora visual em formato de matriz, esta tese de doutorado resulta em métodos de Visual Analytics para a intepretabilidade de modelos Random Forest, suportando explicações de modelos e de dados cobrindo propósitos preditivos e descritivos. Para explicações de modelos (preditivo), ExMatrix dispõe regras lógicas a formar representações visuais globais e locais, fornecendo visões gerais e análises de decisões. Explicações globais podem revelar o conhecimento aprendido pelo modelo a partir de um conjunto de dados rotulados, enquanto explicações locais focam na classificação de uma instância de dados em particular. Para explicações de dados (descritivo), VAX processa regras lógicas resultando na visualização de regras descritivas para insights automáticos dos dados. Explicações de dados permitem a identificação e a interpretação visual de padrões em conjuntos de dados multivariados. Qualquer problema representado por um conjunto de dados rotulados é um potencial caso de uso para os métodos propostos. O método ExMatrix foi aplicado em química analítica e o método VAX empregado em conjuntos de dados reais para análises de dados multivariados. A principal contribuição desta tese de doutorado reside em métodos de Visual Analytics suportando a interpretabilidade de Random Forest para propósitos preditivos e descritivos em explicações de modelo e de dados.

Palavras-chave: Visualização de Regras Lógicas, Random Forest, Intepretailidade de Modelos de Classificação e Explicações de Modelos e de Dados Multivariados.

ABSTRACT

POPOLIN NETO, M. **Random Forest interpretability - explaining classification models and multivariate data through logic rules visualizations**. 2022. 128 p. Tese (Doutorado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2022.

Classification models have immense potential and ubiquitous future, considering the vast number of prediction tasks in different domains where such models are applicable. Models' interpretability may be just as important as performance, providing global and local explanations to interpret the acquired knowledge and audit decisions. In addition to the predictive ability, classification models can also be employed as descriptive tools, where interpretability involves data explanations. Logic rules have been widely used in interpretability solutions, and Decision Trees are well recognized for consistent logic rules generation. The Random Forest approach (Decision Trees ensemble) has been broadly adopted due to its ability to produce accurate results and deal with multivariate datasets. However, Random Forest models' interpretability faces the challenge of handling a substantial number of logic rules. Based on logic rules visualization into a matrix-like visual metaphor, this doctoral thesis leads to Visual Analytics methods for Random Forest models' interpretability, supporting models and data explanations covering predictive and descriptive purposes. For models (predictive) explanations, ExMatrix arranges logic rules towards global and local visual representations, providing overviews and decisions reasoning. Global explanations can unveil the knowledge learned by the model from a class-labeled dataset, whereas local explanations focus on a particular data instance classification. For data (descriptive) explanations, VAX handles logic rules, resulting in descriptive rules visualization for automated data insights. Data explanations support the identification and visual interpretation of patterns in multivariate datasets. Any problem denoted by a class-labeled dataset is a potential use case for the proposed methods. ExMatrix was applied in analytical chemistry, and VAX was used in real-world datasets for multivariate data analyses. The main contribution of this doctoral thesis lies in Visual Analytics methods supporting Random Forest interpretability for predictive and descriptive purposes in model and data explanations.

Keywords: Logic Rules Visualization, Random Forest, Classification Models Interpretability, and Models and Multivariate Data Explanations.

LIST OF FIGURES

LIST OF ALGORITHMS

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

ANN	Artificial Neural Networks
DR	Dimensional Reduction
DT	Decision Tree
EP	Emerging Patterns
FET	Fisher Exact Test
GE	Global Explanation
IDMAP	Interactive Document Mapping
JEP	Jumping Emerging Pattern
LE/SC	Local Explanation Showing Smallest Changes
LE/UR	Local Explanation Showing the Used Rules
MCS	Multidimensional Calibration Space
MDI	Mean Decrease Impurity
PAH	Poly Allylamine Hydrochloride
POC	Point-Of-Care
PVS	Poly Vinyl Sulfonic Acid
RF	Random Forest
RFm	Random Forest miner
SDSN	Sustainable Development Solutions Network
SERS	Surface-Enhanced Raman Scattering
SVM	Support Vector Machines
VA	Visual Analytics
VND	Visual Neural Decomposition
VSD	Visualizing Subgroup Distribution
WBCD	Wisconsin Breast Cancer Diagnostic
XAI	Explainable Artificial Intelligence

CONTENTS

CHAPTER

1

INTRODUCTION

1.1 Context

Many real-world problems can be modeled into prediction tasks in which classification models from machine learning are handy. From class-labeled datasets, such models are capable of learning relationships between variables (a.k.a. attributes, features, or dimensions) and classes, so when a new data instance is provided, it is associated with a particular class. Shifting from the effort to improve models' quantitative metrics like accuracy, Explainable Artificial Intelligence (XAI) has been in the spotlight (ADADI; BERRADA, 2018; LIAO; GRUEN; MILLER, 2020). Several XAI solutions aim at classification models' interpretability, going beyond quantitative metrics analysis (e.g., confusion matrix) (TAN; STEINBACH; KUMAR, 2005) towards making the model's overall logic and its decisions understandable to humans (RIBEIRO; SINGH; GUESTRIN, 2016; GUIDOTTI *et al.*, 2018b; LIAO; GRUEN; MILLER, 2020). For example, from a disease diagnostic model, interpretability may involve understanding the knowledge learned about the disease and the outcome for a particular patient (RIBEIRO; SINGH; GUESTRIN, 2016; CLOUGH *et al.*, 2019). Therefore, model interpretability (i.e., explainability) [1] can be

[1] The terms "interpretability" and "explainability" can be seen as tied concepts in the machine learning community (CARVALHO; PEREIRA; CARDOSO, 2019). Indeed, such terms have been often used interchangeably (GILPIN *et al.*, 2018; CARVALHO; PEREIRA; CARDOSO, 2019; BRONIATOWSKI, 2021; GAUR; FALDU; SHETH, 2021). However, some authors reasoned about the distinctions (GILPIN *et al.*, 2018; BRONIATOWSKI, 2021). Interpretability may be associated with meaningful descriptions for humans to make sense of an algorithm's output, appealing for their cognition, knowledge, and biases (GILPIN *et al.*, 2018; BRONIATOWSKI, 2021). For example, suppose a rental application rejected by a classification model, users could employ their background (domain) knowledge to interpret the data instance (variables values) representing the applicant (BRONIATOWSKI, 2021). On the other hand, explainability can be related to unveiling internal data processing or representation, which relies on the algorithm's technical structure (GILPIN *et al.*, 2018; BRONIATOWSKI, 2021). Thus taking the early example, regarding the applicant rejected by the classification model, users

related to global and local explanations for model overview and classification process reasoning (GUIDOTTI *et al.*, 2018b; DU; LIU; HU, 2019). Given the potential of classification models and the various current applications across human society, models interpretability is expected to be the focus of governmental initiatives, e.g., *European General Data Protection Regulation*, which demands explanations on automated decisions concerning individuals (GUIDOTTI *et al.*, 2018b; CARVALHO; PEREIRA; CARDOSO, 2019) [2].

Logic rules (if-then rules) are commonly applied for model interpretability (MING; QU; BERTINI, 2019; RIBEIRO; SINGH; GUESTRIN, 2018; GUIDOTTI *et al.*, 2018a; LAKKARAJU; BACH; LESKOVEC, 2016) [3], since logic statements leading to a particular class are intrinsically understandable by humans (FüRNKRANZ; GAMBERGER; LAVRAC, 2012; LAKKARAJU; BACH; LESKOVEC, 2016; GUIDOTTI *et al.*, 2018b; MIRANDA; SARDINHA; CERRI, 2021). They can be extracted from the model, like Decision Tree (DT) (GUIDOTTI *et al.*, 2018b), or inferred as surrogates of complex black-box models such as Artificial Neural Networks (ANN) and Support Vector Machines (SVM) (MING; QU; BERTINI, 2019; GUIDOTTI *et al.*, 2018a). DT models are well recognized for generating consistent logic rules (MIRANDA; SARDINHA; CERRI, 2021). Inherent interpretable models like DTs generally impose a trade-off regarding accuracy, especially when compared to ANN and SVM (MING; QU; BERTINI, 2019; DU; LIU; HU, 2019). Random Forest (RF) arranges multiple randomly built DTs (ensemble of DTs) capable of producing accurate results (BREIMAN, 2001; BIAU; SCORNET, 2016). Moreover, it is notably useful for data involving many variables (a.k.a. multivariate or multidimensional) (BREIMAN, 2001; BIAU; SCORNET, 2016). RF is challenging regarding interpretation for ensembling several DTs, originating hundreds or thousands of logic rules, which may be considered even a black-box model (GUIDOTTI *et al.*, 2018a; ADADI; BERRADA, 2018). Logic rules can be represented in visual metaphors (MING; QU; BERTINI, 2019) rather than the standard text format approach (FREITAS, 2014; GUIDOTTI *et al.*, 2018a). Visual representations can be instrumental in supporting model interpretability (RIBEIRO; SINGH; GUESTRIN, 2016; ENDERT *et al.*, 2017), which has been the purpose of Visual Analytics (VA) tools (MING; QU; BERTINI, 2019; ZHAO *et al.*, 2019; DI CASTRO; BERTINI, 2019), where model visualization is a key aspect (KEIM *et al.*, 2010; SACHA *et al.*, 2014). Although VA approaches have been independently proposed for logic rules visu-

would employ explanations to understand how/why the model came to its decision (BRONIATOWSKI, 2021). **Nevertheless, in this thesis the term "interpretability" is used in the broad general sense, encompassing "explainability", as in many works in the literature** (CARVALHO; PEREIRA; CARDOSO, 2019)**.**

[2] The Brazilian General Data Protection Law (LGPD) – Federal Law nº. 13.709/2018 also requires explanations about automated decisions concerning individuals (BRASIL, 2018; BRASIL, 2019).

[3] Despite being useful, other approaches for model interpretability can be employed rather than logic rules, such as features' contribution and used image area (RIBEIRO; SINGH; GUESTRIN, 2016).

alization (MING; QU; BERTINI, 2019) and RF interpretability (ZHAO *et al.*, 2019), they may experience scalability issues on models' global and local explanations. Hence, both concise and dynamic visual representations may lead to proper RF model explanations.

Apart from predictive capability, classification models can also serve as descriptive tools, distinguishing data instances among different classes (TAN; STEINBACH; KUMAR, 2005). Once interpretable, such models are suitable for multivariate data explanation, as recently proposed by VA solutions arranging visual representations of SVM and ANN models (GLEICHER, 2013; KNITTEL *et al.*, 2020). Nevertheless, concepts like Emerging Patterns arrange descriptive logic rules (NOVAK; LAVRAC; WEBB, 2009), where a specific type, called Jumping Emerging Pattern (JEP), provides high discriminative power between classes (KANE; CUISSART; CRÉMILLEUX, 2015; GARCíA-VICO *et al.*, 2018). DTs can be employed to obtain such patterns (descriptive rules) (NOVAK; LAVRAC; WEBB, 2009; GARCíA-VICO *et al.*, 2018; LOYOLA-GONZÁLEZ; MEDINA-PÉREZ; CHOO, 2020), and RF models have proven to produce diversified high-quality patterns (GARCíA-BORROTO; MARTíNEZ-TRINIDAD; CARRASCO-OCHOA, 2015; LOYOLA-GONZáLEZ *et al.*, 2019; LOYOLA-GONZÁLEZ; MEDINA-PÉREZ; CHOO, 2020). The potential of descriptive logic rules lies in understanding the phenomenon represented by data (NOVAK; LAVRAC; WEBB, 2009; GARCíA-VICO *et al.*, 2018). Data explanation approaches based on classification models visual representations support knowledge generation from complex data (KNITTEL *et al.*, 2020), which is the VA's primary goal involving insights and hypotheses (KEIM *et al.*, 2010; SACHA *et al.*, 2014). Although inspiring, the proposed VA solutions for descriptive purposes make use of black-box models (SVM and ANN) (GLEICHER, 2013; KNITTEL *et al.*, 2020), requiring constraints for reaching interpretability. On the other hand, JEPs are naturally interpretable and descriptive, as well as suitable for visualization methods (NOVAK; LAVRAC; WEBB, 2009).

In summary, since classification models can be used as predictive and descriptive tools (TAN; STEINBACH; KUMAR, 2005), model interpretability may involve visual explanations for model (MING; QU; BERTINI, 2019; ZHAO *et al.*, 2019) and data (GLEICHER, 2013; KNITTEL *et al.*, 2020) understanding. Logic rules are helpful in model interpretability solutions (LAKKARAJU; BACH; LESKOVEC, 2016; GUIDOTTI *et al.*, 2018b) and can be extracted from DTs (NOVAK; LAVRAC; WEBB, 2009; MIRANDA; SARDINHA; CERRI, 2021). The RF approach arranges several DTs, producing accurate results and handling datasets with many variables (BREIMAN, 2001; BIAU; SCORNET, 2016). Therefore, this doctoral thesis proposes visualization methods supporting RF interpretability to leverage its predictive and descriptive capabilities in model and data explanations.

1.2 Goals

Within the context set previously, this doctoral project has as goals the following:

Create visualization methods arranging logic rules to support Random Forest classification models' interpretability covering predictive and descriptive purposes. The methods must provide model-centered visual representations for reaching global and local explanations for model overview and decisions reasoning, and data-centered visual representations for achieving multivariate data explanation via data insights and knowledge generation. In other words, the primary goals are model and data (predictive and descriptive) explanations visualizing logic rules extracted from Random Forest.

Therefore, the hypothesis is given by:

Logic rules visualizations can support explanations of multivariate data and Random Forest models' overall logic and outcomes.

The next section highlights the results accomplished.

1.3 Results

In order to fulfill the defined goals, two VA methods were created where logic rules extracted from RF models are visualized using a matrix-like visual metaphor. The ExMatrix (Explainable Matrix) method provides model-centered visual representations, investigating the knowledge learned by the model and auditing instance classification process (predictive purpose) (POPOLIN NETO; PAULOVICH, 2021). The VAX (multiVariate dAta eXplanation) method supports data-centered visual representations, allowing for multivariate data explanation automated data insights (descriptive purpose). The visual metaphor employed in ExMatrix and VAX is based on matrix visualization guidelines (CHEN *et al.*, 2004; WU; TZENG; CHEN, 2008), with rules displayed as rows, features (variables) as columns, and rules predicates as cells. A flowchart-based summarization is found in Figure 30 of Appendix C, presenting both methods' inputs, processes, and outputs.

The ExMatrix global explanations are able to manage a considerable number of rules at once, and local explanations can present used rules or smallest changes rules for a specific data instance classification. Although ExMatrix was conceived focusing on RF models, it can also interpret a single DT. Moreover, the ExMatrix has been applied in analytical chemistry, where data are acquired from samples of liquid or gaseous solutions with

analyte concentrations or distinguishable factors (RIUL JúNIOR *et al.*, 2010). Logic rules inferred from such data by DT models produce a calibration space employing multiple features (dimensions) and, more importantly, interpretable by ExMatrix. This calibration is named Multidimensional Calibration Space (MCS) (POPOLIN NETO *et al.*, 2021).

Additionally to descriptive logic rules visualization, the VAX method integrates JEPs and Dimensional Reduction (DR) techniques (NONATO; AUPETIT, 2019) originating maps for similarity in data instances. The visualization and integration of similarity maps and JEPs lead to automated insights (LAW; ENDERT; STASKO, 2020) used for multivariate data explanation. Clusters and outliers may be revealed in such maps, and JEPs visualization can be employed for further investigations. VAX was used in the analysis of two real-world datasets. One regarding the 2016 US presidential election and the other the 2019 world happiness report produced by the Sustainable Development Solutions Network.

The contributions of this doctoral project are distributed across 6 articles and 1 book chapter either published, accepted, or submitted (preprint). Appendix A provides a more detailed list of publications and submissions, and source code (majority) can be found at <https://gitlab.com/popolinneto/exmatrix>, being also available as code package at <https://pypi.org/project/exmatrix/>.

1.4 Organization

This thesis is organized as a collection of the main articles produced (complete list in Appendix A). Except Chapters 1 for introduction and 5 for conclusion, Chapters 2, 3, and 4 are full papers either published or submitted to journals. The chapters are:

- Chapter 2: POPOLIN NETO, M.; PAULOVICH, F. V. Explainable matrix - visualization for global and local interpretability of random forest classification ensembles. IEEE Transactions on Visualization and Computer Graphics, v. 27, n. 2, p. 1427–1437, 2021. Available: <https://doi.org/10.1109/TVCG.2020.3030354>.

  For models (predictive) explanations, the ExMatrix method is proposed, providing meaningful global and local visual representations.

- Chapter 3: POPOLIN NETO, M.; SOARES, A. C.; OLIVEIRA, O. N.; PAULOVICH, F. V. Machine learning used to create a multidimensional calibration space for sensing and biosensing data. Bulletin of the Chemical Society of Japan, v. 94, n. 5, p. 1553–1562, 2021. Available: <https://doi.org/10.1246/bcsj.20200359>.

  An ExMatrix application in analytical chemistry. The method is employed in Impedance Spectroscopy data obtained from sensing units.

- Chapter 4: POPOLIN NETO, M.; PAULOVICH, F. V. Multivariate Data Explanation by Jumping Emerging Patterns Visualization. arXiv preprint arXiv:2106.11112. 2021.

  For data (descriptive) explanations, the VAX method is proposed, generating automated data insights visualizing JEPs and data instances similarity maps.

CHAPTER

2

EXPLAINABLE MATRIX – EXMATRIX

This chapter (paper *Explainable matrix - visualization for global and local interpretability of random forest classification ensembles* [1]) presents ExMatrix, a VA method for models (predictive) explanations, allowing RF models' interpretability via global and local visual representations. The ExMatrix global explanations provide reasoning about the knowledge learned by the RF model, while local explanations support the classification process reasoning. The ExMatrix method arranges logic rules into a matrix-like visual metaphor. The latter uses matrix visualization guidelines and is more scalable than literature methods, handling a substantial number of logic rules at once, an essential issue for RF models' interpretability.

Abstract: Over the past decades, classification models have proven to be essential machine learning tools given their potential and applicability in various domains. In these years, the north of the majority of the researchers had been to improve quantitative metrics, notwithstanding the lack of information about models' decisions such metrics convey. This paradigm has recently shifted, and strategies beyond tables and numbers to assist in interpreting models' decisions are increasing in importance. Part of this trend, visualization techniques have been extensively used to support classification models' interpretability, with a significant focus on rule-based models. Despite the advances, the existing approaches present limitations in terms of visual scalability, and the visualization of large and complex models, such as the ones produced by the Random Forest (RF) technique, remains a challenge. In this paper, we propose *Explainable Matrix (ExMatrix)*, a novel visualization method for RF interpretability that can handle models with massive quantities of rules. It employs a simple yet powerful matrix-like visual metaphor, where rows are rules, columns are features, and cells are rules predicates, enabling the analysis

[1] POPOLIN NETO, M.; PAULOVICH, F. V. Explainable matrix - visualization for global and local interpretability of random forest classification ensembles. IEEE Transactions on Visualization and Computer Graphics, v. 27, n. 2, p. 1427–1437, 2021. Available: <https://-doi.org/10.1109/TVCG.2020.3030354>.

of entire models and auditing classification results. ExMatrix applicability is confirmed via different examples, showing how it can be used in practice to promote RF models interpretability.

2.1 Introduction

Imagine a machine learning classification model for cancer prediction with 99% accuracy, prognosticating positive breast cancer for a specific patient. Even though we are far from reaching such level of precision, we (researchers, companies, among others) have been trying to convince the general public to trust classification models, using the premise that machines are more precise than humans (CRUZ; WISHART, 2006). However, in most cases, yes or no are not satisfactory answers. A doctor or patient inevitably may want to know why positive? What are the determinants of the outcome? What are the changes in patient records that may lead to a different prediction? Although standard instruments for building classification models, quantitative metrics such as accuracy and error cannot tell much about the model prediction, failing to provide detailed information to support understanding (LIU *et al.*, 2018).

We are not advocating against machine learning classification models, since there is no questioning about their potential and applicability in various domains (ENDERT *et al.*, 2017; BUTLER *et al.*, 2018). The point is the acute need to go beyond tables and numbers to understand models' decisions, increasing trust in the produced results. Typically, this is called model interpretability and has become the concern of many researchers in recent years (YANG; DU; HU, 2019; CARVALHO; PEREIRA; CARDOSO, 2019). Model interpretability is an open challenge and opportunity for researchers (ENDERT *et al.*, 2017) and also a government concern, as the *European General Data Protection Regulation* requires explanations about automated decisions regarding individuals (LIU; WANG; MATWIN, 2018; CARVALHO; PEREIRA; CARDOSO, 2019; GUIDOTTI *et al.*, 2018b).

Model interpretability strategies are typically classified as global or local approaches. Global techniques aim at explaining entire models, while the local ones give support for understanding the reasons for the classification of a single instance (DU; LIU; HU, 2019; CARVALHO; PEREIRA; CARDOSO, 2019). In both cases, interpretability can be attained using inherent interpretable models such as Decision Trees, Rules Sets, and Decision Tables (KOHAVI, 1995a), or through surrogates, where black-box models, like Artificial Neural Networks or Support Vector Machines, are approximated by rule-based interpretable models (GUIDOTTI *et al.*, 2018b; CARVALHO; PEREIRA; CARDOSO, 2019). The key idea is to transform models into logic rules, using them as a mechanism to enable the interpretation of a model and its decisions (LEI *et al.*, 2018; GUIDOTTI *et al.*, 2018a; DI CASTRO; BERTINI, 2019; MING; QU; BERTINI, 2019; RIBEIRO; SINGH;

GUESTRIN, 2018).

Recently, visualization techniques have been used to empower the process of interpreting rule-based classification models, particularly Decision Tree models (DI CASTRO; BERTINI, 2019; ZHAO *et al.*, 2019; VAN DEN ELZEN; VAN WIJK, 2011; SCHULZ, 2011). In this case, given the inherent nature of these models, the usual adopted visual metaphors focus on revealing tree structures, such as the node-link diagrams (GRAHAM; KENNEDY, 2010; ZHAO *et al.*, 2019; MING; QU; BERTINI, 2019). However, node-link structures are limited when representing logic rules (FREITAS, 2014; HUYSMANS *et al.*, 2011; LIMA; MUES; BAESENS, 2009), and present scalability issues, supporting only small models with few rules (GRAHAM; KENNEDY, 2010; SCHULZ; HADLAK; SCHUMANN, 2011; ZHAO *et al.*, 2019). Matrix-like visual metaphors have been used (MING; QU; BERTINI, 2019; DI CASTRO; BERTINI, 2019) as an alternative, but visual scalability limitations still exist, and large and complex models cannot be adequately visualized, such as the Random Forests (BREIMAN, 2001; BIAU; SCORNET, 2016). Among rule-based models, Random Forests is one of the most popular techniques given their simplicity of use and competitive results (BIAU; SCORNET, 2016). However, they are very complex entities for visualization since multiple Decision Trees compose a model, and, although attempts have been made to overcome such a hurdle (ZHAO *et al.*, 2019), the visualization of entire models is still an open challenge.

In this paper, we propose *Explainable Matrix (ExMatrix)*, a novel method for Random Forest (RF) interpretability based on the visual representations of logic rules. ExMatrix supports global and local explanations of RF models enabling tasks that involve the overview of models and the auditing of classification processes. The key idea is to explore logic rules by demand using matrix visualizations, where rows are rules, columns are features, and cells are rules predicates. ExMatrix allows reasoning on a considerable number of rules at once, helping users to build insights by employing different orderings of rules/rows and features/columns, not only supporting the analysis of subsets of rules used on a particular prediction but also the minimum changes at instance level that may change a prediction. Visual scalability is addressed in our solution using a simple yet powerful compact representation that allows for overviewing entire RF models while also enables focusing on specific parts for details on-demand. In summary, the main contributions of this paper are:

- A new matrix-like visual metaphor that supports the visualization of RF models;
- A strategy for Global interpretation of large and complex RF models supporting model overview and details on-demand; and
- A strategy to promote Local interpretation of RF models, supporting auditing models' decisions.

2.2 Related Work

Typically, visualization techniques aid in classification tasks in two different ways. One is on supporting parametrization and labeling processes aiming to improve model performance (ANKERST *et al.*, 1999; TEOH; MA, 2003; DO, 2007; VAN DEN ELZEN; VAN WIJK, 2011; TALBOT *et al.*, 2009; HöFERLIN *et al.*, 2012; LEE; JOHNSON; CHENG, 2016; LIU *et al.*, 2018). The other is on understanding the model as a whole or the reasons for a particular prediction. In this paper, our focus is on the latter group, usually named model interpretability.

Interpretability techniques can be divided into pre-model, in-model, or post-model strategies, regarding support to understand classification results before, during, or after the model construction (CARVALHO; PEREIRA; CARDOSO, 2019). Pre-model strategies usually give support to data exploration and understanding before model creation (PAIVA *et al.*, 2015; CHOO *et al.*, 2010; MIGUT; WORRING, 2010; CARVALHO; PEREIRA; CARDOSO, 2019). In-model strategies involve the interpretation of intrinsically interpretable models, such as Decision Trees, and post-model strategies concerns interpretability of complete built models, and they can be model-specific (RAUBER *et al.*, 2017; WU *et al.*, 2018) or model-agnostic (DI CASTRO; BERTINI, 2019; MING; QU; BERTINI, 2019; RIBEIRO; SINGH; GUESTRIN, 2018; GUIDOTTI *et al.*, 2018a). Both in-model and post-model approaches aim to provide interpretability by producing global and/or local explanations (DU; LIU; HU, 2019).

2.2.1 Global Explanation

Global explanation techniques produce overviews of classification models aiming at improving users' trust in the model (RIBEIRO; SINGH; GUESTRIN, 2016). For inherently interpretable models, the global explanation is attained through visual representations of the entire model. For more complex non-interpretable black-box models, such as Artificial Neural Networks or Support Vector Machines, interpretability can be attained through a surrogate process where such models are approximated by interpretable ones (MING; QU; BERTINI, 2019; DI CASTRO; BERTINI, 2019; HALL, 2018). Decision Trees (BREIMAN *et al.*, 1984; TAN; STEINBACH; KUMAR, 2005; LOH, 2014) are commonly used as surrogate models (DI CASTRO; BERTINI, 2019; HALL, 2018), and whether a surrogate or a classification model per se, the most common visual metaphor for global explanation is the node-link (MING; QU; BERTINI, 2019; ZHAO *et al.*, 2019), such as the BaobaView technique (VAN DEN ELZEN; VAN WIJK, 2011). The node-link metaphor's problem is scalability (GRAHAM; KENNEDY, 2010; SCHULZ; HADLAK; SCHUMANN, 2011; ZHAO *et al.*, 2019), mainly when it is used to create visual representations for Random Forests, limiting the model to be small in number of trees (STIGLIC *et al.*, 2006). Creating a scalable visual representation for an entire Random Forest model,

presenting all decision paths (root node to leaf node paths), remains a challenge even with a considerably small number of trees (LIU *et al.*, 2018).

Although the node-link metaphor is the straightforward representation for Decision Trees, logic rules extracted from decision paths have also been used to help on interpretation (LIMA; MUES; BAESENS, 2009). Indeed, disjoint rules have shown to be more suitable for user interpretation than hierarchical representations (LAKKARAJU; BACH; LESKOVEC, 2016), and a user test comparing the node-link metaphor with different logic rule representations, showed that Decision Tables (KOHAVI, 1995a) (rules organized into tables) offers better comprehensibility properties (FREITAS, 2014; HUYSMANS *et al.*, 2011). Nonetheless, this strategy uses text for representing rules having as drawback model size (FREITAS, 2014). Similarly to Decision Tables, our method does not lean on the hierarchical property of Decision Trees. However, instead of using text to represent logic rules, we used a matrix-like visual metaphor, where rows are rules, columns are features, and cells are rules predicates, capable of displaying a much larger number of rules than the textual representations.

The idea of using a matrix metaphor to present rules is not new (DI CASTRO; BERTINI, 2019; MING; QU; BERTINI, 2019), and it has been used before by the RuleMatrix technique (MING; QU; BERTINI, 2019). RuleMatrix is a model-agnostic approach to induce logic rules from black-box models, presenting rules in rows, features in columns, and predicates in cells using histograms. As data histograms require a certain display space to support human cognition, the number of rules displayed at once is reduced. Therefore, not being able to present entire or even parts of Random Forest models (notice that their focus is the visualization of surrogate rules, not models). Our approach also uses a matrix metaphor; however, we employ a simpler icon (colored rectangular shape) for the matrix cells, mapping different rule properties (e.g., predicates, class, and others), considerably improving the scalability of the visual representation. Besides the recognized scalability of matrix visualization and custom cells (ALSALLAKH *et al.*, 2014; BEHRISCH *et al.*, 2016; ALPER *et al.*, 2013), rows and columns order is an important principle (WU; TZENG; CHEN, 2008; CHEN; SINICA; TAIPEI, 2002; CHEN *et al.*, 2004; BEHRISCH *et al.*, 2016), and in our approach rules and features can be organized using different criteria, promoting analytical tasks not supported by the RuleMatrix, such as the holistic analysis of Random Forest models through complete overviews. Worthy mentioning that different from usual matrix visual metaphors for trees and graphs that focus on nodes (BEHRISCH *et al.*, 2016; GRAHAM; KENNEDY, 2010), our approach focus on decision paths, which is the object of analysis on Decision Trees (LIMA; MUES; BAESENS, 2009; FREITAS, 2014; HUYSMANS *et al.*, 2011), so representing a different concept.

2.2.2 Local Explanation

Unlike the model overview of global explanations, local explanation techniques focus on a particular instance classification result (RIBEIRO; SINGH; GUESTRIN, 2018; ZHAO *et al.*, 2019), aiming to improve users' trust in the prediction (RIBEIRO; SINGH; GUESTRIN, 2016). As in global strategies, local explanations can be provided using inherently interpretable models or using surrogates of black-boxes (RIBEIRO; SINGH; GUESTRIN, 2018; GUIDOTTI *et al.*, 2018a; STRUMBELJ; KONONENKO, 2010). In general, local explanations are constructed using the logic rule applied to classify the instance along with its properties (e.g., coverage, certainty, and fidelity), providing additional information for prediction reasoning (MING; QU; BERTINI, 2019; LAKKARAJU; BACH; LESKOVEC, 2016).

One example of a visualization technique that supports local explanation is the RuleMatrix (MING; QU; BERTINI, 2019). RuleMatrix was applied to support the analysis of surrogate logic rules of Artificial Neural Networks and Support Vector Machine models. Local explanations are taken by analyzing the employed rules, observing the instance features values coupled with rules predicates and properties. Another interactive system closely related to our method is the iForest (ZHAO *et al.*, 2019), combining techniques for Random Forest models local explanations. The iForest system focuses on binary classification problems, and for each instance, it allows the exploration of decision paths from Decision Trees using multidimensional projection techniques. A summarized decision path is built and displayed as a node-link diagram by selecting decision paths of interest (circles in the projection).

As discussed before, node-link diagrams are prone to present scalability issues. Although iForest reduces the associate issues by summarizing similar decision paths, it fails to present the overall picture of Random Forest classification models' voting committees. Our approach shows the voting committee by displaying all rules (decision paths) used by a model when classifying a particular instance, allowing insights into the feature space and class association by ordering rules and features in different ways. Also, our approach can be applied to multi-class problems, not only binary classifications, and, as iForest, it supports counterfactual analysis (GOMEZ *et al.*, 2020; LIAO; GRUEN; MILLER, 2020) by displaying the rules that, with the smallest changes, may cause the instance under analysis to switch its final classification.

2.3 ExMatrix

In this section, we present *Explainable Matrix (ExMatrix)*, a visualization method to support Random Forest global and local interpretability.

2.3.1 Overview

To create a classifier, classification techniques take a labelled dataset $X = \{x_1, ..., x_N\}$ with N instances and their classes $Y = \{y_1, ..., y_N\}$, where $y_n \in C = \{c_1, ..., c_J \geq 2\}$ and x_n consists of a vector $x_n = [x_n^1, ..., x_n^M]$ with M features $F = \{f_1, ..., f_M\}$ values, and build a mathematical model to compute a class y_n when new instances $x_n \notin X$ are given as input. In this process, X is usually split into two different sets, one X_{train} to build the model and one X_{test} to test it. The existing techniques have adopted many different strategies to build a classifier. The Random Forest (RF) is an ensemble approach that creates multiple Decision Tree (DT) models $DT_1, ..., DT_K$ of randomly selected subsets of features and/or training instances, and combines them to classify an instance using a voting strategy (TAN; STEINBACH; KUMAR, 2005; BREIMAN *et al.*, 1984; BREIMAN, 2001; BIAU; SCORNET, 2016). Therefore, a RF model is a collection of decision paths, belonging to different DTs, combined to classify an instance.

Aiming at supporting users to examine RF models and enable results audit, ExMatrix presents the decision paths extracted from DTs as logic rules using a matrix visual metaphor, supporting global and local explanations. ExMatrix arranges logic rules $R = \{r_1, ..., r_Z\}$ as rows, features $F = \{f_1, ..., f_M\}$ as columns, and rule predicates $r_z = [r_z^1, ..., r_z^M]$ as cells, inspired by similar user-friendly and powerful matrix-like solutions (WU; TZENG; CHEN, 2008; CHEN; SINICA; TAIPEI, 2002; CHEN *et al.*, 2004). Figure 1 depicts our method overview, composed mainly of two steps. One involving the vector rules extraction, where all decision paths of each DT_k in the RF model are converted into vectors, and a second one where these vectors are displayed using a matrix metaphor to support explanations. The next sections detail these steps, starting with the vector rule extraction process.

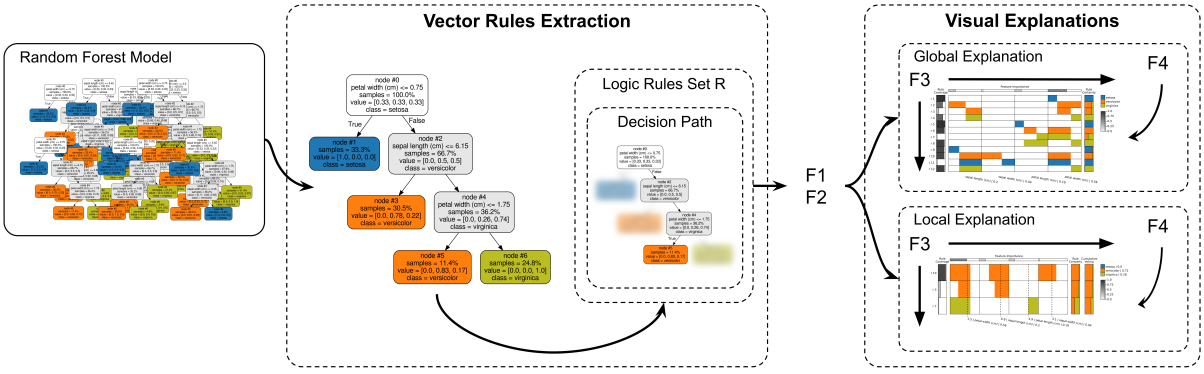


Figure 1 – *Explainable Matrix (ExMatrix)* overview. ExMatrix is composed of two main steps. In the first, decision paths of the RF model under analysis are converted into logic rules. Then, in the second, these rules are displayed using a matrix metaphor to support global and local explanations.

2.3.2 Vector Rules Extraction

As mentioned, ExMatrix first step involves the transformation of each decision path, the path from a DT root node to a leaf node, into a vector rule representing the features' intervals for which the decision path is true. The resulting vectors present dimensionality equal to the number of features M, with coordinates composed of pairs representing the features' minimum and maximum interval values. In more mathematical terms, this process transforms, for every tree DT_k, each decision path $p_{(o,d)}$ (from the root node o to the leaf node d) into a disjoint logic rule (vector) r_z. Let $p_{(o,d)} = \{(f_o \bigotimes \theta_o), ..., (f_v \bigotimes \theta_v)\}$ denotes a decision path, where each node i contains a logic test $\bigotimes \in \{\text{“} \leq \text{”}, \text{“} > \text{”}\}$ bisecting the feature f_i using a threshold $\theta_i \in \mathbb{R}$, and that the node v is the parent of the leaf node d (ZHAO *et al.*, 2019). To convert $p_{(o,d)}$ into a vector rule $r_z = [r_z^1, ..., r_z^M]$, each element $r_z^m = \{\alpha_z^m, \beta_z^m\}$ is computed representing the intervals covered by $p_{(o,d)}$ if and only if $f^m \in p_{(o,d)}$. Otherwise, $r_z^m = \varnothing$. Considering $f^m \in p_{(o,d)}$, the lower limit α_z^m is the maximum $\theta_i \in p_{(o,d)}$ for the feature f^m and logic test $\bigotimes = \text{“} > \text{”}$. If such combination does not exist in $p_{(o,d)}$, α_z^m is set to the minimum value of feature f^m in X, that is

$$\alpha_z^m = \begin{cases} \max(\theta_i | f_i = f^m, \bigotimes = \text{“} > \text{”}) & \text{if } (f_i = f^m > \theta_i) \in p_{(o,d)} \\ \min(x^m | x^m \in X) & \text{Otherwise.} \end{cases} \tag{2.1}$$

Similarly, the upper limit β_z^m is the minimum $\theta_i \in p_{(o,d)}$ for the feature f^m and logic test $\bigotimes = \text{“} \leq \text{”}$. If such combination does not exist in $p_{(o,d)}$, β_z^m is set to the maximum value of feature f^m in X, that is

$$\beta_z^m = \begin{cases} \min(\theta_i | f_i = f^m, \bigotimes = \text{“} \leq \text{”}) & \text{if } (f_i = f^m \leq \theta_i) \in p_{(o,d)} \\ \max(x^m | x^m \in X) & \text{Otherwise.} \end{cases} \tag{2.2}$$

Beyond predicates, three other properties are extracted for each logic rule r_z, being certainty, class, and coverage. The rule certainty r_z^{cert} is a vector of probabilities for each class $c_j \in C$, obtained from the decision path (leaf node value). The rule class r_z^{class} is the $c_j \in C$ with the highest probability on the rule certainty r_z^{cert}. The rule coverage r_z^{cov} is [2] the number of instances in X_{train} of class r_z^{class} for which r_z is valid divided by the total number of instance of r_z^{class} in X_{train}. The vector rules extraction process results in a set of disjoint logic rules $R = \{r_1, ..., r_Z\}$, where each rule r_z classifies an instance x_n belonging to class r_z^{class} if its predicates $r_z = [r_z^1, ..., r_z^M]$ are all true for the feature values in x_n.

