

8th International Conference on Business Intelligence,
July 19 - 21, 2023 (CBI'23), Istanbul, Turkey

The 8th
International Conference on Business Intelligence,
July 19 - 21, 2023 (CBI'23), Istanbul, Turkey



Springer

BOOK ABSTRACT



جامعة السلطان مولاي سليمان
Université Sultan Moulay Slimane



Contents

Preface.....	4
Papers Abstracts	5
Section_01: Artificial Intelligence and Business Intelligence.....	5
Deep learning models for cybersecurity in IoT networks	5
Diabetes prediction by machine learning algorithms and risks factors	6
Machine learning prediction of weather-induced road crash events for experienced and novice drivers: Insights from a driving simulator study	6
Predicting driving license applicant’s performance for Car Reverse Test System	7
A business intelligence system for governing risks in SMEs	8
A Systematic Analysis for Machine Learning Based Cow Price Prediction	8
Ontology-based Mediation with Quality Criteria	9
Section_02: Optimization and Decision Support.....	10
A Comparative Study on the implementation of Blockchain in Supply Chain Models	10
A Hybrid Approach of Dijkstra’s algorithm and A* search, with an optional adaptive threshold heuristic	11
Localization and Navigation of ROS-based Autonomous Robot in Hospital Environment	12

Three-Phase Hybrid Evolutionary Algorithm for the Bi-Objective Travelling Sales-man Problem.....	13
Causal Discovery and features importance analysis: What can be inferred about at-risk students?	13
Computational features and applications of an inhomogeneous Gompertz diffusion process?.....	14
Understanding Data Journalism Acceptance among Social TV Users: A Case Study of Twitter in the United Arab Emirates.....	15
Unveiling the Performance Insights: Benchmarking Anomaly-Based Intrusion Detec-tion Systems using Decision Tree Family Algorithms on the CICIDS2017 Dataset213	16

Preface

In 2023, we had the great opportunity to organize the 8th edition of the International Conference on Business Intelligence (CBI 2023). The conference has grown to be a reputable event for the scientific and business communities. This book of proceedings collects papers accepted for presentation at CBI 2023.

CBI 2023 was organized by the Faculty of Sciences and Techniques (FST) and the laboratory of Information Processing and Decision Support (TIAD) at Sultan Moulay Slimane University along with the Association of Business Intelligence (AMID), and held during **July 19 - 21, 2023 (CBI'23), Istanbul, Turkey**.

For this edition, we received 50 submissions, which were reviewed by a Program Committee of 90 international experts in various fields related to business intelligence and decision support. Out of these submissions, the Program Committee decided to accept 15 regular papers, yielding an acceptance rate of 30%. The contributions are organized in topical sections: Decision Support and Artificial Intelligence; Business Intelligence and Database; and Optimization and Dynamic Programming.

As program chairs of CBI 2023 and editors of these proceedings, we would like to thank the President of Sultan Moulay Slimane University and the Dean of the Faculty of Sciences and Techniques for their support to the conference. In addition, we want to warmly thank again all the authors for their high-quality contributions and all the Program Committee members for their invaluable hard work. We also sincerely thank our keynote speakers for sharing their precious insights and expertise. Finally, our special thanks go to the Organizing Committee and to all the local arrangements coordinators. We cordially invite you to visit the CBI website at <https://www.cbi-bm.com/> and to join us at future CBI conferences.

Organizing committee Chairs

Rachid EL AYACHI

Mohamed FAKIR

Mohamed BASLAM

Papers Abstracts

Section_01: Artificial Intelligence and Business Intelligence

Deep learning models for cybersecurity in IoT networks

NAJI Meriem¹ and ZOUAGH Hicham²

¹ *Computer Sciences Department, Sultan Moulay Slimane University, Beni-Mellal, Morocco.*

meriemnaji39@gmail.com

² *Computer Sciences Department, Sultan Moulay Slimane University, Beni-Mellal, Morocco*

