National College of Ireland

National College of Ireland

Project Submission Sheet

Student Name:	Dhruv Sharma
Student ID:	x22228268
Programme:	MSc in Data Analytics **Year:** 2023-2024
Module:	Data Mining and Machine Learning 2
Lecturer:	Anu Sahni
Submission Due Date:	19/05/2024
Project Title:	DMML2-TABA
Word Count:	8458

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the references section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature:	Dhruv Sharma
Date:	19/05/2024

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. Projects should be submitted to your Programme Coordinator.
3. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
4. You must ensure that all projects are submitted to your Programme Coordinator on or before the required submission date. **Late submissions will incur penalties.**
5. All projects must be submitted and passed in order to successfully complete the year. **Any project/assignment not submitted will be marked as a fail.**

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Volcanic Eruption Prediction

Dhruv Sharma
MSc Data Analytics
National College of Ireland
Dublin, Ireland
x22228268@student.ncirl.ie

***Abstract*— The main objective of this project report is to provide a synthesis and compilation of recent studies so that the latest development toward the prediction of volcanic eruptions may be looked into. The report looks into various methodologies and technologies for the prediction of volcanic eruptions, including machine learning algorithms, remote sensing techniques, and integration with different data sources such as seismic data, SAR imagery, and ionospheric anomalies. These approaches show the diversity and innovativeness of the current methods for prediction. Most of these prediction approaches, along with many technologies, have their pros and cons. This report will show a balanced discussion on the benefits, from enhanced accuracy to real-time monitoring capabilities, and limitations, with more of the computational demands and issues in good quality data, of various approaches, including machine learning, satellite monitoring, and multispectral imaging. This critical analysis will allow the identification of what works best and what needs enhancement. Such insights have been drawn from most of the recent studies, therefore emphasizing the trends in the increased usage of AI and machine learning, integration with multisource data, and remote-sensing technologies, hence proposing potential future research directions that need to be followed to address the prevailing challenges and further improve the predictive capability. Summarizing these insights, the current report is likely to serve the scientific community with the current state of volcanic eruption prediction, in advancing risk management and preparedness for these disasters, to cater to areas of protection for human life and property, as well as the health of people who dwell in or near the vicinity of volcanoes.**

Keywords*— *Volcanic eruption, seismic data, machine learning.

I. Introduction

Volcanic eruptions are one of the most disastrous natural phenomena that occur on Earth. In addition to immediate destruction leading to human and material casualties, eruptions also result in severe and widespread economic consequences and long-term damage to the environment. Identifying ways to predict these events with decent accuracy is paramount to minimizing the impact, as it will allow creating more appropriate risk response measures. Due to the serious implications of being unprepared, developing reliable predictive methods should be a priority in disaster management. Fortunately, modern technology and improved data analysis capabilities make it easier to predict such events accurately, and hence we are hopeful of improving readiness against such natural disasters.

In summary, the material in this report demonstrates how machine learning, remote sensing, and other technologies have all played pivotal roles in the recent prodigious advances in the prediction of volcanic eruption. Machine learning is necessary for interpreting the bewildering amount of refined seismic data required to predict eruptions. Remote sensing methods allow for the monitoring of the activity of a volcano nearly in real-time from hundreds of miles overhead in space. Together, these tools have paved the way for a dramatically improved understanding of volcanic activity and the possibility to predict it.

It proceeds with relevant methodology to predict volcanic eruptions using machine learning techniques in line with all available data sources and up-to-date analytical tools that should be used in reducing risks related to this type of natural disaster. The report introduces a framework that will enhance greatly the state of preparedness related to the response of disasters through good integration between machine learning models, data pre-processing, and novel analysis strategies. This will ensure that the right tools and, more importantly, information on how to manage and respond to volcanic threats have been reached by the local government and all the stakeholders acknowledged in disasters up to the disaster response teams if identified.

II. Exploratory Data Analysis

An EDA and cleaning process of data is an important stage in a data science project. The latter helps in getting an idea of what underlies the information in the data. They help clean the data, leading to appropriate and correct predictive models. This subsection describes the main assumptions pertaining to data, approach, and techniques used for EDA and data cleaning in this study related to predicting volcanic eruptions. The environment is such that it is highly volatile and unpredictable, so seismic information gathered using sensors near volcanic areas inherently contains data that include noise, missing values, and possibly anomalies. In order to be sure that the developed models for prediction have to be credible and accurate, it is important to understand these inherent issues while developing robust data preprocessing and analysis techniques.

The first step of data preprocessing in predicting volcanic eruptions is a solid inspection of the dataset. This includes identifying the data type of the features, whether they are numeric or categorical; detecting missing values; and reviewing summary statistics, such as mean, median, and standard deviation. Period of the data: Understand how relevant the features are, such as seismic activity and gas emissions, to the predictions. The missing data imputation process for continuous variables is done using the median, whereas imputations are done using the mode for categorical variables which is effective in removing bias with large standard deviation, regardless of type of data. The dataset can be wavelet transformed to denoise the information. This will, therefore, help to suppress random noise, hence keeping the signal clear. Outliers should be flagged and treated using different outlier treatment processes such as the Interquartile Range (IQR) method utilising Z scores or transformed with techniques like log transformation or Winsorization to lessen their influence on the model predictions.

Therefore, the next move should portrait what standardization entails and touch on how in data preprocessing the tantrums of having a wide featured dataset are curtailed with the view of making sure that the distance metrics won't experience much higher dimensionalities than those make other sense at all in it; as well as aiding practitioners who employ models based on this kind by excluding their concernfulness about features whose values lie far apart from each other. Methods suggested for prediction of volcano eruptions lie within different scale transformation methodologies like Min-Max Scaling where one re-scales attribute values into the interval [0,1], It is sensitive to outliers due to this reason aloneserd; Z-Score Normalization is a technique used to make features' means zero and standards deviations one. Folideists by scaling brings all featurers close together thus handling them differently from each other. This helps in reducing prejudice and guaranteeing equilibrium between positive or negative forecasts regarding future massive lava flow events; thus z-score normalization should be preferred over range normalization in this particular geological situation.

III. Dimensionality Reduction

The estimation of model performance requires a dimension reduction of non-essential features to reduce overfitting. Hence, dimensionality reduction and the features selection process shape the data into the form that renders the predictive model both efficient and interpretable. This section brings forth how dimensionality of the features can decreased and some of the applied methods of feature selection in the context of volcanic eruption data.