As an example of vector rule extraction, consider the zoomed DT in Figure 1 from a RF for the Iris dataset (FISHER, 1936), with 150 instances in three classes $C = \{setosa, versicolor, virginica\}$ and 4 features $F = \{sepal\ length\ ,\ sepal\ width\ ,\ petal\ length\ ,\ petal\ width\}$. From this tree, the decision

[2] The rule coverage formulation used here equals rule support definition.

path $p_{(\#0,\#5)}$ is transformed into the vector rule $r_3 = [\{6.15, 7.9\}, \varnothing, \varnothing, \{0.75, 1.75\}]$ with $r_3^{class} = versicolor$, since rule certainty equals to $r_3^{cert} = [0.0, 0.83, 0.17]$ (leaf node #5 value), indicating that r_3 is valid for 83% of the *versicolor* instances and 17% of *virginica* instances in X_{train}. The rule coverage $r_3^{cov} = 0.28$ as r_3 is valid for 10 out of 35 *versicolor* instances in X_{train}.

2.3.3 Visual Explanations

Once the vector rules are extracted, they are used to create the matrix visual representations for global and local interpretation. To guide our design process we adopted the iForest design goals (G1 - G3) (ZHAO *et al.*, 2019) and the RuleMatrix target questions (Q1 - Q4) (MING; QU; BERTINI, 2019) summarized on Table 1. These goals and questions consider classification model reasoning beyond performance measures (e.g., accuracy and error), focusing on the model internals. For global explanations, where the focus is an overview of a model, ExMatrix displays feature space ranges and class associations (**G1** and **Q1**), and how reliable these associations are (**Q2**). For local explanations, where the focus is the classification of a particular instance x_n, ExMatrix allows the analysis of x_n values and features space ranges that resulted into the assigned class y_n (**G2** and **Q3**), and the inspection of the changes in x_n that may lead to a different classification (**G3** and **Q4**).

Table 1 – ExMatrix design goals.

Global	Local
G1 Reveal the relationships between features and predictions (ZHAO *et al.*, 2019).	**G2** Uncover the underlying working mechanisms (ZHAO *et al.*, 2019).
Q1 What knowledge has the model learned? (MING; QU; BERTINI, 2019).	**G3** Provide case-based reasoning (ZHAO *et al.*, 2019).
Q2 How certain is the model for each piece of knowledge? (MING; QU; BERTINI, 2019).	**Q3** What knowledge does the model utilize to make a prediction? (MING; QU; BERTINI, 2019).
	Q4 When and where is the model likely to fail? (MING; QU; BERTINI, 2019).

ExMatrix implements these goals using a set of four functions:

F1 – Rules of Interest. Function $R' = f_{rules}(R, \ldots)$ returns a subset of rules of interest $R' \subseteq R$. For global explanations $f_{rules}(R, \ldots)$ returns the entire vector rules set $R' = R$ or a subset $R' \subset R$ defined by the user, while for local explanations $f_{rules}(R, x_n, \ldots)$ returns a subset $R' \subset R$ related to a given instance x_n.

F2 – Features of Interest. Function $F' = f_{features}(R', \ldots)$ returns features of interest $F' \subseteq F$ considering a set of rules of interest R'. For global explanations $f_{features}(R', ...)$ returns all features used by the RF model, whereas for local explanations $f_{features}(R', x_n, ...)$ returns the features used to classify a given instance x_n.

F3 – Ordering. Function $L' = f_{ordering}(L, criteria, \ldots)$ returns an ordered version L' of a input set L following a given criterion, where L can be rules R' or features F'. This is used for both global and local explanations aiming at revealing patterns, a key property in matrix-like visualizations (WU; TZENG; CHEN, 2008; CHEN; SINICA; TAIPEI, 2002; CHEN *et al.*, 2004), where rows and columns can be sorted in different ways, following, for instance, elements properties (KRAUSE *et al.*, 2017) or similarity measures (CHOI; CHA, 2010; TZENG; WU; CHEN, 2009; BEHRISCH *et al.*, 2016; FUJIWARA; KWON; MA, 2019).

F4 – Predicate Icon. Function $f_{icon}(r_z^m, \ldots)$ returns a cell icon (visual element) for a predicate r_z^m of the rule r_z and feature f_m. For global and local explanations, a cell icon is a color-filled rectangular element, allowing our visual metaphor to display a substantial number of logic rules at once. This is an important aspect since matrix-like visualizations can display a massive number of rows and columns relying on such icons not requiring many pixels (CHEN *et al.*, 2004).

Figure 1 shows how these four functions are used in conjunction to build the visual representations for global and local interpretation. Functions **F1** and **F2** are used to select and map rules and features of interest. Function **F3** is used to change the rows and columns order to help in finding interesting patterns, and function **F4** is used to derive the predicate icon that can vary depending on the type of interpretation task (global or local). In the next section, we detail how these functions are used to build ExMatrix visual representations.

2.3.3.1 Global Explanation (GE)

Our first visual representation is an overview of RF models called *Global Explanation (GE)*. To build this matrix, $R' = f_{rules}(R, \ldots)$ returns all logic rules R or a subset $R' \subset R$ defined by the user, and $F' = f_{features}(R', \ldots)$ returns all features used by at least one rule $r_z \in R'$. As previously explained, matrix rows represent logic rules, columns features, and cells rules predicates (icons). Rows and columns can be ordered using different criteria ($L' = f_{ordering}(L, criteria, \ldots)$). The rows can be ordered by rules' coverage, certainty, class & coverage, and class & certainty, while columns can be ordered by feature importance, calculated using the Mean Decrease Impurity (MDI) (BREIMAN, 2002).

For the ExMatrix GE visualization, the matrix cell icon representing the rule predicate r_z^m consists of a rectangle ($f_{icon}(r_z^m,\ldots)$) colored according to the rule class r_z^{class}, positioned and sized inside the matrix cell proportional to the predicate limits $\{\alpha_z^m,\beta_z^m\}$, where the left side of the matrix cell represents the value $\min(x^m|x^m \in X)$ and the right side $\max(x^m|x^m \in X)$ (goals **G1** and **Q1**). The cell background not covered by the predicate limits can be either white or be filled using a less saturated color. If no predicate is present, the matrix cell is left blank.

Rules and features properties are also exhibited using additional rows and columns (goal **Q2**). The rule coverage r_z^{cov} is shown using an extra column on the left side of the table with cells' color (grayscale) and fill proportional to the coverage. The rules certainty r_z^{cert} is shown in an extra column in the right side of the table with cells split into colored rectangles with sizes proportional to the probability of the different classes. The feature importance is shown in an extra row on the top of the table with cells' color (grayscale) and fill proportional to the importance. Also, labels are added below the matrix, combining feature name and importance value.

Figure 2 presents a ExMatrix GE visualization of a RF model for the Iris dataset with 3 trees with maximum depth equals to 3. In this example, the rows (rules) are ordered by extraction order, and the columns (features) follows the dataset order. The logic rule $r_3 = [\{6.15, 7.9\}, \varnothing, \varnothing, \{0.75, 1.75\}]$ extracted from the decision path $p_{(\#0,\#5)}$ (see Figure 1) is zoomed in. It is colored in orange since this is the color we assign to the *versicolor* class and it classifies 83% of the training instances as belonging to this class (17% belonging to *virginica*). Also, its coverage is $r_3^{cov} = 0.28$.

2.3.3.2 Local Explanation Showing the Used Rules (LE/UR)

The second visual representation, called *Local Explanation Showing the Used Rules (LE/UR)*, is a matrix to help in auditing the results of a RF model providing explanations for the classification of a given instance x_n. In this process, $R' = f_{rules}(R, x_n)$ returns all logic rules used by the model to classify x_n (goals **G2** and **Q3**). As in the ExMatrix GE visualization, $F' = f_{features}(R')$ returns all features used by logic rules R', $f_{icon}(r_z^m, X)$ returns a cell icon representing predicates limits, and $f_{ordering}(L, criteria)$ can order rules R' by coverage, certainty, class & coverage, and class & certainty, and features F' by importance.

In addition to the coverage and certainty columns, in the ExMatrix LE/UR visualization, an extra column is added to represent the committee's cumulative voting. In this column, the cell at the i^{th} row is split into colored rectangles with sizes proportional to the different classes' probability considering only the first i rules. In this way, given a matrix order (e.g., based on the rule coverage), it is possible to see from what rule the committee reaches a decision that is not changed even if the remaining rules are used to

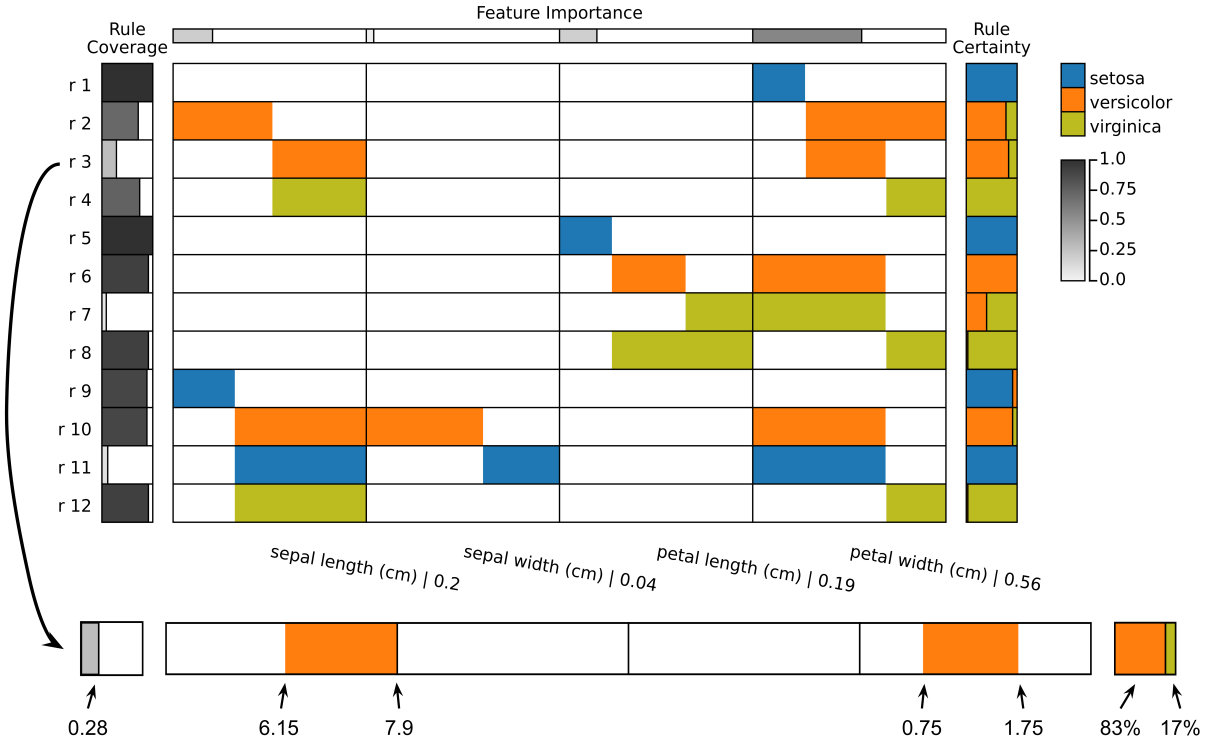


Figure 2 – ExMatrix Global Explanation (GE) of a RF model for the Iris dataset containing 3 trees with maximum depth equal to 3. Rows represent logic rules, columns features, and matrix cells the predicates. Additional rows and columns are also used to represent rule coverage and certainty. One matrix row is highlighted to exemplify how the rules' information is transformed into icons.

classify x_n (indicated by a black line). Notice that this column's last cell always represents the committee's final decision regardless of rule ordering.

Figure 3 presents the ExMatrix LE/UR representation for instance $x_{13} = [6.9, 3.1, 4.9, 1.5]$. We use the same RF model of Figure 2 with 3 trees, so the RF committee uses 3 rules in the classification. The resulting matrix rows are ordered by rule coverage and columns by feature importance. The (optional) dashed line in each column indicates the values of the features of instance x_{13}. According to the committee, the probability of x_{13} to be *versicolor* is 72% and 28% to be *virginica*. Most of the *virginica* probability comes from the rule r_7, which holds the lowest coverage.

2.3.3.3 Local Explanation Showing Smallest Changes (LE/SC)

Our final matrix representation, called *Local Explanation Showing Smallest Changes (LE/SC)*, is also designed to support results audit when classifying a given instance x_n. In this visualization, for each DT_k in the RF model, we display the rule requiring the smallest change to make DT_k to change the classification of x_n. Let r_z be the rule extracted from DT_k that is true when classifying x_n, in this process we seek for the rule r_e from DT_k with $r_e^{class} \neq r_z^{class}$ that presents the minimum summation of changes to the

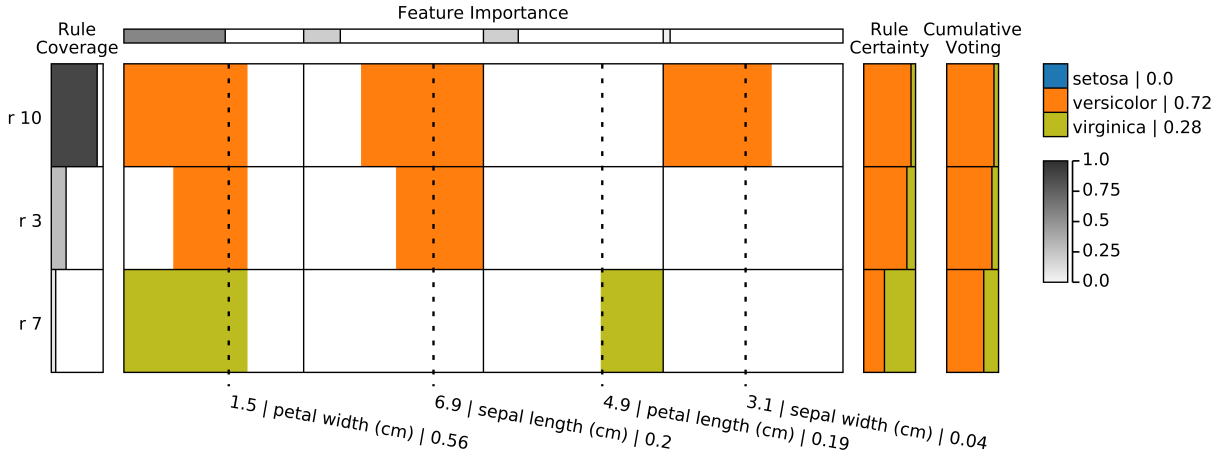


Figure 3 – ExMatrix Local Explanation showing the Used Rules (LE/UR) visualization. Three rules are used by the RF committee to classify a given instance as belonging to the *versicolor* class with 72% of probability. The dashed line in each column indicates the features' values of the instance.

values of x_n that makes r_e true and r_z false, that is, $\Delta_{(r_e,x_n)} = \sum_{m=1}^{M}(\Delta^m_{(r_e,x_n)})$, where

$$\Delta^m_{(r_e,x_n)} = \begin{cases} \frac{\min(|\alpha^m_e - x^m_n|, |\beta^m_e - x^m_n|)}{|\max(x^m | x^m \in X_{train}) - \min(x^m | x^m \in X_{train})|} & \text{if } x^m_n \notin \{\alpha^m_e, \beta^m_e\} \\ 0 & \text{Otherwise.} \end{cases} \tag{2.3}$$

Using this formulation, function $R' = f_{rules}(R, x_n)$ returns the list of logic rules that can potentially change the classification process outcome requiring the lowest changes (goals **G3** and **Q4**), and function $F' = f_{features}(R', x_n)$ returns the features used by the rules in R'. Beyond the ordering criteria for rules and features previously discussed, function $f_{ordering}(L, criteria)$ also allows ordering using the change summation $\sum_{m=1}^{M}(\Delta^m_{(r_e,x_n)})$. Finally, function $f_{icon}(r^m_e, x_n)$ returns a rectangle positioned and sized proportional to the change $\Delta^m_{(r_e,x_n)}$, with positive changes colored in green and negative in purple, with the cell matrix background filled using a less saturated color. If $\Delta^m_{(r_e,x_n)} = 0$, the cell matrix is left blank. To help understand the class swapping, we add another column to the right of the table indicating the classification returned by the original rule r_z, showing the difference to the similar rule r_e that cause the DT_k to change prediction.

Figure 4 shows the ExMatrix LE/SC visualization for instance $x_{13} = [6.9, 3.1, 4.9, 1.5]$ from the same RF model of Figure 2. Features F' are ordered by importance and rules by change sum. The dashed lines represent the instance x_{13} values. As an illustration, rule r_6 presents the smallest change in the feature "petal length" to replace a rule of majority class *virginica* for a rule of class *versicolor*, potentially increasing the RF original outcome of 72% for class *versicolor* on instance x_{13}.

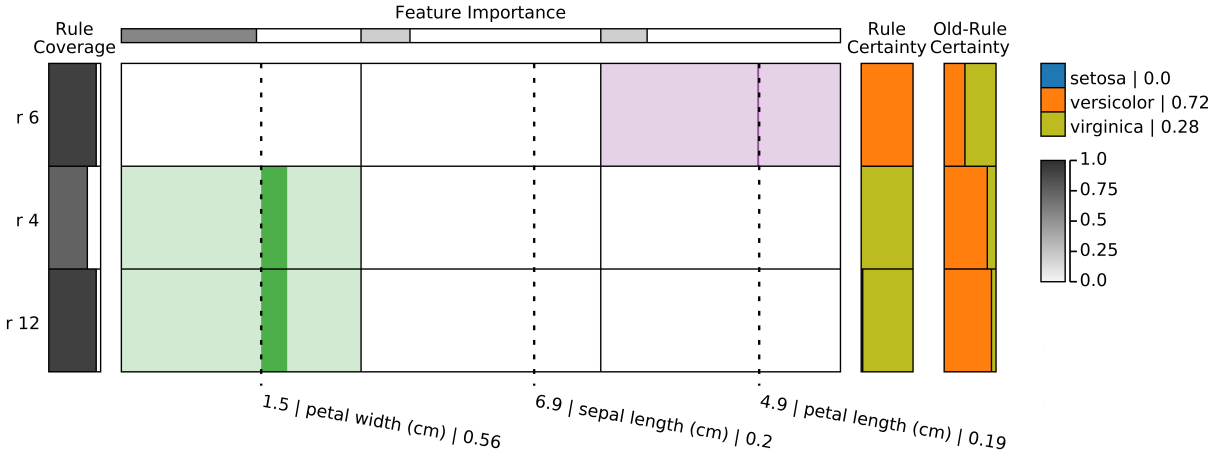


Figure 4 – ExMatrix Local Explanation Showing Smallest Changes (LE/SC) visualization. Three rules with the smallest change to make the DTs to change class decisions are displayed. The rule in the first row presents the smallest change. Small perturbations may change the RF classification decision.

2.4 Results and Evaluation

In this section, we present and evaluate our method through a use-case [3] discussing the proposed features, two usage-scenarios [4,5] showing ExMatrix being used to explore RF models, finishing with a formal user test. All datasets employed (see Table 2) in this section were downloaded from the *UCI Machine Learning Repository* (DHEERU; TANISKIDOU, 2017), and the ExMatrix implementation is publicly available as a Python package at <https://pypi.org/project/exmatrix/> to be used in association with the most popular machine learning packages.

Table 2 – Datasets used for ExMatrix evaluation.

Name	Source	Preprocessing
Wisconsin Diagnostic Breast Cancer (WDBC)	Dua and Graff (2017)	-
German Credit Data	Dua and Graff (2017)	Zhao *et al.* (2019)
Contraceptive Method Choice	Dua and Graff (2017)	-

2.4.1 Use Case: Breast Cancer Diagnostic

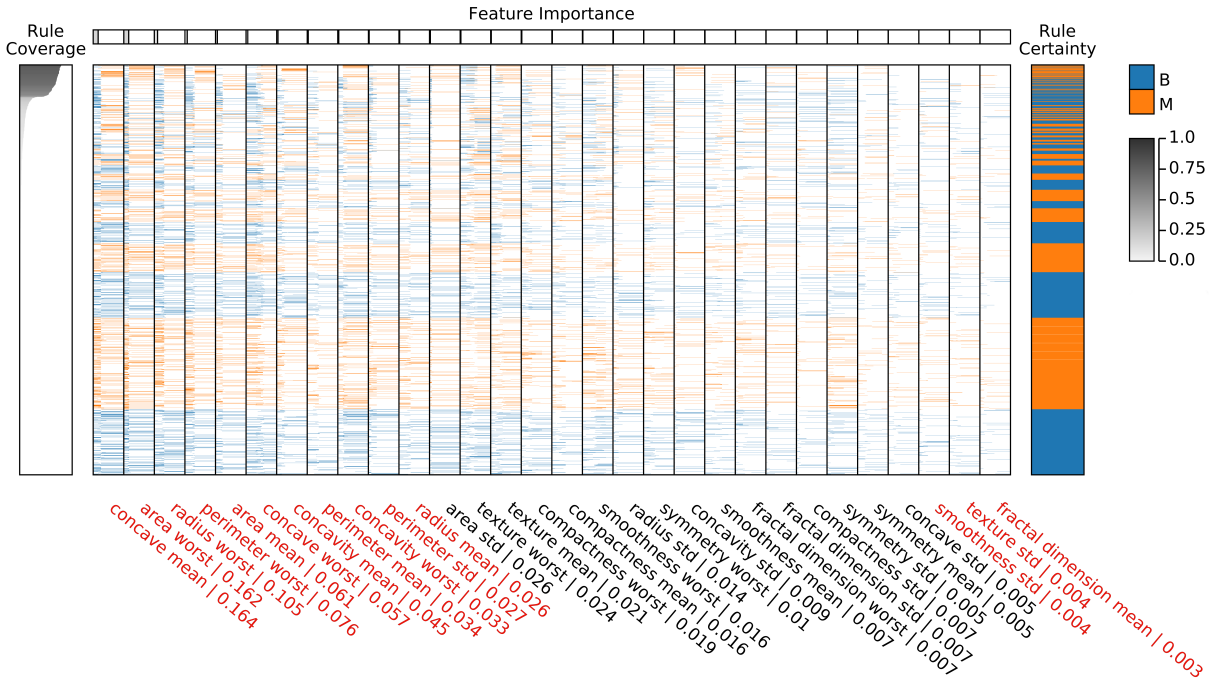
In this use case, we utilize the *Wisconsin Breast Cancer Diagnostic (WBCD)* dataset to discuss how to use ExMatrix global and local explanations to analyze RF models. The WDBC dataset contains samples of breast mass cells of $N = 569$ patients, 357 classified as benign (B) and 212 as malignant (M), with $M = 30$ features (cells properties).

[3] <https://popolinneto.gitlab.io/exmatrix/papers/2020/ieeevast/usecase/>
[4] <https://popolinneto.gitlab.io/exmatrix/papers/2020/ieeevast/usagescenarioi/>
[5] <https://popolinneto.gitlab.io/exmatrix/papers/2020/ieeevast/usagescenarioii/>

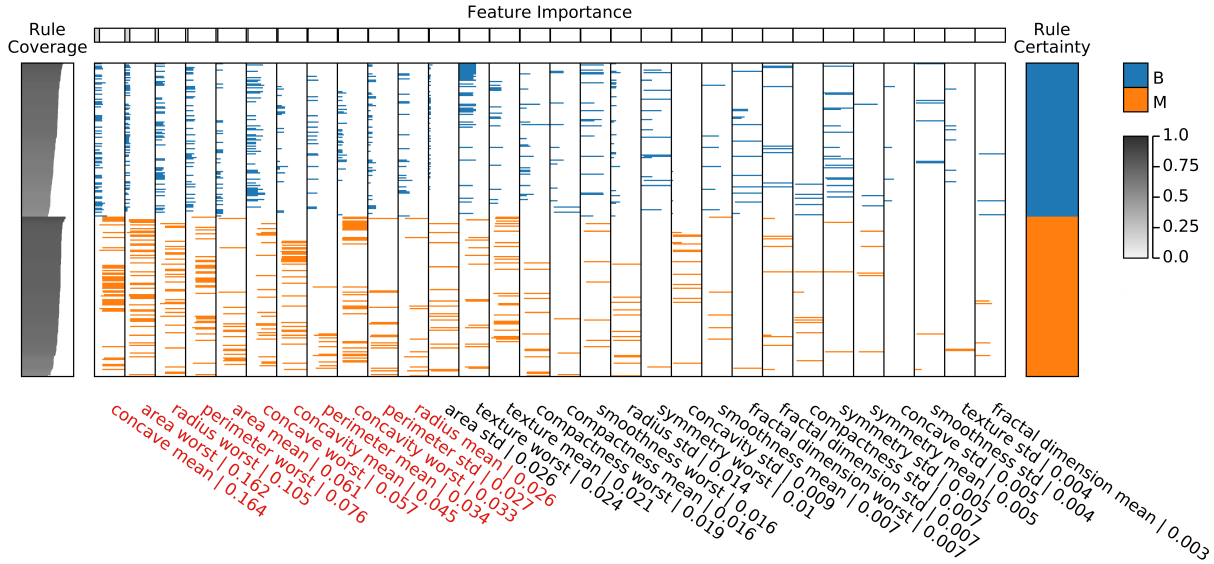
The RF model used as example was created randomly selecting 70% of the instances for training and 30% for testing and setting the number of DTs to $K = 128$, not limiting their depths. The result is a model with 3,278 logic rules, 25.6 rules per DT, and an accuracy of 99%.

An overview of this model is presented in Figure 5a using the ExMatrix GE representation (see subsubsection 2.3.3.1). In this visualization, rules are ordered by coverage and features by importance. Using this ordering scheme, it is possible to see that "concave mean", "area worst", and "radius worst" are the three most important features, whereas "smoothness std", "texture std", and "fractal dimension mean" are less important, and that the RF model used all 30 features. Also, taking only the high coverage rules and features with more importance ("concave mean' to "radius mean"), some patterns in terms of predicate ranges emerge. To help verify these patterns, low-coverage rules can be filtered out, resulting in a new visualization containing only high-coverage rules. Figure 5b presents the resulting filtered visualization with rules ordered by class & coverage facilitating the comparison between the two dataset classes. In this new visualization, it is apparent that low feature values appear to be related to class B whereas higher values to class M (goals **G1**, **Q1**, and **Q2**). In this example, filtering aids in focusing on what is important regarding the overall model behavior, removing unimportant information and reducing cluttering, relying on the so-called Schneiderman's visualization mantra (SHNEIDERMAN, 1996).

The error rate of 1% in this model is due to the misclassification of only one instance of the test set. Instance x_{29} was wrongly classified as class B with a probability of 55%. Figure 6a shows the ExMatrix LE/UR representation (see subsubsection 2.3.3.2) using x_{29} as target instance. In this visualization, the matrix is ordered by class & coverage to focus on the difference between classes, and some interesting patterns are visible. For instance, predicate ranges of both classes B and M overlap for most features, except for "fractal dimension std" and "concave std". Also, these two features, along with "symmetry std", "concave mean", "compactness std", and "symmetry mean" are more related to class B (blue) since rules of such class heavily use them and sparsely used by rules of class M (orange) showing what is actively used by the model to make the prediction (goals **G2** and **Q3**). Besides, analyzing ExMatrix LE/SC visualization on Figure 6b, one can notice that positive changes on features "concave mean" and "perimeter worst" may tie or alter the prediction of x_{29} to class M since many green cells can be observed in the respective columns for rules of class M, while negative changes on "area worst" and "concavity mean" increases its classification as class B since many purple cells can be observed in the respective columns for rules of class B (goals **G3** and **Q4**).



(a) ExMatrix GE visualization.

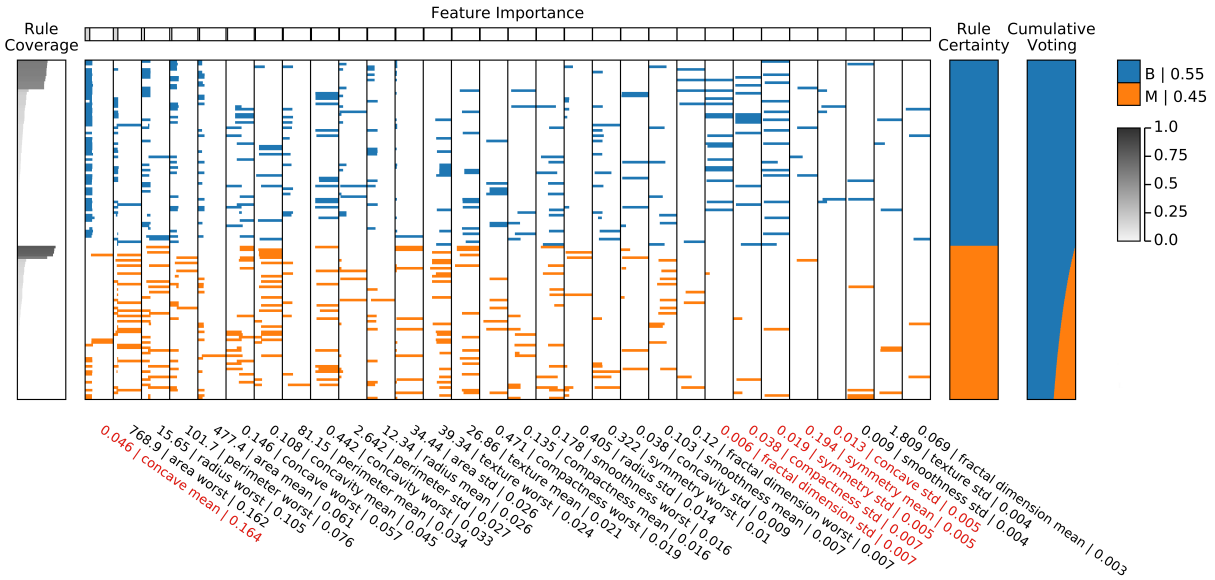


(b) ExMatrix GE representation with filtered rules (only high-coverage rules).

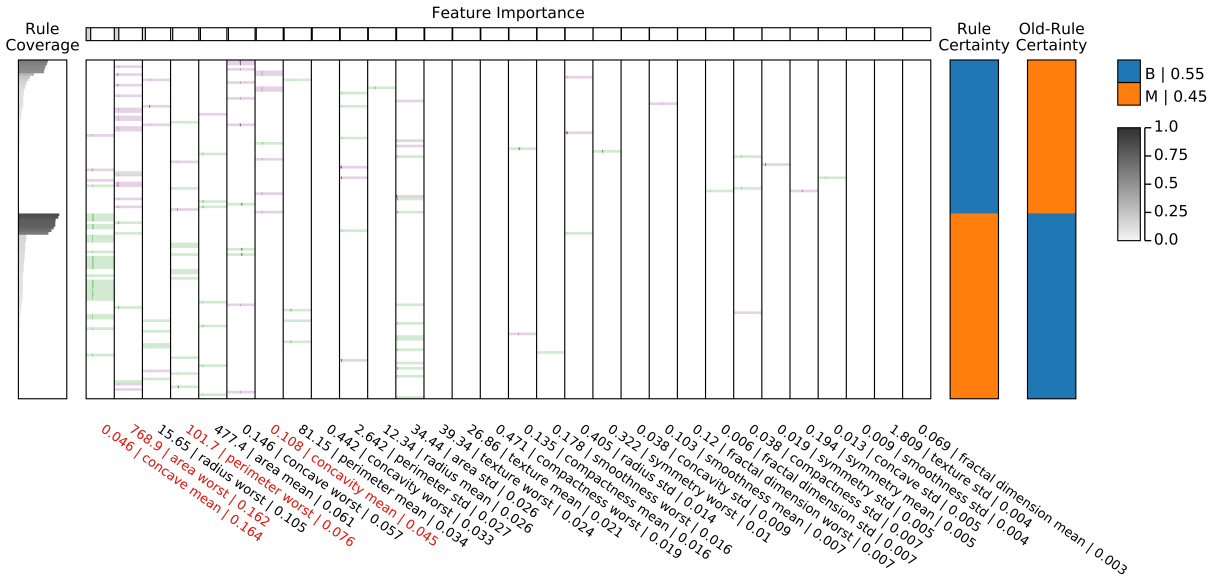
Figure 5 – ExMatrix GE representations of the WDBC RF model. In (a), giving the ordering scheme by rule coverage and feature importance, some patterns emerge in terms of predicates ranges. In (b) the low-coverage rules are filtered-out to help focusing the analysis on what is important. Low feature values appear to be more related to class B whereas higher values to class M for the most important features.

2.4.2 Usage Scenario I: German Credit Bank

As a first hypothetical usage scenario, we describe a bank manager Sylvia incorporating ExMatrix in her data analytics pipeline. To speed up the evaluation of loan applications, she sends her dataset of years of experience to a data science team and asks for a classification system to aid in the decision-making process. Such dataset contains 1,000



(a) ExMatrix LE/UR for instance x_{29}, showing the used rules on the classification process.



(b) ExMatrix LE/SC for instance x_{29}, presenting changes in the instance feature values to make the DTs to change class prediction.

Figure 6 – ExMatrix local explanations of the WDBC RF model. Two different visualizations are displayed, one showing the rules employed in the classification of a target instance (a), and one presenting the smallest changes to make the trees of the model to change the prediction of that instance (b). In both cases, the target instance is the only misclassified instance.

instances (customers profiles) and 9 features (customers information), with 700 customers presenting rejected applications and 300 accepted (here we use a pre-processed (ZHAO *et al.*, 2019) version of German Credit Data from UCI). For the implementation of such a system, Sylvia has two main requirements: (1) the system must be precise in classifying loan applications, and; (2) the classification results must be interpretable so she can explain the outcome.

To fulfill the requirements, the data science team builds an RF model setting the number of DTs to 32 with a maximum depth of 6. The produced model's accuracy was 81%, resulting in 1,273 logic rules, 38.7 rules per DT. Using the ExMatrix GE representation (omitted due to space constraints, see supplemental material) [6], she observes that the features "Account Balance", "Credit Amount", and "Duration of Credit" are the three most important, whereas "Value Savings/Stocks", "Duration in Current address", and "Instalment percent" are the three less. Also, by inspecting the most generic knowledge learned by the system (patterns formed by high-coverage rules) using a filtered representation of the ExMatrix GE visualization on Figure 7, she notices that applications that request a credit to be paid in more extended periods (third column) tend to be rejected, matching her expectations. However, unexpectedly, customers without account ("Account Balance": 1 - No account, 2 - No balance, 3 - Below $200 , 4 - $200 or above) have less chance to have their application rejected (first column), something she did not anticipate (goals **G1**, **Q1**, and **Q2**). Although confronting some of her expectations and bias, she trusts her data, and the classification accuracy seems convincing, so she decides to put the system in practice.

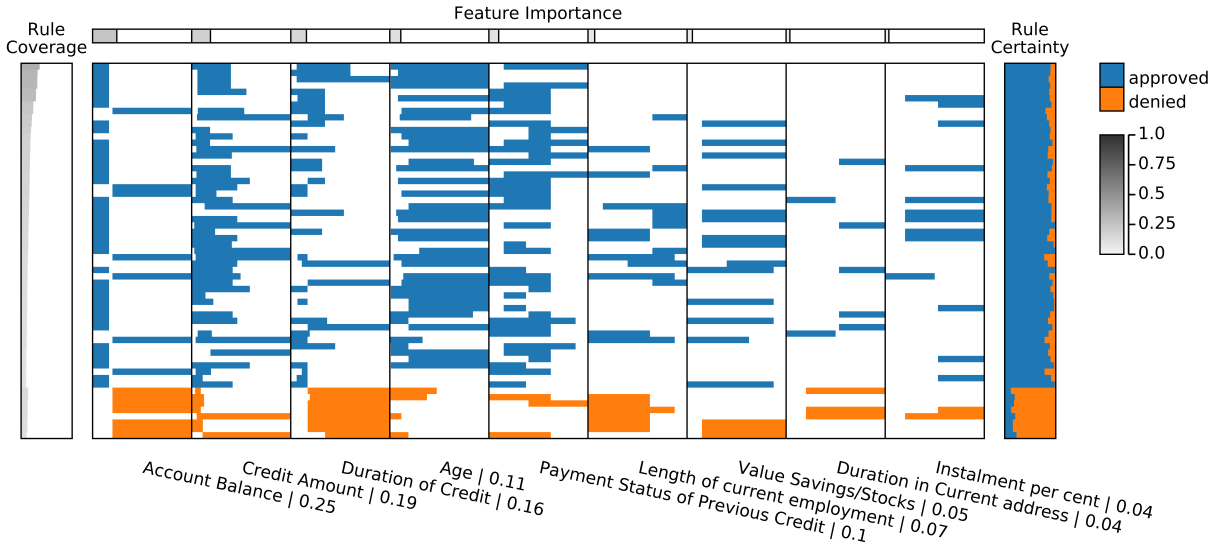


Figure 7 – ExMatrix GE representation (rules filtered by coverage and certainty) of the RF model for the German Credit Data UCI dataset. Based on the most generic knowledge learned by the RF model (rules with high coverage), it is possible to conclude that applications requesting credit to be paid in longer periods tend to be rejected.

