The Islamia University of Bahawalpur

Final year Report of Project
BS in Artificial Intelligence

Identifying and detecting misinformation spread in Social Networks

Mam Quratulain Qureshi

Lecturer
Supervisor

Muhammad Aoun Cheema

F20BARIN1E02011

Fall 2020-2024

Table of Contents

CERTIFICATE OF APPROVAL

It is to certify that the final year project of BS (AI) “Project title” was developed by STUDENT NAME , Roll No, SESSION under the supervision of “SUPERVISOR NAME” and that in (his/her) opinion; it is fully adequate, in scope and quality for the degree of Bachelors of Science in Artificial Intelligences.

Supervisor

External Examiner

Chairman Department of Artificial Intelligence

Chapter 1

1 Introduction:

With the internet industry development, online social media, like Twitter and Facebook have become important sources of information, providing tremendous convenience to our daily life. However, as online social media have a low threshold, large numbers of users and real-time dissemination of information, rumors in social media can soon trigger a mass effect. to trust rumors and propagate them. The "explosive" propagation of rumors in social media may cause serious harm to society, especially during epidemic outbreaks. The spread of rumors can cause serious Social harm, especially in social networks.

Purpose:

The main objective is to develop a robust software solution that employs advanced algorithms and techniques to identify, track, and analyze evolving rumors and misinformation across various online social media platforms. We will be focused on achieving a rumor detection accuracy of at least 80%, the system utilizes advanced Machine learning as well as Deep learning algorithms to detect rumors. Recognizing the dynamic nature of misinformation, the project emphasizes continuous algorithm updates and incorporates user customization features, ensuring adaptability and providing a personalized experience. Ultimately, the objective is to contribute to a trustworthy and secure digital information landscape, addressing the issues arising from the swift spread of inaccurate information in today's online environment.

Scope:

The project's scope centers on developing a Natural Language Processing (NLP)-centered system tailored to identify rumors and misinformation within the dynamic environment of social networks. Focused on major modules like NLP Project, Data Manager, Language Model, and Evaluation Module, the project emphasizes efficient data handling, robust language model training, and continuous adaptation to evolving linguistic patterns. The scope extends to the responsible use of AI, adhering to ethical considerations. Overall, the project aims to make a positive impact on information integrity by providing a sophisticated and user-centric solution to detect rumors and misinformation.

1.1 Problem statement:

The extensive use of modern social media networks has led to a significant challenge – the rampant spread of rumors. This phenomenon not only jeopardizes cyber security but also poses a substantial threat to social stability. A notable incident occurred on April 23rd, 2013, when a fabricated news report, disseminated through a compromised Twitter account, falsely claimed explosions at the White House and injuries to then-President Barack Obama. Despite swift clarifications, the misinformation caused widespread panic, causing a staggering $136.5 billion loss in the stock market. This event underscores the urgent need for automated systems that can predict and counteract the accuracy of information, particularly at its early stages, to mitigate the far-reaching consequences of false rumors.

The project aims to address this pressing issue by utilizing advanced techniques in Natural Language Processing (NLP) and machine learning. The aim is to create a sophisticated system capable of precisely predicting the trustworthiness of information spreading on social media platforms. By identifying and classifying rumors in their nascent stages, the system seeks to provide a proactive intervention mechanism, offering an opportunity to curtail the widespread dissemination of false information.

Through this endeavor, the project aims to enhance cyber security, safeguard social stability, and minimize the adverse impacts associated with the rapid spread of misinformation in the digital age.

1.2 Objectives:

- Detect the diffusion of evolving rumors in social networks.
- Early rumors Detection in social networks.

1.3 Rumors:

Rumor is typically defined in one of two ways:

1. *One can define rumor as "distorted, exaggerated, irrational and inauthentic information" (Miller 1992), which is a commonly-held view in practice.*

2. *However, in academic research, rumor is typically defined as an "unverified or unconfirmed message".*

For instance, ***Buckner (1965)*** *define rumor as an "unconfirmed message passed from one person to another that refers to an object, person, or situation",*

We adopt," ***Rumors are defined as unauthenticated statements or reports related to common interests and are widely spread by various means of dissemination"*****.**

Propagation of rumors is dependent on the below:
Assume that a spreader propagates a rumor through direct touch with others. According to the characteristics of the propagation of rumors, the total population is divided into four categories:

IDSRI (Ignorance–Discussant–Spreader–Remover–Ignorance) rumor propagation model.

Table 1: Different types of spreaders

Type of spreader	*Explanation of spreader*
Ignorance	*A person who does not know rumors.*
Discussant	*A person who knows rumors but does not spread them, just participates in the discussion.*
Spreader	*A person who knows rumors and spreads them.*
Remover	*A person who knows rumors and no longer spreads them.*

Chapter 2
Background Study

2.1 Background study

"Evaluation of machine learning and deep learning models on fake and real news datasets using various text representation techniques. It introduces a novel stacking model with a random forest meta classifier, achieving high testing accuracies of 99.94% and 96.05% on ISOT and KDnugget datasets, surpassing baseline methods. The study employs a corrected version of McNemar's test to assess and select the best-performing individual model for fake news detection". [1]

"A novel big data and machine learning approach for fake news detection, utilizing the FNC-1 dataset. Employing a Spark distributed cluster, it creates a stacked ensemble model with Random Forest, Logistic Regression, and Decision Tree classifiers, extracting features through N-grams, Hashing TF-IDF, and count vectorizer. The proposed model achieves a superior 92.45% F1-score, outperforming state-of-the-art techniques by 9.35%, while the evaluation includes metrics such as accuracy, precision, recall, and F1-score, comparing against baseline methods". [2]

"This paper introduces a hybrid neural network model (CNN-LSTM) for fake news stance detection, utilizing convolutional neural network (CNN) and long short-term memory (LSTM) layers. Employing dimensionality reduction techniques (PCA and chi-square) to select relevant features, the model achieves 97.8% accuracy and 76.3% F1-score on the Fake News Challenge (FNC-1) dataset, surpassing state-of-the-art methods, while addressing limitations and suggesting future research directions". [3]

"A novel fake news detection framework that can detect fake news in the early phase by exploiting the information from the news articles and the social contexts. The proposed model is based on a Transformer architecture, which has two parts: the encoder part to learn useful representations from the fake news data and the decoder part that predicts the future behavior based on past observations. The authors also incorporate many features from the news content and social contexts into their model to help classify the news better. They propose an effective labeling technique to address the label shortage problem and show that their model can detect fake news with higher accuracy within a few minutes after it propagates than the baselines". [4]

"This paper introduces a fake news detection classifier employing recurrent neural network (RNN) models, including vanilla RNN, GRU, and LSTM, trained on the LIAR benchmark dataset. Comparative analysis involves traditional machine learning methods such as logistic regression, SVM, and CNN. Despite the absence of transformer learning or architecture, the study contributes to understanding the effectiveness of RNN variants in discerning fake news content". [5]

"This paper presents a model for fake news detection in Twitter posts, utilizing machine learning and deep learning algorithms, including Support Vector Machine, Naïve Bayes, Logistic Regression, Recurrent Neural Network, and Long Short-Term Memory. Evaluated on a dataset of 20,360 tweets related to the Chile earthquake 2010, SVM and Naïve Bayes outperform other algorithms with 89.34% and 89.06% accuracy, respectively, using TF-IDF features". [6]

“This paper employs a machine learning approach, utilizing NLTK for text processing on a dataset of 12,165 news articles. It applies Naive Bayes, SVM, and Logistic Regression algorithms, combining their outputs through averaging, and experiments with a one-layer neural network. Achieving an 83% accuracy on the training set with the average hypothesis model, the study discusses challenges, limitations, and proposes future directions for improvement, suggesting the use of n-grams, contextual word tokens, and more robust packages”. [7]

“This paper introduces a rumor detection system for Twitter, utilizing the PHEME dataset and applying machine learning algorithms (Naive Bayes, Decision Tree, SVM, and Random Tree). With feature selection and pruning, it reports that Random Tree and Decision Tree achieve the highest accuracies at 95.21% and 95.71%, respectively”. [8]

“BERT-based deep learning approach called FakeBERT for fake news detection in social media. The model combines different parallel blocks of the single-layer deep Convolutional Neural Network (CNN) having different kernel sizes and filters with the BERT. The proposed model outperforms the existing models with an accuracy of 98.90% 1”. [9]

“Comprehensive comparison of a few transformer architecture-based pre-trained models for text summarization. The authors used the BBC news dataset for analysis and comparison, which contains text data that can be used for summarization and human-generated summaries for evaluating and comparing the summaries generated by machine learning models. The paper also discusses the two methods of generating summaries: extractive summarization and abstractive summarization”. [10]

