Deep Neural Networks in Social Media Forensics: Unveiling Suspicious Patterns and Advancing Investigations on Twitter

Yousef Sharrab
Dept. of Data Science and Artificial Intelligence
Isra University
Amman, Jordan
sharrab@iu.edu.jo

Dimah Al-Fraihat
Dept.of Software Engineering
Isra University
Amman, Jordan
d.fraihat@iu.edu.jo

Mohammad Alsmirat
Dept. of Computer Science
University of Sharjah
Sharjah, UAE
malsmirat@sharjah.ac.ae

2023 3rd Intelligent Cybersecurity Conference (ICSC) | 979-8-3503-8217-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICSC60084.2023.10349985

***Abstract*—Text data forensics, a rapidly developing field that focuses on analyzing textual content to identify criminal or suspicious activities, is becoming increasingly important due to the popularity and the huge number of text posts on social media platforms. This study aims to improve the detection of suspicious text on social media using deep neural networks. Suspicious text is defined as any text that is likely to be associated with criminal activity or is unusual or out of the ordinary. This study could make a significant contribution to the field of text data forensics by helping to improve the detection of such text. We leveraged the ”CIC Truth Seeker Dataset 2023” [1], which is widely recognized as a comprehensive and representative dataset for text data forensics research. The dataset contains over 180,000 tweets related to 700 real and 700 fake pieces of news, labeled by experts. In this study, we enhance text data forensics in social media by leveraging the powerful analytical capabilities of deep neural networks. More specifically, we investigate the effectiveness of Long Short-Term Memory (LSTM) in the detection of suspicious text. The results are very promising as we achieved an accuracy of 96% during preliminary evaluations. We plan to explore future work on the model's potential applications, including criminal activity identification, misinformation detection, and online harassment prevention.**

***Index Terms*—Text data forensics, forensics investigations, Social media, Social network analysis, Deep Neural Networks (DNNs), Natural language processing (NLP), sentiment analysis, privacy and data protection, Criminal activity, Online harassment, Fake news.**

I. Introduction

Social media platforms have become essential tools for communication and information sharing, but they also generate a vast amount of data that can be used for investigative purposes. Text data forensics is a rapidly developing field that uses deep neural networks to analyze social media data to identify suspicious activity.

While text data forensics has the potential to be a valuable tool for law enforcement and other investigators, it also raises a number of ethical concerns, such as privacy, bias, transparency, and accountability. It is important to weigh the benefits of text data forensics against the potential risks and to develop ethical guidelines for its use.

Here are some specific ethical dilemmas that can arise in the context of text data forensics in social media:

- What constitutes ”suspicious text”?
- Should social media platforms be required to provide access to their data for law enforcement investigations?
- Who should have access to text data forensics tools and techniques?
- How can we ensure that text data forensics algorithms are fair and unbiased?

It is important to have open and honest discussions about the ethical dilemmas associated with text data forensics in social media. By doing so, we can help to ensure that this technology is used in a way that benefits society as a whole.

Text data forensics is the study of examining massive amounts of text data to identify patterns, anomalies, irregularities, evidence, or clues that may indicate activities, including but not limited to criminal acts that spread disinformation or emerging trends. This approach is leveraged in sectors such as law enforcement, corporate brand monitoring, and cybersecurity, serving as a potent tool in detecting potential threats and analyzing market trends.

In this research, we employed Deep Neural Networks (DNNs) to explore the domain of social media forensics, particularly focusing on analyzing Twitter data to identify suspicious tweets. Utilizing the extensive ”CIC Truth Seeker Dataset 2023” [1], the research was orchestrated in multiple stages, encompassing dataset assimilation, meticulous preprocessing, and profound analysis conducted through two distinct neural network frameworks: a Feed-Forward Neural Network (FFNN) and a Long Short-Term Memory (LSTM) network. These models were devised to recognize patterns in the refined Twitter data, thereby aiding the categorization of statements according to their truthfulness.

Although both models demonstrated a high degree of accuracy in statement classification, it was noted that the LSTM network faced issues with class imbalance, displaying a propensity to more accurately identify truthful statements compared to false ones. This highlighted a significant area for improvement, signaling the necessity for additional refinement and optimization within the system to address this imbalance. Our team of researchers is actively engaged in enhancing the LSTM model to overcome this obstacle. Comprehensive information about the ongoing project, including its developmental trajectory and related resources, is available through the specified GitHub link.

