Volume 9, Issue 6, June – 2024 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUN1107

Machine Learning-Powered Earthquake Early Warning System

Vijaya Saraswathi R
Department of CSE
VNRVJIET, Hyderabad

Abstract:- The most devastating natural disasters on earth are earthquakes that causes long-term effects on geography, civilization, and human life. These unpredictable events pose a serious threat to infrastructure. Furthermore, the current Earthquake Early Warning (EEW) systems are facing issues such as limited warning times, false alarms, maintenance costs, high construction costs, and data interpretation. Highlighting these as an urgent need for mitigation measures, there is a need to improve the effectiveness of electronic alerts and public safety measures. For this transformative machine learning techniques and the integration of disparate data, can embark on creating social security and lives protecting from major environmental disasters like earthquakes. This paper has compared various Machine Learning (ML) techniques by training them by using two datasets: one from India and another from India United States Geological from Research World Database to improve the robustness and generality of the earthquake prediction model in the Earthquake Early Warning (EEW) framework. This represents a major advance for earthquake detection and promises to reduce response time. Among various ML Techniques, Random Forest has performed well in earthquake warning with 96.06% accuracy and 98.6% precision.

Keywords:- *Earthquake Early Warning System, Random Forest (RF), Support Vector Machine (SVM), Naïve Bayes, Decision Tree, K-Nearest Neighbors (KNN), Logistic Regression.*

I. INTRODUCTION

Earthquakes occur due to the intense forces caused by shifting ground rocks. This is the result of the continuous movement of tectonic plates beneath the Earth's surface. These massive plates are always in motion, colliding and interacting along fault lines. Over time, this interaction builds up pressure. Eventually, this pressure reaches a critical point, leading to a sudden release of energy in the form of seismic waves. These waves then travel through the ground, causing it to shake and tremble. While the primary reason for earthquakes is the movement of tectonic plates, there are other contributing factors as well. These include volcanic eruptions, the movement of underground magma, significant earthquakes, and even human activities like hydraulic fracturing, though to a lesser extent. Earthquakes can have significant financial consequences, often leading to long-term economic challenges in the affected regions. Additionally, the psychological impact on survivors can be profound, with many experiencing anxiety and distress. To decrease earthquake harm, using early alert systems and following earthquake safety rules is crucial. These alert systems employ sensor networks in quake zones that steadily track seismic motion. When an earthquake strikes, the initial fastest "P waves" are identified by sensors that relay the data to a central site. By analyzing this information, the earthquake's location, strength, and anticipated arrival of larger tremors can be estimated. Alerts are then issued to impacted areas, enabling people time to seek refuge. However, as near the quake's epicenter, these early warning systems weaken in effectiveness, and false warnings detrimentally. Machine Learning has potential ways to improve earthquake detection, giving warnings faster. By looking at earthquake data records and current readings, special computer programs called algorithms learn patterns that may signal tremors happening. Looking back at past earthquake info helps ML models pinpoint locations, strengths, and arrival timings more accurately, resulting in better alert plans and less false alarms. ML algorithms also work through seismic readings quicker than old methods, meaning rapid analysis, and sending out alerts sooner, especially near epicenters. Machine learning models can better detect real earthquakes. They use extra data like GPS and ground changes. This helps tell ground quakes from other causes, cutting false alarms and boosting reliability for early warning. "Fig 1" shows stations tracking global seismic activity worldwide, on main-lands and islands. New stations face money troubles, though plans exist to build more. "Fig 2" maps the latest quake location from April 2024.

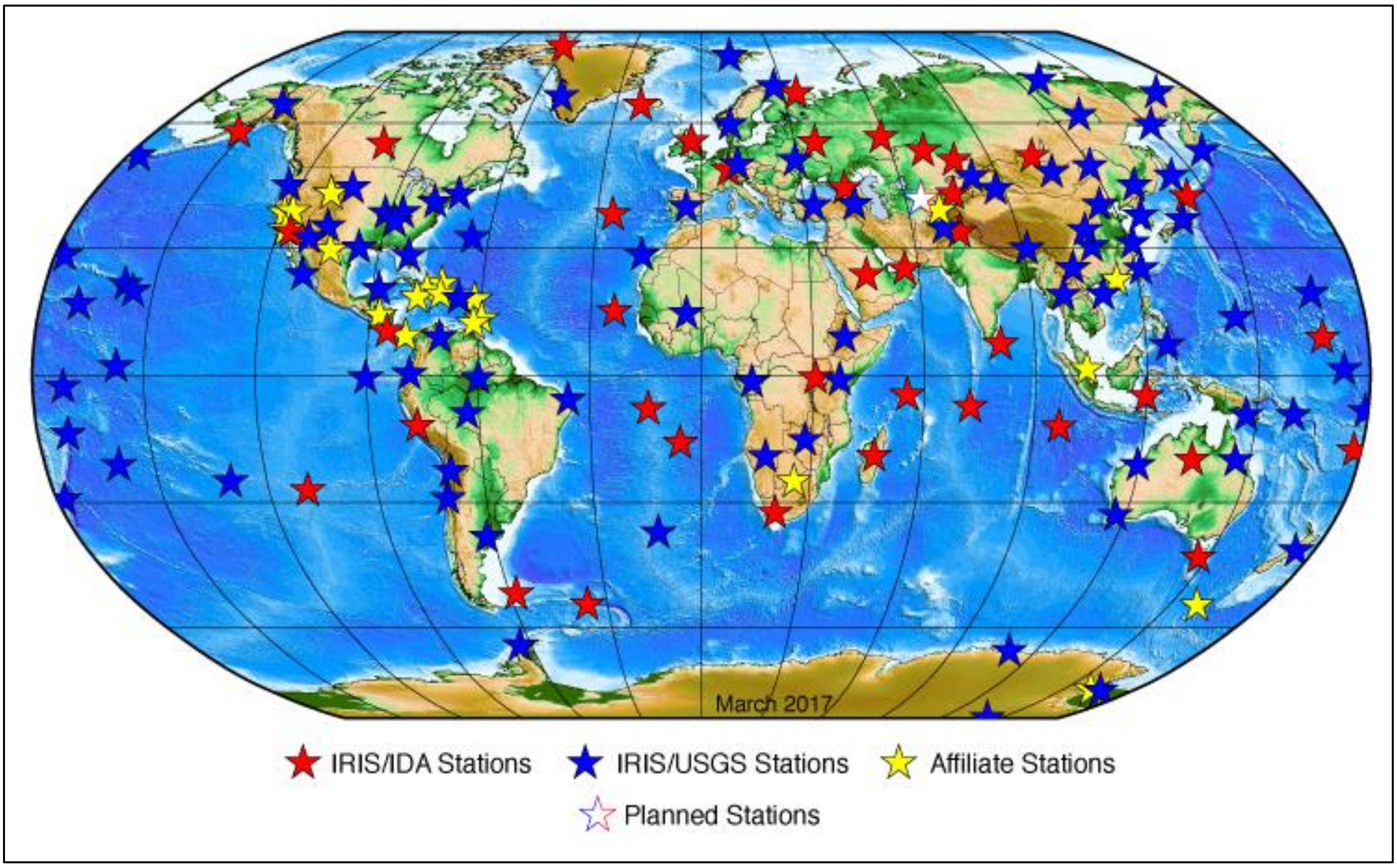

Fig 1 Map of Station Locations of the Global Seismographic Network [https://iris.edu/hq/programs/gsn].