Dimensionality reduction and feature selection techniques play a huge part in the approach to high-dimensional data handling in this project. The curse of dimensionality is a major problem in datasets having more features than observations which contribute to poor generalisation and high computational expenses. An important reduction method is PCA, which can help in reducing the dimensions of while keeping the significant variance of the data, which makes the data less complex for further processing. After doing the reduction, feature selection methods can further be applied to reduce the dataset's complexity with the aid of methods like the Random Forest feature importance and Recursive Feature Elimination by which the most predictive features can be selected. Hence, the proposed technique will ensure finer generalization, better interpretability as well as lower complexity.

IV. Feature Engineering

Feature extraction and engineering are two crucial stages in converting raw seismic data into informative input information to be fed into a machine learning model. Rolling statistics, such as rolling averages, standard deviations, and rolling skewness, are the quantitative features designed to capture trends and patterns quantitatively. Rolling averages remove these short-term fluctuations to extract the long-term trends, while lagged features bring in the past moments, therefore capturing temporal dependence. Spectral features are derived from signal characteristics representing Fourier Transformations from time-series data into a frequency domain. This makes it good for detecting existing frequencies that may indicate volcanic activity, hidden patterns, and periodicities within the data to give good insights into prediction.

Feature engineering and extraction are then carried out through the analysis of the spectrogram using CNN. Spectrograms are, in fact, the graphical representative of the frequency content of the signal over time. CNN is found to be very strong in classifying complex spatial–temporal characteristics from an analysed spectrogram in the frequency domain of the signal. CNN is capable of detecting complex features and interrelation among the data with an image-like structure and is perfect for analysing the spectrogram. It will provide a good increase in interpretability of the model and the accuracy and reliability of results.

Specifically, in this work, the expert domain knowledge about volcanic activity is used to extract specific features of the domain in order to ensure more accurate prediction models. For example, the critical gas emission ratios indicate the changes in behaviour, which cannot be inferred from the absolute values of gas emission alone. The windowed sum of seismic events that records the frequency of seismic events at different time scales, hourly, daily, or even weekly, is recorded since high-frequency volcanic activity should precede the eruption. The following is a calculation of the cumulative seismic energy, that is, the accumulated sum of energy released from seismic activity in the time window, since the total amount of released energy in the window provides the general overview of seismic activity as well as its intensity in the period preceding the eruption. These specific features make the models deep and insightful enough to ensure that the models are accurate and reliable in predicting the future states of volcanic phenomena.

Hence, it may be the rationale behind including all these features—temporal, spectral, spectrogram, and domain-specific features in this project, as it allows the capturing of the full range of patterns and relationships latent in the seismic data, necessary for the correct prediction of volcanic eruptions. This holistic approach considering different engineered features will be robust and generate reliable models having much-enhanced accuracy in the prediction of volcanic eruptions.

V. Choice of Modelling Techniques

Another important factor for precise prediction is the choice of the correct model. Advanced deep learning models and ensemble learning techniques need to be applied, considering the complexity in the prediction of a volcanic eruption. This section outlines the methodologies used in the four reference papers as well as some fresh strategies explaining how the proposed techniques can be implemented and utilised in the context of volcanic eruption prediction.

A. Random Forest

The technique randomly picks the next predictor at each node in the tree and further builds numerous unpruned trees to result in what is known as extremely randomized trees. This makes Random Forest an ideal approach that is good for both regression and classification tasks with datasets involving larger dimensions, being resistant to overfitting. It builds up an ensemble of decision trees while training and provides the mode of the class labels (in classification) or the average prediction (in regression) for every individual tree. The model inherently generates feature importance scores that really help one identify which features contribute to making a prediction. The Random Forest deals with continuous as well as categoric data, generates important feature scores, while reducing overfitting by averaging, making it versatile and robust—particularly useful during the early modelling phases (Shao et al., 2024; Tutumlu & Saraç, 2024). However, Random Forest is computationally intensive and less interpretable than simpler models.

Besides, the characteristic of a random forest model to be able to handle mixed data types and resist overfitting makes it a nearly perfect selection for initial model building. The feature importance scores of random forest models aid in understanding those variables that are predictive and enhance interpretability. Simple models such as linear and logistic regression could not be applied since the data in seismic belong to complex and inseparable nonlinear relationships.

B. *Gradient Boosting Machines (GBM)*

Use multiple Gradient Boosting Machines such as XGBoost and LightGBM to predict with high accuracy. Capable of capturing complex and interesting patterns in the data, GBM builds models in a sequential manner. In that process, the new model corrects the error that was possibly made by the previous one. Missing values can inherently be handled by techniques like XGBoost, which makes feature selection really very convenient. Further, feature importance ranking can be done with GBM models, and after that, a good insight into the data would be possible, followed by effective feature picking.

Such models as XGBoost and LightGBM within this category, with high accuracy and performance, respectively, treat missing data well and capture complicated interactions among features (Ghafari et al., 2023; Senjab et al., 2023). Although high in performance, these models are computationally expensive and risk inducing overfitting if careful tuning is not done. Gradient boosting machines, for instance, XGBoost and LightGBM, are known to have strong predictive performances by ensuring that the model can recover from missing data owing to complex interactions of the features. These models have high tolerance for heterogeneous data and nonlinear relationships, generally outperforming simpler models. Hence, GBMs generate robust models comparing with XGBs thus by providing an elaborate ranking of feature importance predictive models can be enriched, but hyperparameter tuning is still difficult.

C. *Deep Learning Models: Long Short-Term Memory (LSTM)*

LSTMs are a type of Recurrent Neural Network designed to capture time dependencies in data series. LSTMs are well suited for data that are structured in a sequential manner, giving some relationship to the time series based on the order of observations. LSTMs are designed to track dependencies over long input sequences, and this can be very efficient for time series prediction in the prediction of volcanic activity, where the past can have a lot of influence on the future. In this case, LSTMs could capture important long-range dependencies and patterns which are crucial for the prediction of volcanic eruptions, based on the historical activity data.