“It incorporates AANE for user attributed network learning, DeepWalk for news-user topological network reconstruction, and a fusion framework combining CNN-extracted news content features and shallow perceptron-learned news-user network features. Evaluated on BuzzFeed and PolitiFact datasets, the framework achieves state-of-the-art performance without employing transformer learning or architecture, using network embedding methods (AANE, DeepWalk, Node2vec, LINE), binary cross-entropy loss, and evaluation metrics such as precision, recall, and F1”. [11]

“This paper introduces a novel early fake news detection model using a two-level convolutional neural network (TCNN) to extract semantic features from news articles and a user response generator (URG) to generate user responses based on historical data. Evaluated on Weibo and Twitter datasets, the model outperforms baselines, relying on convolutional neural networks and variational autoencoders for feature extraction and response generation”. [12]

“Comprising multi-modal feature extraction, modal-independent prediction, and cross-modal similarity modules, the method utilizes Text-CNN for textual features, a pre-trained image2sentence model, and Text-CNN for visual features. Evaluated on PolitiFact and GossipCop datasets, SAFE outperforms baselines using softmax and cross-entropy for fake news”. [13]

“This paper presents LIAR, a dataset for fake news detection with manually labeled statements from POLITIFACT.COM. Introducing a hybrid convolutional neural network (CNN) architecture integrating text and meta-data features, the study demonstrates superior performance, outperforming baselines with the

hybrid CNNs achieving the best accuracy of 0.274 on the test set". [14]

"This paper examines hyperpartisan and fake news writing styles using a dataset of 1,627 articles fact-checked by BuzzFeed. Introducing a novel assessment method called Unmasking, the study finds effective discrimination of hyperpartisan news and satire from mainstream but notes the inefficacy of style-based fake news detection. Utilizing the BuzzFeed-Webis Fake News Corpus and a dataset, the paper employs Unmasking and a random forest classifier with style and topic features, revealing commonalities in writing style between left-wing and right-wing news". [15]

"The paper uses transfer learning to evaluate how well models trained on two different Twitter datasets (CREDBANK and PHEME) can predict the accuracy of tweets in a third dataset (BuzzFeed). The authors also use a classifier trained on a combination of PHEME and the Internet Argument Corpus to assign disagreement labels to tweets". [16]

"This paper surveys fake news detection on social media, categorizing methods into news content and social context models from a data mining perspective. It addresses challenges in intentional falsehood and dynamic social context, reviews evaluation metrics, datasets, and discusses future directions, employing natural language processing, machine learning, and network analysis techniques". [17]

"The paper introduces TI-CNN, a novel approach for fake news detection that leverages both text and image information. Utilizing convolutional neural networks (CNNs), the model extracts explicit and latent features from a dataset of 20,015 news articles, achieving classification between 11,941 fake and 8,074 real news articles". [19]

"This paper critically assesses the 2017 Fake News Challenge Stage 1 (FNC-1) on stance detection, reproducing top system results, proposing a new evaluation metric, and introducing a cross-domain dataset. Utilizing FNC-1 and a new dataset, the study analyzes models based on MLPs, CNNs, and gradient-boosted decision trees, introducing a feature-rich stacked LSTM model. Evaluation metrics include FNC metric, macro-averaged F1 score, and comparison with a human upper bound". [20]

"The paper introduces a satirical news detection model using five features (Absurdity, Humor, Grammar, Negative Affect, and Punctuation) and a dataset of 360 articles. Employing an SVM classifier, it achieves 90% precision, 84% recall, and 87% F-score, with publication in the Proceedings of NAACL-HLT in 2016". [22]

"A hybrid deep learning model for rumor detection in social networks, leveraging post-wise features from user, content, lexical, and text aspects. Utilizing Twitter data and benchmarked datasets, employing preprocessing techniques, word embedding, PCA, and standard scalar methods, the proposed BiLSTM_UCL model achieves a superior 97% accuracy, surpassing state-of-the-art approaches, as reported in the online publication from March 2022". [23]

"The paper introduces a content-based fake news detection method utilizing knowledge graphs and embedding techniques. Employing TransE and B-TransE models, it represents triples from news articles and knowledge graphs, utilizing bias functions and fusion strategies for classification. The approach utilizes Kaggle's 'Getting Real about Fake News' dataset, incorporates true articles from mainstream media, and leverages DBpedia as an external knowledge graph for fake news detection without explicit

use of transfer learning". [24]

"Automatic fake news detection system using machine learning, deep learning, and natural language processing techniques. Employing data preprocessing and various models, including logistic regression, decision tree, naive Bayes, support vector machine, BERT, and LSTM, it achieves high accuracy. Evaluating on a Kaggle dataset, BERT demonstrates the highest accuracy at 98%, followed by LSTM at 95%, showcasing the effectiveness of transfer learning and deep learning methods in fake news detection, as reported in the December 2022 publication". [25]

"This paper conducts a thorough performance evaluation of eight machine learning algorithms (L1 regularized logistic regression, SVM, Gaussian naive Bayes, multinomial naive Bayes, decision trees, random forests, MLP, and CNN) for fake news detection. Using three datasets, it compares algorithms based on accuracy, F1-measure, training time, and classification time". [26]

"The paper is a comprehensive survey on fake news and rumor detection, examining methodologies covering definitions, data collection, feature extraction, and techniques, while also addressing challenges in data collection. It highlights publicly available datasets like LIAR, CREDBANK, and PHEME, discusses potential future directions including transfer learning, multimodal analysis, and explainable models, without reporting specific accuracy results. Submitted to Elsevier in May 2019, the publication status is unclear". [27]

"This paper provides a survey of methods and systems for detecting fake news, emphasizing concepts and techniques like data collection, preprocessing, vectorization, and classification. It compares Naïve Bayes and Logistic Regression classifiers". [28]

"This paper proposes a hybrid approach, integrating linguistic cues and network metadata for fake news detection using a random forest classifier. Utilizing the FakeNewsCorpus and GRFN datasets, the study implements four sub-pipelines for feature engineering. The model achieves high accuracy and F1-score (99.9% and 99.99% on FakeNewsCorpus, 87.9% on GRFN)". [29]

"Investigates the application of three common models to classify articles as reliable or unreliable based on structured content. Employing the Fake News Corpus (over 9 million articles) and the Fake and Real News Dataset (45,000 labeled articles), the study uses bag of words and term frequency-inverse document frequency for feature extraction. It trains and evaluates multilayer perceptron, random forest, and multinomial naive Bayes models, utilizing metrics like accuracy, precision, recall, and F1-score". [31]

"This paper introduces a deep learning model for Arabic fake news detection, incorporating both news content and user social context. Trained on a self-created Arabic dataset, the study compares CNN and BiLSTM with five word embedding models. Evaluating the model's generalization on a public dataset, the paper highlights MARBERT with CNN as the top-performing model in terms of accuracy and F1-score". [33]

"Leveraging a pre-trained multilingual BERT, the method generates synthetic fake news data to address dataset imbalance, achieving the highest accuracy of 92.45%. Utilizing the BanFakeNews dataset, the study employs a two-stage approach: data augmentation with BERT and embedding/classification with a smaller BERT model. Evaluation metrics include accuracy, precision, recall, and F1-score, demonstrating

superiority over twelve baseline models. The BERT model, based on transformer architecture, plays a pivotal role in data augmentation and fake news detection". [34]

"Utilizing machine learning and deep learning models on two datasets. Employing term frequency, term frequency-inverse document frequency, and embedding techniques, and using a random forest meta-classifier, it achieves an accuracy of 99.94% on the ISOT dataset and 96.05% on the KDnugget dataset, surpassing state-of-the-art methods, as reported in the 2021 publication".[35]

"A Knowledge-driven Multimodal Graph Convolutional Network (KMGCN) for fake news detection, leveraging textual, visual, and knowledge information as graphs. Evaluated on PHEME and WEIBO datasets, the model outperforms state-of-the-art baselines, achieving the best performance with accuracies of 0.8863 on WEIBO and 0.8399 on PHEME, as reported in the 2020 publication". [36]

"A Chinese COVID-19 dataset comprising fact-checked microblogs from Weibo, with 2,104 entries containing ground-truth labels and various information. Evaluating five neural network methods for fake news detection, the best-performing TextCNN achieves a macro F1 score of 0.938, as reported in the 2021 publication in the journal Social Network Analysis and Mining". [37]

"Using LUN and SLN datasets, the study employs a two-layer heterogeneous graph attention network and an entity comparison network for feature representation and semantic consistency. The fused features are fed into a softmax layer for classification, enhancing detection efficiency by considering different writing styles of fake news". [38]

"A deep attention-based recurrent neural network for early rumor detection on social media, utilizing sequential posts to focus on distinctive textual features. Evaluated on Twitter and Weibo datasets, the model outperforms five state-of-the-art methods in terms of precision, recall, and F-measure, demonstrating superior performance in detecting rumors more quickly and accurately, as reported in the 2017 publication". [40]

