Journal for Re Attach Therapy and Developmental Diversities
eISSN: 2589-7799
2023 December 6 (10s): 1728-1736

Detection Of Fake Online Reviews Using Semi-Supervised And Supervised Learning

G.Shaheen Firdous[1*], G.Prathibha Priyadarshini[2], K.Bindu Sree[3], K.Yeshaswi[4]

[1*]Assistant Professor, Department of CSE, Ravindra College of Engineering for Women
[2]Assistant Professor, Department of CSE, Ravindra College of Engineering for Women
[3]Student, Ravindra College of Engineering for Women
[4]Student, Ravindra College of Engineering for Women

***Corresponding Author:** G.Shaheen Firdous
*Assistant Professor, Department of CSE, Ravindra College of Engineering for Women

Abstract:-
In the era of booming social media and e-commerce, the authenticity of online reviews has become a critical factor influencing consumer decision-making. This study presents an innovative approach to combating the proliferation of fake online reviews, combining semi-supervised and supervised learning techniques. Our method involves extracting pertinent features from review text and employing them to train a supervised classifier, augmented by a semi-supervised algorithm designed to identify and eliminate potentially fraudulent reviews. We leverage a comprehensive dataset comprising genuine and fake reviews, and our results demonstrate the superior performance of our proposed method compared to existing techniques. It achieves a high accuracy rate in identifying fake reviews while effectively minimizing false positives. in parallel, the increasing importance of social media monitoring highlights the significance of analyzing social data to gain insights into consumer behaviour. To this end, we conduct sentiment analysis on Twitter tweets, focusing on user reviews of movies. This paper introduces a novel combined dictionary that incorporates social media keywords and online review data. Additionally, we uncover hidden relationship patterns among these keywords, shedding light on the intricate dynamics of online sentiment expression.
As online shopping continues to gain popularity, online reviews have emerged as pivotal factors in e-commerce decision-making. A substantial portion of consumers routinely consult product and store reviews before making purchasing decisions. Unfortunately, this heightened reliance on reviews has given rise to a surge in fake, spam, or poorly constructed reviews that can mislead consumers. Therefore, the need to distinguish genuine from fraudulent reviews is more critical than ever.
Our study addresses this issue by leveraging machine learning techniques, including supervised and semi-supervised learning, to accurately detect fake reviews. Recognizing that obtaining labelled data can be challenging, we emphasize developing models requiring minimal labelled data. Our approach ensures efficiency, providing rapid results while maintaining high accuracy.
In this paper, we explore various classification algorithms, including support vector machines (SVM), random forests (RF), and deep neural networks, to discern the authenticity of online reviews. By combining the power of these algorithms and the use of unlabeled data through semi-supervised learning, we strive to enhance the credibility of online reviews and promote informed decision-making in the realms of e-commerce and social media.

Introduction:

By using the Internet, obtaining any kind of data from any source worldwide has become incredibly simple. The increased popularity of social networks allows users to obtain a lot of data and information. Huge amounts of information available on these websites also attract the attention of fictitious customers. Twitter has swiftly become a hub for securing ongoing client data online. Customers can share anything without restriction on Twitter, an online social network (OSN), including news, rumours, and even their temperaments. A few arguments can be made regarding a variety of topics, such as governmental issues, ongoing projects, and notable events. When a customer tweets something, it is immediately forwarded to his or her followers, enabling them to spread the received information much broader. With the growth of OSNs, there is a pressing need to consider and analyse how clients behave in online social settings.
Online reviews have become a crucial resource for consumers when making purchasing decisions. However, this reliance on reviews has given rise to the issue of fake reviews, which can mislead consumers and harm businesses. Approximately 4% of online reviews are estimated to be fake, resulting in significant financial losses totalling around $152 billion. Fake reviews can be generated by individuals or groups and automated bots, making their detection a pressing concern.
Consumers frequently compare products from different companies before making a purchase, with reviews playing a central role in their decision-making process. Unfortunately, opportunistic individuals have exploited this by posting fake reviews to tarnish genuine products and promote lower-quality alternatives. This poses a threat to customers, businesses, and companies alike, as the integrity of the information they rely on is compromised.

The prevalence of fake online reviews has grown alongside the surge in e-commerce, particularly during the COVID-19 pandemic, where online shopping has become more common. As a result, even a small percentage of affected users can lead to substantial costs. Preparing for future challenges is essential, making the detection of fake reviews crucial not only for the present but also for future scenarios.
To address this issue, machine learning techniques, including supervised and semi-supervised methods, have been employed to identify fake reviews that could deceive consumers. This project aims to tackle this problem head-on and provides a robust approach to distinguishing genuine from fraudulent reviews, safeguarding the interests of both consumers and businesses.