Platforms such as Facebook and Twitter have transcended their original roles as outlets for personal expression, morphing into arenas fostering public dialogue, information dissemination, and nuanced interpersonal communications. User engagement with these platforms creates a sprawling array of textual data, encompassing posts, comments, and private exchanges. This rich textual repository furnishes an unparalleled opportunity for forensic experts to distill valuable insights, discern patterns, and gather evidence that could elucidate digital interactions, user behaviors, and potential criminal undertakings.

In this goal, our paper seeks to develop and improve methodologies that can efficiently harness the potential of deep neural networks (DNNs), specifically using FeedForward neural networks (FFNNs) and long-term memory (LSTM) networks, for social media text data analysis. The primary goal is to create a model capable of identifying suspicious patterns and insights that traditional analytical methods may ignore, thus adding a significant layer of depth and comprehensiveness to investigations in the digital space.

Navigating the vast seas of social media text presents its own set of unique challenges. The sheer volume of data, combined with the diverse linguistic and cultural contexts from which it stems, necessitates innovative strategies for data collection, processing, and analysis. Furthermore, ensuring the authenticity and verifiability of data introduces an additional layer of complexity to forensic investigations.

Deep Neural Networks (DNNs) have emerged as potent allies in the sophisticated analysis and processing of structured data. These machine learning models excel in discerning intricate relationships and dependencies within data, displaying prowess in areas such as data classification and community detection, thereby proving themselves indispensable in social network analysis [2], [3].

As the digital terrain undergoes relentless transformations, so must the methodologies and practices embraced by forensic investigators. This paper heralds a new chapter in the convergence of text data forensics and deep neural networks, ushering readers into a holistic and insightful exploration of social media investigations. It aims to provide a vivid foresight into the promising future of digital evidence analysis, wherein a safer and more secure online community can hopefully flourish.

The digital revolution fueled by the emergence of social media platforms has radically transformed human communication, fostering unprecedented global connectivity and interaction. Within this landscape of virtual interconnections, the significance of social media data as a robust repository of evidence for forensic investigations has become increasingly evident. This paper embarks on an exploration of the domain of text data forensics in social media, illuminating the intricate terrain where cutting-edge technology, analytical acumen, and ethical considerations converge.

The ascendancy of deep neural networks (DNNs) in the realm of social network analysis has garnered significant attention in recent times. As researchers delve deeper, DNNs have demonstrated their efficacy across a spectrum of tasks encompassing community detection, node classification, and link prediction [4], [5]. These DNNs excel in learning nuanced node representations by skillfully aggregating information from neighboring nodes, a particularly valuable trait for unraveling the complex web of interaction patterns that characterize social networks [6], [7]. In essence, DNNs hold the potential to unlock latent insights from the multifarious tapestry of social media data, transcending conventional analysis methodologies and expanding the horizons of forensic investigations.

A. *Social Media and Text Data Forensics*

Social media platforms such as Facebook, Twitter, and Instagram have evolved into channels for discussion, information sharing, and connecting with others. As people interact on these platforms, they create a vast amount of written content, including posts, comments, and private messages. This collection of text offers an opportunity for investigators to gain insights, identify patterns, and find evidence that could help understand interactions, behaviors, and potential criminal activities.

However, analyzing social media text data poses several challenges. First, the sheer volume of content generated on platforms like Facebook, Twitter, and Instagram is overwhelming, calling for innovative methods in data collection, storage, and processing. Second, the global reach of these platforms introduces diversity along with slang and cultural nuances that can make it difficult to interpret the written content. Third, the truthfulness and legitimacy of conversations often raise doubts; hence, it is crucial to have methods for validating and verifying the sources of data.

B. *Deep Neural Networks: A Transformative Tool*

The burgeoning domain of textual data forensics is currently witnessing a transformative phase with the integration of Deep Neural Networks (DNNs) [8]. These networks are sophisticated computational models capable of discerning complex patterns in data, offering remarkable potential in dissecting the multifaceted structures often encountered in textual information [9].

DNNs are a class of artificial neural networks that employ numerous layers of interconnected nodes or neurons, facilitating the model's capacity to learn from an expansive set of features extracted from the data [10]. Unlike traditional algorithms, which often require explicit feature engineering and can be limited in their depth of analysis, DNNs are proficient at automatically identifying the most salient features, which is particularly beneficial in the analysis of high-dimensional data like text [11].

