I.J. Information Engineering and Electronic Business, **2025, 5, 62-79**
Published Online on October 8, 2025 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijieeb.2025.05.05

Impact of 2023 Turkey Earthquake Price Hikes: Insightful Socio-Economic Analysis Using Transformer Models and Explainable AI

Muhammed Yaseen Morshed Adib
Southeast University, Dhaka, 1208, Bangladesh
E-mail: yaseen.morshed@seu.edu.bd
ORCID iD: https://orcid.org/0009-0003-0012-8901

Md. Tauhid Bin Iqbal
East West University, Dhaka, 1212, Bangladesh
E-mail: tauhid.iqbal@ewubd.edu
ORCID iD: https://orcid.org/0000-0001-7126-8969

Farig Yousuf Sadeque*
BRAC University, Dhaka, 1212, Bangladesh
E-mail: farig.sadeque@bracu.ac.bd
ORCID iD: https://orcid.org/0000-0001-6797-7826
*Corresponding Author

Received: 24 November, 2024; Revised: 10 January, 2025; Accepted: 12 March, 2025; Published: 08 October, 2025

Abstract: Natural disasters cause economic instability, leading to severe financial hardships for affected communities. The rapid surge in essential goods prices during such events significantly burdens vulnerable populations, highlighting the critical need for timely policy interventions. While understanding public sentiment on economic distress is crucial for effective data-driven policy generation, research specifically analyzing public sentiment on price hikes in such contexts remains limited, often due to a lack of dedicated datasets. To address this, this paper first introduces a novel dataset of social media comments on price hikes related to the 2023 Turkey earthquake. Second, to support data-driven policy-making by quantifying public sentiment, we applied a range of AI models and identified transformer-based models like DistilBERT as particularly effective for sentiment classification. Furthermore, we employ Explainable AI techniques to enhance model trust, enabling policymakers to confidently use these insights to support disaster recovery and economic stabilization in affected regions.

Index Terms: Price hike, Transformer, XAI, Sentiment Analysis, Earthquake

1. Introduction

Natural disasters, such as earthquakes, cause not only physical destruction but also economic instability, leading to severe financial hardships for affected communities. Following the 2023 earthquake in Turkey, the prices of essential goods surged by up to 30% [20], further exacerbating the burden on vulnerable populations. The World Bank estimated direct damages of 34.2 billion USD [21] along with a significantly higher recovery costs, highlighting the urgent need for targeted and timely policy interventions. Understanding public sentiment during such crises is crucial for effective policy responses, and social media serves as a real-time source for capturing public concerns in this regard [23]. Despite growing research on sentiment analysis in disaster contexts, the socio-economic impact of post-disaster price hikes remains largely unexplored. While prior studies have analyzed general disaster-related emotions, they have not adequately captured the interplay between economic instability, price surges, and public perception. This oversight limits the ability of policymakers to effectively address the economic dimensions of disaster recovery.

In this work, we aim to bridge this gap by introducing a novel disaster-specific dataset that tracks sentiment on price hikes before, during, and after the 2023 Turkey earthquake. This dataset categorizes sentiment across several dimensions, including relevance to price hikes, temporal shifts, and regional variations, offering a detailed picture of

public reactions. By analyzing sentiment shifts over time and accurately classifying public sentiment regarding price hikes, policymakers can immediately adjust necessary policies before concerns escalate, ensuring that disaster recovery efforts align with the needs of affected communities. Therefore, to classify these sentiments, we utilize a range of machine learning models, in- cluding traditional and recent approaches, complemented by Explainable AI techniques to enhance model interpretability. Our experiments show that transformer-based models, particularly DistilBERT, achieve the best performance in classifying the sentiments in disaster-related discussions. Applying Explainable AI techniques on the top of transformer model further provides both instance-specific insights and global views of feature importance, ensuring that decision-makers can trust the model's predictions and use them effectively in disaster recovery strategies.

In summary, our key contributions include: (1) the development of a sentiment analysis framework for assessing post-disaster economic conditions, aiming to aid policymakers in understanding public sentiment and formulating targeted interventions (2) the introduction of a structured dataset for tracking public sentiments on economic distress, and

(3) the application of advanced AI techniques to extract actionable insights. These insights can aid decision-makers in formulating data-driven policies to support disaster recovery and economic stabilization.

1.1. Backgrounds

In recent years, sentiment analysis and language technologies have become vital for understanding public reactions to natural disasters and socio-economic issues, with Twitter data serving as a valuable resource. After the Izmir earthquake, researchers analyzed Twitter to examine tweet frequency, recurring themes, and sentiments, providing critical insights into public emotions post-disaster [5]. This analysis helps clarify how individuals react in crises and highlights prominent concerns among affected populations. Similarly, studies of Twitter reactions to earthquakes in Turkey and Syria revealed emotional responses such as empathy and fear, aiding disaster management efforts by informing relief organizations and government agencies about public sentiments [6].

NLP and machine learning techniques have been widely applied across various fields, demonstrating effectiveness in sentiment analysis. For instance, KNN and decision tree algorithms have been successfully used for analyzing social media sentiment and detecting spam emails in cybersecurity. These methods showcase the adaptability of NLP models in diverse applications [7, 8]. Moreover, XGBoost classifier has proven valuable for identifying depression and suicide risks on social media [9]. Additionally, approaches like SVM and random forest model have improved accuracy in phishing email detection. These methods underline the potential of advanced classifiers in tackling complex challenges across digital platforms [10, 11]. In this study, we also utilized RNN-based architectures like Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM) due to their impressive ability to capture semantics [12]. The application of sentiment analysis extends to price hikes, with studies leveraging LSTM-ANN approaches to analyze social media comments and comparing various machine learning algorithms to understand public sentiment regarding price increases [13, 14].

Sentiment analysis has also been applied to socio-economic issues, like rising energy prices. From January 2021 to June 2022, Twitter data was analyzed using transformer-based models and topic modeling techniques to classify sentiments and identify relevant themes [15]. Advanced methods like BERT capture emotions and cluster related conversations, offering insights into public concerns about inflation and government policies. Comparative analyses of emotion recognition models like BERT, RoBERTa, DistilBERT, and XLNet help determine the most effective techniques for sentiment analysis [16]. Techniques like LIME and SHAP have been systematically reviewed for their effectiveness in interpreting AI models, enhancing model transparency and interpretability [17].

Moreover, sentiment analysis supports policymaking by helping policymakers, identify emerging issues, and assess policy impacts. By analyzing social media sentiment, policymakers gain real-time insights into public perceptions, enabling more responsive decision-making [18]. Furthermore, Explainable AI has become essential in supporting public policymaking, enabling policymakers to make more informed and accountable decisions through impactful insights [19]. Several studies have examined sentiment analysis for disaster impact assessment using machine learning and deep learning models. Research on Twitter-based disaster sentiment analysis has highlighted the importance of real-time data in understanding public emotions [5, 6]. However, most prior studies have focused on general disaster responses rather than the economic implications of such events. Some researchers have analyzed price fluctuations, but few have combined sentiment analysis with socio-economic impact studies [13, 14, 15]. Furthermore, the application of Explainable AI (XAI) techniques in this domain remains relatively unexplored. We like to minimize these gaps gaps by utilizing transformer- based models and XAI to assess sentiment trends, highlight key socio-economic concerns, and provide interpretable insights into price hike relatability in disaster contexts.

1.2. Workflow of the study

Our research workflow, as depicted in Figure 1, involved several key stages. Initially, we gathered data from various social media platforms (Facebook, YouTube, Twitter) and online news sources. Multiple annotators labeled the data to ensure accurate sentiment classification. We then conducted exploratory data analysis (EDA) to extract summary statistics and meaningful insights from the dataset, followed by essential preprocessing. Subsequently, we evaluated the data using a range of machine learning models (SVM, Random Forest), deep learning architectures (LSTM, BiLSTM), and advanced transformer models (DistilBERT, XLNet). Model performance was rigorously assessed using Precision, Recall, and F1 score. To enhance interpretability, we employed Explainable AI techniques (LIME and SHAP) on the selected model.