One day she receives a new customer interest in a loan. After filling the system with his data, unfortunately, the application got rejected by the classification system. Based on the new *European General Data Protection Regulation* (LIU; WANG; MATWIN, 2018; CARVALHO; PEREIRA; CARDOSO, 2019; GUIDOTTI *et al.*, 2018b) that requires explanations about decisions automatically made, the customer requests clarification. By inspecting the ExMatrix LE/UR visualization on Figure 8a, she notices, besides the

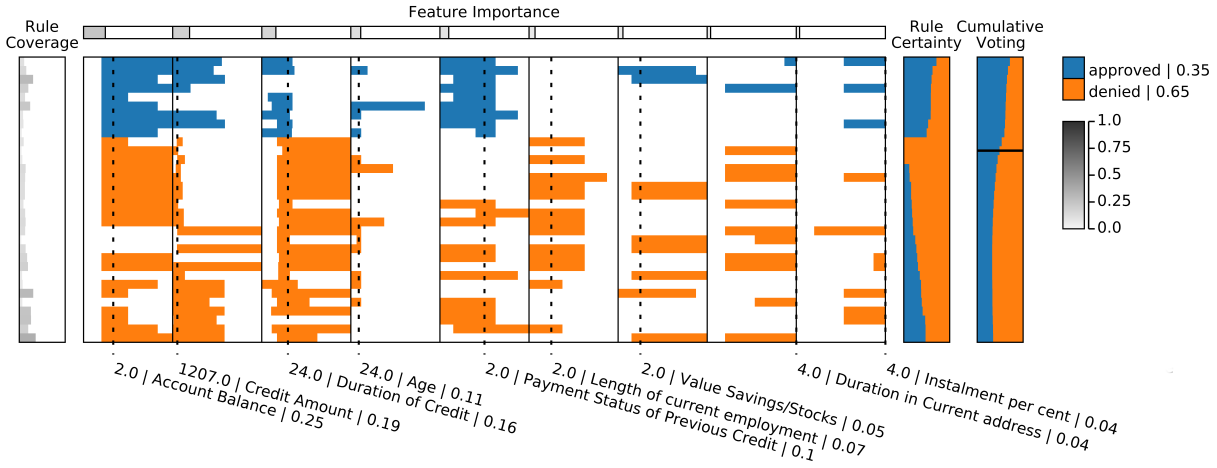
[6] Figure 24 of Appendix B.

denied probability of 65%, that even if all "approved" rules (blue) are used, very few high-certainty "denied" rules (orange) define the final decision of the model (see the Cumulative Voting and Rule Certainty columns), indicating that those rules, and the related logic statements, have a strong influence in the loan rejection. Also, she sees that the feature "Length of current employment" is the most directly related to the denied outcome since it is used only by rules that result in rejection (goals **G2** and **Q3**). Using this information, she explains to the customer that since he is working for less than one year in the current job (2 as "Length of current employment" corresponds to less than 1 year), the bank recommends denying the application. However, analyzing the ExMatrix LE/SC representation in Figure 8b, she realizes that negative changes in features "Credit Amount" and "Duration of Credit" may turn the outcome to approved (goals **G3** and **Q4**).Thereby, as an alternative, she suggests lowering the requested amount and the number of installments. Based on the observable differences to make the rules change class, she notices that upon reducing the credit application from $1,207 to $867 and the number of payments from 24 to 15, the system changes recommendation to "approved". Figure 8c presents the ExMatrix LE/UR visualization if such suggested values are used, changing the final classification.

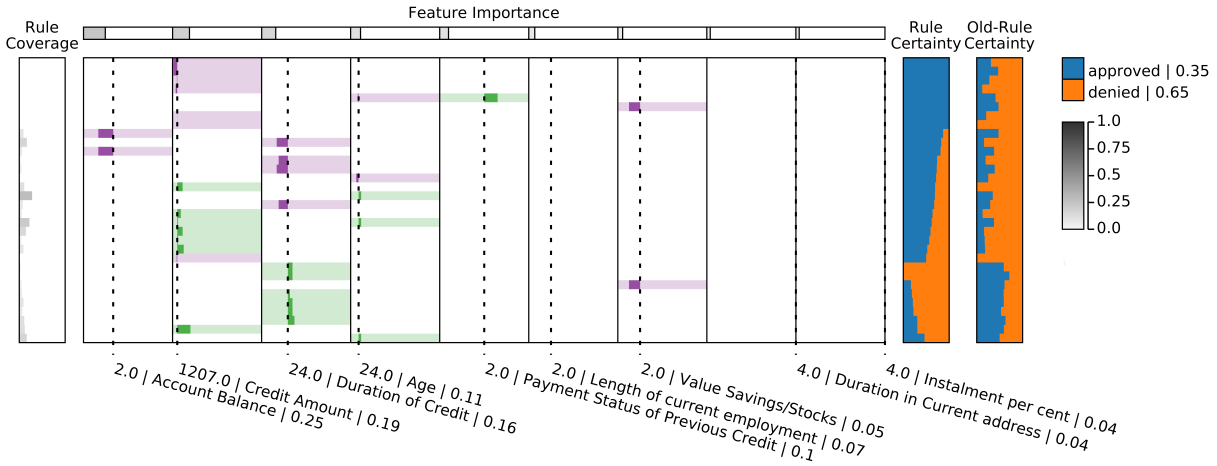
2.4.3 Usage Scenario II: Contraceptive Method

This last usage scenario presents Christine, a public health policy manager who wants to create a contraceptive campaign to advertise a new, safer drug for long term use. To investigate married wives' preferences, Christine's data science team creates a prediction model using a data set with information about contraceptive usage choices her office collected past year (here we use the Contraceptive Method Choice dataset from UCI). The dataset contains 1,473 samples (married wives profiles) with 9 features, where each instance belongs to one of the classes "No-use", "Long-term", and "Short-term", regarding the contraceptive usage method, with 42.7% of the instances belonging to class No-use, 22.6% to Long-term, and 34.7% to Short-term.

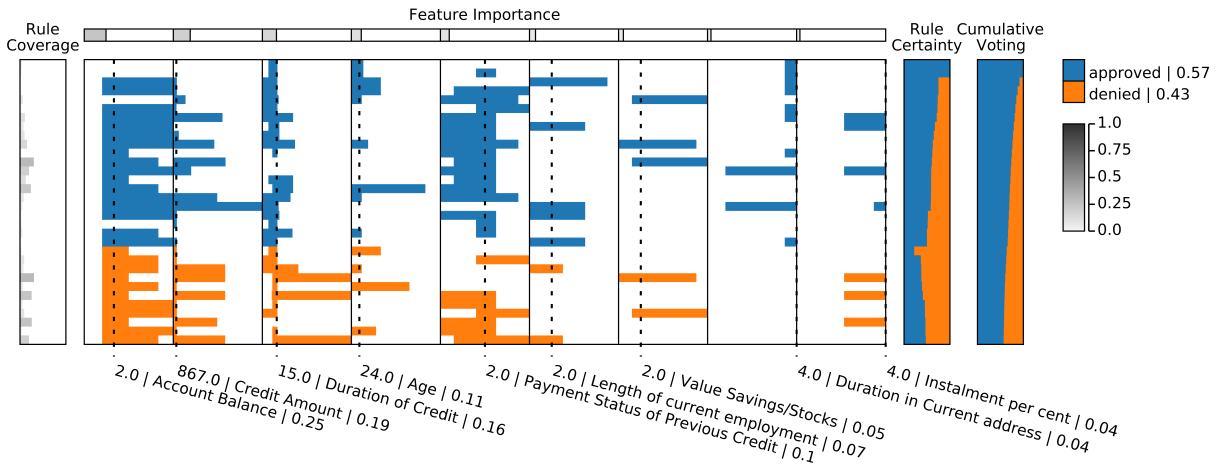
Since interpretability is mandatory in this scenario, allowing the results to be used in practice, the data science team creates an RF model and employs ExMatrix to support analysis. To create the model, the team set the number of DTs to 32 and maximum depth to 6, resulting in 1,383 logic rules, 43.2 rules per DT. The RF model accuracy is 63%, and, although not ideal for individual classifications, can be used to understand general knowledge learned by the model from the dataset.



(a) ExMatrix LE/UR for instance x_{154}.



(b) ExMatrix LE/SC for instance x_{154}.



(c) ExMatrix LE/UR modifying instance x_{154}, which changes RF's decision.

Figure 8 – ExMatrix local explanations of a RF model for the German Credit Data UCI dataset. Analyzing one sample (instance x_{154}) of rejected application (a), it is possible to infer that it is probably rejected due to the (applicant) short period working in the current job. However, lowering the requested amount as well as the number of instalments can change the RF's decision (b) and (c).

By inspecting the ExMatrix GE representation of the model (omitted due to space constraints, see supplemental material) [7], she readily understands that the features "Number of children ever born", "Wife age", and "Wife education" are the three most relevant for defining the contraceptive method class, while "Media exposure", "Wife now working?", and "Wife religion" are the three less. Also, further exploring a filtered version of the ExMatrix GE representation on Figure 9, to focus only on high-coverage and high-certainty rules ordered by class & coverage, she notices some interesting patterns regarding features space ranges and classes. For instance, lower values for the feature "Number of children ever born" (first column) are more related to class No-use and rarely related to class Long-term. For contraceptive method usage, higher values for the feature "Wife age" (second column) are related to class Long-term, while average and lower values are more related to class Short-term. Also, higher values for "Wife education" (third column) are more related to class Long-term (goals **G1**, **Q1**, and **Q2**). Based on these observations, and given the modest budget she received for the campaign, Christine decides to focus on the group of older and highly educated wives with at least one child to target the campaign's first phase.

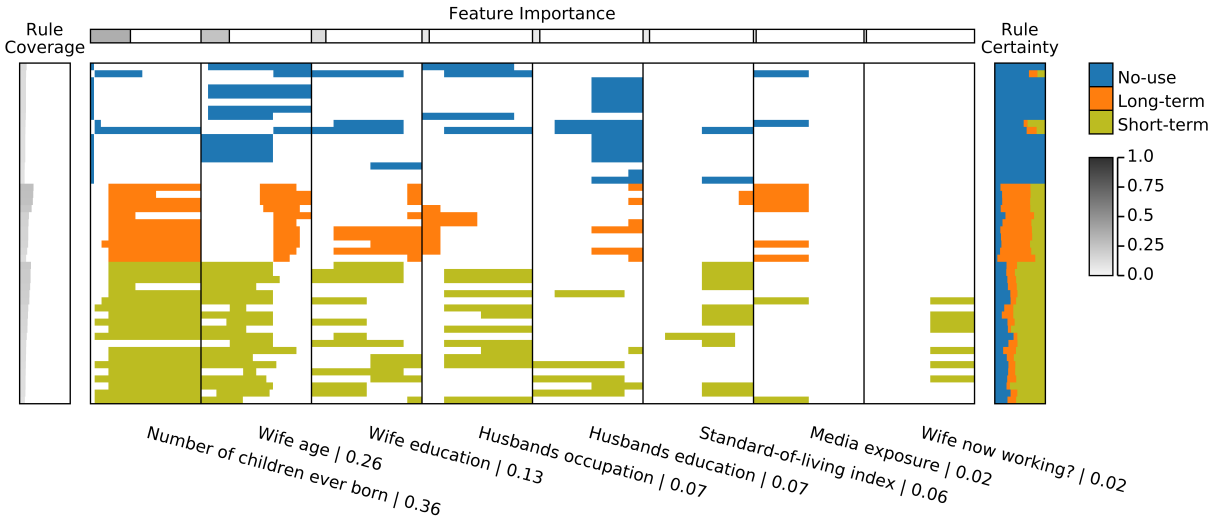


Figure 9 – ExMatrix GE representation (rules filtered by coverage and certainty) of the RF model for the Contraceptive Method Choice UCI dataset. Based on high-coverage high-certainty rules, some interesting patterns can be observed. For instance, on contraceptive method usage, older women tend to use long-term contraceptive methods.

2.4.4 User Study

To evaluate the ExMatrix method, we performed a user study to assess the proposed visual representations for global and local explanations. In this study, we asked four different questions based on the ExMatrix visualizations created for the use-case presented in subsection 2.4.1, focusing on evaluating the goals presented in Table 1.

[7] Figure 25 of Appendix B.

The study started with video tutorials about RF basic concepts and how to use ExMatrix to analyze RF models and classification results through the proposed explanations. A total of 13 users participated, 69.2% male and 30.8% female, aged between 24 to 36, all with a background in machine learning. The participants were asked to analyze the explanations of Figure 5a, Figure 6a, and Figure 6b, where each analysis was followed by different question(s) (see Table 3). On the visualizations, features names were replaced by "Feature 1" to "Feature 30" and classes names by "Class A" and "Class B", aiming at removing any influence of knowledge domain in the results, since our focus is to assess the visual metaphors.

Table 3 – User study questions.

Question	Goals	Visualization
Qst 1 - About features space ranges and class ASSOCIATIONS. Considering rules with HIGH COVERAGE, and features with HIGH IMPORTANCE, select your answer: (three options of associations)	G1, Q1, and Q2	Figure 5a
Qst 2 - Instance 29 is classified as Class A with a probability of 55%, against 45% for Class B. What feature is more related to Class A and less related to Class B? (four options of features)	G2 and Q3	Figure 6a
Qst 3 - Select the pair of features where DELTA CHANGES on instance 29 will potentially INCREASE Class A probability, and by that may SUPPORT its classification as Class A. (four options of features pairs)	G3 and Q4	Figure 6b
Qst 4 - Select the pair of features where DELTA CHANGES on instance 29 will potentially INCREASE Class B probability, and by that may ALTER its classification as Class A. (four options of features pairs)	G3 and Q4	Figure 6b

Using the ExMatrix GE representation (Figure 5a), 76.9% of the participants were able to identify patterns involving feature space ranges and classes, where, for high coverage rules and high importance features, low features values are more related to class B, while features with large values are more related to class M (**Qst 1**). Using the ExMatrix LE/UR (Figure 6a), also 76.9% of the participants were able to recognize that feature "concave std" is the most related to class B for instance x_{29} classification outcome (**Qst 2**). Using the ExMatrix LE/SC (Figure 6b), 61.5% of the participants were able to identify that negative changes on instance x_{29} features "area worst" and "concavity mean" values would better support the class B outcome (**Qst 3**), and 46.2% were able to identify that positive changes on features "concave mean" and "perimeter worst" values may alter the outcome from class B to class M (**Qst 4**).

In general, the results were promising for the first two analyses, but the participants

present worse results when interpreting the ExMatrix LE/SC visualization. This is not surprising since this representation requires a much better background about RF theory. The ExMatrix GE and LE/UR visualizations are more generic and involve much fewer concepts about how RF models work internally. In contrast, the ExMatrix LE/SC requires a good level of knowledge about ensembles models and how the voting system work when making a prediction. Although most of the users self-declared with a background in machine learning, only 30% are RF experts.

We also have asked subjective, open questions, and, in general, users gave positive feedbacks about ExMatrix explanations, where the visualizations were classified as visually pleasing and useful for understanding RF models.

2.5 Discussion and Limitations

Although the natural choice to visualize a tree collection is to use tree structure metaphors, two main reasons make disjoint rules organized into tables a better option when analyzing DTs and especially RFs. First, using tree structure metaphors, the visual comparison of logic rules (decision paths) can be overwhelming since different paths from the root to the leaves define different orders of attributes, slowing down users when searching within a tree to answer classification questions (FREITAS, 2014; HUYSMANS *et al.*, 2011). An issue that is amplified in RFs, since multiple DTs are analyzed collectively. In contrast, in a matrix metaphor, the attributes are considered in the same order easing this process (FREITAS, 2014; HUYSMANS *et al.*, 2011). Second, given the constraints of usual DT inference methods (non-overlapping predicates with open intervals), features can be used multiple times in a single decision path resulting in multiple nodes (one per test) using the same feature. Consequently, if tree structures are employed, each feature's decision intervals need to be mentally composed by the user, and nodes using the same feature can be far away in the decision path. The decision intervals are explicit in the matrix representation and can be easily compared across multiples rules and trees. Therefore, although tree structure visual metaphors are the usual choice when hierarchical structures are the focus (GRAHAM; KENNEDY, 2010; SCHULZ; HADLAK; SCHUMANN, 2011), on DTs and RFs, the decision paths are the object of analysis (TAN; STEINBACH; KUMAR, 2005; FREITAS, 2014; HUYSMANS *et al.*, 2011; LIMA; MUES; BAESENS, 2009) and transforming paths into disjoint rules organized into tables emphasize what is essential (see supplemental material) [8].

Considering the above points, it is clear that scalability for RFs visualization is not just a choice of getting a visual metaphor that can represent millions of nodes, but getting a visual representation that is scalable and still properly supports essential analytical tasks

[8] Section B.2 "Why logic rules in a matrix-like visual metaphor instead of node-link diagrams?" of Appendix B.

(see Table 3). Something much more complex than merely visualizing a forest of trees. In this scenario, ExMatrix renders a promising solution, supporting the analysis of many more rules concomitantly than the existing state-of-the-art techniques. However, it is not a perfect solution. ExMatrix covers two different perspectives of RFs, conveying Global and Local information. In the Local visualization, scalability is not a problem since one rule is used per DT, so even for RFs with hundreds or even thousands of trees, ExMatrix scales well. However, for Global visualization, scalability can be an issue since the number of rules substantially grows with the number of trees. Although we can represent one rule per line of pixels, we are limited by the display resolution, and, even when the display space suffices, ExMatrix layouts can be cluttered and tricky to explore.

The solution we adopt to address scalability was to implement the so-called Schneiderman's visualization mantra (SHNEIDERMAN, 1996), allowing users to start with an overview of the model, getting details-on-demand by filtering rules to focus on specific sets of interest. Although users are free to select any subset of rules, considering that the goal of the Global visualization is to generate insights about the overall models' behavior, here we mainly explore filtering low-coverage rules since they are only valid for a few specific data instances (that is the coverage definition). Although simple, such a solution makes the analysis of entire models easier by removing unimportant information and reducing cluttering. Another potential solution is to make the rows' height proportional to coverage or certainty so that the rules with the lowest coverage or certainty are less prominent (visible) and could even be combined in less than one line of pixels. We have not tested this approach and left it as future work.

Regarding the user study, although the results were satisfactory and within what we expect for the ExMatrix GE and LE/UR visualizations, the results for the ExMatrix LE/SC representation were sub-optimal, and the *XAI Question Bank* (LIAO; GRUEN; MILLER, 2020) can help us to shed some light about the reasons. According to this bank, the GE addresses the leading question "*How (global)*" as "*What are the top rules/features it uses?*", whereas the LE/UR addresses the leading question "*Why*", enabling to answer inquiries such as "*Why/how is this instance given this prediction?*". However, the LE/SC involves three leading questions, "*What If*", "*How to be that*", and "*How to still be this*", where the changes on instance features values are presented supporting hypotheses (not answers), which shown to be too complex for the users. We believe that designing visual representations to answer each of these questions individually would be more effective and may reach better results.

Nevertheless, as discussed in the User Study section, participants' low performance not only resulted from the visual metaphor but also the expertise on RF models. Among the participants, few know the RF technique in detail, indicating that people with less expertise can use ExMatrix GE and LE/UR visualizations, but the LE/SC representation

is more suitable for experts. In general, despite the complexity of the questions we ask participants to solve, they acknowledged the ExMatrix potential, expressing encouraging remarks, including "*... this solution ... allows a deeper understanding of how each particular rule or feature impacted on the final the decision/classification.*" or "*I think the ExMatrix can be used in a variety of domains, from E-commerce to Healthcare...*".

Although we design ExMatrix with RF interpretability in mind, it can be readily applied to DT models, such as the ones used as surrogates for black-box models as Artificial Neural Networks and Support Vector Machines, or approaches based on logic rules such as Decision Tables since the core of our method is the visualization of rules. Another compelling scenario that can be explored is the engineering of models. In this case, through rule selection and filtering, simplified models could be derived where, for instance, only high coverage rules are employed or any other subset of interest. Also, model construction and improvement can be supported. The visual metaphors we propose can be easily applied to the analysis and comparison of RF models resulting from different parametrizations, such as different numbers of trees and their maximum depth. Therefore, allowing machine learning engineers to go beyond accuracy and error when building a model.

2.6 Conclusions and Future Work

In this paper, we present *Explainable Matrix (ExMatrix)*, a novel method for Random Forest (RF) model interpretability. ExMatrix uses a matrix-like visual metaphor, where logic rules are rows, features are columns, and rules predicates are cells, allowing to obtain overviews of models (Global Explanations) and audit classification results (Local Explanations). Although simple, ExMatrix visual representations are powerful and support the execution of tasks that are challenging to perform without proper visualizations. To attest ExMatrix usefulness, we present one use-case and two hypothetical usage scenarios, showing that RF models can be interpreted beyond what is granted by usual metrics, like accuracy or error rate. Although our primary goal is to aid in RF models global and local interpretability, the ExMatrix method can also be applied for the analysis of Decision Trees, such as the ones used as surrogates models, or any other technique based on logic rules, opening up new possibilities for future development and use. We plan as future work to create new ordering and filtering criteria along with aggregation approaches to improve the current ExMatrix explanations and, more importantly, to conceive new ones. Another fascinating forthcoming work is creating optimized rule-based models from complex RF models, which we also intend to investigate.

CHAPTER

3

MULTIDIMENSIONAL CALIBRATION SPACE – MCS

This chapter (paper *Machine learning used to create a multidimensional calibration space for sensing and biosensing data* [1]) presents an ExMatrix (Chapter 2) application in analytical chemistry called MCS (Multidimensional Calibration Space). Sensors and biosensors based on Impedance Spectroscopy generate multidimensional (multivariate) data from samples with analyte concentrations or distinguishable factors. DT models can be employed for sensing units calibration, being interpretable by ExMatrix. The latter can be applied in several domains, being analytical chemistry a significant example. **Due to copyright issues, the published paper [1] introducing the MCS concept is not fully arranged in this chapter. Nevertheless, the sections omitted do not harm the context of this Ph.D. thesis.**

Abstract: Calibration curves are essential constructs in analytical chemistry to determine parameters of sensing performance. In the classification of sensing data of complex samples without a clear dependence on a given analyte, however, establishing a calibration curve is not possible. In this paper we introduce the concept of a multidimensional calibration space which could serve as reference to classify any unknown sample as in determining an analyte concentration from a calibration curve. This calibration space is defined from a set of rules generated using a machine learning method based on trees applied to the dataset. The number of attributes employed in the rules defines the dimension of the calibration space and is established to warrant full coverage of the dataset. We demonstrate the calibration space concept with impedance spectroscopy data from sensors, biosensors and an e-tongue, but the concept can be extended to any type of sensing

[1] POPOLIN NETO, M.; SOARES, A. C.; OLIVEIRA, O. N.; PAULOVICH, F. V. Machine learning used to create a multidimensional calibration space for sensing and biosensing data. Bulletin of the Chemical Society of Japan, v. 94, n. 5, p. 1553–1562, 2021. Available: <https://-doi.org/10.1246/bcsj.20200359>.

data and classification task. Using the calibration space should allow for the correct classification of unknown samples, provided that the data used to generate rules via machine learning can cover the whole range of sensing measurements. Furthermore, an inspection in the rules can assist in the design of sensing systems for optimized performance.

3.1 Introduction

Analytical curves are ubiquitous in analytical chemistry, with well-established procedures to determine analytical parameters recommended by IUPAC (CURRIE, 1995; CURRIE, 1999). The same can be said of calibration curves and processes for instruments in general, some of which are utilized not only to determine a given physical quantity but also to adjust the instruments for correct functioning (MOOSAVI; GHASSABIAN, 2018; CASES; LÓPEZ-LORENTE; LÓPEZ-JIMÉNEZ, 2018). In sensors and biosensors, in particular, using analytical curves one can "transform" the task of classifying the set of samples under analysis into a predictive exercise where the concentration of a given analyte in an unknown sample can be determined precisely. In other types of sensors, as in the case of electronic tongues (SHIMIZU *et al.*, 2017; BRAUNGER *et al.*, 2017; SHIMIZU; BRAUNGER; RIUL, 2019; DAIKUZONO *et al.*, 2015; MENDEZ; PREEDY, 2016; TERMEHYOUSEFI, 2018; OLIVEIRA *et al.*, 2013) and electronic noses (MENDEZ; PREEDY, 2016; FARRAIA *et al.*, 2019; PATEL, 2014; DI NATALE *et al.*, 2000), this determination may not be possible and no analytical curves can be established. This happens because these sensors may be used to classify different types of liquids such as wine (RIUL *et al.*, 2004; RUDNITSKAYA *et al.*, 2017) or coffee (ALESSIO *et al.*, 2016) without determining the concentration of any specific analyte. Calibration procedures can nevertheless still be employed for e-tongues and e-noses (LEGIN; RUDNITSKAYA; VLASOV, 2003; GRABOSKI *et al.*, 2020), but these are related to the multivariate calibration that allows for studying quantitative and qualitative aspects of simple and complex solutions (VLASOV; LEGIN; RUDNITSKAYA, 2002). For the so-called complex samples that contain multiple analytes, use has been made of multivariate analysis (PODRAŻKA *et al.*, 2018) and other statistical and computational methods (DI ROSA *et al.*, 2017), including information visualization and machine learning techniques. These methods are advantageous for the evaluation of large volumes of data, providing predictions about food and water contaminants, diagnosis and prognosis. For example, Daikuzono *et al.* (2017) applied an information visualization technique referred to as Interactive Document Mapping (IDMAP) in electrical impedance spectra to detect gliadin in food samples contaminated by gluten. IDMAP was also used to treat data from a microfluidic electronic tongue to detect petrochemical compounds, heavy metals and basic flavors (SHIMIZU *et al.*, 2017) and from a biosensor to detect pathogenic bacteria in food samples (WILSON *et al.*, 2019). Examples of machine learning applied to diagnosis include data processing of very distinct

natures. In image analysis, for instance, the superiority of computer-assisted diagnosis (compared to human experts) is well established, which includes the use of deep learning for analyzing magnetic resonance images (LUNDERVOLD; LUNDERVOLD, 2019). For cardiac diseases, datasets with varied medical parameters and results from clinical exams, including from immunosensors, have been used for diagnosis with various machine learning strategies (VASHISTHA *et al.*, 2018). Optical biosensing data from paper-based point-of-care (POCs) have been used in conjunction with machine learning to diagnose cardio-vascular diseases (BALLARD *et al.*, 2020). Cancer biomarkers were detected with higher specificity in serum samples of patients by applying classification algorithms to SERS (surface-enhanced Raman scattering) data in a biosensor made with a microfluidic chip (BANAEI *et al.*, 2019). SERS spectra from other biosensors were treated with machine learning algorithms for the diagnosis of liver cancer and liver cirrhosis (LI *et al.*, 2015). Even for glucose detection using amperometry has a genetic algorithm been useful to improve diagnosis (GONZALEZ-NAVARRO *et al.*, 2016). Calibration curves are not much useful in these cases, though they are sometimes employed for comparison purposes and to obtain sensing performance.

The usefulness of the various statistical and computational methods above mentioned is irrefutable, but their limitation remains in not being interpretable in the classification tasks. In this paper, we introduce the concept of a multidimensional calibration space which we believe addresses this limitation. The definition and exemplification of the multidimensional calibration space is given in the following section.

3.2 Methodology

The concept of a multidimensional calibration space is introduced by applying Decision Tree (DT) models (BREIMAN *et al.*, 1984; TAN; STEINBACH; KUMAR, 2005) to the impedance spectroscopy data of a sensor made with layer-by-layer films of polyelectrolytes to detect different concentrations of phytic acid (MORAES *et al.*, 2010). This problem was chosen because we knew from the latter reference that the sensor was not selective for phytic acid and a calibration curve could not be established. On the other hand, there was some distinction between the spectra for different concentrations and therefore classification should be a simple task with a small number of rules. As it will be shown, this allows the calibration space to be represented with only three dimensions.

The choice of DT models is justified by the possibility of establishing predictive rules for calibration, as DTs are today the most prevalent interpretable classification approach (HALL, 2018; GUIDOTTI *et al.*, 2018b). In classification with machine learning techniques, a computational model is built to predict the class of a given data instance. A set of data instances $X = \{x_1, x_2, ..., x_N\}$ and their associated classes $Y = \{y_1, y_2, ..., y_N\}$

where $y_i \in C = \{c_1, c_2, ..., c_M\}$, usually called the training set, are employed to infer a function $f(.)$ that maps each training instance x_i into its class y_i, that is $f(x_i) \rightarrow y_i$. If X is comprehensive and represents the phenomena under analysis in their entirety, $f(.)$ can be used to predict the class of any new instance x_j that was not originally in X. The techniques to infer $f(.)$ may be split into two distinct groups (HALL, 2018; GUIDOTTI *et al.*, 2018b), viz. the black-boxes such as Artificial Neural Networks (RIUL *et al.*, 2004) and Support Vector Machines (SONG *et al.*, 2013; KUMAR *et al.*, 2012) and the inherently interpretable models such as DTs (BREIMAN *et al.*, 1984; TAN; STEINBACH; KUMAR, 2005) and Rule Sets (HALL, 2018; GUIDOTTI *et al.*, 2018b). Although typically less accurate, the techniques of the latter group allow for the reasoning of a given classification. In other words, interpretable models support the understanding of how the attributes of a given instance, that is, the values describing it, contributed to its classification. With this scheme one may generate interpretable models not only as predictive tools but also as descriptive strategies where intrinsic relationships among data attributes and classes can be revealed (TAN; STEINBACH; KUMAR, 2005).

Using DT models as the classification approach for impedance spectroscopy data, we may select a subset of available attributes (frequencies) without requiring dimensionality reduction as in data pre-processing or manual frequency selection. By using the visualization method ExMatrix (POPOLIN NETO; PAULOVICH, 2021), the Multidimensional Calibration Space created by DT models can be explored, displaying space ranges and classes associations. The classification of individual instances/samples can be analyzed for reasoning about the class assignment. As the name indicates, DT techniques create a tree-like structure where internal nodes contain test functions (or predicates) based on the data attribute values to recursively split the training data into non-overlapping subgroups so that each final subgroup contains only instances of the same class. Figure 10 displays an example of a DT inferred from a training dataset containing 35 instances of 5 phytic acid concentrations (7 instances per concentration at 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, and 10^{-6} M). Each instance is described by its capacitance at 3 frequencies (100, 10 and 1 Hz, referred to as F100, F10, and F1, respectively) obtained with a sensor made of a layer-by-layer (LbL) film with poly(allylamine hydrochloride) (PAH) and poly(vinyl sulfonic acid) (PVS) deposited onto an interdigitated gold electrode. The DT in Figure 10 is validated through a testing set containing 15 instances (3 instances per concentration) not present in the training set, reaching 100% accuracy. In the nomenclature adopted here, we use FX to refer to the capacitance value in nF at the frequency X in Hz. Using a DT, IF/THEN logic rules can be extracted to represent the combination of attribute ranges that best describe a specific class of instances. Each tree path from the root to a leaf defines one distinct rule, and the entire set of rules can be used as a descriptive model of the (training) data. For instance, in Figure 10 the path from node #0, node #1, node #5, and node #6 defines the rule IF C(nF)@F100 $\leq$ 225.81 AND C(nF)@F1 $>$ 487.22

AND C(nF)@F1 $\leq$ 509.96 THEN 10^{-4}. If a node predicate is true, we navigate to the left branch; right otherwise. Note that different rules can result in the same class, everything depending on the complexity of the input data.

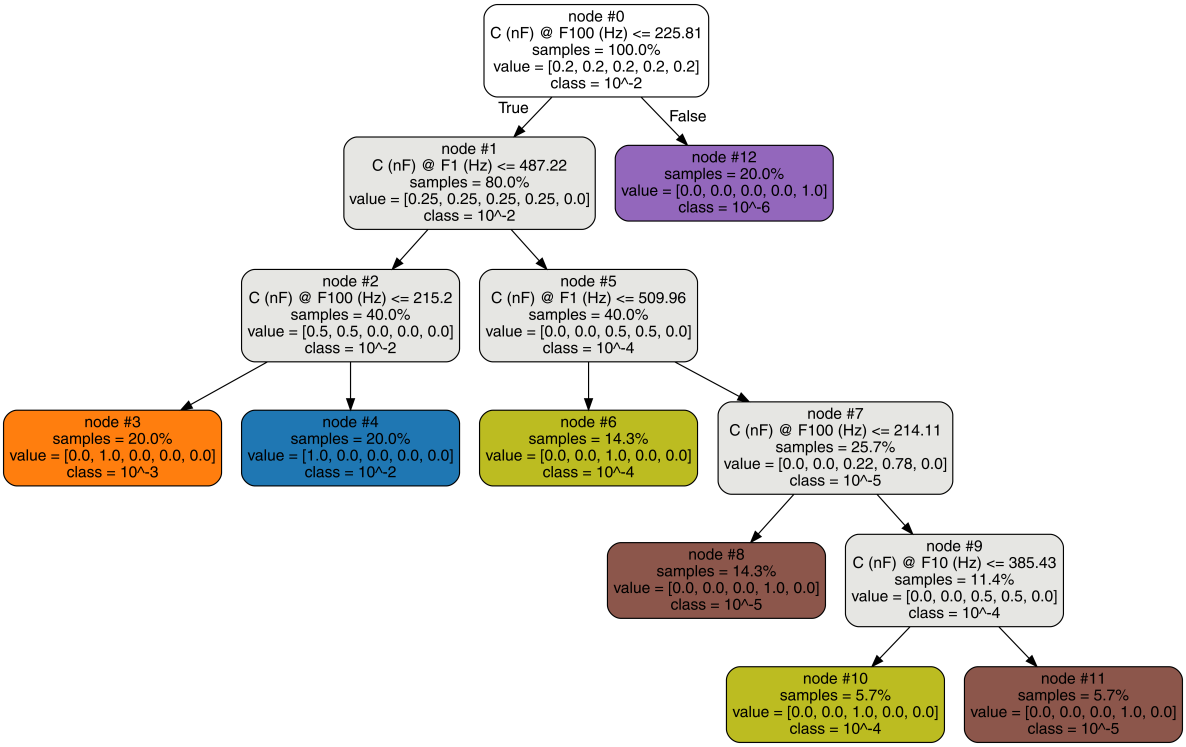


Figure 10 – Example of DT for a dataset containing the capacitance at 3 frequencies (F100, F10, and F1), measured with a PAH/PVS sensor (MORAES *et al.*, 2010) on samples of phytic acid concentrations (10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, and 10^{-6} M).

For X corresponding to data from sensors or biosensors (or other types of data as explained later on), a multidimensional calibration space visualization can be created using the ExMatrix (POPOLIN NETO; PAULOVICH, 2021) technique to display an overview of the logic rules extracted from a DT model inferred from X. DT models can be complex to interpret as the number of nodes increases, producing deep trees. With the ExMatrix visual representations, a DT model is arranged into a matrix-like visualization where rules are rows, attributes are columns, and the predicates are the matrix cells. Figure 11 shows a visual representation of the DT in Figure 10. The resulting matrix has 7 rows, one per rule, and 3 columns for the different attributes. The matrix cell is colored to reflect the inferred class and filled so that darker colors represent the range of each attribute used by a rule. In a cell, the left-most side represents the minimum value for an attribute considering the entire dataset, whereas the right-most side is the maximum. For instance, in the rule r3 depicted on the third matrix row represents the rule IF C(nF)@F100 $\leq$ 225.81 AND C(nF)@F1 $>$ 487.22 AND C(nF)@F1 $\leq$ 509.96 THEN 10^{-4}. The F100 cell is filled representing the range [167.58, 225.81], where 167.58 is the minimum value admitted by the attribute F100. F1 is filled to represent the interval

[487.22, 509.96]. In addition to obtaining the relationships among attributes ranges and classes conveyed by each rule, the coverage of the rules is also calculated. This coverage is [2] the percentage of instances in the training set belonging to the same inferred class for which the rule is true. The rule coverage is mapped to one additional column on the left side of the matrix. The rule r3 on the third row in Figure 11 extracted from the path finishing at node #6 in Figure 10 has coverage of 0.71, whereas the rule r5 on the fourth row from the path finishing at node #10 has coverage of 0.29. This indicates that the first is more generic, being valid for a higher number of instances. The last rule is more specific, valid for a small number of instances. The attribute importance is added as a row on the top of the table and reflects attributes capability to differentiate classes (BREIMAN, 2002). The attribute name is placed at the bottom, along with the attribute importance value.