LSTMs are well designed for the problem of predicting time series, since they can model sequenced data and capture dependencies that are long-termed. However, these variations in input sequence lengths are handled, but they are computationally expensive and hard to interpret. The LSTMs should successfully capture long-range persistence and long-term patterns because time-series data will be included in volcanic prediction. Traditional RNN and even simple feedforward neural network architectures are rejected because they have problems with long-term dependencies and can suffer a vanishing gradient problem, for which LSTMs are designed.

In order to forecast volcanic eruptions with greater precision, we suggest that different advanced approaches like Random Forest, Gradient Boosting Machines (GBM), and Long Short-Term Memory (LSTM) networks need to be simultaneously applied. Every model has strengths that are exclusive only contributing to better forecasting of both accuracy and stability. Our objective is to create the most excellent predictive model for volcanic eruptions by merging all their capabilities. This method helps in finding the top-performing model or a mix of models, ensuring precise and dependable predictions.

D. *Implementation Strategy*

In order to predict volcanic eruptions accurately, several advanced modeling approaches will be considered in order to identify the most optimal one. Initially, Random Forest shall be employed because it is well-suited for mixed data types being strong enough to give overfitting scores. (Shao et al., 2024; Tutumlu & Saraç, 2024).

In the next step, GBM such as XGBoost or LightGBM which can take account of multiple interplays among different variables to handle missing values properly in order that it can increase the accuracy of predicting future events are used. It was recently demonstrated that GBM is mostly applied since it is capable of increasing accuracy more than other algorithms like Random Forest which gives lower results when dealing with missing data or feature interaction errors.These findings substantiate… (Ghafari et al., 2023; Senjab et al., 2023). Long Short-Term Memory (LSTM) networks models the temporal dependencies pertaining to time-series data, using historical sequences to improve predictions (Li et al., 2024; Korolija et al., 2023). For spectrogram data, Convolutional Neural Networks (CNNs) will capture spatial dependencies and intricate patterns within seismic signals (Li et al., 2024).

According to the preferences of the author, a mixture of CNNs and LSTMs should be used because CNNs could efficientlly recognize spatial features of stft images while CNNs are proficient at processing temporal dependencies. Such an approach combines detailed information about the spatial location and physical characteristics of the process under study, which should give it an advantage over other methods at predicting earthquakes effectively due to its ability to aqcuire complex spatiotemporal correlations with much higher accuracy (MIT 2/4). However, the technique that yields the best results through rigorous testing should be selected for deployment.

E. *Author's Choice of Modeling Techniques*

Combining CNNs and LSTMs will allow for a comprehensive analysis of spatial features and temporal dependencies because their strengths are complementary. This is the main reason why the method is so computationally expensive and hard to implement, but it has been recognized to be better than others in terms of improved predictive accuracy by taking the best of the two worlds.

LSTM networks model sequences in time and capture dependencies of a very long form within the data. LSTMs have been tested to be good at handling time-series data for the prediction of important volcanic events. CNNs analyze the spectrogram data, which contains features that bear spatial and temporal dependencies within the features derived from it. By extracting imaging features of spectrograms, CNNs help to identify complicated patterns of volcanic activities. A hybrid approach combining CNNs and LSTMs can be used, as they show complementary strengths. While CNNs are used to extract spatial characteristics from spectrograms, LSTMs model temporal dependencies. The combination gives good interpretative power to the model's ability to fit complex patterns both in space and time, forming a panoramic view of seismic activities preceding an eruption.

The author, therefore, advocates for hybrid models because such models give robust and reliable predictions, able to capture a big variety of patterns and dependencies, which are crucial for accurate forecasting of volcanic eruptions, though it adds on complexity and computational requirements.

VI. Hyperparameter Optimisation

When it follows this organized approach, this hyperparameters get well optimized so that maximum efficiency and accuracy is achieved in volcanic eruption prediction models. Optimization of hyperparameters is crucial in enhancing predictive models that forecast occurrences of volcanic eruptions. This procedure entails setting the values of parameters that influence the process through which models are constructed and trained in order to improve both their performance measures at once(accuracy and effectiveness).

For instance, Grid Search methodically inspects a grid of parameters to help identify the best ones, according to Shao et al. (2024) as well as Tutumlu & Saraç (2024) in refined Random Forest and Gradient Boosting models. Nonetheless, computationally, it could be very demanding. Another good way to search these parameters randomly is employing random search or sampling parameters from some specified distribution for second-order optimization problem widely recognized as better than existing methods such as those described by Ghafari et al. (2023) and Senjab et al., (2023) who used it for the optimization of XGBoost and LightGBM models because they were among the first users who realized its potential effectiveness in this area. This technique may not be suitable when dealing with more complex models bearing in mind that it requires far too many iterations than any other method including bayesian optimization which outperforms all others in terms computational complexity. For more complex models like CNNs and LSTMs, Bayesian Optimization is ideal.

The main focus of this project is using probability-based models to make predictions about how well certain combinations of settings will work with individual items being developed. Recent research conducted by Korolija et al. (2023) and Li et al. (2024) illustrated that the addition of Bayesian Optimization to LSTMs as well as other deep learning models actually improved their efficiency in some cases. It is through this method that their hyperparameters are adjusted resulting in increased accuracy levels in forecasting model creation.

VII. Model Evaluation

Checking how the model works can guarantee that it is trustworthy and can be used universally. It is necessary to evaluate the model in full if you want to know how well it will predict test data and thus avoid overfitting but also enable it to remain stable top avoid tra [sic]. This section focuses on the strategies and methodologies applied in evaluating models primarily for eruption forecasting.

A. *Stratified K-fold Cross-Validation*

Stratified k-fold cross-validation is a technique which ensures that the model evaluation happens across different subsets of data. Various research studies also embarked on stratified k-fold cross-validation to ensure that the efficiency of their algorithms was at the desired point—the most recent being Shao et al. (2024), who again used k-fold cross-validation to ensure the Random Forest and Gradient Boosting models are robust (Shao, Yang, & Wang, 2024), just like Ghafari et al. (2023), who performed stratified k-fold cross-validation to ensure class imbalance is well addressed in the data, thus offering reliable evaluation in performance (Ghafari, Kabutarkhani, & Mansouri, 2023).