"Using geometric deep learning, specifically graph convolutional neural networks. Integrating heterogeneous features such as user profile, social network, news propagation, and content, the method learns task-specific patterns". [42]

"Using PHEME 2017 and PHEME 2018 datasets, the model comprises three modules: TRM encodes tweets with a deep BiLSTM, CPM propagates vectors through the conversation graph using GraphSAGE, and CM predicts rumor labels with a softmax function. Results reveal superior performance over baselines, including RDM, on both datasets". [45]

"The evaluation employs transformer models (BERT and SBERT), utilizing information retrieval and semantic similarity models for retrieval and natural language inference models for stance detection. A dataset for COVID-19 misinformation detection on social media used". [46]

"Evaluation on English and Chinese datasets. The framework incorporates a task-conditioned prompt-wise hypernetwork (TPHNet) for bidirectional knowledge transfer and is based on p-tuning v2, utilizing multilayer soft-prompts for natural language understanding and knowledge probing tasks". [47]

"The PHEME dataset is used for evaluation. The model, based on a BERT encoder and classification head,

outperforms static setup baselines, showcasing positive backward transfer. Rehearsal-based continual learning methods, including REPLAY and GEM, are applied in their approach". [48]

"Utilizing self-supervised learning to capture relations between propagation patterns in social networks and semantic patterns in post content. The methods employ graph neural networks (GNNs) to encode propagation trees and convolutional neural networks (CNNs) to encode semantic content. Evaluation on Twitter, Weibo, and PHEME datasets". [49]

"Three text classifiers (LR, BERT, and BERT+) are compared under chronological and random splits, employing BERT and its variants (BERTweet and ERNIE) as pre-trained transformer models for rumor detection. Twitter 15, Twitter 16, PHEME, Weibo datasets used". [50]

Table 2: Summary of background study

Paper No	*Year of publish*	*Datasets*	*Methodology*	*Performance Accuracy*
1	*2021*	*ISOT dataset*	*Logistic regression, SVM, K-NN, Decision tree, Random forest, LSTM*	*99.63% 99.63% 68.65% 99.62% 99.84% 99.74%*
2	*2023*	*FNC-1*	*Hashing, TF_IDF, Essembler, Lr, Bigram, Lr_Glove, SVM_CV,*	*93.45% 88.45% 73.45% 91.75%*
3	*2020*	*benchmark dataset, FNC dataset*	*CNN-LSTM*	*97.8%*
4	*2021*	*Fake News, LAIR,*	*CNN, Deep CNN, Fake BERT, GRU, LSTM, StackedCNN, CNN,*	*98.3 98.36 98.90 47.2 46.8 48.5 27*

		FNC_1, PHEME	RCNN, FDML, CNN+LSTM, Dense Neural Network, CNN, BiLSTM-CNN,	33.7 50.8 97.8 94.31 87.1 86.12
5	2018	LAIR dataset,	SVM Logistic Regression Bi-LSTM LSTM	0.255 0.247 0.233 0.2166
6	2019	Chile earthquake Twitter	SVM, Naïve Bayes, RCNN	89.34% 89.06%
7	2018	Fake news	Naive Bayes, SVM, Logistic Regression,	83.02%
8		Pheme dataset	SVM, Decision Tree, Random Tree, Naïve Bayes	92.14 95.71 95.21 91.42
9	2019	TI-CNN, Fake News Corpus	LSTM, BERT, convolutional layers	0.91 0.98 0.937
10	2020	Dataset of 6460 tweets	Bag of Words, Graph Neural Networks (GNNs)	
11	2019	FakeNewsNet, BuzzFeed, PolitiFact	RST+SVM, DeepWalk, Node2vec, Bi-LSTM	
12	2018	Weibo dataset, Twitter dataset	TCNN, TCNN_URG,	88.08 89.84
13	2020	PolitiFact, GossipCop	VGG-19, LIWC LIWC	64.9% 82.2% 83.6%

			VGG-19	77.5%
14	2017	LIAR	SVM, Logistic Regression, Bi-LSTM	
16	2017	PHEME, CREDBANK	featuring accuracy prediction, aligning the three datasets, selecting which features to use,	
17	2017	BuzzFeedNews, LIAR, BS Detector, CREDBANK	Feature extraction, Model construction	
18	2019	Twitter data, Facebook data, Reddit, 4chan, Sina Weibo	SVMs, CrowdSourcing Platforms, h-index,	
20	2018	FNC-1 dataset,	FNC-1 evaluation methodology	
21	2019	PolitiFact BuzzFeedNews	RST+SVM, TriFN, UCEM, RST+SVM, TriFN, UCEM,	61.06% 86.4% 89.5% 57.1% 87.8% 932%
22	2016	360 news articles one in 2015 and one in 2016	SVM,	
23	2022	Real-world, Benchmarked dataset	SVM, UCL_PCA, BiLSTM_Embed	
24	2018	Fake news article base (FAB), The true news article base (TAB)	Knowledge Stream, TransE FML	
25	2022	open-source fake news dataset	Logistic Regression, Naive Bayes, Decision Tree, SVM, LSTM, BERT	74 74 90 77 95 98
26	2019	LIAR, The Signal Media	logistic regression, C-support vector	

		One-Million News Articles Dataset, Getting Real about Fake News	*classification, Gaussian naive Bayes, Multinomial naive Bayes, Random forests, Multi-layer perceptron, Convolutional neural networks (CNNs)*	
27	*2019*	*LIAR, CREDBANK, PHEME, FakeNewsNet*	*Data collection, Feature extraction, Detection techniques, Future directions*	
28	*2019*	*fake_or_real_news*	*logistic regression, Naïve Bayes,*	
29	*2018*	*Fake News Corpus, Getting Real about Fake News (GRFN) dataset*	*Machine learning pipeline, Random Forest algorithm,*	*99.9% 87.9%*
30	*2018*			
31	*2023*	*Liar, Liar Pants on Fire, Fake and real news dataset, BuzzFeed-Webis Fake News, Fake News or Truth, Fake News Corpus*	*Multi-Layer Perceptron, Random Forest (RF) Classification, Multinomial Naïve Bayes (MNB)*	*27.40% 94% 90.19% 75% 82%*
32	*2018*	*LIAR*	*Multi-source Multi-class Fake news Detection framework (MMFD), CNN, LSTM, Multi-class Discriminative Function (MDF)*	*29.06 34.77 38.81*
33	*2023*	*800 Tweets, 177 Tweets, 1862 Tweets ArCOV19-Rumors,*	*Tweet Collection, Tweets Preprocessing, Features Extraction, Proposed Model Architecture BiLSTM, MARBERT-CNN CNN-RNN, MARBERT-CNN*	*0.899 0.76 0.8000 0.8630*

		COVID-19 misinformation		*0.7000* *0.8707*
34	*2022*	*BanFakeNews,* *Augmented Dataset*	*CNN,* *LSTM,* *Bi-LSTM,* *CNN-LSTM,* *M-BERT,* *AugFake-BERT*	*83.34* *84.69* *85.10* *88.83* *88.69* *92.45*
35	*2021*	*ISOT dataset,* *KDnugget dataset*	*SVM,* *k-NN,* *CNN,* *GRU,* *LSTM*	*99.94%* *96.05 %*
36	*2020*	*PHEME,* *WEIBO*	*SVM-TS,* *CNN,* *EANN,* *GRU,* *TextGCN,* *KMGCN* *SVM-TS,* *CNN,* *EANN,* *GRU,* *TextGCN,* *KMGCN*	*0.6312 0.7112* *0.7212* *0.7927* *0.8571* *0.8863* *0.6399 0.7007 0.7177* *0.8282* *0.8374* *0.8756*
37	*2020*	*CHECKED dataset*	*Data collection,* *Data analysis,* *Fake news detection*	
38	*2023*	*LUN,* *SLN*	*BERT,* *BERT+LSTM,* *GCN+attn*	
39	*2019*	*MisInfoText,* *Buzzfeed*	*Text classification,* *Data collection,* *neural network models*	
40	*2018*	*Twitter,* *Sina Weibo*	*DT-Rank,* *LK-RBF,* *SVM-TS,* *CERT,* *ML-GRU*	
41	*2019*	*Sina Weibo,*	*TF-IDF,* *GRU-2,* *CAMI,* *CED(0.875),* *CED(0.975),*	*0.859* *0.920* *0.896* *0.938* *0.946*