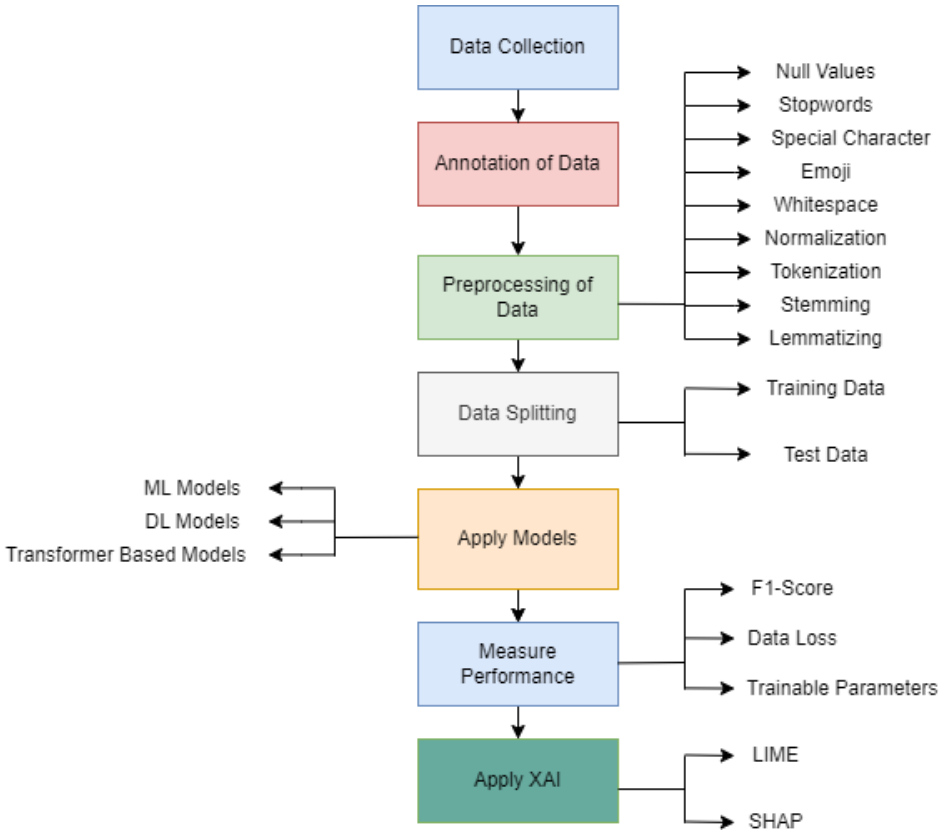


Fig. 1. Workflow of this Research

2. Dataset Creation and Analysis

2.1 Dataset Description

Initially, we selected relevant online sources and social media discussions that provided insights into the Turkey Earthquake 2023 and the subsequent price hikes. After gathering these sources, we structured the dataset around four core attributes such as sentiment, time of posting comments, region of the commenters, and price hike relatability of the comments. These attributes were chosen to capture both the socio-economic impact and the public's sentiment surrounding the event. Subsequently, the dataset was created and these columns were combined to comprehend the social and economic effects of the Turkey Earthquake 2023, especially with reference to the subsequent price increases. Table 1 provides a detailed description of each attribute in the dataset.

Here, the term "Comments" refers to user-provided language that expresses their thoughts about the Turkey Earthquake of 2023 and its effects on the economy. The Sentiment column represents the final sentiment classification assigned to each comment. To determine this final sentiment, three annotators initially reviewed and labeled each comment with their interpretations of its sentiment. These annotations were then aggregated, and based on majority voting among the annotators, a single sentiment label: negative, positive, or neutral was assigned as the final value in the sentiment column. This approach ensures that the sentiment assigned reflects the reliability and accuracy of the dataset. These comments have one of three sentiments: 0, 1, or 2. A score of 0 indicates negativity, a score of 1 indicates positivity, and a score of 2 indicates neutrality. Whether the statement was made before or after the earthquake is indicated by the second property, Time. In this case, 0 indicates remarks made prior to the earthquake, 1 indicates remarks made during the earthquake, and 2 indicates remarks made following the earthquake. The Region feature helps to explain regional differences in public opinion by indicating the comment's geographic origin. In this case, 1 represents comments from Turkey and 0 represents comments from other countries. Price Hike Relatability, the last characteristic, establishes whether or not the comment is related to the price increases brought on by the earthquake. Here, a 1 indicates that the criticism is about the price increase, while a 0 indicates it is not.

Table 1. Attribute Description from the Dataset

Attribute Name	Attribute Description
Comments	Contains textual comments from social media users and news portal readers in English.
Sentiment	Represents the polarity of the comments such as Positive, Negative, and Neutral.
Time (Before, After)	Indicates the period when the comment was made in relation to the earthquake.
Region	Specifies the geographic origin of the comment.
Price Hike Relatability	Indicates whether the comment is related to the price hike following the earthquake.

2.2 Exploratory Data Analysis

Exploratory Data Analysis (EDA) is essential for understanding a dataset's core characteristics before proceeding with modeling, as it helps to identify patterns and relationships that guide data preparation and improve model accuracy. For our analysis, we conducted different EDA Analysis, including statistical analysis of the Dataset, Data Quality Checking using Kappa Score, ensure annotation consistency; and Data Exploration and Analysis.

2.2.1 Statistical Analysis of the Dataset

We present the dataset's statistical overview by outlining its key attributes such as mean, standard deviation, minimum, and maximum values. Table 2 provides a detailed summary using these statistical measures. During this process, the "Comment" attribute was excluded, as it holds no relevance to the statistical analysis.

Table 2. Dataset Description

Statistic	**Attribute**			
	Sentiment	**Time (Before, During, After)**	**Region**	**Price Hike Relatability**
Count	5,052	5,052	5,052	5,052
Mean	1.00	1.55	0.76	0.65
Std	0.81	0.65	0.43	0.48
Min	0.00	0.00	0.00	0.00
25%	0.00	1.00	0.00	0.00
50%	1.00	2.00	1.00	1.00
75%	1.00	2.00	1.00	1.00
Max	2.00	2.00	1.00	1.00

The correlation matrix presented in Table 3 provides insights into the relationships between key attributes in the dataset. Notably, The primary correlation observed is between sentiment and time, suggesting that the earthquake had a significant impact on public sentiment. Additionally, The moderate negative correlation between sentiment and price hike relatability indicates that price hikes were a concern and were more likely to obtain negative sentiments. The strong correlation between region and price hike relatability suggests that Turkey might be more affected by price increases or have more public discussion about them.

Table 3. Correlation among Data Points

Attributes	**Correlation Coefficient**			
	Sentiment	**Time (Before, After)**	**Region**	**Price Hike Relatability**
Sentiment	1.000	-0.500	-0.400	-0.600
Time (Before, After)	-0.500	1.000	0.300	0.600
Region	-0.400	0.300	1.000	0.700
Price Hike Relatability	-0.600	0.600	0.700	1.000

2.2.2 Data Exploration and Analysis

We analyzed social media comments related to the 2023 Turkey earthquake across three periods: before, during, and after the event. The majority of comments (3174) were made after the earthquake, with 1453 during, and 425 before the event, showing a surge in public engagement post-disaster which is represented in Figure 2.

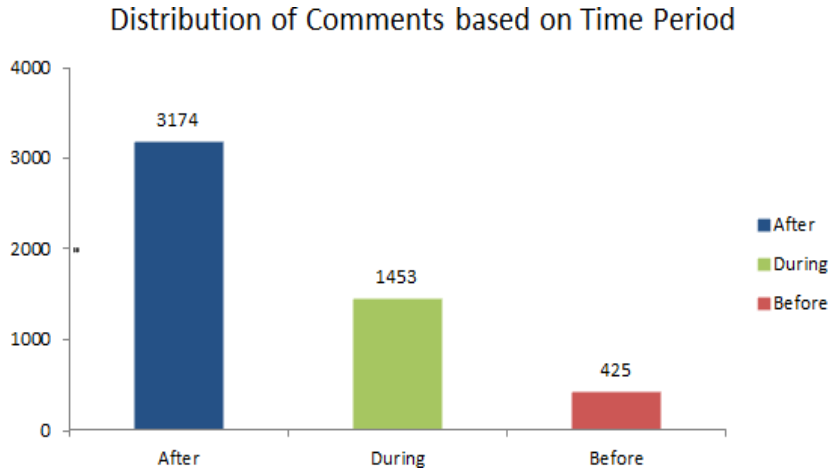


Fig. 2. Distribution of Social Media Comments by Time Period

Next, we examined the comments' focus on the price hike. Of the total, 3277 comments directly addressed the price hike, while 1775 were general. Figure 3 indicates widespread public concern over economic issues following the earthquake.

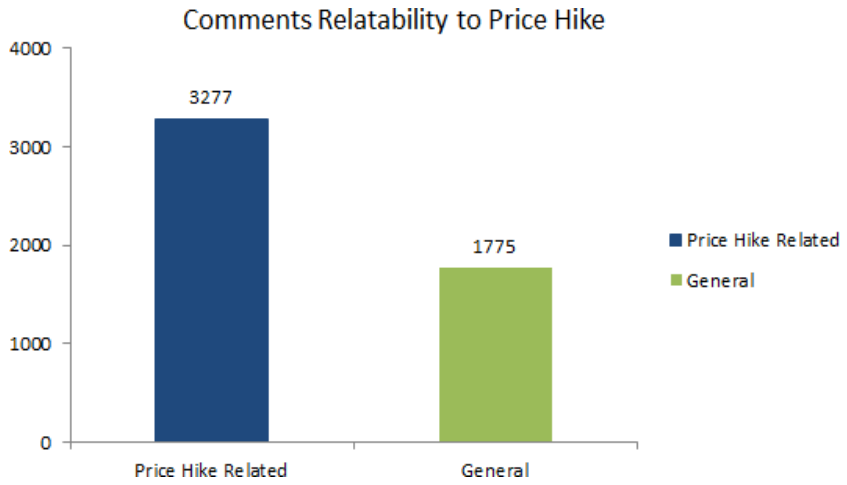


Fig. 3. Comments Repeatability to Price Hike

We analyzed the geographic distribution of comments, finding that most (3,813) came from individuals in Turkey, while 1,239 originated from outside the country, as shown in Figure 4. This indicates both local and international interest in the earthquake's impact.

Social media platforms have different user behaviors, which can introduce biases in sentiment analysis. Twitter users post short, real-time reactions, often expressing immediate emotions. Facebook comments are usually longer and allow for deeper discussions, while YouTube comments vary based on the content. To reduce platform bias, we included data from Twitter, Facebook, and YouTube, ensuring a balanced sentiment analysis. This approach helped capture both quick reactions (Twitter) and more detailed opinions (Facebook and YouTube). However, platform differences still affect sentiment trends. For example, Twitter users may react angrily to price hikes instantly, while Facebook discussions develop over time. Future studies could refine this by weighing sentiment trends based on engagement or adapting models to each platform.

In Figure 5 we present the distribution of comments collected from various social media sources.