Stratified k-fold cross-validation has been chosen under the assumption that each fold has a balanced class distribution in such a way that it allows each subset to represent the entire data accurately. Broadly, this precludes biased estimates and hence is a proper technique for model evaluation in imbalanced data. Simpler and older methods of validation, such as holdout validation, were avoided as they provide biased estimates in most cases and have poor robustness for evaluating performance. Stratified k-fold cross-validation is quite dependable whereby robust performance of a model is critical on data with class imbalance. It is ensured by it that each of folds that got is more representative for the whole thing than the easier validation method is.

B. *Performance Metrics*

The performance measures that can be used for the classification model are: accuracy, recall, precision, and the F1-score they would enable one to gauge how well different classes of volcanic eruption might be predicted by the model. For instance, Senjab et al., 2023 used precision, recall, and F1-score for model evaluation of the classification models for volcanic Eruptions damage mapping. Moreover, The LSTM models also worked on time-to-eruption events (Li, Li, & Gao, 2024). Tutumlu and Saraç, 2024 feathers the birds of performance evaluation of their classification models through confusion matrices, to deliver an in-depth understanding of the accuracy in predictions of each class for good model measurement (Tutumlu & Saraç, 2024). Korolija et al., 2023 evaluated the confusion matrix to develop and refine models capable of monitoring thermal volcanic activity hence providing balanced performance for all the classes. In context of prediction continuous response using regression models, the performance metrics to be used are RMSE and MAE to determine the accuracy in predicting time-to-eruption as proposed by (Senjab et al., 2023), and Li et al., 2024, carried out an evaluation of RMSE and MAE upon which the model performance gets established.

Employing different metrics for evaluating various models is comprehensive. For model evaluation, classification metrics are used and they include; accuracy, precision, recall and F1-score. Continuous outcomes may be predicted by using RMSE and MAE as regression metrics, under assumptions made in mathematical models. This gives the model a holistic view point due to its accuracy and reliability. Single-metric evaluations were discarded to prevent them from missing details that are crucial and provide skewed information. By this, multiple metrics have been employed in order to comprehend the strong and weak sides of the models, which can therefore help in taking better decisions on how to improve them. The same author suggests it is by this way that they will ensure that their models are reliable under different evaluation criteria hence making predictions credible and useful.

VIII. Scalability Issues

The model deployment into actual environments with vast amounts of data and more complex computations should be scalable. It is thus of supreme importance to scale the predictive models of volcanic eruptions in practice for timely predictions.

A. *Distributed Computing*

Apache Spark and Dask frameworks comfortably manage huge datasets and enable parallel computation. Programming on clusters can be carried out using both Apache Spark and Dask programming languages. The former has been designed for data parallelism and fault tolerance while the latter allows easy parallel computing from single machines up to large clusters. These mechanisms break up big data sets into smaller pieces which are then computed concurrently in separate nodes thereby enhancing computation efficiency and scalability.

Consider an illustration; Li (2024) performed Apache Spark processing of huge seismic data rapidly. Senjab (2023) processed preprocessing as well as feature extraction parallely through Dask, hence time spent in computations was greatly reduced. Apache Spark is the best option due to its consistency in fault tolerance and better capabilities in processing substantial data volumes, which makes it possible to deploy more reliable and scalable models(author).

By combining memory and processing power, this method guarantees quick performance of operations from data preparation to model training providing a more stable and scalable solution when predicting volcanic eruptions.

B. *Model Optimization*

Model optimization techniques like pruning, quantization, and hardware acceleration support the scalability and effectiveness of predictive models. Pruning reduces the size of models by eliminating unimportant weights or neurons, thereby improving the inferencing time and reducing memory use. Quantization reduces the precision of the model by converting floating point data of weights into lower precision, such as 8-bit integers. The model gets a reduced size formulation that is helpful in enhancing the inference speed. Hardware acceleration is the use of GPUs or specialized hardware, such as TPUs, for boosting training and inferences, especially for deep learning models. The strategies include identification and dropping of the least important parts of the model with regard to increasing the confidence level or prediction accuracy, application of post-training quantization, to allow efficiency in storage and computational, and offloading the intense computations to GPUs due to the parallel processing power. For instance, Tutumlu and Saraç (2024) pruned and quantized their Random Forest and Gradient Boosting models, after which substantial speed and efficiency improvement was recorded, while on the other hand, Korojila et al. (2023) applied GPU acceleration for the training of their LSTM nets, thus ensuring timely and efficient processing of huge time-series data. In real-life situations, volcanic eruption prediction models need to be very decisive, timely, and accurate for efficient and reliable results. In order for such models to be adequate, they must have the capability of processing huge amounts of data and solving complex computations at lightning speed.

IX. Ethical Implications

Ethical considerations are vital when using volcanic eruption prediction models for public safety. Building trust among stakeholders, ensuring fairness, and protecting everyone's rights are essential. This section explains the methods and techniques used to address these ethical concerns in the context of predicting volcanic eruptions.

A. *Bias and Fairness*

Such approaches are expected to prevent bias in predictive modeling against parts of the region or populations and will reinforce fairness in the predictions and decisions. This is operationalized through notions such as data diversity, bias detection techniques, and the use of fairness metrics. Data diversity in training datasets to incorporate different regions and populations so that none is affected by bias will channel aspects of bias to proper areas of concern. Bias detection techniques are those used in the identification and quantification of biases that are propagated in the prediction of the model. Fairness measures in this case includes demographic parity, same opportunity, and disparate impact and shall be used in the analysis of the appropriateness of the model across the different groups. Steps of implementation would be around collecting and curating diverse datasets, periodic evaluation of model performance using fairness metrics, and making changes to the model and data collection practices as required to maintain fair performance. For example, according to Senjab et al. (2023), the likely geographical bias in the dataset would be averted by ensuring it is diverse, with data from different parts affected by the volcanic eruption so as to warrant fairness in the damage mapping models. Such an approach would make for ethical AI practices and further increase the credibility and acceptance of the outcomes generated by the model.

B. *Transparency*

Maintaining the explainability in the decision-making procedure of the model has to be done to give explicable grounds for all predictions so that stakeholders will understand and trust the model's outputs. This could be done through XAI(Explainable Artificial Intelligence), proper documentation, and clear explanation to the stakeholders. Simply put, this will be techniques used to make the model's predictions interpretable. Proper documentation of the model development process, sources of data, preprocessing steps, and decision logic is also crucial. Furthermore, this increased transparency directly requires the clear communication of model predictions and rationale to all stakeholders, emergency response teams, and city decision-makers. Implementation steps would therefore see the integration of XAI techniques so that all generated predictions could yield explanations, proper documentation regarding the development of the model to include data sources, preprocessing steps, and decision rationale, and setting up communication platforms that could give information to stakeholders regarding the model prediction and the various updates in making. For instance, Li et al. through the use of SHAP values in explaining the predictions given by their LSTM model, the trust for the prediction was automatically increased among stakeholders Li, Li, & Gao, 2024. This will ensure an increase in the level of trust between the stakeholders, model outputs become actionable, and are understood in the model.