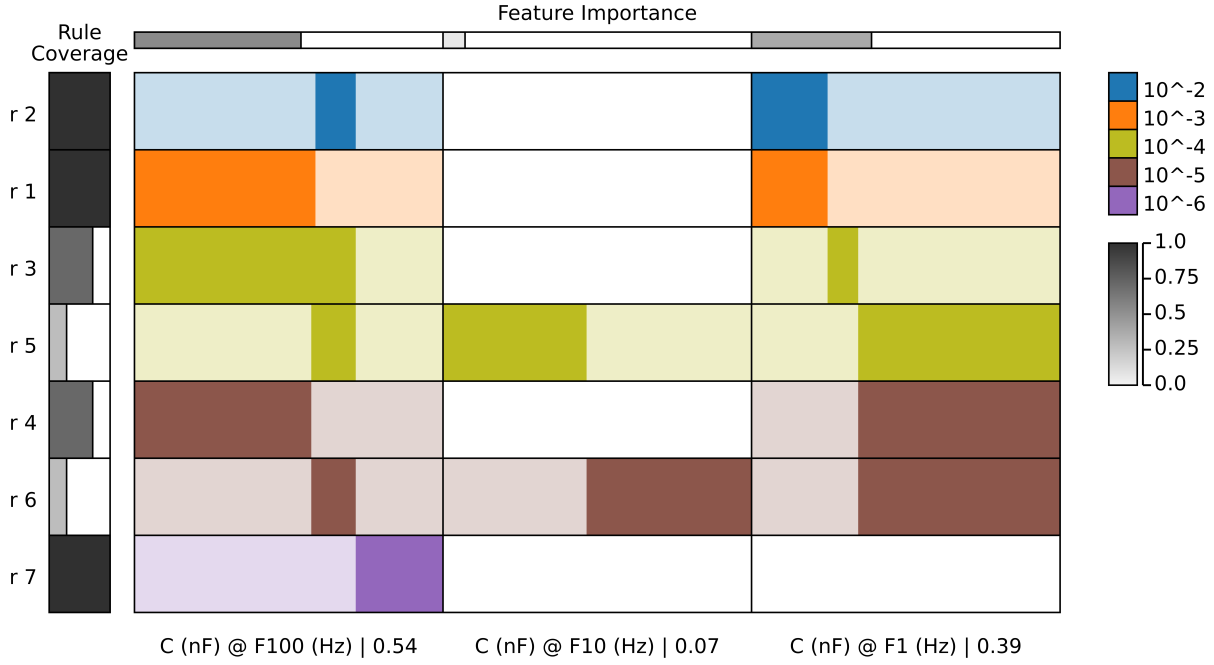


Figure 11 – Multidimensional calibration space visualization using ExMatrix. The seven rules defined in the DT of Figure 10 are represented as rows, the frequencies are in the columns and the cells indicate the ranges in each frequency "used" to predict the different concentrations. The leftmost column represents the rule coverage, with rules r2, r1, and r7 exhibiting maximum values, while r3 and r4 give intermediate values and r5 and r6 have small values. This indicates that for the concentrations 10^{-2}, 10^{-3}, 10^{-6} M, the data is easier to separate (classify) since only one rule can represent those concentration classes on the three frequencies used. However, for the concentrations 10^{-4} and 10^{-5} M, multiple rules are necessary, i.e., the data is more complex for these concentrations. By comparing the ranges in different classes, one may infer the parts of the space that best define a concentration.

Although simple, ExMatrix visual representation allows for an informative analysis. For example, an instance/sample with a higher value of real capacitance at F100 has a

[2] The rule coverage formulation used here equals rule support definition.

high probability of being 10^{-6} M (lilac color), given the high coverage of the rule r7 at the last row (Figure 11). By analyzing rules r2 and r1 at the first two rows, one notes that the 10^{-2} and 10^{-3} M concentrations (blue and orange colors) are similar at F1 and distinguished at F100, where 10^{-2} M (blue color) holds a small range of values, but higher than the values for 10^{-3} M (orange color). By inspecting rules predicting the 10^{-4} and 10^{-5} M concentrations (olive and brown colors), we observe overlaps of attribute ranges at F100 and F1, which are different in terms of F10. These overlaps and rules coverage reveal a certain space split complexity to separate the concentrations (instances/samples). It should be remarked that in simple cases feature selection can be made manually, for example noting the frequencies at which distinction among samples is higher. With DTs, on the other hand, this selection is performed in a systematic, non-arbitrary manner selecting features that present the best separability between classes. This is reflected on the feature (or attribute) importance values displayed at the top of the ExMatrix representation.

The calibration space in Figure 11 can also be represented by checking the ranges at which the different rules apply. As an example, Figure 12 shows dashed lines for an instance/sample with values 169.23, 336.54, and 532.82 for the frequencies F100, F10, and F1. This instance is classified as 10^{-5} M since it falls into the darker colored area of the rule in the first row (brown color/rule r4). In Figure 12 the rules are ordered according to the proximity to the instance under analysis, where the used rule to classify the instance is in the first row (brown/rule r4). Proximity here means the smallest modifications (gaps between dotted lines and ranges) needed to apply to the instance in order to change its class. For instance, in the second row (rule r3) one notes that a small decrease in capacitance at F1 would make the instance to be classified as belonging to the 10^{-4} M class (olive). On the other hand, larger modifications are required to make it switch, for example, to the class 10^{-2} M (blue/rule r2 on the sixth row), where capacitance values at frequencies F100 and F1 need positive and negative increments, respectively.

In the example chosen, the number of rules to cover 100% of the dataset is 7, close to the minimum possible of 5 rules for the five classes. Since only three attributes (F100, F10, and F1) are used in these rules, one may establish a 3D calibration space as shown in Figure 13 for the rules in Figure 11 (or Figure 12). The colored boxes represent the space where each concentration is different from the others. From this plot one infers the difficulty in distinguishing the 10^{-4} and 10^{-5} M concentrations (olive and brown), while the others occupy more defined parts of the 3D calibration space. Also interesting is to consider the instance classified as 10^{-5} M and discussed in connection with Figure 12, represented as a red circle in Figure 13. This circle needs to "travel" very little in the 3D space to move to class 10^{-4} M, whereas to move to class 10^{-2} M the distance is much larger (and in two axes).

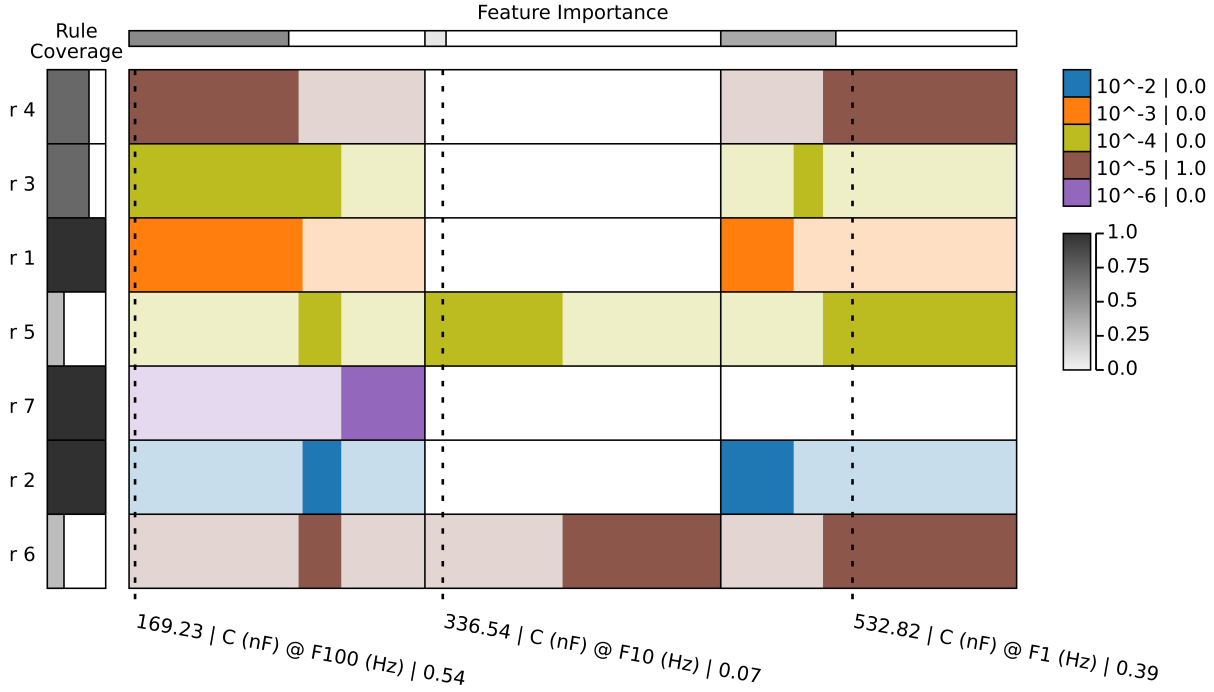


Figure 12 – Multidimensional calibration space visualization using ExMatrix for the classification of a specific instance/sample. Dashed lines indicate the instance values in each attribute (frequency). In this matrix, the rules are ordered by proximity to the instance under analysis, where the rule in the first row (brown/rule r4) is used to classify the instance as class 10^{-5} M. To change the instance's classification from 10^{-5} to 10^{-4} M requires the smallest modification (olive/rule r3 on the second row), while the modification required to change the classification to 10^{-2} M is the largest (blue/rule r2 on the sixth row).

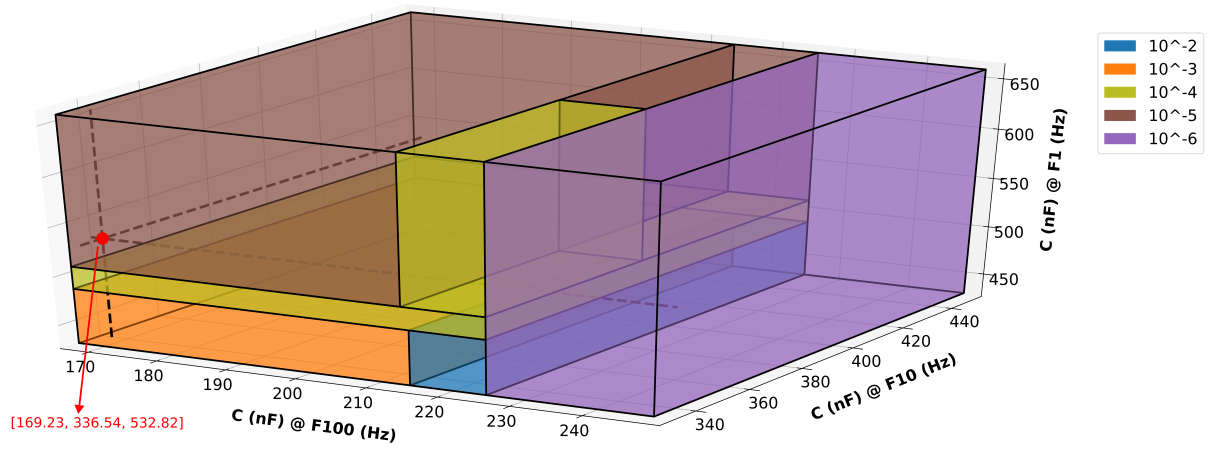


Figure 13 – Multidimensional calibration space shown as a 3D plot. For most concentrations, the parts of the space "used" by the different concentrations are simple. Only for 10^{-4} and 10^{-5} concentrations (olive and brown) the space splitting is more complex.

The coverage of the logic rules obtained from DT models are related to the dataset complexity and DT inference approach. In this paper, we use the Classification And Regression Trees (CART) (BREIMAN *et al.*, 1984) technique to derive DTs. It should be

noted that different trees are obtained by varying the inference parameters. In a model selection [3] experiment (JAMES *et al.*, 2013; KOHAVI, 1995b; JAHANGIRI; RAKHA, 2015), we intentionally varied these parameters to generate multiple (and many) trees and select the one that provides rules with the highest accuracy on KFold Cross Validation using the training set. Then, the parameters selected are used to create a DT model considering the whole training set, and the resulting DT is tested using the test set (unknown samples). In doing so, we selected the best tree in an unbiased manner, its performance can be evaluated, and through the ExMatrix method the Multidimensional Calibration Space can be analyzed.

One limitation that may be inferred from inspecting Figure 13 is in the discretization inherent in the classification rules defined by the DT algorithm, transforming regression into a classification problem (SALMAN; KECMAN, 2012). The samples corresponding to 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, and 10^{-6} M would be located on the "boxes" of round concentrations, and any new sample with an intermediate concentration would be assigned to one of these boxes. This limitation would obviously be circumvented if the dataset contained a much larger number of concentrations.

3.3 Final Remarks

The concept of a multidimensional calibration space introduced here exploits the immense potential from the use of machine learning methods to analyze data. It allows for a predictive power sensors, biosensors and e-tongues. Two features to be highlighted are the use of Decision Trees (DT) algorithms, which permit the generation of interpretable rules, and the visualization of such rules with ExMatrix software. In some aspects this calibration space resembles the multivariate calibration space, and indeed for simple cases with small dimensions they may coincide. However, the concept of a multidimensional calibration space is broader, not only because rules are generated with machine learning but also because different types of data may be treated and visualized. Since the multidimensional calibration space may be applied to any type of data, including images, videos, text, in addition to scientific data, applications beyond pure analytical chemistry can be envisaged. Examples can be surveillance and monitoring systems of various kinds, computer-assisted clinical diagnostics and natural language processing.

[3] Nested KFold Cross-Validation can also be used to assess model performance and choose the best building parameters (VARMA; SIMON, 2006; TSAMARDINOS; RAKHSHANI; LAGANI, 2014).

CHAPTER

4

MULTIVARIATE DATA EXPLANATION – VAX

This chapter (paper *Multivariate Data Explanation by Jumping Emerging Patterns Visualization* [1]) presents VAX, a VA method for data (descriptive) explanations by automated data insights, allowing JEPs visualization and data instances similarity maps. The VAX data explanations involve visualizing descriptive logic rules (JEPs) along with clusters and outliers investigation in DR layouts (instances similarity maps). The VAX method uses RF models' interpretability to explain data extracting, processing, and visualizing JEPs. The application of visual representations of classification models for data explanation is incipient, and the existing approaches employ black-box models. In contrast, JEPs are intrinsically interpretable and yield great descriptive capabilities.

Abstract: Visual Analytics (VA) tools and techniques have been instrumental in supporting users to build better classification models, interpret models' overall logic, and audit results. In a different direction, VA has recently been applied to transform classification models into descriptive mechanisms instead of predictive. The idea is to use such models as surrogates for data patterns, visualizing the model to understand the phenomenon represented by the data. Although very useful and inspiring, the few proposed approaches have opted to use low complex classification models to promote straightforward interpretation, presenting limitations to capture intricate data patterns. In this paper, we present VAX (multiVariate dAta eXplanation), a new VA method to support the identification and visual interpretation of patterns in multivariate datasets. Unlike the existing similar approaches, VAX uses the concept of Jumping Emerging Patterns to identify and aggregate several diversified patterns, producing explanations through logic combinations of data variables. The potential of VAX to interpret complex multivariate datasets is demonstrated through use-cases employing two real-world datasets covering different scenarios.

[1] POPOLIN NETO, M.; PAULOVICH, F. V. Multivariate Data Explanation by Jumping Emerging Patterns Visualization. arXiv preprint arXiv:2106.11112. 2021.

4.1 Introduction

Visualization plays an essential role in multivariate exploratory data analysis (KEIM *et al.*, 2010; SACHA *et al.*, 2014), allowing users to find interesting patterns and formulate hypotheses. In this process, data mining and machine learning techniques can be instrumental, supporting patterns discovery in the visualization or in the data to be displayed by the visualization (DANG; WILKINSON, 2014). In the last decades, lots of effort has been in a different direction, focusing on visualizing data mining and machine learning models, not the data. In general, model visualization focuses on model creation/optimization (ANKERST *et al.*, 1999; TEOH; MA, 2003; DO, 2007; VAN DEN ELZEN; VAN WIJK, 2011; TALBOT *et al.*, 2009; HöFERLIN *et al.*, 2012; LEE; JOHNSON; CHENG, 2016; LIU *et al.*, 2018) or results' interpretation (MING; QU; BERTINI, 2019; DI CASTRO; BERTINI, 2019; ZHAO *et al.*, 2019; RIBEIRO; SINGH; GUESTRIN, 2018; RIBEIRO; SINGH; GUESTRIN, 2016; POPOLIN NETO; PAULOVICH, 2021; CHAN *et al.*, 2020), where global and local explanations aim to support model overview and classification process reasoning (DU; LIU; HU, 2019; LIAO; GRUEN; MILLER, 2020).

The idea of joining these two concepts has been suggested (GLEICHER, 2013; KNITTEL *et al.*, 2020), using machine learning classification models as surrogates to explore an underlying phenomenon represented by data. The core concept is based on the fact that classification models can be used not only for predictive analysis but also for descriptive purposes (TAN; STEINBACH; KUMAR, 2005). If a model is transparent and understandable (LIAO; GRUEN; MILLER, 2020), for instance, through explanatory strategies (CHAN *et al.*, 2020), it can be used as a proxy to understand the patterns in the data. In other words, the idea is to use classification models for descriptive purposes as the primary goal, being data-centered instead of a model hub, where prediction capability is not the target (GLEICHER, 2013; KNITTEL *et al.*, 2020).

In data mining and machine learning fields, this has a long history, with the supervised descriptive rule discovery framework (NOVAK; LAVRAC; WEBB, 2009) unifying concepts such as Emerging Patterns (EP) (DONG; LI, 1999), supporting class differentiation and emerging trends (NOVAK; LAVRAC; WEBB, 2009; LOYOLA-GONZÁLEZ; MEDINA-PÉREZ; CHOO, 2020; GARCíA-VICO *et al.*, 2018; ANWAR; WARNARS; SANCHEZ, 2017). Although visualization is recognized as a powerful component when analyzing a phenomenon through data patterns (NOVAK; LAVRAC; WEBB, 2009; LOYOLA-GONZÁLEZ; MEDINA-PÉREZ; CHOO, 2020), the use of classification models as descriptive tools is still in its infancy in the visualization field. To the best of our knowledge, one solution based on Support Vector Machines (GLEICHER, 2013) and another leveraging Artificial Neural Networks (KNITTEL *et al.*, 2020). Although inspiring approaches, both are based on visual representations of so-called black-box models (MING; QU; BERTINI, 2019; DI CASTRO; BERTINI, 2019; RIBEIRO; SINGH;

GUESTRIN, 2016), which may not provide the descriptive power (GARCíA-VICO *et al.*, 2018; LOYOLA-GONZÁLEZ; MEDINA-PÉREZ; CHOO, 2020) of EP.

This paper presents VAX (multiVariate dAta eXplanation), a new visual analytics method for multivariate data interpretation that employs prediction models to leverages the descriptive power of Jumping Emerging Patterns (JEPs), a special type of EP (KANE; CUISSART; CRÉMILLEUX, 2015; GARCíA-VICO *et al.*, 2018; LOYOLA-GONZÁLEZ; MEDINA-PÉREZ; CHOO, 2020). Using JEPs, class-associated inherent interpretable logic statements are extracted, focusing on data description, not model precision. These are then concisely displayed using a compact matrix metaphor and dimensionality reduction layouts, supporting different analytical tasks involving pattern analysis and data content, revealing intricate and complex information that may be otherwise challenging to discover using usual exploratory approaches.

In summary, the main contributions of this paper are:

- A new method for JEPs visualization, where a matrix metaphor is used to display patterns as rows, variables as columns, and data information through histograms in the cells;
- A new strategy for JEPs selection and aggregation from random decision trees that helps to summarize large sets of patterns while representing the entire data set; and
- An instance map for analyzing data instances from the perspective of the discovered patterns, composing an analytical cycle that goes from data to patterns and from patterns to data.

The remainder of the paper is organized as follows. Section 4.2 covers the literature in classification model visualization for descriptive analysis, discussing the current limitations and positioning our solution. Section 4.3 details our proposed approach, showing how JEPs are extracted, aggregated, and visualized. Section 4.4 presents two different use-cases explaining how to use our solution for data explanation. Finally, Section 4.5 lists our limitations and Section 4.6 outlines our conclusions and future work.

4.2 Related Work

In the visualization literature, the idea of using classification models as descriptive tools instead of predictive engines, using them as proxies to understand or describe multivariate data patterns, is a new concept. Here, we arrange the existing approaches into two groups, model specific (GLEICHER, 2013; KNITTEL *et al.*, 2020) and Emerging Patterns (DONG; LI, 1999; NOVAK; LAVRAC; WEBB, 2009; GARCíA-VICO *et al.*, 2018; LOYOLA-GONZÁLEZ; MEDINA-PÉREZ; CHOO, 2020).

4.2.1 Model Specific

Different classification models have been used for descriptive purposes, for instance, Support Vector Machines (SVM) (GLEICHER, 2013) and Artificial Neural Networks (ANN) (KNITTEL *et al.*, 2020). Since these models are black-boxes (MING; QU; BERTINI, 2019; DI CASTRO; BERTINI, 2019; RIBEIRO; SINGH; GUESTRIN, 2016), they require model-specific solutions to reach interpretability and serve for multivariate data explanation.

In Explainers (GLEICHER, 2013), Dimensionality Reduction (DR) (NONATO; AUPETIT, 2019) layouts are created using linear functions from SVM models. These functions combine different variables, and a heuristic is applied to narrow down the potential combinations so the analyst can filter and select the functions of interest. Explainers allows analysts to reason about the projected data points arrangement analyzing the linear function used to create the layout. It is an inspiring and pioneer approach but is limited to present patterns resulting from linear combinations of up to three variables, missing patterns in more complex non-linear associations. In our approach, the patterns extracted can involve more than three variables and represent non-linear relationships among instances and their classes. To allow that, we use the concept of Emerging Patterns (DONG; LI, 1999; NOVAK; LAVRAC; WEBB, 2009; GARCíA-VICO *et al.*, 2018; LOYOLA-GONZÁLEZ; MEDINA-PÉREZ; CHOO, 2020) to extract patterns with different variables combinations. Furthermore, patterns are also used to create instances DR layouts, enabling analysis involving patterns and data instances.

Visual Neural Decomposition (VND) (KNITTEL *et al.*, 2020) enables multivariate data explanation by visually presenting ANN decompositions. In VND, neural network hidden node weights displayed through stacked bars are used to show the relations between variable ranges and classes (e.g., class A with a threshold probability). The nodes are organized in cards containing variables ordered by importance to the node. Although VND can show non-linear relationships among instances and a particular class, supporting analysis with more than three variables, the captured patterns' complexity is bounded by the simple neural network architecture (one hidden layer) employed to allow interpretability. Also, their ordering and class-specific visualizations can make difficult the analysis of multiple classes at once. In our approach, since we use inherent interpretable logic statements from Emerging Patterns, complex relations can be captured, and the patterns for multiple classes can be concisely displayed since they are arranged in a compact way. Another positive aspect of our approach is capturing patterns with maximum confidence (and statistical significance) so that a specific class's patterns do not support instances from another class. Thus different from VND, patterns are not restricted to a class probability threshold.

Explainers and VND are inspiring approaches but lack the explanatory power of

Emerging Patterns (DONG; LI, 1999; NOVAK; LAVRAC; WEBB, 2009; GARCíA-VICO *et al.*, 2018; LOYOLA-GONZÁLEZ; MEDINA-PÉREZ; CHOO, 2020), where one of the main goals is to obtain patterns for data explanation.

4.2.2 Emerging Patterns

Emerging Patterns (EP) consist of class associated relational statements among variables, providing classes differentiation and emerging trends (DONG; LI, 1999; NOVAK; LAVRAC; WEBB, 2009; GARCíA-VICO *et al.*, 2018; ANWAR; WARNARS; SANCHEZ, 2017). Decision Trees (BREIMAN *et al.*, 1984; TAN; STEINBACH; KUMAR, 2005) can be employed for extracting these patterns (NOVAK; LAVRAC; WEBB, 2009; DONG; LI, 1999; GARCíA-VICO *et al.*, 2018; LOYOLA-GONZÁLEZ; MEDINA-PÉREZ; CHOO, 2020). To obtain many diversified (GARCíA-VICO *et al.*, 2018; LOYOLA-GONZÁLEZ; MEDINA-PÉREZ; CHOO, 2020) and expressive patterns, our approach extracts Jumping Emerging Patterns (JEPs) (KANE; CUISSART; CRÉMILLEUX, 2015; GARCíA-VICO *et al.*, 2018; LOYOLA-GONZÁLEZ; MEDINA-PÉREZ; CHOO, 2020) from random Decision Trees based on Random Forest (GARCíA-BORROTO; MARTíNEZ-TRINIDAD; CARRASCO-OCHOA, 2015; LOYOLA-GONZáLEZ *et al.*, 2019; LOYOLA-GONZÁLEZ; MEDINA-PÉREZ; CHOO, 2020), post-processing these patterns by a selection and aggregation strategy. JEPs are a particular case of EP where the confidence is maximum, that is, when the mined patterns are class-exclusive, not supporting instances of different classes.

Visualization is a powerful component when analyzing a phenomenon through data patterns (NOVAK; LAVRAC; WEBB, 2009; LOYOLA-GONZÁLEZ; MEDINA-PÉREZ; CHOO, 2020). In this context, two crucial aspects should be supported, content and multi-class investigation, being the ability to show patterns content (e.g., data distribution) and be used on multi-class problems (more than two classes) (NOVAK; LAVRAC; WEBB, 2009). Some approaches address the visualization of pattern properties (e.g., support) through visual markers (NOVAK; LAVRAC; WEBB, 2009), but lacking patterns' content representation. Visualizing Subgroup Distribution (VSD) (GAMBERGER; LAVRAC; WETTSCHERECK, 2002) presents patterns as line plots for continuous variables and binary class problems (two classes). Despite being intuitive, VSD (GAMBERGER; LAVRAC; WETTSCHERECK, 2002) is not suitable for multi-class domains, and multiple variables visualization is an issue (NOVAK; LAVRAC; WEBB, 2009). Our approach satisfies both content and multi-class requirements, presenting patterns content using histograms and classes mapped as categorical colors.

EPs can also be seen as descriptive logic rules (GARCíA-VICO *et al.*, 2018; NOVAK; LAVRAC; WEBB, 2009), and Decision Trees are a well-recognized method for consistent logic rules generation (MIRANDA; SARDINHA; CERRI, 2021). Disjoint and

consistent logic rules are very interpretable (MIRANDA; SARDINHA; CERRI, 2021; LAKKARAJU; BACH; LESKOVEC, 2016; FüRNKRANZ; GAMBERGER; LAVRAC, 2012), therefore popular in visual analytics solutions for model interpretability (POPOLIN NETO; PAULOVICH, 2021; MING; QU; BERTINI, 2019; RIBEIRO; SINGH; GUESTRIN, 2018; GUIDOTTI *et al.*, 2018a; LAKKARAJU; BACH; LESKOVEC, 2016). Such solutions usually support global and local model interpretation aiming to explain the model itself, and its decisions (DU; LIU; HU, 2019). RuleMatrix (MING; QU; BERTINI, 2019) and ExMatrix (POPOLIN NETO; PAULOVICH, 2021) are two approaches that visually present rules in a matrix format. Despite great solutions, both are model-centered, where rules are primarily used in explaining models. Our solution also leverages a matrix metaphor, but our goal is to support explanations of multivariate data, not models. Although serving as an inspiration to our solution, RuleMatrix only presents global histograms, and more important, variables (columns) order are fixed. ExMatrix and VAX are based on matrix visualization, in which rows and columns order plays a significant role (CHEN *et al.*, 2004; WU; TZENG; CHEN, 2008). However, ExMatrix does not convey data distribution. VAX supports global and local histograms visualization, where patterns and variables can be ordered to produce meaningful visual representations. Furthermore, VAX integrates visualization of descriptive logic rules (JEPs) with DR layouts using dataset extension (PéREZ *et al.*, 2015). Our solution provides a powerful visual analytics method focusing on data analysis, instead of model explanations as RuleMatrix and ExMatrix.

4.3 Methodology

This section presents VAX (multiVariate dAta eXplanation), a new multidimensional data explanation approach that combines JEPs visualization and DR layouts to support data pattern discovery and interpretation. To reach the general objective – multivariate data explanation, VAX enables the five automated data insights presented in Table 4. In this way, VAX can support exploratory tasks (LAW; ENDERT; STASKO, 2020) inside the visual analytics process (KEIM *et al.*, 2010; SACHA *et al.*, 2014). We implement **I1 - Visual motifs** and **I2 - Distribution** using a matrix-like visual metaphor based on matrix visualization guidelines (CHEN *et al.*, 2004; WU; TZENG; CHEN, 2008). Since the objects of analysis are descriptive patterns (JEPs), we set these as rows (**I1**) and place the used variables as columns with matrix cells presenting global and local histograms (**I2**). Such arrangement combines the strengths of model explanations approaches (POPOLIN NETO; PAULOVICH, 2021; MING; QU; BERTINI, 2019), improving them towards data interpretation. The insights **I3 - Cluster** and **I4 - Outlier** are addressed by instances similarity maps, projecting data instances as viewed through the JEPs lens using a DR technique, where clusters (**I3**) and outliers (**I4**) can be observed.

Finally, **I5 - Compound fact** is reached through combining the different proposed visual representations where clusters and outliers (**I3** and **I4**) can be explained by visualizing the patterns along with variables distributions (**I1** and **I2**). Figure 14 shows VAX pipeline, where **1**: JEPs are extracted using random Decision Trees (DTs) based on Random Forest (GARCíA-BORROTO; MARTíNEZ-TRINIDAD; CARRASCO-OCHOA, 2015; LOYOLA-GONZÁLEZ; MEDINA-PÉREZ; CHOO, 2020), and the more relevant patterns are selected and aggregated following a well-defined strategy; **2**: The resulting aggregated patterns are then visualized using a matrix metaphor (**I1** and **I2**); **3**: Data instances, as viewed by the patterns, are displayed using a DR technique creating an instances similarity map (**I3** and **I4**); By exploring the visualizations, **4**: how patterns are related to instances and **5**: how instances are connected with patterns (**I5**). Hence, VAX involves two key aspects, JEPs and instances similarity maps, further discussed in the following subsections.

Table 4 – Automated data insights.

Insight Type (LAW; ENDERT; STASKO, 2020)	
I1	**Visual motifs.** Unique/special/specific patterns, being but not only custom visual metaphors, representing a particular notion/structure on data.
I2	**Distribution.** Variables values distribution, such as histograms plots.
I3	**Cluster.** Instances group, like a set of points relative closed to each other on a scatter plot.
I4	**Outlier.** Particular instance with distinct variables values to the distribution, such as an instance relative a part from other instances in a scatter plot.
I5	**Compound fact.** Meaningful composition of two or more insights types.

4.3.1 Jumping Emerging Patterns

JEP (Jumping Emerging Pattern) is a particular type of EP (Emerging Pattern) (GARCíA-VICO *et al.*, 2018; LOYOLA-GONZÁLEZ; MEDINA-PÉREZ; CHOO, 2020). VAX arranges JEPs extraction, selection, aggregation, and visualization.

4.3.1.1 Definitions

EP (Emerging Pattern) was defined by Dong and Li (1999). In formal terms, given a class-labeled dataset X and its set of variables V, a pattern p is a conjunction of selectors $p = \{i_1, i_2, ..., i_{|V|}\}$ (logical complex), each one defining a relational statement in the form of $i_v = v \# S$. In the statement defined by a selector i_v, S consists one or more possible values for variable $v \in V$, whereas the relational operator $\#$ can be $\in, \notin, >, <, \geq$, or

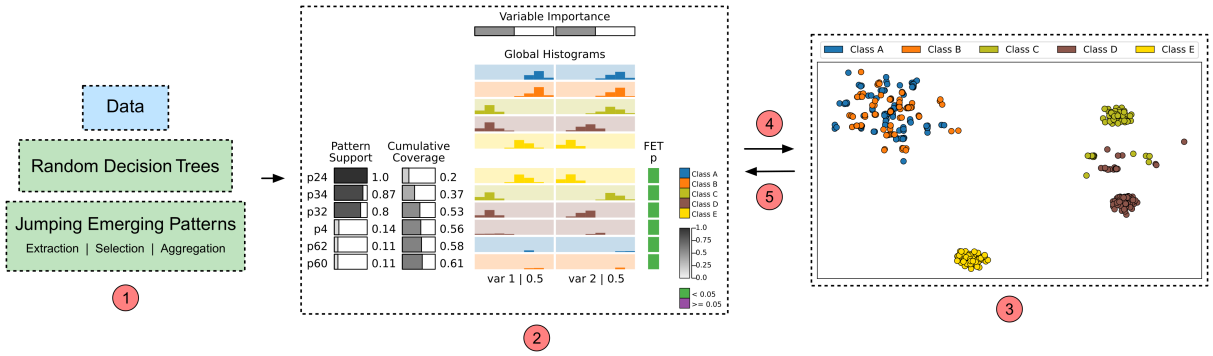


Figure 14 – Data explanation pipeline based on automated data insights (Table 4). **1**: JEPs (Jumping Emerging Patterns) are extracted from random Decision Trees and then selected and aggregated following a well-defined strategy. **2**: JEPs are visualized into a matrix-like visual metaphor using global and local histograms (**I1** and **I2**). **3**: Two-dimensional instances maps are built to show the instances' similarity relationships from the patterns' perspective (**I3** and **I4**). **4**: From the matrix visualization, JEPs can be further analyzed by inspecting the supported instances on the map (**I5**). **5**: From the map, instances can be further investigated by inspecting the JEPs in which they are supported (**I5**).

$\leq$ (MICHALSKI; STEPP, 1982; GARCíA-VICO *et al.*, 2018). In this paper we work with patterns arranging selectors in the form $v \in S$, being $S = [a,b]$ with a and $b \in \mathbb{R}$, so defining a real set bounded by a and b. An instance $x \in X$ (a.k.a. transaction, sample, example, or item) is supported by a pattern p if it satisfies all selectors of p (MICHALSKI; STEPP, 1982; DONG; LI, 1999; GARCíA-VICO *et al.*, 2018). Moreover, pattern p is considered an EP by having a Growth Rate $GR(p)$ higher than a given threshold (≥ 1) (MICHALSKI; STEPP, 1982; DONG; LI, 1999; GARCíA-VICO *et al.*, 2018), defined by

$$GR(p) = \begin{cases} 0 & \text{If } Supp_{X_1}(p) = Supp_{X_2}(p) = 0 \\ \infty & \text{If } Supp_{X_2}(p) \neq 0 \wedge Supp_{X_1}(p) = 0 \\ \frac{Supp_{X_2}(p)}{Supp_{X_1}(p)} & \text{otherwise} \end{cases} \tag{4.1}$$

where $Supp_{X_1}(p)$ is the support of p in the dataset $X_1 \subset X$, and $Supp_{X_2}(p)$ the support of p in the dataset $X_2 \subset X$, given by

$$Supp_{X_o}(p) = \frac{count_{X_o}(p)}{|X_o|} \tag{4.2}$$

with X_o being the subset of instances from class o, $count_{X_o}(p)$ the number of instances from X_o supported by p, and $|X_o|$ the cardinality of X_o (DONG; LI, 1999; NOVAK; LAVRAC; WEBB, 2009; GARCíA-VICO *et al.*, 2018). On binary problems (two classes), X_1 contains instances of one class, while X_2 is composed of other class instances. In multi-class problems (more than two classes), *One-vs-All* strategy can be used (GARCíA-VICO *et al.*, 2018), where X_2 contains the instances for a particular class, and X_1 contains the

instances of all remaining classes. In summary, EP core idea is to discover patterns whose support increases (Growth Rate) from a dataset X_1 to a dataset X_2 (DONG; LI, 1999; NOVAK; LAVRAC; WEBB, 2009; GARCíA-VICO *et al.*, 2018).

For illustration, we take the synthetic dataset X_S presented in Figure 15. This synthetic dataset X_S is composed of 500 instances $X_S = \{x_1, x_2, ..., x_{500}\}$, described by 2 real variables (var_1, var_2), and equally distributed among Class A, B, C, D, and E (100 instances per class). Three clusters can be spotted among data instances. Class A and B strongly overlap, Class C and D are adjacent, and Class E is completely separated.