1.1 Feasibility Study
In this stage, the project's viability is evaluated, and a business proposal is presented with a very generic project plan and some cost projections. The feasibility assessment of the suggested system must be completed during system analysis. This will guarantee that the suggested solution won't burden the business. Understanding the main system requirements is crucial for the feasibility analysis.
Three key considerations involved in the feasibility analysis are:

1.1.1 Economical Feasibility
1.1.2 Technical Feasibility
1.1.3 Social Feasibility
1.1.1 Economical Feasibility
This study is being done to see what kind of financial impact the system will have on the company. The corporation has a finite amount of money to invest in the system's research and development. There must be proof to support the costs. Because the majority of the technologies were free to use, the constructed system was also within the budget. Only customised goods needed to be bought.

1.1.2 Technical Feasibility
This study is being done to evaluate the system's technical requirements for reviews of hype about restaurants based on sentiment analysis or technical feasibility. Any system created must not place a heavy burden on the technical resources at hand. The amount of technological resources available will be heavily strained as a result. As a result, the client will have high expectations. The created system must have low demand because its implementation merely necessitates little or no adjustments.

1.1.3 Social Feasibility
The goal of the study is to determine how much the user accepts the system. This includes the instructions needed for the user to operate the system effectively. The system shouldn't make the user feel threatened; instead, they should view it as a need. The techniques used to inform and acquaint the user with the system are the only factors that affect the level of acceptance by the users. As the system's ultimate user, his confidence must be increased so that he may offer some helpful criticism, which is encouraged.

2 Literature review
Mohawesh et. al [3] offered a thorough analysis of the most significant efforts on machine learning-based fake review identification to date. They have first reviewed the feature extraction strategies employed by numerous researchers. They then described the creation techniques for the existing datasets. Then, using summary tables, they described various conventional machine learning models and neural network models used for false review detection. By increasing feature extraction and classifier construction, traditional statistical machine learning improves the performance of text categorization models. Deep learning, on the other hand, enhances the presentation learning method, the structure of the algorithm, and new knowledge to improve performance. They also gave a comparison of a few transformers and neural network model-based deep learning techniques that haven't been applied to the identification of false reviews. Roberta demonstrated the maximum accuracy on both datasets, according to the results. Additionally, recall precision and the F1 score demonstrated Roberta's effectiveness in spotting bogus reviews. Finally, they evaluated the current research gaps and potential future directions to produce reliable results in this field.
Ahmed et. al [4] proposed an n-gramme analysis and machine learning-based fake news detection methodology. Six alternative machine classification algorithms are examined and compared, as well as two different feature extraction strategies. The Term Frequency-Inverted Document Frequency (TF-IDF) feature extraction technique and Linear Support Vector Machine (LSVM) classifier produced the best results in the experimental evaluation, with an accuracy of 92%.
Atefeh Heydari et. al [5] proposed a reliable spam detection method for reviews. A thorough assessment of the literature revealed the timing element's potential in this field and inspired the creation of a review spam detection algorithm based on time series analysis techniques. In this experiment, they suggest a review spam detection strategy that looks into the

falsity of reviews that have been submitted, based on the idea that capturing burst patterns during the reviewing process can increase detection accuracy.

Paul et al [6] examine the research on the detection of fake online platform reviews (FRD). It addresses the causes of the limited success that the present procedures and rules have had in averting damage due to false reviews and covers both basic research and commercial solutions.

Deng et. al [7] The wording of the review always informs us of reality, according to an analysis of all the traits of hype-filled phoney reviews. They suggested an algorithm to detect online phoney Using sentiment analysis, review the excitement surrounding restaurants. because hype reviews are either good or negative. Reviews are taken into account in our experiment in four different ways: sensitivity, atmosphere, service, and general attitude. A review will be labelled as a hype review if the four dimensions' analysis results are consistent. Our test results showed that our algorithm has an accuracy of roughly 74%, and the approach suggested in this article is also applicable in other contexts, such as sentiment analysis of online opinion in emergency case management.

Rathore et al [8] offer a top-down structure finding fake reviewer groups created using the DeepWalk algorithm using the graph data from the reviewers and a (modified) semi-supervised clustering method that can take into account some prior knowledge. They use an actual review dataset from the Google Play Store to test our suggested approach. Of the 38,123 reviewer IDs in the dataset, 2207 fraud reviewer IDs have partial ground-truth information available. Our testing findings show that the suggested method may detect the potential spammer groups quite well. The suggested method can be further developed to identify social media groups of opinion spammers (such as fake comments or fake postings) using temporal affinity, semantic traits, and sentiment analysis.

Khan et al [9] developed a method Using supervised learning to identify false reviews in Internet literature. To distinguish between bogus and real reviews, the study uses machine learning classifiers. The performance of the proposed system is compared to benchmark works, and experimental findings are assessed using several evaluation metrics.