h.zougagh@usms.ac.ma

Abstract— The idea of the Internet of Things (IoT) was developed to improve people’s lives by providing diversity range of interconnected smart devices and applications in several areas. However, ensuring security in IoT environments remains a critical challenge, primarily due to the various security threats that devices face. While several approaches have been proposed to secure IoT devices, there is always room for improvement. One promising avenue is leveraging machine learning, which has shown its ability to identify patterns even in situations where traditional methods fail. Deep learning, in particular, offers an advanced approach to enhancing IoT security. One transparent option for anomaly-based detection is the utilization of deep learning techniques. This article introduces several approaches based on Recurrent Neural Networks (RNNs) using Long Short-Term Memory (LSTM), Autoencoders, and Multilayer Perceptrons. By harnessing the power of IoT, these anomaly-based Intrusion Detection Systems (IDS) offer the capability to effectively analyze all traffic flowing through the IoT network. The proposed model exhibits the ability to detect any potential intrusions or abnormal traffic behavior. To validate its effectiveness, the model is trained and tested using the NSL-KDD datasets, achieving an impressive accuracy of 97.85% for binary classification and 97.98% for multiclass classification.

Keywords—: IA, Network Intrusion Detection System, Deep Learning, Security, NSL-KDD.

Diabetes prediction by machine learning algorithms and risks factors

Youssef FAKIR

Sultan Moulay Slimane University, Beni Mellal, Morocco

Youssef.fakir@usms.ma

Abstract— Diabetes is a chronic disease that can have a serious impact on one's health; moreover, the risk of getting it can be decreased with early detection and care. For predicting diabetes, this study aims to compare the performance of six algorithms which are artificial neural networks (ANNs), decision tree (DT), support vector machine (SVM), K-Nearest Neighbors (K-NN), Naive Bayes (NB) and Random Forests models using common risk factors. These models are evaluated in terms of their accuracy, sensitivity, specificity, precision and F-measure. The algorithms were tested using three processes: three factors (glucose, BMI and age), five factors (glucose, BMI, age, insulin and skin) and for the last process all the patterns were used. The variables having the greatest impact on diabetic patients are identified from the association rules extracted, after the extraction of frequent variables by FP-Growth algorithm. By application of the algorithms mentioned above, the results showed that the random forest algorithm is considered as the best machine learning algorithm for the case of all factors but for the cases (3 factors) or (5 factors) Naive Bayes is better compared to the Random Forests algorithm.

Keywords— Artificial Neural Networks, Random Forest, K-nearest Neighbors, Decision Tree, Support Vector Machine, Naive Bayes, FP-Growth

Machine learning prediction of weather-induced road crash events for experienced and novice drivers: Insights from a driving simulator study

Zouhair Elamrani Abou Elassad¹, Mohamed Ameksa², Dauha Elamrani Abou Elassad¹, Hajar Mousannif²

¹ SARS Research Team, ENSAS, Cadi Ayyad University, Safi, Morocco

² LISI Laboratory, FSSM, Cadi Ayyad University, Marrakesh, Morocco

z.elamrani@uca.ma

Abstract— Road crashes are one of the most critical issues that pose a serious threat to our daily life; Crash occurrences prediction is a key role in designing efficient intelligent transportation systems. In this study, we aim to analyze

road crash events for experienced and novice drivers under several weather conditions during multiple driving simulations that have been conducted using a desktop driving simulator. This work outlined the effect of snow and rain conditions on driver behavior by endorsing real-time driver data namely: wheel angle position, throttle pedal position and brake pedal position. Moreover, optimized modeling strategies using the deep learning algorithm Multilayer Perceptron (MLP) and Support Vector Machine (SVM) along with Bayesian Networks (BN) models have been developed to analyze crash events. To the authors' knowledge, there has been a limited interest at assessing the impact of both snow and rainy weather conditions on the occurrence of crash events while providing a critical analysis for experienced and novice drivers based on driver entries; this approach fill the research gap of the combined effect of driving experience and weather conditions on road crash occurrence. The findings depict superior performances have been obtained when adopting the proposed strategy. As a whole, new insights into weather-induced crash events' investigation for experienced and novice drivers have been acquired and can be endorsed for designing effective crash avoidance/warning systems.

Keywords— Crash prediction, Machine learning, Novice drivers, Experienced drivers, Weather conditions, Data balancing, Driving simulator.

