ISSN No:-2456-2165

Fraudulent Text Detection System Using Hybrid Machine Learning and Natural Language Processing Approaches

Muhammad Dawaki[1]
Department of Computer Science Gombe State University Gombe, Nigeria

Ahmed Mohammed[2]
Department of Computing Technologies SRM Institute of Science and Technology Tamil Nadu, India

Dr. Mustapha Ismail[3]
Department of Computer Science Gombe State University Gombe, Nigeria

Abstract:- In today's digital age, fraudulent behaviour is becoming increasingly common. Many of the fraudulent actions have been carried out by sending text messages with malicious links attached, which can disrupt a system and potentially steal confidential personal information from a user. A system capable of identifying and classifying fraudulent content within a text string was developed in this project using machine learning algorithms and natural language processing libraries. Due to the ever-changing and sophisticated nature of fraudulent activity, detecting fraud is a difficult task that necessitates the use of cutting-edge technology to combat fraud. However, this research looked at the potential of developing a cutting-edge machine learning model. The fraudulent detection model was trained and tested using many machine learning algorithms utilizing an SMS spam dataset in this study. Three of the eleven algorithms used, K-Nearest Neighbor, Naive Bayesian Classifier, and Random Forest Classifier, outperformed the others, with performance accuracy and precision of 90% and 100% for K-Nearest Neighbor, 96% and 100% for Nave Bayesian Classifier, and 97% and 100% for Random Forest Classifier, respectively. The count vectorizer technique was used to select and extract the best features. The final optimal model performance obtained was 97% accuracy and 100% precision using accuracy, precision, recall, and f1-measure as metrics. The results obtained are promising, and the model was deployed using the streamlit framework.

Keywords:- *Fraudulent, Machine Learning, Dataset, Algorithm, Natural Language Processing.*

I. INTRODUCTION

Fraud can be defined as “intentional deception to secure unfair or unlawful gain”. There are many ways to carry out a finance fraud due to the convenience of Internet. The most common fraud in daily life is email, SMS, internet, credit card or debit card fraud. To address this problem, many researchers include data scientists and software engineers tried to develop a learning algorithm to find patterns from the fraud transaction and thus detect the potential frauds using this system [1].

With the popularity of the Internet, email is a part of our daily life. It is the most widely used medium for communication worldwide because of its cost effectiveness, reliability, quickness and easy accessibility. Email is prone to spam emails because of its wide usage and all of its benefits as a genuine medium of communication. Internet Spam is one or more unsolicited messages sent or posted as a part of larger collection of messages, all having substantially identical content. Most spam messages take the form of advertising or promotional materials like debt reduction plans, getting rich quick schemes, gambling opportunities, pornography, online dating, health-related products etc. The major technical disadvantages of spam messages are wastage of network resources (bandwidth), wastage of time, damage to the PC‟s & laptops due to viruses. Spammers generally have a designed personalized template emails to deliver their messages using a bulk mailing software. It is widely assumed that most of the spam messages are sent directly from a collection of bots [2].

In this study, an algorithm will be developed and optimized to accurately predict a fraudulent content within a text string, being whether email text, SMS text, internet content and the like.

ISSN No:-2456-2165

II. LITERATURE SURVEY

This section featured a literature review as well as the Research's current approaches.

A. Performance of machine learning techniques in the detection of financial frauds

In this paper, I. Sadgali, N. Sael and F. Benabbou [3] proposed Machine Learning Algorithm to detect financial fraud. Financial fraud is on the rise, posing a severe threat to the financial industry. As a result, financial institutions are under pressure to upgrade their fraud detection systems on a regular basis. Several studies have employed machine learning and data mining approaches to find solutions to this challenge in recent years. We present a state-of-the-art on numerous fraud strategies as well as detection and prevention techniques such as classification, clustering, and regression recommended in the literature in the work. Due to the nature of unstructured data, deep learning may be suitable for this kind of problem.

B. Email Spam Detection and Data Optimization using NLP Techniques

In this paper, S. T. Dhivya, S. Nithya, G. S. Priya, E. Pugazhendi [4] Proposed email spam detection and data optimization with NLP techniques System based on Latent Dirichlet Allocation (LDA) and email spam detection based on NLP-N-gram model and opinion ranking. Here the mail data optimization is achieved by deleting the advertising e-mail with attachments. The optimization process involves the following steps: Dynamic input emails are taken from various platforms such as Gmail, Yahoo, Live Mail via the Java Mail API and then data preprocessing is carried out and various steps such as tokenization, lemmatization and stemming are carried out. The TF-IDF vectorization is found for the words and stored in a document matrix. Then the percentage of the punctuation to that of the sentence is found. These above-mentioned procedures are carried out to identify whether the specified email is spam or ham. The mail data is optimized by classifying topics using LDA and deleting classified advertising and spam e-mails. The percentage of data saved is 7.3%. Latent Dirichlet Allocation (LDA) Algorithm is an unsupervised approach, but supervised learning should also be considered.