The map illustrates the distribution of stations comprising the Global Seismographic Network (GSN), an international network of seismographic stations aimed at monitoring and studying seismic activity worldwide. These strategically positioned stations enable comprehensive seismic data collection for earthquake research and hazard assessment.

The remaining paper is organized as follows. The second section consist of literature review providing an overview of recent developments in earthquake detection techniques including machine learning algorithms, science related to earthquakes, seismic sensors, and data analysis techniques.

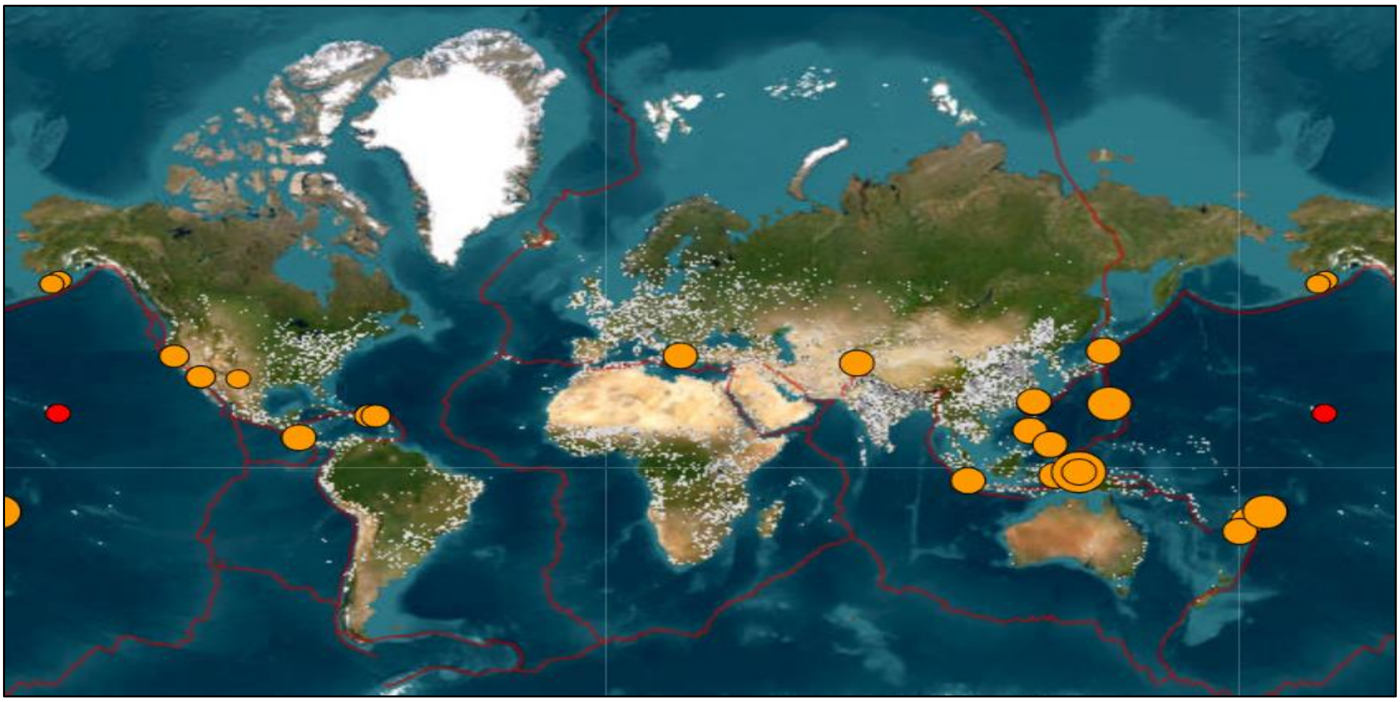

Fig 2 Recent Earthquakes as of Apr 2024.

The third section consist of effectiveness of earthquake research model development and training is explained and discusses the selection of algorithms, mathematical expressions, and computational methods used at the implementation level. The fourth section is results providing summary of the results obtained from testing and evaluation of earthquake models. Finally, the conclusion summarizes the main findings and insights of the study and the contributions of the work. It also suggests some avenues for future research, such as improving the detection algorithms, optimization of sensor networks, and data mining.

II. LITERATURE SURVEY

The literature review offers a concise overview of the many research materials, procedures, and methods employed in early earthquake research. CNNs along with other neural network models have been shown to be promising for seismic data analysis and earthquake prediction. Developing early warning systems that can deliver timely alerts to lessen the impact of earthquakes on buildings and society can be facilitated by combining machine learning and deep learning with sensor technology.

Pal et al. (2023) [1] investigated the issue of seismic hotspots and coldspots, looking at patterns in space and the underlying reasons for variations in seismic activity. By applying geospatial analysis and data modeling, the study determines areas with significantly reduced seismicity (coldspots) and areas with increased seismic activity (hotspots). The study's findings improve our understanding of the spatial distribution of seismic activity and make it possible to put more targeted and effective mitigation measures in place in earthquake-prone areas.

Cremen et al. (2022) [2] conducted a study to investigate the possible impact of seismic early warning systems across Europe. They undertook a rigorous analytical and modeling process to examine the viability of implementing these technologies in a variety of environments. Cremen et al.'s findings enable wise resource allocation and decision-making, opening the door for greater disaster resilience across the continent. A public earthquake early warning system based on smartphones [3] explored user input and system performance indicators to highlight the value of user-centered design and show how smartphone-based solutions can support public safety and preparedness against seismic events. Through the analysis of user feedback and system performance, their research provides insight into the dependability and usability of these systems. In order to determine the ability to predict of deep learning models, Shokouhi et al. (2021) [4] set out to examine their capacity to forecast laboratory earthquakes using seismic data from active sources. They demonstrated the untapped potential of deep learning algorithms for interpreting intricate seismic patterns and precisely forecast the occurrence of earthquakes through extensive experimentation and model training.

A revolutionary journey was initiated by Munchmeyer et al. (2021) [5] when they presented a transformer network-based technique for real-time seismic waveform-based earthquake location and size estimation. They used real seismic data to train their model and assess its performance, demonstrating how deep learning may improve the speed and accuracy of earthquake parameter estimate. Using the groundbreaking metrics τc and Pd, Kumar (2020) [6] started developing earthquake early warning systems specifically for Kachchh, Gujarat, India. Before formulating a plan to strengthen the region's seismic preparedness, which should decrease the consequences of earthquakes and safeguard people and property, Kumar carried out extensive research. A revolutionary machine learning method for assessing earthquake magnitudes—a crucial part of seismic event analysis—was introduced by Mousavi and Beroza (2020) [7]. Their research concentrated on applying machine learning techniques to enhance the accuracy and reliability of magnitude estimates. By using seismic waveform data and advanced algorithms, they were able to achieve notable accuracy advances that set the stage for more precise seismic hazard assessment and risk management.

The Earthquake Transformer is a new deep learning model that was introduced by Mousavi et al. (2020) [8] that can concurrently identify and select different phases of earthquakes. Their work uses deep learning techniques to produce performance never before observed, which constitutes a significant leap in seismic event analysis.

Using seismic station networks, Zhang et al. (2020) [9] presented a novel deep learning approach for the detection of induced earthquakes. Their method tackles the difficult problem of accurately locating generated seismic events, especially in regions that are prone to them. Their work improves our knowledge of and ability to control induced seismic risks by precisely localizing seismic data using deep learning.