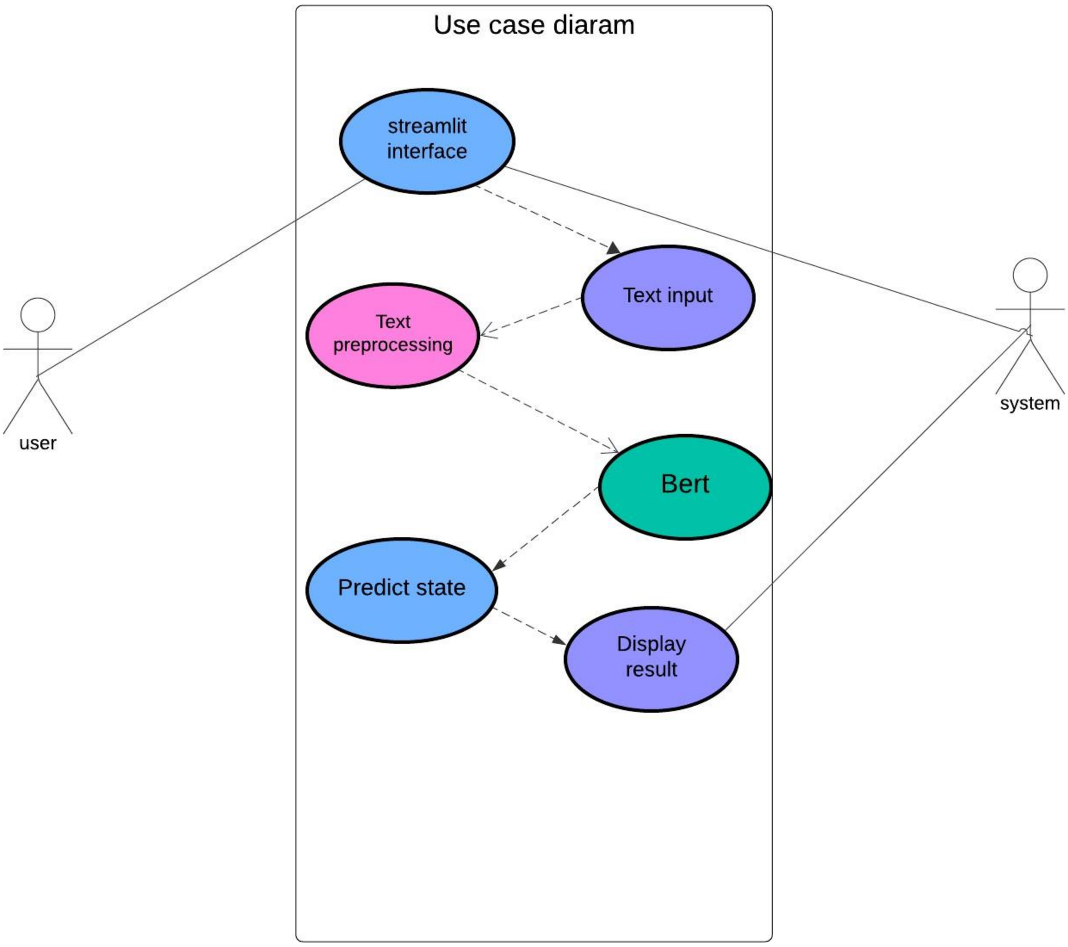
		 Twitter	*CED-CNN(0.875),* *CED-CNN(0.975)* *TF-IDF,* *GRU-2,* *CAMI,* *CED(0.875),* *CED(0.975),* *CED-CNN(0.875),* *CED-CNN(0.975)*	*0.900* *0.912* *0.587* *0.672* *0.595* *0.717* *0.744* *0.721* *0.704*
42	*2019*	*Snopes, PolitiFact,* *Buzzfeed*	*Data collection,* *Geometric deep learning,* *Model evaluation*	 *92.7% ROC AUC*
43	*2019*	*Weibo,* *Twitter15,* *Twitter16*	*GLAN,* *w/o LRE,* *w/o GRE* *GLAN,* *w/o LRE,* *w/o GRE* *GLAN,* *w/o LRE,* *w/o GRE*	*94.6* *86.8* *88.7* *90.5* *82.7* *83.8* *90.2* *86.4* *87.5*
44	*2014*	*Twitter dataset*	*Binary logistic Regression*	
45	*2020*	*PHEME 2017,* *PHEME 2018*	*GNN-U,* *GNN-M,* *GNN-MU,* *GNN-S,* *GNN-SB,* *TRM* *GNN-U,* *GNN-M,* *GNN-MU,* *GNN-S,* *GNN-SB,* *TRM*	*0.606* *0.630* *0.622* *0.737* *0.816,* *0.900* *0.581* *0.588* *0.646* *0.731* *0.798* *0.919*
46	*2020*	*COVIDLIES*	*Data collection,* *Annotation process.* *Evaluation metrics*	

47	2022	PHEME + Twitter15&16 + Covid19, Weibo	CPT-RD, TPHNet, simInit	
48	2021	PHEME dataset,	BERT-BL, M2-BL, REPLAY, REPLAY+TTOKENS, GEM, GEM+TTOKENS	33.8% 44.3% 70.2% 75.7% 62.0% 68.5%
49	2023	Twitter, Weibo, PHEME	SRD-PSCD SRD-PSID SRD-PSCD SRD-PSID SRD-PSCD SRD-PSID	0.837 0.857 0.949 0.962 0.839 0.838
50	2023	PHEME, Weibo	logistic regression, BERT, BERT+	
51	2022	Pheme,, Weibo	DDGCN DDGCN	0.855 0.948
52	2022	BuzzFeed, PolitiFact	DTCN,	93.670 95.280
53	2023	Chinese dataset (DatasetCN), English dataset (DatasetEN)	CNFRD, BERT + BiLSTM,	88.92% 87.07%
54	2018	Weibo dataset Twitter Dataset	HAS-BLSTAM HAS-BLSTAM	0.943 0.844
55	2021	Weibo,	GRU, TEXT-CNN, HAN, RvNN, RvNN, EANN, GAN-GRU,	0.839 0.807 0.833 0.903 0.847 0.866 0.867

			GLAN, *CGAT*	*0.902* *0.940*
		Twitter	*GRU,* *TEXT-CNN,* *HAN,* *RvNN,* *RvNN,* *EANN,* *GAN-GRU,* *GLAN,* *CGAT*	*0.852* *0.839* *0.851* *0.789* *0.824* *0.794* *0.783* *0.853* *0.892*

Chapter 3
Requirement Analysis

3.1 Use Case Diagram



3.2 Functional Requirements

Text Preprocessing:

System must process the following:

- Convert text to lower case
- Tokenize the text
- Remove punctuation and special characters.
- Get English stopwords
- Remove stopwords
- Reducing words to their base or root form.

Text Representation:

- Bag of Words
- TF-IDF

Model Building:

- SVM
- Naïve Bayes
- Random Forest
- Gradient Boosting
- Bert

Model Evaluation:

- Accuracy
- F1 score

3.3 Non-Functional Requirements:

- **Performance**
  The model should be fast and responsive, able to process text and provide results timely according to the accuracy of trained model.

- **Reliability**
  The model must be reliable all available all the time.

- **Usability**
  The model should be user friendly abstracting all the complex details and only giving them the understandable procedure.

- **Scalability**
  The model must be scalable, able to handle large amount of rumors without compromising on performance and reliability.

- **Maintainability**
  The model should be easy to maintain and update, with well-organized manner that should be easily understandable to others.

3.4 Operating environments:

- **Google Colab**
  The application requires a powerful computer system with sufficient processing power, memory and storage to process large volume of data. in order to meet these requirement we will use Google colab

- **Software dependencies**
  - Pandas
  - NumPy
  - Seaborn
  - matplotlib
  - Scikit-learn
  - NLTK
  - Genism
  - torch
  - transformers
  - hugging face datasets
  - transformers[torch] accelerate -U

- **Internet Connectivity**
  A reliable internet connection is required. The minimum speed should be 15 to 20 mbps.

