IFSEAI: Interpretable Feature Selection for Explainable Artificial Intelligence Special track in DBKDA 2025

Basabi Chakraborty
Madanapalle Institute of Technology and Science
Andhra Pradesh, India
Iwate Prefectural University
Iwate, Japan
email:basabi@iwate-pu.ac.jp

Ashis Kumar Mandal
Department of Computer Science and Engineering
Hajee Mohammad Danesh Science and Technology University
Dinajpur, Bangladesh
e-mail:ashis@hstu.ac.bd

***Abstract*—This paper summarizes the research presentations in two sessions of the special track "IFSEAI: Interpretable Feature Selection for Explainable Artificial Intelligence" along with The Seventeenth International Conference on Advances in Databases, Knowledge, and Data Applications, DBKDA 2025, held from March 09, 2025 to March 13, 2025 in Lisbon, Portugal. The research studies of this track were expected to deal with the issue of interpretability and explainability in the development of machine learning models. The contributions in this track proposes various model agnostic tools for improving interpretability of general machine learning models as well as presents studies on several explainable machine learning algorithms for specific real world applications.**

Keywords-eXplainable Artificial Intelligence (XAI); Interpretability; Feature Selection.

I. Introduction

In recent years, eXplainable Artificial Intelligence (XAI) has gained significant attraction in developing trustworthy AI systems, especially for risk sensitive or safety critical systems [1]. Performance evaluation in terms of only classification accuracy, a uni-dimensional metric, can not satisfy users in the area of medical diagnosis, financial prediction or airlines traffic control decision systems. The fundamental problem of trust leads to questions such as whether humans do more mistakes or are the mistakes harmful or how to react when things go wrong. Interpretability of the decision system is one important factor to deal with these questions. Explainability of a model summarizes the model behavior and explains the causes of decisions while interpretability is comprehension of the model's actions. Building interpretable model is the first step towards development of XAI models.

Feature selection is the key preprocessing step behind the design of successful decision system as it directly impacts the Machine Learning (ML) model's performance and clarity. Interpretable feature selection brings transparency to the black box process, providing scope of designing ML models, which in turn leads to the development of explainable and trustworthy AI system. The objective of this special track, IFSEAI, is to provide a platform to explore theoretical advancements and present new, efficient algorithms for interpretable feature selection as well as explainable ML models. The track also aims to bridge the gap between theoretical research and real-world applications by encouraging research works on theoretical exploration, algorithm development as well as practical applications. The next section represents a short summary of the contributions in this special track followed by the conclusion in the final section.

II. Summary of Submissions

In the first paper on "Exploring Latent Concepts in SHAP Values - A New Approach Using Singular Value Decomposition" by Yukari Shirota and Tamaki Sakura [2], the authors integrated SHAP (SHaply Additive exPlanations) with SVD (Singular Value Decomposition) to propose SHAP-SVD, a tool for analyzing the latent concepts expressed by SHAP values of several explanatory variables in regression problem. The authors applied the proposed tool for analysis of stock price growth rate data of 67 Japanese and Indian automakers. The target variable growth rate is dependent on four explanatory variables. The SHAP value of four variables of each of the companies are calculated and SHAP matrix (67×4) is formed. Singular Value Decomposition (SVD) of SHAP matrix leads to the diagonal matrix defining the principal components. The first and second principal components are extracted and analyzed to find the underlying latent concepts expressed by the SHAP values. Finally the companies were grouped into two categories based on the concepts associated with the principal components of the SHAP matrix. The proposed SHAP-SVD algorithm presents a way to uncover the underlying similarities of the companies in categorization by defining latent features.

The second submission on "Route Planning in Wildfire Areas by Integrating a Modified A^* Algorithm with Deep Learning" by Manavjit Singh Dhindsa et al. [3] presents a framework for safe path planning that integrates wildfire spread predictions from state-of-the-art deep learning models with an optimized $A^*(OA^*)$ algorithm. The proposed approach utilizes binary fire masks to generate safe and efficient evacuation routes while adhering to strict safety constraints, such as maintaining buffer zones around fire-affected regions. The binary fire mask is generated by prediction of wildfire spread from training samples using a combination of two deep learning neural network architectures. The proposed framework offers a robust solution for real time evacuation planning, contributing to the broader efforts of wildfire management and disaster mitigation.