Predicting driving license applicant's performance for Car Reverse Test System

Abderrahman AZI¹ and Abderrahim SALHI² and Mostafa JOURHMANE³

*Information Processing and Decision Laboratory Sultan Moulay Slimane
University Beni-Mellal, Morocco*

¹ abderrahman.azi@gmail.com

² ab.salhi@gmail.com

³ m.jourhmane@usms.ma

Abstract— This research is focuses on evaluating the performance of applicants during driving license exam. The study utilizes the generated paths from the last developed part of the Car Reverse Test (CRT) system, which are categorized accordingly. The main objective is to interpret the path chosen by applicants and their vehicle control techniques in order to predict a score to be proposed to the supervisor. The collected applicant data are classified using data mining techniques using Multi-Layer Perceptron (MLP), Naïve Bayes (NB), Support Vector Machine (SVM) and Decision Trees algorithms. The

classification results are generally acceptable, with the MLP classifier achieving the highest accuracy percentage of 93% compared to other classifiers. Thus, this technique is strongly recommended regarding to its accuracy, followed by NB, SVM and DT, which achieved classification accuracies of 88%, 80% and 67 % respectively.

Keywords— Driving license exam, CRT system, Multi-Layer Perceptron, test score prediction, Naïve Bayes, Support Vector Machine, Decision Trees.

A business intelligence system for governing risks in SMEs

Abdelaziz Darwiesh¹, Mohamed Elhoseny², Reem Atassi³ and A. H. El-Baz⁴

¹ Faculty of Science, Damietta University, New Damietta 34517, Egypt

² Faculty of Computing and Informatics, University of Sharjah, Sharjah 26666, UAE

³ Faculty of Computer Information System, Higher Colleges of Technology, Dubai 00000, UAE

⁴ Faculty of Computers and Artificial Intelligence, Damietta University, New Damietta 34517, Egypt

adarwiesh@isods.org, melhoseny@ieee.org

Abstract— This article suggests a new business intelligence system for risk management in small and medium-sized enterprises (SMEs). It depends on social media perceptions. It uses a lexicon approach to analyze, identify and assess risks. It will help SMEs managers in making best decisions. Also, a mathematical formulation and explicit formulas for the proposed system are provided. In addition, Book Nook enterprise is studied as a use case to examine the potential risks through the tweets of customers. Further, various performance indicators are calculated to validate the effectiveness of the suggested model.

Keywords— Business Intelligence, Risk Management, Social Media, SMEs.

A Systematic Analysis for Machine Learning Based Cow Price Prediction

**A.K.M. Tasnim Alam¹, Zahid Hasan Nirob¹, Afrin Jahan Urme¹, Rashidul Hasan Hridoy¹, Md. Tarek Habib² and Farruk Ahmed²*

¹ Daffodil International University, Dhaka, Bangladesh

^{1*}tasnim15-12239@diu.edu.bd, zahid15-12438@diu.edu.bd
afrin1512209@diu.edu.bd, rashidul15-8596@diu.edu.bd

² Independent University, Bangladesh, Dhaka, Bangladesh

²md.tarekhabib@yahoo.com, 2farruk60@gmail.com

Abstract— Cow meets a significant number of demands for meats in South Asian countries, and a huge number of cows were sold in Bangladesh on the eve of Eid al-Adha. Cow prices depend on several factors, and determining the price of a cow is a cumbersome task for an inexperienced individual. Nowadays machine learning algorithms are significantly used for accurate price estimation. This study presents an efficient and accurate tool for determining cow prices using several characteristics of cows based on machine learning. Sixteen characteristics of a cow are considered in this study to determine its price. Four different machine learning algorithms were used and evaluated in this study for generating an accurate price prediction model. In this machine learning based systematic analysis, several experimental studies are conducted for evaluating models more precisely. Performance matrices were also used to evaluate machine learning algorithms.

Keywords— Systematic analysis, cow price prediction, machine learning, gradient boosting regression, decision tree regression, random forest regression, linear regression.

Ontology-based Mediation with Quality Criteria

Muhammad Fahad ¹

¹ Univ Lyon, Univ Lyon 2, UR ERIC 5 avenue Mendès France, 69676 Bron
Cedex, France

f.muhammad@univ-lyon2.fr

Abstract— Information integration has long history since human started using and collecting information. But, it has been strong focus of IT research since many recent years. It deals with providing a unified and transparent access to a collection of heterogeneous data sources. In information integration, the formulization of a global schema is a difficult task that manages multiple, autonomous and heterogeneous data sources. This paper presents a semantic system named OntMed for an ontology-based data integration of heterogeneous data sources to achieve interoperability between them. Our

system is based on the quality criteria (consistency, completeness and conciseness) for building the reliable analysis contexts to provide an accurate unified view of data to the end user. The generation of an error-free global analysis context with the semantic validation of initial mappings generates accuracy, and provides the means to access and exchange information in semantically sound manner. In addition, data integration in this way becomes more practical for dynamic situations and helps decision maker to work within more consistent and reliable virtual data warehouse.

