Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871

Vol. 6, No. 2, April 2025, Page. 755-774
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.2378

Comparative Analysis of Decision Tree, Random Forest, Svm, and Neural Network Models for Predicting Earthquake Magnitude

Turino[*1], Rujianto Eko Saputro[2], Giat Karyono[3]

[1,2]Magister of Computer Science, Universitas Amikom Purwokerto, Indonesia
[3]Informatics, Universitas Amikom Purwokerto, Indonesia

Email: [1]23MA41D002@students.amikompurwokerto.ac.id

Received : Jun 25, 2024; Revised : Sep 25, 2024; Accepted : Oct 14, 2024; Published : Apr 26, 2025

Abstract

This study conducts a comparative analysis of four machine learning algorithms—Decision Tree, Random Forest, Support Vector Machine (SVM), and Neural Network—to predict earthquake magnitudes using the United States Geological Survey (USGS) earthquake dataset. The analysis evaluates each model's performance based on key metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). The Random Forest model demonstrated superior performance, achieving the lowest MAE (0.217051), lowest RMSE (0.322398), and highest R^2 (0.574261), indicating its robustness in capturing complex, non-linear relationships in seismic data. SVM also showed strong performance, with competitive accuracy and robustness. Decision Tree and Neural Network models, while useful, had comparatively higher error rates and lower R^2 values. The study highlights the potential of ensemble learning and kernel methods in enhancing earthquake magnitude prediction accuracy. Practical implications of the findings include the integration of these models into early warning systems, urban planning, and the insurance industry for better risk assessment and management. Despite the promising results, the study acknowledges limitations such as reliance on historical data and the computational intensity of certain models. Future research is suggested to explore additional data sources, advanced machine learning techniques, and more efficient algorithms to further improve predictive capabilities. By providing a comprehensive evaluation of these models, this research contributes valuable insights into the effectiveness of various machine learning techniques for earthquake prediction, guiding future efforts to develop more accurate and reliable predictive models.

Keywords : *Earthquake Prediction, Machine Learning, Predictive Modeling, Random Forest, Seismic Data, SVM.*

1. INTRODUCTION

1.1. Background Information

Earthquakes are natural phenomena that have the potential to cause widespread devastation, including loss of life, economic damage, and environmental degradation. The importance of predicting earthquakes cannot be overstated, as accurate predictions can lead to timely warnings and the implementation of necessary measures to mitigate their impact. Earthquake prediction encompasses various aspects, including forecasting the occurrence, location, and magnitude of future seismic events. Among these, predicting the magnitude of an earthquake is particularly crucial as it directly correlates with the potential damage and necessary preparedness levels. Understanding the patterns and behaviors of earthquakes through data analysis is a fundamental step in developing reliable predictive models, ultimately aiding in disaster risk reduction and preparedness planning.

The United States Geological Survey (USGS) earthquake data provides a comprehensive repository of seismic events, offering valuable insights into the characteristics and patterns of earthquakes. This dataset includes a wide range of attributes such as the time, location (latitude and

Vol. 6, No. 2, April 2025, Page. 755-774
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.2378

longitude), depth, and magnitude of the earthquakes, along with additional metadata like the number of seismic stations used in the calculation, the type of seismic event, and various error estimates. The richness and granularity of this dataset make it an indispensable resource for researchers and scientists aiming to study seismic activities and develop predictive models. By leveraging this data, it is possible to analyze historical earthquake occurrences, identify trends, and build models that can forecast future seismic events with higher accuracy.

1.2. Importance of Earthquake Prediction

Earthquake prediction is a critical aspect of seismology due to its significant social and economic implications. Researchers emphasize the importance of developing effective earthquake prediction programs to minimize the loss of life and property resulting from seismic events [1], [2]. Recurrence models that identify cycles of repeating earthquakes have shown predictive power, contributing to determining earthquake risk [3]. While various methodologies, such as thermal infrared anomaly analysis and machine learning models, have been explored for earthquake prediction, the effectiveness and precision of these approaches remain areas of ongoing research [4], [5]. Additionally, the study of seismic temporal distribution and subsurface structures plays a vital role in earthquake prediction and damage mitigation [6], [7].

Efforts in earthquake prediction have also extended to exploring factors like gravity changes, induced stress fields, and ionospheric total electron content to enhance prediction accuracy and determine seismic risk areas [8], [9], [10]. Machine learning techniques have been applied to predict earthquakes with high accuracy, aiding in forecasting the time-to-mainshock and improving early warning systems [11], [12]. Furthermore, understanding the relationship between seismic activity patterns and climatic data through deep learning models has shown promise in earthquake magnitude prediction [13].

The prediction of earthquakes is not only about forecasting the event itself but also involves estimating potential damage, casualties, and injuries to enable proactive measures and disaster reduction [14], [15]. Moreover, the intention to prepare for earthquakes among residents can be influenced by attitudes, subjective norms, and supportive behaviors, highlighting the importance of community engagement in earthquake preparedness [16]. Addressing misinformation and debunking popular beliefs related to earthquake prediction is essential to ensure accurate information dissemination in earthquake seismology [17].

1.3. Data Mining Techniques in Earthquake Prediction

The Decision Tree is a popular machine learning algorithm used for both classification and regression tasks. It operates by recursively partitioning the data space into subsets based on the value of input features, creating a tree-like model of decisions. Each internal node in the tree represents a decision based on the value of a single attribute, while each leaf node represents a final prediction or outcome. One of the key advantages of Decision Trees is their interpretability. The hierarchical structure of the tree provides a clear and intuitive visualization of the decision-making process, allowing researchers and practitioners to understand how predictions are made.

Decision Trees have been widely applied in the field of earthquake prediction, leveraging their ability to handle complex, multi-dimensional datasets. Various studies have demonstrated their effectiveness in predicting different earthquake-related parameters, such as the occurrence, magnitude, and impact of seismic events. Notable application of Decision Trees in earthquake prediction is the work by [5]. Their model was trained on a comprehensive dataset from the United States Geological Survey (USGS), incorporating various attributes such as the depth, location, and previous seismic activity. The

Vol. 6, No. 2, April 2025, Page. 755-774
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.2378

study found that the Decision Tree model outperformed several other machine learning algorithms in terms of prediction accuracy and computational efficiency.

Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the average prediction (for regression) or the majority vote (for classification) of the individual trees. This technique was introduced by Leo Breiman and Adele Cutler in the early 2000s as a way to improve the accuracy and robustness of decision tree models. One of the primary advantages of Random Forest is its ability to handle high-dimensional data with a large number of features and instances. The algorithm is also known for its robustness to overfitting, especially when compared to single decision tree models. This is achieved through techniques such as bootstrap aggregating (bagging) and feature randomness. Bagging involves training each tree on a different random subset of the training data, while feature randomness involves selecting a random subset of features for each split in the trees. Together, these techniques ensure that the Random Forest model is both powerful and generalizable, making it suitable for a wide range of predictive tasks, including earthquake prediction.

Random Forest has been extensively applied in the field of earthquake prediction due to its versatility and high predictive accuracy. Several studies have demonstrated the effectiveness of Random Forest models in analyzing seismic data and predicting various earthquake-related parameters. For instance, research [18] demonstrated the use of Random Forest in earthquake forecasting, emphasizing its capability in predicting an earthquake's latitude, longitude, magnitude, and depth.

Support Vector Machine (SVM) is a supervised learning algorithm that is widely used for classification and regression tasks. Introduced by Vladimir Vapnik and his colleagues in the 1990s, SVM aims to find the optimal hyperplane that separates data points of different classes with the maximum margin. The margin is defined as the distance between the hyperplane and the nearest data points from each class, known as support vectors.