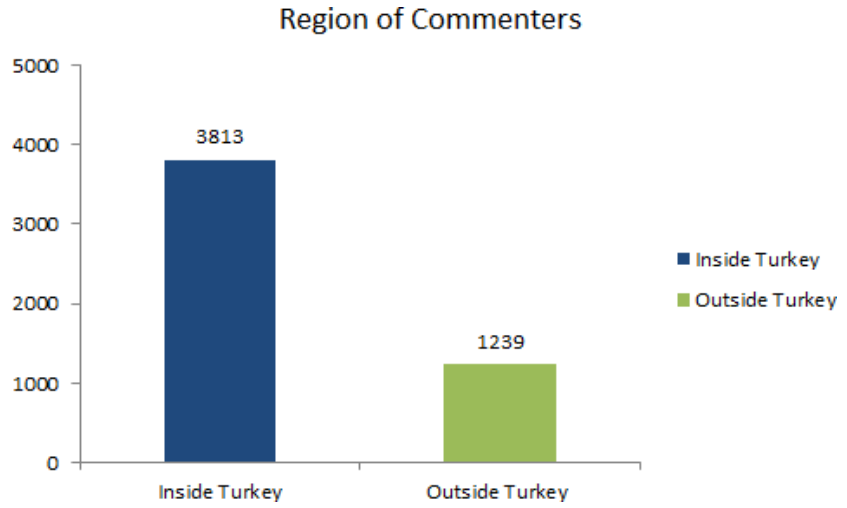


Fig. 4. Region of Commenters

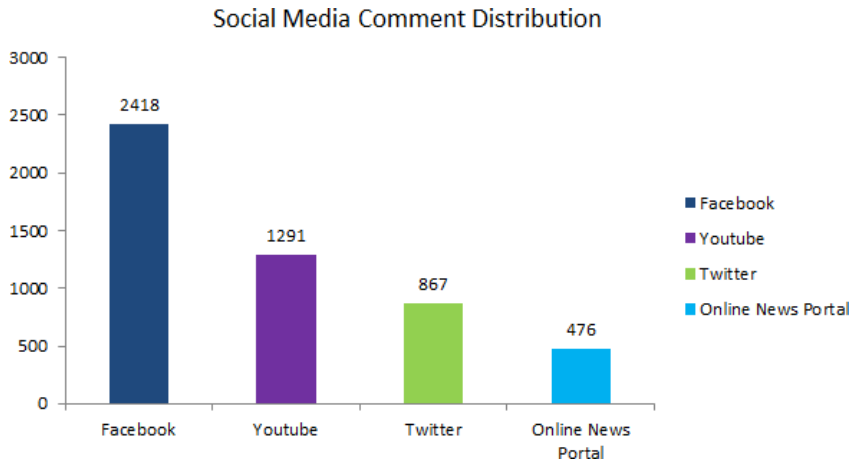


Fig. 5. Ratio of Number of Sources in the Dataset

2.2.3 Data Categorization and Annotation

For the purpose of this data categorization and annotation, we took help from a linguistic expert who suggested guide- lines for categorizing sentences into three classes: negative, positive, and neutral. Negative sentiment (0) was assigned to comments that expressed criticism of price increases resulting from the earthquake, descriptions of personal or collective losses, negative emotions such as sadness, anger, or frustration, and dissatisfaction or blame directed towards authorities for inadequate responses to the post-earthquake challenges. Positive sentiment (1) was applied to statements that conveyed support or hope for recovery, praised resilience or efforts in providing aid, or expressed gratitude for assistance received during or after the earthquake. Neutral sentiment (2) was used for comments that were purely informational, presented a balanced view without strong emotional expression, or discussed the situation without making any clear judgment. These guidelines ensured accurate and consistent labeling of social media comments related to the 2023 Turkey Earthquake and its economic consequences, enhancing the overall quality of our dataset. We collaborated with some trained annotators who carefully followed these guidelines to label the dataset.

The kappa statistic is widely used to measure annotation reliability, providing insight into the consistency of ratings beyond chance agreement. In our study, we have applied the Fleiss' Kappa score as a way to quantify the agreement among multiple raters [24]. This score is highly effective for monitoring and assessing how well raters are in sync. It serves as an effective method to ensure the consistency and reliability of the data we are analyzing. Moreover, it aids in decision-making by illustrating the degree of agreement present. In fields like socio-economics, Fleiss' Kappa is particularly useful for identifying areas of inconsistency. In our analysis, we used Fleiss' Kappa to measure the level of agreement among different reviewers, calculated using the following mathematical formula.

Table 4. Interpretation of Fleiss Kappa Score

Kappa Score Range (%)	Interpretation
0.01 - 0.09	Poor Agreement
0.10 - 0.20	Slight Agreement
0.21 - 0.40	Fair Agreement
0.41 - 0.60	Moderate Agreement
0.61 - 0.80	Substantial Agreement
0.81 - 1.00	Almost Perfect Agreement

Table 4 represents the values associated with the data quality. In our case, the Fleiss' Kappa score is 83%, which resides in the category of almost perfect agreement. This reflects that the inter-annotator agreement is strong enough to process the dataset. Consequently, the comments in the dataset can be reliably fed into various models.

However, to address class imbalance, we balanced the dataset with 1685 negative, 1684 positive, and 1683 neu- tral comments, avoiding resampling techniques. This approach preserved the diversity of opinions while enhancing the model's ability to analyze sentiment accurately across different contexts. We avoid resampling to train our model with unique data, which helps maintain the integrity of the original dataset and ensures that the model learns from a wide range of perspectives.

A key aspect of our data set development was to ensure a balanced representation of sentiment labels – positive, negative, and neutral to mitigate potential bias and improve model accuracy. This equal distribution prevents any single sentiment from disproportionately influencing the model's learning and prediction capabilities. As a result, the model can more effectively discern sentiment patterns within disaster-related content, leading to consistent performance. We deliberately chose to avoid resampling techniques, such as oversampling or under sampling, as these methods can introduce artificial data points and potentially distort the underlying sentiment trends, affecting the reliability of our analysis for disaster recovery planning and economic policymaking. Our approach prioritizes preserving the authenticity of the data set.

3. Experimentation Details

In this section, we describe the details of our experimental settings, including data preprocessing steps, description of ML/DL models, XAI methods, training details etc.

3.1 Data Preprocessing

Text preprocessing is an essential step that includes various techniques to ensure the data is clean and ready for analysis [25]. For our experimental settings, we conduct several data preprocessing tasks on our dataset. We summarize them here:

- **Dropping Null Values:** The first step we perform is removing null values from the dataset, ensuring data integrity before it is fed into machine learning (ML), deep learning (DL), and transformer-based models.
- **Converting Text to Lowercase:** To maintain uniformity and avoid case sensitivity issues, all text is converted to lowercase. This standardization allows models to recognize words without case distinctions, enhancing sentiment analysis accuracy.
- **Removing HTML Tags:** HTML tags are eliminated from the data, as they do not provide meaningful information for analysis.
- **Removing Special Characters and Punctuation:** All special characters and punctuation marks are removed to ensure cleaner input data, enhancing the quality of the dataset.
- **Handling Emojis:** Although emojis can visually express emotions, they do not significantly aid in understanding text sentiment. Emojis are converted into their corresponding text descriptions, preserving their expressive value while simplifying the analysis.
- **Creating a Word Dictionary:** A dictionary is generated to identify unique words within the dataset, which contains 29,172 words in total, with 3,854 unique words—many of which are related to earthquakes and their aftereffects.
- **Lemmatization:** Lemmatization is applied to normalize the text data by converting words to their base forms. This process enhances consistency, reduces dimensionality, and treats various inflected forms of a word as a single entity, improving the coherence of the feature set [26].
- **Word Embeddings:** Word embedding techniques, primarily Word2Vec, are employed to represent data in dense vectors, enabling contextual learning from a large corpus.Word2Vec is an effective word embedding technique that captures semantic relationships between words by mapping them into continuous vector spaces, enhancing the performance of various natural language processing tasks [27]. Although other embeddings like GloVe and TF-IDF were considered, Word2Vec demonstrated superior performance in terms of accuracy and relevance. For the DistilBERT and XLNet models, no additional embedding techniques are required, as these models incorporate contextual embeddings directly within their architecture [28].

3.2 Overview of the Implemented Models

In our sentiment analysis task, we implemented a diverse set of models to ensure comprehensive evaluation and performance comparison. This study employs a structured methodology to analyze sentiment and socio-economic concerns using multiple modeling approaches. Traditional machine learning models such as Random Forest, Support Vector Machines (SVM), and XGBoost are applied to establish baseline results. Additionally, deep learning architectures such as Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM) are implemented to evaluate the role of sequential dependencies in sentiment classification. The study further integrates transformer-based models (DistilBERT and XLNet) to capture contextual understanding in sentiment classification. To ensure a transparent and explainable analysis, XAI techniques (LIME and SHAP) are incorporated. The results from each model are systematically compared using accuracy, precision, recall, and F1-score, allowing for a clear evaluation of the strengths and limitations of each approach. The parametric details of the LSTM, GRU, BiLSTM and BiGRU, XLNet, DistilBERT architecture that we impelemented are presented in Table 5.

We employed a DistilBERT-based architecture for sentiment analysis classification tasks, leveraging the efficiency and performance of transformer-based models. DistilBERT, a compact and faster variant of BERT, maintains effective language understanding capabilities while being computationally efficient. The model was trained and evaluated on a pre-processed dataset, which was split into training and testing sets and tokenized using the DistilBERT tokenizer. The text data was converted into input IDs and attention masks suitable for the model, with a maximum sequence length of

128. A custom dataset class named TurkeyEarthquakeDataset, was defined to manage the encodings and labels. The architecture includes an initial DistilBERT layer, followed by a pre-classifier and classifier layer, with dropout applied for regularization. Fine-tuning was performed for three sentiment classes: negative, positive, and neutral, utilizing the AdamW optimizer and a dynamic learning rate scheduler.