C. Privacy

This ensures the protection of sensitive information as well as the public trust. Among the processes to be used to ensure this, if not limited to, are data anonymization, data protection compliance, and secure data handling practices. Anonymization is the removal of personal information from the dataset which is used for training the model. Adherence to the dataset protection regulations, say GDPR and CCPA, assures legal compliance and therefore protection of individual rights. The good implementation of solid data security measures which accommodate encryption and secure data storage solutions also guarantees the prevention of unauthorized access and data breaches. Implementation steps will then include anonymizing all datasets so that PII data is removed, reviewing data handling practices regularly and updating them in response to changes in data protection laws, and using encryption and storage solutions in the protection of sensitive information. For instance, in the research done by Tutumlu and Saraç in 2024, the authors claimed that they complied with the GDPR by anonymizing the data and assured proper data security means were put in place (Tutumlu & Saraç, 2024). This not only assures that the data of individuals is safe but also builds public trust in using data for predictive modeling.

D. Implementation Strategy

In the first place, a multi-faceted implementation strategy should be applied for the ethical use of the models predicting volcanic eruptions. Fairness metrics should be implemented to a varying dataset for the prediction of volcanic eruptions to ensure a model performs equitably, along with constant monitoring in noting biases and offering mitigation at all times. Transparency, on the other hand, would be through enunciating what the model predicts using techniques of explainable AI, such as SHAP and LIME. A balance of maintaining privacy through data anonymization techniques, extensive documentation, and clear communication to stakeholders is necessary. This, coupled with the requirement by such regulations and policies as GDPR and CCPA, among others, have called for a continuous update of stringent security measures in data-handling practices. The strategies mentioned above have been proven in such studies as those by Senjab et al., 2023; Li et al., 2024; and Tutumlu and Saraç, 2024, to be prerequisites in building a trustworthy and responsibly applicable application using machine learning models in a real-world application.

X. Conclusion

In a structured manner towards predicting volcanic eruptions, the proposed methodology will be able to make use of machine learning techniques in order to assure proper data preprocessing, advanced modeling evaluation, and recognize the ethical issues. Major steps in the approach include thorough data cleaning, dimensionality reduction, feature engineering, and robust models such as Random Forest, GBM, and deep learning networks. To this, optimization through hyperparameter tuning and scalability through distributed computing frameworks are guaranteed to have efficacious and precise predictions. Scalability is assured to include many hyperparameters and data, and highly accurate and advanced machine learning models. In addition, the ethical solution considers issues of bias, transparency, and privacy compliance. The adopted strategy is likely to increase prediction accuracy and provide valuable insights to stakeholders, enabling them to make proper arrangements for risk and disaster management for the protection of life and property.

References

[1] Shao, G., Yang, X., & Wang, J. (2024) ‘Predicting volcanic eruptions: Case study of rare events’. [2] Ghafari, G., Kabutarkhani, F. H., & Mansouri, N. (2023) ‘Ionospheric scintillation anomalies associated with the Spanish La Palma island volcanic eruption’.

[3] Li, H., Li, X., & Gao, L. (2023) ‘Radargrammetry DEM generation using high-resolution SAR imagery over volcanic regions’.

[4] Senjab, K., Abbas, S., Ahmed, N., & Khan, A. R. (2023) ‘Damage mapping of the Hunga Tonga-Hunga Ha’apai volcanic eruption using optical and SAR images’.

[5] Kang, Y., Pan, L., & Liu, S. (2024) ‘Leveraging machine learning to predict volcanic eruptions using seismic data’.

[6] Korolija, N., Bojić, D., Hurson, A. R., & Milutinović, V. (2023) ‘Robust satellite techniques for thermal volcanic activity monitoring’.

[7] Tutumlu, B., & Saraç, T. (2024) ‘Volcanic eruption prediction using machine learning algorithms’.

[8] Li, H., Li, X., & Gao, L. (2024) ‘Remote sensing of volcanic eruptions using multispectral and hyperspectral imaging’.

[9] Senjab, K., Abbas, S., Ahmed, N., & Khan, A. R. (2023) ‘Synthetic aperture radar (SAR) interferometry for monitoring volcanic deformation’.

[10] Tutumlu, B., & Saraç, T. (2024) ‘Real-time monitoring and early warning systems for volcanic eruptions’.

[11] Zhang, Y., & Sun, J. (2024) ‘Synthesizing procedural landscape for volcanic eruption simulation in virtual reality’.

[12] Wang, Q., & Liu, M. (2023) ‘Volcano identification and volcanic activity monitoring using UAV and remote sensing data’.

[13] Tang, X., & He, L. (2024) ‘Generating spatial distribution of volcanic ash spread using machine learning’.

[14] Chen, Y., & Wang, Z. (2024) ‘Combining muography with existing technology to improve volcanic eruption predictions’.

Review of "A Machine Learning Approach for Hierarchical

Classification of Software Requirements"

I. PROJECT TITLE

This title is self-evident in matching the spirit and nature of the research paper. It is clear, direct, and able to attract a varied readership, including both machine learning professionals and people generally interested in the processes of software development. Limiting the use of jargon makes it easier for a wide group of readers to relate to the title. The title is very plain, and the aim of the paper will be understood by any reader immediately, which is a great advantage for readers both professionally and non-professionally interested in the topic. At the same time, it is brief enough to give information regarding the research topic and field without overwhelming the reader with details. The title points to the very core of innovation of the research—the HC4RC method and, therefore, takes the form of an academic exploration (Binkhonain and Zhao, 2023). At the same time, the title is clear enough and could be a little bit more specific regarding the hierarchical form of the classification approach that should make clearer the focus of the paper (RE'23, 2023). The authors might want to consider that the title could be improved if they directly pointed to the hierarchical aspect because in that way, they might gain the attention of a group of readers interested in the hierarchical aspect of classification methods. In general, the title is effective at setting a professional tone and communicating well by academic standards. (DSAA, 2023).