Keywords— Information Integration, Ontology based Mediation, Ontology, Quality Criteria.

Section_02: Optimization and Decision Support

A Comparative Study on the implementation of Blockchain in Supply Chain Models

rof. Dhanashri Joshi¹, Atharva Naikwadi², Rohit Mokashi³, Mohit Pande⁴ and Sitanderpal Singh⁵

¹JSPM Rajarshi Shahu College of Engineering, Pune

¹dmjoshi_comp@jspmrscoe.edu.in

²JSPM Rajarshi Shahu College of Engineering, Pune

²atharva.s.naikwadi@gmail.com

³JSPM Rajarshi Shahu College of Engineering, Pune

³rohitvm1009@gmail.com

⁴JSPM Rajarshi Shahu College of Engineering, Pune

⁴mohitpande18@gmail.com

⁵JSPM Rajarshi Shahu College of Engineering, Pune

⁵sunnysingh94064@gmail.com

Abstract— This research intends to investigate the existing state, prospective uses in supply chain, future scope of this technology, dangers, and obstacles that we may encounter. We discuss the importance of blockchain in supply chain in Indian context. Each customer should be assured of the quality and credibility of items they would be using. This technology is most known for its application in crypto currency. It has the potential of increasing its reach to much more domains. The blockchain network is comparatively a better option than a traditional approach as each new block of data is attached to the

existing blockchain in a linear and chronological way. With each block connected to the preceding one, any bad actors wishing to tamper with data would need to decode prior blocks before reaching their desired data and rewriting the chain ahead while manipulating the rest of the nodes in a similar approach. As tampering with the data is both technologically difficult and commercially unfeasible, the security of a blockchain becomes two-fold.

Keywords— Blockchain, Solidity, Ethereum, Smart Contracts, Ethereum Virtual Machine

A Hybrid Approach of Dijkstra's algorithm and A* search, with an optional adaptive threshold heuristic

Lhoussaine AIT BEN MOUH¹, Mohamed OUHDA², Yousef EL MOURABIT and Mohamed BASLAM³

¹ Moulay Slimane University, FST Beni Mellal, Morocco

¹aitbenmouh@yahoo.fr, ²ouhda.med@gmail.com, ³m.baslam@usms.ma

Abstract— This study is a part of the trajectory planning applied to harvest system work where mobile robots must be able to navigate safely the environment to look for palmer crops. Many constraints can be faced, such as crop selection as maturity changes over time, searching for the most mature palmer, avoiding different kinds of obstacles, robot speed control, and the cost of moving from an initial point to a goal target. After studying different trajectory planning approaches and their applications [8], we conclude that some of these methods can be combined to design a new, powerful approach based on the accurate property of Dijkstra and the heuristic function of A Star. Dijkstra is known as a powerful algorithm based on graph mapping and reducing the path cost, and A Star on the other side is one of the best guides for path searching due to the heuristic function that avoids exploring all environment nodes and only those leading to the goal. Combining Dijkstra's algorithm and the A* (A-star) algorithm can lead to a more efficient pathfinding approach. Dijkstra's algorithm [4] is a well-known method for finding the shortest path between two nodes in a graph, while the A* algorithm is an extension of Dijkstra's algorithm that uses heuristic estimates to guide the search towards the goal node. By combining these two algorithms, we can use Dijkstra's algorithm to explore the graph and generate a good initial estimate of the path cost, then use the A* algorithm to refine the estimate and guide the search towards the goal node. This paper explores the utilization of trajectory planning in a harvesting system. By employing both the Dijkstra and

A* algorithms, we propose a hybrid approach to ensure optimal timing for finding a path. We conduct a comparative analysis to evaluate the performance of the new approach by comparing the application of a single algorithm versus the hybrid approach across various graph sizes.