The performance of each model is assessed based on multiple evaluation metrics, highlighting both their strengths and weaknesses. Among traditional machine learning models, Random Forest achieved the highest accuracy (71.60%), in handling structured sentiment data. However, it struggles with complex linguistic patterns and lacks contextual under-

standing. Deep learning models, particularly BiLSTM, performed significantly better, achieving an accuracy of 74.26%, which highlights its ability to capture sequential dependencies in text data. Nevertheless, BiLSTM requires more computational resources and is prone to overfitting when dealing with limited data. The most substantial improvement is observed with transformer-based models, where XLNet achieved an accuracy of 81.20%, effectively capturing contextual depen- dencies across long-range texts. However, it is computationally expensive and requires a large amount of training data to generalize well. DistilBERT outperformed all other models with an accuracy of 82.20%, benefiting from both efficiency and high accuracy. While it delivers superior results, it may still lack domain-specific adaptation for disaster-related sentiment. The comparative analysis highlights the superiority of transformer-based models in capturing complex linguistic patterns and contextual sentiment, but also acknowledges the trade-offs in computational complexity and domain-specific adaptability.

Table 5. Hyperparameter and Parametric Details of Models

Hyperparameter	LSTM	GRU	BiGRU	BiLSTM	XLNet	DistilBERT
Number of epochs	50	50	50	50	30	30
Activation function	Softmax	Softmax	Softmax	Softmax	N/A	N/A
Units	64, 32	64, 32	128, 64, 32	128, 64, 32	N/A	N/A
Embedding vector	Word2Vec	Word2Vec	Word2Vec	Word2Vec	Contextual Embedding	Contextual Embedding
Embedding vector length	64	64	64	64	N/A	N/A
Dropout	0.4	0.4	0.4	0.4	0.2	0.2
Recurrent Dropout	0.25	0.15	N/A	N/A	N/A	N/A
Dense Layers	3	3	3	3	N/A	N/A
Batch Size	16	8	16	16	16	16
Validation Split	0.2	0.2	0.2	0.2	N/A	N/A
Maximum sequence length	N/A	N/A	N/A	N/A	Tokenizer-specific	128
Learning rate	N/A	N/A	N/A	N/A	2e-5	5e-5
Number of classes	3 (neg, pos, neu)	3 (neg, pos, neu)	3 (neg, pos, neu)	3 (neg, pos, neu)	3 (neg, pos, neu)	3 (neg, pos, neu)

Training occurred over 30 epochs with a batch size of 16, incorporating loss calculation via the CrossEntropyLoss function, backpropagation, and parameter updates using the optimizer. A linear learning rate scheduler was also employed. Figure 6 shows the methodology of the modified model. Evaluation on the test set involved comparing predictions to true labels to assess classification performance, with results measured using accuracy and classification reports. This training and evaluation process demonstrated the model's ability to effectively capture and classify sentiments from text data, showcasing the utility of transformer-based models while optimizing for computational efficiency. Table 6 refers to the modifications made in the DistilBERT model.

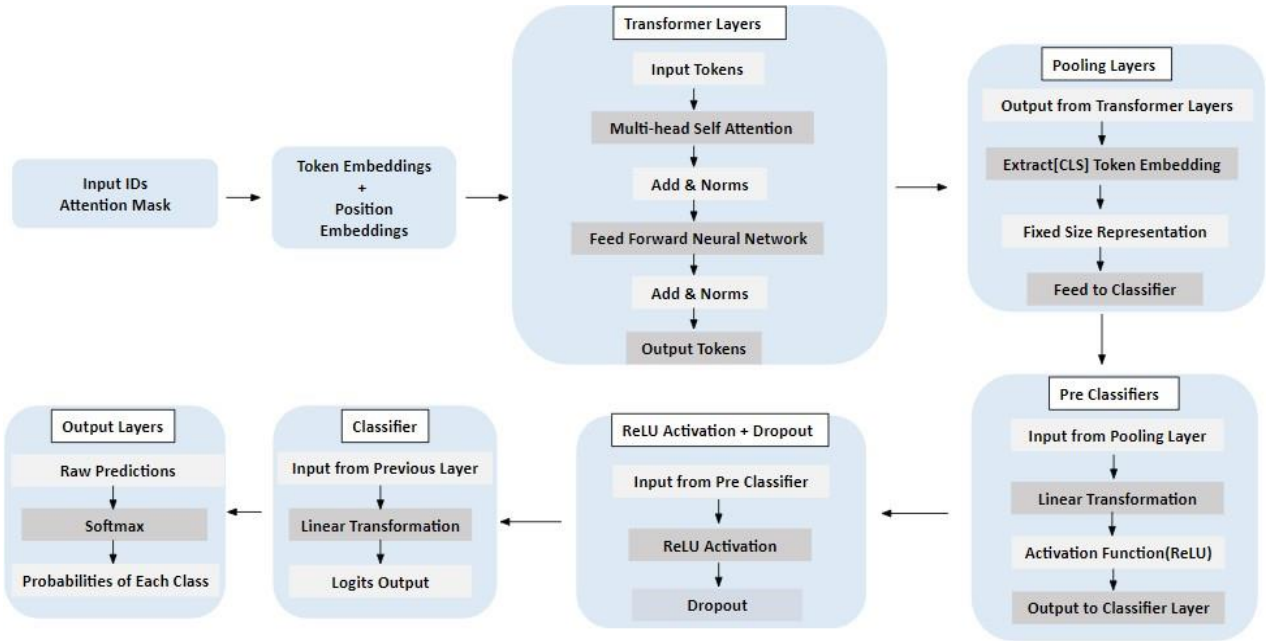


Fig. 6. Methodology of Modified DistilBERT Model

3.3 Explainable AI Methods

In our research, we utilize Explainable AI (XAI) methods to enhance the transparency of our machine learning models. These techniques provide insights into how specific features, such as sentiment, price hike relatability, and region, influence the model's predictions. By offering interpretable explanations, XAI improves trust in the model's outputs, essential.

Table 6. Differences between Traditional DistilBERT and Modified DistilBERT Code

Aspect	Traditional DistilBERT	Modified DistilBERT
Dataset Splitting	Not specified	Explicitly splits the dataset into train and test sets using traintestsplit
Tokenizer Initialization	Uses DistilBertTokenizer	Same as traditional
Tokenization	Tokenizes input data using DistilBertTokenizer	Same as traditional
Dataset Class	Uses standard data loading and tokenization methods	Custom TurkeyEarthquakeDataset class to handle encodings and labels
DataLoader Batch Size	Typically varies, often larger	Uses a batch size of 16
Model Definition	Distil Bert For Sequence Classification from Hugging Face	Custom DistilBERTForSentimentAnalysis class with additional pre-classifier layer and ReLU activation
Pre-Classifier Layer	Not present	Added pre-classifier layer with ReLU activation
Dropout Rate	Defined within DistilBertConfig	Custom dropout defined within the model class
Optimizer	Uses AdamW	Same as traditional
Learning Rate Scheduler	Uses learning rate scheduler	Same as traditional
Training Loop	Standard training loop	Custom training loop with explicit loss calculation and backward pass
Loss Function	Cross-Entropy Loss (implicit within Hugging Face model)	Cross-Entropy Loss (explicitly calculated using nn.CrossEntropyLoss())
Evaluation Method	Standard evaluation loop	Custom evaluation loop with accuracy and classification report
Training Epochs	Typically varies	Set to 30 epochs

For deriving credible insights into socio-economic impacts. We use Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) to explain the models.

LIME explains individual predictions by approximating the behavior of a complex model with a simpler one in the local region around a particular instance. By making slight modifications to the data and analyzing the resulting changes in predictions, LIME provides interpretable insights into which features drive the model's output for a specific example.

$$\theta_{next\ step} = \theta_{current} - \eta.\nabla_{\theta} J(\theta, x^{(i)}, y^{(i)}) \tag{1}$$

LIME's explanation model $\xi(x)$ is trained to approximate the complex model $\boldsymbol{f}$ locally using an interpretable model g, where the loss function L ensures g closely follows $\boldsymbol{f}$ around the point of interest, and $\Omega(g)$ promotes simplicity. LIME enhances the interpretability of the sentiment analysis models. LIME provides local explanations by approximating the model for individual predictions. For example, if a sentence is classified as negative, LIME highlights key features (e.g., "price hikes," "frustrating") that contributed to the sentiment score, validating the model's output and explaining its reasoning.

SHAP, based on cooperative game theory, computes the Shapley values that quantify how much each feature contributes to the model's prediction. It provides both local and global interpretability by fairly distributing the contribution of each feature across all possible subsets of features. SHAP helps understand the overall model behavior by aggregating feature contributions across multiple predictions.

$$\emptyset_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)] \tag{2}$$

Here, ϕ_i is the Shapley value for feature i, $f(S)$ represents the model's prediction for subset S, and N denotes the set of all features. The term $\frac{|S|!(|N| - |S| - 1)!}{|N|!}$ assigns a weight to ensure fair attribution.

SHAP computes Shapley values to measure each feature's marginal contribution across all possible feature subsets. By aggregating these contributions, SHAP provides a global perspective on feature importance, revealing how factors like "region" or "economic concerns" consistently drive sentiment classification. These techniques ensure that stakeholders can understand and trust the model's predictions, helping inform better decision-making in disaster recovery strategies.

Through XAI techniques like LIME and SHAP, we gain deeper insights into our models, making the outputs more interpretable and trustworthy.

3.4 Performance Metrics

In sentiment analysis, we focus on performance metrics that reflect the model's classification ability, such as precision, recall, accuracy, F1-score. A probabilistic interpretation of precision, recall, and F-score is crucial for understanding the evaluation of machine learning models [29]. For deep learning and transformer-based architectures, relying only on precision, recall, and accuracy can be limiting. Thus, we also use the macro F1-score to evaluate performance across all sentiment classes.