Li et. al [10] To the best of our knowledge, this is the first time the idea of a review group has been put forth and employed. The purpose of the review grouping algorithm is to efficiently divide the reviews of reviewers into groups that take part in creating cutting-edge grouping models to recognise both positive and unfavourable false reviews. Based on the group concept, several new language-independent features are created. To construct a model of reviewer group collusion, they also investigate the relationships between reviewers who collude. The review group technique and reviewer group collusion models have been shown to significantly increase precision in the false reviews classification job when compared to baselines, especially when reviews are provided by professional review spammers, by 4% to 7%.

3.0 MODULES

The project work is divided into 3 modules.

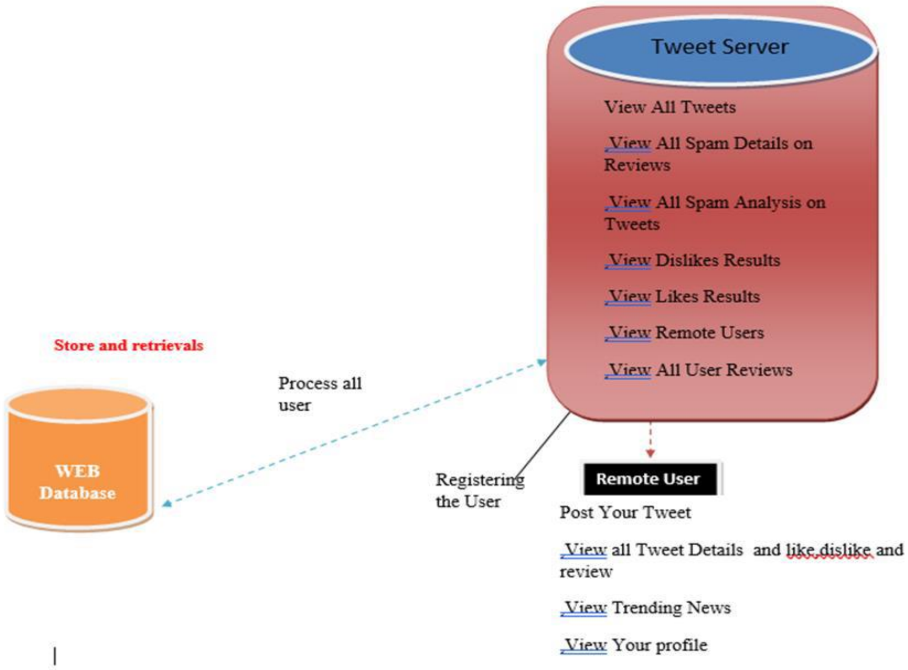


Fig3.2 Architecture Diagram

3.1.1 Data Pre-Processing

On Twitter, spammers can be challenging to detect. The recommended method combines information from informal organisations with the elimination of features from textual content. Before utilising lattice factorization to establish the underlined highlight grid or the tweets, the authors first created a social regularisation using the association coefficient to show how the underlining framework was factorised. To represent the current state of Twitter, the researchers produced the UDI Twitter dataset by combining data with social regularisation and factorization grid measurements, performing probes, and combining the results described utilising a hiddeMarkovov model to differentiate spam that was discovered after it had already been caught. The proposed scientific categorization is divided into four basic categories, including counterfeit client ID, URL-based spam recognition, counterfeit material, and identifying spam in drifting spots. Every class of differentiating proof techniques is based on a certain model, technique, and location computation. entities charged with enforcing the law must take proactive steps to protect women.

3.1.2 Support Vector Machine (SVM)

A Support Vector Machine (SVM) is a popular machine-learning technique for classification and regression issues. Data classification using SVMs, a supervised learning technique, is possible in both linear and non-linear ways.SVMs provide a variety of benefits over other machine learning methods like decision trees and logistic regression. They are less prone to overfitting, which occurs when a model is extremely complex and matches the training data too closely, and are more resistant to data noise. SVMs are computationally effective and can handle large datasets. In the project's sentiment analysis of tweets on women's safety in Indian cities, SVMs can be used. The SVM algorithm can be taught using a dataset of tweets that have been labelled.SVM can be trained on a set of labelled reviews in the context of detecting phoney online reviews to find patterns and features that differentiate between real and fake reviews. Word frequency, readability score, and sentiment analysis are a few variables that can be employed to train the SVM model. The SVM model can be used to categorise incoming reviews as authentic or false after it has been trained. SVM is also applicable to semi-supervised learning to increase the model's capacity for accurate categorization. The SVM model is first trained on a collection of labelled data in semi-supervised learning. The unlabeled data is then classified using the model, and the classifications are utilised to update the model. This procedure is repeated until the model's accuracy is adequate. SVM is a potent machine learning technique that can be used to identify false internet reviews, to sum up. SVM can be used in semi-supervised learning to increase the precision of the classification model and can be trained on labelled data to differentiate between real and fraudulent reviews.