In 2020, Chin et al. (2020) [10] presented a cutting-edge earthquake detection system driven by recurrent neural networks (RNNs). This ground-breaking research not only shows the potential of state-of-the-art technology, but it also ushers in a new era of adaptable and responsive seismic monitoring, prepared to accurately and adaptively protect communities from seismic hazards. A novel deep learning method was created by Saad et al. (2020)[11] to categorize earthquake parameters in early warning systems. Their work aims to efficiently classify seismic features by utilizing deep learning's capabilities. By using seismic data to train deep neural networks, they show amazing advancements in classification precision that increase the dependability of early warning systems.

A cutting-edge FPGA-based hardware solution was unveiled by Basu et al. (2019) [12] with the goal of enhancing seismic event identification and noise reduction. Their groundbreaking work highlights the importance of efficient hardware implementations for real-time seismic data processing, which offers significant technological breakthroughs in earthquake monitoring. Their work suggests

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUN1107

more robust earthquake early warning systems by demonstrating better performance and reliability in seismic event detection through the use of FPGA technology. A novel approach to earthquake detection was given by Chin et al. (2019) [13], which addressed the shortcomings of existing methods. Their research uses machine learning and data-driven strategies to increase detection accuracy. In 2016, Kong et al. (2016) [14] introduced MyShake, a revolutionary concept in seismic identification. MyShake created a network that surpasses traditional monitoring methods by converting cellphones into earthquake sensors that can give out real-time notifications. MyShake brought in a new era of community protection through technology by enabling citizen-driven earthquake monitoring.

III. PROPOSED METHDOLOGY

The proposed architecture shown in "fig 3" follows a structured and sequential procedure, initially by following the seismic data collection from indian and usgs worldwide databases. then the collected seismic data undergoes a preprocessing to remove noisy data, null values, to normalize and transform the data into an appropriate format for further analysis. then the machine learning techniques like naive bayes, svm, random forest, logistic regression, and knn, are trained, compared to learn and understand the underlying patterns and relationships. this leads to provide efficient machine learning technique to detect earthquake and aims to reduce the response time. these techniques are thoroughly tested on the two datasets to evaluate their performance. performance metrics, including accuracy, precision, recall, and f1-score are performed to assess the effectiveness of each model in earthquake detection.

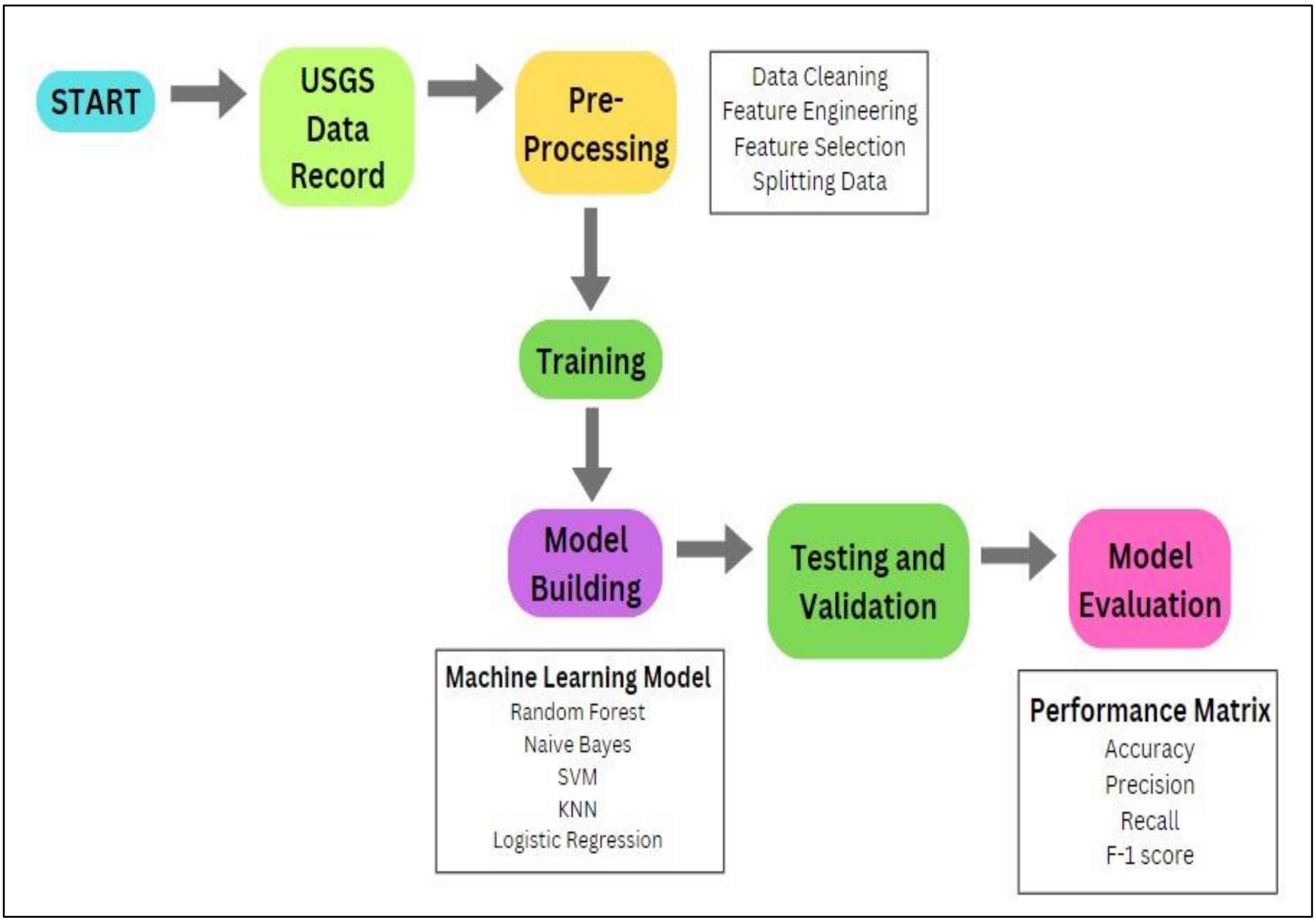

Fig 3 Proposed Architecture

A. Data Collecion

A vast network of seismic stations furnished with sensitive seismographs detects ground movement and converts it into electrical signals to capture the earthquake. These signals are transmitted to the USGS via satellites, internet, phone lines, or microwave telemetry in the remote areas. At USGS facility the powerful computers filter out noise and analyze the seismic wave characteristics. Finally, scientists pinpoint the earthquake's location and measure its magnitude. Then triangulating the data from multiple stations, this processed data is then readily available to the public through numerous USGS platforms for application of various technologies to understand the underlying patterns. The dataset for the Earthquake is collected from the United states geological survey (USGS). The size of overall dataset is 1525 KB consisting of 8000+ records and it consists of 19 attributes.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUN1107

B. Data Preprocessing and Model Traning

The initial step involves inspecting the dataset for any missing values, commonly denoted as NaN values. Should such values be present, various techniques are employed to handle them effectively. These techniques include backfill , forward fill , mode, and mean imputation methods. This paper has opted to utilize the forward fill technique. This method involves replacement of missing values with the last observed non-null values along the corresponding column. By implementing this preprocessing step, it ensures the integrity and completeness of the dataset, thereby facilitating robust and accurate analysis in the subsequent stages of traning the model. In the training phase, the dataset is split into two distinct parts: training and testing which comprise of two variables “X” and “Y”. X representing the features utilized to predict the target variable Y. In the training phase, the dataset is split into two distinct parts: training and testing which comprise of two variables “X” and “Y”. X representing the features utilized to predict the target variable Y. Subsequently, .fit() on training and .predict() on testing the dataset to calculate necessary mathematical operations and predict the occurances of the earthquake to minimize the response time.