The formulas for precision, recall, accuracy, F1-score, and macro F1-score are listed below:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \tag{3}$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{4}$$

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative} \tag{5}$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{6}$$

$$Macro\ F1\ Score = \frac{1}{n} \sum_{i=1}^{n} F1\ Score\ for\ class\ i \tag{7}$$

For our three-class classification task (negative, positive, and neutral sentiments), we use the macro F1-score to ensure balanced performance across all classes. Micro, macro, and weighted F1-scores are commonly used performance metrics for multilabel emotion classification, providing insights into model effectiveness in multi-class scenarios [30]. Our balanced dataset supports fair evaluation using this metric.

Above-mentioned performance metrics like accuracy, precision, recall, and F1-score provide meaningful insights in assessing model performance particularly into real-world scenario. High accuracy and F1-score mean that the models reliably capture public sentiment, which is crucial for policymakers making informed decisions. As an instance, a high recall in detecting negative sentiment shows that the model effectively identifies public frustration, signaling the need for urgent action to that area, such as price controls or financial aid. Precision ensures that only genuine concerns are flagged, helping policymakers focus on the most pressing issues. Thus, by connecting these metrics to economic decision-making, this study helps policymakers understand public sentiment better and take data-driven actions that improve disaster response and recovery efforts.

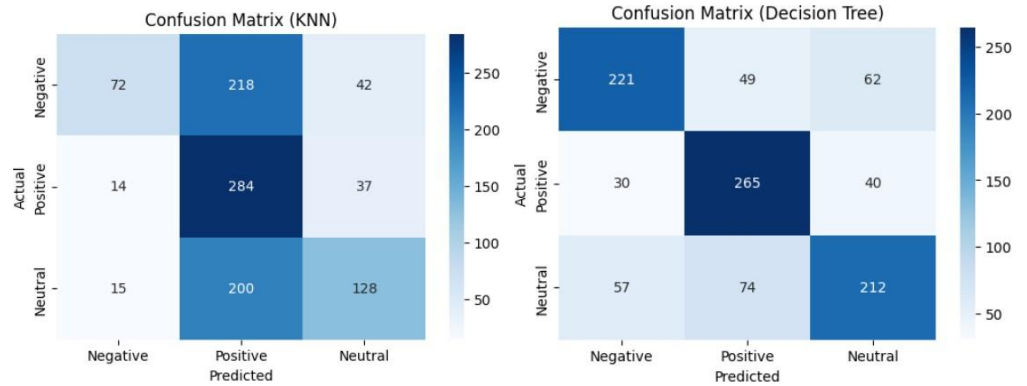
4. Result Analysis

4.1 Results Traditional Machine Learning Models

In this analysis, several machine learning models were compared for sentiment classification based on accuracy, precision, recall, and F1-score. The Random Forest model outperformed the others, achieving the highest accuracy of 71.60% with balanced performance across all sentiment categories. The SVM also demonstrated strong performance, particularly in precision and recall for positive and neutral sentiments. K-Nearest Neighbors (KNN) was the worst performer, with an accuracy of 60.24%. Overall among the machine learning models, Random Forest was the best performing model, while KNN lagged behind. Figure 7 presents the comparative analysis of the results of applied ML models by confusion matrices. The comparative analysis of the results among all the applied models is presented in Table 7.

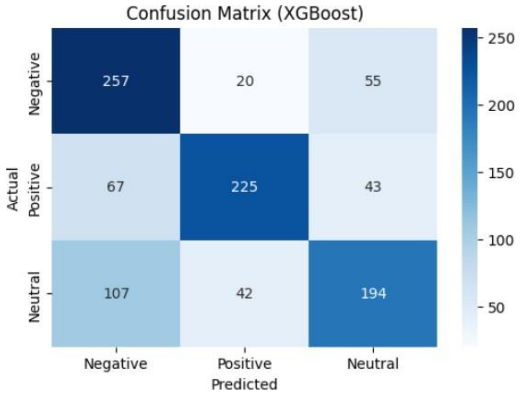
4.2 Results Deep Learning Models

Several deep learning models, including Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), Bidirectional LSTM (BiLSTM), and Bidirectional GRU (BiGRU), were implemented to compare their performance in a sentiment analysis task. Key evaluation metrics such as accuracy, precision, recall, and F1-score were calculated for each model. The BiLSTM model outperformed the others, achieving the highest accuracy of 74.26% with balanced performance across sentiment categories. The BiGRU model followed closely with an accuracy of 73.30%, exhibiting balanced results in terms of precision and recall. The GRU model showed consistent performance with 72.29% accuracy, demonstrating slightly better recall than LSTM. Lastly, the LSTM model achieved an accuracy of 72.60%, but had lower recall for positive sentiments. In Figure 8, the confusion matrices illustrate the prediction strengths and weaknesses across sentiment classes, identifying BiLSTM and BiGRU as the most effective architectures for sentiment classification in this study among DL models.

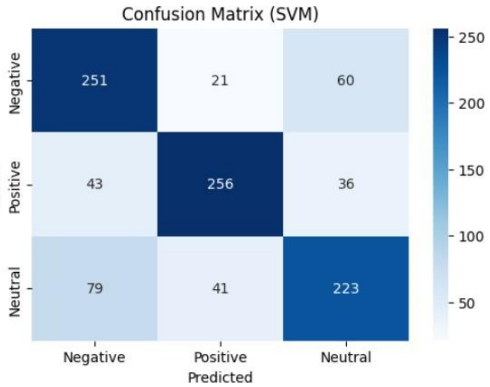


(a) Confusion Matrix of K-Nearest Neighbor

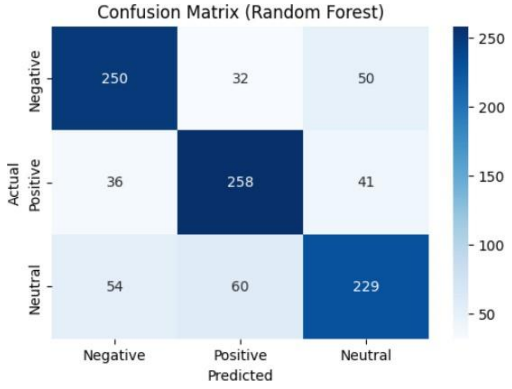
(b) Confusion Matrix of Decision Tree



(c) Confusion Matrix of XGBoost



(d) Confusion Matrix of SVM

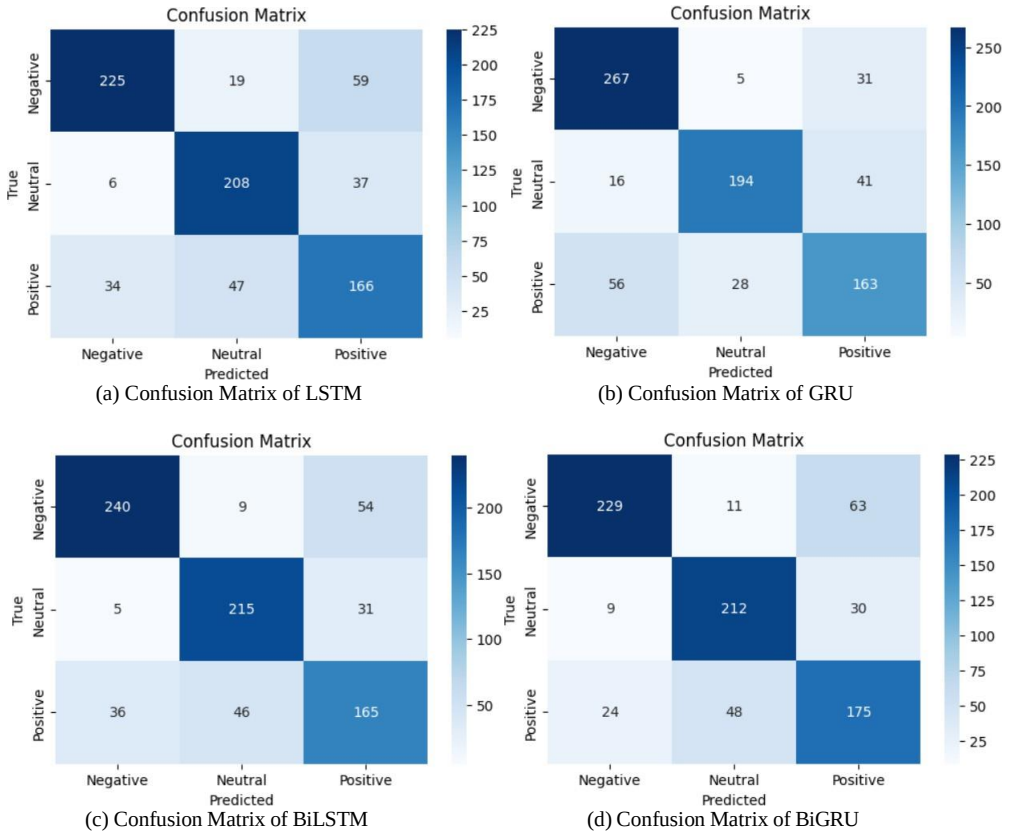


(e) Confusion Matrix of Random Forest

Fig. 7. Comparison of Confusion Matrices of Applied ML Models

4.3 Result Analysis of the XLNet

We evaluated various models, including traditional machine learning (ML), deep learning (DL), and the transformer based model XLNet, for sentiment classification. XLNet emerged as the top performer, achieving an accuracy of 81.20%, along with precision, recall, and F1-scores around 82%, significantly surpassing the other models. The ML and DL models, while effective, did not match XLNet's performance. Deep learning models, such as LSTM, GRU, BiLSTM, and BiGRU, demonstrated balanced classification performance across sentiment categories. In particular, the BiLSTM model had the highest recall for positive sentiments among the DL models. The confusion matrices for these models reveal the distribution of true and false predictions, providing valuable insights into each model's strengths and weaknesses. Overall, XLNet's superior precision, recall, and accuracy make it the most suitable model for sentiment classification in this context, outperforming both traditional ML and DL models.