3.1.3 Simple Bayes algorithm

The popular machine learning algorithm Naive Bayes can be used to recognise fake internet using supervised and semi-supervised learning, and review techniques. It uses the Bayes theorem. by Naive Bayes, a probabilistic approach, to assess if a set of qualities suggests that a review is valid or not. In the case of identifying false internet reviews, Naive Bayes can be trained on a set of labelled reviews to identify patterns and attributes that distinguish between real and phoney reviews. Several characteristics, such as sentiment analysis, word frequency, and readability score, can be used to train the Naive Bayes model. After being trained, the Naive Bayes model can be used to classify incoming reviews as true or false.
The unfounded assumption of Naive Bayes is that the features used to train the model are not reliant on one another. Despite this restriction, naive Bayes is frequently used in text classification issues, such as the finding of bogus internet reviews. Naive Bayes is quite simple to construct and may be easily taught on large datasets. In conclusion, fraudulent internet reviews can be detected using the Naive Bayes approach, a well-known machine learning algorithm. Naive Bayes can be trained on labelled data to discern between genuine and fake reviews. It is also applicable to semi-supervised learning to improve the precision of the classification model. The average proportions of confirmed records that were spam and those that weren't, as well as the number of fervent backers of the client accounts. The following parameters were used to determine the manufacturing of the phoney substance: social notoriety, global commitment, theme commitment, agreeability, and believability. From that point forward, the fake drug's future development and the overall impact of individuals who disseminated it at the time were both predicted using the model of projected relapse.

3.2 Algorithm for Support Vector Machines (SVM)

One of the most well-liked supervised learning algorithms, Support Vector Machine, or SVM, is used to solve Classification and Regression problems. However, it is largely employed in Machine Learning Classification issues.
The SVM algorithm's objective is to establish the ideal line or border for a decision that can classify n-dimensional space, allowing us to quickly classify fresh data points in the future. A hyperplane is the name given to this optimal decision boundary.
SVM selects the extreme vectors and points that aid in the creation of the hyperplane. Support vectors, which are used to represent these extreme instances, are what give the Support Vector Machine method its name. Take a look at the diagram below, where two distinct categories are separated using a decision boundary or hyperplane.

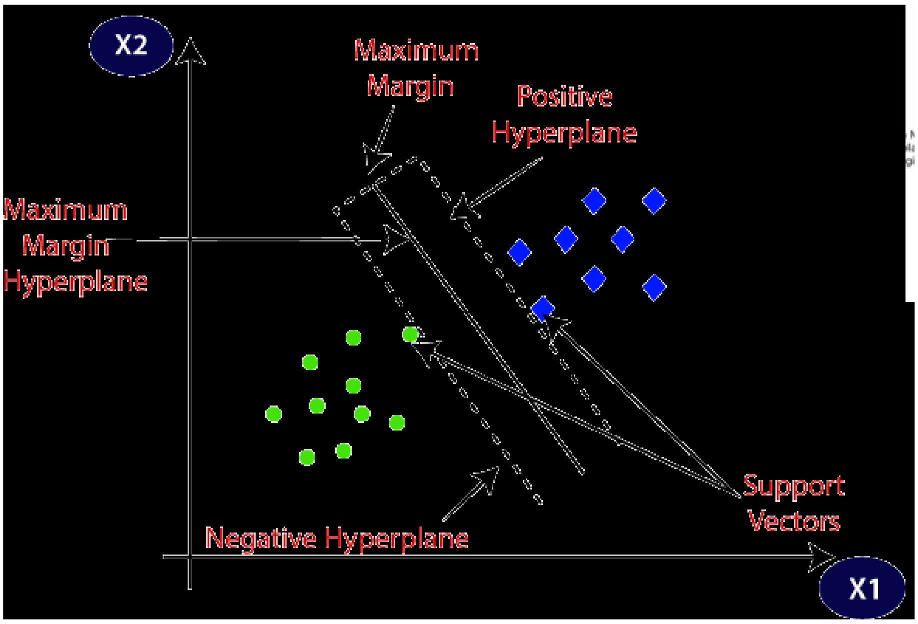


Fig-4.1 Support Vector Machine

4. Methodology

4.1 Data Gathering

The gathering of data sets is the initial phase. Millions of song datasets are used to develop the model. It includes numerous audio tracks in the WAV file type. 30-second chunks make up each audio file. The datasets are made up of 100 songs in each of 10 different genres. Genres including blues, classical, country, disco, hip hop, jazz, metal, pop, reggae, and rock are all represented in the million song data sets. 20% of the data is used for testing, while the remaining 80% is used for training.

4.2 Data Preparation

Data preparation is one way to get the data ready and make it suitable for a machine-learning model. It is the first and most important stage when creating a machine-learning model. It's not always the case when creating a machine learning project that we tend to find clear and formatted knowledge. And while performing any activity with knowledge, it's important to clean it and arrange it in a very organised way. Consequently, we frequently apply a knowledge preparation task to this.

