International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 01 | Jan 2023 www.irjet.net p-ISSN: 2395-0072

A GUI BASED GRADING APPROACH FOR QUORA QUERIES AND MESSAGES USING MACHINE LEARNING TECHNIQUES

Dr. M Srinivasa Sesha Sai[1], Gopi Bapanapalli[2], Vineeth Marri[3], Lokesh Vadlamudi[4]

[1]Professor,Department of Information Technology, KKR & KSR Institute of Technology and Sciences, Guntur, India.
[2,3,4]Undergraduate Students, Department of Information Technology, KKR & KSR Institute of Technology and Sciences, Guntur, India.

Abstract - *Humans are powerful and smart enough to invent new technologies such as blockchain, cloud computing, web applications like Quora, Stack Overflow and more. Today, people are expected to take advantage of the internet to send messages and ask unknown questions on Quora. Therefore, our team focused on classifying whether a given message is a useful message or useless message and reviews the questions asked on the Quora to ensure that they are genuine and decided that it was a dishonest question. The existing system has poor accuracy in the grading of questions and messages. The proposed system eliminates the shortcomings and improves accuracy by up to 99.87%. The proposed system uses different kinds of machine learning techniques, such as Naive Bayes, Logistic Regression, Support Vector Machine algorithms.*

Key Words: Naïve Bayes, Support Vector Machine, spam, Insincere questions, Regression, Flask, Python, web application, Server, ham, Sincere.

1.INTRODUCTION

With so many new technologies emerging today, it is most widely used to send messages on Android apps such as WhatsApp, Telegram, Instagram, Twitter, and the Messages app on mobile phones. Today, using the Internet, we can send any number of messages from one place to another, so we need to know if the message from the other side is a useful message. With technology, spammers are ready to steal data, money and useful his information by sending spam his messages to his mobile application apps like messages and emails. Therefore, it is important to evaluate the message in order to save the data. Another important thing is to classify the questions asked on the Quora website as honest or dishonest. Because people are so smart you can ask any kind of question about it. or false. There are so many solutions for this, for example, [1] Abeer Alsadoon propose a solution using a Gain and mining algorithm that gives 100% accuracy but the problem with that implementation is the time taken for classifying the email whether it is spam or not and he is not implementing the Quora query classification using same algorithms but in our proposed using regression algorithms we can classify the SMS messages and Quora queries whether they are useful or not. Let us dive into another solution for this problem which was proposed by the [3] Aakash Atul Alurkar, the author wants to classify only email messages as spam or not spam but what about user interaction mode because customer satisfaction is the major priority of us, we need to satisfy the customer so for consumer purpose, our proposed system developed a GUI for interaction with the computer system by building a web application using trending technology like python framework flask. Our proposed system uses different types of machine learning technologies and deploys those algorithms into a fully pledged web application for providing a user interface. Infidelity is defined as words and actions that people do not really feel, have no meaning to, or are not based on their true feelings. Fraud is one of the most serious problems in Internet forums, especially Q&A forums. This is because it can affect the quality of Internet forums. Popular content (messages, posts, etc.) typically does not follow forum rules and can annoy other users. A machine learning model was implemented to detect questions from users in Q&A forums. Therefore, our goal is to develop a GUI based web application for layering queries made in the Quora application and messages received in the app. Next, we have also a feature scope for extending this proposed system to next level for security of consumer or customer or user by using trending technologies in the real world so that we make application which can satisfy the user. Today we have so many technologies been there for sending spam messages and steal the user's data by using those spam messages so it is important to know the messages coming to us are spam or not and protect our self from the spammers.

2. LITERATURE SURVEY

[2] Muhana Magboul Ali Muslam, Mansoor RAZA and Nathali Dilshani Jayasinghe proposed system for classifying the mails as spam or ham using K- means Clustering machine learning algorithm, K-nearest Neighbour machine learning algorithm (KNN), Decision tree as result Out of the total emails, more than 55 percent is identified as spam. But when consider the time taken to estimate or classify the message is spam or ham is very high and they can't classify the Quora queries but out proposed system is able classify both messages and Quora queries with less time and good accuracy.

[3] Aakash Atul Alurkar, Shreeya Vijay Joshi, Siddhesh Sanjay Ranade, Piyush A. Sonewar, Sourabh Bharat Ranade proposed a system for detecting the message is ham or not so in that they did using machine learning techniques they conclude that We then propose to classify emails into spam and ham using a machine learning approach. This allows the algorithm to detect desirable characteristics more accurately than manually setting desirable characteristics. conduct. The main idea is to classify the user's incoming mail based on various parameters commonly used by spammers. Its main purpose is to group important emails and block spam emails. With a variety of simple Internet domains openly available, it is a waste of effort for system administrators to block potential spammers from predefinedlists. This document also considers email bodies containing commonly used keywords and punctuation.