(a) Confusion Matrix of LSTM (b) Confusion Matrix of GRU

(c) Confusion Matrix of BiLSTM (d) Confusion Matrix of BiGRU

Fig. 8. Comparison of Confusion Matrices of Applied DL Models

4.4 Result Analysis of the Modified DistilBERT

We assessed the performance of various models, including traditional machine learning, deep learning, and the transformer-based model, DistilBERT. DistilBERT emerged as the best-performing model, achieving an average accuracy of 82.20%. It also demonstrated impressive precision (84.81%), recall (83.79%), and F1-score (84.30%), outperforming both machine learning and deep learning models across all sentiment categories.

Here, we present the confusion matrices for the transformer-based models utilized in Figure 9. These matrices provide a clear depiction of each model's performance by displaying how they classify the sentiment categories of Negative, Positive, and Neutral. Through the confusion matrices, we can evaluate how well these transformer architectures manage the complexity of our dataset.

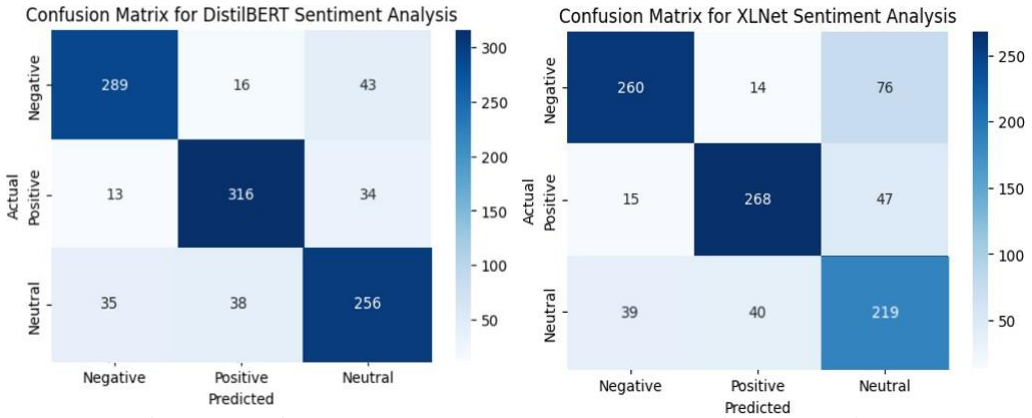
We would like to note that the pre-trained DistilBERT and XLNet in this study were fine-tuned using our earthquake related sentiment dataset. However, future studies could explore further improvements by incorporating additional disaster datasets from diverse geographic locations to examine how sentiment varies across different socio-economic conditions and cultural contexts. Expanding the dataset to include multiple disasters would allow for a more generalizable model capable of analyzing sentiment trends across various crisis scenarios.

4.5 Performance Comparison: DistilBERT vs. XLNet Using McNemar's Test

In our sentiment analysis task, we evaluated two transformer-based models, DistilBERT and XLNet, finding that DistilBERT significantly outperforms XLNet according to McNemar's test. The McNemar test is a statistical method for assessing significant differences between paired nominal data, often used in machine learning to compare model performances [31]. This statistical test compares model performance based on prediction disagreements, indicating a statistically significant difference in accuracy. The results of the McNemar's test were as follows:

- **Test Statistic**: 85.8
- **p-value**: 0.03918

Specifically, DistilBERT classified a higher proportion of data points correctly, highlighting its superiority in this task, particularly in the context of analyzing social media comments related to the 2023 Turkey earthquake.



(a) Confusion Matrix of DistilBERT (b) Confusion Matrix of XLNet

Fig. 9. Comparison of Confusion Matrices of Transformer Based Models

4.6 Overall Comparative Analysis

Table 7 presents a overall comparison of the results of all the above-mentioned models. The comparison highlights the superior performance of transformer-based models, particularly DistilBERT and XLNet, over both traditional machine learning (ML) and deep learning (DL) models across all evaluated metrics. DistilBERT achieved the highest accuracy at 82.20%, followed closely by XLNet with an accuracy of 81.20%. DistilBERT excelled with balanced and superior precision 84.81%, recall 83.79%, and F1-score 84.30%.

In summary, DistilBERT notably outperformed both ML and DL models, achieving an accuracy of 82.20% and also DistilBERT excelled with balanced and superior precision 84.81%, recall 83.79%, and F1-score 84.30% While XLNet also showed strong performance with an accuracy of 81.20%. Therefore, transformer-based models, especially DistilBERT, emerged as the most efficient choices for sentiment classification tasks in this context.Figure 10 show the details of the F1-Score Comparison for all the applied Models.

Table 7. Comparison of Performance Metrics Across ML, DL, DistilBERT, and XLNet Models

Model	Accuracy	Precision	Recall	F1 Score
KNN	60.24%	65.60%	64.60%	65.10%
Decision Tree	67.40%	70.90%	70.30%	70.60%
XGBoost	67.30%	70.52%	70.20%	70.36%
SVM	71.32%	72.60%	72.90%	72.20%
Random Forest	71.60%	74.12%	73.40%	73.76%
LSTM	72.60%	73.93%	74.67%	74.30%
GRU	72.29%	73.50%	72.76%	73.13%
BiLSTM	74.26%	77.36%	76.02%	76.69%
BiGRU	73.30%	74.98%	74.60%	74.79%
XLNet	81.20%	82.36%	82.24%	82.30%
DistilBERT	82.20%	84.81%	83.79%	84.30%

4.7 Model Interpretation Using Explainable AI

In modern machine learning applications, models like DistilBERT offer powerful predictions, but their complexity can hinder understanding of decision-making processes. Explainable AI (XAI) techniques aim to bridge this gap, providing insights into model behavior, which is essential for trust, fairness, and informed decision-making in sensitive areas like policy making. XAI allows interpretation of individual predictions, aiding in the identification of patterns, biases, and key drivers in large datasets. By implementing different XAI methods, we can gain deeper insights into the data, refine models, and achieve a nuanced understanding of underlying phenomena. In this part, we implemented two XAI techniques: LIME and SHAP to interpret our DistilBERT model. These methods enhance model transparency and clarify the factors influencing sentiment classifications. LIME is a widely used method for providing local explanations for machine learning model predictions, offering insights into model decision-making [32]. On the other hand, SHAP enhances interpretability in models by quantifying the contribution of each feature to the model's predictions [33].

LIME (Local Interpretable Model-agnostic Explanations) explains predictions by approximating model behavior lo- cally around specific instances. It perturbs input data and observes changes in predictions, creating a simpler model for that instance. Figure 11 shows LIME applied to a test sentence, highlighting the contributions of individual words to the negative classification. Key words like "heartbreaking*"* and *"devastating"* significantly influenced the model's prediction, providing transparency to the decision-making process.