One of the key strengths of SVM is its effectiveness in high-dimensional spaces, making it suitable for applications involving complex datasets with many features. Additionally, SVM is particularly well-suited for problems where the data points are not linearly separable in the original feature space. By using kernel functions, SVM can map the input features into a higher-dimensional space where a linear separation is possible. This capability allows SVM to capture intricate patterns and relationships within the data, making it a powerful tool for predictive modeling.

Support Vector Machine has been extensively applied in earthquake prediction research due to its ability to handle complex, non-linear relationships in seismic data. Various studies have demonstrated the effectiveness of SVM in predicting earthquake occurrences, magnitudes, and other related parameters. In a study focusing on forecasting earthquake magnitude categories in the Flores Sea, various approaches including SVM were utilized, highlighting the relevance of SVM in earthquake prediction research [19].

Neural Networks, inspired by the human brain's structure and function, are a class of machine learning algorithms designed to recognize patterns and relationships within data. They consist of layers of interconnected nodes, or neurons, where each connection has an associated weight. These layers include an input layer, one or more hidden layers, and an output layer. The neurons in each layer perform weighted sums of their inputs, apply an activation function, and pass the result to the next layer.

One of the primary advantages of Neural Networks is their ability to model complex, non-linear relationships in data. They are particularly well-suited for tasks where traditional linear models fail to capture the underlying patterns. Neural Networks can handle large amounts of data and are capable of learning from it without explicit feature engineering, making them powerful tools for a variety of applications, including image and speech recognition, natural language processing, and, importantly, earthquake prediction.

Vol. 6, No. 2, April 2025, Page. 755-774
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.2378

Neural Networks have been extensively utilized in earthquake prediction research, leveraging their ability to learn complex patterns from large datasets. Various studies have demonstrated the effectiveness of Neural Networks in predicting earthquake occurrences, magnitudes, and other related parameters. For instance, research by [20] developed a deep neural network-based seismic response prediction framework that merges structural information with earthquake ground motion data to predict structural responses accurately.

1.4. Problem Statement

Accurately predicting the magnitude of earthquakes remains a significant challenge within the field of seismology and data science. The inherent complexity of seismic activities, influenced by numerous geological and environmental factors, makes it difficult to develop models that can reliably predict earthquake magnitudes. Earthquake magnitude prediction involves understanding and analyzing vast amounts of historical seismic data, which includes various parameters such as the location, depth, time, and characteristics of previous earthquakes. Despite advancements in data collection and processing technologies, the unpredictable nature of earthquakes means that even the most sophisticated models can struggle to achieve high levels of accuracy. This unpredictability is further compounded by the lack of consistent patterns in seismic data, leading to difficulties in identifying the key predictors of earthquake magnitude.

One of the primary challenges in earthquake magnitude prediction is the variability in the geological conditions of different regions. Factors such as tectonic plate boundaries, fault lines, and soil composition can vary widely, affecting the behavior of earthquakes in unpredictable ways. Additionally, the quality and completeness of the available data can also impact the accuracy of predictive models. In many cases, data from seismic events may be incomplete or contain errors, further complicating the modeling process. This necessitates the use of robust data preprocessing and handling techniques to ensure that the models are trained on reliable and relevant data. Moreover, the dynamic and evolving nature of the Earth's crust requires models to be adaptive and capable of incorporating new data as it becomes available.

Given the challenges associated with earthquake magnitude prediction, there is a clear need for a comparative analysis of different predictive algorithms to identify the most effective approaches. While numerous algorithms have been developed and applied in the field of earthquake prediction, their relative performance in terms of accuracy, reliability, and computational efficiency varies. Traditional statistical methods, machine learning techniques, and advanced neural networks each offer distinct advantages and limitations. For instance, decision trees and random forests are known for their interpretability and ability to handle non-linear relationships, while support vector machines (SVM) and neural networks can model complex patterns but may require more computational resources and expertise to implement effectively.

A comprehensive comparative analysis can provide valuable insights into the strengths and weaknesses of various predictive algorithms, guiding the selection of the most appropriate methods for earthquake magnitude prediction. By systematically evaluating and comparing different models on a standardized dataset, researchers can determine which algorithms offer the best balance of accuracy and efficiency for this specific application. This analysis can also highlight areas where existing models fall short and identify opportunities for further refinement and innovation. Ultimately, such comparative studies are crucial for advancing the state of the art in earthquake prediction and enhancing our ability to mitigate the impacts of these natural disasters.

1.5. Research Objective and Significance

The field of earthquake prediction has garnered significant attention due to the catastrophic impact earthquakes can have on human life and infrastructure. Accurate prediction models are essential for mitigating these impacts, enabling timely disaster preparedness and response. Various machine learning algorithms have been employed in recent years to enhance the accuracy of earthquake magnitude predictions. Among these, Decision Tree, Random Forest, Support Vector Machine (SVM), and Neural Network models have shown promise due to their ability to handle complex patterns in seismic data. However, the performance of these models can vary widely based on the characteristics of the dataset and the specific implementation of each algorithm.

The primary objective of this research is to conduct a comprehensive comparative analysis of Decision Tree, Random Forest, SVM, and Neural Network models in predicting earthquake magnitude. This involves systematically evaluating each model's performance in terms of accuracy, computational efficiency, and robustness. By leveraging a standardized dataset from the United States Geological Survey (USGS), which includes a wide range of seismic attributes such as location, depth, magnitude, and various error estimates, we aim to provide a clear understanding of the strengths and limitations of each model. This analysis is critical for identifying the most effective predictive techniques and guiding future research and practical applications in earthquake prediction.

Through this comparative study, we seek to answer several key questions: Which model offers the highest accuracy in predicting earthquake magnitudes? How do the computational requirements and processing times of these models compare? What are the specific advantages and drawbacks of each algorithm in the context of seismic data? By addressing these questions, our research aims to contribute valuable insights to the field of earthquake prediction, supporting the development of more reliable and efficient predictive models. Ultimately, the findings of this study will aid in the advancement of data-driven approaches to disaster risk reduction, enhancing our ability to anticipate and respond to seismic events.

The significance of this study lies in its potential to contribute meaningfully to the field of earthquake prediction. Earthquakes, by their nature, are complex and unpredictable, making it difficult for researchers and practitioners to develop accurate predictive models. This study addresses this challenge by providing a comparative analysis of four widely-used machine learning algorithms: Decision Tree, Random Forest, Support Vector Machine (SVM), and Neural Network models. By evaluating these models' performance in predicting earthquake magnitudes using a comprehensive dataset from the United States Geological Survey (USGS), this research offers critical insights into which algorithms are most effective for this application. The findings will advance the scientific understanding of seismic data analysis and help refine existing predictive techniques, paving the way for more accurate and reliable earthquake predictions.

Moreover, this study's contribution extends to methodological advancements in the application of machine learning to geophysical phenomena. The rigorous comparative framework used in this research can serve as a blueprint for future studies aiming to evaluate different predictive models across various domains. By demonstrating the strengths and limitations of each algorithm in the context of earthquake magnitude prediction, the study provides a detailed reference for researchers and practitioners in seismology, data science, and related fields. This contribution is particularly valuable as it not only enhances the current state of knowledge but also stimulates further research into innovative approaches for tackling the inherent challenges in earthquake prediction.