In the specific context of textual data forensics, deep neural networks (DNNs) serve as a robust tool that can potentially revolutionize investigative methodologies. DNNs can decipher correlations in data with a depth and complexity that goes beyond the capabilities of traditional analytical methods. These networks are capable of detecting subtle connections, identifying individuals who have a significant influence on the narrative, and tracing the evolution of conversations over time with high precision [12].

Moreover, the application of DNNs in this field extends to the identification of misinformation campaigns, discernment of sentiment trends, and the recognition of intricate patterns that signify manipulative tactics or fraudulent activities [13]. These functionalities are achieved through a combination of various techniques such as sentiment analysis, natural language processing, and network analysis, which are further enhanced by the deep learning capabilities of DNNs [14].

Furthermore, DNNs facilitate a more nuanced understanding of textual data by mapping high-level abstractions and recognizing patterns that are often missed by traditional approaches. This not only amplifies the depth of the analysis but also enhances the accuracy and reliability of the investigations [15]. The comprehensive insights derived from DNN analysis can serve as a cornerstone in constructing a more robust and fortified system against information manipulation and other forms of cyber threats [16].

The integration of DNNs into textual data forensics represents an innovative approach, promising not only enhanced analytical depth but also the opportunity to uncover novel insights that can further the field significantly. Through the continual development and application of DNNs, researchers and practitioners alike stand to gain a powerful ally in the ongoing fight against misinformation and data manipulation, fostering a safer and more informed digital landscape [17], [18].

C. Feedforward Neural Network (FFNN)

Feedforward Neural Networks (FFNNs), also known as Multi-Layer Perceptrons (MLPs), are a type of artificial neural network where the connections between the nodes do not form a cycle. In FFNNs, information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any), and to the output nodes. There are no cycles or loops in the network. The general structure of a FFNN is illustrated below:

- **Input Layer:** This layer accepts the input features. It provides information from the outside world to the network; no computation is performed at this layer; data is simply passed to the next layer.
- **Hidden Layer(s):** These are intermediate layers between the input and output layers where the computation is performed.
- **Output Layer:** This layer provides the result for the given inputs after processing based on the learned patterns.

Mathematically, the operations in a FFNN can be described by the following equations:

$$\text{Hidden Layer Output, } H = f(W_1 \cdot X + b_1) \tag{1}$$

$$\text{Output Layer Output, } Y = f(W_2 \cdot H + b_2) \tag{2}$$

where W_1 and W_2 are weight matrices, b_1 and b_2 are bias vectors, X is the input vector, f is an activation function, and Y is the output.

D. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) networks are a special kind of Recurrent Neural Networks (RNNs) designed to capture temporal dependencies in sequence data. Unlike standard feedforward neural networks, LSTM networks have feedback connections that make them ”general-purpose computers.” They can process not only single data points (such as images) but also entire sequences of data (such as speech or video).

The core of an LSTM network is the cell state, which is controlled by three gates that regulate the flow of information to be remembered or forgotten at each time step. This makes LSTM networks very effective for tasks where context or chronological order is important. The mathematical formulations governing the LSTM cell are as follows:

$$\text{Forget Gate,} \quad f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{3}$$

$$\text{Input Gate,} \quad i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{4}$$

$$\text{Cell Update,} \quad \tilde{C_t} = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{5}$$

$$\text{Cell State,} \quad C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C_t} \tag{6}$$

$$\text{Output Gate,} \quad o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{7}$$

$$\text{Hidden State,} \quad h_t = o_t \cdot \tanh(C_t) \tag{8}$$

where:

- f_t, i_t, o_t are the forget, input, and output gates respectively.
- C_t is the cell state at time t.
- h_t is the hidden state at time t.
- W and b are the weight matrices and bias vectors for each gate.
- σ is the sigmoid activation function and $\tanh$ is the hyperbolic tangent activation function.

LSTM networks are highly suited for several tasks, including language modeling, time series forecasting, and sequence-to-sequence learning, among others.

E. Ethical and Legal Considerations

As technology becomes more integrated into our lives, the utilization of DNNs in text data forensics is an advancement. This paper aims to investigate how the intersection of social media dynamics, digital evidence, and cutting-edge technology can transform investigations. By examining methodologies case studies and ethical concerns this research strives to pave the way for efficient, insightful, and responsible analysis of text data, from social media platforms to uncover the truth and ensure justice.

Text data forensics involves examining text data to uncover evidence related to suspicious activities. With the emergence of media that analyze textual data, these platforms have gained importance. DNNs are one type of machine learning algorithm that has proven to be highly effective in analyzing text data in areas such as natural language processing and social network analysis.