[7] Hendri Priyambowo, Mirna Adriani proposed a system for stratification of Quora questions as sincere or insincere using Nearest Neighbor, Decision Tree, Random Forest. The purpose of this study is to compare machine learning algorithms and find out which algorithm and function can provide the best results in detecting dishonest question task. In this study, we also want to explore how feature selection techniques affect classification results on average, unigram features were basically the most important features used in classification tasks. Combine Unigram features with other features. It uses deep learning technology.

Additionally, some experiments, such as detecting dishonest questions, suffer from imbalanced data, so data resampling was one of the challenges in this experiment. Therefore, a comparison with other his data resampling algorithms can be made for another re-search.

[6] Suresh Babu, C. V. Guru Rao, P. U. Anita proposed system for classifying the mails as spam mails and ham mails using Neighbor Probability based Naïve Bayes Algorithm. This system takes so much of time to classify the message as useful or useless and it can't classify the Quora questions. But our proposed system can do both classifications using regression algorithms.

[5] De Rosal Ignatius Moses Setiadi, Christy Atika Sari, Eko Hari Rachmawanto, Niken Larasati Octaviani proposed a system for grading the email spams using Multinomial Naive Bayes Classifier, Support Vector Machine, and Recurrent Neural Network. The accuracy obtained by using the MNBC algorithm is 93%, while those using the SVM algorithm is 96%, and using the RNN algorithm is 74%. The method and algorithm that has the highest or best accuracy results is the machine learning method using the Support Vector Machine algorithm with an accuracy of 0.96 or 96%. But this system also only classifies the mails but not Quora queries. Our proposed system can do both classifications.

[4] D. Karthika Renuka, V. Sri Vinitha proposed system of Performance Analysis of E-Mail Spam Classification using different Machine Learning Techniques. This system uses the following algorithms like K-Nearest Neighbor, Naive Bayes, Artificial Neural Network, Support Vector Machine, and Random Forests algorithm.

3. PROPOSED SYSTEM

There are so many technologies in this dynamic world, and there are pros and cons to using them. Some drawbacks are sending deceptive her messages to users and using spam messages to steal user data. Another point is to ask dishonest questions on the Quora website. Users cannot identify messages sent by hackers or spammers, so our team focuses on that, identifying SMS spam messages and classifying messages posted on the Quora web application. develop a system to error. The proposed system uses various types of his machine learning techniques, such as the machine learning-based test classification approach. The downside of the survival system is accuracy and time complexity. The proposed system has excellent accuracy and time complexity compared with existing systems.

3.1 STEPS INVOLVED IN PROPOSED SYSTEM

1. Gathering the data set information
2. Preprocessing
3. Train the model
4. Deploy the model into web pages

3.1.1 Gathering the data set information

The dataset is downloaded from the Kaggle. The data sets considered in this project are SMS Spam Collection and Quora insincere dataset. The datasets have 1 feature and 1 target. The target is binary type like either class1 or class2. All the target labels are equally distributed for the training and testing data. 75% data is used for training, remaining for testing.

3.1.2 preprocessing

Preprocessing is one of the most important steps for any kind of text model. Required preprocessing includes lowercase conversion, punctuation removal, stop word removal, and lemmatization/de-emphasis. Feature extraction applied to data. Convert the data into vector form so that it can be understood by machines.

3.1.3 Train the model

After performing feature extraction, the data will be going to train the model to predict the results. How the training is performed and implemented using Machine Learning algorithms.

3.1.4 deploy the model into web pages

Flask is a micro web framework written in Python. It is classified as a micro framework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where preexisting third-party libraries provide common functions.
Finally to run the model have to use app.run() method.

3.2 Proposed System Architecture

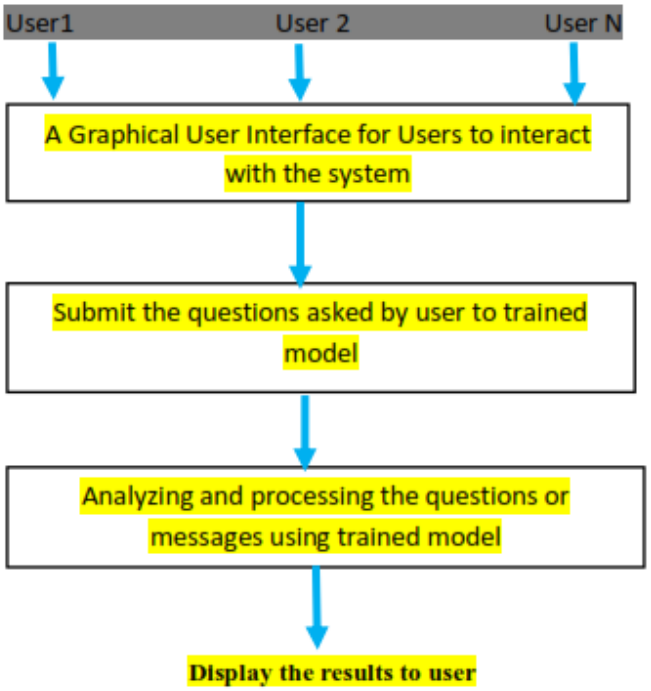

Fig 3.2.1 Architecture of proposed system.