Keywords— Trajectory Planning, Path Planning, Mobile Robot, Artificial Intelligence

Localization and Navigation of ROS-based Autonomous Robot in Hospital Environment

Hamza Ben roummane¹ and Cherki Daoui²

^{1,2} TIAD Laboratory, Sciences and Technology Faculty, Sultan Moulay

Slimane University, Beni Mellal, Morocco.

hamza.benroummane@gmail.com, daouic@yahoo.com

Abstract— This work is part of a research project during the COVID-19 pandemic that aims to design and develop a mobile autonomous robot for hospitals. In practice, implementing a navigation program directly on a physical robot is both expensive and hazardous. The solution is to perform a simulation using ROS (Robot Operating System), which offers several advantages that make it an appealing option for testing and development. In an unknown hospital environment, this paper presents a simulation of the navigation process of the autonomous robot Turtlebot3 by employing the Simultaneous Localization and Mapping (SLAM) algorithm, specifically the GMapping method, utilizing the distributed software framework of ROS. In a known hospital environment, we utilize trajectory planning algorithms designed for deterministic models. However, considering the inherent uncertainty in the environment and the inevitable inaccuracies of the models, we integrate the Markov decision process (MDP) by applying the classical Q-Learning algorithm. Through these simulations, our aim is to test and refine the navigation algorithms to enhance the performance of our mobile robot. Ultimately, the proposed simulation approach contributes to the development of robotic solutions that can assist in performing various routine tasks remotely. This saves time for healthcare personnel and, most importantly, ensures their safety.

Keywords— ROS, SLAM, GMapping algorithm, Turtlebot3 robot, MDP, Simulation, Q-Learning algorithm.

Three-Phase Hybrid Evolutionary Algorithm for the Bi-Objective Travelling Salesman Problem

Omar Dib^{1,}*

*Wenzhou-Kean University, Wenzhou, China
odib@kean.edu*

Abstract— In this research paper, we address the Bi-objective Traveling Salesman Problem (BTSP), which involves minimizing two conflicting objectives: travel time and monetary cost. To tackle this problem, we propose a novel three-Phase Hybrid Evolutionary Algorithm (3PHEA) that combines the Lin-Kernighan Heuristic, an enhanced Non-Dominated Sorting Genetic Algorithm, and a Pareto Variable Neighborhood Search. We conduct a comparative study comparing our approach with three existing methods specifically designed for solving BTSP. Our evaluation includes 14 instances of varying degrees of difficulty and different sizes. To assess the performance of the algorithms, we employ multi-objective performance indicators. The results of our study demonstrate that 3PHEA outperforms the existing approaches by a significant margin. It achieves coverage of up to 80% of the true Pareto fronts, indicating its superiority in solving the BTSP.

Keywords— bi-objective traveling salesman problem; hybrid evolutionary algorithms; non-dominated sorting memetic algorithm; pareto variable neighborhood search; multi-objective performance indicators.

Causal Discovery and features importance analysis: What can be inferred about at-risk students?

Ismail OUAADI¹ and Aomar IBOURK²

¹ *Laboratory of Research in Social and Solidarity Economy, Governance and Development (LARESSGD), Cadi Ayyad University, Marrakech, Morocco
ismail.ouaadi@gmail.com*

² *Laboratory of Research in Social and Solidarity Economy, Governance and Development (LARESSGD), Cadi Ayyad University, Marrakech, Morocco
aomaribourk@gmail.com*

Abstract— In this paper, we introduced machine learning and causal discovery algorithms that can be used as a way to determine the relevant characters of

students with low performances issues and to analyze their implications to highlight this type of students. Through this, we aim to provide some new useful insights that can allow to predict and explain the inherent relationships. By using six machine learning algorithms (Gradient Boosting, K-nearest neighbors, SVM, Random Forest, and Decision Tree) and four causal discovery algorithms (PC, GES, LinGAM, and GOLEM), we try to develop and use these models to analyze and draw conclusions from patterns and data. In this study, we present these algorithms to show the performance of the developed models in explaining the effect of variables and the nature of their relationship with low performing students. The results revealed that these models produce useful insights and highlight the existing relationship among students with low performances in reading and other student characters.

Keywords— Machine Learning · Feature importance analysis · Causal Discovery · at-risk student · Reading skills

Computational features and applications of an inhomogeneous Gompertz diffusion process?

Nadia Makhlouki¹, Ahmed Nafidi^{2,3}, and Achchab Boujemâa³

Hassan First University of Settat, Berrechid National School of Applied Sciences, Mathematics and Informatics Department, LAMSAD, B.P. 218, 26103 Berrechid, Morocco. n.makhlouki@uhp.ac.ma

Mohammed VI Polytechnic University, Lot 660, Hay Moulay Rachid Ben Guerir, 43150, Morocco.