C. Machine Learning Techniques

➢ *Naive Bayes:*

A theorem with a "naive" assumption of feature independence, it assumes that features are conditionally independent given the class label. This assumption simplifies the computation, making Naive Bayes highly efficient and suitable for large datasets. The algorithm calculates the probability of each class given a set of feature values using Bayes' theorem.

$$P(A/B) = P(B/A) * P(A)/P(B) \tag{1}$$

$\mathrm{P(A/B)}$: Probability of event A occurring given that event B has already occurred.

$\mathrm{P(B/A)}$: Probability of event B occurring given that event A has already occurred.

$\mathrm{P(A)}$: Probability of event A occurring, and is the probability of event B occurring.

➢ *K-Nearest Neighbors (KNN):*

In earthquake early warning systems, KNN plays a crucial role in analysing seismic data to predict potential earthquakes. By utilizing historical seismic data and its associated features, it categorizes new seismic events based on similarity to the past instances. This similarity-based classification enables the system to identify whether incoming seismic activity signifies an earthquake or not. KNN operates in real-time, swiftly processing incoming data and issuing warnings promptly when predefined thresholds for earthquake classification are met.

$$dist = sqrt((x2 - x1)^2 + (y2 - y1)^2) \tag{2}$$

Where:

$dist$is the distance between the two data points

$(x1, y1)$ are the coordinates of the first data point

$(x2, y2)$ are the coordinates of the second data point

➢ *Suppot Vector Machine(SVM):*

Earthquake detection, can be applied to classify seismic events based on features extracted from seismic wave recordings. Its ability to handle nonlinear relationships and robustness to outliers make it well-suited for accurately identifying earthquake events amidst noisy data. SVM remains a popular choice in earthquake detection systems due to its versatility, effectiveness, and ability to generalize well to unseen data. The equation of the main separator line is called a hyperplane equation.

$$H: wT(x) + b = 0 \tag{3}$$

The distance of any line, $ax + by + c = 0$ from a given point say, $(x0 , y0)$ is given by d.

Similarly, the distance of a hyperplane equation:

$H: wT\Phi(x) + b = 0$ from a given point vector $\Phi(x0)$ can be easily written as

$$dH\big(\phi(x_0)\big) = (|w\text{^}T\ (\phi(x\_0\ )) + b|)/||w||\_2 \tag{4}$$

Were $||w||\_2$ is the Euclidean norm for the length of w given by :

$$||w||_2 = \sqrt{w_1^2 + w_2^2 + w_3^2 + \cdots w_n^2} \tag{5}$$

➢ *Logistic Regression:*

In earthquake detection system, it can be applied to classify seismic events as either earthquake or non-earthquake based on features extracted from seismic wave recordings. These features may include characteristics such as amplitude, frequency, and duration of seismic signals. One of the key advantages of Logistic Regression is its simplicity and interpretability. The model estimates the probability of occurrence of a particular event, allowing for straightforward interpretation of results. Additionally, it can also handle both linear and nonlinear relationships between predictor variables and the outcome, making it versatile for various types of data.

$$y = e\text{^}(b0 + b1 * x) \,/\, (1 + e\text{^}(b0 + b1 * x)) \tag{6}$$

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUN1107

Where:

x is the input value

y is the predicted output

b0 is the bias or intercept term

b1 is the coefficient for the single input value (x)

- *Random Forest:*

In earthquake detection systems, offers exceptional performance in accurately classifying seismic events, leveraging its ability to capture nonlinear relationships and handle noisy data. Its versatility, scalability, and robustness make it a popular choice for a wide range of classification tasks, including real-time earthquake detection within Earthquake Early Warning (EEW) frameworks.

$$Gini\ Index\ =\ 1-\Sigma\ (pi)^2 \tag{7}$$

$$Entropy\ =\ -\sum pi * log2(pi) \tag{8}$$

IV. RESULTS AND DISCUSSION

After the text edit In the proposed system for earthquake prediction, seismic data from India and globally were analysed using various machine learning algorithms. Among SVM, Naive Bayes, Logistic Regression, Random Forest, and KNN, Random Forest consistently outperformed others in accuracy and precision. It demonstrated high accuracy and low false positives. Random Forest was recommended for earthquake prediction tasks, suggesting the exploration of recall and F1 score for deeper insights. Additionally, feature importance analysis was proposed to identify key factors influencing earthquake occurrences.

A. Preprocessing Results

Fig.4 showcases the accuracy and precision metrics for a Support Vector Machine (SVM) model. With an accuracy rate of 94.48%. Additionally, the precision of the SVM model is recorded at 95.33%. This suggests a maximum level of confidence in the model's ability to discern earthquake occurrences accurately from the dataset, and by offering valuable insights of earthquake prediction tasks.

Model: Support Vector Machine
Accuracy: 0.9448529411764706
Precision: 0.9533568578342
Classification Report:

	precision	recall	f1-score	support
no	0.92	1.00	0.96	366
yes	1.00	0.83	0.91	178
accuracy			0.94	544
macro avg	0.96	0.92	0.93	544
weighted avg	0.95	0.94	0.94	544

Fig 4 SVM for Indian Dataset

Fig.5 showcases the report of classification that reveals a well-performing SVM model with accuracy around 95%. Indicating that the model makes very few mistakes in the classifications. Precision and recall, both averaging around 95-96%. Naive Bayes model was trained on Indian and world wide datasets and the results are obtained. Fig.6 shows the Naive Bayes model has an accuracy of 94.4%.The model correctly classified nearly 94.4% of the data points and the precision is 1.0.This means that all the positive predictions made by the model were truly positive for the "earthquake" class.

Model: Support Vector Machine
Accuracy: 0.9540441176470589
Precision: 0.9243568578342
Classification Report:

	precision	recall	f1-score	support
no	0.94	1.00	0.97	366
yes	1.00	0.86	0.92	178
accuracy			0.95	544
macro avg	0.97	0.93	0.95	544
weighted avg	0.96	0.95	0.95	544

Fig 5 SVM for Worldwide Dataset

Fig.6 shows the Naive Bayes model has an accuracy of 94.4%.The model correctly classified nearly 94.4% of the data points and the precision is 1.0.This means that all the positive predictions made by the model were truly positive for the "earthquake" class.

NaiveBayes
Accuracy: 0.94375
Precision: 1.0
Classification Report:

	precision	recall	f1-score	support
earthquake	1.00	0.94	0.97	800
nuclear explosion	0.00	0.00	0.00	0
accuracy			0.94	800
macro avg	0.50	0.47	0.49	800
weighted avg	1.00	0.94	0.97	800

Fig 6 Naïve Bayes for Indian Dataset

Fig.7 shows the performance of a Naive Bayes model on a classification task. The overall accuracy of the model is 0.95.The precision for the class "earthquake" is 0.99, which means that out of all the instances classified as earthquakes by the model, 99% were actually earthquakes.

Naive Bayes
Accuracy: 0.9557356190986406
Precision: 0.9865271848659963
Classification Report:

	precision	recall	f1-score	support
earthquake	0.99	0.96	0.98	14834
explosion	0.00	0.00	0.00	1
nuclear explosion	0.00	0.00	0.00	97
volcanic eruption	0.00	0.00	0.00	1
accuracy			0.96	14933
macro avg	0.25	0.24	0.24	14933
weighted avg	0.99	0.96	0.97	14933

Fig 7 Naïve Bayes for Worldwide Dataset

Fig.8 shows a Random Forest model's performance in classifying seismic events. With an accuracy of 0.9619, the model seems effective, achieving nearly 96% correct predictions. However, a closer look reveals a potential data imbalance issue. The model excels at identifying earthquakes, but struggles with explosions. This suggests the training data might favor earthquakes. Fig.9 shows the performance metrics for a Random Forest classification model, likely used for earthquake detection.