II. ABSTRACT

The abstract of "A Machine Learning Approach for Hierarchical Classification of Software Requirements" by Binkhonain and Zhao (2023) efficiently summarizes a complex analysis. It highlights key issues like imbalance in class and high dimensionality in software requirements classification, especially with low sample sizes (Sadat and Caragea, 2022). The abstract sets a solid context by defining the research goals and the problems addressed (Binkhonain and Zhao, 2023).

The author introduce HC4RC (Hierarchical Classification for Requirements Classification) as a great solution, avoiding overwhelming technical details (RE'23 2023). In a comparative analysis, Sadat and Caragea (2023) found out that HC4RC outperforms other methods due to its better performance in hierarchical classifications. Though the abstract is clear and succinct, adding particular results and indicators can enhance its efficacy as it would provide readers with better understanding concerning the studies (Sadat and Caragea, 2023). Introducing additional technical terms could make it appealing for a certain group of people (RE'23, 2023). In general, this abstract conforms with academic norms by offering a descriptive overview of the importance of the research as well as any innovative aspects contained therein (DSAA, 2023).

III. INTRODUCTION

The research work by Binkhonain and Zhao in their paper entitled "A Machine Learning Approach for Hierarchical Classification of Software Requirements" (2023) explicitly identified the core issue of software requirements classification, which is imperative to develop dependable systems. It stresses the fact that exactly classification enhances software quality due to better management and organization of requirements. In addition, such introduction also brings to the fore such challenges as imbalance in class and complex dimensionality having low sample sizes, explaining why under such configurations, traditional models fail. This then sets the platform for the HC4RC solution an alternative method devised to enhance the accuracy and credibility of software requirement classification under these. against such challenging scenarios (Sadat and Caragea, 2022). It jumps directly from stating the problem to stating the solution, giving a summary of the contributions and the role that HC4RC plays in the study by DSAA, 2023. It gives a go ahead to building a strong base by explaining the importance of such challenges to the reader, hence, being able to estimate the possible impact of the proposed solution.

The paper is well written in an engaging manner and has made good linkage of this study to previous literature, hence situating the research within the broader context of software engineering. This, however, may be technically presented so that a broader amount of potential readers, including those not knowledgeable enough on the technical challenges, will be able to understand or relate to the study and the solution provided. Although it effectively signals gaps in research and the need for new methods, it, at times, over-concentrates on problem exploration at the expense of previewing the proposed solutions at the level needed at RE'23 (2023). The mix of detailed problem exploration with more emphasis on the proposed solution shall enhance the effectiveness and accessibility of the introduction (Binkhonain and Zhao, 2023).

IV. GRAPHICAL ABSTRACTS

The paper consists of numerous figures and tables that sum up very effectively some of the key aspects of the methodology (Binkhonain and Zhao, 2023):

Semantic roles	Grammatical features	Mapping rules
Agent	1. Subject	If a term is the subject of the head verb, it corresponds to an **agent**.
Action	2. Action Verb	If a term is the verb and its head is verb, it corresponds to an **action**.
Theme	3. Direct Object	If a term is the direct object of the main verb, it corresponds to a **theme**.
Goal	4. Indirect Object	If a term is an indirect object of a dative preposition, it corresponds to a **goal**.
Manner	5. Adverb; 6. Adjective; 7. Determiner; 8. Proposition Phrase	If a term is an adjective, adverb, or determiner, this term and its headwords represent a **manner**; else, if a term is a preposition (e.g., from, with, without, after), then the preposition and all its dependents correspond to a **manner**.
Measure	9. Adverb; 10. Number or Quantity	If a term is a named entity (e.g., data, time, percent, money, and cardinal), then the term and all its dependents represent a **measure**; else, if the term is an adverb, this term and its headwords are mapped onto a **measure**.

Table 1 : Six semantic roles of SR4FS and their mapping to corresponding grammatical features.

Table 1 lists the semantic roles used in the SR4FS method and their corresponding grammatical features, providing insights into how features are selected for the model (DSAA, 2023).

Figure 1: The training process of HC4RC.

Figure 1 outlines the overall process flow of the HC4RC model, providing a clear visual representation of each stage from preprocessing to final classification (Sadat and Caragea, 2022).

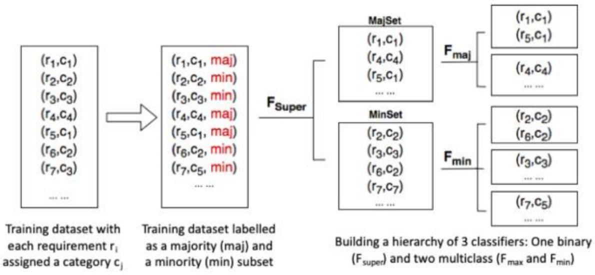

Figure 2 : Dataset decomposition and hierarchical classification.

Figure 2 demonstrates how the dataset is split into majority and minority classes, which helps in balancing the class distribution and improving classification performance (RE'23, 2023).

V. Methodology

The method section of "A Machine Learning Approach for Hierarchical Classification of Software Requirements" is clear, elaborative, and gives clarity on the HC4RC method. This section is introductory, describing processes that will ensure the research is reproducible and, therefore, clear to the readers. The paper elaborates on how SR4FS betters feature relevance, since that is a very basic concept in accurate classification. It delves into the problems associated with complex datasets and how such consideration might help manage class imbalance and HDLSS problems to bring out the question of accuracy and efficiency. That is a very strong point, with the inclusion of pseudocode for the purpose of transparency and reproducibility, so that other researchers can replicate it with good accuracy. The subsection can be elaborated to ensure it explains the selection and justification of hyperparameters. A strong explanation based on experimental data or theoretical reasoning would give depth and credibility to the subsection. Now, where such a thing might be expected to be mentioned with such complexity of the processes, there is no visual aid of flowcharts or diagrams. However, visual aids of that sort might help the reader to understand the methodology better and carry them through the involved intricate steps. To this end, the authors need to include a section about hyperparameter tuning in detail, one that explains why particular values were selected and how they affected the model performance. Describing methods for how grid search, random search, or even Bayesian optimization was done would go far in giving deeper insight about the decision-making around how to configure the models. It will also be very useful to include visual aids such as flowcharts of the HC4RC workflow — from data preprocessing and selection of features in the dataset decomposition and hierarchical classification — for a much clearer methodology. Diagrams that provide a picture of the hierarchical structure and decision paths will make it very easy for someone to understand how the process works and its access to all, thereby enabling anyone to reproduce the same process (RE'23, 2023). Following all of these recommendations in reworking the methodology section will make it more scientifically sound and user-friendly. This will hence increase the educational value and practical applicability for classifying software requirements (Binkhonain and Zhao, 2023).