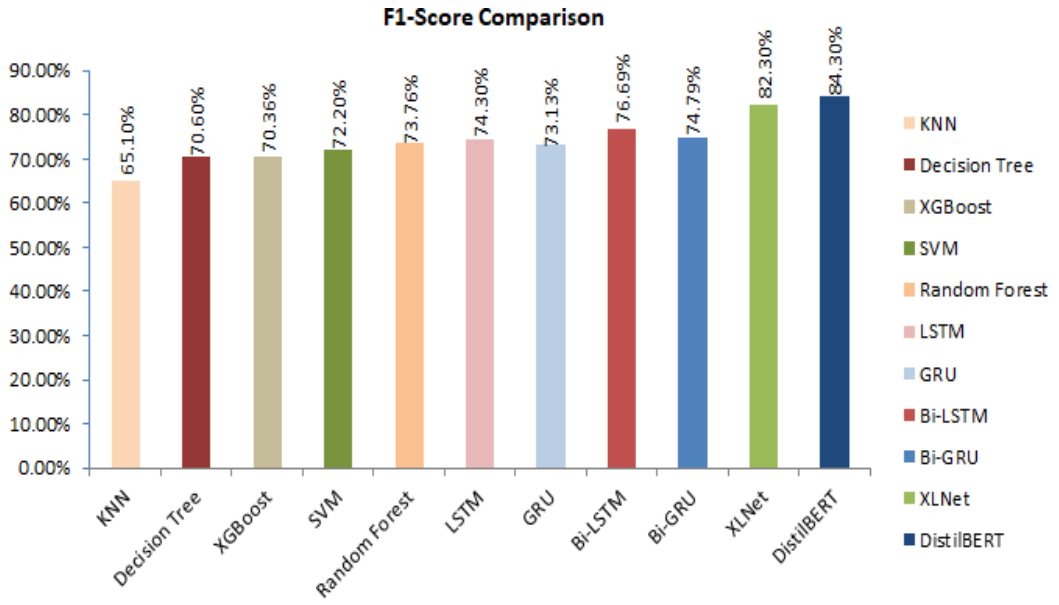


Fig. 10. F1-Score Comparison of different models

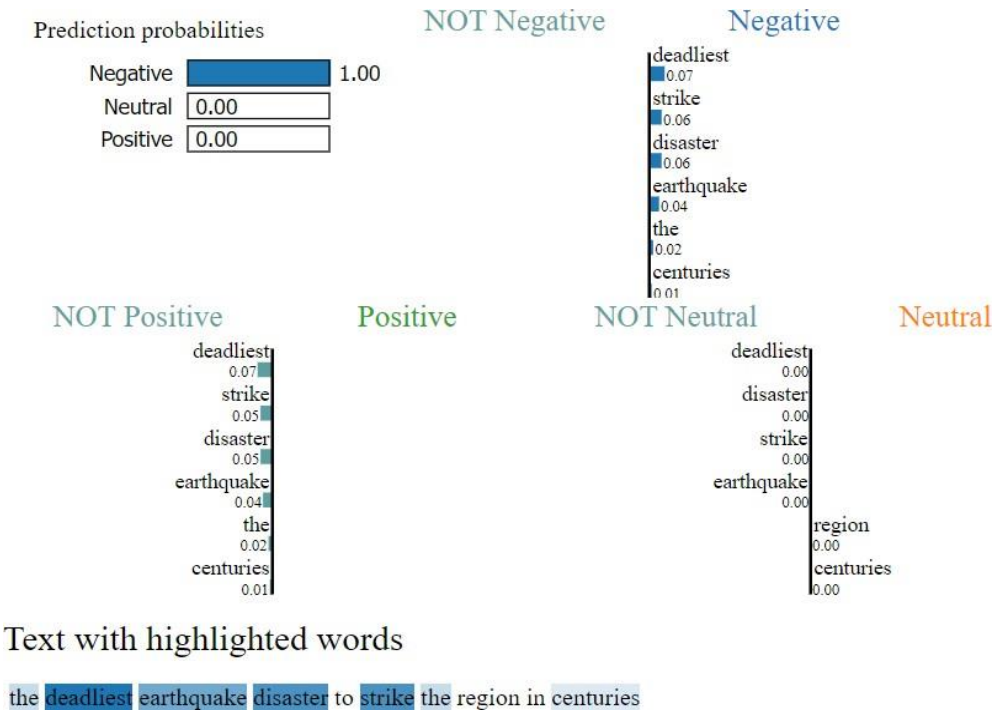


Fig. 11. Model Interpretation Using Explainable AI - LIME

In Figure 12, we consider the sentence *"The devastating earthquakes in Turkey are surely very heartbreaking news."* SHAP calculates the contribution of each word to sentiment classification, with *"heartbreaking"* and *"devastating"* having the highest SHAP values. This quantifies the importance of features in the final classification, with other words like *"earthquakes"* also contributing significantly. By applying both LIME and SHAP, we gained a comprehensive under-standing of the model's behavior. LIME provides local interpretability, while SHAP offers local and global insights into feature contributions, enhancing transparency and enabling more informed insights from the model's predictions.

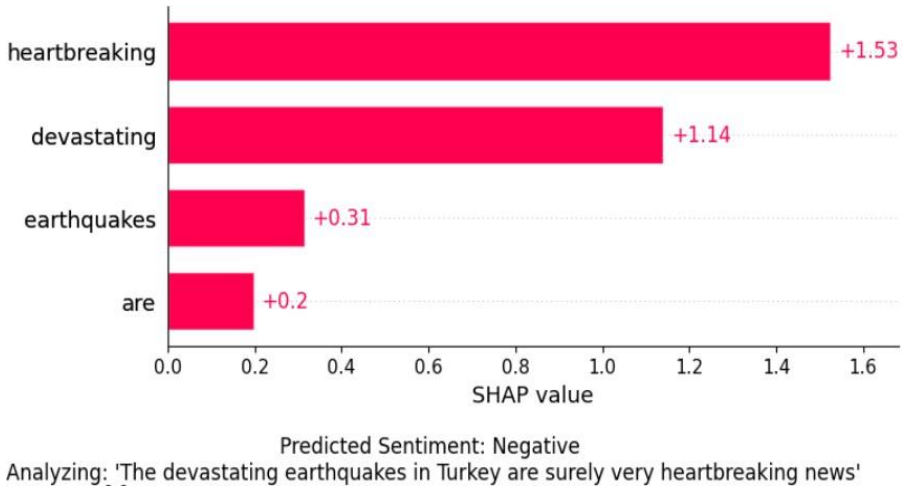


Fig. 12. Model Interpretation Using Explainable AI - SHAP

5. Discussion

In the context of disaster management, effective strategies are essential for building resilience and responding quickly [34, 35]. Furthermore, social media comments can provide real-time insights, enabling authorities to assess public sentiment and identify urgent needs, thus supporting effective decision-making during crises. This study contributes significantly to the field by providing a unique dataset derived from real-time social media comments following the 2023 Turkey earthquake, which captures public sentiment on socio-economic issues, particularly price hikes. This dataset serves as a valuable resource for sentiment analysis and policy-making, especially in disaster management and economic recovery contexts. In our research, we collected and analyzed high-quality, reliable custom-collected data from social media responses after the 2023 Turkey earthquake. Policymakers can significantly benefit from the dataset used in this study, as it provides real-time insights into public sentiment and economic concerns following a disaster. The dataset helps policymakers assess the immediate and long-term socio-economic impact of price hikes, ensuring that economic interventions are tailored to the needs of affected populations. Moreover, the dataset offers a historical reference for understanding how public sentiment evolves over time, helping authorities refine future disaster response strategies. By integrating sentiment analysis into policy planning, governments can develop more adaptive and community-focused recovery programs, ensuring that interventions align with public concerns and expectations. While this study highlights the role of sentiment analysis in understanding public concerns post-disaster, it also demonstrates the efficiency of the dataset in informing disaster recovery strategies. The structured collection and classification of social media comments enable policymakers to identify key socio-economic challenges, such as price surges, economic distress, and economic disparities. By lever- aging these insights, authorities can tailor relief efforts, allocate resources effectively, and implement targeted economic interventions.

By focusing on sentiment analysis and socio-economic impacts, we provide valuable insights into public sentiment regarding price hikes. The use of advanced analytics and explainable AI in our study further reinforces the transparency and accountability of model predictions. This clarity helps stakeholders understand the key factors driving the insights, contributing to a more informed approach to policy-making.

While data-driven policy-making is a complex task requiring extensive analysis and comprehensive data integration, our research demonstrates a small yet significant contribution. By ensuring the effective utilization of our dataset, we provide a strong foundation for future policy decisions that can be more accurate, equitable, and responsive to real-time public needs. Our study represents an important step towards achieving impactful, evidence-based policy-making. For example, by analyzing social media data, we categorized public sentiment into negative, positive, and neutral responses using models like DistilBERT and XLNet. Our findings reveal a significant correlation between sentiment and socio- economic factors, highlighting DistilBERT's efficacy in guiding policy measures to mitigate disaster impacts.

6. Limitations and Future Direction

To this point, we summarize the limitations of our work here. First, relying on social media data may not accurately reflect the views of everyone affected by the disaster, as some people might not have access to social media or may choose not to share their opinions online. This can lead to a bias in the sentiment captured. Additionally, the effectiveness of sentiment analysis depends on the quality of social media posts; the presence of bots, fake accounts, or misinformation can distort the results. While models like DistilBERT and XLNet perform well in classifying sentiment, they may have difficulty understanding the context of certain posts, especially those that use sarcasm, idioms, or local expressions. Although we trained these models on disaster-relevant data, incorporating more diverse and extensive

datasets could further enhance their ability to accurately capture sentiment in disaster-related contexts. The timing of when data is collected can also affect sentiment; for example, immediate reactions may differ from those expressed weeks or months later. Lastly, sentiment analysis tools can sometimes misinterpret the context, leading to incorrect classifications that could influence the overall findings.

While this study offers useful insights, there are several ways future research can improve sentiment analysis in disaster situations. Expanding the dataset to cover a wider range of disasters would help create models that work well across different crises. Additionally, combining data from multiple disaster events could provide a clearer understanding of how public sentiment changes in response to various emergencies, helping policymakers and disaster response teams make better decisions. Including data from multiple social media platforms, news reports, and government sources would also improve accuracy and reduce bias. Another important area for future work is improving model performance by adding more language and context-based features. This could include techniques like topic modeling, sarcasm detection, and multimodal learning, which combines text analysis with images and audio from disaster-related content. Finally, developing transformer models specifically trained for disaster-related sentiment could further enhance accuracy and provide deeper insights into public reactions, economic concerns, and policy-making. Addressing these areas would make sentiment analysis a more powerful tool for disaster response and economic planning.

Acknowledgment

We sincerely thank Sumaiya Hoque for her invaluable assistance in formulating data annotation guidelines, which greatly enhanced the quality of our dataset.

We also extend our gratitude to Tanvir Rahman, Nahid Hasan, Nazrul Islam, Tanver Hossain Refat, and Shanta Maria Shithil for their voluntary contributions to the annotation process, ensuring consistency and accuracy.

Their collective efforts significantly contributed to the success of this research.

References

[1] M. A. Sit, C. Koylu, and I. Demir, "Identifying disaster-related tweets and their semantic, spatial and temporal context using deep learning, natural language processing and spatial analysis: a case study of Hurricane Irma," *Social Sensing and Big Data Computing for Disaster Management*, pp. 8–32, 2020.

[2] L. Hagen, Ö . Uzuner, C. Kotfila, T. M. Harrison, and D. Lamanna, "Understanding citizens' direct policy sugges- tions to the federal government: A natural language processing and topic modeling approach," *2015 48th Hawaii International Conference on System Sciences*, pp. 2134–2143, 2015.

[3] I. Lauriola, A. Lavelli, and F. Aiolli, "An introduction to deep learning in natural language processing: Models, techniques, and tools," *Neurocomputing*, vol. 470, pp. 443–456, 2022.

[4] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, and oth- ers, "Transformers: State-of-the-art natural language processing," *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations*, pp. 38–45, 2020.

[5] O ¨. Ag˘rali, H. So¨ku¨n, and E. Karaarslan, "Twitter data analysis: Izmir earthquake case," *Journal of Emerging Com- puter Technologies*, vol. 2, no. 2, pp. 36–41, 2022.