The potential impact of this study on disaster preparedness and response is substantial. Accurate earthquake magnitude predictions can significantly enhance the effectiveness of early warning systems, allowing communities and authorities to implement timely and appropriate measures to mitigate the impacts of seismic events. By identifying the most reliable predictive models, this research helps improve the precision of these warnings, thereby reducing the risk of casualties and economic losses. In

Vol. 6, No. 2, April 2025, Page. 755-774
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.2378

regions prone to earthquakes, such advancements are crucial for ensuring public safety and resilience. Enhanced predictive capabilities can inform better urban planning, infrastructure development, and emergency response strategies, making communities more robust in the face of natural disasters.

Furthermore, the findings of this study have practical implications for policymakers, emergency management agencies, and other stakeholders involved in disaster risk reduction. By providing a clearer understanding of which machine learning models offer the highest accuracy and efficiency in earthquake magnitude prediction, this research supports informed decision-making and resource allocation. This, in turn, enables more strategic investments in technologies and initiatives that bolster disaster preparedness and response. Ultimately, the study underscores the importance of leveraging advanced data analytics to enhance our ability to anticipate and respond to seismic events, contributing to safer and more resilient societies.

2. METHOD

The overall research method for this study follows a structured machine learning workflow, commonly used in predictive modeling studies. It includes key steps such as data collection, exploratory data analysis (EDA), data preprocessing, model implementation, and evaluation metrics calculation (Figure 1). This methodology is in line with established practices in data science research, where a systematic approach is employed to ensure the robustness and accuracy of the predictive models [21]. This structured process ensures that all stages of the research contribute effectively to the model's predictive performance.

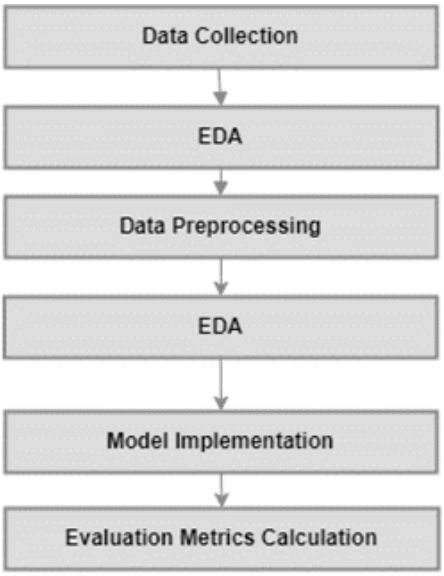

Figure 1. Research Method Flowchart

2.1. Data Collection

The dataset utilized in this study is sourced from the United States Geological Survey (USGS), a cornerstone in seismic research that provides a comprehensive repository of earthquake data. The USGS dataset encompasses a wide range of seismic events recorded globally, making it highly regarded for its thoroughness and accuracy. This extensive catalog has been widely employed in various studies to analyze earthquake patterns, fault interactions, and seismic hazards. For instance, [22] utilized the USGS records, combined with local data, to enhance the analysis of spatio-temporal variations in East Java. Similarly, [23] examined interactions between the Hayward and Calaveras Faults based on historical USGS seismic data. This diversity of applications demonstrates the dataset’s

Vol. 6, No. 2, April 2025, Page. 755-774
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.2378

critical importance, making it an ideal choice for this research focused on predictive modeling of earthquake magnitudes and occurrences. Its inclusion of key features—such as time, location, depth, and magnitude—allows for a detailed analysis of the factors influencing seismic events.

The dataset contains 102,452 entries and 23 columns. Each entry represents a recorded seismic event, and the columns provide detailed information about each event. This dataset contains multiple columns, each representing a specific attribute of the recorded earthquakes. These attributes are crucial for developing and evaluating predictive models. The columns in the dataset are as follows:Data Preprocessing

a. time: This column records the exact time of the earthquake, represented as the number of milliseconds since the Unix epoch (January 1, 1970, 00:00:00 UTC). This precise timestamp is essential for time-series analysis and understanding temporal patterns in seismic activity.
b. latitude: The latitude of the earthquake's epicenter, expressed in decimal degrees. This geographical coordinate helps in mapping and analyzing the spatial distribution of earthquakes.
c. longitude: The longitude of the earthquake's epicenter, also in decimal degrees. Along with latitude, this information is crucial for identifying the locations most affected by seismic events.
d. depth: The depth at which the earthquake occurred, reported in kilometers. Depth is a significant factor in assessing the potential impact of an earthquake, as shallow earthquakes generally cause more damage.
e. mag: The magnitude of the earthquake, measured on various scales such as Richter, moment magnitude, etc. This is the target variable for our predictive models, representing the intensity of the seismic event.
f. magType: This column specifies the type of magnitude measurement used, such as "mb" (body wave magnitude), "ml" (local magnitude), or "mw" (moment magnitude). Understanding the type of magnitude is important for standardizing the data and ensuring consistent comparisons.
g. nst: The total number of seismic stations that recorded the earthquake. This count can influence the accuracy of the recorded data and is an indicator of the data's reliability.
h. gap: The largest azimuthal gap between adjacent seismic stations, measured in degrees. A smaller gap generally indicates better coverage and potentially more accurate location estimates.
i. dmin: The distance to the nearest seismic station, reported in degrees. This distance can affect the precision of the recorded data, with closer stations typically providing more accurate measurements.
j. rms: The root-mean-square of the residuals of the earthquake's hypocenter location. This value provides an estimate of the fit quality of the seismic model used to determine the earthquake's location.
k. net: The ID of the seismic network that located the earthquake. Different networks may use varying methodologies, which can affect the data's consistency and comparability.
l. id: A unique identifier assigned to each earthquake event, ensuring that each record can be distinctly referenced.
m. updated: The time when the earthquake event was most recently updated in the catalog, also represented as the number of milliseconds since the Unix epoch. This column helps track the currency of the data.
n. place: A human-readable description of the earthquake's location. This qualitative information complements the quantitative latitude and longitude data, providing context for the earthquake's impact area.
o. type: The type of seismic event, such as "earthquake," "quarry blast," or "explosion." Differentiating between these types is crucial for filtering and focusing the analysis on natural seismic events.

Vol. 6, No. 2, April 2025, Page. 755-774
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.2378

p. horizontalError: The horizontal error, in kilometers, of the location reported in the latitude and longitude columns. This metric indicates the uncertainty in the earthquake's reported epicenter.
q. depthError: The error, in kilometers, associated with the reported depth of the earthquake. This value helps gauge the reliability of the depth measurement.
r. magError: The estimated standard error of the reported earthquake magnitude. This error estimate is important for assessing the confidence in the magnitude values.
s. magNst: The number of seismic stations used to calculate the earthquake magnitude. A higher number of stations can lead to a more accurate magnitude estimate.
t. status: The status of the earthquake event in the USGS earthquake catalog, such as "reviewed" or "automatic." This status indicates whether the event has been manually reviewed by a seismologist or automatically recorded by the system.
u. locationSource: The ID of the agency or network that provided the earthquake location. This information is crucial for understanding the origin of the data and ensuring its reliability.
v. magSource: The ID of the agency or network that provided the earthquake magnitude. Knowing the source helps in verifying the consistency and accuracy of the magnitude data.

By leveraging this comprehensive dataset, the study aims to build and evaluate predictive models that can accurately estimate earthquake magnitudes. Understanding each column's role and significance ensures that the data is effectively utilized, providing a solid foundation for developing robust and reliable predictive models.