Random Forest
Accuracy: 0.9618964709033684
Precision: 0.9949204725546983
Classification Report:

	precision	recall	f1-score	support
earthquake	1.00	0.96	0.98	14834
explosion	0.00	0.00	0.00	1
nuclear explosion	0.25	0.99	0.40	97
volcanic eruption	0.00	0.00	0.00	1
accuracy			0.96	14933
macro avg	0.31	0.49	0.34	14933
weighted avg	0.99	0.96	0.98	14933

Fig 8 Random Forest for Indian Dataset

The model achieves a high overall accuracy of 0.9705, indicating it correctly classifies nearly 97% of the cases. However, it's important to consider precision as well. While the model has high precision for the "no" class (meaning it accurately identifies most negative cases), the precision for the "yes" class is lower. This suggests the model might make some false positive predictions for the "yes" class (e.g., classifying something as an earthquake when it's not).

Random Forest
Accuracy: 0.9705882352941176
Precision: 0.9684616988066226
Classification Report:

	precision	recall	f1-score	support
no	0.97	0.98	0.98	175
yes	0.97	0.95	0.96	97
accuracy			0.97	272
macro avg	0.97	0.97	0.97	272
weighted avg	0.97	0.97	0.97	272

Fig 9 Random Forest for Worldwide Dataset

Fig.10 reveals the performance of a Random Forest model for earthquake classification. It achieves a high accuracy of 0.9619, successfully classifying almost 96.2% of seismic events. However, the precision for explosions is concerning at 0.00, indicating all predicted explosions were wrong. This highlights a potential bias or limited data for classes like explosions.

Logistic regression
Accuracy: 0.9613970588235294
Precision: 0.92
Classification Report:

	precision	recall	f1-score	support
no	0.95	1.00	0.97	366
yes	0.99	0.89	0.94	178
accuracy			0.96	544
macro avg	0.97	0.94	0.95	544
weighted avg	0.96	0.96	0.96	544

Fig 10 Logistic Regression for Indian Dataset

Fig.11 shows the performance of a Logistic Regression model on a classification task. The accuracy is 0.9559, indicating the model predicts correctly 95.59% of the time. The table also includes a classification report that provides more details on the model's performance for different classes, but the specific classes are not shown in the part of the image.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUN1107

```
Logistic regression
Accuracy: 0.9558823529411765
Precision: 0.92
Classification Report:
              precision    recall  f1-score   support

          no       0.94      0.99      0.97       175
         yes       0.99      0.89      0.93        97

    accuracy                           0.96       272
   macro avg       0.96      0.94      0.95       272
weighted avg       0.96      0.96      0.96       272
```

Fig 11 Logistic Regression for Worldwide Dataset

Fig.12 depicts that KNN classifier model has an accuracy of 0.96, which is 96%. This means that the model correctly classified 96% of the 272 instances it was tested on.The precision of the model is 0.93, or 93%. Precision refers to the model's ability to identify only relevant instances, and not incorrectly classify irrelevant instances as relevant. In this case, the model is good at both accuracy and precision.

```
KNN
Accuracy: 0.9632352941176471
Precision: 0.9306930693069307
Classification Report:
              precision    recall  f1-score   support

          no       0.98      0.96      0.97       175
         yes       0.93      0.97      0.95        97

    accuracy                           0.96       272
   macro avg       0.96      0.96      0.96       272
weighted avg       0.96      0.96      0.96       272
```

Fig 12 KNN for Indian Dataset

Fig.13 shows a classification report for a KNN model.The accuracy of the model is 0.9522.In simpler terms, the model correctly classified nearly 95% of the data points.The precision is 0.9286. This means that out of all the positive predictions made by the model, 92.86% were truly positive.

```
KNN
Original Accuracy: 0.9522058823529411
Precision: 0.9285714285714286
Classification Report:
              precision    recall  f1-score   support

          no       0.97      0.96      0.96       175
         yes       0.93      0.94      0.93        97

    accuracy                           0.95       272
   macro avg       0.95      0.95      0.95       272
weighted avg       0.95      0.95      0.95       272
```

Fig 13 KNN for Worldwide Dataset

Table 1 summarizes the performance of various machine learning models. It compares their accuracy (overall correctness) and precision (correctness of positive predictions) across different models. Random Forest seems to be the best performer with highest accuracy and precision.The classification report evaluates the performance of our SVM model. The high accuracy (around 97%) indicates the model can effectively differentiate between the positive and negative classes in our dataset. This is further supported by the precision (around 96%) and recall (around 95%) metrics, which suggest the model is good at identifying true positives (correctly classifying positive cases) and minimizing false positives (incorrectly classifying negative cases as positive). In the context of our specific classification task (briefly describe your task here), this strong performance signifies the model's ability to accurately classify data points, making it a promising tool for (describe potential applications).

Table 1 Machine Learning Models Summary

Models	**Smmay**	
	Accuracy	**Precision**
SVM	0.9448	0.9533
Naïve Bayes	0.9437	1
RF	0.9618	0.986
Logisric Regression	0.9558	0.92
KNN	0.9522	0.92

The bar graph in Fig.14 depicts a comparison of accuracy and precision for five machine learning models: SVM, Naive Bayes, Random Forest, Logistic Regression, and KNN. The x-axis categorizes the models, while the y-axis represents two separate values: accuracy and precision. Each model is represented by a pair of bars, one for accuracy and another for precision. Random Forest stands out as the best performing model, achieving the highest accuracy of 0.96 that means Random Forest can identify the events of the earthquake better than the other Machine Learning Techniques.