Chapter 4

Experimentation and Results

4.1 Data acquisition

The dataset we have used is given as follows:

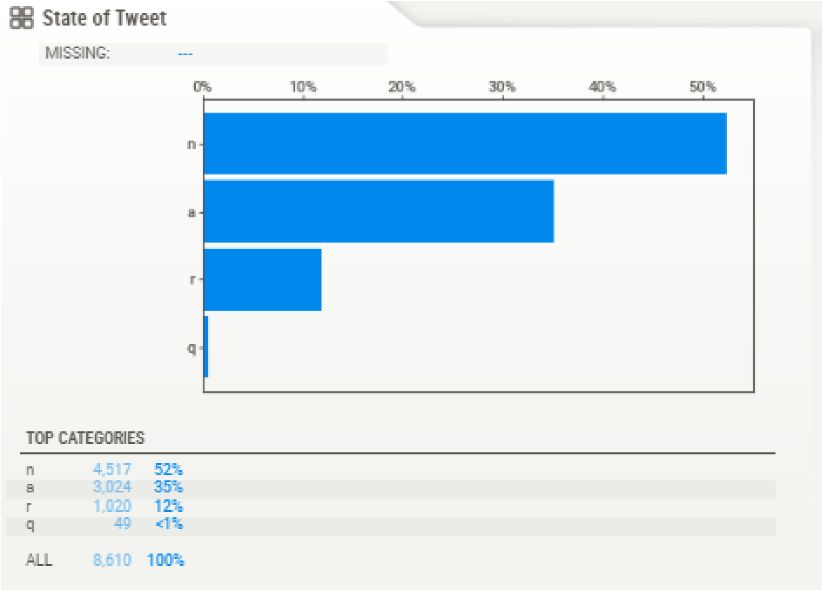
Dataset link:
https://www.kaggle.com/datasets/muhammadmrsaleen/rumor-and-non-rumor-dataset

Dataset Details

#	Column	Data type	Description
0	ID	Int	Unique identifier for each social media post
1	Description	Obj	Text content of the social media post.
2	#Tweets	Int	Number of tweets by the user.
3	Date_time_creation_account	Obj	Date and time when the user's account was created
4	Language	Obj	Language of the post.
5	#Followers	Int	Number of followers of the user.
6	#Friends	Int	Number of friends or accounts the user is following.
7	Date&Time	Obj	Date and time of the post.
8	#Favorite	Int	Number of favorites or likes for the post.
9	#Retweet	Int	Number of retweets for the post.
10	Another Tweet Inside	Int	Indicates if there's another tweet inside this one.
11	Source	Obj	Source device or platform used to post the content.
12	Tweet ID	Int	Unique identifier for the tweet.
13	Retweet ID	Obj	ID of the retweet, if applicable.
14	Quote ID	Obj	ID of the quoted tweet, if applicable.
15	Reply ID	Float	ID of the reply tweet, if applicable.
16	Frequency of tweet occurrences	Obj	Frequency of occurrences for the tweet.
17	State of Tweet	obj	Indicates the nature of the tweet

4.2 State of tweet

r	**The tweet/retweet is a rumor post**
a	**The tweet/retweet is an anti-rumor post**
q	**The tweet/retweet is a question about the rumor, however neither confirm nor deny it**
n	**The tweet/retweet is not related to the rumor (even though it contains the queries related to the rumor, but does not refer to the rumor)**



4.3 Text preprocessing techniques used:

TFIDF:

Model	Accuracy	F1 Score
SVM	0.82	0.77
Naïve Bayes	0.82	0.75
Random forest	0.83	0.77
Gradient Boosting	0.79	0.77

Bag of words:

Model	Accuracy	F1 Score
SVM	0.83	0.79
Naïve Bayes	0.66	0.70
Random forest	0.85	0.80
Gradient Boosting	0.83	0.80

Chapter 5
Model selection and evaluation:

The experimentations is done on the following machine learning model:

- SVM
- Naïve Bayes
- Gradient boosting
- Random forest
- Bert

Support Vector Machine (SVM):
Support Vector Machine (SVM) is a supervised learning algorithm that can be used for classification or regression challenges. It works by finding the hyper plane that best divides a dataset into classes, aiming to maximize **the margin between the data points of different classes.**

Naive Bayes:
Naive Bayes is a probabilistic classifier based on Bayes' Theorem, assuming strong independence between features. Despite its simplicity and the naive assumption of feature independence, it performs remarkably well on a variety of tasks, particularly those involving text classification.

Gradient Boosting:
Gradient Boosting is an ensemble technique that builds a model in a stage-wise fashion from weak learners, typically decision trees. It optimizes for a loss function by iteratively adding models to correct errors made by the previous ones, enhancing predictive accuracy and robustness.

Random Forest:
Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes (classification) or mean prediction (regression) of individual trees. This technique improves predictive performance and reduces over fitting by averaging the results of numerous trees.

BERT (Bidirectional Encoder Representations from Transformers):
BERT is a pre-trained transformer model developed by Google that has revolutionized natural language processing tasks. By considering the context from both directions (left-to-right and right-to-left) in all layers, BERT achieves state-of-the-art results in various NLP benchmarks, making it highly effective for tasks such as text classification, question answering, and sentiment analysis.