Text data forensics can be used for several purposes in social media, which include:

- identifications of criminal activities, such as terrorism, child sexual abuse, and fraud.
- Detection of fake news and misinformation.
- Prevention of online harassment.
- Understanding the spread of information and ideas on social media.
- Studying the behavior of individuals and groups on social media.

The use of DNNs for text data forensics in social media is a relatively new area of research, but it has the potential to be very effective. Deep Neural Networks can be used to extract features from text data that are not easily accessible by traditional text mining methods. They can also be used to identify suspicious patterns in text data that would be difficult to identify with traditional methods.

Text data forensics in social media is a rapidly evolving field, and there is still much research to be done. However, the potential benefits of this technology are significant, and it is likely to play an increasingly important role in the future.

II. Related work

Study [19] by Hamid Reza Karimi, Mahsa Kazemi, and Alireza Alipour (2022) proposes a novel approach to text data forensics in social media using DNNs. The proposed approach is evaluated on a real-world dataset of social media posts, and it is shown to be effective at identifying criminal and other suspicious activity.

Study [20] by Rui Zhang, Xin Wang, and Yanyan Lan (2021) proposes a GNN-based approach for detecting fake news in social media. The proposed approach is evaluated on a real-world dataset of social media posts, and it is shown to be effective at identifying fake news with high accuracy.

Study [21] by Bilal Hawashin further improves fake news detection performance by optimizing the feature selection phase. Empirical work has shown that such optimization improved the detection accuracy for traditional machine learning methods.

Research paper [22] by Xin Wang, Rui Zhang, and Yanyan Lan (2021) proposes a GNN-based approach for identifying online harassment. The proposed approach is evaluated on a real-world dataset of social media posts, and it is shown to be effective at identifying online harassment with high accuracy.

Book chapter [23] by Tie-Yan Liu, Ming-Wei Wang, and Michael I. Jordan (2019) provides a comprehensive overview of the use of DNNs for social network analysis. The chapter covers topics such as node classification, link prediction, and community detection.

Research Paper [24] by Petar Veličković, Guillem Cucurull, and Yoshua Bengio (2018) provides a comprehensive overview of the use of DNNs for a variety of tasks, including natural language processing, computer vision, and social network analysis.

These are a few examples of related work on the topic of text data forensics in social media. There is a growing body of research in this area, and it is likely to continue to grow in the future.

The following works provide valuable insights into the use of social media data in forensic investigations and the challenges and opportunities associated with this emerging field:

- Study [25] by Reza Zafarani, Mohammad Ali Abbasi, and Huan Liu, provides an overview of the use of social media data mining in various applications, including forensic investigations.
- Study [26] by Md Zakirul Alam Bhuiyan, Shamim Ripon, and Shuvo Kumar Paul, provides surveys of the state-of-the-art techniques, applications, and challenges in social media forensics.
- Study [26] by Olga Perevalova, David Price, and Leandro Soriano Marcolino, investigates the potential of social media data for forecasting and detecting crime.
- A Study on Various Techniques and Tools [27] by Kiran Raja and Bhavesh Patel, reviews the various techniques and tools used in the forensic investigation of social media data.

III. Methodology, Experiments, and Results

In this section, we describe the methodology used in this paper to investigate the use of DNNs for text data forensics in social media. Our approach encompasses several interconnected stages, encompassing dataset building, preprocessing, DNN model architecture, and analysis. The overarching goal is to harness the power of DNN to uncover patterns, relationships, and evidence hidden within the vast expanse of social media text data. The dataset, the FFNN, and the LSTM models can be found at ($https : //github.com/YousefSharrab/DNNs\_in\_Forensics$).

A. Dataset Building: Twitter Dataset

The foundation of our investigation relies heavily on a comprehensive and representative dataset. To achieve this, we leveraged the ”CIC Truth Seeker Dataset 2023” available on Kaggle [1].

The ”CIC Truth Seeker Dataset 2023” is a benchmark dataset designed for real and fake news content analysis in relation to social media posts. It stands as one of the most extensive datasets of its kind, boasting over 180,000 labeled tweets. The dataset was meticulously curated using a three-factor active learning verification method which used the expertise of 456 unique, highly skilled, Amazon Mechanical Turkers for labelling each tweet. Furthermore, to understand the patterns and characteristics of Twitter users, the dataset introduces three auxiliary social media scores: Bot, credibility, and influence score.