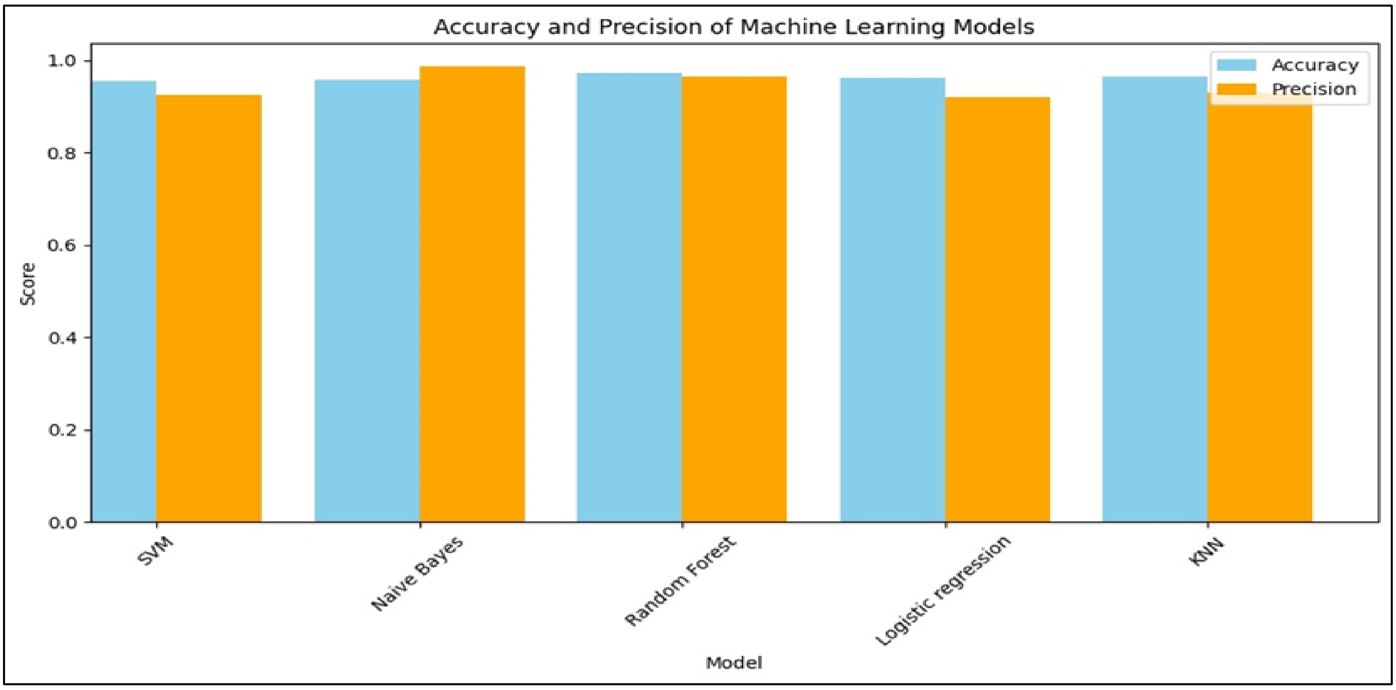

Fig 14 Comparison of Accuracy and Precision for Indian Dataset

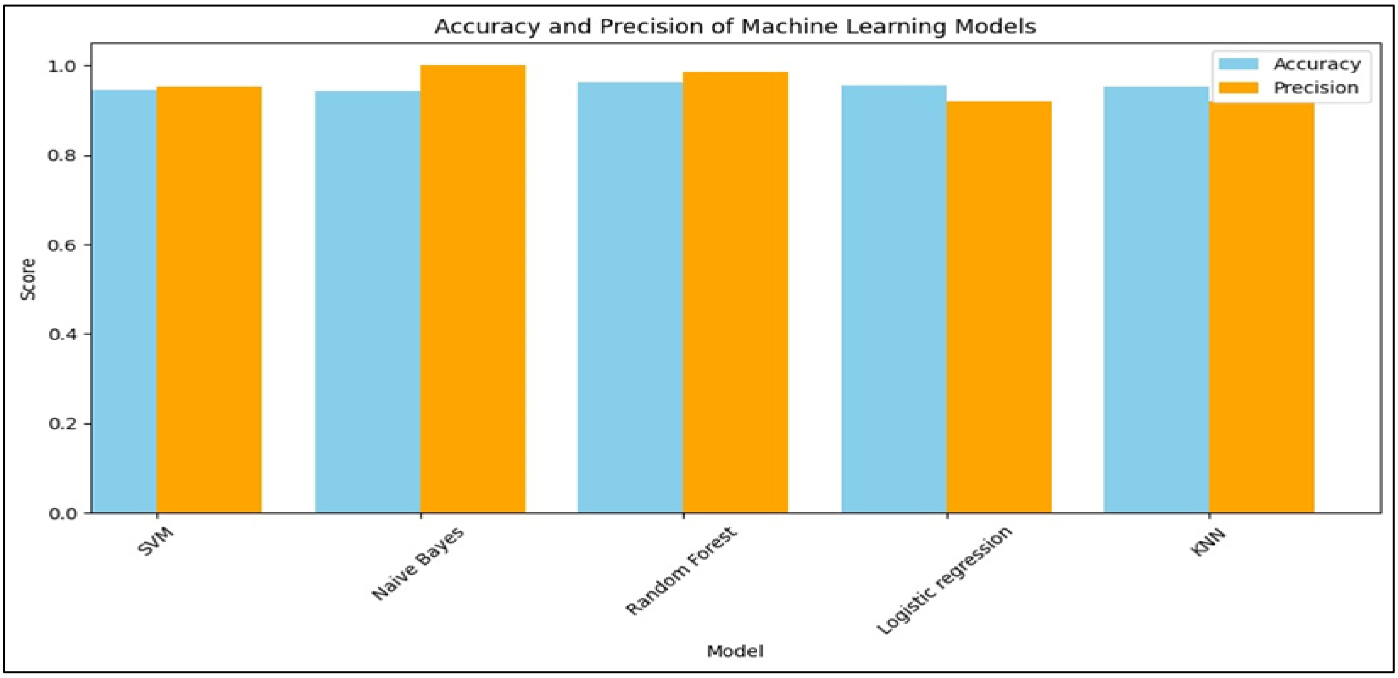

Fig 15 Comparison of Accuracy and Precision for World Dataset

Fig.15 depicts the accuracy and precision of five Machine learning models used for earthquake prediction in terms of bar graph. Models on X-axis, accuracy and precision values on Y-axis. This visualization is instrumental in understanding the strengths and weaknesses of each model. It reveals whether a model prioritizes making a high number of overall correct classifications (earthquake vs non-earthquake) or focuses on highly precise classifications, even if fewer overall. This insight is crucial for selecting the optimal model for specific earthquake prediction needs.

Fig 16 is a line graph showing the accuracy and precision of four machine learning models: SVM, Naive Bayes, Random Forest, and KNN. The x-axis of the graph is labeled "Model" and the y-axis is labeled "Score". The y-axis ranges from 0.92 to 0.98. The graph shows that the Random Forest model has the highest accuracy, followed by SVM, Naive Bayes, and KNN. The accuracy of the Random Forest model is approximately 0.96, while the accuracy of the SVM model is approximately 0.94. The precision of the models is more varied. The Naive Bayes model has the highest precision, followed by Random Forest, KNN, and SVM. The precision of the Naive Bayes model is 1.0, while the precision of the SVM model is approximately 0.92.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUN1107