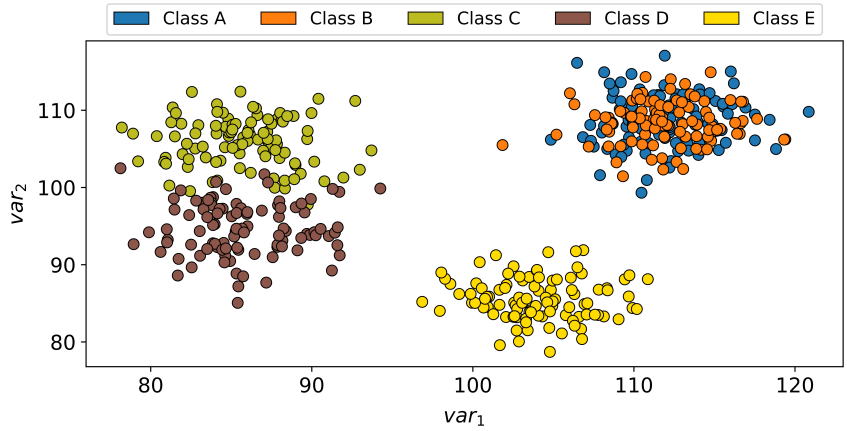


Figure 15 – The synthetic dataset X_S used to illustrate our approach.

For the synthetic dataset $X_S = X_A \cup X_B \cup X_C \cup X_D \cup X_E$, the EP $p_{ex} = \{var_1 \in [96, 120], var_2 \in [78, 95]\}$ for Class E has $GR(p_{ex}) = \infty$, since $Supp_{X_E}(p_{ex}) = 1.0$ and $Supp_{X_{A \cup B \cup C \cup D}}(p_{ex}) = 0$. In other words, pattern p_{ex} is an EP by increasing the support from dataset $X_{A \cup B \cup C \cup D}$ to dataset X_E. The pattern p_{ex} supports all instances from Class E ($count_{X_E}(p_{ex}) = |X_E| = 100$), but none instances from Class A, B, C, or D ($count_{X_{A \cup B \cup C \cup D}}(p_{ex}) = 0$ and $|X_{A \cup B \cup C \cup D}| = 400$).

There are different types of EPs based on the relationships between variables (GARCíA-VICO *et al.*, 2018; LOYOLA-GONZÁLEZ; MEDINA-PÉREZ; CHOO, 2020). **JEPs (Jumping Emerging Patterns) are EPs with GR $= \infty$, that is, patterns that support instances of a single class, providing a great discriminative power among classes** (GARCíA-VICO *et al.*, 2018; KANE; CUISSART; CRÉMILLEUX, 2015)**.** The pattern $p_{ex} = \{var_1 \in [96, 120], var_2 \in [78, 95]\}$ is a JEP, since $GR(p_{ex}) = \infty$. It is worth mentioning that JEPs also present maximum value (1.0) of confidence (pattern precision) once they do not support instances of different classes (GARCíA-VICO *et al.*, 2018).

From this point forward, we use the notations in Table 5 and the synthetic dataset X_S in Figure 15 to help explain our approach and demonstrate how it performs in overlap, adjacent, and separated data clusters.

Table 5 – Summary of notation.

Notation	Description
X	Class-labeled dataset.
X_o	The subset of instances from class o in X, that is $X_o \subset X$.
V	Set of variables from X.
v	A variable $v \in V$.
X^v	All values for variabel v in X.
p_j	An EP (Emerging Pattern) j arranging a conjunction of selectors $p_j = \{i_1, i_2, ..., i_{\|V\|}\}$.
i_v	A selector in the form $v \in S$, being S a real set, as $v \in [a,b]$ with a and $b \in \mathbb{R}$.
$S_v^{p_j}$	The real set from the selector for variable v in pattern p_j.
p_j^{class}	The resulting (associated) class of pattern p_j.
P	Set of EPs (Emerging Patterns).

4.3.1.2 Extraction

Mining EPs is an NP-Hard problem resulting from an exponential number of candidate patterns if the number of variables grows (WANG *et al.*, 2004; LI *et al.*, 2004; GARCíA-VICO *et al.*, 2018; LOYOLA-GONZÁLEZ; MEDINA-PÉREZ; CHOO, 2020). EPs can be mined using different strategies (NOVAK; LAVRAC; WEBB, 2009; GARCíA-VICO *et al.*, 2018; LOYOLA-GONZÁLEZ; MEDINA-PÉREZ; CHOO, 2020), such as DT (Decision Tree) models (BREIMAN *et al.*, 1984; TAN; STEINBACH; KUMAR, 2005). The idea is to build varied DTs using diversity factors and then extract a pattern from each decision path (root to leaf node) (NOVAK; LAVRAC; WEBB, 2009; DONG; LI, 1999; GARCíA-VICO *et al.*, 2018; LOYOLA-GONZÁLEZ; MEDINA-PÉREZ; CHOO, 2020). DT models built upon random aspects have proved to be very useful on classification tasks (HO, 1995; HO, 1998; WANG; WANG; ZHAO, 2010; JAMES *et al.*, 2013), like in Random Forest (RF) models (BREIMAN, 2001; BIAU; SCORNET, 2016).

In our approach, to extract EPs from a class-labeled dataset X, we use the Random Forest miner (RFm) (GARCíA-BORROTO; MARTíNEZ-TRINIDAD; CARRASCO-OCHOA, 2015; LOYOLA-GONZáLEZ *et al.*, 2019; LOYOLA-GONZÁLEZ; MEDINA-PÉREZ; CHOO, 2020) for building random DT models from X. Being V the variables set of X, the RFm method (GARCíA-BORROTO; MARTíNEZ-TRINIDAD; CARRASCO-OCHOA, 2015; LOYOLA-GONZáLEZ *et al.*, 2019) builds k unpruned DTs by selecting and analyzing a random subset of variables at each internal node creation, where variables subset size equals $\log_2 |V|$. Unlike the RF proposed in Breiman (2001), the bagging process (random selection of training instances) (JAMES *et al.*, 2013) is not used in RFm, to avoid hidden dependencies among patterns and data instances (GARCíA-BORROTO; MARTíNEZ-TRINIDAD; CARRASCO-OCHOA, 2015; LOYOLA-GONZÁLEZ; MEDINA-PÉREZ; CHOO, 2020). The RFm has been proved

to be an excellent strategy for obtaining diversified patterns (GARCíA-BORROTO; MARTíNEZ-TRINIDAD; CARRASCO-OCHOA, 2015).

The Algorithm 1 presents the extraction process, where a class-labeled dataset X and the number of trees k are input parameters. Function DecisionTree() creates a DT model (BREIMAN *et al.*, 1984; BREIMAN, 2001; TAN; STEINBACH; KUMAR, 2005; JAMES *et al.*, 2013) while ExtractPatterns() extracts a pattern p for each decision path (root to leaf node) from a DT (LOYOLA-GONZÁLEZ; MEDINA-PÉREZ; CHOO, 2020). Since data explanation is the target, the entire X is employed to build the random DT models. The whole dataset usage is found in data explanations solutions like VND (KNITTEL *et al.*, 2020) and Explainers (GLEICHER, 2013), rather than splitting into training and test sets for predictive model evaluation (JAMES *et al.*, 2013).

Algorithm 1: EPs extraction using random DTs.

Input: Class-labeled Dataset - X, Number of Trees - k
Output: Emerging Patterns - P
$P \leftarrow \varnothing$;
for *1 to k* **do**
  $DT \leftarrow DecisionTree(Dataset = X,\ Subset\ Size = \log_2)$;
  $P \leftarrow P \cup ExtractPatterns(DT)$;
end

Table 6 presents 6 patterns among 13975 extracted by Algorithm 1 execution taking as inputs the synthetic dataset X_S and number of trees $k = 128$.

Table 6 – EPs (Emerging Patterns) extracted using Algorithm 1 on the synthetic dataset X_S. A total of 13975 were extracted using 128 random DTs (Decision Trees). The EPs presented here are for Class B, with $GR = \infty$, support of 0.11, and p-value for statistical significance of 10^{-9}.

Pattern	Selectors
p_{2920}	$\{var_1 \in \{95.91, 111.29\},\ var_2 \in \{108.77, 109.75\}\}$
p_{3134}	$\{var_1 \in \{95.55, 111.29\},\ var_2 \in \{108.77, 109.75\}\}$
p_{6001}	$\{var_1 \in \{97.27, 111.29\},\ var_2 \in \{108.77, 109.75\}\}$
p_{8470}	$\{var_1 \in \{106.57, 111.29\},\ var_2 \in \{108.77, 109.75\}\}$
p_{9100}	$\{var_1 \in \{107.43, 111.29\},\ var_2 \in \{108.77, 109.74\}\}$
p_{12834}	$\{var_1 \in \{98.78, 111.29\},\ var_2 \in \{108.76, 109.75\}\}$

One interesting factor is that the Fisher Exact Test (FET) can be applied to compute statistical significance per pattern (BOULESTEIX; TUTZ; STRIMMER, 2003; NOVAK; LAVRAC; WEBB, 2009; LOEKITO; BAILEY, 2009). Values above the significance p-value (usually 0.05) imply on the null-hypothesis acceptation that there is no association between the pattern and a class (LOEKITO; BAILEY, 2009).

The 6 patterns of Table 6 are associated to Class B, with $GR = \infty$ (that is JEP and confidence of 1.0), support of 0.11 (11 out 100 Class B instances, Equation 4.2), and p-value for statistical significance of 10^{-9} (< 0.05 means significant). Moreover, such patterns support the same instances. Despite the ability to provide diversified patterns (GARCíA-BORROTO; MARTíNEZ-TRINIDAD; CARRASCO-OCHOA, 2015; LOYOLA-GONZÁLEZ; MEDINA-PÉREZ; CHOO, 2020), DT models created over random factors can extract redundant patterns along with finding the varied ones (GARCíA-BORROTO; MARTíNEZ-TRINIDAD; CARRASCO-OCHOA, 2015; LOYOLA-GONZáLEZ *et al.*, 2019).

4.3.1.3 Selection and Aggregation

Once all patterns are extracted from k random DT models (Algorithm 1), originating a set of patterns P, these must be selected and aggregated. Based on a selection method (LOYOLA-GONZáLEZ *et al.*, 2019), we analyze patterns in set P from the highest to the lowest support value, choosing a subset so each pattern provides 1.0 for confidence value ($GR = \infty$) and does not support data instances supported by other patterns. Thus, we select JEPs ($GR = \infty$ and confidence value 1.0) that support data instances individually. The main difference between our strategy and the one presented in Loyola-González *et al.* (2019) is that we do not discard supplementary patterns (different patterns that support the same instances). The supplementary patterns are aggregated, not losing their information.

The selection and aggregation process is summarized in Algorithm 2. Using an iterative greedy process on set of patterns P (resulted from Algorithm 1) ordered by decreasing support, we select as a candidate the first pattern (highest support) $p_{candidate}$. If $p_{candidate}$ confidence value equals to 1.0 it is selected as p_{pivot} and aggregated with the subset of patterns $P' \subset P$ that support the same instances.

For aggregation procedure, given a pattern $p_a \in P'$ supporting the same instances of p_{pivot}, all patterns selectors for the same variable are aggregated as the intersection of the two defined real sets ($S^{p_{pivot}}$ and S^{p_a}). So, supposing p_{pivot} and p_a having selectors for a variable $v \in V$, the aggregation is $S_v^{p_{pivot}} \cap S_v^{p_a}$. For the case where the selector for a variable $v \in V$ is found in only one pattern (p_{pivot} or p_a), the missing real set is $[min(X^v), max(X^v)]$, being $min(X^v)$ and $max(X^v)$ the minimum and maximum values for variabel v in X. After P' patterns aggregation with p_{pivot}, P' patterns are removed from P. In this way, complementary patterns turn into a single pattern.

The next candidate pattern $p_{candidate} \in P$ is then analyzed. If $p_{candidate}$ confidence value differs from 1.0, $p_{candidate}$ is discarded, as result of not being a JEP. However, if confidence value equals to 1.0, the $p_{candidate}$ supported instances are investigated. If $p_{candidate}$ supports at least one instance already supported by the patterns selected and aggregated

before, $p_{candidate}$ is discarded, removing a redundant pattern. If $p_{candidate}$ supports only instances not supported by the patterns selected and aggregated before, it is selected as p_{pivot} and aggregated with the set of patterns P' that support the same instances, removing from P this set of patterns after aggregation. This process is repeated until the end of P is reached, selecting and aggregating complementary patterns and discarding the redundant ones.

Algorithm 2: JEPs selection and aggregation.

```
Input: P - Emerging Patterns
Output: P - Jumping Emerging Patterns
SI ← ∅;
P ← OrderByDecreasingSupport(P);
while p_candidate ← Next(P) do
    if Confidence(p_candidate) = 1.0 and SupportedInstances(p_candidate) ∉ SI then
        p_pivot ← p_candidate;
        SI ← SI ∪ SupportedInstances(p_pivot);
        foreach p_a ∈ P do
            if SupportedInstances(p_pivot) = SupportedInstances(p_a) then
                p_pivot ← Aggregate(p_pivot, p_a);
                Remove(p_a, P);
            end
        end
    else
        Remove(p_candidate, P);
    end
end
```

The selection and aggregation process summarized in Algorithm 2 results on meaningful JEPs supporting all instances from a dataset X, explaining it through high support patterns, discarding the redundant ones, and using low support patterns to explain outlier instances.

The 13975 extracted patterns by Algorithm 1 for the synthetic dataset X_S (number of trees $k = 128$) were processed by Algorithm 2 for selection and aggregation. 67 patterns are taken as pivots, 2073 were aggregated, and 11835 were discarded. Therefore, Algorithm 2 resulted in 67 JEPs supporting all instances from the synthetic dataset X_S. The patterns in Table 6 were aggregated, leading to the pattern $p_{60} = \{var_1 \in [107.43, 111.29], var_2 \in [108.77, 109.74]\}$. As presented in Figure 16, the 67 JEPs delimit regions in the bidimensional space formed by variables var_1 and var_2. The region delimited by pattern p_{60} has edges colored in black.

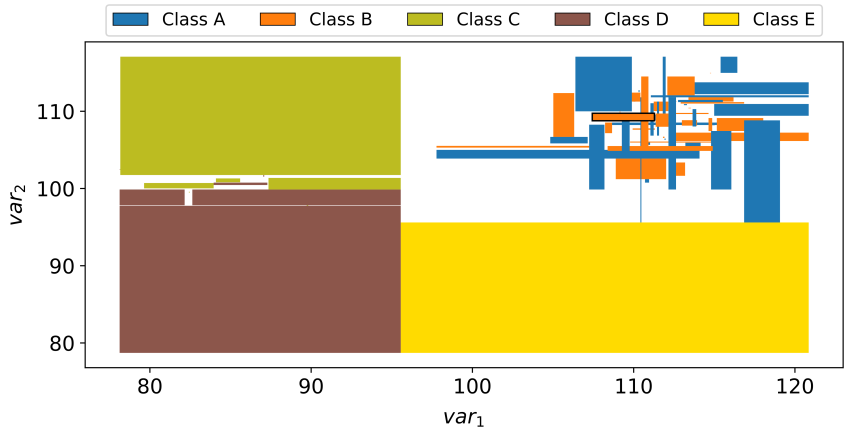


Figure 16 – The regions delimited by the 67 JEPs resulted from the selection and aggregation process via Algorithm 2, taking as input 13975 patterns extracted by Algorithm 1 with synthetic dataset X_S and number of trees $k = 128$. The region with edges colored in black is delimited by pattern $p_{60} = \{var_1 \in [107.43, 111.29], var_2 \in [108.77, 109.74]\}$.

4.3.1.4 *Visualization*

The selected and aggregated JEPs (Algorithm 2) are displayed using the matrix visual metaphor (insight **I1**) in Figure 17. It arranges patterns as rows (**1**) and variables as columns (**2**). Matrix cells (**3**) are divided into bins presenting (local) variables' histograms (MUNZNER, 2014) (insight **I2**) considering instances supported by a particular pattern, where classes are mapped to categorical colors. Cells' width outlines the minimum and maximum (left to right) values for a variable. Similarly, global histograms (insight **I2**) for each class are laid out on the top of the matrix (**4**), exhibiting all instances from a specific class per row. In both global and local histograms, the bins are normalized between 0 and 1, giving global and local ratios of instances for a specific class. Cells' height maps the minimum and maximum (bottom to top) ratio value (0 to 1). The number of bins is determined based on the Freedman-Diaconis rule (FREEDMAN; DIACONIS, 1981; CORRELL *et al.*, 2019), although it can also be freely defined. If a pattern (row) does not have a selector for a variable (column), the respective cell is left empty (no histogram) by default (also available if needed). The patterns support is mapped to a column placed on the matrix left side (**5**). The cumulative dataset coverage is also mapped into a column on the matrix left side (**6**), representing the collective percentage of instances covered from the dataset considering the matrix order (top to bottom). The variable importance is outlined above global histograms and in variables name at the bottom (**7**). Pattern support, cumulative coverage, and variable importance are mapped to color (linear grayscale) and size (rectangular shape width). The FET p-value for each pattern is displayed in a column to the matrix right side (**8**) using a binary color scheme, green for values below p-value = 0.05 (statistically significant) and purple otherwise. The employed matrix-like visual metaphor provides more custom features, such as choosing empty cells color and frames presenting selectors real set.

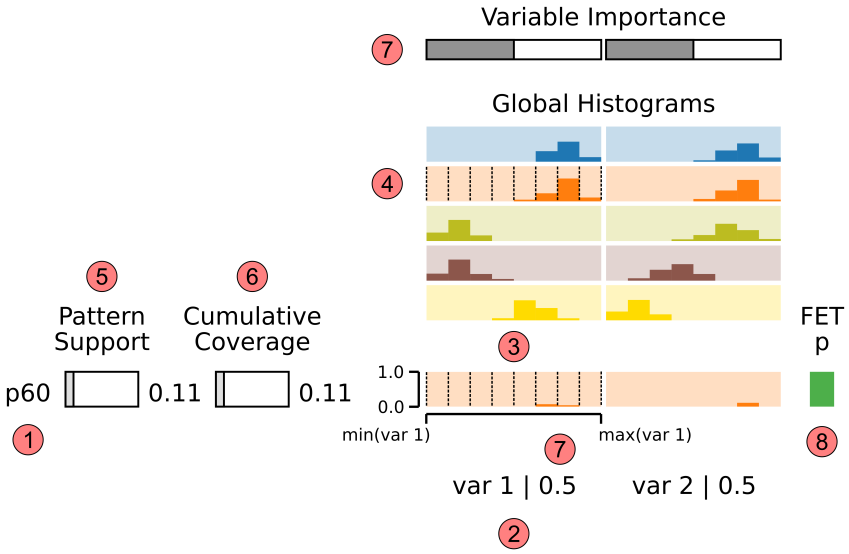


Figure 17 – The matrix-like visual metaphor. **1**: JEPs are displayed as rows. **2**: Variables are arranged as columns. **3**: Cells are divided into bins showing local normalized histograms. **4**: Global histograms (one per class) are placed on the top, also being normalized. **5**: Pattern support. **6**: Cumulative coverage taking the matrix order (top to bottom). **7**: Variable importance. Both pattern support, cumulative coverage, and variable importance are mapped to size and color (grayscale). **8**: FET (Fisher Exact Test) significance value colored as green (statistically significant) or purple (not significant).

There are two key aspects for creating meaningful visual representations of JEPs displayed as a matrix: filtering and ordering. Filtering is a common strategy when several patterns are available (GARCíA-VICO *et al.*, 2018; LOYOLA-GONZÁLEZ; MEDINA-PÉREZ; CHOO, 2020). In our approach, JEPs can be filtered by support value, data coverage, class, and supported instance(s). The latter requires an instance of interest or a subset of instances. Rows and columns order plays a significant role in matrix visualization (CHEN *et al.*, 2004; WU; TZENG; CHEN, 2008). Thus, JEPs (rows) can be ordered by support, class, and class & support. Furthermore, variables (columns) can be ordered by importance, which is calculated based on Paja (2018) following

$$Imp(v,P) = \sum_{j=1}^{|P|} \begin{cases} Supp_{X_o}(p_j) \mid class\ o = p_j^{class} & \text{If } i_v \in p_j \\ 0 & \text{otherwise} \end{cases} \tag{4.3}$$

where the importance of a variable v given a set of JEPs P is the summation of the support from each pattern $p \in P$ having a selector for v. After calculating the importance for all $v \in V$, these are normalized between 0 and 1.

Figure 18 displays 12 JEPs out the 67 resulted from Algorithm 1 and 2 for the synthetic dataset X_S (number of trees $k = 128$). The 67 patterns were filtered by data coverage (p_{24} to p_{57}) and supported instance (p_5). The 12 resulting patterns are ordered by support. The first row contains the pattern p_{24} (selectors $\{var_1 \in [95.55, 120.85], var_2 \in$

[78.72, 95.6]}). This pattern support is maximum, indicated by the filled rectangle on the support column. The cumulative coverage reflects the value of 0.20, as pattern p_{24} supports all 100 instances of Class E from the dataset X_S out of 500 instances. Once pattern p_{24} has selectors on variables var_1 and var_2, its cells show histograms for these variables considering only the instances (local) supported by p_{24}. For this case, global and local histograms are equals since p_{24} supports all instances from Class E. The second row arranges the pattern p_{34}, having a support value of 0.87, meaning that this pattern supports 87% of Class C instances (87 instances out 100). The cumulative coverage indicates the value of 0.37, stating that patterns p_{24} and p_{34} together cover 37% of the instances from dataset X_S (187 out 500). The last pattern p_5 has low support (0.01 supporting 1 Class D instance out 100), being not statistically significant (purple for FET p-value). All the remaining 11 patterns are otherwise significant. From the first pattern p_{24} to the last p_5, about 69% of the dataset X_S is covered (cumulative coverage of 0.69).

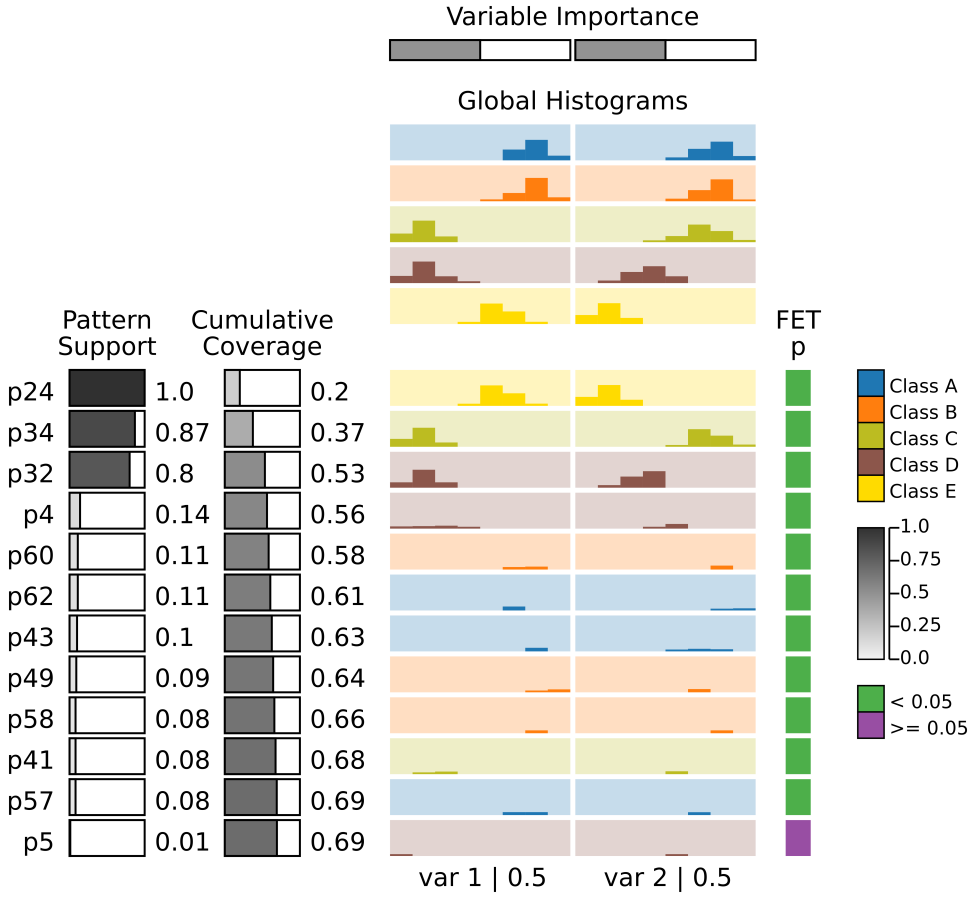


Figure 18 – The 12 JEPs filtered out the 67 obtained for the synthetic dataset X_S by Algorithm 1 (number of trees $k = 128$) and Algorithm 2. Variables values combinations and classes associations are presented, such as the strong pattern (p_{24}) for Class E instances. The pattern p_{24} supports all Class E instances, having median values for variable var_1 and low values for variable var_2. The 12 JEPs cover about 69% of the dataset X_S (cumulative support), where only the low support pattern p_5 is not statistically significant (purple for FET).

Analyzing the JEPs in Figure 18 makes it possible to observe variables values combinations associated with classes (insights **I1** and **I2**). Class E has a strong pattern (unique with the highest support) on dataset X_S. The pattern p_{24} supports 100% of Class E instances (support value of 1.0). The instances from this class have median values for variable var_1 and low values for variable var_2, as presented on pattern p_{24} histograms. The patterns p_{34} and p_{32} complete the list of high support patterns. The p_{34} supports 87% of Class C instances, having low values for variable var_1 and high values for variable var_2. The p_{32} supports 80% of Class D instances, arranging low values for variable var_1 and median values for variable var_2. These 3 patterns (p_{24}, p_{34}, and p_{32}) cover more than half (53%, cumulative coverage of 0.53) of the instances from dataset X_S. The pattern p_4 refers to a small instances subset from Class D (14%), differing from pattern p_{32} on variable var_2 (higher values). The patterns providing the highest support for Class A and B (p_{60} and p_{62}) attend only 11% of each class. Class A and B strongly overlap in dataset X_S (Figure 15), so high support patterns for these classes are not attainable. The pattern p_{41} leads to a narrow instances subset from Class C (8%), differing from pattern p_{34} on variable var_2 (lower values). For Class D, the pattern p_5 supports only one instance (1%), being an exception against patterns p_{32} and p_4 on variable var_2 (higher value).

4.3.2 Instances Similarity Map

In order to support clusters and outliers analyses (insights **I3** and **I4**) from a class-labeled dataset X, we use DR layouts leveraging the space extension approach proposed in Pérez *et al.* (2015) to incorporate JEPs perspectives. Being $X \in \mathbb{R}^{n \times d}$, where $n = |X|$ is the number of instances in X, and $d = |V|$ the number of variables of X, the key idea is to create an extended dataset $X' \in \mathbb{R}^{n \times 2d}$ as

$$X' = [X|\widetilde{X}] \tag{4.4}$$

with $\widetilde{X} \in \mathbb{R}^{n \times d}$ composed by repeatedly class centroids (mean values). Then, for the instances subset $X_h \subset X$ belonging to class h, the extension equals to

$$\widetilde{x} = \frac{1}{|X_h|} \sum_{x \in X_h} x \tag{4.5}$$

Since all instances $x \in X$ are supported by only one JEP selected and aggregated by Algorithm 2, we use JEPs as additional classes to extend X. Hence, X' incorporates a new set of variables $\widetilde{V}$, arranging the mean of the instances subset supported by each pattern obtained from Algorithm 2.

A real parameter $\lambda \in [0,1]$ is used to control the gradual transition between the dataset X and the extended part $\widetilde{X}$ (PéREZ *et al.*, 2015), having $X_{weight} = X'W_\lambda$ where

matrix $W_\lambda \in \mathbb{R}^{2d \times 2d}$ is defined by

$$W_\lambda = \begin{pmatrix} (1-\lambda)I & 0 \\ 0 & \lambda I \end{pmatrix} \tag{4.6}$$

After data standardization (z-score), dataset X_{weight} is used along a DR technique to create an instances similarity map (DR layout). By varying the parameter λ, it is possible to get maps from the original dataset X with $\lambda = 0$ to only the extended $\widetilde{X}$ part setting $\lambda = 1$. In this paper, we use the DR technique MDS (KRUSKAL, 1964) for instances maps creation.

For creating the extended version of the synthetic dataset X_S, the 67 JEPs resulting from Algorithm 2 are used as additional classes of instances $x \in X_S$. Therefore, 67 classes are considered in the process (PéREZ *et al.*, 2015). The synthetic dataset extension X'_S has then four variables $V'_S = \{var_1, var_2, \widetilde{var}_1, \widetilde{var}_2\}$, where $\widetilde{var}_1$ and $\widetilde{var}_2$ are the mean values of var_1 and var_2 from each instances subset supported by the 67 patterns. The 11 instances from Class B supported by pattern p_{60} have values 109.59 and 109.19 for variables $\widetilde{var}_1$ and $\widetilde{var}_2$.

Figure 19 presents the instances map applying MDS on $X_{weight} = X'_S W_\lambda$ with $\lambda = 0.70$ and standardization by z-scores. Three clusters can be identified (insight **I3**), one for each high support pattern (p_{24}, p_{34}, and p_{32}). Reasoning about them (insight **I5**), the one formed (insight **I3**) by pattern p_{24} (insight **I1**) for Class E contains instances with higher values for variable var_1 and lower values (majority) variable var_2 (insight **I2**) compared to those established by patterns p_{34} and p_{32} (insight **I3**). The difference between these two latter clusters is found in variable var_2 (insight **I2**), where Class C instances (p_{34}) have higher values than Class D instances (p_{32}). Furthermore, an outlier for Class D can also be analyzed (insight **I4**). It comes from the low support pattern p5 (insight **I1**), and it yields a low value for variable var_1 and a high value for variable var_2 (insight **I2**) in contrast (exception) to the Class D cluster settled by pattern p_{32} (insight **I5**).

The 11 instances supported by pattern p_{60} are also highlighted in Figure 19. Although instances from Class A and B are minor grouped in the similarity map, no major cluster is found for such classes. This behavior is expected since both classes strongly overlap in the synthetic dataset X_S (Figure 15).

The clusters and outliers investigation into instances similarity maps can be used to filter JEPs visualization, where maps are browsed selecting instances of interest obtaining the patterns supporting such instances. The instances similarity maps play an essential role in dealing with many patterns.

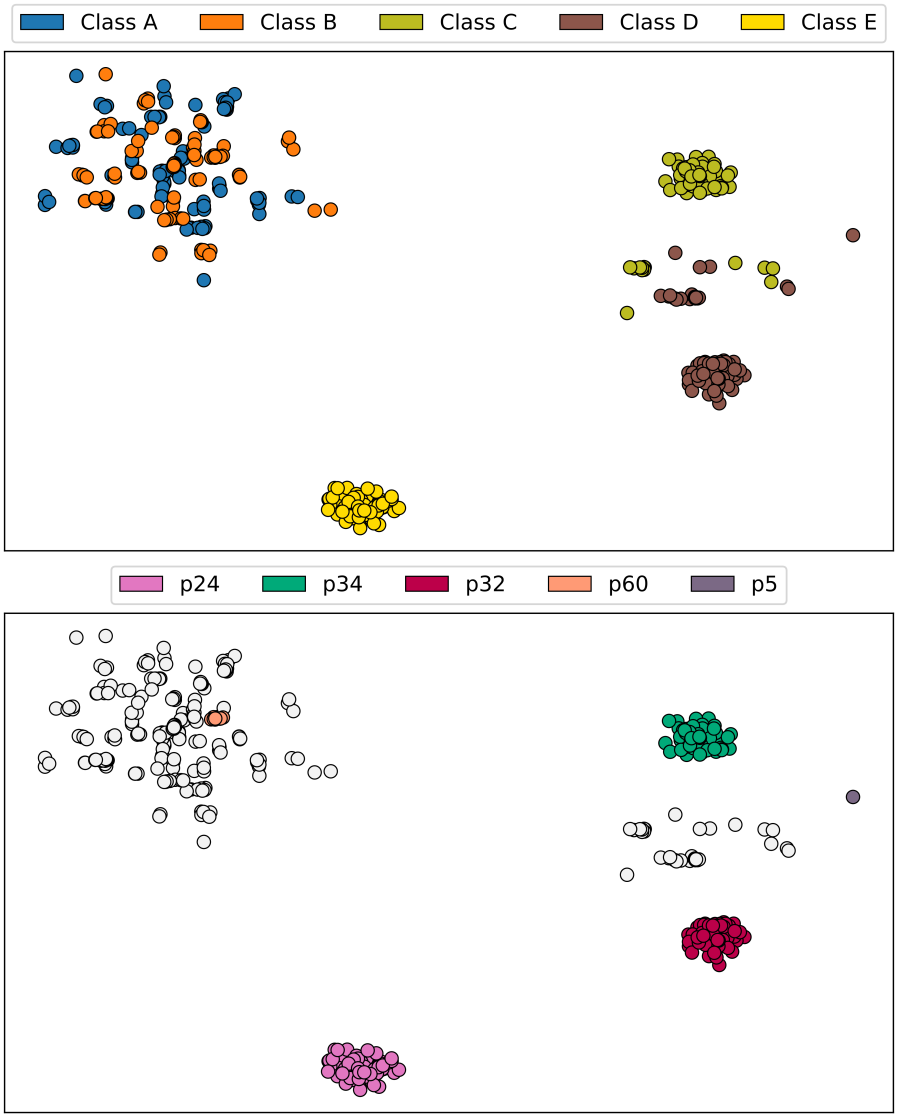


Figure 19 – Instances Similarity Map for the synthetic dataset X_S under JEPs perspectives (67 from Algorithm 2): MDS application in the dataset extension $X_S^{'}$ with JEPs as classes and $\lambda = 0.70$. Clusters can be spotted, formed by high support patterns p_{24}, p_{34}, and p_{32}. A Class D outlier can also be seen, mapped by pattern p_5. The instances subset supported by pattern p_{60} is also emphasized.

4.4 Use Cases

This section presents use-cases involving real-world datasets (see Table 7), showing how to use VAX in different analytical scenarios of multivariate datasets' exploratory analyses. The two use-cases involve datasets with and without ground truth labels (classes). The JEPs and instances similarity maps visualizations provide data explanation, revealing statistically significant patterns considering different variables values combinations along with clusters and outliers investigation. VAX is implemented using Python programming language, being also available as a code package [2] for imminent usage. The source code

[2] <available after publication acceptance>

for the two use-cases is accessible as Python notebook pages [3,4].

Table 7 – Datasets used for VAX evaluation.

Name	Source	Preprocessing
AP VoteCast 2018	Tompson and Benz (2018)	Knittel *et al.* (2020)
World Happiness Report 2019	Helliwell, Layard and Sachs (2019)	-

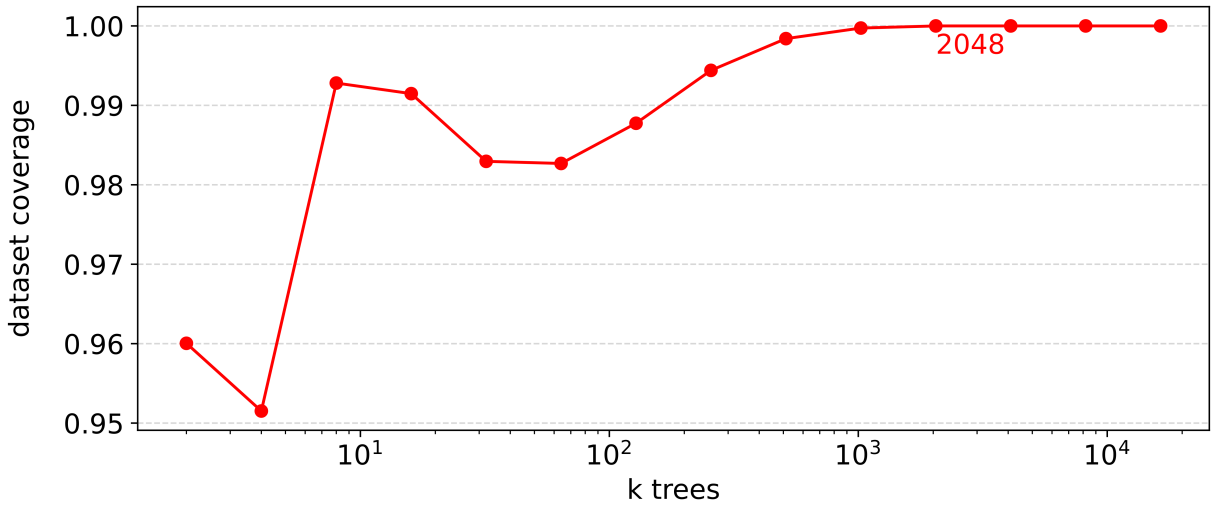
4.4.1 Use Case I – US Presidential Election

The first use case involves the analysis of the 2016 US presidential election using the dataset from a survey conducted in 2018 by the independent social research organization NORC of the University of Chicago (TOMPSON; BENZ, 2018; KNITTEL *et al.*, 2020). It contains the participants' answers about different political and societal aspects of the United States and what candidate they support (Donald Trump or Hillary Clinton). Following the steps described in Knittel *et al.* (2020), the nationally representative subset is used, resulting in 4,913 data instances (registered voters) and 67 variables. After removing missing values keeping only data instances with a revealed vote (Donald Trump or Hillary Clinton), the number of variables is reduced to 60 and instances to 3,754, 43.3% pro-Donald Trump (1,625), and 56.7% pro-Hillary Clinton (2,129).