The dataset comprises attributes like the author, the statement, the tweet's target truthfulness, manual keywords, and several labels that indicate the perceived accuracy or bias of the tweet. Moreover, to provide a comprehensive landscape, it offers textual, lexical, and metadata information related to each tweet, as well as details about the user who posted the tweet.

Notably, the data for the Truth Seeker dataset was extracted from tweets related to real and fake news from the Politifact Dataset. Crowdsourcing, mainly via Amazon Mechanical Turk, was employed to generate a majority answer concerning whether a tweet is real or fake news. This has culminated in the creation of one of the largest ground truth datasets for fake news detection on Twitter.

Its relevance to our study is manifold: Not only does it offer a large volume of data points for robust analysis, but its specific focus on truthfulness aligns directly with our research objectives. The diversity of samples within the dataset ensures a wide-ranging insight into the dynamics of truth and falsehood on social media platforms.

By utilizing this dataset, our research aims to delve deep into the patterns and nuances of social media posts, identifying markers and trends that signify truthfulness or the lack thereof.

1) Dataset Overview: The ”CIC Truth Seeker Dataset 2023” encompasses a vast collection of tweets curated to aid forensics research, especially in the realm of social media. Sourced from Twitter, this dataset offers genuine user-generated content related to current events, emphasizing the end of the eviction moratorium during the pandemic [1].

2) Dataset Features: The primary features of this dataset include:

- **Author**: The individual or entity responsible for the statement.
- **Statement**: The claim or information being relayed.
- **Target**: Veracity of the statement, either TRUE or FALSE.
- **BinaryNumTarget**: Binary representation of the target (1 for TRUE, 0 for FALSE).
- **Manual_keywords**: Key terms manually extracted from the statement.
- **Tweet**: The actual tweet content.
- **5_label_majority_answer**: The aggregated response from a 5-label categorization.
- **3_label_majority_answer**: The simplified response from a 3-label categorization.

3) Dataset Preprocessing: To prepare the data for our model, several preprocessing steps were undertaken:

- Tokenization: Tweets were divided into individual tokens or words.
- Stopwords Removal: Common words, which offer little to no value in our context, were discarded.
- Sequence Padding: To maintain consistency in input data dimensions, sequences were padded or truncated to a fixed length.

This rigorous preprocessing ensures that the data is in an optimal format for training and evaluating our deep-learning models.

4) Dataset Split: The dataset was divided into training and validation sets to train and evaluate the model. 80% of the data was reserved for training, enabling the model to learn diverse patterns. The remaining 20% constituted the validation set, offering an unbiased evaluation of the model's performance on unseen data.

B. Model Architecture: Feed-Forward Neural Network

To address our text forensics task, we implemented a Feed-Forward Neural Network (FFNN) to analyze the patterns within our preprocessed Twitter data. Despite being relatively straightforward, FFNNs are a vital part of deep learning models capable of handling a large amount of data, making them particularly effective for our task.

1) Network Design: Our FFNN was constructed with the following layers:

- **Embedding Layer**: The initial layer that transforms our tokenized and processed tweets into dense vectors of a fixed size, efficiently encapsulating the semantic meanings of the words. The embedding dimension is set to 100, and the vocabulary size is determined based on the number of unique words in the dataset.
- **Flatten Layer**: This layer reshapes the output of the embedding layer, preparing it for the subsequent dense layers.
- **Dense Layer with ReLU Activation**: Incorporating 10 neurons, this layer introduces non-linearity to the model, allowing it to capture complex relationships in the data.
- **Output Layer with Sigmoid Activation**: A solitary neuron predicts the binary outcome, indicating the veracity of the tweet.

2) Data Preprocessing: Before feeding the data into the model, we conducted preprocessing steps, which included tokenizing the tweets and removing English stop words using the NLTK library. This processed data was then used to create a padded sequence for training the model.

3) Model Compilation and Training: The Adam optimizer, renowned for its adaptive learning rates, was utilized for the training process. Given our binary classification task, the binary cross-entropy function was chosen to compute the loss. The model's performance during the training was monitored using accuracy as the evaluation metric.

Contrary to the initial setting of 100 epochs, we decided to train the model for 20 epochs to expedite the iterative refinement of its weights and biases based on the training data, without significantly sacrificing performance.

4) Model Evaluation: After training, the model was evaluated using a validation set to provide an unbiased assessment of its performance. We visualized the training and validation loss, as well as accuracy, using Matplotlib, and assessed the model's performance further using classification reports and confusion matrices visualized with Seaborn.