4.3 Model training

Once the model has been built, the model coaching must be completed. T the frequency domain and temporal domain the information preprocessing stage is essential since feature selection enhances model correctness. The time domain and frequency domain options have been manually extracted. The supervised classification algorithmic programme uses the classifiers random forest, support vector machine, and CNN. The number of decision trees that are formed is the main factor. Accuracy is also added in addition to the decision tree.

4.5 Support Vector Machine

A supervised learning formula may be the support vector machine. The SVM may function as a classifier that generates several hyperplanes in an associated infinite-dimensional space. The SVM space units are utilised for a variety of tasks, including regression and classification. In a highly robust support vector machine, it is crucial to create the best boundary that can be seen as a hyperplane. Support vector machine space units have the advantage of altering huge dimensional data sets; they are frequently used to categorise improved biological identification of bogus review data.

4.6. LSTM (Long Short-Term Memory)

Long short-term memory (LSTM) is a recurrent neural network (RNN) architecture used in deep learning. Unlike conventional feedforward neural networks, LSTM has feedback connections. For instance, LSTM can be used for tasks like text categorization, connected, unsegmented handwriting identification, speech recognition, network traffic anomaly detection, and IDS (intrusion detection systems), among others.

The three "regulators" of the data flow within an LSTM unit, which are typically referred to as gates in common designs, are an input gate, an output gate, and a forget gate. The memory element of an LSTM unit is a cell. Some variants of the

Journal for Re Attach Therapy and Developmental Diversities
eISSN: 2589-7799
2023 December 6 (10s): 1728-1736

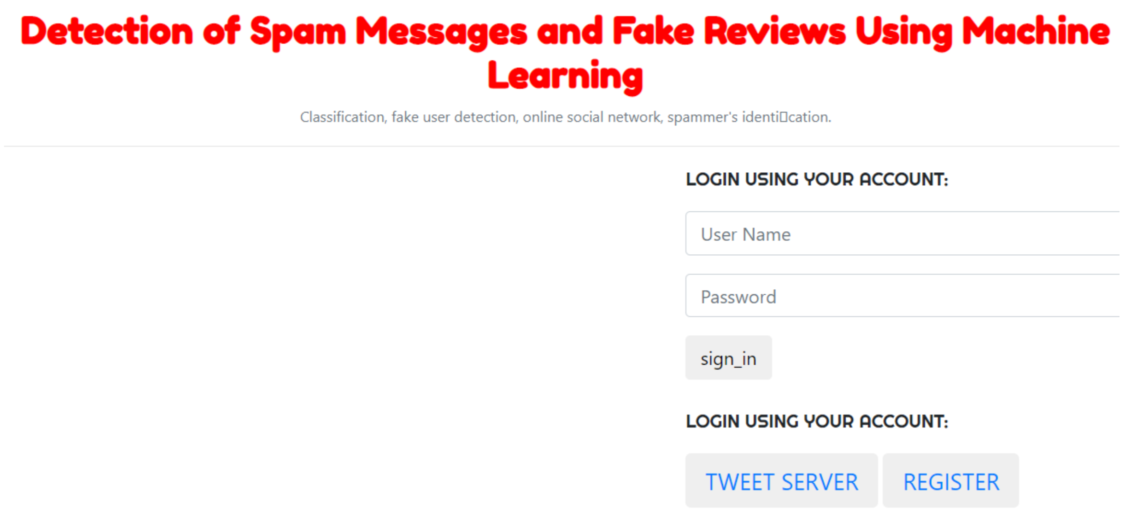
LSTM unit don't produce any, a few, or even all of those gates. For instance, GRUs (gated repeating units) are deficient in an output gate.