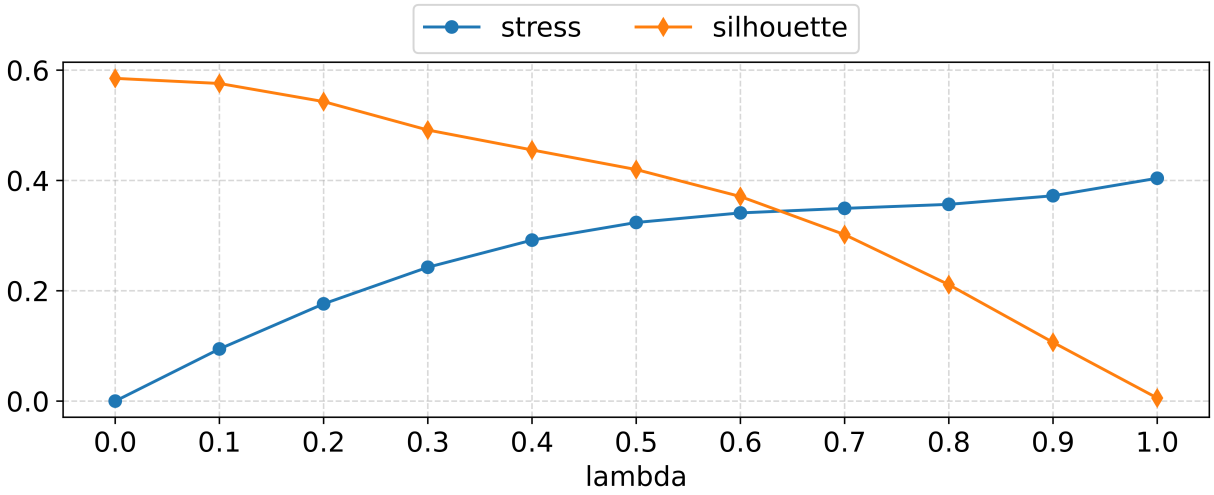
For patterns extraction process (Algorithm 1), considering the voters' orientation as classes, the number of trees k must be set. We aim at the minimum number of trees capable of extracting enough patterns to our selection and aggregation strategy (Algorithm 2) resulting in the coverage of all data instances. Figure 20a presents the obtained data coverage vs. number of trees from 2 to 16384 (2^1 to 2^{14}). 2048 is the minimum number of DTs capable of providing enough diversified patterns resulting in 100% of data instances coverage after selection and aggregation. From 856,900 patterns extracted by Algorithm 1 ($k = 2048$), 255 were selected, 3,042 aggregated, and 853,603 discarded by Algorithm 2.

For instances similarity map creation, the parameter λ must be chosen regarding dataset extension (PéREZ *et al.*, 2015). We select the λ value by analyzing the Kruskal Stress (KRUSKAL, 1964) and Silhouette Coefficient (ROUSSEEUW, 1987). Varying λ, Figure 20b shows the Kruskal Stress (from 0 for none stress to 1 for maximum stress) and the transformed Silhouette Coefficient (cluster consistency). The latter is normalized and inverted into values between 0 and 1 (with 0 being the best value and 1 the worst). Stress and silhouette values are calculated on the dataset extended (120 variables against

[3] <available after publication acceptance>
[4] <available after publication acceptance>



(a) The Dataset coverage vs. Number of trees k (2^1 to 2^{14}). After extracting, selecting, and aggregating (Algorithm 1 and Algorithm 2) the patterns from 2 trees, 96% of the data instances are covered. By using 2048 (2^{11}) trees 100% of coverage is reached.



(b) The Kruskal Stress and Silhouette Coefficient values on dataset extensions varying the λ parameter. The 255 patterns resulting from selection and aggregation (Algorithm 1 with number of trees $k = 2048$ and Algorithm 2) are taken as classes for the extension (subsection 4.3.2). Kruskal Stress lies among 0 and 1, whereas Silhouette Coefficient is normalized and inverted into values between 0 and 1. The minimum stress and silhouette values are found for λ between 0.60 and 0.70.

Figure 20 – The parameters definition: Number of trees k for Algorithm 1 and λ for dataset extension.

the original with 60) using as classes the 255 JEPs resulting from the Algorithm 2. It is possible to spot that the minimum stress and silhouette values are found between 0.60 and 0.70. Hence, we take $\lambda = 0.65$.

Figure 21 presents 14 JEPs from the resulting 255, representing 70% of the dataset. The patterns are ordered by support and variables by importance. Despite the number of variables and complexity of the dataset, about half of it (48%) is described by only

two patterns, p_{184} and p_{152} (first two rows). These are the highest support patterns for Hillary and Trump voters, respectively. Interestingly, the difference between them is on three questions/variables (first three columns): "HEALTHLAW" with majority answers "Expand the law" for Hillary and "Repeal the law entirely" for Trump, "RUSSIA" with answers "Yes" for Hillary and "No" for Trump, and "IMMWALL" with answers "Strongly oppose" for Hillary and "Strongly favor" for Trump. Therefore, half of Hillary voters (50% from pattern p_{184}) are in favor of expanding the Affordable Care Act (Obamacare), against the wall with Mexico, and believe that the Trump election campaign was coordinated with Russia. On the other hand, about half of Trump voters (45% from pattern p_{152}) are against the Affordable Care Act, supporting the wall, and do not believe Russia had a role in Trump's campaign.

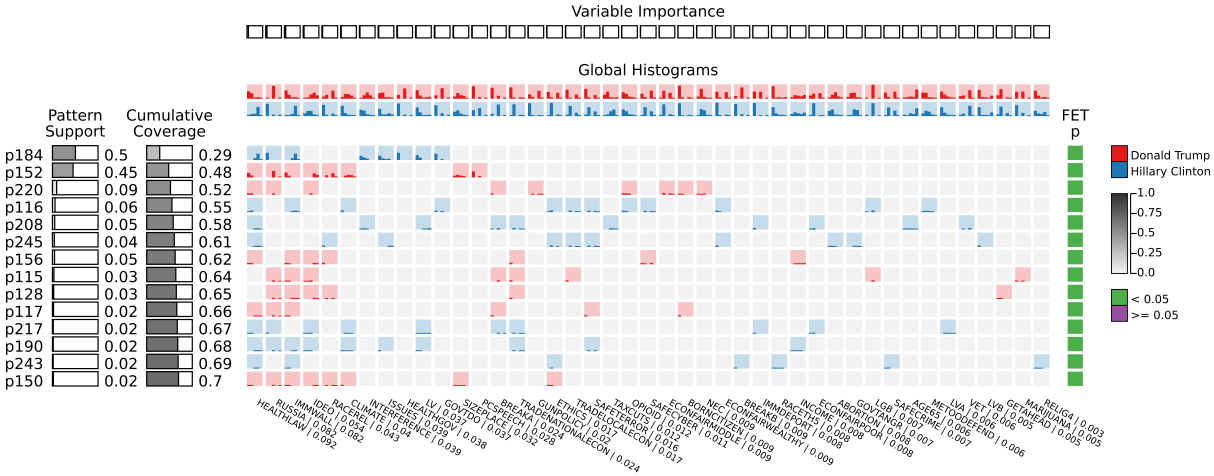
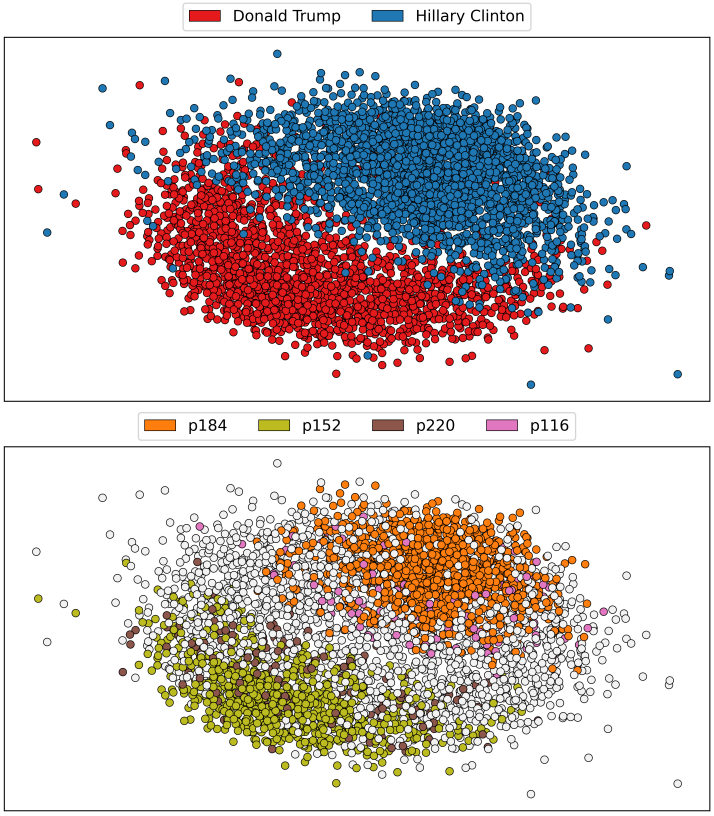


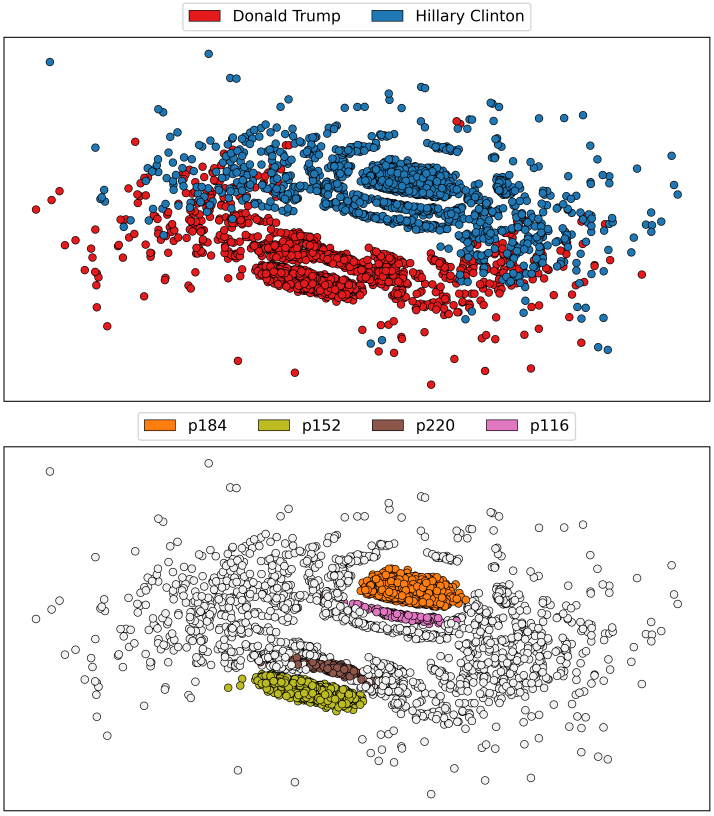
Figure 21 – The 14 JEPs filtered out the 255 obtained by Algorithm 1 ($k = 2048$) and Algorithm 2. JEPs are ordered by support and variables by importance. About half (48%) of the electorate can be described by only two patterns (p_{184} and p_{152}), and these diverge in three points: The support to the Affordable Care Act, the construction of the wall with Mexico, and the Russian participation in Trump's campaign.

The support drops considerably for the subsequent patterns. However, revealing a more heterogeneous scenario for the other half of voters. For instance, pattern p_{220} (third row) supporting 9% of Trump voters diverges from the strong pattern p_{152} on political ideology (variable "IDEO" fourth column) while considering themselves as moderate and not conservative. The pattern p_{116} (fourth row) supporting 6% of Hilary voters differs from the relevant pattern p_{184} regarding the Affordable Care Act (variable "HEALTHLAW" first column), advocating to repeal the law at least in parts. The last pattern p_{150} supports only 2% of instances from its associated class. All the remaining 241 patterns (255 in total and 14 displayed in Figure 21) have equal or less than 2% of support.

The patterns found for half of the voters are readily analyzed and compared to others. These high support patterns can be seen as clusters in the instances similarity map in Figure 22b, but not in Figure 22a when the map is created with the original dataset.



(a) The instances similarity map by $\lambda = 0.0$.



(b) The instances similarity map by $\lambda = 0.65$.

Figure 22 – The Instances similarity maps visualizations for the 2016 US election dataset. The 4 highest support patterns (p_{184}, p_{152}, p_{220}, and p_{116} from Figure 21) can be seen as clusters in the similarity map (b) but not in (a).

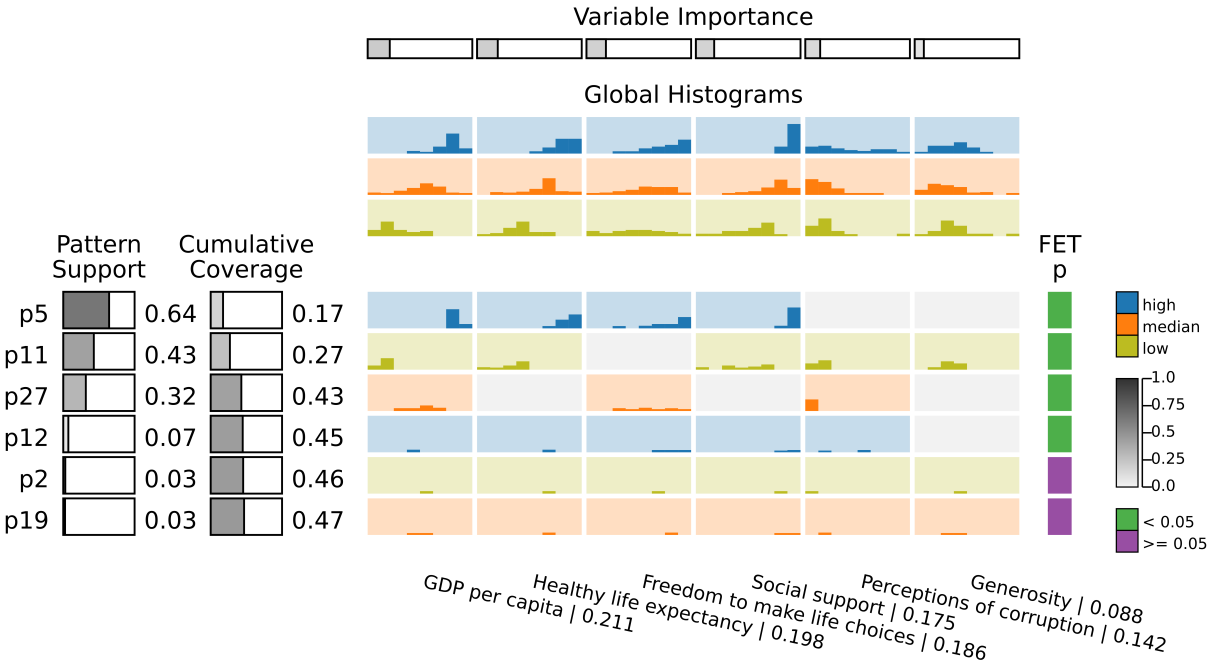
Thus, extending the dataset over JEPs can reveal clusters in DR layouts otherwise not found using only the original dataset. The data explanations show a fascinating picture of the US, indicating that the simple division between the two extremes is much more complex in practice, with several subgroups diverging from the most common behavior found by JEPs visualization. This is also clear in the instances similarity map. The major groups are spotted, but mixed subgroups also exist. A valid hypothesis is that "both sides are more similar than may think".

4.4.2 Use Case II – World Happiness

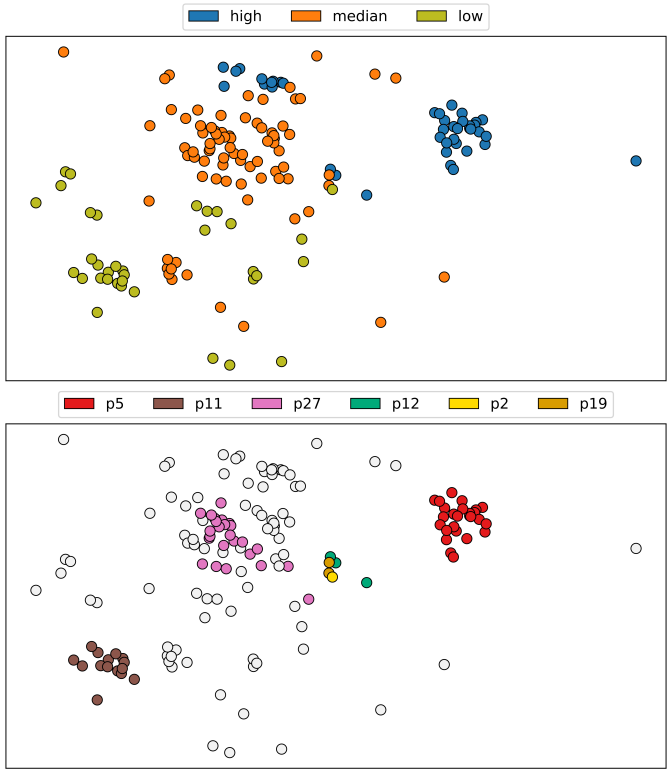
The second use case concerns the dataset analysis of the World Happiness Report 2019 by the Sustainable Development Solutions Network (SDSN) (HELLIWELL; LAYARD; SACHS, 2019). This dataset presents a ranking of 156 countries based on an index representing how happy their citizens perceive themselves. It also contains six other variables along with the "Happiness Score" variation across countries. These variables are "GDP per capita", "Social support", "Healthy life expectancy", "Freedom to make life choices", "Generosity", and "Perceptions of corruption". Since the world happiness dataset does not present class labels, a strategy must be followed to produce labels for data instances (countries). We have chosen variable discretization (SALMAN; KECMAN, 2012), but approaches such as clustering (JAMES *et al.*, 2013) and instances subset selection (CAO; BROWN, 2020) are also suitable.

For the analysis, we chose the happiness score to be discretized to create the labels. The idea is to verify the differences between countries, regarding the other six variables, based on perceived happiness levels. In this process, the happiness score was discretized in three equally sized bins, where instances class is the instances assigned bin ("high", "media", and "low"). In other words, we are transforming a regression problem into classification (SALMAN; KECMAN, 2012). The "high happy" class encloses 42 countries with happiness scores from 6.13 to 7.76, the "median happy" contains 79 countries from 4.49 to 6.13, and the "low happy" class groups 35 countries from 2.85 to 4.49. Since the variable "Happiness Score" is employed for class derivation (discretization), it is not used to extract patterns. The parameters number of trees k and λ for map creation are set following the specified procedures in subsection 4.4.1, that is the minimum number of trees, and Kruskal Stress and Silhouette Coefficient optimization. From $2,893$ patterns extracted by Algorithm 1 ($k = 64$), 29 were selected, 252 aggregated, and $2,612$ discarded by Algorithm 2. For instances similarity map creation it was used $\lambda = 0.65$.

Figure 23a presents 6 JEPs from the resulting 29, filtered by the highest support patterns for each class together with instances of interest from the map shown in Figure 23b.



(a) The 6 JEPs filtered out the 29 obtained by Algorithm 1 ($k = 64$) and Algorithm 2. JEPs are ordered by support and variables by importance.



(b) The instances similarity map by $\lambda = 0.65$.

Figure 23 – The JEPs and instances similarity maps visualizations for the World Happiness Report 2019. The first 3 patterns (p_5, p_{11}, and p_{27}) in (a) describe the general behavior of high, median, and low happy countries, with a clear difference between them in terms "GDP per capita", "'Healthy life expectancy", and "Social support". The instance similarly map (b) allows to reason about clusters and outliers. The 6 countries placed about the map center were selected to analyze their respective patterns in (a).

The patterns in Figure 23a represent 47% of the dataset. The patterns are ordered by support, whereas variables by importance. Pattern p_5 (first row) represents 64% of the "high happy" countries, indicating that those countries have high values in "GDP per capita", "Health life expectancy", and "Social support". Moreover, p_5 supports countries with varied values for "Freedom to make life choices". The second pattern with the highest support (p_{11} at the second row) describes "low happy" countries, accounting for 43% of that class. These countries have low values for "GDP per capita" and "Perceptions of corruption", and median values for "Health life expectancy", "Social support", and "Generosity". Compared to "high happy" countries in pattern p_5, it is clear that those "low happy" countries present lower values for "GDP per capita", "Health life expectancy", and "Social support". The pattern p_{27} supports 32% of "median happy" countries, having median values for "GDP per capita", low values for "Perceptions of corruption", and about flat distributed values for "Freedom to make life choices". These "median happy" countries are between "high happy" and "low happy" countries from patterns p_5 and p_{11} in "GDP per capita". The above mentioned patterns bring us to an interesting hypothesis that "low perceptions of corruption does not necessary make citizens happy".

The first three patterns p_5, p_{11}, and p_{27} in Figure 23a encode countries' general behavior (43% of the dataset). The patterns p_5 and p_{11} led to two clusters in the instances similarity map in Figure 23b. However, instances from pattern p_{27} are spread between such clusters. The map allows reason about clusters and outliers by selecting instances of interest and visualizing their respective patterns. From Figure 23b, the 6 countries placed about the map center have been selected, presenting their patterns in Figure 23a. The 3 "high happy" countries supported by pattern p_{12} differ from pattern p_5 by having median values for "GDP per capita". The single "low happy" country supported by pattern p_2 diverges from pattern p_{11} by holding high values in "GDP per capita", "Healthy life expectancy", and "Social support". The 2 "median happy" countries supported by pattern p_{19} contrast pattern p_{27} by yielding high values for "Perceptions of corruption". Furthermore, these 6 close countries with different levels of happiness are similar in "Healthy life expectancy". Thus, browsing the instances similarity map and visualizing patterns from selected instances makes it possible to reason about countries' differences and similarities under the optics of happiness levels.

4.5 Discussion and Limitations

VAX provides a powerful VA method that focuses on data analysis via automated insights (LAW; ENDERT; STASKO, 2020) instead of pure classification model explanation. For multivariate data explanation, VAX presents JEPs (Jumping Emerging Patterns) using a matrix-like visual metaphor. The employed metaphor arranges patterns as rows and variables as columns, meeting content and multi-class requirements (NO-

VAK; LAVRAC; WEBB, 2009) by showing global and local histograms (matrix cells) along classes coded as categorical colors. Since VAX follows matrix visualization guidelines (CHEN *et al.*, 2004; WU; TZENG; CHEN, 2008), meaningful visual representations can be reached by filtering and ordering patterns (rows) and variables (columns).

VAX involves extracting, selecting, aggregating, and visualizing JEPs. Many diversified patterns are usually extracted from random DTs, defining multiple variables combinations. The RFm method (GARCíA-BORROTO; MARTíNEZ-TRINIDAD; CARRASCO-OCHOA, 2015; LOYOLA-GONZÁLEZ; MEDINA-PÉREZ; CHOO, 2020) is used for building DT models (Algorithm 1). Based on RF (Random Forest) (BREIMAN, 2001; BIAU; SCORNET, 2016), RFm builds k DT models from a class-labeled dataset using a random subset of variables (with size $\log_2|V|$) on each internal node creation. However, instances bagging (BREIMAN, 2001; JAMES *et al.*, 2013) is not used (GARCíA-BORROTO; MARTíNEZ-TRINIDAD; CARRASCO-OCHOA, 2015; LOYOLA-GONZÁLEZ; MEDINA-PÉREZ; CHOO, 2020).

The RFm has been proved to be a valuable approach to mine diversified patterns (GARCíA-BORROTO; MARTíNEZ-TRINIDAD; CARRASCO-OCHOA, 2015). The JEPs selection and aggregation process (Algorithm 2) is inspired by an existent approach (LOYOLA-GONZáLEZ *et al.*, 2019), reducing the initial set of patterns extracted. These are selected by their confidence and support and aggregated with other compelling patterns. Those not fulfilling the specifications of being representative (high support) or an exception (for particular instances) are discarded. In this way, complementary patterns are aggregated, not losing their information, whereas redundant patterns are discarded. There is no consensus about the total number of DT models k (GARCíA-BORROTO; MARTíNEZ-TRINIDAD; CARRASCO-OCHOA, 2015; LOYOLA-GONZáLEZ *et al.*, 2019; GARCíA-VICO *et al.*, 2018). We aim at the minimum number of trees capable of extracting enough patterns to our selection and aggregation strategy resulting in the coverage of all data instances. Furthermore, a threshold of data coverage can also be used for setting the number of DTs k. The quality of extracted JEPs is related to DTs' ability to obtain variables and class relations from a particular dataset. Low-quality patterns are associated with the acquisition of only low support relations. This issue derives from the DT limitation of learning generic patterns or their nonexistence in the dataset in question, which is desirable since it is better not to create artifacts. By focusing on JEPs, ambiguous instances (equally variables values but different classes) are not supported by the available patterns.

As a limitation on JEPs visualization, since histograms require a certain vertical (height) display space, it is impossible to visualize a significant number of patterns at once. Nevertheless, displaying many patterns may hamper the analysis. Potential solutions are selecting groups of interest, like the patterns that combined attain minimum dataset sup-

port (e.g., 70%) or focusing on particular instances (e.g., outliers). Regarding horizontal space for displaying variables, the visual metaphor is still limited by the number of variables used by a single pattern. For this matter, filtering variables by importance may be an option.

DR (Dimensional Reduction) layouts are great for identifying and analyzing patterns in multidimensional (multivariate) data (NONATO; AUPETIT, 2019; LIU *et al.*, 2017; PéREZ *et al.*, 2015). Still, the ability to produce such patterns is directly related to the DR technique's efficiency in revealing clusters and outliers into a lower-dimensional space from a much bigger and complex original space. Clusters and outliers may not be identified on poor visual quality projections, with points and groups overlaps (PéREZ *et al.*, 2015), which is a problem related to the curse of dimensionality (DONOHO, 2000). Extending the original space (PéREZ *et al.*, 2015) over JEPs perspectives enables identifying and analyzing patterns in two-dimensional projections into DR layouts. Clusters and outliers not found in the original space projection can be investigated in projections from data extended using JEPs. We select the λ parameter for dataset extension based on Kruskal Stress (KRUSKAL, 1964) and Silhouette Coefficient (ROUSSEEUW, 1987), maximizing the coefficient along minimizing the stress. Although not a constrain, we used the statistical mean for variable creation, as in the space extension original proposition (PéREZ *et al.*, 2015), and the classic MDS technique (KRUSKAL, 1964) for generating DR layouts. The studies of multiple layouts from different DR techniques along with varying statistic measures for creating variables are fascinating but out of the scope of this work. We included these investigations as future work.

VAX integrates descriptive logic rules (JEPs) visualization with instances similarity maps (DR layouts). High support patterns tend to produce clusters into DR layouts. Global histograms are references for the local ones from patterns cells, which can be used to explain such clusters. In the other direction, the DR layout can be browsed for clusters and outliers, presenting the patterns supporting these data instances. With JEPs matrix visualization and the instances maps, compound facts (LAW; ENDERT; STASKO, 2020) can be generated, enabling linked data insights involving descriptive patterns, clusters, and outliers. Compound facts are highly desirable, providing more nuanced insights about multivariate data (LAW; ENDERT; STASKO, 2020).

4.6 Conclusions

In this paper, we present VAX *(multiVariate dAta eXplanation)*, a new method for analyzing multivariate datasets. VAX employs aggregated Jumping Emerging Patterns (JEPs) to capture intricate patterns in a class-labeled dataset. A matrix-like visual metaphor is used for JEPs visualization, where patterns are rows, variables are columns,

and data distribution conveyed using histograms are matrix cells. Based on matrix visualization, meaningful visual representations can be reached by filtering and ordering patterns (rows) and variables (columns). Furthermore, instances similarity maps via (Dimensional Reduction) DR layouts aim better understanding of dataset's overall image (e.g., clusters and outliers) using JEPs lens. VAX allows JEPs and instances maps visualization that can be applied to different domains, addressing phenomenon comprehension through knowledge acquisition, showing a valuable tool for creating hypotheses based on data insights. We plan as future work new approaches for filtering and ordering JEPs and variables to enhance VAX visual representations. Moreover, we intend to pursue new methods for diversified patterns extraction, and comparisons involving different DR techniques and various statistic measures for instances maps creation.

CHAPTER

5

CONCLUSION

5.1 Contributions

This doctoral thesis presented two VA methods displaying logic rules extracted from RF models into a matrix-like visual metaphor. The methods are ExMatrix and VAX for RF models interpretability, covering predictive and descriptive purposes, respectively. Based on matrix visualization guidelines, both methods arrange rules as rows, features (variables) as columns, and rules predicates as cells (ranges or histograms). Rules (rows) and features (columns) filtering and ordering are crucial to create meaningful visual representations. ExMatrix aims at global and local explanations for overviewing the model and auditing the classification process for a particular data instance. VAX aims at multivariate data explanation, extracting and processing logic rules from an RF model to obtain JEPs (Jumping Emerging Patterns). Such expressive data patterns are also employed for generating data instances similarity maps. The visual representations obtained from ExMatrix are model-centered, and those achieved by VAX are data-centered. A flowchart-based summarization is found in Figure 30 of Appendix C, exhibiting inputs, processes, and outputs for ExMatrix and VAX.

Besides enabling filtering and ordering, the matrix-like visual metaphor used in ExMatrix is more concise and scalable for logic rules visualization than node-link diagrams (common DT representation) and state-of-the-art techniques. Features order (columns) overcomes the effort of mentally composing features ranges on node-link diagrams. The rules predicates (features ranges) mapped as rectangular shapes provide scalability requiring fewer display pixels. Scalability is especially necessary when dealing with RF models. Once arranging several DTs, many logic rules can be generated. In addition to RF models, ExMatrix can interpret a single DT taken as a classification model. For sensing units calibration in analytical chemistry, the ExMatrix application to interpret DT models built over sensing data creates MCS (Multidimensional Calibration Space). Such interpretable

calibration is based on logic rules extracted from DT models, and it may arrange several features (dimensions) and data from heterogeneous sources.

The RF-based approach for mining descriptive logic rules (JEPs) used in VAX has been noted in the literature for extracting diversified patterns. The selection and aggregation process results in compelling JEPs, aggregating complementary patterns (keeping their information), and discarding the redundant ones. Therefore, JEPs can be expressive by providing high support or an exception supporting distinct instances. DR layouts are handy for exploratory analyses of multivariate data, but limited to the DR technique's efficiency in unfolding the data structure. The data instances similarity maps provided by VAX allow investigations on clusters and outliers from a DR layout built over data extended (new variables) using JEPs. VAX enables compound facts (linked data insights) integrating JEPs visualization and instances similarity maps. Two real-world datasets were analyzed by VAX, leading to hypothesis-making and knowledge acquisition.

The proposed methods fulfilled the goals defined in Chapter 1, confirming the hypothesis of logic rules visualizations being able to support explanations of multivariate data and RF models' overall logic and outcomes. Appendix A contains a list of publications and submissions, and source code (majority) is available at <https://gitlab.com/popolinneto/exmatrix> and as code package at <https://pypi.org/project/exmatrix/>. Although achieving the goals and validating the hypothesis, some limitations are found in ExMatrix and VAX.

5.2 Limitations

Although scalable, display size (resolution) imposes a limitation for ExMatrix visual representations, impacting global and local explanations differently. Global explanations handle all logic rules, whereas local explanations deal with one per DT model. A possible solution is rules filtering. The explanations for global and local showing used rules are readily useful visual representations, whereas those showing smallest changes rules may be more appropriate for RF experts. Local explanations with the smallest changes rules lead to hypotheses, rather than straightforward answers. These latter explanations may reflect the classification process stability for a particular instance. Scalability is also an issue in VAX visual representations of JEPs, where histograms are used in matrix cells requiring a vertical display space. The matrix-like visual metaphor is also limited by the number of variables (columns) employed in a single pattern, needing horizontal space. Thus, patterns filtering is crucial for JEPs visualization, and variables may also need filtration.

The DT models employed in ExMatrix and VAX were created by the CART algorithm (BREIMAN *et al.*, 1984). Other DT algorithms can generate different logic rules.

Both ExMatrix and VAX are bounded by DT models limitations in obtaining variables and class relations from a class-labeled dataset. For ExMatrix, these limitations can lead to poor accuracy models interpretation, whereas for VAX, no expressive JEPs extraction. For the cases where complex black-box models such ANN or SVM are needed, overcoming RF models performance, ExMatrix may be used to create explanations of surrogate logic rules. Besides DTs deficiencies, the dataset under analysis may contain no variables and class associations. For the latter scenario, it is beneficial VAX failure in extracting expressive JEPs and producing clusters into instances maps, considering the creation of patterns where there are none.

5.3 Future work

There are many possibilities for ExMatrix and VAX applications since any problem represented by a class-labeled dataset is a potential use case. Therefore, both methods can be employed in several domains as well as in complete VA systems. Regarding ExMatrix, it is worth investigating RF model optimization (e.g., editing) assisted by model explanations. About VAX, it is interesting experiments with different DR techniques along with various statistic measures for dataset extension. Moreover, new filtering and ordering criteria may produce better explanations, improving both ExMatrix and VAX.

The main challenge for ExMatrix is the ensemble factor. Global and especially local explanations must be analyzed under ensemble optics. The user needs to have in mind that knowledge is acquired from several unique models (DTs) and a voting process is taken for data instance classification. A complete user test with RF experts is required for ExMatrix's further improvements and evaluation. Visual explanations focusing on the voting committee (e.g., features impact) may be the next step. On the other hand, VAX's principal challenge is the descriptive concept. Visual explanations from JEPs visualization and data instances similarity maps must be seen as phenomenon descriptions. The user must analyze VAX representations for knowledge acquisition. The efforts for significant advances are required in approaches for patterns extraction and combined visual representations of data instances and patterns.

BIBLIOGRAPHY

ADADI, A.; BERRADA, M. Peeking inside the black-box: A survey on explainable artificial intelligence (xai). **IEEE Access**, v. 6, p. 52138–52160, 2018. Citations on pages 29 and 30.

ALESSIO, P.; CONSTANTINO, C. J. L.; DAIKUZONO, C. M.; RIUL, A.; OLIVEIRA, O. N. de. Analysis of coffees using electronic tongues. In: MéNDEZ, M. L. R. (Ed.). **Electronic Noses and Tongues in Food Science**. San Diego: Academic Press, 2016. p. 171–177. ISBN 978-0-12-800243-8. Available: <https://www.sciencedirect.com/science/article/pii/B9780128002438000172>. Citation on page 62.

ALPER, B.; BACH, B.; RICHE, N. H.; ISENBERG, T.; FEKETE, J.-D. Weighted graph comparison techniques for brain connectivity analysis. In: **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2013. (CHI '13), p. 483–492. ISBN 9781450318990. Available: <https://doi.org/10.1145/2470654.2470724>. Citation on page 39.

ALSALLAKH, B.; MICALLEF, L.; AIGNER, W.; HAUSER, H.; MIKSCH, S.; RODGERS, P. Visualizing sets and set-typed data: State-of-the-art and future challenges. In: BORGO, R.; MACIEJEWSKI, R.; VIOLA, I. (Ed.). **EuroVis - STARs**. [S.l.]: The Eurographics Association, 2014. ISBN -. Citation on page 39.

ANKERST, M.; ELSEN, C.; ESTER, M.; KRIEGEL, H.-P. Visual classification: An interactive approach to decision tree construction. In: **Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: ACM, 1999. (KDD '99), p. 392–396. ISBN 1-58113-143-7. Available: <http://doi.acm.org/10.1145/312129.312298>. Citations on pages 38 and 72.

ANWAR, N.; WARNARS, H. L. H. S.; SANCHEZ, H. E. P. Survey of emerging patterns. In: **2017 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom)**. [S.l.: s.n.], 2017. p. 11–18. Citations on pages 72 and 75.

BALLARD, Z. S.; JOUNG, H.-A.; GONCHAROV, A.; LIANG, J.; NUGROHO, K.; CARLO, D. D.; GARNER, O. B.; OZCAN, A. Deep learning-enabled point-of-care sensing using multiplexed paper-based sensors. **npj Digital Medicine**, v. 3, n. 1, p. 66, May 2020. ISSN 2398-6352. Available: <https://doi.org/10.1038/s41746-020-0274-y>. Citation on page 63.

BANAEI, N.; MOSHFEGH, J.; MOHSENI-KABIR, A.; HOUGHTON, J. M.; SUN, Y.; KIM, B. Machine learning algorithms enhance the specificity of cancer biomarker detection using sers-based immunoassays in microfluidic chips. **RSC Adv.**, The Royal Society of Chemistry, v. 9, p. 1859–1868, 2019. Available: <http://dx.doi.org/10.1039/C8RA08930B>. Citation on page 63.

BEHRISCH, M.; BACH, B.; RICHE, N. H.; SCHRECK, T.; FEKETE, J.-D. Matrix reordering methods for table and network visualization. **Computer Graphics Forum**, v. 35, n. 3, p. 693–716, 2016. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.12935>. Citations on pages 39 and 44.

BIAU, G.; SCORNET, E. A random forest guided tour. **TEST**, v. 25, n. 2, p. 197–227, Jun 2016. ISSN 1863-8260. Available: <https://doi.org/10.1007/s11749-016-0481-7>. Citations on pages 30, 31, 37, 41, 80, and 97.

BOULESTEIX, A.-L.; TUTZ, G.; STRIMMER, K. A CART-based approach to discover emerging patterns in microarray data. **Bioinformatics**, v. 19, n. 18, p. 2465–2472, 12 2003. ISSN 1367-4803. Available: <https://doi.org/10.1093/bioinformatics/btg361>. Citation on page 81.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. **Diário Oficial da República Federativa do Brasil**, Brasília, DF, 2018. ISSN 1677-7042. Available: <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm>. Citation on page 30.