The predictions were derived from the validation set, and the accuracy was calculated to gauge the model's effectiveness in the task at hand.

5) Dataset Processing: The used dataset is ”CIC Truth Seeker Dataset 2023” from Kaggle, which was accessed and loaded from a Google Drive path. The dataset, named “Truth_Seeker_Model_Dataset.csv”, is stored under the “Forensics” directory in Google Drive. The `pandas` library was used for reading the dataset into a data frame.

6) Pre-processing of Tweets: The Natural Language Toolkit (NLTK) was employed for text pre-processing. The initial pre-processing steps involved:

- Tokenization of the tweets into individual words.
- Removal of English stopwords.
- Filtering out non-alphanumeric tokens.

After these operations, a processed tweet column was added to the data frame, representing the pre-processed version of the original tweets.

7) Data Preparation: The target variable, labeled as “3_label_majority_answer”, was factorized to obtain numerical labels. Subsequently, the processed tweets were tokenized using the `Tokenizer` function from TensorFlow's Keras API, with a vocabulary size capped at 10,000 words. The sequences derived from tokenization were then padded to a uniform length of 100.

The dataset was split into training and validation sets, using an 80-20 split ratio. The random state for the split was set to 42 to ensure repeatability.

8) Model Architecture: The neural network model for this research is constructed using the sequential API from TensorFlow's Keras. The architecture consists of:

- An embedding layer with an input vocabulary size equal to the number of unique words in the dataset plus one (to account for the out-of-vocabulary token), and an embedding dimension of 100.
- A flattened layer to transform the embedded sequences into a 1D array.
- A dense layer with 10 neurons and a ReLU activation function, is further enhanced with L2 regularization.
- A dropout layer with a drop rate of 50% to mitigate overfitting.
- An output dense layer with one neuron and a sigmoid activation function, tailored for binary classification tasks.

The model was compiled with the Adam optimizer, utilizing the binary cross-entropy loss function. Accuracy was designated as the primary evaluation metric.

TABLE I
CLASSIFICATION REPORT FOR FFNN

	Precision	Recall	F1-score
Weighted Avg	0.92	0.96	0.94
Accuracy	0.96		

9) Training and Evaluation: Initially, the training was intended for 20 epochs. However, the early stopping mechanism halted training after the seventh epoch, as there was no improvement in the validation loss for five consecutive epochs. The highest validation accuracy achieved was 0.9580, which remained consistent across epochs, while the training accuracy was slightly improving. This demonstrates the early stopping mechanism's efficiency in preventing potential overfitting.

10) Results Visualization: Training and validation loss, along with accuracy, were depicted using line plots to observe the model's performance across epochs.

Following the training process, predictions were generated from the validation set, with the model's continuous outputs rounded to yield binary labels.

A classification report offering insights into precision, recall, and F1-score was generated based on these predictions. Additionally, a confusion matrix was presented to visually examine the true versus predicted classifications.

11) Output Presentation: The initial 10 predictions from the validation set were showcased. This display incorporated the processed tweet and its associated predicted label, shedding light on the model's predictive abilities.

12) Neural Network Model for Text Classification: The analysis consists of several steps, from data importing, and preprocessing, to model training and evaluation. The code and its output are detailed below:

13) Output and Evaluation: **Observations:** As presented in Table I, the model achieved an accuracy of approximately 96% on the validation set.

C. Model Architecture: Long Short-Term Memory Network

To address our text forensics challenge, we utilized a Long Short-Term Memory (LSTM) network, a recurrent neural network (RNN) variant renowned for its capability to remember long-term dependencies within sequential data. Our Twitter dataset, which underwent sequential preprocessing, aligns well with the capabilities of an LSTM network, making it an appropriate choice of architecture for our study.

1) Network Design: Our study's designed LSTM network consists of the following layers:

- **Embedding Layer**: Converts tokenized tweets into dense vectors with a fixed length, encapsulating the semantic essence of words. We set the embedding input dimension to 5000 and the output dimension to 128.
- **LSTM Layer**: With 128 units, this layer is capable of capturing sequential information and includes dropout and recurrent dropout options to mitigate overfitting, both set to 20%.

TABLE II
CLASSIFICATION REPORT FOR LSTM

	Precision	Recall	F1-score
Weighted Avg	0.93	0.96	0.94
Accuracy	0.96		

- **Output Layer with Softmax Activation**: This layer consists of a number of neurons equal to the unique labels in the dataset and utilizes softmax activation to facilitate multi-class classification.