2.2. Data Preprocessing

Handling missing values is a crucial step in data preprocessing to ensure the integrity and accuracy of predictive models. In the USGS earthquake dataset, missing values can occur in several columns, including latitude, longitude, depth, magnitude, and other features. To address missing values, we applied different imputation techniques based on the nature of the data. For numerical columns such as latitude, longitude, depth, and magnitude, mean imputation was used. This method involves replacing missing values with the mean value of the respective column, a technique that is widely used due to its simplicity and ability to maintain data size. Although mean imputation is effective in preserving the overall distribution of data, it may underestimate variance and introduce bias compared to more sophisticated methods, such as multiple imputation [24], [25]. Despite these limitations, it has proven to be one of the more reliable approaches in improving predictive performance in similar datasets, such as in software effort estimation[26].

For categorical variables such as magType, net, type, locationSource, and magSource, mode imputation was employed. This technique replaces missing values with the most frequent category in the respective column. Mode imputation is particularly useful for categorical data as it preserves the most common occurrences, ensuring that the imputed values remain representative of the dataset. The importance of selecting appropriate imputation techniques has been underscored by research showing that improper imputation can degrade the quality of predictions, emphasizing the need to align the method with the data’s missingness mechanism [25], [27]. By applying these imputation techniques, we aimed to minimize data loss and maintain the dataset's robustness for subsequent analysis and model training.

Encoding categorical variables is essential for converting non-numeric data into a numerical format that machine learning algorithms can process. In the USGS earthquake dataset, several columns contain categorical data, such as `magType`, `net`, `type`, `locationSource`, and `magSource`. To encode these variables, the One-Hot Encoding technique was used. This method transforms each categorical value into a separate binary column, indicating the presence or absence of each category.

Vol. 6, No. 2, April 2025, Page. 755-774
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.2378

One-Hot Encoding was implemented using the `OneHotEncoder` class from the `sklearn.preprocessing` module. This approach ensures that categorical variables are encoded without introducing any ordinal relationships, preserving the integrity of the data. By converting categorical variables into a suitable numerical format, the dataset becomes fully compatible with various machine learning algorithms, enabling them to learn effectively from the data.

Normalizing or scaling numerical features is a critical preprocessing step to ensure that all features contribute equally to the model's learning process. In the USGS earthquake dataset, numerical columns such as latitude, longitude, depth, nst, gap, dmin, rms, horizontalError, depthError, and magError were normalized using standard scaling. This technique, implemented through the `StandardScaler` class from the `sklearn.preprocessing` module, transforms the data to have a mean of zero and a standard deviation of one.

Standard scaling is particularly important for algorithms that are sensitive to the scale of input data, such as Support Vector Machines (SVM) and Neural Networks. By normalizing the numerical features, we ensure that the model training process is not biased towards features with larger magnitudes, thereby improving the model's convergence and performance. This preprocessing step helps create a balanced dataset, allowing the machine learning algorithms to learn more effectively and make more accurate predictions.

2.3. Exploratory Data Analysis (EDA)

EDA plays a critical role in understanding the underlying patterns and distributions within the USGS earthquake dataset. One of the primary steps in EDA involves visualizing key features to gain insights into their distributions and relationships. For instance, the distribution of earthquake magnitudes is a crucial aspect to examine. Using histograms, we can visualize the frequency of different magnitude values, revealing the most common earthquake intensities and identifying any skewness in the data. Such visualizations help in understanding the range and central tendency of the earthquake magnitudes, which are essential for developing accurate predictive models.

2.4. Model Development

In this study, four widely recognized machine learning algorithms were utilized to predict earthquake magnitudes: Decision Tree, Random Forest, Support Vector Machine (SVM), and Neural Network. Each of these algorithms brings unique strengths to the predictive modeling process, catering to different aspects of the data's complexity and structure.

The Decision Tree algorithm operates by recursively splitting the dataset into subsets based on the value of input features. This hierarchical approach results in a tree-like model of decisions, where each node represents a feature and each branch represents a decision rule. The final nodes, or leaves, represent the output variable. Decision Trees are intuitive and easy to interpret, making them a popular choice for initial exploratory analysis. Decision Trees have demonstrated their effectiveness in handling complex datasets, such as in energy consumption prediction for buildings [28].

Random Forest, an ensemble learning method, builds multiple decision trees and merges their predictions to improve accuracy and robustness. Each tree in the forest is trained on a random subset of the data, and feature selection is randomized for each split within the trees. This technique helps in reducing overfitting and variance, providing a more generalized model compared to a single decision tree. This technique has been widely applied in various domains, including building energy predictions [28] where it has shown significant performance improvements

The Support Vector Machine (SVM) is a powerful algorithm used for both classification and regression tasks. In regression, SVM attempts to find the optimal hyperplane that best fits the data within a margin of tolerance. It is particularly effective in high-dimensional spaces and is known for its ability to model non-linear relationships through the use of kernel functions. SVM has demonstrated its robustness in diverse applications, such as predicting temperature changes in steel processing, where it outperformed other models like Artificial Neural Networks [29].

Vol. 6, No. 2, April 2025, Page. 755-774
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.2378

Neural Networks, inspired by the structure and function of the human brain, consist of layers of interconnected nodes (neurons), including input, hidden, and output layers. Each connection between neurons has an associated weight, and the network learns by adjusting these weights to minimize prediction errors. Neural Networks are highly flexible and can model complex non-linear relationships, making them suitable for a wide range of predictive tasks. They have been successfully applied in air pollution prediction, where they integrated various environmental parameters to achieve high accuracy [30].

To optimize the performance of each algorithm, specific parameter settings and configurations were applied based on best practices and preliminary experiments. For the Decision Tree, the criterion used was 'mse' (mean squared error) for regression tasks, with the best splitter, no maximum depth to allow full growth unless pruned, a minimum of two samples required to split an internal node, and a minimum of one sample required to be at a leaf node.

In the Random Forest model, the number of estimators was set to 100 (the number of trees in the forest), using 'mse' as the criterion. The maximum number of features considered for splitting a node was set to 'auto' (the square root of the number of features), with no maximum depth, a minimum of two samples required to split an internal node, a minimum of one sample required to be at a leaf node, and bootstrap sampling enabled.

For the SVM, the kernel used was 'rbf' (radial basis function) for non-linear regression, with a regularization parameter (C) of 1.0, an epsilon of 0.1 in the epsilon-SVR model, and a gamma value set to 'scale' (default scaling factor).

The Neural Network was configured as a Multi-layer Perceptron (MLP) with one hidden layer comprising 100 neurons. The activation function used was 'relu' (rectified linear unit), with 'adam' as the solver (stochastic gradient-based optimizer). The alpha value (L2 regularization term) was set to 0.0001, the learning rate was 'adaptive', and the maximum number of iterations was set to 1000.

These configurations were chosen to balance model complexity with computational efficiency, ensuring that each model could be effectively trained on the dataset without overfitting. By systematically tuning these parameters, the models were optimized to achieve the best possible performance in predicting earthquake magnitudes. The use of Decision Tree, Random Forest, SVM, and Neural Network algorithms, along with their specific parameter settings and configurations, provided a comprehensive approach to developing robust predictive models for earthquake magnitude prediction. Each algorithm's unique strengths contributed to a diverse ensemble of models capable of capturing the intricate patterns in the seismic data.

2.5. Training and Evaluation

The first step in the training and evaluation process involves splitting the dataset into training and testing sets. This is essential to evaluate the model's performance on unseen data, ensuring that it generalizes well to new instances. In this study, the dataset was divided using an 80-20 split, where 80% of the data was allocated to the training set and 20% to the testing set. This split ratio is commonly used in machine learning to provide a sufficient amount of data for training while reserving a significant portion for evaluation. The training set is used to train the machine learning models, allowing them to learn the patterns and relationships within the data, while the testing set is used to assess the model's predictive performance.