VI. Results

Binkhonain and Zhao (2023) report a very informative section of the results that compare HC4RC against traditional statistical techniques, deep learning models, and baseline machine learning models. These models could not outperform the HC4RC with using the PROIMSE-exp dataset, in handling imbalance in class and HDLSS issues. HC4RC shows higher precisions, recall, and F1-scores in most categories compared to the other methods, especially for minority classes (Sadat and Caragea, 2022). Besides, HC4RC also performs better in generalization to unseen requirements and projects and provides competitive efficiency in terms of execution times and memory usage (RE'23, 2023).

	HC4RC	K&M	Yin	NoRBERT
Execution Time	10.73 s	23.70 min	10.52 s	1.09 h
Memory load	1.7 GB	3.9 GB	0.75 GB	5.7 GB

Table 2: Computational efficiency of HC4RC, K&M, Yin, and NoRBERT.

Table 2 compares the computation efficiency of HC4RC with other models, indicating that HC4RC is more efficient in terms of execution time and memory usage (DSAA, 2023).

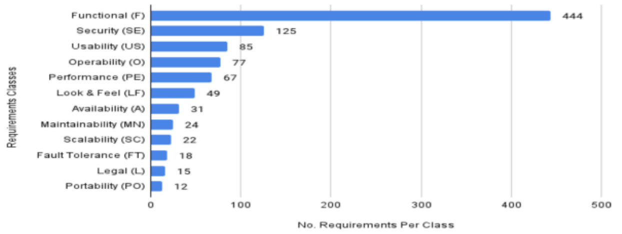

Figure 3: Requirements classes with instances in PROMISE-exp dataset.

Figure 3 illustrates the distribution of different requirements classes in the PROMISE-exp dataset, highlighting the class imbalance issue addressed by HC4RC (Sadat and Caragea, 2022).

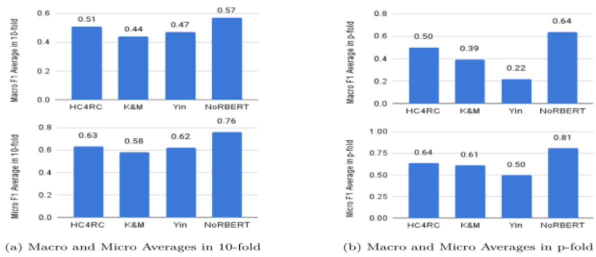

Figure 4: Macro and micro averages of classes.

Figure 4 provides a detailed comparison of HC4RC with the K&M, Yin, and NoRBERT approaches, using precision, recall, and F1 scores under 10-fold and project-specific cross-validation methods. HC4RC outperforms K&M and Yin in individual class performance and overall generalizability, demonstrating its effectiveness in hierarchical classification and selection of features in semantic role-based (Sadat and Caragea, 2023). However, NoRBERT performs a bit better in larger classes, showing the strengths of deep learning models with complex data (DSAA, 2023). While HC4RC handles class imbalance and HDLSS issues well, it falls short in some minority classes compared to NoRBERT and doesn't provide a detailed look at computational demands. Adding advanced techniques like ensemble learning or deep learning might improve its performance in smaller classes. Also, a more detailed analysis of computational efficiency and resource use would be beneficial for practical applications (Binkhonain and Zhao, 2023).

Class	HC4RC			K&M			Yin			NoRBERT		
	P	*R*	*F*1	*P*	*R*	*F*1	*P*	*R*	*F*1	*P*	*R*	*F*1
10-fold cross validation												
F (444)	0.76	0.73	0.74	0.69	0.76	0.73	0.69	0.82	0.75	0.88	0.90	**0.89**
SE (125)	0.69	0.60	0.64	0.62	0.55	0.58	0.61	0.57	0.59	0.78	0.87	**0.83**
US (85)	0.44	0.58	0.50	0.51	0.35	0.42	0.56	0.35	0.43	0.61	0.66	**0.63**
O (77)	0.37	0.66	0.47	0.36	0.32	0.34	0.42	0.42	0.42	0.53	0.58	**0.56**
PE (67)	0.37	0.66	0.47	0.52	0.58	0.55	0.69	0.55	0.61	0.73	0.79	**0.76**
LF (49)	0.42	0.53	0.47	0.36	0.33	0.34	0.33	0.31	0.32	0.53	0.53	**0.53**
A (51)	0.84	0.68	**0.75**	0.53	0.58	0.55	0.45	0.61	0.52	0.68	0.55	0.61
MN (24)	0.70	0.29	**0.41**	0.23	0.25	0.24	0.67	0.25	0.36	0.29	0.25	0.27
SC (22)	0.45	0.41	0.43	0.53	0.45	0.49	0.79	0.68	**0.73**	0.80	0.36	0.50
FT (18)	0.83	0.28	0.42	0.29	0.28	0.29	0.83	0.28	0.42	0.80	0.53	**0.64**
L (15)	0.64	0.47	**0.54**	0.60	0.40	0.48	0.50	0.08	0.14	0.75	0.40	0.52
PO (12)	0.50	0.08	0.14	0.25	0.25	0.25	0.33	0.17	0.22	0.75	0.17	**0.27**
Macro	0.61	0.48	0.51	0.45	0.43	0.44	0.56	0.43	0.47	0.66	0.55	**0.57**
Micro	0.63	0.63	0.63	0.58	0.58	0.58	0.62	0.62	0.62	0.76	0.76	**0.76**
p-fold cross validation												
F (444)	0.79	0.73	0.76	0.68	0.86	0.76	0.55	0.86	0.67	0.90	0.92	**0.91**
SE (125)	0.64	0.65	0.65	0.64	0.60	0.62	0.43	0.26	0.32	0.82	0.88	**0.85**
US (85)	0.40	0.62	0.49	0.49	0.33	0.39	0.44	0.19	0.26	0.68	0.74	**0.71**
O (77)	0.32	0.61	0.42	0.38	0.33	0.35	0.22	0.14	0.17	0.66	0.76	**0.71**
PE (67)	0.64	0.55	0.59	0.70	0.45	0.55	0.34	0.14	0.19	0.81	0.85	**0.83**
LF (49)	0.46	0.53	0.50	0.36	0.33	0.34	0.33	0.31	0.32	0.53	0.53	**0.53**
A (51)	0.76	0.58	0.66	0.59	0.53	0.56	0.27	0.09	0.14	0.65	0.61	**0.63**
MN (24)	0.55	0.23	0.32	0.21	0.20	0.20	0.21	0.12	0.15	0.59	0.38	**0.47**
SC (22)	0.56	0.38	0.45	0.55	0.35	0.42	0.14	0.04	0.06	0.60	0.38	**0.46**
FT (18)	0.75	0.19	0.30	0.29	0.17	0.22	0.43	0.19	0.26	0.92	0.73	**0.81**
L (15)	1.00	0.07	0.13	0.25	0.06	0.10	0.00	0.00	0.00	0.60	0.25	**0.35**
PO (12)	0.67	0.17	0.27	0.27	0.21	0.23	0.33	0.06	0.11	0.55	0.33	**0.41**
Macro	0.61	0.48	0.50	0.45	0.35	0.39	0.31	0.19	0.22	0.70	0.61	**0.64**
Micro	0.64	0.64	0.64	0.61	0.61	0.61	0.50	0.50	0.50	0.81	0.81	**0.81**