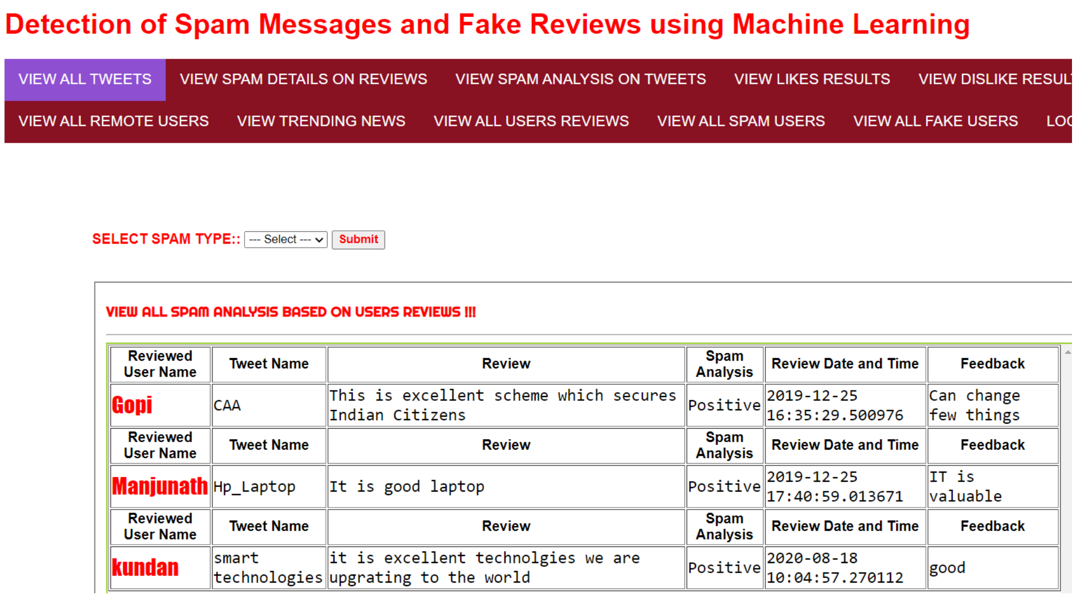
4.7 Model evaluation and testing

After the model coaching, checking the model is the final stage. The testing phase requires the use of datasets. Utilising learned knowledge and test data sets, the model is assessed. As a result, each component of the model is evaluated, and its functioning is verified throughout the testing phase. The main goal of testing is to determine whether each component is operating successfully or unsuccessfully. The training phase is finished making use of support vector machines, CNN, and random forest classifiers. The testing is completed using 20% of the dataset to predict the results.