- **Accuracy Using TFIDF**

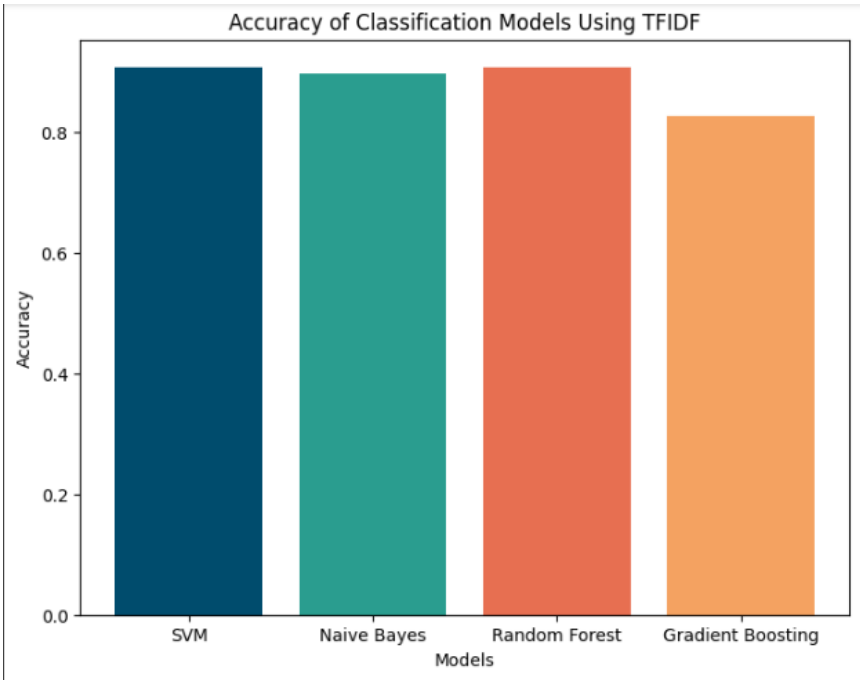


Figure 5.1

- **F1 score using TFIDF:**

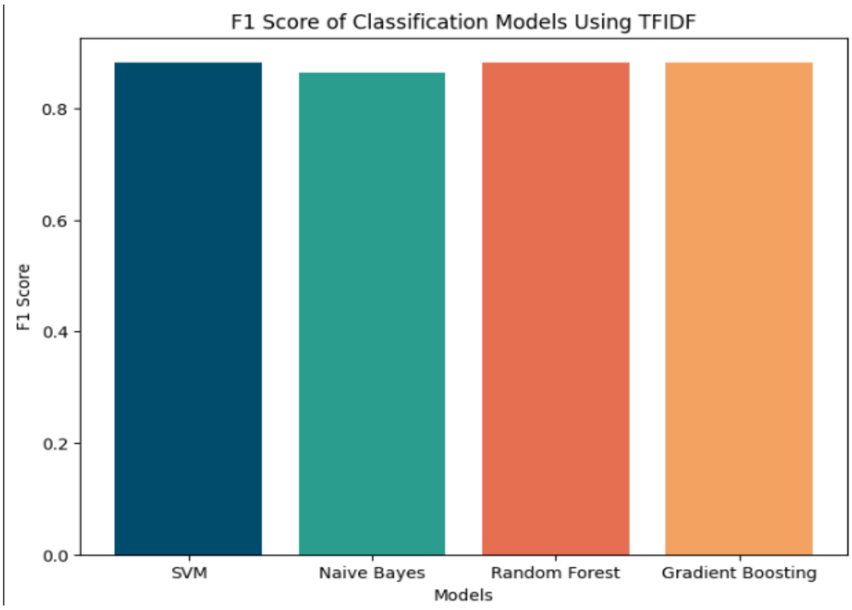


Figure 5.2

- **Accuracy Using Bag of words:**

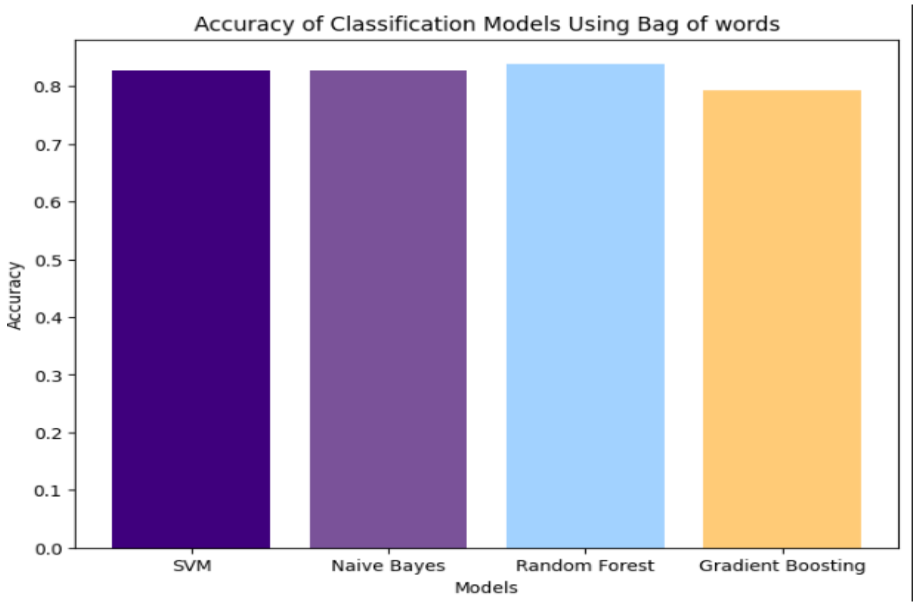


Figure 5.3

- **F1 score using Bag of words:**

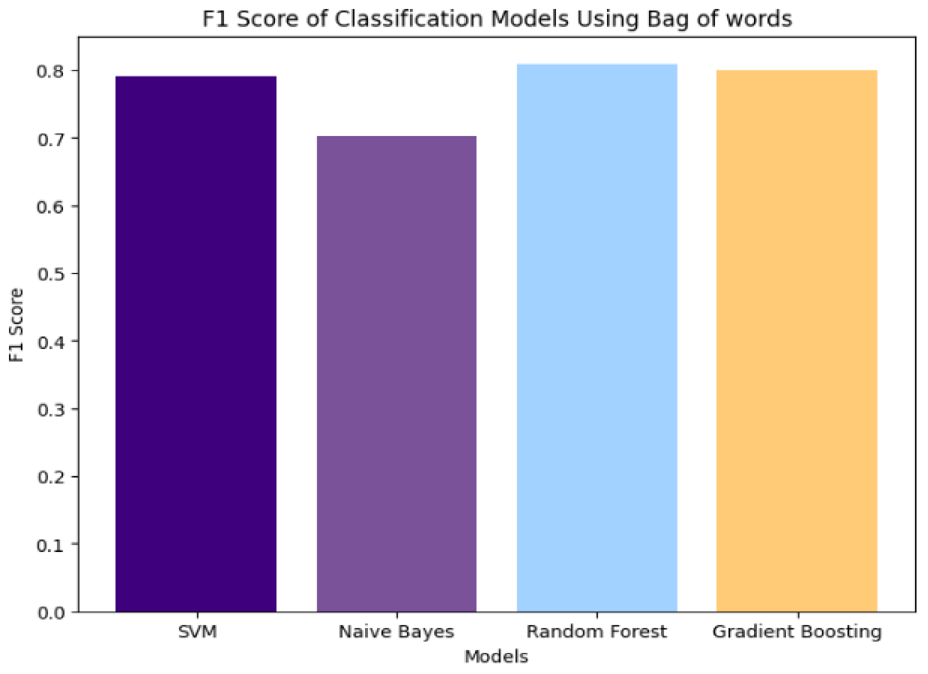


Figure 5.4

Bert Model:

Epoch	Training Loss	Validation loss	Accuracy	F1 Score
1	0.60	0.44	0.89	0.84
2	0.49	0.44	0.89	0.84
3	0.39	0.47	0.90	0.88

Loss over epochs:

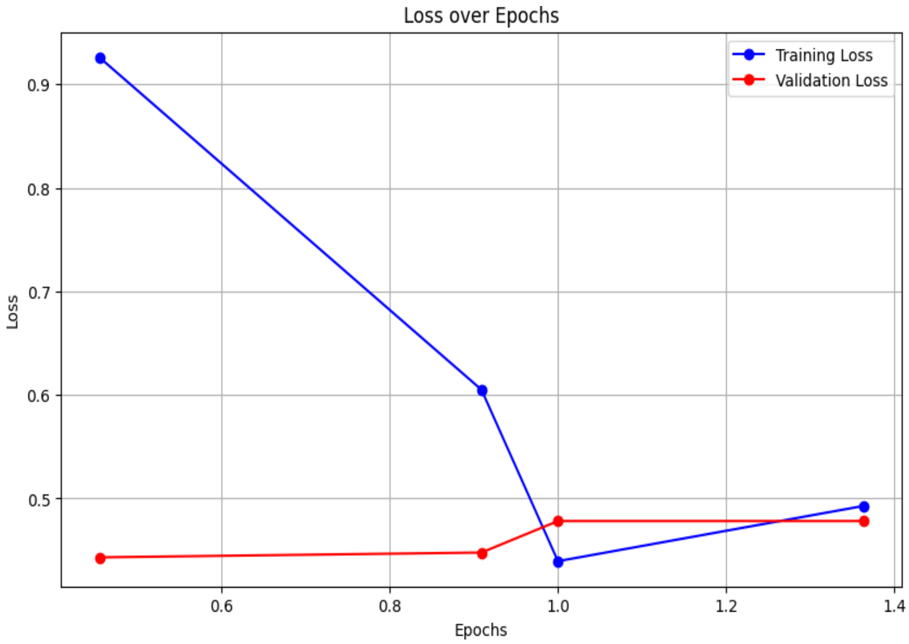
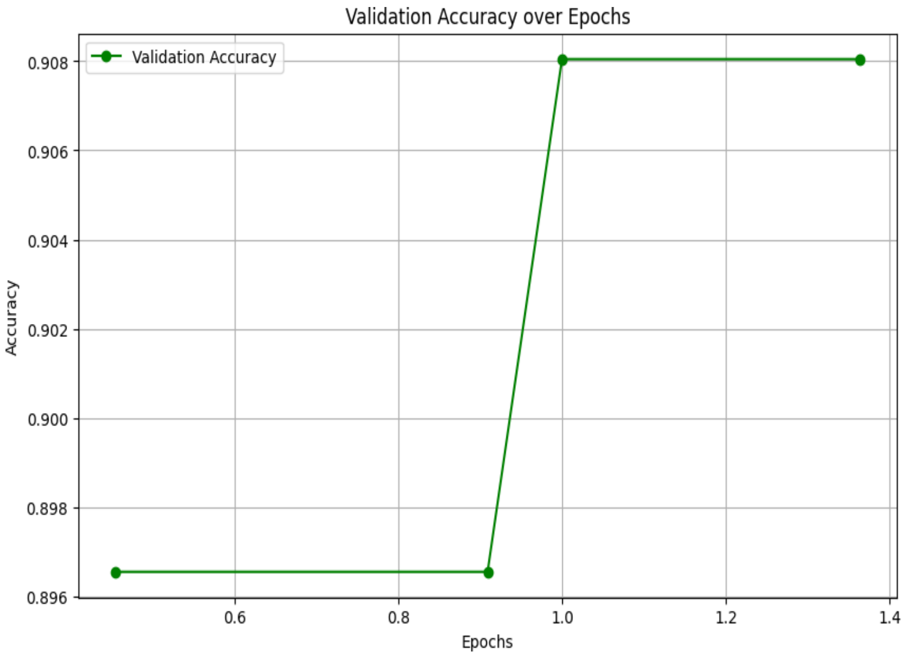


Figure 5.5

Accuracy Graph:



F1 Score:

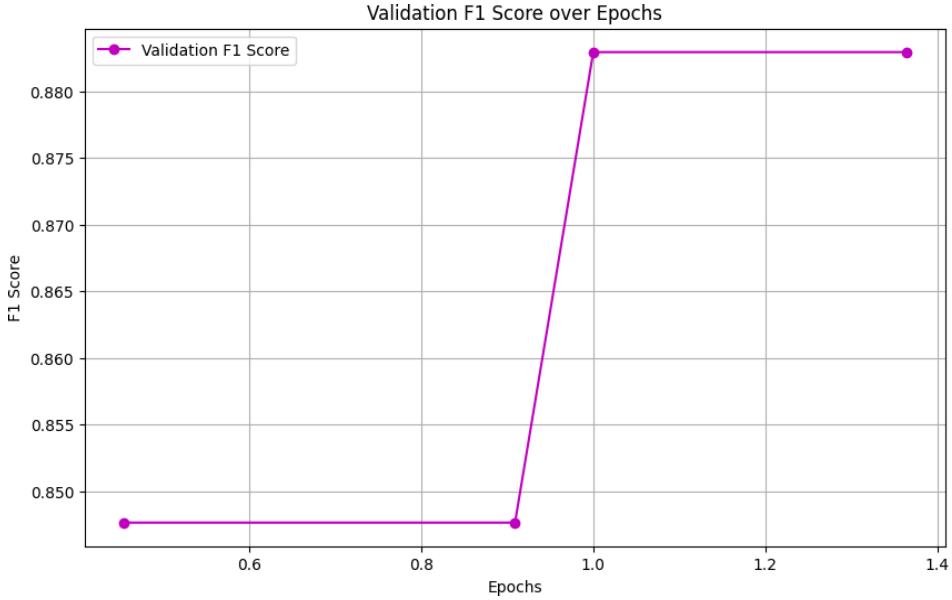


Figure 5.6

6. Future Work:

As of new, this is the current class distribution which clearly indicates the state “n” is dominating the dataset used to train (fine-tune) the model.
n “the tweet not related to rumors even though it contains the queries related to the rumors but does not defers to the rumor”.