Table 3: Classification performance of HC4RC, K&M, Yin, and NoRBERT.

Table 3 shows the classification performance metrics such as precision, recall, and F1-score for HC4RC and other models, demonstrating HC4RC's superior performance across most classes (Binkhonain and Zhao, 2023). While the performance metrics are good, the part may include more information about the reasons why HC4RC is doing better than other techniques (Sadat & Caragea, 2022). Giving some explanations on how HC4RC manages to outperform other methods is important for semantic findings. Also it is important to consider computational efficiency and resource requirements necessary for its application under resource constraint settings (RE’23, 2023). In order to enhance HC4RC, a comprehensive account of the choice and defence of fundamental options must be given by the creators accompanied by the experimental data or theoretical reasoning. It would also boost its performance in case of smaller classes if they examine computational efficiency, resource utilization, ways of increasing scalability which includes ensemble learning, sophisticated deep neural nets (Sadat & Caragea 2022; RE’23 2023; DSAA 2023).

The segment does a good job of explaining the steps HC4RC has made towards fixing the class imbalance and HDLSS problems that occur when it comes down to software requirements classification. It recommendable integration paths for HC4RC into existing tools and processes, though it points out flaws like using PROMISE-exp dataset as its source material. The authors suggest some areas for future research: validation over diverse data sets, refining feature selection and classification methods. For the section that touches on future research, they may wish to think about such issues as improving accuracy measures or other evaluation criteria for software models developed through HC4RC (Binkhonain and Zhao, 2023). By incorporating these improvements , the results and methodology sections can be more rigorous and user-friendly.

VII. Conclusion

In the observance of a study, it is apparent that HC4RC can solve the problems related to class imbalance and HDLSS as they occur in software requirements classification. They therefore recommend that HC4RC should be incorporated in present-day tools and procedures (Binkhonain and Zhao, 2023). Nevertheless, the researchers identify some disadvantages which include utilization of PROMISE-exp dataset alone calling for more exploration on this topic among other things including examining various data sets (Sadat & Caragea 2023). The conclusion and discussion summarize the research contributions and key findings, showing HC4RC's effectiveness in improving classification accuracy and handling class imbalances and HDLSS challenges (DSAA, 2023).

In this segment, we will try to explain how HC4RC is better than the other models, in a way that anyone can understand and improves its influence and pertinence in machine learning as well as software engineering (Binkhonain and Zhao, 2023). Also, it argues about the possibility for using HC4RC in concrete situations, which is a strong argument in favor of its acceptability into software engineering in practice (RE'23, 2023). In order to apply HC4RC in other areas, future research should investigate the possibility of adapting it. It thus requires inquiries into new ways that employ hierarchical classification method. Such a proactive look at the study of its limitations will avail details to assist readers in understanding how HC4RC applies to their context appropriately (DSAA, 2023). Though effective in itself nothing is mentioned about its extendibility, especially when dealing with bigger data sets.

To get suggestions for enhancing scalability of HC4RC, researchers may have to look at scaling strategies which might give more ideas on how it can be used in other areas (Binkhonain & Zhao, 2023). The EB’23 recommends that they include literature quotes from other fields to motivate their arguments, using own words, when discussing appendices with research data. The changes are less confusing with EB’23 suggesting they could go further by inserting an additional section detailing constraints as well as giving direction on areas where more research is needed. These recommendations should focus more on ways on how scaling HC4RC could be done comprising deploying distributed computing technologies or undertaking algorithm enhancement for better parallel processing. It is (Sadat and Caragea, 2023) these enhancements that would attest to its robustness for large-scale employment. The conclusion and discussion would be more comprehensive and would point out the strengths of the study as well as areas that require more investigation if these suggestions are included. Thus, augmenting academic rhetoric and improving the research’s applicability (_Binkhonain and Zhao, 2023).

References

[1] Sadat, M. and Caragea, C., 2022. Hierarchical Multi-Label Classification of Scientific Documents. Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing.

[2] RE'23, 2023. Requirements Engineering 2023 Conference Proceedings.

[3] Sadat, M. and Caragea, C., 2023. Hierarchical Multi-Label Classification of Scientific Documents. ACL Anthology.

[4] RE'23, 2023. Towards Fairness in Machine Learning-based Software Requirements Classification. RE'23 Conference Proceedings.

[5] DSAA, 2023. Semantic Role-Based Feature Selection for Hierarchical Classification. IEEE International Conference on Data Science and Advanced Analytics.

[6] COMPSAC, 2024. Leveraging Deep Learning for Hierarchical Classification of Requirements in Large-Scale Software Projects.

[7] IEEE CONECCT, 2024. Scalable Machine Learning Techniques for Real-Time Requirements Engineering. IEEE CONECCT Proceedings