C. Spam Detection Approach for Secure Mobile Message Communication Using Machine Learning Algorithms

In this paper, A. U. Haq, L. G. Jun, S. Nazir,Habib and U. Khan [5] Proposed that machine learning-based spam detection can be used for accurate spam detection. To classify spam and ham messages in mobile device communications, they employed Logistic Regression, K-nearest neighbor, and Detection Tree. The strategies are tested using a collection of SMS datasets. The dataset is also divided into two halves, one for testing and the other for training. Furthermore, 70 percent of the data is used for training, while 30 percent is used for testing. The Logistic Regression is a classifier that computes the prediction y in a binary classification problem as 0 or 1, indicating whether it belongs to a negative or positive class. It predicts values for the variable in the multiple classification. The decision tree is a supervised machine learning algorithm in the form of a tree, with each node representing a decision or leaf node. In this tree the nodes are linked to one another. The K-nearest neighbor classification is also a supervised learning algorithm, but the performance is not good enough.

D. Opinion Rank: Trustworthy Website Detection using Three Valued Subjective Logic

In this paper, G. Liu, Q. Yang and X. Niu [6] proposed the Opinion Rank algorithm to calculate the trustworthiness of each available website and identify the trustworthy ones with high trust values. This algorithm is based on a broad search algorithm that starts with an existing set of trusted websites. Since these websites play an important role in opinion ranking, they have also used other algorithms like High PageRank and Inverse PageRank to rank the websites based on their trustworthiness. Using the public dataset, they validated the Opinion Rank and the HarMean PageRank which analyze the impact of website selection. The opinion rank algorithm calculates the trustworthiness of all websites, trust propagation and trust combination operations defined in the three-valued trust model of subjective logic. The HarMean PageRank combines the results of PageRank and Inverse PageRank. The convergence and performance of this algorithm are better than those of the Trust-Rank and Good-Rank algorithms. Opinion Rank identifies more trusted websites and less spam websites in less time. One of the drawbacks is that it is very difficult to identify the subset of websites that are needed to update their trustworthiness.

E. Improved email spam classification method using integrated particle swarm optimization and decision tree

In this paper, A. Sharma and H. Kaur [7] proposed a machine learning model for spam detection that combined the Naive Bayes algorithm and intelligence-based Particle Swarm Optimization. The Bayes theorem, which has a high probability distribution property, is the foundation of the Naive Bayes algorithm. Particle Swarm Optimization, on the other hand, is based on the behavior of fish and birds. Based on the keywords included in the email data, the Naive Bayes algorithm calculates the mail class and non-spam class. The Particle Swarm Optimization approach is also used to improve the accuracy and classification process by optimizing the parameters of the Naive Bayes algorithm.

F. Drawbacks of the Existing work

There are many systems designed for fraud detection, but they lack Graphical User Interface (GUI) for easy user interaction.

S.No	Paper	Methodology/Algorithm Used	Advantages	Disadvantages
1	[3]	Machine Learning	Classification, clustering, and regression techniques are pretty good in performing fraud detection.	Due to the nature of unstructured data, deep learning may be suitable for this kind of problem.
2	[4]	NLP Latent Dirichlet Allocation (LDA) and TF-IDF	NLP TF-IDF Is excellent in features selection and extraction.	Latent Dirichlet Allocation (LDA) Algorithm is an unsupervised learning approach, but supervised learning should also be considered.
3	[5]	Machine Learning	The decision tree as a supervised machine learning model perform better than KNN with good accuracy scores.	The K-nearest neighbor classification is also a supervised learning algorithm, but the performance is not good enough.
4	[6]	Three Valued Subjective Logic	Opinion Rank identifies more trusted websites and less spam websites in less time.	One of the drawbacks is that it is very difficult to identify the subset of websites that are needed to update their trustworthiness
5	[7]	Machine Learning & integrated particle swarm optimization	The Particle Swarm Optimization approach was used to improve the accuracy and classification process by optimizing the parameters of the Naive Bayes algorithm.	The study did not take into account deep learning approach with the particle swarm optimization.

Table 1: Comparison of the existing works

III. PROPOSED APPROACH

In this research, a hybrid machine learning approach to build fraudulent content detection and classification system form text data was proposed. The proposed system will be able to classify a text as fraudulent or non-fraudulent, which will be extremely useful for the security of our digital interactions as well as the safety of our personal and confidential information. The proposed system will be implemented with the use of a dataset and machine learning algorithms for training, testing, and prediction. The dataset will be split into two portions, one for training and the other for testing, with a 7:3 or 8:2 ratio. Both the training and testing data splits will be used to determine the accuracy of the classification using accuracy, precision, and recall and f1-measure metrics respectively.