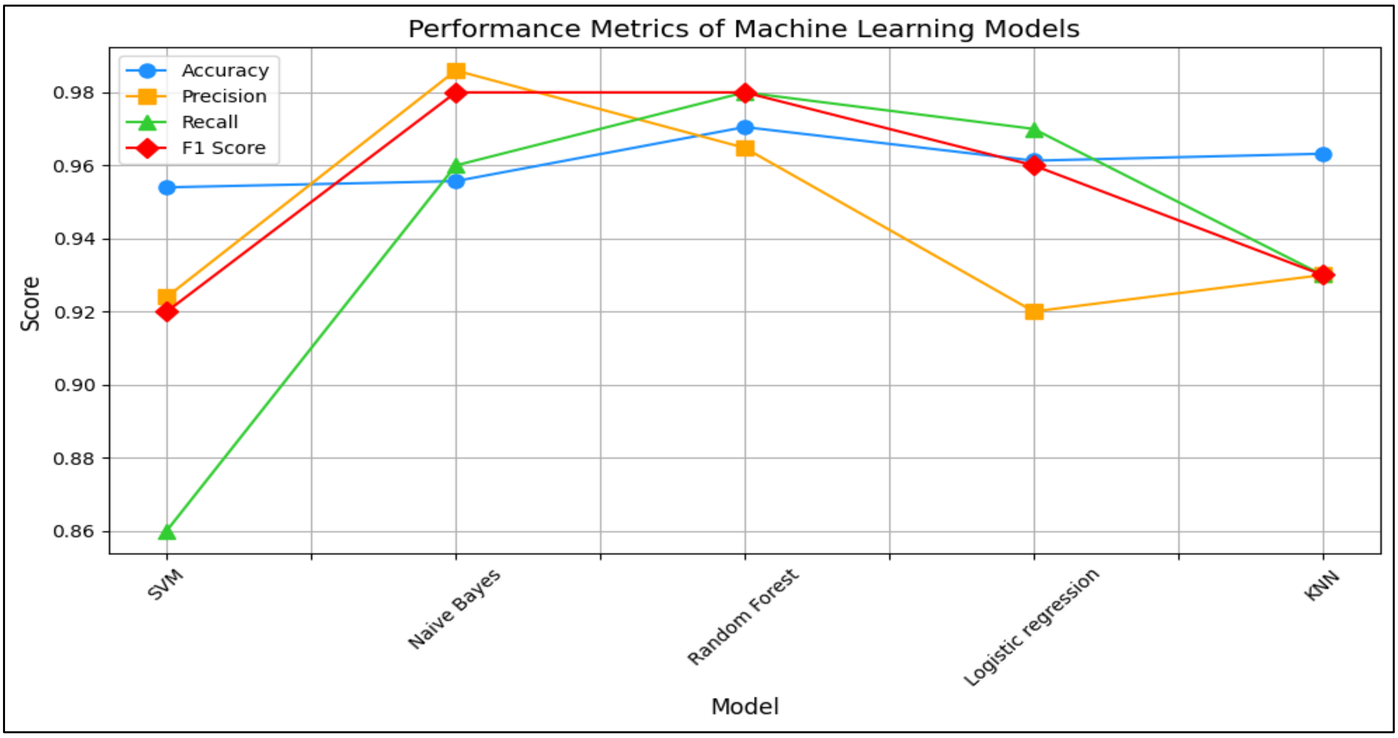

Fig 16 Comparison of Accuracy and Precision for Worldwide Dataset

Fig.17 is a line graph depicting the Accuracy, Recall, F1 Score and precision of four machine learning models on a classification task. The models are Support Vector Machine (SVM), Naive Bayes, Random Forest, and K-Nearest Neighbors (KNN). The x-axis labels the models, while the y-axis represents the "Score" which ranges from 0.92 to 0.98. The graph indicates that Random Forest achieves the highest accuracy (around 0.96), followed by SVM (around 0.94), Naive Bayes (around 0.93), and KNN (around 0.92). Precision scores vary more across the models. Naive Bayes has the highest precision (1.0), followed by Random Forest (around 0.95), KNN (around 0.94), and SVM (around 0.92). Overall, the graph suggests that Random Forest performs best for this specific classification task, while Naive Bayes delivers the most precise classifications. The bar graph in Fig.18 compares the performance of the best models obtained from training on two separate earthquake datasets: Indian and Worldwide.

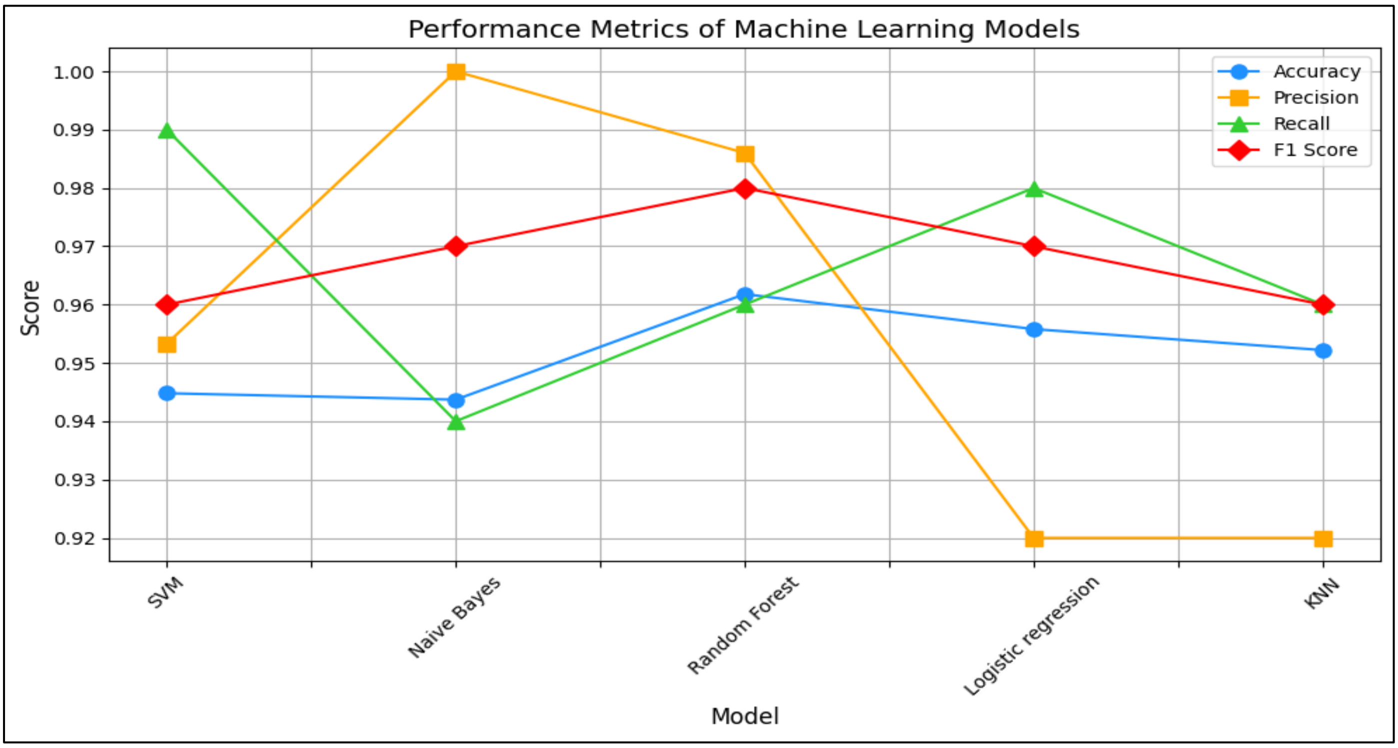

Fig 17 Comparison of Accuracy and Precision for Indian Dataset

In bar graph X-axis represents the dataset of Indian/ Worldwide Dataset, whereasY-axis represents Accuracy/Precession. The height of the bars in the graph indicates metrics for each of the model and also the model which has achived higher accuracy and precision in classifying the earthquake events. This comparison can also helps in understanding whether dataset selection would significantly impact the model performance for earthquake prediction.