_____. Lei nº 13.853, de 08 de julho de 2019. **Diário Oficial da República Federativa do Brasil**, Brasília, DF, 2019. ISSN 1677-7042. Available: <http://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2019/Lei/L13853.htm>. Citation on page 30.

BRAUNGER, M. L.; SHIMIZU, F. M.; JIMENEZ, M. J. M.; AMARAL, L. R.; PIAZZETTA, M. H. d. O.; GOBBI, A. L.; MAGALHãES, P. S. G.; RODRIGUES, V.; OLIVEIRA, O. N.; RIUL, A. Microfluidic electronic tongue applied to soil analysis. **Chemosensors**, v. 5, n. 2, 2017. ISSN 2227-9040. Available: <https://www.mdpi.com/2227-9040/5/2/14>. Citation on page 62.

BREIMAN, L. Random forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001. ISSN 1573-0565. Available: <https://doi.org/10.1023/A:1010933404324>. Citations on pages 30, 31, 37, 41, 80, 81, and 97.

_____. Manual on setting up, using, and understanding random forests v3. 1. **Statistics Department University of California Berkeley, CA, USA**, v. 1, p. 58, 2002. Citations on pages 44 and 66.

BREIMAN, L.; FRIEDMAN, J.; OLSHEN, R.; STONE, C. **Classification and Regression Trees**. [S.l.]: Chapman and Hall/CRC, 1984. Citations on pages 38, 41, 63, 64, 68, 75, 80, 81, and 102.

BRONIATOWSKI, D. **Psychological Foundations of Explainability and Interpretability in Artificial Intelligence**. NIST Interagency/Internal Report (NISTIR), National Institute of Standards and Technology, Gaithersburg, MD, 2021. Available: <https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=931426>. Citations on pages 29 and 30.

BUTLER, K. T.; DAVIES, D. W.; CARTWRIGHT, H.; ISAYEV, O.; WALSH, A. Machine learning for molecular and materials science. **Nature**, v. 559, n. 7715, p. 547–555, 2018. ISSN 1476-4687. Available: <https://doi.org/10.1038/s41586-018-0337-2>. Citation on page 36.

CAO, F.; BROWN, E. T. Dril: Descriptive rules by interactive learning. In: **2020 IEEE Visualization Conference (VIS)**. [S.l.: s.n.], 2020. p. 256–260. Citation on page 94.

CARVALHO, D. V.; PEREIRA, E. M.; CARDOSO, J. S. Machine learning interpretability: A survey on methods and metrics. **Electronics**, v. 8, n. 8, 2019. ISSN 2079-9292. Available: <https://www.mdpi.com/2079-9292/8/8/832>. Citations on pages 29, 30, 36, 38, and 52.

CASES, M. V.; LÓPEZ-LORENTE, Á. I.; LÓPEZ-JIMÉNEZ, M. Á. **Foundations of Analytical Chemistry: A Teaching–Learning Approach**. 1. ed. Springer International Publishing, 2018. 487 p. ISBN 978-3-319-62871-4. Available: <https://doi.org/10.1007/978-3-319-62872-1>. Citation on page 62.

CHAN, G. Y.-Y.; BERTINI, E.; NONATO, L. G.; BARR, B.; SILVA, C. T. Melody: Generating and visualizing machine learning model summary to understand data and classifiers together. **arXiv preprint arXiv:2007.10614**, 2020. Citation on page 72.

CHEN, C.-H.; HWU, H.-G.; JANG, W.-J.; KAO, C.-H.; TIEN, Y.-J.; TZENG, S.; WU, H.-M. Matrix visualization and information mining. In: ANTOCH, J. (Ed.). **COMPSTAT 2004 — Proceedings in Computational Statistics**. Heidelberg: Physica-Verlag HD, 2004. p. 85–100. ISBN 978-3-7908-2656-2. Citations on pages 32, 39, 41, 44, 76, 85, and 97.

CHEN, C.-h.; SINICA, A.; TAIPEI. Generalized association plots: information visualization via iteratively generated correlation matrices. **Statistica Sinica**, v. 12, p. 7–29, 01 2002. Citations on pages 39, 41, and 44.

CHOI, S. seok; CHA, S. hyuk. A survey of binary similarity and distance measures. **Journal of Systemics, Cybernetics and Informatics**, p. 43–48, 2010. Citation on page 44.

CHOO, J.; LEE, H.; KIHM, J.; PARK, H. ivisclassifier: An interactive visual analytics system for classification based on supervised dimension reduction. In: **2010 IEEE Symposium on Visual Analytics Science and Technology**. [S.l.: s.n.], 2010. p. 27–34. Citation on page 38.

CLOUGH, J. R.; OKSUZ, I.; PUYOL-ANTÓN, E.; RUIJSINK, B.; KING, A. P.; SCHNABEL, J. A. Global and local interpretability for cardiac mri classification. In: SHEN, D.; LIU, T.; PETERS, T. M.; STAIB, L. H.; ESSERT, C.; ZHOU, S.; YAP, P.-T.; KHAN, A. (Ed.). **Medical Image Computing and Computer Assisted Intervention – MICCAI 2019**. Cham: Springer International Publishing, 2019. p. 656–664. ISBN 978-3-030-32251-9. Citation on page 29.

CORRELL, M.; LI, M.; KINDLMANN, G.; SCHEIDEGGER, C. Looks good to me: Visualizations as sanity checks. **IEEE Transactions on Visualization and Computer Graphics**, v. 25, n. 1, p. 830–839, 2019. Citation on page 84.

CRUZ, J. A.; WISHART, D. S. Applications of machine learning in cancer prediction and prognosis. **Cancer Informatics**, v. 2, p. 117693510600200030, 2006. Citation on page 36.

CURRIE, L. A. Nomenclature in evaluation of analytical methods including detection and quantification capabilities (IUPAC recommendations 1995). Walter de Gruyter GmbH, v. 67, n. 10, p. 1699–1723, Jan. 1995. Available: <https://doi.org/10.1351/pac199567101699>. Citation on page 62.

_____. Nomenclature in evaluation of analytical methods including detection and quantification capabilities 1: (iupac recommendations 1995). Elsevier BV, v. 391, n. 2, p. 105–126, May 1999. Available: <https://doi.org/10.1016/s0003-2670(99)00104-x>. Citation on page 62.

DAIKUZONO, C. M.; DANTAS, C. A.; VOLPATI, D.; CONSTANTINO, C. J.; PIAZZETTA, M. H.; GOBBI, A. L.; TAYLOR, D. M.; OLIVEIRA, O. N.; RIUL, A. Microfluidic electronic tongue. **Sensors and Actuators B: Chemical**, v. 207, p. 1129–1135, 2015. ISSN 0925-4005. A Special Issue in Honour of Professor Yu. G. Vlasov. Available: <https://www.sciencedirect.com/science/article/pii/S0925400514012027>. Citation on page 62.

DAIKUZONO, C. M.; SHIMIZU, F. M.; MANZOLI, A.; RIUL, A.; PIAZZETTA, M. H. O.; GOBBI, A. L.; CORREA, D. S.; PAULOVICH, F. V.; OLIVEIRA, O. N. Information visualization and feature selection methods applied to detect gliadin in gluten-containing foodstuff with a microfluidic electronic tongue. **ACS Applied Materials & Interfaces**, v. 9, n. 23, p. 19646–19652, 2017. PMID: 28481518. Available: <https://doi.org/10.1021/acsami.7b04252>. Citation on page 62.

DANG, T. N.; WILKINSON, L. Scagexplorer: Exploring scatterplots by their scagnostics. In: **2014 IEEE Pacific Visualization Symposium**. [S.l.: s.n.], 2014. p. 73–80. Citation on page 72.

DHEERU, D.; TANISKIDOU, E. K. **UCI Machine Learning Repository**. University of California, Irvine, School of Information and Computer Sciences, 2017. Available: <http://archive.ics.uci.edu/ml>. Citation on page 48.

DI CASTRO, F.; BERTINI, E. Surrogate decision tree visualization interpreting and visualizing black-box classification models with surrogate decision tree. **CEUR Workshop Proceedings**, v. 2327, 1 2019. ISSN 1613-0073. Citations on pages 30, 36, 37, 38, 39, 72, 73, and 74.

DI NATALE, C.; PAOLESSE, R.; MACAGNANO, A.; MANTINI, A.; D'AMICO, A.; LEGIN, A.; LVOVA, L.; RUDNITSKAYA, A.; VLASOV, Y. Electronic nose and electronic tongue integration for improved classification of clinical and food samples. **Sensors and Actuators B: Chemical**, v. 64, n. 1, p. 15–21, 2000. ISSN 0925-4005. Available: <https://www.sciencedirect.com/science/article/pii/S0925400599004773>. Citation on page 62.

DI ROSA, A. R.; LEONE, F.; CHELI, F.; CHIOFALO, V. Fusion of electronic nose, electronic tongue and computer vision for animal source food authentication and quality assessment – a review. **Journal of Food Engineering**, v. 210, p. 62–75, 2017. ISSN 0260-8774. Available: <https://www.sciencedirect.com/science/article/pii/S0260877417301796>. Citation on page 62.

DO, T.-N. Towards simple, easy to understand, an interactive decision tree algorithm. **College Inf. Technol., Can tho Univ., Can Tho, Vietnam, Tech. Rep**, p. 06–01, 2007. Citations on pages 38 and 72.

DONG, G.; LI, J. Efficient mining of emerging patterns: Discovering trends and differences. In: **Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: Association for Computing Machinery, 1999. (KDD '99), p. 43–52. ISBN 1581131437. Available:

<https://doi.org/10.1145/312129.312191>. Citations on pages 72, 73, 74, 75, 77, 78, 79, and 80.

DONOHO, D. L. High-dimensional data analysis: The curses and blessings of dimensionality. In: **American Mathematical Society Conference on Mathematical Challenges of the 21st Century**. [S.l.: s.n.], 2000. Citation on page 98.

DU, M.; LIU, N.; HU, X. Techniques for interpretable machine learning. **Commun. ACM**, Association for Computing Machinery, New York, NY, USA, v. 63, n. 1, p. 68–77, Dec. 2019. ISSN 0001-0782. Available: <https://doi.org/10.1145/3359786>. Citations on pages 30, 36, 38, 72, and 76.

DUA, D.; GRAFF, C. **UCI Machine Learning Repository**. University of California, Irvine, School of Information and Computer Sciences, 2017. Available: <http://archive.ics.uci.edu/ml>. Citation on page 48.

ENDERT, A.; RIBARSKY, W.; TURKAY, C.; WONG, B. W.; NABNEY, I.; BLANCO, I. D.; ROSSI, F. The state of the art in integrating machine learning into visual analytics. **Computer Graphics Forum**, v. 36, n. 8, p. 458–486, 2017. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13092>. Citations on pages 30 and 36.

FARRAIA, M. V.; RUFO, J. C.; PACIÊNCIA, I.; MENDES, F.; DELGADO, L.; MOREIRA, A. The electronic nose technology in clinical diagnosis: A systematic review. **Porto Biomedical Journal**, v. 4, n. 4, 2019. ISSN 2444-8664. Available: <https://doi.org/10.1097/j.pbj.0000000000000042>. Citation on page 62.

FISHER, R. A. The use of multiple measurements in taxonomic problems. **Annals of Eugenics**, v. 7, n. 2, p. 179–188, 1936. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-1809.1936.tb02137.x>. Citation on page 42.

FREEDMAN, D.; DIACONIS, P. On the histogram as a density estimator:l2 theory. **Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete**, v. 57, n. 4, p. 453–476, Dec 1981. ISSN 1432-2064. Available: <https://doi.org/10.1007/BF01025868>. Citation on page 84.

FREITAS, A. A. Comprehensible classification models: A position paper. **SIGKDD Explor. Newsl.**, ACM, New York, NY, USA, v. 15, n. 1, p. 1–10, Mar. 2014. ISSN 1931-0145. Available: <http://doi.acm.org/10.1145/2594473.2594475>. Citations on pages 30, 37, 39, 57, 122, and 123.

FUJIWARA, T.; KWON, O.-H.; MA, K.-L. Supporting analysis of dimensionality reduction results with contrastive learning. **IEEE Transactions on Visualization and Computer Graphics**, v. 26, n. 1, p. 1–1, 2019. Citation on page 44.

FüRNKRANZ, J.; GAMBERGER, D.; LAVRAC, N. **Foundations of Rule Learning**. 1. ed. Springer Berlin Heidelberg, 2012. 334 p. ISSN 1611-2482. ISBN 978-3-540-75196-0. Available: <https://doi.org/10.1007/978-3-540-75197-7>. Citations on pages 30 and 76.

GAMBERGER, D.; LAVRAC, N.; WETTSCHERECK, D. Subgroup visualization: A method and application in population screening. In: **In Proceedings of the International Workshop on intelligent Data Analysis in Medicine and Pharmacology, IDAMAP**. [S.l.: s.n.], 2002. Citation on page 75.

GARCíA-BORROTO, M.; MARTíNEZ-TRINIDAD, J. F.; CARRASCO-OCHOA, J. A. Finding the best diversity generation procedures for mining contrast patterns. **Expert Systems with Applications**, v. 42, n. 11, p. 4859–4866, 2015. ISSN 0957-4174. Available: <https://www.sciencedirect.com/science/article/pii/S0957417415001359>. Citations on pages 31, 75, 77, 80, 81, 82, and 97.

GARCíA-VICO, A.; CARMONA, C.; MARTíN, D.; GARCíA-BORROTO, M.; JESUS, M. del. An overview of emerging pattern mining in supervised descriptive rule discovery: taxonomy, empirical study, trends, and prospects. **WIREs Data Mining and Knowledge Discovery**, v. 8, n. 1, p. e1231, 2018. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/widm.1231>. Citations on pages 31, 72, 73, 74, 75, 77, 78, 79, 80, 85, and 97.

GAUR, M.; FALDU, K.; SHETH, A. Semantics of the black-box: Can knowledge graphs help make deep learning systems more interpretable and explainable? **IEEE Internet Computing**, v. 25, n. 1, p. 51–59, 2021. Citation on page 29.

GILPIN, L. H.; BAU, D.; YUAN, B. Z.; BAJWA, A.; SPECTER, M.; KAGAL, L. Explaining explanations: An overview of interpretability of machine learning. In: **2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)**. [S.l.: s.n.], 2018. p. 80–89. Citation on page 29.

GLEICHER, M. Explainers: Expert explorations with crafted projections. **IEEE Transactions on Visualization and Computer Graphics**, v. 19, n. 12, p. 2042–2051, 2013. Citations on pages 31, 72, 73, 74, and 81.

GOMEZ, O.; HOLTER, S.; YUAN, J.; BERTINI, E. Vice: Visual counterfactual explanations for machine learning models. In: **Proceedings of the 25th International Conference on Intelligent User Interfaces**. New York, NY, USA: Association for Computing Machinery, 2020. (IUI '20), p. 531–535. ISBN 9781450371186. Available: <https://doi.org/10.1145/3377325.3377536>. Citation on page 40.

GONZALEZ-NAVARRO, F. F.; STILIANOVA-STOYTCHEVA, M.; RENTERIA-GUTIERREZ, L.; BELANCHE-MUñOZ, L. A.; FLORES-RIOS, B. L.; IBARRA-ESQUER, J. E. Glucose oxidase biosensor modeling and predictors optimization by machine learning methods. **Sensors**, v. 16, n. 11, 2016. ISSN 1424-8220. Available: <https://www.mdpi.com/1424-8220/16/11/1483>. Citation on page 63.

GRABOSKI, A. M.; ZAKRZEVSKI, C. A.; SHIMIZU, F. M.; PASCHOALIN, R. T.; SOARES, A. C.; STEFFENS, J.; PAROUL, N.; STEFFENS, C. Electronic nose based on carbon nanocomposite sensors for clove essential oil detection. **ACS Sensors**, v. 5, n. 6, p. 1814–1821, 2020. PMID: 32515185. Available: <https://doi.org/10.1021/acssensors.0c00636>. Citation on page 62.

GRAHAM, M.; KENNEDY, J. A survey of multiple tree visualisation. **Information Visualization**, Palgrave Macmillan, v. 9, n. 4, p. 235–252, Dec. 2010. ISSN 1473-8716. Available: <https://doi.org/10.1057/ivs.2009.29>. Citations on pages 37, 38, 39, 57, and 123.

GUIDOTTI, R.; MONREALE, A.; RUGGIERI, S.; PEDRESCHI, D.; TURINI, F.; GIANNOTTI, F. Local rule-based explanations of black box decision systems. **arXiv preprint arXiv:1805.10820**, 2018. Citations on pages 30, 36, 37, 38, 40, and 76.

GUIDOTTI, R.; MONREALE, A.; RUGGIERI, S.; TURINI, F.; GIANNOTTI, F.; PEDRESCHI, D. A survey of methods for explaining black box models. **ACM Comput. Surv.**, ACM, New York, NY, USA, v. 51, n. 5, p. 93:1–93:42, Aug. 2018. ISSN 0360-0300. Available: <http://doi.acm.org/10.1145/3236009>. Citations on pages 29, 30, 31, 36, 52, 63, and 64.

HALL, P. On the art and science of machine learning explanations. **arXiv preprint arXiv:1810.02909**, 2018. Citations on pages 38, 63, and 64.

HELLIWELL, J.; LAYARD, R.; SACHS, J. **World Happiness Report 2019**. New York: Sustainable Development Solutions Network, 2019. ISBN 978-0-9968513-9-8. Available: <https://worldhappiness.report/ed/2019/>. Citations on pages 90 and 94.

HO, T. K. Random decision forests. In: **Proceedings of 3rd International Conference on Document Analysis and Recognition**. [S.l.: s.n.], 1995. v. 1, p. 278–282 vol.1. Citation on page 80.

_____. The random subspace method for constructing decision forests. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 20, n. 8, p. 832–844, 1998. Citation on page 80.

HUYSMANS, J.; DEJAEGER, K.; MUES, C.; VANTHIENEN, J.; BAESENS, B. An empirical evaluation of the comprehensibility of decision table, tree and rule based predictive models. **Decision Support Systems**, v. 51, n. 1, p. 141 – 154, 2011. ISSN 0167-9236. Available: <http://www.sciencedirect.com/science/article/pii/S0167923610002368>. Citations on pages 37, 39, 57, 122, and 123.

HöFERLIN, B.; NETZEL, R.; HöFERLIN, M.; WEISKOPF, D.; HEIDEMANN, G. Interactive learning of ad-hoc classifiers for video visual analytics. In: **2012 IEEE Conference on Visual Analytics Science and Technology (VAST)**. [S.l.: s.n.], 2012. p. 23–32. Citations on pages 38 and 72.

JAHANGIRI, A.; RAKHA, H. A. Applying machine learning techniques to transportation mode recognition using mobile phone sensor data. **IEEE Transactions on Intelligent Transportation Systems**, v. 16, n. 5, p. 2406–2417, 2015. Citation on page 69.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An Introduction to Statistical Learning with Applications in R**. Springer New York, 2013. ISSN 1431-875X. ISBN 978-1-4614-7138-7. Available: <https://doi.org/10.1007/978-1-4614-7138-7>. Citations on pages 69, 80, 81, 94, and 97.

KANE, B.; CUISSART, B.; CRÉMILLEUX, B. Minimal jumping emerging patterns: Computation and practical assessment. In: CAO, T.; LIM, E.-P.; ZHOU, Z.-H.; HO, T.-B.; CHEUNG, D.; MOTODA, H. (Ed.). **Advances in Knowledge Discovery and Data Mining**. Cham: Springer International Publishing, 2015. p. 722–733. ISBN 978-3-319-18038-0. Citations on pages 31, 73, 75, and 79.

KEIM, D.; KOHLHAMER, J.; ELLIS, G.; MANSMANN, F. **Mastering the information age solving problems with visual analytics**. Eurographics Association, 2010. ISBN 978-3-905673-77-7. Available: <http://diglib.eg.org/handle/10.2312/14803>. Citations on pages 30, 31, 72, and 76.

KNITTEL, J.; LALAMA, A.; KOCH, S.; ERTL, T. Visual neural decomposition to explain multivariate data sets. **IEEE Transactions on Visualization and Computer Graphics**, p. 1–1, 2020. Citations on pages 31, 72, 73, 74, 81, and 90.

KOHAVI, R. The power of decision tables. In: **Proceedings of the 8th European Conference on Machine Learning**. Berlin, Heidelberg: Springer-Verlag, 1995. (ECML'95), p. 174–189. Citations on pages 36 and 39.

_____. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: **Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1995. (IJCAI'95), p. 1137–1143. ISBN 1558603638. Citation on page 69.

KRAUSE, J.; DASGUPTA, A.; SWARTZ, J.; APHINYANAPHONGS, Y.; BERTINI, E. A workflow for visual diagnostics of binary classifiers using instance-level explanations. In: **2017 IEEE Conference on Visual Analytics Science and Technology (VAST)**. [S.l.: s.n.], 2017. p. 162–172. ISSN null. Citation on page 44.

KRUSKAL, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. **Psychometrika**, v. 29, n. 1, p. 1–27, Mar 1964. ISSN 1860-0980. Available: <https://doi.org/10.1007/BF02289565>. Citations on pages 88, 90, and 98.

KUMAR, R.; BHONDEKAR, A. P.; KAUR, R.; VIG, S.; SHARMA, A.; KAPUR, P. A simple electronic tongue. **Sensors and Actuators B: Chemical**, v. 171-172, p. 1046–1053, 2012. ISSN 0925-4005. Available: <https://www.sciencedirect.com/science/article/pii/S0925400512006065>. Citation on page 64.

LAKKARAJU, H.; BACH, S. H.; LESKOVEC, J. Interpretable decision sets: A joint framework for description and prediction. In: **Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: Association for Computing Machinery, 2016. (KDD '16), p. 1675–1684. ISBN 9781450342322. Available: <https://doi.org/10.1145/2939672.2939874>. Citations on pages 30, 31, 39, 40, 76, and 122.

LAW, P.-M.; ENDERT, A.; STASKO, J. Characterizing automated data insights. In: **2020 IEEE Visualization Conference (VIS)**. [S.l.: s.n.], 2020. p. 171–175. Citations on pages 33, 76, 77, 96, and 98.

LEE, T.; JOHNSON, J.; CHENG, S. An interactive machine learning framework. **arXiv preprint arXiv:1610.05463**, 2016. Citations on pages 38 and 72.

LEGIN, A.; RUDNITSKAYA, A.; VLASOV, Y. Electronic tongues: new analytical perspective for chemical sensors. In: **Integrated Analytical Systems**. Elsevier, 2003, (Comprehensive Analytical Chemistry, v. 39). p. 437–486. Available: <https://www.sciencedirect.com/science/article/pii/S0166526X03801150>. Citation on page 62.

LEI, J.; WANG, Z.; FENG, Z.; SONG, M.; BU, J. Understanding the prediction process of deep networks by forests. In: **2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM)**. [S.l.: s.n.], 2018. p. 1–7. Citations on pages 36 and 37.

LI, J.; MANOUKIAN, T.; DONG, G.; RAMAMOHANARAO, K. Incremental maintenance on the border of the space of emerging patterns. **Data Mining and Knowledge Discovery**, v. 9, n. 1, p. 89–116, Jul 2004. ISSN 1573-756X. Available: <https://doi.org/10.1023/B:DAMI.0000026901.85057.58>. Citation on page 80.

LI, X.; YANG, T.; LI, S.; JIN, L.; WANG, D.; GUAN, D.; DING, J. Noninvasive liver diseases detection based on serum surface enhanced raman spectroscopy and statistical analysis. **Opt. Express**, OSA, v. 23, n. 14, p. 18361–18372, Jul 2015. Available: <http://www.osapublishing.org/oe/abstract.cfm?URI=oe-23-14-18361>. Citation on page 63.

LIAO, Q. V.; GRUEN, D.; MILLER, S. Questioning the ai: Informing design practices for explainable ai user experiences. In: **Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2020. (CHI '20), p. 1–15. ISBN 9781450367080. Available: <https://doi.org/10.1145/3313831.3376590>. Citations on pages 29, 40, 58, and 72.

LIMA, E.; MUES, C.; BAESENS, B. Domain knowledge integration in data mining using decision tables: case studies in churn prediction. **Journal of the Operational Research Society**, Taylor & Francis, v. 60, n. 8, p. 1096–1106, 2009. Available: <https://doi.org/10.1057/jors.2008.161>. Citations on pages 37, 39, 57, and 123.

LIU, S.; MALJOVEC, D.; WANG, B.; BREMER, P.-T.; PASCUCCI, V. Visualizing high-dimensional data: Advances in the past decade. **IEEE Transactions on Visualization and Computer Graphics**, v. 23, n. 3, p. 1249–1268, 2017. Citation on page 98.

LIU, S.; XIAO, J.; LIU, J.; WAN, X.; WU, J.; ZHU, J. Visual diagnosis of tree boosting methods. **IEEE Transactions on Visualization and Computer Graphics**, v. 24, n. 1, p. 163–173, Jan 2018. Citations on pages 36, 38, 39, and 72.

LIU, X.; WANG, X.; MATWIN, S. Interpretable deep convolutional neural networks via meta-learning. In: **2018 International Joint Conference on Neural Networks (IJCNN)**. [S.l.: s.n.], 2018. p. 1–9. Citations on pages 36 and 52.

LOEKITO, E.; BAILEY, J. Using highly expressive contrast patterns for classification - is it worthwhile? In: THEERAMUNKONG, T.; KIJSIRIKUL, B.; CERCONE, N.; HO, T.-B. (Ed.). **Advances in Knowledge Discovery and Data Mining**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. p. 483–490. ISBN 978-3-642-01307-2. Citation on page 81.

LOH, W.-Y. Fifty years of classification and regression trees. **International Statistical Review**, v. 82, n. 3, p. 329–348, 2014. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/insr.12016>. Citation on page 38.

LOYOLA-GONZÁLEZ, O.; MEDINA-PÉREZ, M. A.; CHOO, K.-K. R. A review of supervised classification based on contrast patterns: Applications, trends, and challenges. **Journal of Grid Computing**, Oct 2020. ISSN 1572-9184. Available: <https://doi.org/10.1007/s10723-020-09526-y>. Citations on pages 31, 72, 73, 74, 75, 77, 79, 80, 81, 82, 85, and 97.

LOYOLA-GONZáLEZ, O.; LóPEZ-CUEVAS, A.; MEDINA-PéREZ, M. A.; CAMIñA, B.; RAMíREZ-MáRQUEZ, J. E.; MONROY, R. Fusing pattern discovery and visual analytics approaches in tweet propagation. **Information Fusion**, v. 46, p. 91–101,

2019. ISSN 1566-2535. Available: <https://www.sciencedirect.com/science/article/pii/S1566253517307716>. Citations on pages 31, 75, 80, 82, and 97.

LUNDERVOLD, A. S.; LUNDERVOLD, A. An overview of deep learning in medical imaging focusing on mri. **Zeitschrift für Medizinische Physik**, v. 29, n. 2, p. 102–127, 2019. ISSN 0939-3889. Special Issue: Deep Learning in Medical Physics. Available: <https://www.sciencedirect.com/science/article/pii/S0939388918301181>. Citation on page 63.

MENDEZ, M. R.; PREEDY, V. **Electronic noses and tongues in food science**. [S.l.]: Academic Press, 2016. 332 p. ISBN 9780128002438. Citation on page 62.

MICHALSKI, R.; STEPP, R. Revealing conceptual structure in data by inductive inference. In: MICHIE, D.; HAYES, J. E.; PAO, Y.-H. (Ed.). **Machine Learning 10**. [S.l.]: John Wile and Sons Publishing, 1982. Citation on page 78.

MIGUT, M.; WORRING, M. Visual exploration of classification models for risk assessment. In: **2010 IEEE Symposium on Visual Analytics Science and Technology**. [S.l.: s.n.], 2010. p. 11–18. Citation on page 38.

MING, Y.; QU, H.; BERTINI, E. Rulematrix: Visualizing and understanding classifiers with rules. **IEEE Transactions on Visualization and Computer Graphics**, v. 25, n. 1, p. 342–352, Jan 2019. ISSN 1077-2626. Citations on pages 30, 31, 36, 37, 38, 39, 40, 43, 72, 73, 74, and 76.

MIRANDA, T. Z.; SARDINHA, D. B.; CERRI, R. Preventing the generation of inconsistent sets of crisp classification rules. **Expert Systems with Applications**, v. 165, p. 113811, 2021. ISSN 0957-4174. Available: <http://www.sciencedirect.com/science/article/pii/S0957417420306254>. Citations on pages 30, 31, 75, and 76.

MOOSAVI, S. M.; GHASSABIAN, S. Linearity of calibration curves for analytical methods: A review of criteria for assessment of method reliability. In: . InTechOpen, 2018. Available: <https://doi.org/10.5772/intechopen.72932>. Citation on page 62.

MORAES, M. L.; MAKI, R. M.; PAULOVICH, F. V.; FILHO, U. P. R.; OLIVEIRA, M. C. F. de; RIUL, A.; SOUZA, N. C. de; FERREIRA, M.; GOMES, H. L.; OLIVEIRA, O. N. Strategies to optimize biosensors based on impedance spectroscopy to detect phytic acid using layer-by-layer films. **Analytical Chemistry**, American Chemical Society, v. 82, n. 8, p. 3239–3246, Apr 2010. ISSN 0003-2700. Available: <https://doi.org/10.1021/ac902949h>. Citations on pages 16, 63, and 65.

MUNZNER, T. **Visualization Analysis and Design**. [S.l.]: CRC Press, 2014. (AK Peters Visualization Series). ISBN 9781466508934. Citation on page 84.

NONATO, L. G.; AUPETIT, M. Multidimensional projection for visual analytics: Linking techniques with distortions, tasks, and layout enrichment. **IEEE Transactions on Visualization and Computer Graphics**, v. 25, n. 8, p. 2650–2673, 2019. Citations on pages 33, 74, and 98.

NOVAK, P. K.; LAVRAC, N.; WEBB, G. I. Supervised descriptive rule discovery: A unifying survey of contrast set, emerging pattern and subgroup mining. **J. Mach. Learn. Res.**, JMLR.org, v. 10, p. 377–403, Jun. 2009. ISSN 1532-4435. Citations on pages 31, 72, 73, 74, 75, 78, 79, 80, 81, and 97.

OLIVEIRA, J. E.; GRASSI, V.; SCAGION, V. P.; MATTOSO, L. H. C.; GLENN, G. M.; MEDEIROS, E. S. Sensor array for water analysis based on interdigitated electrodes modified with fiber films of poly(lactic acid)/multiwalled carbon nanotubes. **IEEE Sensors Journal**, v. 13, n. 2, p. 759–766, 2013. Citation on page 62.

PAIVA, J. G. S.; SCHWARTZ, W. R.; PEDRINI, H.; MINGHIM, R. An approach to supporting incremental visual data classification. **IEEE Transactions on Visualization and Computer Graphics**, v. 21, n. 1, p. 4–17, Jan 2015. Citation on page 38.

PAJA, W. A decision rule based approach to generational feature selection. In: PERNER, P. (Ed.). **Advances in Data Mining. Applications and Theoretical Aspects**. Cham: Springer International Publishing, 2018. p. 230–239. ISBN 978-3-319-95786-9. Citation on page 85.

PATEL, H. K. **The Electronic Nose: Artificial Olfaction Technology**. 1. ed. Springer, New Delhi, 2014. 247 p. ISBN 978-81-322-1547-9. Available: <https://doi.org/10.1007/978-81-322-1548-6>. Citation on page 62.

PODRAżKA, M.; BąCZYńSKA, E.; KUNDYS, M.; JELEń, P. S.; NERY, E. W. Electronic tongue-a tool for all tastes? **Biosensors**, v. 8, n. 1, 2018. ISSN 2079-6374. Available: <https://www.mdpi.com/2079-6374/8/1/3>. Citation on page 62.

POPOLIN NETO, M.; PAULOVICH, F. V. Explainable matrix - visualization for global and local interpretability of random forest classification ensembles. **IEEE Transactions on Visualization and Computer Graphics**, v. 27, n. 2, p. 1427–1437, 2021. Available: <https://doi.org/10.1109/TVCG.2020.3030354>. Citations on pages 32, 64, 65, 72, and 76.

POPOLIN NETO, M.; SOARES, A. C.; OLIVEIRA, O. N.; PAULOVICH, F. V. Machine learning used to create a multidimensional calibration space for sensing and biosensing data. **Bulletin of the Chemical Society of Japan**, v. 94, n. 5, p. 1553–1562, 2021. Available: <https://doi.org/10.1246/bcsj.20200359>. Citation on page 33.

PéREZ, D.; ZHANG, L.; SCHAEFER, M.; SCHRECK, T.; KEIM, D.; DíAZ, I. Interactive feature space extension for multidimensional data projection. **Neurocomputing**, v. 150, p. 611–626, 2015. ISSN 0925-2312. Special Issue on Information Processing and Machine Learning for Applications of Engineering Solving Complex Machine Learning Problems with Ensemble Methods Visual Analytics using Multidimensional Projections. Available: <https://www.sciencedirect.com/science/article/pii/S0925231214012879>. Citations on pages 76, 87, 88, 90, and 98.

RAUBER, P. E.; FADEL, S. G.; FALCãO, A. X.; TELEA, A. C. Visualizing the hidden activity of artificial neural networks. **IEEE Transactions on Visualization and Computer Graphics**, v. 23, n. 1, p. 101–110, Jan 2017. Citation on page 38.

RIBEIRO, M. T.; SINGH, S.; GUESTRIN, C. "why should i trust you?": Explaining the predictions of any classifier. In: **Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: Association for Computing Machinery, 2016. (KDD '16), p. 1135–1144. ISBN 9781450342322. Available: <https://doi.org/10.1145/2939672.2939778>. Citations on pages 29, 30, 38, 40, 72, 73, and 74.

_____. Anchors: High-precision model-agnostic explanations. In: . [s.n.], 2018. Available: <https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16982/15850>. Citations on pages 30, 36, 37, 38, 40, 72, and 76.

RIUL, A.; de Sousa, H. C.; MALMEGRIM, R. R.; dos Santos, D. S.; CARVALHO, A. C.; FONSECA, F. J.; OLIVEIRA, O. N.; MATTOSO, L. H. Wine classification by taste sensors made from ultra-thin films and using neural networks. **Sensors and Actuators B: Chemical**, v. 98, n. 1, p. 77–82, 2004. ISSN 0925-4005. Available: <https://www.sciencedirect.com/science/article/pii/S0925400503007512>. Citations on pages 62 and 64.

RIUL JúNIOR, A.; DANTAS, C. A. R.; MIYAZAKI, C. M.; JR., O. N. O. Recent advances in electronic tongues. **Analyst**, The Royal Society of Chemistry, v. 135, p. 2481–2495, 2010. Available: <http://dx.doi.org/10.1039/C0AN00292E>. Citation on page 33.

ROUSSEEUW, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. **Journal of Computational and Applied Mathematics**, v. 20, p. 53–65, 1987. ISSN 0377-0427. Available: <https://www.sciencedirect.com/science/article/pii/0377042787901257>. Citations on pages 90 and 98.

RUDNITSKAYA, A.; SCHMIDTKE, L. M.; REIS, A.; DOMINGUES, M. R. M.; DELGADILLO, I.; DEBUS, B.; KIRSANOV, D.; LEGIN, A. Measurements of the effects of wine maceration with oak chips using an electronic tongue. **Food Chemistry**, v. 229, p. 20–27, 2017. ISSN 0308-8146. Available: <https://www.sciencedirect.com/science/article/pii/S0308814617302017>. Citation on page 62.

SACHA, D.; STOFFEL, A.; STOFFEL, F.; KWON, B. C.; ELLIS, G.; KEIM, D. A. Knowledge generation model for visual analytics. **IEEE Transactions on Visualization and Computer Graphics**, v. 20, n. 12, p. 1604–1613, Dec 2014. ISSN 1077-2626. Citations on pages 30, 31, 72, and 76.

SALMAN, R.; KECMAN, V. Regression as classification. In: **2012 Proceedings of IEEE Southeastcon**. [S.l.: s.n.], 2012. p. 1–6. Citations on pages 69 and 94.

SCHULZ, H. Treevis.net: A tree visualization reference. **IEEE Computer Graphics and Applications**, v. 31, n. 6, p. 11–15, Nov 2011. ISSN 0272-1716. Citation on page 37.

SCHULZ, H.; HADLAK, S.; SCHUMANN, H. The design space of implicit hierarchy visualization: A survey. **IEEE Transactions on Visualization and Computer Graphics**, v. 17, n. 4, p. 393–411, 2011. Citations on pages 37, 38, 57, and 123.

SHIMIZU, F. M.; BRAUNGER, M. L.; RIUL, A. Heavy metal/toxins detection using electronic tongues. **Chemosensors**, v. 7, n. 3, 2019. ISSN 2227-9040. Available: <https://www.mdpi.com/2227-9040/7/3/36>. Citation on page 62.

SHIMIZU, F. M.; TODÃO, F. R.; GOBBI, A. L.; OLIVEIRA, O. N.; GARCIA, C. D.; LIMA, R. S. Functionalization-free microfluidic electronic tongue based on a single response. **ACS Sensors**, American Chemical Society, v. 2, n. 7, p. 1027–1034, Jul 2017. Available: <https://doi.org/10.1021/acssensors.7b00302>. Citation on page 62.