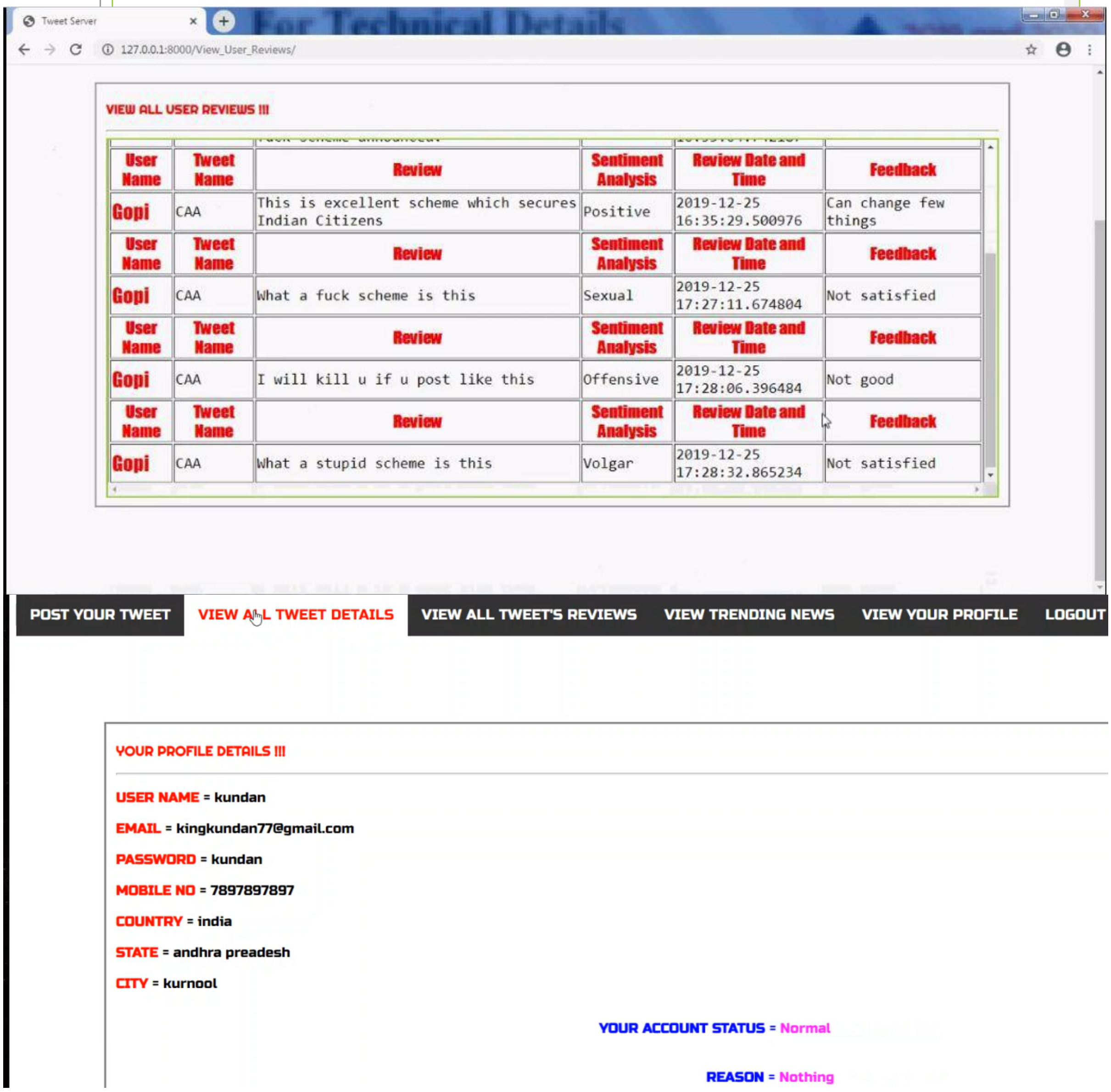
Result

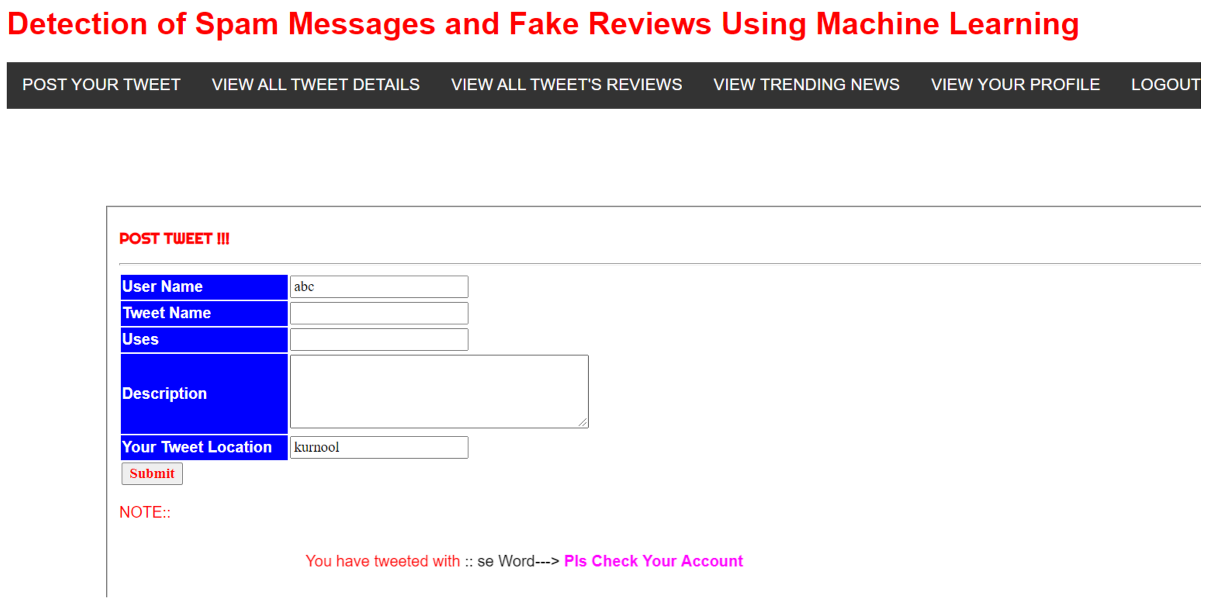
Step 1: Type the python app.py command on the Command Prompt after choosing the path to the file.
Step 2: If you're a new user, click the register option and fill out the form.
Step 3: After entering your username and password, click the login button.
Step 4: Type the text into the text field and press the "submit" button.
Step 5: Determine whether the review is fake or not using the RF, SVM, and CNN-LSTM algorithms.
Step 6: Ultimately, compare the accuracy of three algorithms and determine which has the highest accuracy.






Detection of Spam Messages and Fake Reviews using Machine Learning
VIEW ALL TWEETS
VIEW SPAM DETAILS ON REVIEWS
VIEW SPAM ANALYSIS ON TWEETS
VIEW LIKES RESULTS
VIEW DISLIKE RESULTS
VIEW ALL REMOTE USERS
VIEW TRENDING NEWS
VIEW ALL USERS REVIEWS
VIEW ALL SPAM USERS
VIEW ALL FAKE USERS
LOGOUT
SELECT SPAM TYPE::
--- Select ---
Submit
VIEW ALL SPAM ANALYSIS BASED ON USERS REVIEWS !!!
Reviewed User Name
Tweet Name
Review
Spam Analysis
Review Date and Time
Feedback
Gopi
CAA
What a stupid scheme is this
Volgar
2019-12-25 17:28:32.865234
Not satisfied