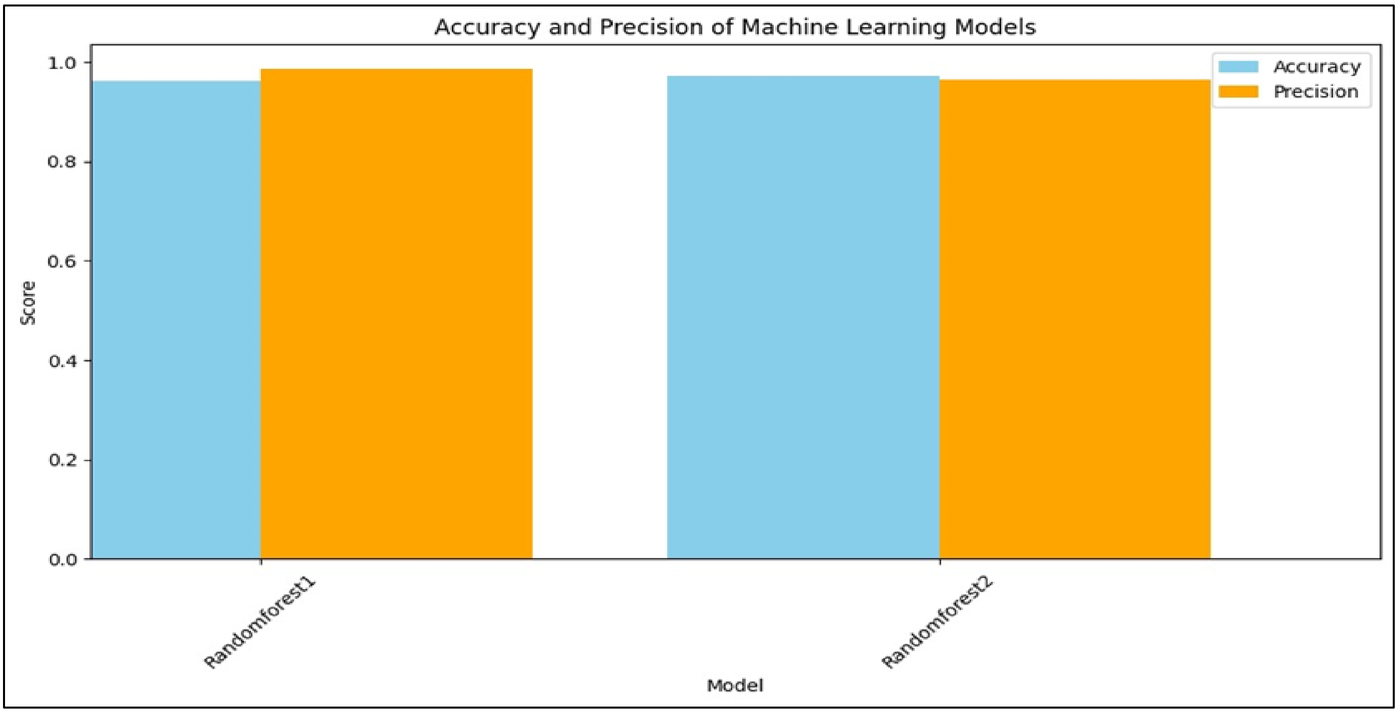

Fig 18 Comparison of best models of Indian & Worldwide Dataset.

It is constantly demonstrated that Random Forest has significantly achieved highest accuracy around 96% amongst other machine learning techniques, including SVM, Naive Bayes, Logistic Regression, and KNN, with respective to the datasets of Indian dataset and the worldwide dataset. The dominance of accuracy can be attributed to several other factors like ensemble learning, integration of predictions from various decision trees to mitigate the overfitting and variance in more robust and accurate predictions. Furthermore, the model can effectively identify relevant features through feature importance calculation, to capture complex relationships in data and make accurate predictions. Moreover, Random Forest exhibits robustness and is suitable for analyzing seismic data prone to such challenges. Overall, Random Forest emerges as the preferred choice for earthquake prediction tasks due to its consistent high accuracy and robust performance across diverse datasets and scenarios.

V. CONCLUSION

In conclusion, the paper focus on providing an efficient machine Learning Technique for earthquake early warning. For this, seismic data from the USGS and Indian databases are used to examine the efficacy of each machine learning models. The data was then used to train the Machine learning models, that includes K-Nearest Neighbors, Support Vector Machines, Random Forest, Naive Bayes, and Logistic Regression. These models were compared to identify the efficient model as earthquake early warning system with performance metrics like Accuracy and Precession. As a result, Random Forest has attained the highest accuracy exceeding 96%. There is also further scope to improve the system by building more complex warning system, by investigating additional sensors data, such as accelerometers from smartphones. For improving the accuracy and shorten warning times can be achieved by looking into the integration of deep learning systems with real-time processing capabilities by integrating advanced data fusion techniques, including combining seismic data with other types of geospatial data such as land use, topography, and geological features, can provide a more comprehensive understanding of earthquake dynamics and improve prediction accuracy to mitigate the impact of seismic events on society.

REFERENCES

[1]. Pal, S. C., Saha, A., Chowdhuri, I., Ruidas, D., Chakrabortty, R., Roy, P., & Shit, M. (2023). Earthquake hotspot and coldspot: Where, why and how? Geosystems and Geoenvironment, 2(1), 100130. https://doi.org/10.1016/j.geogeo.2022.100130
[2]. Cremen, G., Galasso, C., & Zuccolo, E. (2022). Investigating the potential effectiveness of earthquake early warning across Europe. Nature Communications, 13(1), 1–10. https://doi.org/10.1038/s41467-021-27807-2
[3]. Bossu, R., Finazzi, F., Steed, R., Fallou, L., & Bondár, I. (2022). "shaking in 5 seconds!"—Performance and user appreciation assessment of the earthquake network smartphone-based public earthquake early warning system. Seismological Research Letters, 93(1), 137–148. https://doi. org/10.1785/0220210180

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUN1107

[4]. Shokouhi, P., Girkar, V., Rivière, J., Shreedharan, S., Marone, C., Giles, C. L., & Kifer, D. (2021). Deep learning can predict laboratory quakes from active source seismic data. Geophysical Research Letters, 48(12), e2021GL093187. https://doi.org/10.1029/2021gl093187

[5]. Münchmeyer, J., Bindi, D., Leser, U., & Tilmann, F. (2021). Earthquake magnitude and location estimation from real time seismic waveforms with a transformer network. Geophysical Journal International, 226(2), 1086–1104. https://doi.org/10.1093/gji/ggab139

[6]. Kumar, S. (2020). Development of earthquake early warning systems for Kachchh, Gujarat, in India using τc and Pd. Arab. J. Geosci.

[7]. Mousavi, S. M., & Beroza, G. C. (2020). A Machine Learning Approach for Earthquake Magnitude Estimation. Geophysical Research Letters, 47(1), e2019GL085976.

[8]. Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., & Beroza, G. C. (2020). Earthquake transformer: An attentive deep-learning model for simultaneous earthquake detection and phase picking. Nature Communications, 11(1), 1-12.

[9]. Zhang, X., Zhang, J., Yuan, C., Liu, S., Chen, Z., & Li, W. (2020). Locating induced earthquakes with a network of seismic stations in Oklahoma via a deep learning method. Scientific Reports, 10(1), 1-12.

[10]. Chin, T.-L., Chen, K.-Y., Chen, D.-Y., & Lin, D.-E. (2020). Intelligent real-time earthquake detection by recurrent neural networks. IEEE Transactions on Geoscience and Remote Sensing, 58(8), 5440-5449.

[11]. Saad, O. M., Hafez, A. G., & Soliman, M. S. (2020). Deep learning approach for earthquake parameters classification in earthquake early warning system. IEEE Geoscience and Remote Sensing Letters.

[12]. Basu, S., Pandit, S., Chakrabarti, A., & Barman, S. (2019, December). FPGA based hardware design for noise suppression and seismic event detection. In Proceedings of 5th IEEE International Symposium on Smart Electronic Systems (iSES) (pp. 1-6). IEEE.

[13]. Chin, T.-L., Huang, C.-Y., Shen, S.-H., Tsai, Y.-C., Hu, Y. H., & Wu, Y.-M. (2019). Learn to detect: Improving the accuracy of earthquake detection. IEEE Transactions on Geoscience and Remote Sensing, 57(11), 8867-8878.

[14]. Kong, Q., Allen, R. M., Schreier, L., & Kwon, Y.-W. (2016). MyShake: A smartphone seismic network for earthquake early warning and beyond. Science Advances, 2(2), e1501055.