SHNEIDERMAN, B. The eyes have it: a task by data type taxonomy for information visualizations. In: **Proceedings 1996 IEEE Symposium on Visual Languages**. [S.l.: s.n.], 1996. p. 336–343. Citations on pages 49 and 58.

SONG, H.; WANG, Y.; ROSANO, J. M.; PRABHAKARPANDIAN, B.; GARSON, C.; PANT, K.; LAI, E. A microfluidic impedance flow cytometer for identification of differentiation state of stem cells. **Lab Chip**, The Royal Society of Chemistry, v. 13, p. 2300–2310, 2013. Available: <http://dx.doi.org/10.1039/C3LC41321G>. Citation on page 64.

STIGLIC, G.; MERTIK, M.; PODGORELEC, V.; KOKOL, P. Using visual interpretation of small ensembles in microarray analysis. In: **19th IEEE Symposium on Computer-Based Medical Systems (CBMS'06)**. [S.l.: s.n.], 2006. p. 691–695. Citation on page 38.

STRUMBELJ, E.; KONONENKO, I. An efficient explanation of individual classifications using game theory. **J. Mach. Learn. Res.**, JMLR.org, v. 11, p. 1–18, Mar. 2010. ISSN 1532-4435. Available: <http://dl.acm.org/citation.cfm?id=1756006.1756007>. Citation on page 40.

TALBOT, J.; LEE, B.; KAPOOR, A.; TAN, D. S. Ensemblematrix: Interactive visualization to support machine learning with multiple classifiers. In: **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**. New York, NY, USA: ACM, 2009. (CHI '09), p. 1283–1292. ISBN 978-1-60558-246-7. Available: <http://doi.acm.org/10.1145/1518701.1518895>. Citations on pages 38 and 72.

TAN, P.-N.; STEINBACH, M.; KUMAR, V. **Introduction to data mining**. 1. ed. [S.l.]: Pearson, 2005. 792 p. ISBN 0321321367. Citations on pages 29, 31, 38, 41, 57, 63, 64, 72, 75, 80, 81, and 123.

TEOH, S. T.; MA, K.-L. Paintingclass: Interactive construction, visualization and exploration of decision trees. In: **Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: ACM, 2003. (KDD '03), p. 667–672. ISBN 1-58113-737-0. Available: <http://doi.acm.org/10.1145/956750.956837>. Citations on pages 38 and 72.

TERMEHYOUSEFI, A. **Nanocomposite-Based Electronic Tongue**. 1. ed. Springer International Publishing, 2018. 101 p. ISBN 978-3-319-66847-5. Available: <https://doi.org/10.1007/978-3-319-66848-2>. Citation on page 62.

TOMPSON, T.; BENZ, J. **AP VoteCast 2018**. ICPSR - Interuniversity Consortium for Political and Social Research, 2018. Available: <https://www.openicpsr.org/openicpsr/project/109687/version/V2/view>. Citation on page 90.

TSAMARDINOS, I.; RAKHSHANI, A.; LAGANI, V. Performance-estimation properties of cross-validation-based protocols with simultaneous hyper-parameter optimization. In: LIKAS, A.; BLEKAS, K.; KALLES, D. (Ed.). **Artificial Intelligence: Methods and Applications**. Cham: Springer International Publishing, 2014. p. 1–14. ISBN 978-3-319-07064-3. Citation on page 69.

TZENG, S.; WU, H.; CHEN, C. Selection of proximity measures for matrix visualization of binary data. In: **2009 2nd International Conference on Biomedical Engineering and Informatics**. [S.l.: s.n.], 2009. p. 1–9. ISSN 1948-2922. Citation on page 44.

VAN DEN ELZEN, S.; VAN WIJK, J. J. Baobabview: Interactive construction and analysis of decision trees. In: **2011 IEEE Conference on Visual Analytics Science and Technology (VAST)**. [S.l.: s.n.], 2011. p. 151–160. Citations on pages 37, 38, and 72.

VARMA, S.; SIMON, R. Bias in error estimation when using cross-validation for model selection. **BMC Bioinformatics**, v. 7, n. 1, p. 91, Feb 2006. ISSN 1471-2105. Available: <https://doi.org/10.1186/1471-2105-7-91>. Citation on page 69.

VASHISTHA, R.; DANGI, A. K.; KUMAR, A.; CHHABRA, D.; SHUKLA, P. Futuristic biosensors for cardiac health care: an artificial intelligence approach. **3 Biotech**, v. 8, n. 8, p. 358, Aug 2018. ISSN 2190-5738. Available: <https://doi.org/10.1007/s13205-018-1368-y>. Citation on page 63.

VLASOV, Y.; LEGIN, A.; RUDNITSKAYA, A. Electronic tongues and their analytical application. **Analytical and Bioanalytical Chemistry**, v. 373, n. 3, p. 136–146, Jun 2002. ISSN 1618-2650. Available: <https://doi.org/10.1007/s00216-002-1310-2>. Citation on page 62.

WANG, L.; WANG, Y.; ZHAO, D. Building emerging pattern (ep) random forest for recognition. In: **2010 IEEE International Conference on Image Processing**. [S.l.: s.n.], 2010. p. 1457–1460. Citation on page 80.

WANG, L.; ZHAO, H.; DONG, G.; LI, J. On the complexity of finding emerging patterns. In: **Proceedings of the 28th Annual International Computer Software and Applications Conference, 2004. COMPSAC 2004.** [S.l.: s.n.], 2004. v. 2, p. 126–129 vol.2. Citation on page 80.

WILSON, D.; MATERóN, E. M.; IBáñEZ-REDíN, G.; FARIA, R. C.; CORREA, D. S.; OLIVEIRA, O. N. Electrical detection of pathogenic bacteria in food samples using information visualization methods with a sensor based on magnetic nanoparticles functionalized with antimicrobial peptides. **Talanta**, v. 194, p. 611–618, 2019. ISSN 0039-9140. Available: <https://www.sciencedirect.com/science/article/pii/S0039914018311342>. Citation on page 62.

WU, H.-M.; TZENG, S.; CHEN, C.-h. Matrix visualization. In: _____. **Handbook of Data Visualization**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. p. 681–708. ISBN 978-3-540-33037-0. Available: <https://doi.org/10.1007/978-3-540-33037-0_26>. Citations on pages 32, 39, 41, 44, 76, 85, and 97.

WU, M.; HUGHES, M.; PARBHOO, S.; ZAZZI, M.; ROTH, V.; DOSHI-VELEZ, F. Beyond sparsity: Tree regularization of deep models for interpretability. In: **AAAI Conference on Artificial Intelligence**. [s.n.], 2018. Available: <https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16285/15867>. Citation on page 38.

YANG, F.; DU, M.; HU, X. Evaluating explanation without ground truth in interpretable machine learning. **arXiv preprint arXiv:1907.06831**, 2019. Citation on page 36.

ZHAO, X.; WU, Y.; LEE, D. L.; CUI, W. iforest: Interpreting random forests via visual analytics. **IEEE Transactions on Visualization and Computer Graphics**, v. 25, n. 1, p. 407–416, Jan 2019. ISSN 1077-2626. Citations on pages 30, 31, 37, 38, 40, 42, 43, 48, 51, and 72.

APPENDIX

A

LIST OF PUBLICATIONS

This appendix exhibits the complete list of published, accepted, or submitted (preprint) works produced during this doctoral project.

1. POPOLIN NETO, M.; PAULOVICH, F. V. Explainable matrix - visualization for global and local interpretability of random forest classification ensembles. IEEE Transactions on Visualization and Computer Graphics, v. 27, n. 2, p. 1427–1437, 2021. Available: <https://doi.org/10.1109/TVCG.2020.3030354>.

2. POPOLIN NETO, M.; SOARES, A. C.; OLIVEIRA, O. N.; PAULOVICH, F. V. Machine learning used to create a multidimensional calibration space for sensing and biosensing data. Bulletin of the Chemical Society of Japan, v. 94, n. 5, p. 1553–1562, 2021. Available: <https://doi.org/10.1246/bcsj.20200359>.

3. PAULOVICH, F. V.; POPOLIN NETO, M. Roadmap for sensorial data analysis. In: SHIMIZU, F. B.; BRAUNGER, M. L.; RIUL JÚNIOR, A. Electronic Tongues Fundamentals and recent advances. 1. ed. IOP Publishing Ltd, 2021. Available: <https://doi.org/10.1088/978-0-7503-3687-1ch10>.

4. POPOLIN NETO, M.; PAULOVICH, F. V. Multivariate Data Explanation by Jumping Emerging Patterns Visualization. arXiv preprint arXiv:2106.11112. 2021.

5. MAZUMDAR, D.; POPOLIN NETO, M.; PAULOVICH, F. V. Random forest similarity maps: A scalable visual representation for global and local interpretation. Electronics, v. 10, n. 22, 2021. ISSN 2079-9292. Available: <https://www.mdpi.com/2079-9292/10/22/2862>.

6. BONDANCIA, T. J.; SOARES, A. C.; POPOLIN NETO, M.; GOMES, N. O.; RAYMUNDO-PEREIRA, P. A.; BARUD, H. S.; MACHADO, S. A.; RIBEIRO, S. J.; MELENDEZ, M. E.; CARVALHO, A. L.; REIS, R. M.; PAULOVICH,

F. V.; OLIVEIRA, O. N. Low-cost bacterial nanocellulose-based interdigitated biosensor to detect the p53 cancer biomarker. Materials Science and Engineering: C, p. 112676, 2022. ISSN 0928-4931. Available: <https://www.sciencedirect.com/-science/article/pii/S0928493122000364>.

7. SOARES, J. C.; SOARES, A. C.; POPOLIN NETO, M.; PAULOVICH, F. V.; OLIVEIRA, O. N.; MATTOSO, L. H. C. Detection of staphylococcus aureus in milk samples using impedance spectroscopy and data processing with information visualization techniques and multidimensional calibration space. Sensors and Actuators Reports, p. 100083, 2022. ISSN 2666-0539. Available: <https://www.sciencedirect.com/science/article/pii/S2666053922000108>.

APPENDIX

B

EXMATRIX – SUPPLEMENTAL MATERIAL

This appendix presents the paper's [1] supplemental material for Chapter 2.

B.1 Additional Figures and Code Reference Table

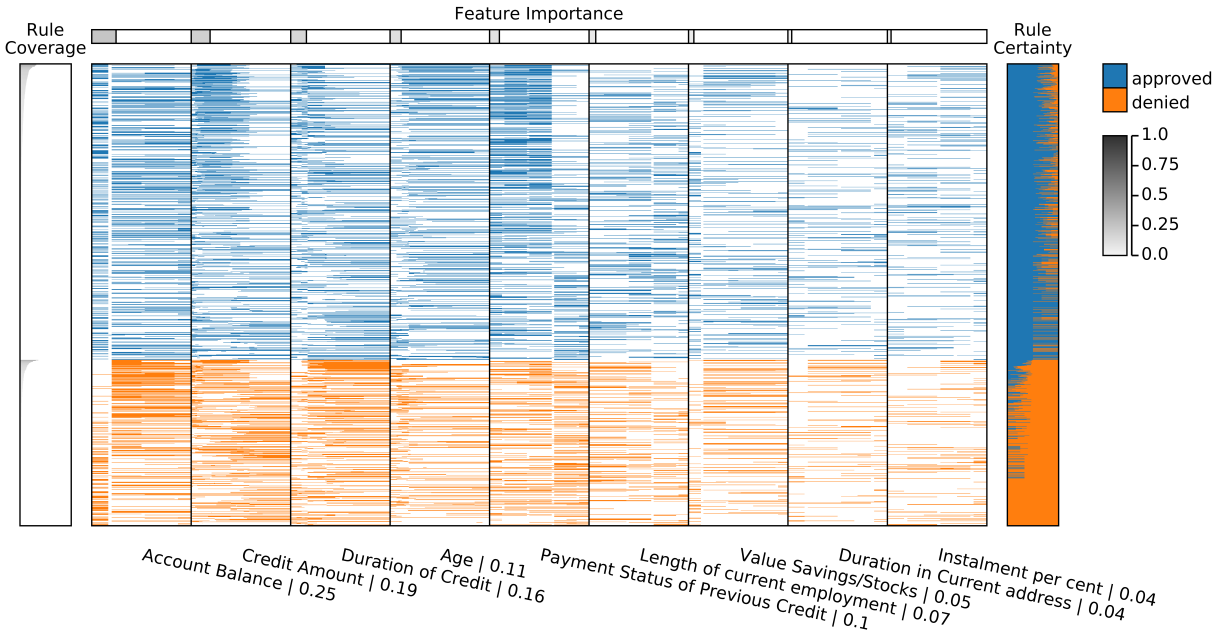


Figure 24 – ExMatrix GE representation of the RF model for the German Credit Data UCI dataset (subsection 2.4.2).

[1] POPOLIN NETO, M.; PAULOVICH, F. V. Explainable matrix - visualization for global and local interpretability of random forest classification ensembles. IEEE Transactions on Visualization and Computer Graphics, v. 27, n. 2, p. 1427–1437, 2021. Available: <https://-doi.org/10.1109/TVCG.2020.3030354>.

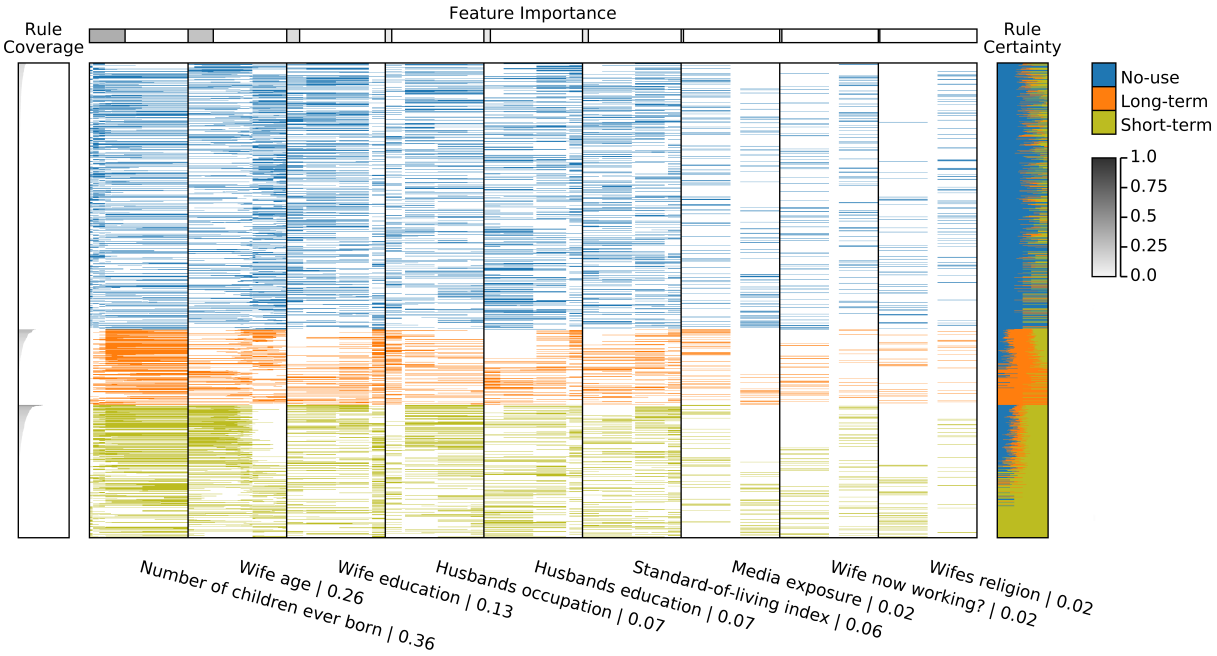


Figure 25 – ExMatrix GE representation of the RF model for the Contraceptive Method Choice UCI dataset (subsection 2.4.3).

Table 8 – Source code references for paper's sections.

Paper Section	Code
ExMatrix: Iris Dataset (section 2.3)	<https://popolinneto.gitlab.io/exmatrix/papers/2020/ieeevast/methodology/>
Use Case: Breast Cancer Diagnostic (subsection 2.4.1)	<https://popolinneto.gitlab.io/exmatrix/papers/2020/ieeevast/usecase/>
Usage Scenario I: German Credit Bank (subsection 2.4.2)	<https://popolinneto.gitlab.io/exmatrix/papers/2020/ieeevast/usagescenarioi/>
Usage Scenario II: Contraceptive Method (subsection 2.4.3)	<https://popolinneto.gitlab.io/exmatrix/papers/2020/ieeevast/usagescenarioii/>
Discussion and Limitations (section 2.5)	<https://popolinneto.gitlab.io/exmatrix/papers/2020/ieeevast/discussion/>

B.2 Why logic rules in a matrix-like visual metaphor instead of node-link diagrams?

In summary, our choice was based on the counterintuitive idea that disjoint rules are better than tree structures when analyzing DTs (LAKKARAJU; BACH; LESKOVEC, 2016). User tests have shown that transforming DTs into rules organized into tables (so-called Decision Tables) offers better comprehensive properties if compared to node-link diagrams (FREITAS, 2014; HUYSMANS *et al.*, 2011). Given the constraints of usual DT inference methods (non-overlapping predicates with open intervals), features can be used multiple times in a single decision path resulting in multiple nodes (one per test) using the same feature. Consequently, if tree structures are employed, the decision (predicate)

intervals for each feature need to be mentally composed by the user, and nodes using the same feature can be far away in the decision path, a hard and prone to error task. In the matrix representation, the decision intervals are explicit and can be easily compared across multiple rules and trees. Node-link diagrams are well recognized for representing tree structures (e.g., nodes hierarchy and connections) (GRAHAM; KENNEDY, 2010; SCHULZ; HADLAK; SCHUMANN, 2011), however, on DTs, the decision paths are the object of analysis (TAN; STEINBACH; KUMAR, 2005; FREITAS, 2014; HUYSMANS *et al.*, 2011; LIMA; MUES; BAESENS, 2009) and transforming paths into disjoint rules focuses on what is essential.

A single DT model can be quite complex regarding the number of nodes and depth. DTs can grow larger and deeper than the number of employed features, presenting redundancies in the tree structure. Moreover, DTs are usually unbalanced, so the number of nodes and tree depth can hamper visual scalability (GRAHAM; KENNEDY, 2010; SCHULZ; HADLAK; SCHUMANN, 2011). Just for illustration, Figure 26 shows only one DT from the RF with 128 trees of the paper use-case (subsection 2.4.1) using a simple node-link metaphor (that focuses on nodes relationships). This DT is unbalanced, and some decision paths are very long with repetitions of features. Figure 27 shows the resulting ExMatrix GE representation for this particular DT. Even with the redundancies of normalization, visual space is efficiently used since matrices are very compact metaphors. The most crucial point here is that by using a matrix representation, as discussed, the features are consistently ordered, and the decision intervals are explicit, as can be observed in Figure 28. Using common tree metaphors as in Figure 29, such information is very laborious to be extracted.

The problem with visual scalability is two-fold when dealing with RFs, as to compose a node-link representation from the RF with 128 trees of the paper use-case (subsection 2.4.1), 128 trees of similar complexity to Figure 26 need to be displayed concomitantly. Our matrix metaphor is a way to represent multiple trees in a single compact visual representation which can be (re)ordered to reveal different patterns - a feature difficult to be supported through usual tree visual metaphors given their focus on nodes (hierarchical) relationships (GRAHAM; KENNEDY, 2010; SCHULZ; HADLAK; SCHUMANN, 2011).

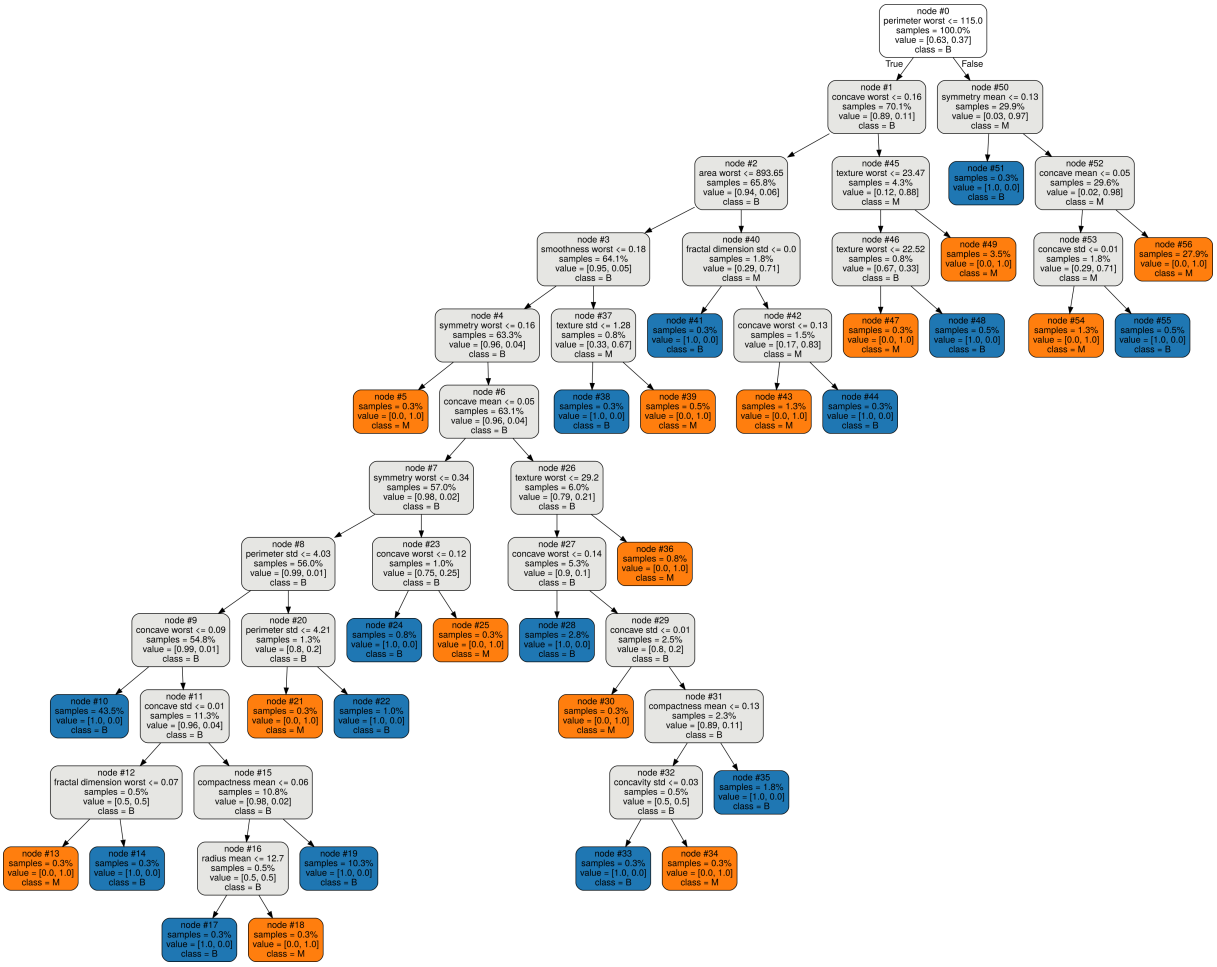


Figure 26 – Node-link diagram of DT_{49} from the RF with 128 trees of the paper use-case (subsection 2.4.1).

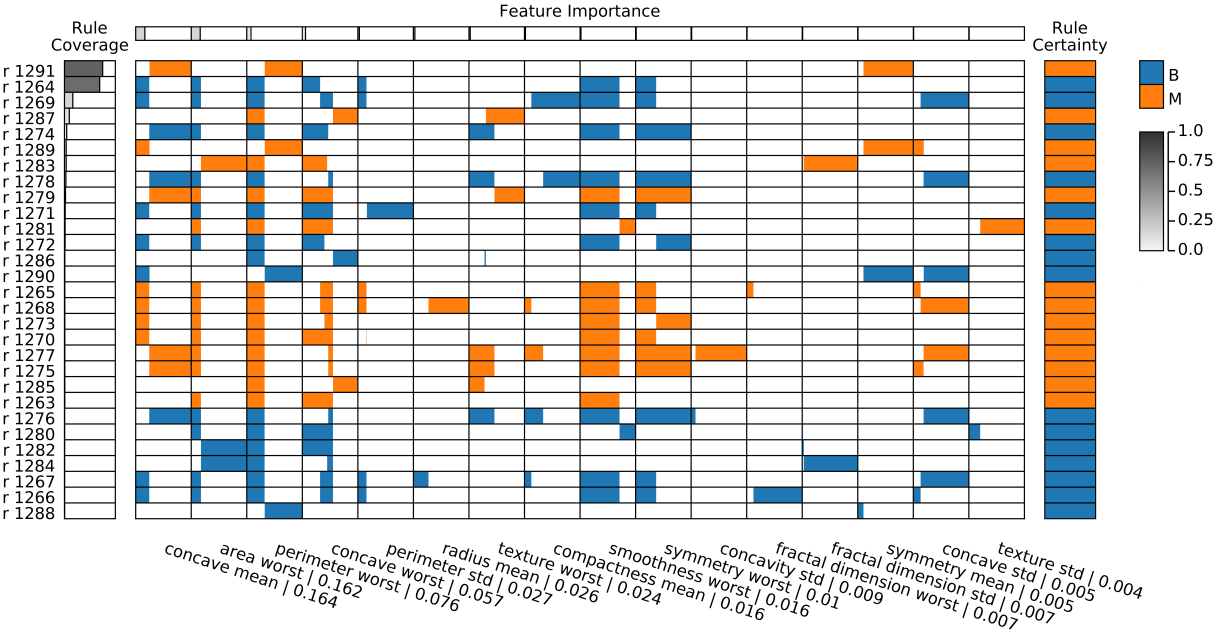


Figure 27 – ExMatrix GE representation of DT_{49} (Figure 26).

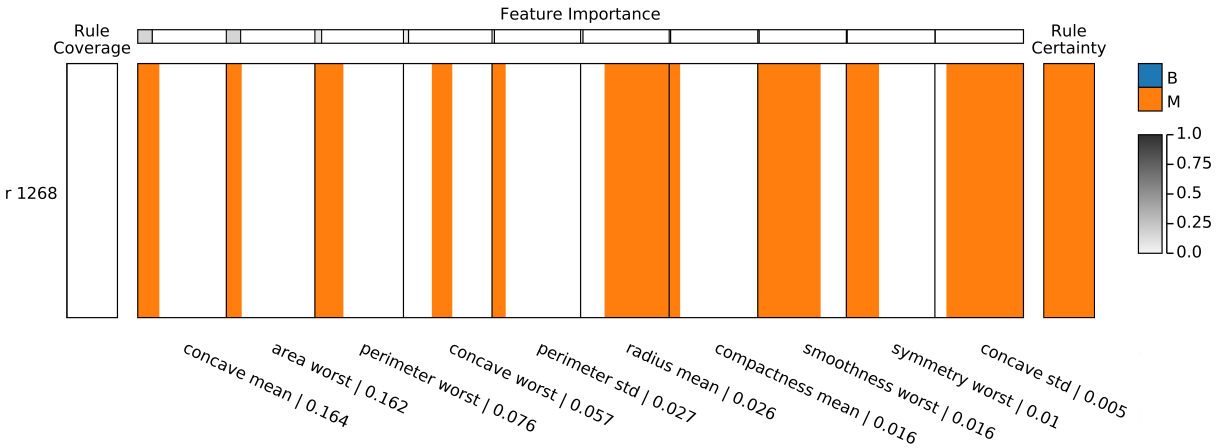


Figure 28 – ExMatrix representation of rule r_{1268} (sixteenth row in Figure 27) extracted from the decision path originating at root node #0 to leaf node #18 of DT_{49} (Figure 26).

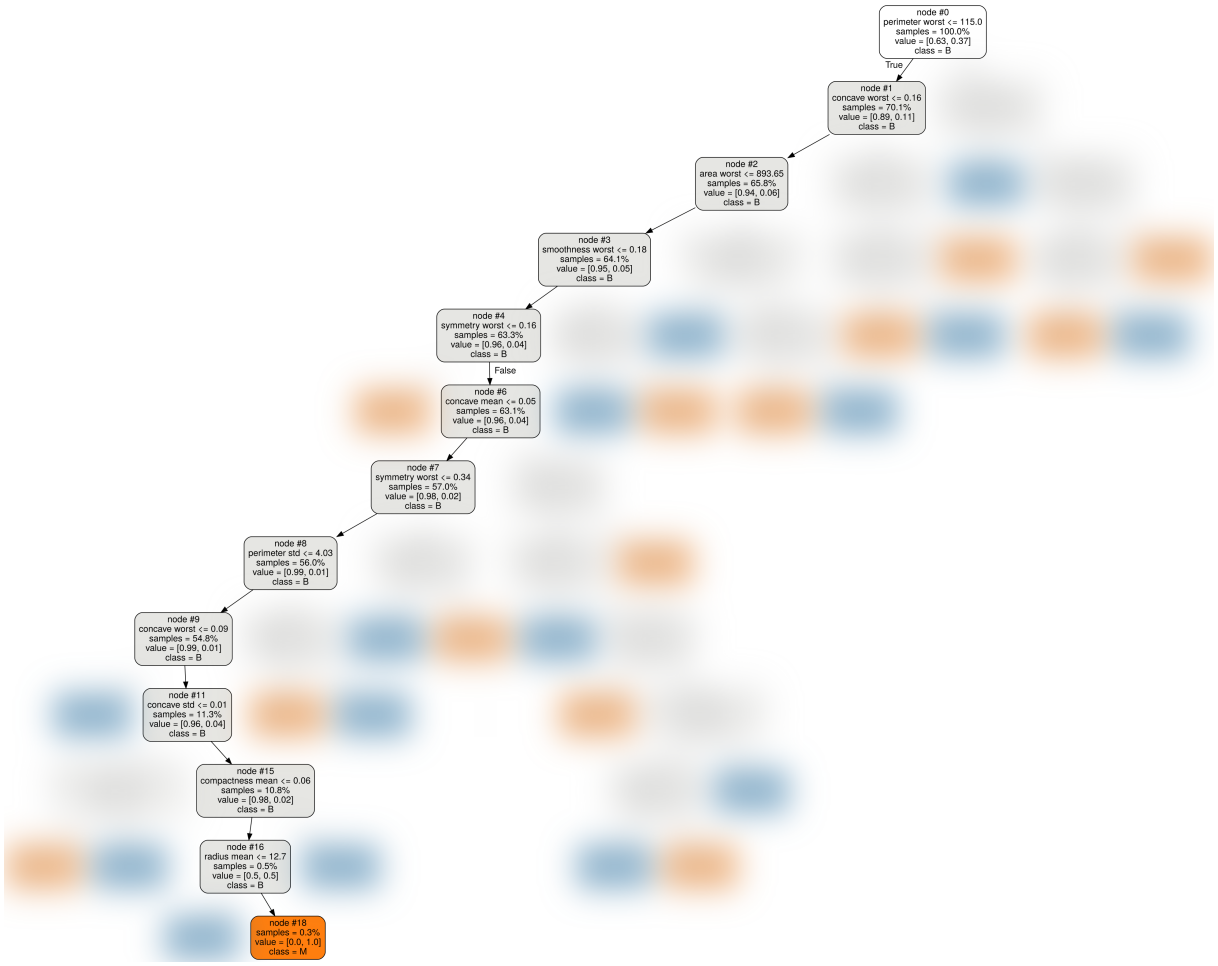


Figure 29 – Decision path originating at root node #0 to leaf node #18 of DT_{49} (Figure 26).

APPENDIX

C

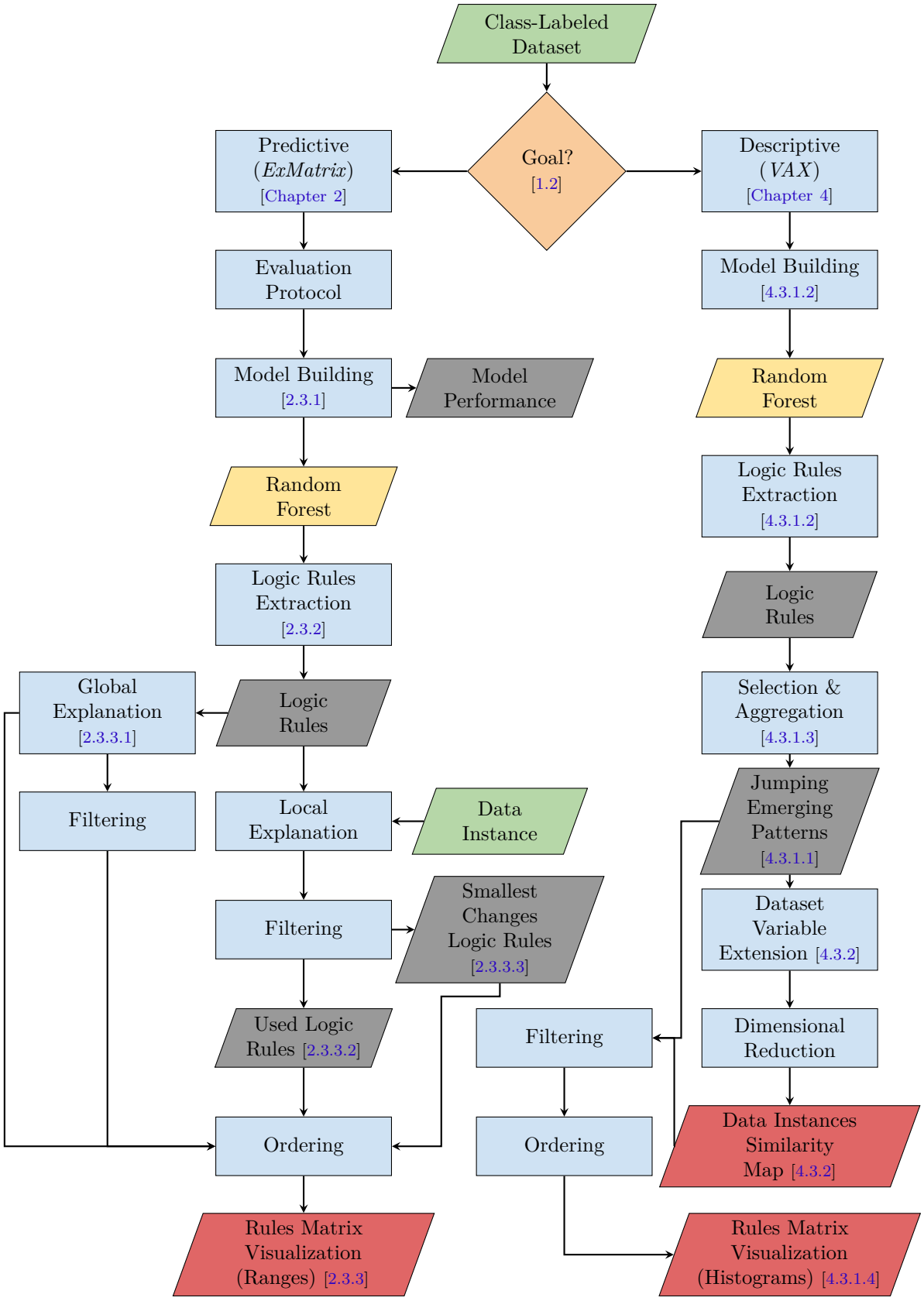
FLOWCHART-BASED SUMMARIZATION

This appendix shows in Figure 30 a flowchart-based summarization arranging inputs, processes, and outputs for ExMatrix and VAX. Both methods employ a matrix-like visual metaphor for logic rules visualization, where rules are rows, features (variables) are columns, and rules predicates are cells.

For model (predictive) explanations with ExMatrix (left goal in Figure 30), an RF model is built from a class-labeled dataset for predicting the class of new data instances (not seen during model training). An evaluation protocol must be chosen, providing model performance. Once all logic rules are extracted from the RF model, global and local explanations can be created. Global explanations provide model overview, where all logic rules may be presented or filtered. Local explanations allow reasoning about the classification process of a specific data instance through used rules and smallest changes rules. A crucial step on global and local explanations is rules (rows) and features (columns) ordering. On rules visual representations for model explanations, the predicates (cells) delimit ranges on features values represented by rectangular shapes.

For data (descriptive) explanations with VAX (right goal in Figure 30), an RF model is built from a class-labeled dataset to explain the data itself. The model must be created using all data instances and without bagging procedure (instances resample during model training) to avoid hidden dependencies. The logic rules extracted from the RF model are selected and aggregated, resulting in JEPs (descriptive logic rules). Histograms represent rules predicates (cells) for data explanations, thus requiring JEPs filtering for display. Ordering is also vital to produce meaningful visual presentations. Moreover, the original dataset is extended, creating new data variables under JEPs perspective. The extended dataset is then applied to DR techniques, enabling multidimensional projection visualization, originating a map for data instances similarity. The map can be browsed to filter JEPs for clusters and outliers investigations.

Figure 30 – A flowchart-based summarization of inputs, processes, and outputs for ExMatrix and VAX methods. From a class-labeled dataset, ExMatrix addresses models global and local (predictive) explanations, whereas VAX multivariate data (descriptive) explanations.



Source: Elaborated by the author.


ICMC
USP
SÃO CARLOS