Tweet Server
127.0.0.1:8000/View_User_Reviews/
VIEW ALL USER REVIEWS !!!
User Name
Tweet Name
Review
Sentiment Analysis
Review Date and Time
Feedback
Gopi
CAA
This is excellent scheme which secures Indian Citizens
Positive
2019-12-25 16:35:29.500976
Can change few things
Gopi
CAA
What a fuck scheme is this
Sexual
2019-12-25 17:27:11.674804
Not satisfied
Gopi
CAA
I will kill u if u post like this
Offensive
2019-12-25 17:28:06.396484
Not good
Gopi
CAA
What a stupid scheme is this
Volgar
2019-12-25 17:28:32.865234
Not satisfied
POST YOUR TWEET
VIEW ALL TWEET DETAILS
VIEW ALL TWEET'S REVIEWS
VIEW TRENDING NEWS
VIEW YOUR PROFILE
LOGOUT
YOUR PROFILE DETAILS !!!
USER NAME = kundan
EMAIL = kingkundan77@gmail.com
PASSWORD = kundan
MOBILE NO = 7897897897
COUNTRY = india
STATE = andhra preadesh
CITY = kurnool
YOUR ACCOUNT STATUS = Normal
REASON = Nothing



Conclusion

I conducted a study of the methods used to identify spammers on online networks. In addition, a scientific taxonomy of Twitter spam identification approaches was introduced, classifying them as fake substance recognition, fake client recognition, URL-based spam discovery, and fake location methods. I considered the methods that were presented in light of some highlights, such as customer highlights, content highlights, chart features, and time highlights.

Additionally, the predefined goals and datasets used by the approaches were also examined. It is anticipated that the newly introduced survey will help analysts uncover the information on the top Twitter spam recognition techniques in a well-organized framework. Despite the development of effective and efficient approaches for the detection of spam and phoney client IDs on Twitter, there are still a few unexplored areas that demand the analysts' full attention. Due to the real effects that false news has on people on an individual as well as a collective level, the difficulties are more briefly featured under False news identifiable proof via web-based media networks. The recognisable verification of discussion sources via web-based media is another related topic worth looking into. Even if a few studies using factual methods have been focused on locating the sources of rumours, more contemporary methodologies, such as informal community-based methodologies, can be used because of their proven effectiveness.

References

[1] B. Erçahin, Ö. Aktaş, D. Kilinç, and C. Akyol, ''Twitter fake account detection,'' in Proc. Int. Conf. Comput. Sci. Eng. (UBMK), Oct. 2017, pp. 388–392.

[2] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, ''Detecting spammers on Twitter,'' in Proc. Collaboration, Electron. Messaging, Anti- Abuse Spam Conf. (CEAS), vol. 6, Jul. 2010, p. 12.

[3] S. Gharge, and M. Chavan, ''An integrated approach for malicious tweets detection using NLP,'' in Proc. Int. Conf. Inventive Commun. Comput. Technol. (ICICCT), Mar. 2017, pp. 435–438.

[4] T. Wu, S. Wen, Y. Xiang, and W. Zhou, ''Twitter spam detection: Sur- vey of new approaches and comparative study,'' Comput. Secur., vol. 76, pp. 265–284, Jul. 2018.

[5] S. J. Soman, ''A survey on behaviors exhibited by spammers in popular social media networks,'' in Proc. Int. Conf. Circuit, Power Comput. Tech- nol. (ICCPCT), Mar. 2016, pp. 1–6

[6] A. Gupta, H. Lamba, and P. Kumaraguru, ''1.00 per RT #BostonMarathon # prayforboston: Analyzing fake content on Twitter,'' in Proc. eCrime Researchers Summit (eCRS), 2013, pp. 1–12.

[7] F. Concone, A. De Paola, G. Lo Re, and M. Morana, ''Twitter analysis for real-time malware discovery,'' in Proc. AEIT Int. Annu. Conf., Sep. 2017, pp. 1–6.

[8] C. Chen, Y. Wang, J. Zhang, Y. Xiang, W. Zhou, and G. Min, ''Statistical features-based realtime detection of drifted Twitter spam,'' IEEE Trans. Inf. Forensics Security, vol. 12, no. 4, pp. 914–925, Apr. 2017.

[9] C. Buntain and J. Golbeck, ''Automatically identifying fake news in popu- lar Twitter threads,'' in Proc. IEEE Int. Conf. Smart Cloud (SmartCloud), Nov. 2017, pp. 208–215.

[10] S. Keretna, A. Hossny, and D. Creighton, ''Recognising user identity in Twitter social networks via text mining,'' in Proc. IEEE Int. Conf. Syst., Man, Cybern., Oct. 2013, pp. 3079–3082.

[11] S. Sadiq, Y. Yan, A. Taylor, M.-L. Shyu, S.-C. Chen, and D. Feaster, ‘‘AAFA: Associative affinity factor analysis for bot detection and stance classification in Twitter,’’ in Proc. IEEE Int. Conf. Inf. Reuse Integr. (IRI), Aug. 2017, pp. 356–365.
[12] M. U. S. Khan, M. Ali, A. Abbas, S. U. Khan, and A. Y. Zomaya, ‘‘Segregating spammers and unsolicited bloggers from genuine experts on Twitter,’’ IEEE Trans. Dependable Secure Comput., vol. 15, no. 4, pp. 551–560, Jul./Aug. 2018.