IV. RESEARCH OBJECTIVES

The goals of this study are to:

- Develop a fraud detection system using hybrid machine learning techniques and the state-of-the-art NLTK library toolkit in Python.
- Use text processing techniques to classify text content in order to detect suspicious or fraudulent text.
- Build a system for fraudulent classification task automation by training and testing algorithms with a dataset.
- Deploy the model for use with the aid of streamlit framework.

V. PROBLEM DEFINITION

It is well known to each and every one of us that fraud is becoming more and more common in our digital society.

Many people have fallen victim to online fraud while surfing the web, receiving email, or receiving text messages. The majority of fraudulent content includes malicious links that, when opened, might harm your system and perhaps steal confidential personal information such as financial transaction information, which is usually the case. The only approach to overcome this problem is to design a method or mechanism of identifying the presence of fraudulent content within a text in the first place by using some of the existing machine learning algorithms and utilizing the state-of-the-art python library known as NLTK (Natural Language Toolkit).

VI. RESEARCH METHODOLOGY

Several machine learning algorithms were employed to the SMS Spam dataset for the implementation based on the below block diagram:

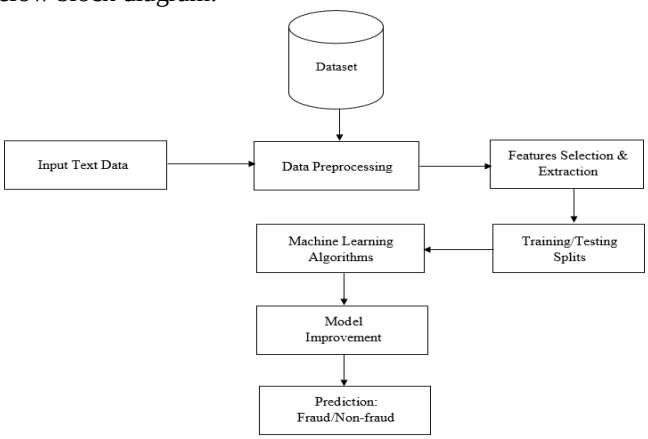

Fig. 1: Proposed System Block Diagram

ISSN No:-2456-2165

The phases of classifying a text into fraud or non-fraud class is as follow; input text string or dataset generation, data pre-processing, data cleaning, text lowercasing and tokenization, stop words and punctuations removal, special characters removal, text stemming, features selection and extraction using either of count vectorizer or TF-IDF vectorizer or both, splitting the dataset into training and testing set in the ratio of 7:3, perform the training and testing using support vector machine, k-neighbour classifier, naïve Bayes classifier, decision tree classifier, linear regression classifier, random forest classifier, extra trees classifier, gradient boosting classifier, XGB classifier, bagging classifier, and adaboost classifier, output classification result, perform model improvement and save the final optimal model for deployment. The figures below depict how those phases were performed.

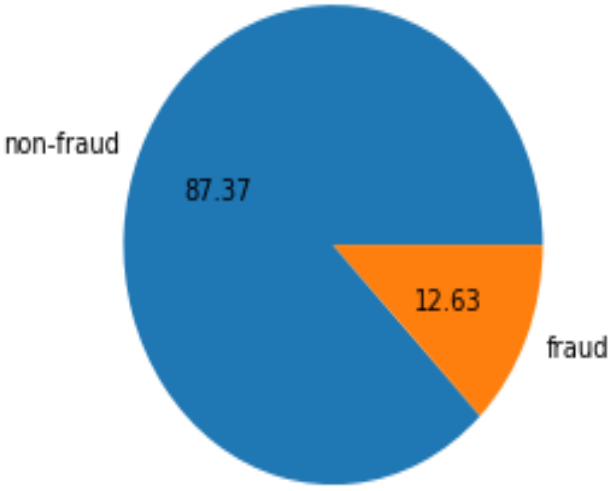

Fig. 2: Dataset fraud and non-fraud classes'percentage distribution

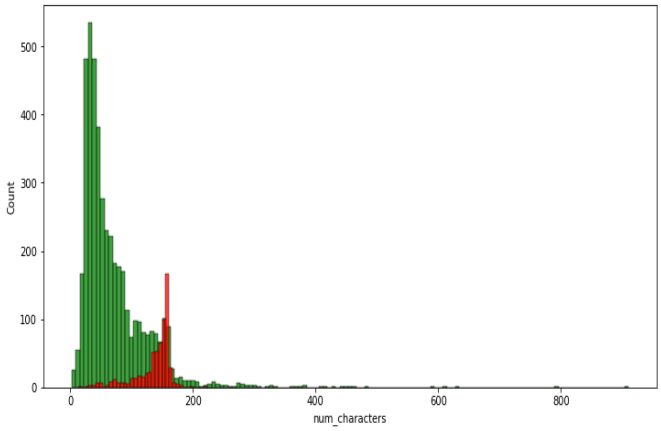

Fig. 3: Distribution of fraud and non-fraud instances based on number of characters