The `train_test_split` function from the `sklearn.model_selection` module was utilized to perform the split, ensuring a random and unbiased division of the data. This function helps in maintaining the integrity of the dataset by preventing any leakage of information from the training set to the testing set. By using a random seed, the split can be reproduced, ensuring the consistency of the results in subsequent runs. This method provides a robust framework for evaluating the performance of the models and ensuring their reliability in real-world applications.

To evaluate the performance of the predictive models, three key metrics were used: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). These metrics provide a comprehensive assessment of the models' accuracy and predictive capabilities.

MAE measures the average absolute difference between the predicted and actual values. It provides a straightforward interpretation of the model's prediction error, with lower values indicating

Vol. 6, No. 2, April 2025, Page. 755-774
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.2378

better performance. MAE is particularly useful for understanding the average magnitude of errors in the predictions, without considering their direction.

RMSE is the square root of the average of squared differences between predicted and actual values. This metric penalizes larger errors more heavily than MAE, making it sensitive to outliers. RMSE provides a more comprehensive measure of the model's accuracy, highlighting any significant discrepancies between predictions and actual outcomes.

R^2 represents the proportion of variance in the dependent variable that is predictable from the independent variables. It ranges from 0 to 1, with higher values indicating a better fit of the model to the data. R^2 provides an overall measure of how well the model captures the underlying patterns in the data.

These evaluation metrics were computed for each model to provide a detailed understanding of their performance. By comparing the metrics across different models, we can identify the most accurate and reliable predictive algorithms for earthquake magnitude prediction.

To ensure the robustness and generalizability of the models, a cross-validation approach was employed. Cross-validation involves partitioning the dataset into multiple subsets, training the model on some subsets while testing it on the remaining ones. This process is repeated multiple times, and the results are averaged to provide a more reliable estimate of the model's performance.

In this study, k-fold cross-validation with k=5 was used. This approach divides the dataset into five equal parts, training the model on four parts while testing it on the fifth part. This process is repeated five times, with each part being used as the testing set once. The average performance across all five iterations is then calculated to obtain a more accurate assessment of the model's predictive capabilities.

The `cross_val_score` function from the `sklearn.model_selection` module was utilized to perform the cross-validation. This function automates the process of partitioning the data, training the model, and computing the evaluation metrics for each fold. By using cross-validation, we can ensure that the models are not overfitting to the training data and are capable of generalizing well to new instances. This approach provides a robust framework for evaluating the performance of the predictive models and ensuring their reliability in real-world applications.

3. RESULT

3.1. EDA Results

The distribution of earthquake magnitudes is a crucial aspect to examine. Using histograms, as shown in Figure 2, we can visualize the frequency of different magnitude values. This reveals the most common earthquake intensities and identifies any skewness in the data. Such visualizations help in understanding the range and central tendency of the earthquake magnitudes, which are essential for developing accurate predictive models.

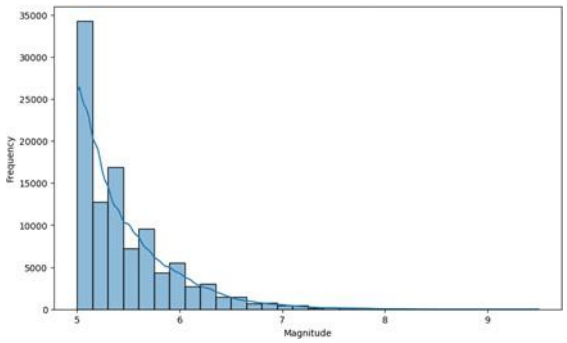

Figure 2. Distribution of Earthquake Magnitudes

Boxplots of depth and magnitude provide insights into the variability and potential outliers in these features. For example, Figure 3 shows a boxplot of earthquake depths, highlighting the distribution and identifying any anomalies in the dataset. These visual tools are instrumental in identifying trends,

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 6, No. 2, April 2025, Page. 755-774
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.2378

anomalies, and regional variations in the dataset, which can inform the feature selection process and model development.

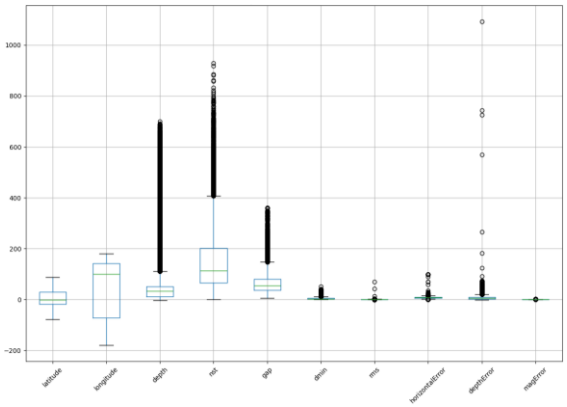

Figure 3. Boxplot of Numerical Features

Correlation analysis is another fundamental component of EDA, aimed at identifying the relationships between different features in the dataset. By calculating correlation coefficients, we can quantify the strength and direction of linear relationships between pairs of numerical variables. For the USGS earthquake dataset, features such as depth, magnitude, latitude, longitude, and others can be analyzed to determine how they are interrelated. A correlation matrix, often visualized using a heatmap as shown in Figure 4, provides a comprehensive view of these relationships, highlighting pairs of features with strong positive or negative correlations.

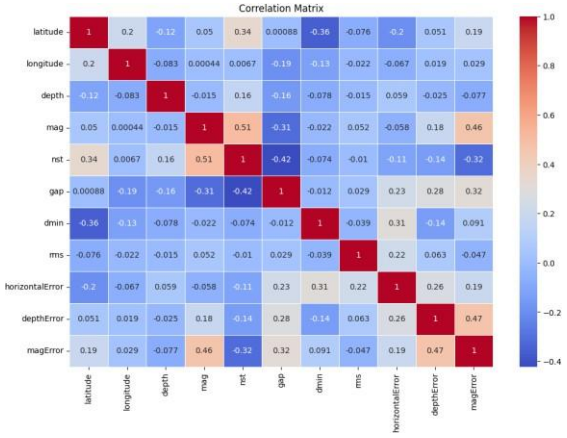

Figure 4. Correlation Matrix

Understanding these correlations is vital for multiple reasons. First, it helps in identifying redundant features that may provide similar information, allowing for more efficient feature selection and reducing the complexity of the model. For example, if two features are highly correlated, one of them might be redundant and could be excluded from the model without significant loss of information. Second, correlation analysis can uncover unexpected relationships that may be important for prediction. For instance, a strong correlation between earthquake depth and magnitude might suggest that deeper earthquakes tend to be of a certain magnitude, which could be a valuable insight for model training. These analyses enable a deeper understanding of the dataset, facilitating the development of more accurate and robust predictive models.

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 6, No. 2, April 2025, Page. 755-774
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.2378

3.2. Model Performance Results

The performance of the four machine learning models—Decision Tree, Random Forest, Support Vector Machine (SVM), and Neural Network—was evaluated using three key metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). These metrics provide a comprehensive understanding of each model's predictive accuracy and robustness.

The Decision Tree model yielded an MAE of 0.282226, an RMSE of 0.435265, and an R^2 score of 0.223995. The cross-validated RMSE (CV_RMSE) was 0.433322. These results indicate that while the Decision Tree model provides a straightforward interpretation, its predictive accuracy is relatively lower compared to the other models.