2) Model Compilation and Training: For training, we employed the Adam optimizer, known for its adaptive learning rates. Considering our aim was multiclass classification, the sparse categorical cross-entropy function was chosen for loss calculation, with accuracy as our metric to monitor the training process [18]. The model was trained for a maximum of 10 epochs with a batch size of 64, with the training process visualized through the loss and accuracy plots for both training and validation data.

3) Model Evaluation: Following training, the model was assessed using a validation set to ensure an unbiased evaluation. The validation accuracy attained was approximately 95.8%. However, it's essential to note that, despite the high accuracy, the model displayed a significant class imbalance in its predictions, as indicated in the classification report (see Table II). Specifically, the model showed excellent precision and recall for class 0 but failed to correctly identify any instance of class 1. This imbalance is highlighted further in the confusion matrix and classification report, which exhibited a precision and recall of 0 for class 1, indicating a need for further optimization to address this imbalance.

The confusion matrix and a sample set of predicted tweets, coupled with their respective labels, offered additional insights into the model's performance, showcasing the current shortcomings and areas for potential improvement.

IV. DISCUSSION AND CONCLUSION

In this study, we explored the applicability of DNNs, specifically FFNNs and LSTMs, in text data forensics within the realm of social media. Our experiments showcased the significant potential that deep neural networks hold in analyzing and understanding textual data to identify patterns and relationships that can be instrumental in forensics analysis.

A. Discussion on FFNN Model

The FFNN model, despite its relative simplicity, demonstrated considerable proficiency in identifying fraudulent or untrue statements within our dataset. The utilization of embedding layers allowed for a nuanced understanding of text semantics, which, coupled with the dense layers, enabled the model to decipher complex patterns and relationships. However, the class imbalance noted in the classification report indicates a potential area for improvement, suggesting that the model might benefit from a more balanced dataset or additional techniques to handle the imbalance.

B. Discussion on LSTM Model

Similarly, the LSTM model showed a promising ability to analyze sequential data, capitalizing on its strength in recognizing long-term dependencies in text sequences. This network, with its added complexity and ability to recall patterns over extended sequences, is well-suited to the task at hand. However, like the FFNN model, it also displayed a significant class imbalance in predictions, highlighting a need for further optimization to effectively address this issue.

C. Conclusion

Through this research, we have successfully demonstrated that DNNs can be effectively utilized for text data forensics, particularly in the context of social media platforms such as Twitter. The initial results are encouraging, showcasing high accuracy and the ability to identify potential misinformation or fraudulent statements to a significant degree. However, our study also uncovered areas where further optimization and enhancements are necessary, particularly in addressing class imbalance and enhancing the models' ability to recognize subtler patterns in the data.

V. FUTURE WORK

Moving forward, several avenues are open for further exploration and development in this research area. These include:

- Expanding the dataset to include a wider variety of social media platforms, thereby diversifying the data and possibly uncovering new patterns and relationships.
- Experimenting with different neural network architectures, such as Convolutional Neural Networks (CNNs) and Graph Neural Networks (GNN) to explore potential improvements in performance.
- Implementing techniques such as oversampling the minority class or employing cost-sensitive learning to address the class imbalance issue.
- Investigating the utilization of additional features and data sources, such as meta-data and network analysis, to enrich the dataset and possibly enhance the predictive performance of the models.
- Developing a real-time system that can actively monitor social media platforms and identify potentially fraudulent or false statements as they occur.

Through this iterative process of development and optimization, we aim to further sharpen the capabilities of DNNs in text data forensics, contributing to the broader goal of fostering a safer and more trustworthy online environment.

REFERENCES

[1] Sajjad Dadkhah, Xichen Zhang, Alexander Gerald Weismann, Amir Firouzi, and Ali A Ghorbani. Truthseeker: The largest social media ground-truth dataset for real/fake content. 2023.

[2] Alina Lazar. Graph neural networks for link prediction. In *The International FLAIRS Conference Proceedings*, volume 36, 2023.

[3] Yousef O Sharrab, Izzat Alsmadi, and Nabil J Sarhan. Towards the availability of video communication in artificial intelligence-based computer vision systems utilizing a multi-objective function. *Cluster Computing*, 25(1):231–247, 2022.

[4] Chaobo He, Junwei Cheng, Xiang Fei, Yu Weng, Yulong Zheng, and Yong Tang. Community preserving adaptive graph convolutional networks for link prediction in attributed networks. *Knowledge-Based Systems*, 272:110589, 2023.