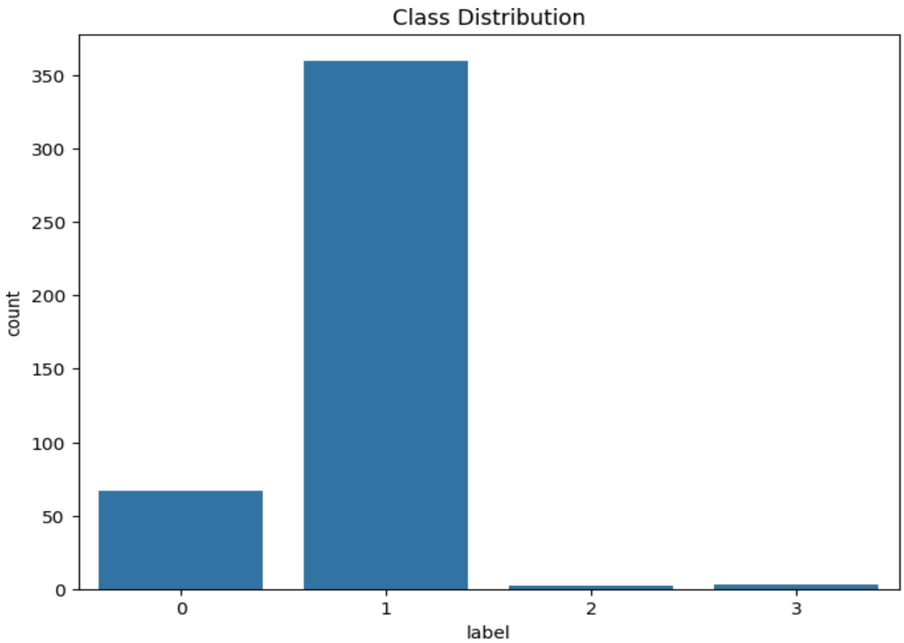
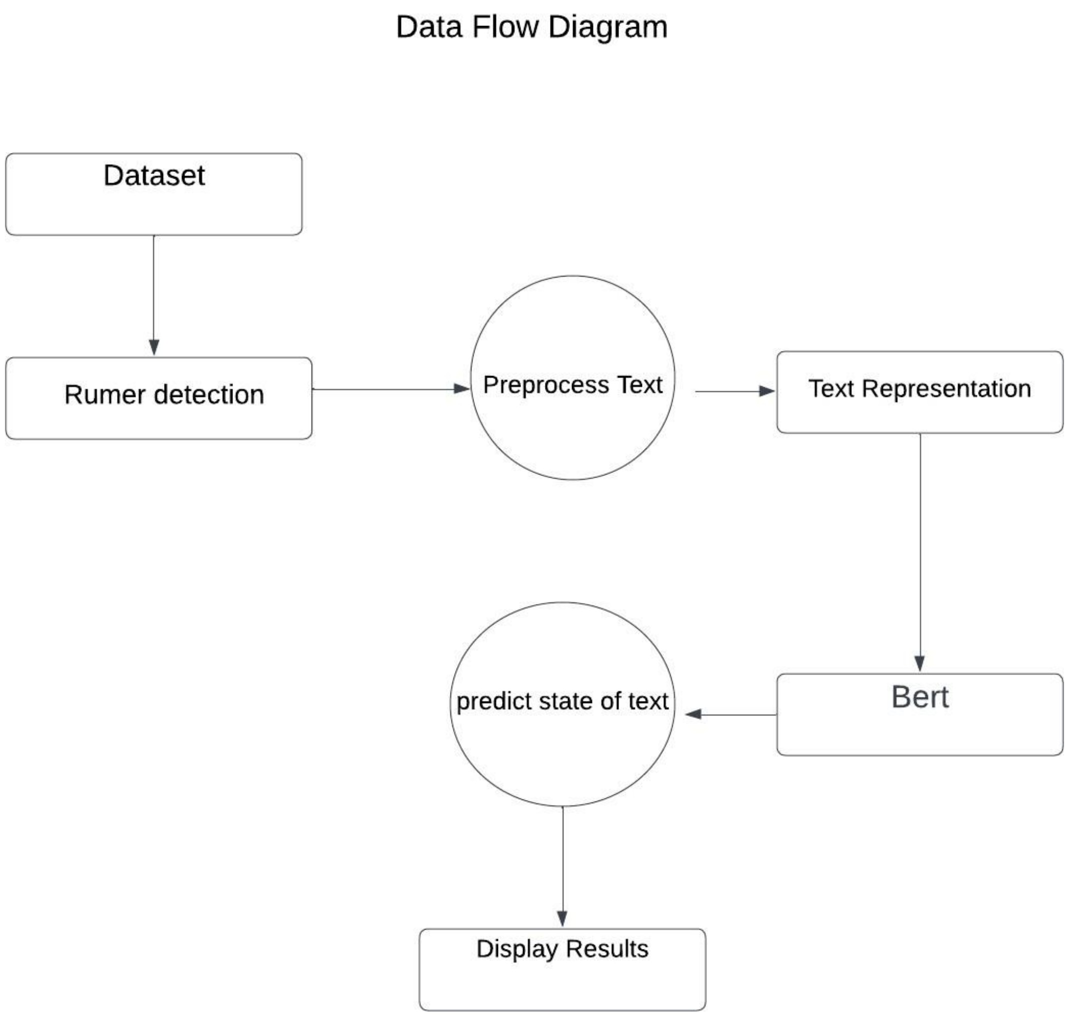


Figure 6.1

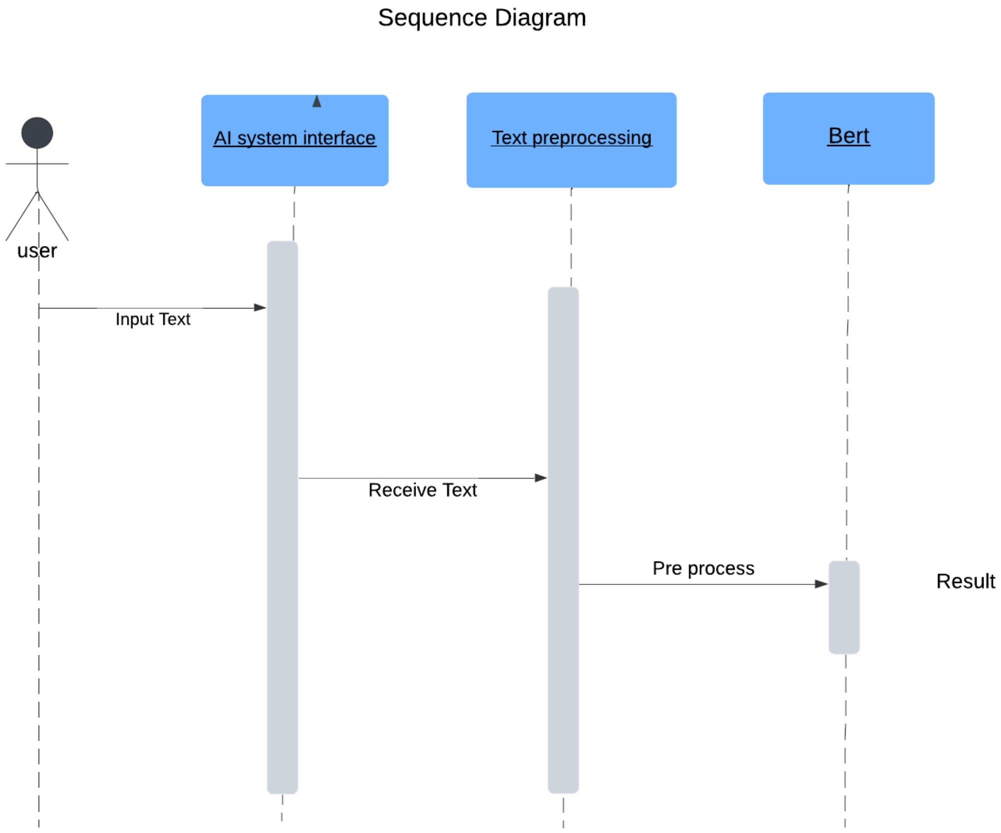
The future plan is to work on data collection making it more reliable and diverse and more accurate.