The Random Forest model outperformed the Decision Tree with an MAE of 0.217051, an RMSE of 0.322398, and an R^2 score of 0.574261. The CV_RMSE for the Random Forest was 0.317599. This improvement can be attributed to the ensemble learning approach of Random Forest, which reduces overfitting and enhances model robustness by averaging the results of multiple decision trees.

The SVM model also demonstrated strong performance, with an MAE of 0.219773, an RMSE of 0.338342, and an R^2 score of 0.531113. The CV_RMSE was 0.331371. SVM's ability to handle high-dimensional spaces and model non-linear relationships contributed to its competitive performance.

The Neural Network model, while flexible and capable of modeling complex relationships, had an MAE of 0.237543, an RMSE of 0.429644, and an R^2 score of 0.243907. The CV_RMSE was 0.337684. Although the Neural Network showed reasonable performance, its results were less impressive than those of the Random Forest and SVM models, possibly due to its higher sensitivity to hyperparameter settings and the need for extensive training data.

These findings are consistent with previous research that highlights the superiority of ensemble methods like Random Forest in predictive tasks involving complex datasets [26], [28]. Similarly, the effectiveness of SVM in high-dimensional spaces has been demonstrated in various domains, including healthcare and environmental prediction[30].

To better illustrate the performance comparison of the models, visualizations shown in Figure 5 represent the evaluation metrics (MAE, RMSE, and R^2) for each model. These visual aids help in clearly depicting the relative strengths and weaknesses of each algorithm.

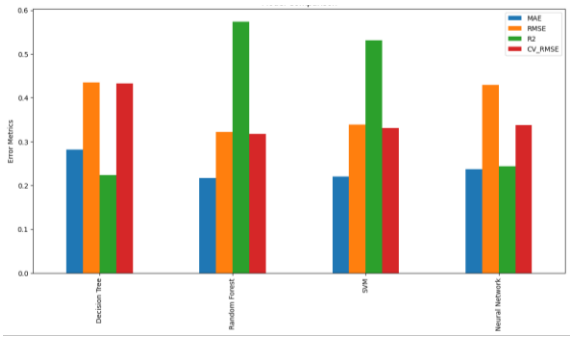

Figure 5. Model Comparison

The bar chart for MAE shows that the Random Forest model had the lowest error, followed closely by SVM, indicating their superior predictive accuracy in terms of absolute error. The Decision Tree and Neural Network models exhibited higher MAE, suggesting larger average deviations from the actual values.

The RMSE bar chart similarly highlights that the Random Forest model had the lowest RMSE, emphasizing its robustness in penalizing larger errors. The SVM model also performed well, while the Decision Tree and Neural Network models had higher RMSE values, indicating greater overall prediction errors.

Vol. 6, No. 2, April 2025, Page. 755-774
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.2378

The R^2 score bar chart provides insight into how well each model explains the variance in the data. The Random Forest model achieved the highest R^2 score, followed by SVM, indicating their strong predictive power. The Decision Tree and Neural Network models had lower R^2 scores, suggesting that they captured less of the underlying data patterns.

These visualizations effectively convey that the Random Forest model consistently outperformed the other models across all metrics, followed by SVM. The Decision Tree and Neural Network models, while still useful, showed comparatively lower performance. These results underscore the importance of selecting appropriate models and tuning their parameters to achieve the best predictive accuracy in earthquake magnitude prediction.

These visualizations provide a clear and comparative view of the model performances, emphasizing the effectiveness of the Random Forest and SVM models in predicting earthquake magnitudes accurately.

4. DISCUSSIONS

4.1. Analysis of Results

The performance of the four predictive models—Decision Tree, Random Forest, Support Vector Machine (SVM), and Neural Network—was evaluated using the metrics Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). The Random Forest model exhibited the best overall performance with the lowest MAE (0.217051), lowest RMSE (0.322398), and highest R^2 (0.574261). This indicates that the Random Forest model was the most accurate and robust in predicting earthquake magnitudes.

The SVM model also performed well, with an MAE of 0.219773, an RMSE of 0.338342, and an R^2 of 0.531113. Although slightly less accurate than the Random Forest, the SVM's performance was still strong, demonstrating its effectiveness in handling non-linear relationships in the data. The Decision Tree and Neural Network models showed comparatively lower performance. The Decision Tree had an MAE of 0.282226, an RMSE of 0.435265, and an R^2 of 0.223995, indicating a higher error rate and less accurate predictions. Similarly, the Neural Network model had an MAE of 0.237543, an RMSE of 0.429644, and an R^2 of 0.243907, suggesting that it struggled to capture the underlying patterns in the data as effectively as the Random Forest and SVM models.

The superior performance of the Random Forest model can be attributed to its ensemble learning approach, which combines the predictions of multiple decision trees to produce a more accurate and robust model. By averaging the results of several trees, the Random Forest reduces overfitting and variance, leading to better generalization on unseen data. This is particularly beneficial in earthquake magnitude prediction, where the data can be complex and noisy.

The SVM model's strong performance is due to its ability to find the optimal hyperplane that separates the data points with maximum margin. The use of the radial basis function (RBF) kernel allows SVM to handle non-linear relationships effectively, making it suitable for the intricate patterns present in seismic data. The decision tree model, while easy to interpret, often suffers from overfitting, which likely contributed to its lower performance. Decision trees can capture complex relationships within the training data but tend to generalize poorly to new data.

The Neural Network model, despite its flexibility and capacity to model complex non-linear relationships, did not perform as well as expected. This could be due to several factors, including the sensitivity of neural networks to hyperparameter settings, the need for extensive training data, and the potential for overfitting if not properly regularized. The relatively higher error rates and lower R^2 values suggest that the Neural Network may have struggled with the specific characteristics of the earthquake dataset.

Vol. 6, No. 2, April 2025, Page. 755-774
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.2378

The Random Forest model's success can be largely attributed to its ability to mitigate overfitting through ensemble learning. By combining multiple decision trees, each trained on different subsets of the data, Random Forest creates a more generalized model that performs well across various scenarios. This robustness is crucial in earthquake prediction, where data variability is high, and capturing the true underlying patterns is challenging.

The SVM model's effectiveness stems from its robust mathematical foundation and ability to handle high-dimensional spaces. The RBF kernel function enables SVM to capture complex, non-linear relationships, making it particularly suitable for seismic data, which often exhibits such patterns. However, SVMs can be computationally intensive, especially with large datasets, which is a consideration for practical applications.

The Decision Tree model's lower performance is likely due to its tendency to overfit the training data. While decision trees are easy to understand and interpret, they can create overly complex models that do not generalize well. Pruning techniques can help mitigate this, but the inherent limitations of single decision trees remain a challenge.

The Neural Network model's performance issues can be attributed to several factors, including its sensitivity to the choice of architecture, learning rate, and other hyperparameters. Neural Networks require significant computational resources and extensive training data to achieve optimal performance. In this study, the complexity of the earthquake data and the potential for overfitting may have hindered the Neural Network's ability to make accurate predictions.

The strong performance of the Random Forest model aligns with findings in other predictive studies, where ensemble learning methods have proven highly effective in managing complex datasets with a mixture of numerical and categorical features [28]. Similarly, SVM’s ability to capture non-linear relationships using kernel functions has been demonstrated in studies on temperature prediction [29]. These findings underscore the adaptability of both models in diverse prediction tasks. Our results are consistent with prior research showing that ensemble models like Random Forest outperform single-tree models in predictive accuracy [26]. In the context of earthquake prediction, similar results have been observed, particularly when incorporating geospatial features. However, the comparatively lower performance of the Neural Network in our study contrasts with its success in other domains, such as air pollution forecasting [30], possibly due to the specific characteristics of the earthquake dataset.