Abstract— here are several uses for stochastic diffusion models. They have received special attention from several scientific disciplines, including biology, physics, chemistry, medical science, and mathematical finance. In this paper, We consider the Gompertz diffusion process based on stochastic inhomogeneous model. We begin by obtaining the analytical formulation for the process's probabilistic properties, the mean functions (conditional and non-conditional). Then, with the maximum likelihood technique and discrete sampling, we estimate the model's parameters. Finally, we used the stochastic inhomogeneous Gompertz diffusion process to analyze the development of the electric power consumption in Morocco in order to assess this method's capacity for modeling actual data.

Keywords— Inhomogeneous Gompertz diffusion model · Stochastic differential equation · Statistical inference · Mean function · Electric power consumption in Morocco.

Understanding Data Journalism Acceptance among Social TV Users: A Case Study of Twitter in the United Arab Emirates

Faycal Farhi^{1*}, Riadh Jeljeli², Abdullah Ali Al Marei³, Khaled Zamoum⁴

¹College of Communication and Media – Al Ain University - United Arab Emirates,

Email: faycal.farhi@aau.ac.ae

²College of Communication and Media – Al Ain University - United Arab Emirates,

Email : Riadh.jeljeli@aau.ac.ae

³Media and Communications Department, Faculty of Humanities, King Khalid University, Abha- Saudi Arabia,

Email: aalmarai@kku.edu.sa

⁴College of Communication, University of Sharjah, United Arab Emirates,

Email: Kzamoum@sharjah.ac.ae

Abstract— Today Social TV users prefer data journalism for different reasons. One of the basic reasons behind this behaviour is audiences' growing interest in transparent and detailed news reports. This research also examined the factors proposed by the Uses and Gratifications theory and the Technology Acceptance Model in the acceptance of data journalism among Emirati Social TV users. Data gathered from structured survey questionnaires were analyzed by using Structural Equation Modelling. Results showed that Information Sharing and Information Seeking are two significant factors linked to users' gratifications. These gratifications, on the other hand, significantly affect their data journalism acceptance. Besides, the mediating effect of perceived usefulness also remained significant, indicating the expected useful outcomes are important determinants of their data journalism acceptance. UOG and TAM significantly predict data journalism acceptance among Emirati Social TV audiences. Thus, it is concluded that data journalism has become a valuable source of information for social TV users, providing detailed insights into complicated issues in an easily understandable and engaging format. Besides, perceived ease of use remains an important factor in social TV users' acceptance of data journalism, as it contributes to their willingness to engage

with this content. Thus, as the social TV audience evolves from Twitter users to other platforms, it is usefully integrated with data journalism.

Keywords— Data Journalism, Social TV, Twitter, United Arab Emirates, Uses and Gratifications, Technology Acceptance Model

Unveiling the Performance Insights: Benchmarking Anomaly-Based Intrusion Detection Systems using Decision Tree Family Algorithms on the CICIDS2017 Dataset

Mohamed AZALMAD, Rachid EL AYACHI and Mohamed BINIZ

Laboratory TIAD, Department of Computer Science, Sultan Moulay Slimane University, MOROCCO

*azalmad.mohamed@gmail.com, rachid.elayachi@usms.ma,
mohamedbiniz@gmail.com*

Abstract— The continuous growth of computer networks and the internet has brought attention to the increasing potential damage caused by attacks. Intrusion Detection Systems (IDSs) have emerged as crucial defense tools against the rising frequency and so-phistication of network attacks. However, effectively detecting new attacks using machine-learning approaches in intrusion detection systems presents challenges.

This study focuses on the CICIDS2017 dataset, which is one of the most recent and updated IDS datasets publicly available. The CICIDS2017 dataset contains both benign and seven common attack network flows, meeting real-world criteria and providing true network traffic data. Furthermore, The CICIDS2017 dataset presents challenges when it comes to measuring the performance of a comprehensive set of machine learning algorithms in order to identify the optimal pattern set for detecting specific attack categories.

This paper contributes to the field of intrusion detection systems by benchmarking the decision tree family. As a result of our study XGBoost achieves the highest accuracy of 99%, followed by Random Forest with 98%, Gradient Boosting Trees with 88%, and Decision Tree with 89%.

Overall, this research provides valuable insights into the performance of decision tree family and feature selection methods, paving the way for the advancement of more reliable and efficient intrusion detection systems.

Keywords— CICIDS-2017, Decision tree family algorithms, IDs, Decision tree, Random Forest, Gradient Boosting, eXtreme Gradient Boosting