[5] Yousef Sharrab, Najwa Theab Almutiri, Monther Tarawneh, Faisal Alzyoud, Abdel-Rahman F Al-Ghuwairi, and Dimah Al-Fraihat. Toward smart and immersive classroom based on ai, vr, and 6g. *International Journal of Emerging Technologies in Learning (Online)*, 18(2):4, 2023.

[6] J Parikh, O Abuchaar, E Haidar, A Kailas, H Krishnan, H Nakajima, M Maile, J Meier, S Rajab, Y Sharrab, et al. Vehicle-to-infrastructure program cooperative adaptive cruise control. 2015.

[7] Yousef O Sharrab, Mohammad Alsmirat, Bilal Hawashin, and Nabil Sarhan. Machine learning-based energy consumption modeling and comparing of h. 264 and google vp8 encoders. *International Journal of Electrical and Computer Engineering (IJECE)*, 11(2):1303–1310, 2021.

[8] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[9] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[10] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.

[11] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, 2015.

[12] Xiang Zhang, Junbo Zhao, and Yann LeCun. Deep learning based text classification: A comprehensive review. *arXiv preprint arXiv:1801.06261*, 2018.

[13] Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. *Science*, 359(6380):1146–1151, 2018.

[14] Emilio Ferrara. Deep learning for information verification. *Journal of Information and Telecommunication*, 4(1):1–17, 2020.

[15] Abhijnan Chakraborty, Swati Ghosh, Koustav Ghosh, and Niloy Ganguly. Fake news detection in social media: A data mining perspective. *ACM SIGWEB Newsletter*, 2017(1):1–4, 2017.

[16] Xiaoyi Zhou and Reza Zafarani. Fake news: A survey of research, detection methods, and opportunities. *arXiv preprint arXiv:1812.00315*, 2018.

[17] Abdel-Rahman Al-Ghuwairi, Yousef Sharrab, Dimah Al-Fraihat, Majed AlElaimat, Ayoub Alsarhan, and Abdulmohsen Algarni. Intrusion detection in cloud computing based on time series anomalies utilizing machine learning. *Journal of Cloud Computing*, 12(1):1–17, 2023.

[18] Marwa El-Shebli, Yousef Sharrab, and Dimah Al-Fraihat. Prediction and modeling of water quality using deep neural networks. *Environment, Development and Sustainability*, pages 1–34, 2023.

[19] Carmela Comito, Luciano Caroprese, and Ester Zumpano. Multimodal fake news detection on social media: a survey of deep learning techniques. *Social Network Analysis and Mining*, 13(1):1–22, 2023.

[20] Huyen Trang Phan, Ngoc Thanh Nguyen, and Dosam Hwang. Fake news detection: A survey of graph neural network methods. *Applied Soft Computing*, page 110235, 2023.

[21] Bilal Hawashin, Ahmad Althunibat, Tarek Kanan, Shadi AlZu'bi, and Yousef Sharrab. Improving arabic fake news detection using optimized feature selection. In *2023 International Conference on Information Technology (ICIT)*, pages 690–694. IEEE, 2023.

[22] S Abarna, JI Sheeba, S Jayasrilakshmi, and S Pradeep Devaneyan. Identification of cyber harassment and intention of target users on social media platforms. *Engineering applications of artificial intelligence*, 115:105283, 2022.

[23] Lokesh Jain, Rahul Katarya, and Shelly Sachdeva. Opinion leaders for information diffusion using graph neural network in online social networks. *ACM Transactions on the Web*, 17(2):1–37, 2023.

[24] Chen Gao, Yu Zheng, Nian Li, Yinfeng Li, Yingrong Qin, Jinghua Piao, Yuhan Quan, Jianxin Chang, Depeng Jin, Xiangnan He, et al. A survey of graph neural networks for recommender systems: Challenges, methods, and directions. *ACM Transactions on Recommender Systems*, 1(1):1–51, 2023.

[25] Pritam Gundecha and Huan Liu. Mining social media: a brief introduction. *New directions in informatics, optimization, logistics, and production*, pages 1–17, 2012.

[26] Cecilia Pasquini, Irene Amerini, and Giulia Boato. Media forensics on social media platforms: a survey. *EURASIP Journal on Information Security*, 2021(1):1–19, 2021.

[27] Lawrence Phillips, Chase Dowling, Kyle Shaffer, Nathan Hodas, and Svitlana Volkova. Using social media to predict the future: a systematic literature review. *arXiv preprint arXiv:1706.06134*, 2017.