4.2. Practical Implications

The findings from this study have significant implications for the field of earthquake prediction. The superior performance of the Random Forest and SVM models in predicting earthquake magnitudes suggests that these algorithms are particularly well-suited for handling the complex and non-linear relationships inherent in seismic data. This indicates that leveraging ensemble learning and advanced kernel methods can greatly enhance the accuracy and reliability of earthquake prediction models. Accurate predictions of earthquake magnitudes are crucial for early warning systems, allowing authorities to issue timely alerts and take necessary precautions to mitigate the impact of seismic events.

Furthermore, the results underscore the importance of model selection and parameter tuning in the development of predictive models.

Bysystematically evaluating different algorithms and optimizing their parameters, this study demonstrates the potential for significant improvements in predictive performance. These findings can inform the practices of researchers and practitioners in the field, guiding the choice of methodologies and techniques for future earthquake prediction efforts. The enhanced predictive capabilities of these models can lead to more effective disaster preparedness and response strategies, ultimately reducing the risk and impact of earthquakes on communities.

Vol. 6, No. 2, April 2025, Page. 755-774
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.2378

The predictive models developed in this study have a wide range of potential applications beyond academic research. For instance, they can be integrated into existing earthquake early warning systems to provide more accurate and timely predictions of seismic events. This can help emergency management agencies to better plan and execute evacuation procedures, allocate resources more efficiently, and reduce the overall response time during an earthquake.

Additionally, the models can be used by urban planners and engineers to design and construct earthquake-resistant infrastructure. By providing reliable predictions of earthquake magnitudes, these models can inform the development of building codes and standards, ensuring that structures are built to withstand the expected seismic forces. This can significantly enhance the resilience of cities and communities in earthquake-prone areas.

The findings also have implications for the insurance industry. Accurate earthquake predictions can help insurers to assess risks more precisely and develop more effective pricing strategies for earthquake insurance policies. This can lead to better risk management and financial planning, benefiting both insurers and policyholders.

4.3. Limitations

Despite the promising results, this study has several limitations that should be acknowledged. One of the primary limitations is the reliance on historical seismic data from the USGS dataset. While this dataset is comprehensive, it may not capture all the nuances and variables that influence earthquake occurrences and magnitudes. Additionally, the models developed in this study are based on the specific features available in the dataset, and their performance may vary when applied to different datasets with different characteristics.

Another limitation is the computational complexity and resource requirements of some of the models, particularly the Neural Network and SVM. These models require significant computational power and extensive hyperparameter tuning to achieve optimal performance, which may not be feasible in all practical applications. Moreover, the study focused on a limited set of machine learning algorithms, and there may be other advanced techniques that could further improve predictive accuracy.

Future research should aim to address these limitations by exploring a broader range of data sources and incorporating additional features that may impact earthquake prediction. For example, integrating real-time seismic data, geological information, and environmental factors could enhance the predictive capabilities of the models. Additionally, future studies could investigate the use of more advanced machine learning techniques, such as deep learning and ensemble methods that combine multiple models to improve accuracy.

Further research is also needed to develop more efficient and scalable algorithms that can handle the large volumes of data required for real-time earthquake prediction. This includes exploring techniques for reducing the computational complexity of existing models and developing new algorithms that can provide accurate predictions with lower resource requirements.

5. CONCLUSION

This study conducted a comprehensive comparative analysis of four machine learning algorithms—Decision Tree, Random Forest, Support Vector Machine (SVM), and Neural Network—to predict earthquake magnitudes using the USGS earthquake dataset. The evaluation was based on key performance metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). The results indicated that the Random Forest model outperformed the other models with the lowest MAE (0.217051), lowest RMSE (0.322398), and highest R^2 (0.574261). The SVM model also demonstrated strong performance, followed by the Decision Tree and

Vol. 6, No. 2, April 2025, Page. 755-774
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.2378

Neural Network models. The findings highlight the effectiveness of ensemble methods and advanced kernel techniques in capturing the complex relationships within seismic data.

The primary contribution of this study lies in its detailed comparative analysis of different machine learning algorithms for earthquake magnitude prediction. By systematically evaluating the performance of these models, the study provides valuable insights into their strengths and limitations. The superior performance of the Random Forest and SVM models underscores the potential of these algorithms in improving the accuracy and reliability of earthquake predictions. This research contributes to the growing body of knowledge in earthquake prediction by demonstrating the applicability of advanced data mining techniques and providing a robust framework for future studies.

Based on the findings of this study, several practical recommendations can be made. First, it is advisable to utilize ensemble learning methods, such as Random Forest, for earthquake magnitude prediction due to their robustness and ability to handle complex, non-linear relationships in the data. Additionally, integrating SVM models can further enhance predictive accuracy, particularly in scenarios where capturing intricate patterns is crucial. It is also recommended to invest in computational resources and infrastructure that support the efficient training and deployment of these advanced models. Furthermore, continuous monitoring and updating of the models with new seismic data can help maintain their accuracy and relevance over time.

To build on the findings of this study, future research should explore the integration of additional data sources, such as real-time seismic data, geological surveys, and environmental variables, to enhance the predictive capabilities of the models. Investigating the use of deep learning techniques, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), could provide further improvements in accuracy and robustness. Additionally, developing more efficient algorithms that can handle large-scale data in real-time settings would be beneficial. Finally, establishing standardized evaluation metrics and benchmarking datasets would facilitate more consistent and meaningful comparisons across studies, advancing the field of earthquake prediction.

REFERENCES

[1] E. A. Al-Heety, "Evaluation of Return Period and Occurrence Probability of the Maximum Magnitude Earthquakes in Iraq and Surroundings," *Iop Conf. Ser. Earth Environ. Sci.*, vol. 1300, no. 1, p. 012001, 2024, doi: 10.1088/1755-1315/1300/1/012001.

[2] H. Uyanık, "A Multi-Input Convolutional Neural Networks Model for Earthquake Precursor Detection Based on Ionospheric Total Electron Content," *Remote Sens.*, vol. 15, no. 24, p. 5690, 2023, doi: 10.3390/rs15245690.

[3] F. Waldhauser and D. P. Schaff, "A Comprehensive Search for Repeating Earthquakes in Northern California: Implications for Fault Creep, Slip Rates, Slip Partitioning, and Transient Stress," *J. Geophys. Res. Solid Earth*, vol. 126, no. 11, 2021, doi: 10.1029/2021jb022495.

[4] N. Genzano, C. Filizzola, K. Hattori, N. Pergola, and V. Tramutoli, "Statistical Correlation Analysis Between Thermal Infrared Anomalies Observed From MTSATs and Large Earthquakes Occurred in Japan (2005–2015)," *J. Geophys. Res. Solid Earth*, vol. 126, no. 2, 2021, doi: 10.1029/2020jb020108.

[5] B. Li, A. Gong, T. Zeng, W. Bao, C. Xu, and Z. Huang, "A Zoning Earthquake Casualty Prediction Model Based on Machine Learning," *Remote Sens.*, vol. 14, no. 1, p. 30, 2021, doi: 10.3390/rs14010030.

[6] W. Xu, X. Li, and M. Gao, "Temporal Distribution Characteristics of Earthquakes in Taiwan, China," *Front. Earth Sci.*, vol. 10, 2023, doi: 10.3389/feart.2022.930468.