In the third contribution on "Time-Series Topic Analysis of Large-Scale Social Media Data using Two-stage Clustering " by Takako Hashimoto [4], the author introduces a two-stage clustering method to extract coarse grained topics from social media text data. In the first step, graph clustering extracts some micro-clusters from graphs which are generated on the basis of the similarity of the posts of the users, with each micro-cluster representing some fine topic. The time series of these micro-clusters are then analyzed in the second stage through time series clustering to reveal more general topics. In this work, the proposed method is applied to Yahoo Japan News Comments related to the election of two specific candidates in Japan to extract people's reactions to the candidates before and after the election. The general opinion topics extracted by the proposed method seem to match the ground truth justifying the methodology to uncover hidden topics in social media.

In the fourth contribution on "Evaluating the Potential of SHAP-Based Feature Selection for Improving Classification Performance" by Ashis Kumar Mandal and Basabi Chakraborty [5], the authors investigate the effectiveness of feature selection using SHAP (SHapley Additive exPlanation) which gas gained attention for its interpretability and ability to quantify the contributions of individual features to model predictions.The research utilizes 10 diverse datasets to evaluate Linear SHAP's capability in identifying relevant features for classification tasks. The performance of Linear SHAP is assessed across varying percentages of selected features and compared to classification models without feature selection. Three popular filter-based feature selection approaches: Chi-square(Chi^2), Mutual Information, and Correlation-based methods are also used for feature selection with the same bench mark data sets. Comparative analysis, supported by statistical significance tests, demonstrates that Linear SHAP performs equally well to the traditional methods while offering the added benefit of interpretability. The findings suggest that Linear SHAP is a viable and promising alternative to popular feature selection techniques for classification tasks.

In the fifth and the final submission on "Visualizing Proximity of Audio Signals from Different Musical Instruments - A Two Step Approach" By Goutam Chakraborty, Cedric Bornand and et.al [6], the authors proposed a method for visualizing audio signals from various musical instruments to identify their characteristic features and quantify their similarities. The work has been conducted in two stages. In the first stage, audio features are extracted and compressed using several feature extraction methods. In the second stage, the audio signals are projected into two dimensional plane using unsupervised visualization tools. The aim of this work is to determine which combination of feature extraction and visualization tool can produce the most interpretable and explainable separation of the audio signals from different sources.

III. Conclusion

In the special track IFSEAI with two sessions IFSEAI I and IFSEAI II, two of the five submissions [2] [5] dealt with proposals of tools that can improve classification accuracy or interpretability of clustering by using interpretable features. Both the paper extended the idea of Linear SHAP for incorporating explainability in machine learning models. Though the initial experiments support the efficiency of SHAP based tools but more experiments are needed for various high dimensional data sets. One submission [4] deals with text data and proposed topic extraction tool for time series text data analysis for interpretation of events. This work is application oriented and demonstrates success in the case study, but needs to be justified for a wide range of applications. The other two papers [3] [6] propose techniques for improved prediction with explainability and visualization of hidden information in the data in two real world applications respectively. Both papers show a direction of research regarding how to combine presently available machine learning tools to develop techniques for enhancing explainability in machine learning models for a particular real world application.

Acknowledgment

We are grateful to the organizers for providing us the opportunity of holding this special track in the conference and help us with all the logistics and administrative resources. Our special thanks to all the reviewers for providing their valuable evaluations and comments on the submitted contributions. Finally we deeply express our gratitude to all the authors for their hard work without whom the track would not exist.

References

[1] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (xai)," *IEEE Access*, vol. 6, pp. 52 138–52 160, 2018.

[2] Y. Shirota and T. Sakura, "Exploring latent concepts in shap values -a new approach using singular value decomposition," *in Special Track, IFSEAI: Interpretable Feature Selection for Explainable Artificial Intelligece, along with DBKDA*, 2025.

[3] M. S. Dhindsa, K. S. Naik, and et. al., "Route planning in wildfire areas by integrating a modified a* algorithm with deep learning," *in Special Track, IFSEAI: Interpretable Feature Selection for Explainable Artificial Intelligece, along with DBKDA*, 2025.

[4] T. Hashimoto, "Time-series topic analysis of large-scale social media data using two-stage clustering," *in Special Track, IFSEAI: Interpretable Feature Selection for Explainable Artificial Intelligece, along with DBKDA*, 2025.

[5] A. K. Mandal and B. Chakraborty, "Evaluating the potential of shap-based feature selection for improving classification performance," *in Special Track, IFSEAI: Interpretable Feature Selection for Explainable Artificial Intelligece, along with DBKDA*, 2025.

[6] G. Chakraborty, C. Bornand, and et. al., "Visualizing proximity of audio signals from different musical instruments - a two step approach," *in Special Track, IFSEAI: Interpretable Feature Selection for Explainable Artificial Intelligece, along with DBKDA*, 2025.