[6] M. M. Hossain, M. S. Amin, F. Khairunnasa, and S. T. H. Rizvi, "Tweet Trends and Emotional Reactions: A Comprehensive Sentiment Analysis of Twitter Responses to the Earthquake in Turkey and Syria," 2023.

[7] F. M. J. M. Shamrat, S. Chakraborty, M. M. Imran, J. N. Muna, M. M. Billah, P. Das, O. M. Rahman, et al., "Sentiment analysis on twitter tweets about COVID-19 vaccines using NLP and supervised KNN classification algorithm," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 23, no. 1, pp. 463–470, 2021.

[8] S. S. I. Ismail, R. F. Mansour, R. M. Abd El-Aziz, and A. I. Taloba, "Efficient E-Mail Spam Detection Strategy Using Genetic Decision Tree Processing with NLP Features," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, pp. 7710005, 2022.

[9] S. Ghosal and A. Jain, "Depression and suicide risk detection on social media using fasttext embedding and xgboost classifier," *Procedia Computer Science*, vol. 218, pp. 1631–1639, 2023.

[10] A. Kumar, J. M. Chatterjee, V. G. D´ıaz, et al., "A novel hybrid approach of SVM combined with NLP and proba- bilistic neural network for email phishing," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 1, pp. 486, 2020.

[11] J. Antony Vijay, H. Anwar Basha, and J. Arun Nehru, "A dynamic approach for detecting the fake news using random forest classifier and NLP," *Computational Methods and Data Engineering: Proceedings of ICMDE 2020, Volume 2*, pp. 331–341, 2020.

[12] P. M. Lavanya and E. Sasikala, "Deep learning techniques on text classification using Natural language process- ing (NLP) in social healthcare network: A comprehensive survey," *2021 3rd international conference on signal processing and communication (ICPSC)*, pp. 603–609, 2021.

[13] S. Chakraborty, M. B. U. Talukdar, M. Y. M. Adib, S. Mitra, and M. G. R. Alam, "LSTM-ANN based price hike sentiment analysis from Bangla social media comments," *2022 25th International Conference on Computer and Information Technology (ICCIT)*, pp. 733–738, 2022.

[14] H. Saputra, R. Rahmaddeni, and F. Fazri, "Comparison of Machine Learning Algorithms in Analyzing Public Opin- ion Sentiments Against Fuel Price Increases," *CESS (Journal of Computer Engineering, System and Science)*, vol. 8, no. 1, pp. 138, 2023, doi:10.24114/cess.v8i1.41911.

[15] Z. Kastrati, A. S. Imran, S. M. Daudpota, M. A. Memon, and M. Kastrati, "Soaring energy prices: Understanding public engagement on Twitter using sentiment analysis and topic modeling with transformers," *IEEE Access*, vol. 11, pp. 26541–26553, 2023.
[16] A. F. Adoma, N.-M. Henry, and W. Chen, "Comparative analyses of bert, roberta, distilbert, and xlnet for text- based emotion recognition," *2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, pp. 117–121, 2020.
[17] V. Vimbi, N. Shaffi, and M. Mahmud, "Interpreting artificial intelligence models: a systematic review on the appli- cation of LIME and SHAP in Alzheimer's disease detection," *Brain Informatics*, vol. 11, no. 1, pp. 10, 2024.
[18] A. Ceron and F. Negri, "Public policy and social media: How sentiment analysis can support policy-makers across the policy cycle," *Rivista Italiana di Politiche Pubbliche*, vol. 10, no. 3, pp. 309–338, 2015.
[19] T. Papadakis, I. T. Christou, C. Ipektsidis, J. Soldatos, and A. Amicone, "Explainable and transparent artificial intelligence for public policymaking," *Data & Policy*, vol. 6, pp. e10, 2024.
[20] A. Evgenidis, M. Hamano, and W. N. Vermeulen, "Economic consequences of follow-up disasters: lessons from the 2011 Great East Japan Earthquake," *Energy Economics*, vol. 104, pp. 105559, 2021.
[21] S. Akar, "Natural Disasters and Syrian Refugees: The Case of Kahramanmaras¸ Earthquakes in Tu¨rkiye," Inte- chOpen, 2024.
[22] T. Wang, J. Chen, Y. Zhou, X. Wang, X. Lin, X. Wang, and Q. Shang, "Preliminary investigation of building damage in Hatay under February 6, 2023 Turkey earthquakes," *Earthquake Engineering and Engineering Vibration*, vol. 22, no. 4, pp. 853–866, 2023.
[23] W. Chung and D. Zeng, "Dissecting emotion and user influence in social media communities: An interaction mod- eling approach," *Information & Management*, vol. 57, no. 1, pp. 103108, 2020.
[24] M. L. McHugh, "Interrater reliability: the kappa statistic," *Biochemia Medica*, vol. 22, no. 3, pp. 276–282, 2012.
[25] S. Kannan, S. Gurusamy, and E. R. Sowmya, "Preprocessing techniques for text mining," *International Journal of Computer Science & Communication Networks*, vol. 5, no. 1, pp. 7–16, 2014.
[26] J. Plisson, N. Lavrac, and D. Mladenic, "A rule based approach to word lemmatization," in *Proceedings of IS*, vol. 3, 2004, pp. 1-8.
[27] K. W. Church, "Word2Vec," *Natural Language Engineering*, vol. 23, no. 1, pp. 155-162, 2017.
[28] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of NAACL-HLT 2019*, 2019.
[29] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and F-score, with implication for evaluation," in *Proceedings of the European Conference on Information Retrieval*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2005.
[30] Hinojosa Lee, Maria Cristina, Johan Braet, and Johan Springael. "Performance Metrics for Multilabel Emotion Classification: Comparing Micro, Macro, and Weighted F1-Scores." *Applied Sciences*, vol. 14, no. 21, 2024, p. 9863.
[31] Pembury Smith, Matilda QR, and Graeme D. Ruxton, "Effective use of the McNemar test," *Behavioral Ecology and Sociobiology*, vol. 74, pp. 1–9, 2020.
[32] Dieber, Ju¨rgen, and Sabrina Kirrane, "Why model why? Assessing the strengths and limitations of LIME," *arXiv preprint arXiv:2012.00093*, 2020.
[33] Roshan, Khushnaseeb, and Aasim Zafar, "Utilizing XAI technique to improve autoencoder based model for com- puter network anomaly detection with shapley additive explanation (SHAP)," *arXiv preprint arXiv:2112.08442*, 2021.
[34] L. Cai and Y. Zhu, "The challenges of data quality and data quality assessment in the big data era," *Data Science Journal*, vol. 14, p. 2, 2015.
[35] P. S. Das, "Good governance strategies for disaster management and risk reduction," in *International Handbook of Disaster Research*, Springer Nature Singapore, 2023, pp. 1–16.

Authors' Profiles

Muhammed Yaseen Morshed Adib was born in Chittagong, Bangladesh, on December 25, 1995. He earned a Bachelor of Science degree in computer science and engineering from Ahsanullah University of Science and Technology, Dhaka, Bangladesh, in 2018, and a Master of Science degree in computer science and engineering from BRAC University, Dhaka, Bangladesh, in 2024. His major field of study is computer science, with a focus on natural language processing, machine learning, deep learning, and sentiment analysis.

He is currently a Lecturer in the Department of Computer Science and Engineering at Southeast University, located at 252, Tejgaon Industrial Area, Dhaka - 1208, Bangladesh. He has previously worked as a full-time faculty member at Stamford University Bangladesh. His recent publications include BiLSTM-ANN Based Employee Job Satisfaction Analysis from Glassdoor Data Using Web Scraping (Elsevier, 2023), LSTM-ANN Based Price Hike Sentiment Analysis from Bangla Social Media Comments (IEEE, 2023), and CNN-GRU Based Fu- sion Architecture For Bengali License Plate Recognition With Explainable AI (IEEE, 2024).

Md. Tauhid Bin Iqbal was born on December 31, 1991, in Mymensingh, Bangladesh. He received the Bachelor's degree in information technology from the University of Dhaka, Bangladesh, in 2012, and the Ph.D. degree in computer science and engineering from Kyung Hee University, South Korea, in 2019.

He conducted postdoctoral research at the Machine Learning and Visual Computing (MLVC) Laboratory, Kyung Hee University, South Korea. He is currently an Assistant Professor with the Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh. His research interests include deep biometric analysis, medical AI, explainable AI, and image processing.

Farig Yousuf Sadeque was born on January 5, 1990, in Dhaka, Bangladesh. He received his Bachelor's degree in computer science and engineering from the Bangladesh University of Engineering and Technology, Dhaka, Bangladesh, and his Ph.D. degree in information science with a minor in computer science from The University of Arizona, Tucson, Arizona, USA, under the supervision of Dr. Steven Bethard.

He is currently an Associate Professor with the Department of Computer Science and Engineering at BRAC University, Dhaka, Bangladesh. Before joining BRAC University, he was an Associate Research Scientist at the Educational Testing Service (ETS) and a postdoctoral research fellow at Harvard University, where he was jointly appointed by the Harvard Medical School and Boston Children's Hospital, working within the Computational Health Informatics Program (CHIP) Laboratory. His research interests include computational linguistics, applied natural language processing, machine learning, computational social science, psycholinguistics, and so- ciolinguistics.

How to cite this paper: Muhammed Yaseen Morshed Adib, Md. Tauhid Bin Iqbal, Farig Yousuf Sadeque, "Impact of 2023 Turkey Earthquake Price Hikes: Insightful Socio-Economic Analysis Using Transformer Models and Explainable AI", International Journal of Information Engineering and Electronic Business(IJIEEB), Vol.17, No.5, pp. 62-79, 2025. DOI:10.5815/ijieeb.2025.05.05