[7] H. Nimiya, T. Ikeda, and T. Tsuji, "Multimodal Rayleigh and Love Wave Joint Inversion for S-Wave Velocity Structures in Kanto Basin, Japan," *J. Geophys. Res. Solid Earth*, vol. 128, no. 1, 2023, doi: 10.1029/2022jb025017.

Vol. 6, No. 2, April 2025, Page. 755-774
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.2378

[8] Y. Zhu, X. Yang, F. Liu, Y. Zhao, S. Wei, and G. Zhang, "Progress and Prospect of the Time-Varying Gravity in Earthquake Prediction in the Chinese Mainland," *Front. Earth Sci.*, vol. 11, 2023, doi: 10.3389/feart.2023.1124573.

[9] J. Lee and T. Hong, "Global Induced Stress Field From Large Earthquakes Since 1900 and Chained Earthquake Occurrence," *Geochem. Geophys. Geosystems*, vol. 22, no. 8, 2021, doi: 10.1029/2021gc009927.

[10] S. Zhu, "Estimation of Seismic Hazard Around the Ordos Block of China Based on Spatial and Temporal Variations of B-Values," *Geomat. Nat. Hazards Risk*, vol. 12, no. 1, pp. 2048–2069, 2021, doi: 10.1080/19475705.2021.1949394.

[11] Y. Wang, Q. Zhao, K. Qian, Z. Wang, Z. Cao, and J. Wang, "Cumulative Absolute Velocity Prediction for Earthquake Early Warning With Deep Learning," *Comput.-Aided Civ. Infrastruct. Eng.*, vol. 39, no. 11, pp. 1724–1740, 2023, doi: 10.1111/mice.13065.

[12] R. Norisugi, "Machine Learning Predicts Earthquakes in the Continuum Model of a Rate-And-State Fault With Frictional Heterogeneities," *Geophys. Res. Lett.*, vol. 51, no. 9, 2024, doi: 10.1029/2024gl108655.

[13] B. Sadhukhan, S. Chakraborty, S. Mukherjee, and R. K. Samanta, "Climatic and Seismic Data-Driven Deep Learning Model for Earthquake Magnitude Prediction," *Front. Earth Sci.*, vol. 11, 2023, doi: 10.3389/feart.2023.1082832.

[14] R. M. Allen and D. Melgar, "Earthquake Early Warning: Advances, Scientific Challenges, and Societal Needs," *Annu. Rev. Earth Planet. Sci.*, vol. 47, no. 1, pp. 361–388, 2019, doi: 10.1146/annurev-earth-053018-060457.

[15] M. R. A. Shahmirani, A. Akbarpour, M. R. A. Ramezani, and E. M. Golafshani, "Buildings, Causalities, and Injuries Innovative Fuzzy Damage Model During Earthquakes," *Shock Vib.*, vol. 2022, pp. 1–11, 2022, doi: 10.1155/2022/4746587.

[16] I. Vrselja, M. Pandžić, and D. Glavaš, "Predicting Earthquake Preparedness Intention Among Croatian Residents: Application of the Theory of Planned Behaviour," *Int. J. Psychol.*, vol. 58, no. 2, pp. 124–133, 2022, doi: 10.1002/ijop.12882.

[17] L. Fallou, M. Corradini, R. Bossu, and J.-M. Cheny, "Preventing and Debunking Earthquake Misinformation: Insights Into EMSC's Practices," *Front. Commun.*, vol. 7, 2022, doi: 10.3389/fcomm.2022.993510.

[18] K. Budiman and Y. Ifriza, "Analysis of earthquake forecasting using random forest," *J. Soft Comput. Explor.*, vol. 2, no. 2, 2021, doi: 10.52465/joscex.v2i2.51.

[19] A. Jufriansah, "Forecasting the magnitude category based on the flores sea earthquake," *J. Resti Rekayasa Sist. Dan Teknol. Inf.*, vol. 7, no. 6, pp. 1439–1447, 2023, doi: 10.29207/resti.v7i6.5495.

[20] T. Kim, J. Song, and O. Kwon, "Pre- and post-earthquake regional loss assessment using deep learning," *Earthq. Eng. Amp Struct. Dyn.*, vol. 49, no. 7, pp. 657–678, 2020, doi: 10.1002/eqe.3258.

[21] "Introduction to Machine Learning," MIT Press. Accessed: Sep. 25, 2024. [Online]. Available: https://mitpress.mit.edu/9780262043793/introduction-to-machine-learning/

[22] F. Hisyam, "Spatio-Temporal Variation Seismicity Pattern in East Java Between 2002 and 2022 Based on the B-Value and Seismic Quiescence Z-Value," *Trends Sci.*, vol. 21, no. 4, p. 7608, 2024, doi: 10.48048/tis.2024.7608.

[23] D. A. Ponce, R. W. Simpson, R. W. Graymer, and R. C. Jachens, "Gravity, Magnetic, and High-precision Relocated Seismicity Profiles Suggest a Connection Between the Hayward and Calaveras Faults, Northern California," *Geochem. Geophys. Geosystems*, vol. 5, no. 7, 2004, doi: 10.1029/2003gc000684.

[24] M. E. Quinteros *et al.*, "Use of Data Imputation Tools to Reconstruct Incomplete Air Quality Datasets: A Case-Study in Temuco, Chile," *Atmos. Environ.*, vol. 200, pp. 40–49, 2019, doi: 10.1016/j.atmosenv.2018.11.053.

[25] D. Zamanzadeh *et al.*, "Autopopulus: A Novel Framework for Autoencoder Imputation on Large Clinical Datasets," 2021, doi: 10.1109/embc46164.2021.9630135.

Vol. 6, No. 2, April 2025, Page. 755-774
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.2378

[26] S. Rafsunjani, R. S. Safa, A. A. Imran, S. Rahim, and D. Nandi, "An Empirical Comparison of Missing Value Imputation Techniques on APS Failure Prediction," *Int. J. Inf. Technol. Comput. Sci.*, vol. 11, no. 2, pp. 21–29, 2019, doi: 10.5815/ijitcs.2019.02.03.

[27] A. U. Chaudhry, W. Li, A. A. Basri, and F. Patenaude, "A Method for Improving Imputation and Prediction Accuracy of Highly Seasonal Univariate Data With Large Periods of Missingness," *Wirel. Commun. Mob. Comput.*, vol. 2019, pp. 1–13, 2019, doi: 10.1155/2019/4039758.

[28] D.-K. Bui, T. N. Nguyễn, T. Ngo, and H. Nguyen-Xuan, "An Artificial Neural Network (ANN) Expert System Enhanced With the Electromagnetism-Based Firefly Algorithm (EFA) for Predicting the Energy Consumption in Buildings," *Energy*, vol. 190, p. 116370, 2020, doi: 10.1016/j.energy.2019.116370.

[29] L. Qiao, J. Zhu, and Y. Wang, "Modeling of Alloying Effect on Isothermal Transformation: A Case Study for Pearlitic Steel," *Adv. Eng. Mater.*, vol. 23, no. 5, 2021, doi: 10.1002/adem.202001299.

[30] M. R. Delavar *et al.*, "A Novel Method for Improving Air Pollution Prediction Based on Machine Learning Approaches: A Case Study Applied to the Capital City of Tehran," *Isprs Int. J. Geo-Inf.*, vol. 8, no. 2, p. 99, 2019, doi: 10.3390/ijgi8020099.

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 6, No. 2, April 2025, Page. 755-774
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.2378