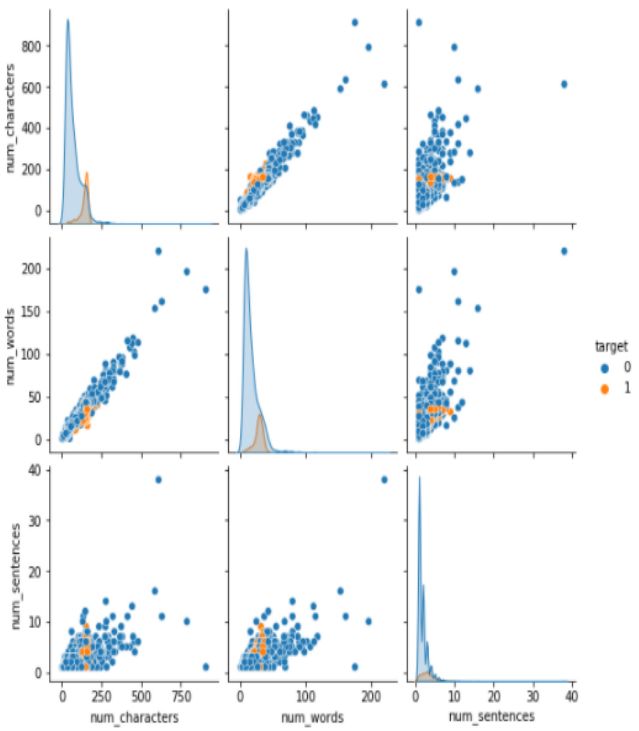

Fig. 4: Pair plot of the corresponding number of characters, words and sentences

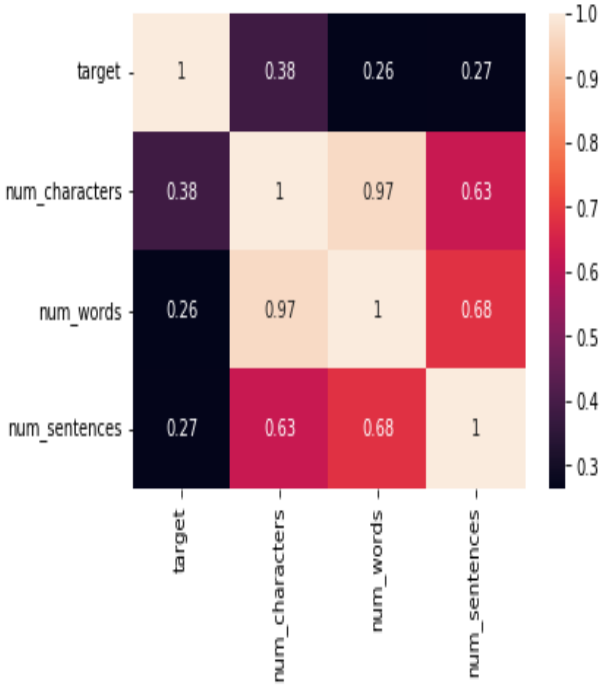

Fig. 5: Correlation matrix of the corresponding number of characters, words and sentences

ISSN No:-2456-2165

	target	text	num_characters	num_words	num_sentences	transformed_text
0	0	Go until jurong point, crazy.. Available only ...	111	24	2	go jurong point crazi avail bugi n great world...
1	0	Ok lar... Joking wif u oni...	29	8	2	ok lar joke wif u oni
2	1	Free entry in 2 a wkly comp to win FA Cup fina...	155	37	2	free entri 2 wkli comp win fa cup final tkt 21...
3	0	U dun say so early hor... U c already then say...	49	13	1	u dun say earli hor u c already say
4	0	Nah I don't think he goes to usf, he lives aro...	61	15	1	nah think goe usf live around though