In future plan, the transfer learning and exploration of open source models available over hugging face for fine tuning on custom dataset to make this project state of the art is included

Data Flow Diagram:



Sequence Diagram:



References:

1. Altheneyan, A. and A. Alhadlaq (2023). "Big data ML-based fake news detection using distributed learning." IEEE Access **11**: 29447-29463.

2. Alyoubi, S., et al. (2023). "The Detection of Fake News in Arabic Tweets Using Deep Learning." Applied Sciences **13**(14): 8209.

3. Arnfield, D. (2023). Enhanced Content-Based Fake News Detection Methods with Context-Labeled News Sources, East Tennessee State University.

4. Bondielli, A. and F. Marcelloni (2019). "A survey on fake news and rumour detection techniques." Information Sciences **497**: 38-55.

5. Buntain, C. and J. Golbeck (2017). Automatically identifying fake news in popular twitter threads. 2017 IEEE International Conference on Smart Cloud (SmartCloud), IEEE.

6. Chen, D., et al. (2023). "CNFRD: A Few-Shot Rumor Detection Framework via Capsule Network for COVID-19." International Journal of Intelligent Systems **2023**.

7. Chen, T., et al. (2018). Call attention to rumors: Deep attention based recurrent neural networks for early rumor detection. Trends and Applications in Knowledge Discovery and Data Mining: PAKDD 2018 Workshops, BDASC, BDM, ML4Cyber, PAISI, DaMEMO, Melbourne, VIC, Australia, June 3, 2018, Revised Selected Papers 22, Springer.

8. Dutta, P. S., et al. (2019). "Fake news prediction: a survey." International Journal of Scientific Engineering and Science **3**(3): 1-3.

9. Gao, Y., et al. (2023). "Rumor detection with self-supervised learning on texts and social graph." Frontiers of Computer Science **17**(4): 174611.

10. Girgis, S., et al. (2018). Deep learning algorithms for detecting fake news in online text. 2018 13th international conference on computer engineering and systems (ICCES), IEEE.

11. Guo, H., et al. (2018). Rumor detection with hierarchical social attention network. Proceedings of the 27th ACM international conference on information and knowledge management.

12. Hamid, A., et al. (2020). "Fake news detection in social media using graph neural networks and NLP Techniques: A COVID-19 use-case." arXiv preprint arXiv:2012.07517.

13. Hanselowski, A., et al. (2018). "A retrospective analysis of the fake news challenge stance detection task." arXiv preprint arXiv:1806.05180.

14. Hossain, T., et al. (2020). COVIDLies: Detecting COVID-19 misinformation on social media. Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020.

15. Imran, M., et al. (2015). "Processing social media messages in mass emergency: A survey." ACM Computing Surveys (CSUR) **47**(4): 1-38.

16. Jiang, S., et al. (2019). User-characteristic enhanced model for fake news detection in social media. Natural Language Processing and Chinese Computing: 8th CCF International Conference, NLPCC 2019, Dunhuang, China, October 9–14, 2019, Proceedings, Part I 8, Springer.

17. Jiang, T., et al. (2021). "A novel stacking approach for accurate detection of fake news." IEEE Access **9**: 22626-22639.

18. Karimi, H., et al. (2018). Multi-source multi-class fake news detection. Proceedings of the 27th international conference on computational linguistics.

19. Katsaros, D., et al. (2019). Which machine learning paradigm for fake news detection? IEEE/WIC/ACM International Conference on Web Intelligence.

20. Keya, A. J., et al. (2022). "AugFake-BERT: handling imbalance through augmentation of fake news using BERT to enhance the performance of fake news classification." Applied Sciences **12**(17): 8398.

21. Lee, N., et al. (2021). "Dynamically addressing unseen rumor via continual learning." arXiv preprint arXiv:2104.08775.

22. Li, J., et al. (2020). Exploiting microblog conversation structures to detect rumors. Proceedings of the 28th International Conference on Computational Linguistics.

23. Liu, F., et al. (2014). "Rumors on social media in disasters: Extending transmission to retransmission."

24. Mahir, E. M., et al. (2019). Detecting fake news using machine learning and deep learning algorithms. 2019 7th international conference on smart computing & communications (ICSCC), IEEE.

25. Mone, S., et al. "FAKE NEWS IDENTIFICATION CS 229: MACHINE LEARNING: GROUP 621."

26. Monti, F., et al. (2019). "Fake news detection on social media using geometric deep learning." arXiv preprint arXiv:1902.06673.

27. Mridha, M. F., et al. (2021). "A comprehensive review on fake news detection with deep learning." IEEE Access **9**: 156151-156170.

28. Mu, Y., et al. (2023). "It's about Time: Rethinking Evaluation on Rumor Detection Benchmarks using Chronological Splits." arXiv preprint arXiv:2302.03147.

29. Pan, J. Z., et al. (2018). Content based fake news detection using knowledge graphs. The Semantic Web–ISWC 2018: 17th International Semantic Web Conference, Monterey, CA, USA, October 8–12, 2018, Proceedings, Part I 17, Springer.

30. Potthast, M., et al. (2017). "A stylometric inquiry into hyperpartisan and fake news." arXiv preprint arXiv:1702.05638.

31. Prachi, N. N., et al. (2022). "Detection of Fake News Using Machine Learning and Natural Language Processing Algorithms [J]." Journal of Advances in Information Technology **13**(6).

32. Qian, F., et al. (2018). Neural User Response Generator: Fake News Detection with Collective User Intelligence. IJCAI.

33. Rodríguez, Á. I. and L. L. Iglesias (2019). "Fake news detection using deep learning." arXiv preprint arXiv:1910.03496.

34. Rubin, V. L., et al. (2016). Fake news or truth? using satirical cues to detect potentially misleading news. Proceedings of the second workshop on computational approaches to deception detection.

35. Shelke, S. and V. Attar (2022). "Rumor detection in social network based on user, content and lexical features." Multimedia Tools and Applications **81**(12): 17347-17368.

36. Shrestha, M. (2018). "Detecting Fake News with Sentiment Analysis and Network Metadata." Earlham College, Fall.

37. Shu, K., et al. (2017). "Fake news detection on social media: A data mining perspective." ACM SIGKDD explorations newsletter **19**(1): 22-36.

38. Singhal, S. "Rumor Detection on Twitter."

39. Song, C., et al. (2019). "CED: credible early detection of social media rumors." IEEE Transactions on Knowledge and Data Engineering **33**(8): 3035-3047.

40. Sun, L. and H. Wang (2023). "Topic-aware Fake News Detection Based on Heterogeneous Graph." IEEE Access.

41. Sun, M., et al. (2022). Ddgcn: Dual dynamic graph convolutional networks for rumor detection on social media. Proceedings of the AAAI conference on artificial intelligence.

42. Torabi Asr, F. and M. Taboada (2019). "Big Data and quality data for fake news and misinformation detection." Big Data & Society **6**(1): 2053951719843310.

43. Umer, M., et al. (2020). "Fake news stance detection using deep learning architecture (CNN-LSTM)." IEEE Access **8**: 156695-156706.

44. Wang, W. Y. (2017). "" liar, liar pants on fire": A new benchmark dataset for fake news detection." arXiv preprint arXiv:1705.00648.

45. Wang, Y., et al. (2020). Fake news detection via knowledge-driven multimodal graph convolutional networks. Proceedings of the 2020 international conference on multimedia retrieval.

46. Wu, F. and X. Luo (2022). "English Text Recognition Deep Learning Framework to Automatically Identify Fake News." Computational Intelligence and Neuroscience **2022**.

47. Yang, X., et al. (2021). Rumor detection on social media with graph structured adversarial learning. Proceedings of the twenty-ninth international conference on international joint conferences on artificial intelligence.

48. Yang, Y., et al. (2018). "TI-CNN: Convolutional neural networks for fake news detection." arXiv preprint arXiv:1806.00749.

49. Yuan, C., et al. (2019). Jointly embedding the local and global relations of heterogeneous graph for rumor detection. 2019 IEEE international conference on data mining (ICDM), IEEE.

50. Zannettou, S., et al. (2019). "The web of false information: Rumors, fake news, hoaxes, clickbait, and various other shenanigans." Journal of Data and Information Quality (JDIQ) **11**(3): 1-37.

51. Zhou, X., et al. (2020). "Safe: similarity-aware multi-modal fake news detection (2020)." Preprint. arXiv **200304981**: 2.

52. Zhou, X. and R. Zafarani (2018). "Fake news: A survey of research, detection methods, and opportunities." arXiv preprint arXiv:1812.00315 **2**.

53. Zuo, Y., et al. (2022). Continually Detection, Rapidly React: Unseen Rumors Detection Based on Continual Prompt-Tuning. Proceedings of the 29th International Conference on Computational Linguistics.