Fig. 6: Cleaned text

Fig. 7: Word cloud for fraud corpus

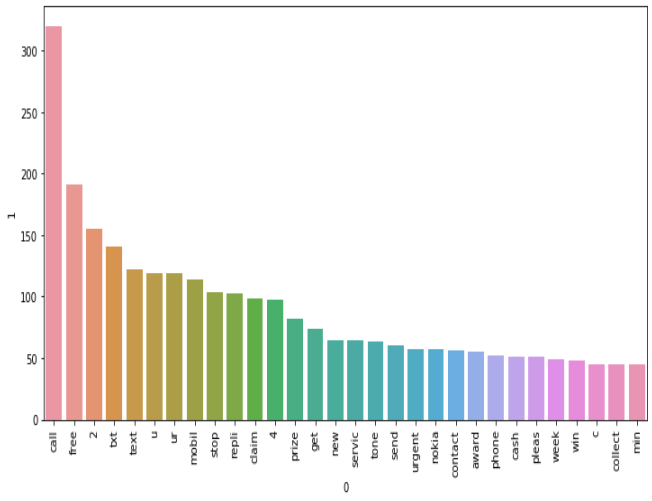

Fig. 8: Bar plot for fraud corpus

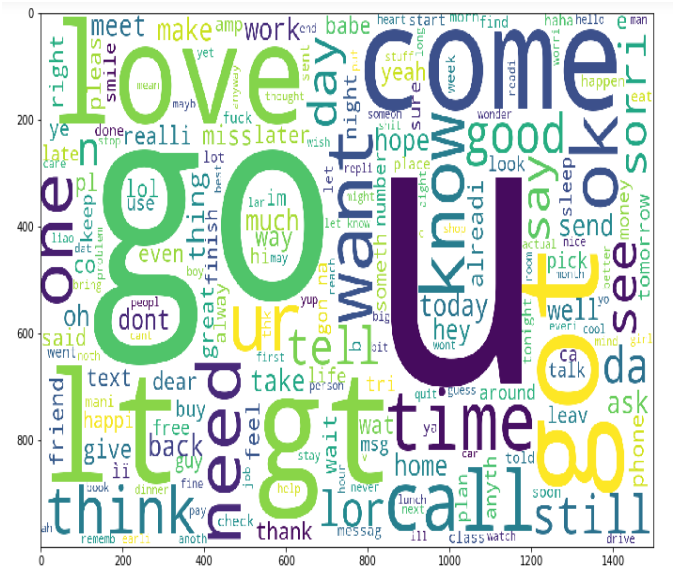

Fig. 9: Word cloud for non-fraud corpus

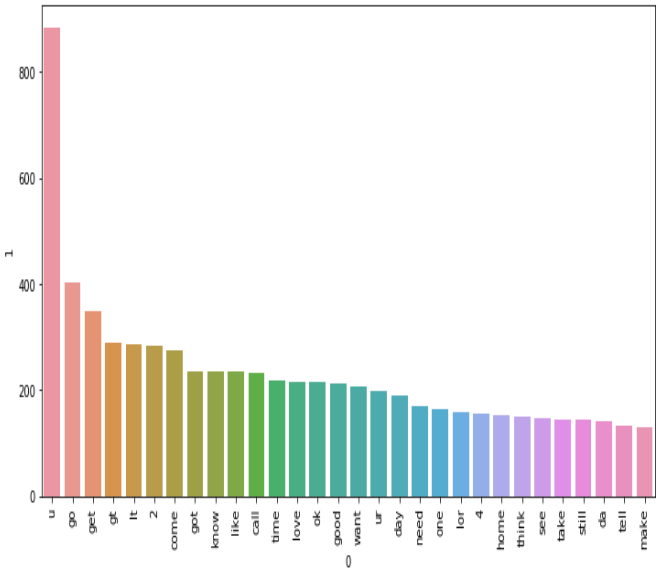

Fig. 10: Bar plot for non-fraud corpus

VII. RESULTS AND DISCUSSIONS

Several machine learning algorithms were used to train and test the fraudulent detection model using an SMS spam dataset in this research. Three of the eleven algorithms used, namely K-Nearest Neighbor, Naive Bayesian Classifier, and Random Forest Classifier, outperformed the others based on their respective performance accuracy and precision of 90% and 100% for K-Nearest Neighbor, 96% and 100% for Naive Bayesian Classifier, and 97% and 100% for Random Forest Classifier, respectively as shown in figure 11. For the optimal feature selection and extraction, the Count Vectorizer technique was used.

ISSN No:-2456-2165

Algorithm	Accuracy	Precision
KNN	0.900387	1.000000
NBC	0.959381	1.000000
RFC	0.971954	1.000000
ETC	0.972921	0.982456
SVC	0.972921	0.974138
AdaBoost	0.961315	0.945455
LRC	0.951644	0.940000
XGB	0.970019	0.934959
GBC	0.952611	0.923810
BC	0.958414	0.862595
DTC	0.936170	0.846154

Fig. 11: Accuracy and precision of the respective Algorithms used sorted based on precision

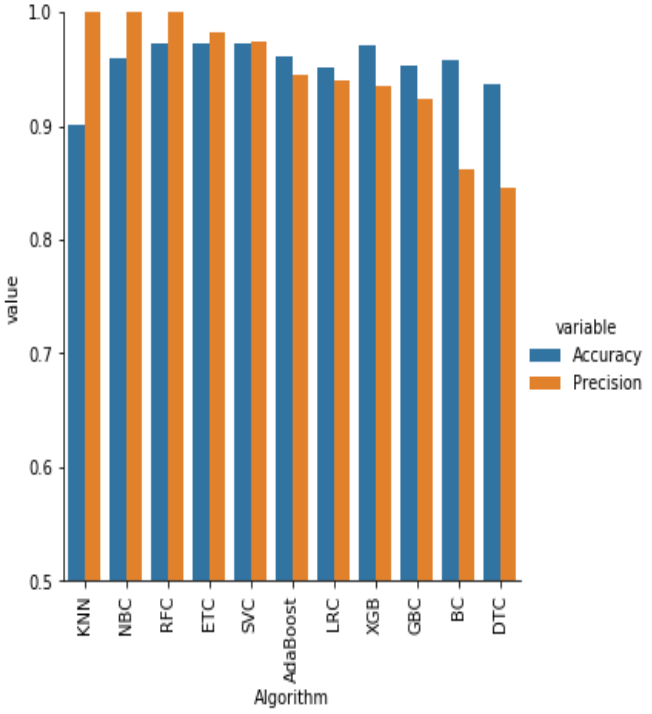

Fig. 12: Accuracy and precision comparison of the respective Algorithms used

K-Nearest Neighbor, Naive Bayesian Classifier, and Random Forest Classifier, outperformed the others based on their respective performance accuracy and precision of 91% and 100% for K-Nearest Neighbor, 97% and 100% for Naive Bayesian Classifier, and 96% and 98% for Random Forest Classifier, respectively as shown in figure 13. For the optimal feature selection and extraction, the TF-IDF Vectorizer technique and maximum features of 3000 were used.

Algorithm	Accuracy_max_ft_3000	Precision_max_ft_3000
KNN	0.905222	1.000000
NBC	0.970986	1.000000
RFC	0.975822	0.982906
SVC	0.975822	0.974790
ETC	0.974855	0.974576
LRC	0.958414	0.970297
XGB	0.967118	0.933333
AdaBoost	0.960348	0.929204
GBC	0.946809	0.919192
BC	0.958414	0.868217
DTC	0.927466	0.811881

Fig. 13: Accuracy and precision of the respective Algorithms used sorted based on precision using 3000 maximum features

Random Forest Classifier, K-Nearest Neighbor, and ExtraTrees Classifier, outperformed the others based on their respective performance accuracy and precision of 98% and98% for Random Forest Classifier, 91% and 98% for K-Nearest Neighbor Classifier, and 97% and 97% for ExtraTrees Classifier, respectively as shown in figure 14. For the optimal feature selection and extraction, the TF-IDF Vectorizer technique and maximum features of 3000 were used after features scaling was done using min-max scaler.

Algorithm	Scaled_Accuracy	Scaled_Precision
RFC	0.975822	0.982906
KNN	0.905222	0.976190
ETC	0.974855	0.974576
LRC	0.967118	0.964286
NBC	0.978723	0.946154
XGB	0.967118	0.933333
AdaBoost	0.960348	0.929204
SVC	0.970019	0.928000
GBC	0.946809	0.919192
BC	0.958414	0.868217
DTC	0.927466	0.811881

Fig. 14:Accuracy and precision of the respective Algorithms used sorted based on precision using 3000 maximum features and min-max features scaling

VIII. OUTCOME OF THE RESEARCH

The final hybrid model report is shown below:

```
Accuracy :  0.9709864603481625
Precision :  1.0
[[896   0]
 [ 30 108]]
              precision    recall  f1-score   support

           0       0.97      1.00      0.98       896
           1       1.00      0.78      0.88       138

    accuracy                           0.97      1034
   macro avg       0.98      0.89      0.93      1034
weighted avg       0.97      0.97      0.97      1034
```

Fig. 15: Optimal model performance report

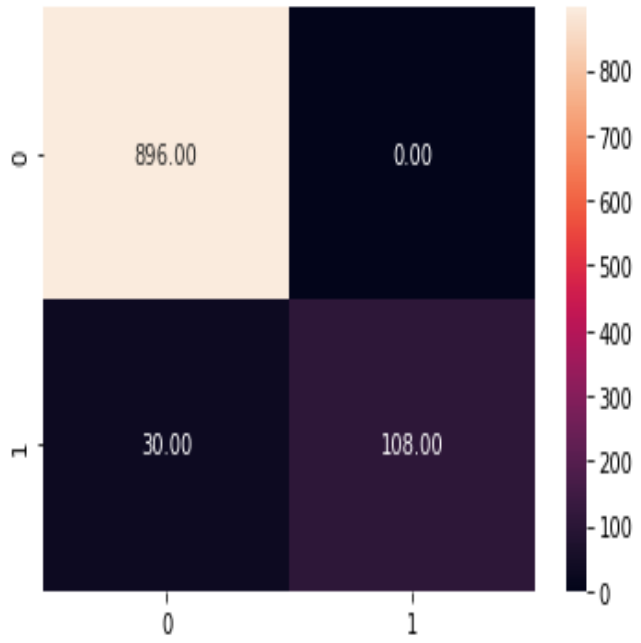

Fig. 16: Confusion matrix report of the optimal model

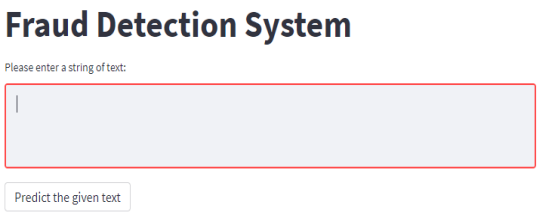

Fig. 17: Screenshot of the fraud detection system

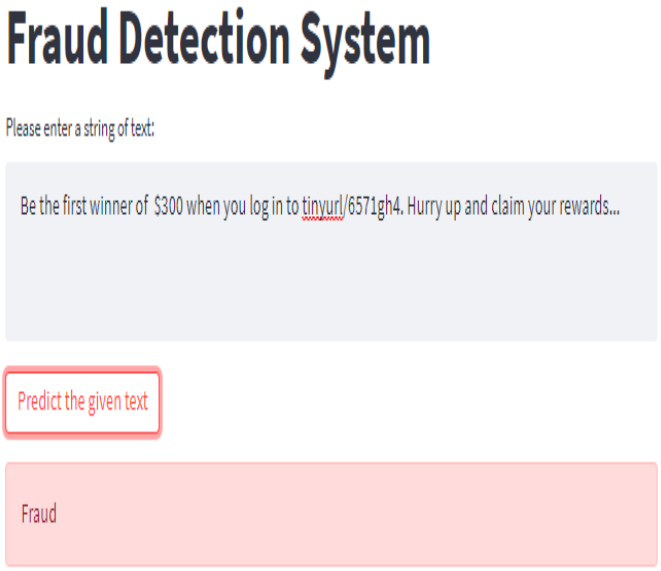

Fig. 18: Text string detected as fraud

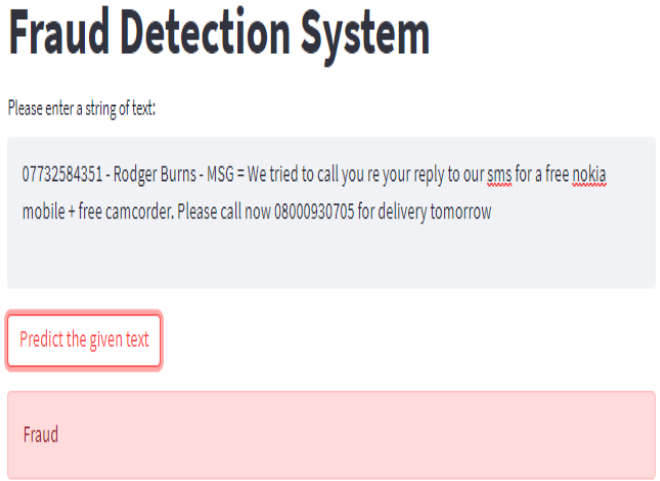

Fig. 19: Text string detected as fraud

ISSN No:-2456-2165

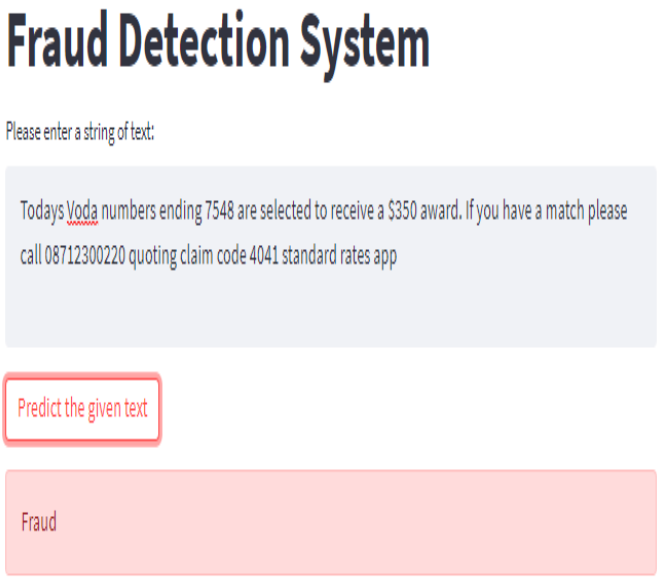

Fig. 20: Text string detected as fraud

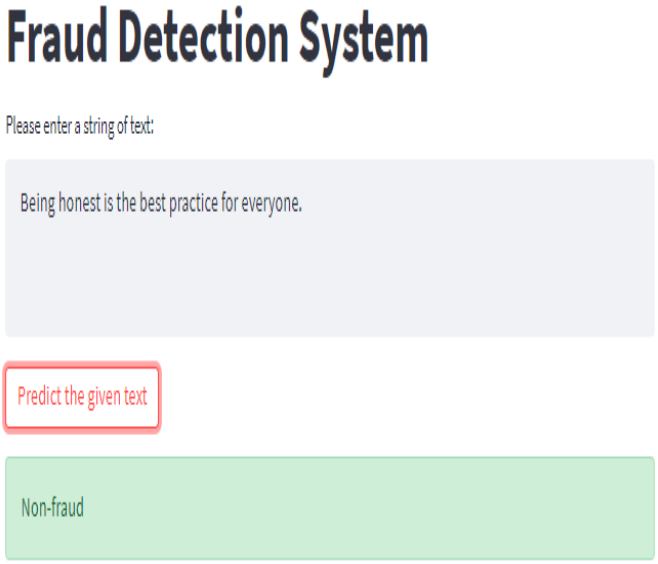

Fig. 21: Text string detected as non-fraud

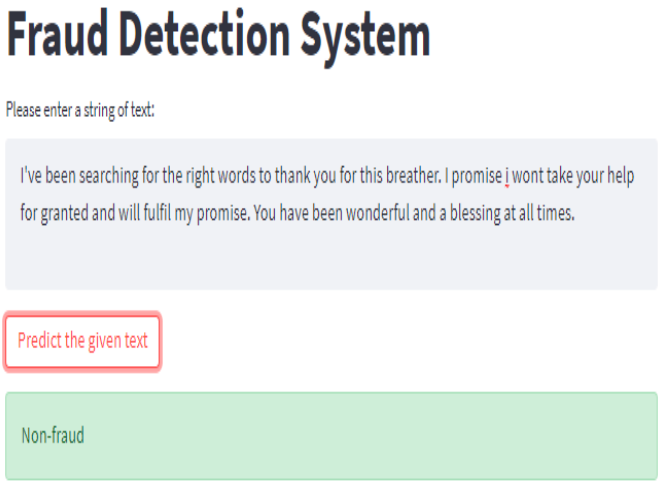

Fig. 22: Text string detected as non-fraud

Fig. 23: Text string detected as non-fraud

IX. FUTURE RESEARCH

The working model will be improved in the future by extending this work to include deep learning approaches, as deep learning is becoming more popular due to its adaptability and flexibility in the ecosystem.

X. CONCLUSION

The detection of fraudulent content within a text string is a challenging task. The continually adapting and complex nature of fraudulent activities necessitates the application of the latest technologies to confront fraud. This research focused on using the state-of-the-art machine learning model to build a fraud detection system. The fraudulent detection model was trained and tested using many machine learning algorithms, utilizing an SMS spam dataset in this study. The final optimal model performance obtained was 97% accuracy and 100% precision using accuracy, precision, recall, and f1-measure as metrics. The results obtained are promising, and the model was deployed using the streamlit framework.

REFERENCES

[1.] Y. YANG, R. CHEN, X. BAI and D. CHEN, "Finance Fraud Detection with Neural Network", E3S Web of Conferences, vol. 214, p. 03005, 2020. Available: 10.1051/e3sconf/202021403005.

[2.] B. Pragna and M. RamaBai, "Spam Detection using NLP Techniques", International Journal of Recent Technology and Engineering, vol. 8, no. 211, pp. 2423-2426, 2019. Available: 10.35940/ijrte.b1280.0982s1119.

[3.] I. SADGALI, N. SAEL and F. BENABBOU, "Performance of machine learning techniques in the detection of financial frauds", Procedia Computer Science, vol. 148, pp. 45-54, 2019. Available: 10.1016/j.procs.2019.01.007.

[4.] T.S Dhivya, S Nithya, S. G. Priya, E. Pugazhendi, 2021, Email Spam Detection and Data Optimization

using NLP Techniques, International Journal Of Engineering Research & Technology (IJERT) Volume 10, Issue 08 (August 2021).

[5.] A. U. Haq, L.GuangJun, S.N.,Habib and U. Khan (2020), "Spam Detection Approach for Secure Mobile Message Communication Using Machine Learning Algorithms", Security and Communication Networks Volume 2020, pp. 1-6, Article ID 8873639, DOI 10.1155/2020/8873639.

[6.] G. Liu, Q. Yang and X.Niu (2020), "OpinionRank: Trustworthy Website Detection using Three Valued Subjective Logic",IEEETransactions on Big Data, pp. 1-1 DOI 10.1109/TBDATA.2020.2994309.

[7.] A. Sharma and H. Kaur (2016), "Improved email spam classification method using integrated particle swarm optimization and decision tree." In Next Generation Computing Technologies (NGCT), 2nd International Conference on pp. 516-521, DOI: 10.1109/NGCT.2016.7877470.