The above diagram shows the architecture of proposed system. It is three tier architecture and it contains input taking and processing the input then give the processing input to trained model for getting the accurate results. We can develop the trained model using data set called SMS Spam Collection and machine learning algorithms like Naïve Bayes classifier, Non-Probabilistic algorithm and regression algorithms like logistic regression. After getting the results we need to deploy those results into a web page. In order to display the results, we used scripting languages like html, cascading style sheets, Python framework flask. Using Flask, we can develop a web application which runs on the local server or non-local server for providing the interface for the user. This architecture accepts multiple users and give accurate results to the user. In order to display the results our team uses the cascading style sheets for styling the website or web page. Finally, by doing the all steps we can get the correct results. We can also get the less time for classifying the messages and queries into two groups using proposed system. The architecture involves the deploying a machine learning algorithm into a well good web application using the flask framework.

3.3 Classifying the messages

We can classify or grade the SMS messages into two types typically they are useful messages and not use full messages or junk messages and ham messages. By using some set words or parameters we can classify them as two groups. Those parameters are shown below.

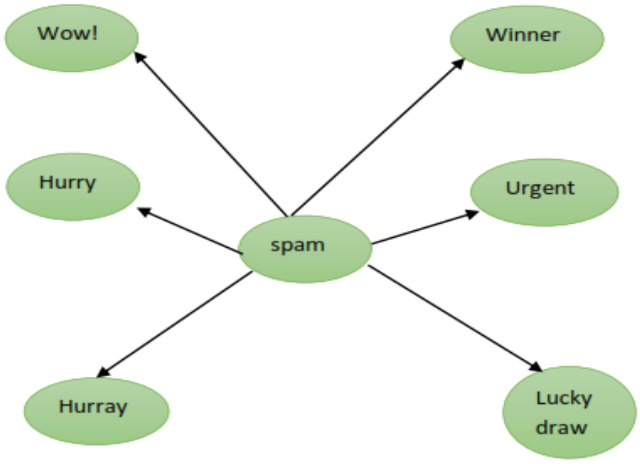

Fig 3.3.1 keywords for junk messages

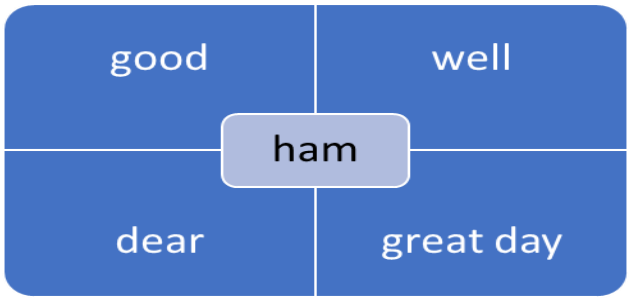

Fig 3.3.2 keywords for ham messages

3.4 Classification of Quora queries

Fig 3.4.1 Indicates the Sincere questions list.

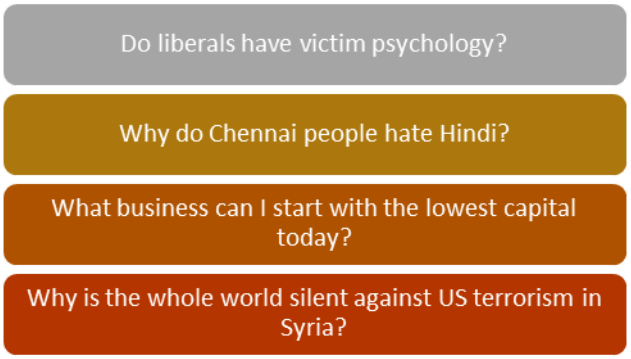

Fig 3.4.2 Indicates some Insincere questions.

4. IMPLEMENTATION

Implementation of proposed system involves the learning different types of machine learning algorithms like logistic regression, support vector machine and Naïve Bayes classifier.

Logistic Regression

Logistic regression is one of the most popular supervised machine learning algorithms used for classification and solving the regression problems. Algorithm for classifying the Quora queries and SMS messages:

Step1: Read the data from csv file (typically a data set).

Step2: Vectorizing the data.

Step3: Split the data using input data.

Step4: Choose the logistic regression classifier.

Step5: Fit the trained data using classifier.

Step7: Predict the data using classifier.

Step8: Calculating the accuracy and time.

Step9: Plot the confusion matrix.

Step10: Visualize the data.

Similarly, we can implement the algorithms of support vector machine and naïve bayes classifier for grading the Quora queries and SMS messages. The difference is to a choose a classifier with respect the algorithm which we want to develop. After doing the all-algorithms implementation we need to analyze the results of 3 algorithms for better accuracy and better time complexity or better throughput. Naive Bayes is a simple learning algorithm that uses Bayes' law with the strong assumption that the attributes of a given class are conditionally independent. This independence assumption is often violated in practice, yet Naive Bayes often provides competitive classification accuracy.A support vector machine (SVM) is a supervised machine learning model that uses a classification algorithm for the two-group classification problem. After giving the SVM model a set of categorically labelled training data, new text can be classified.

5. RESULTS

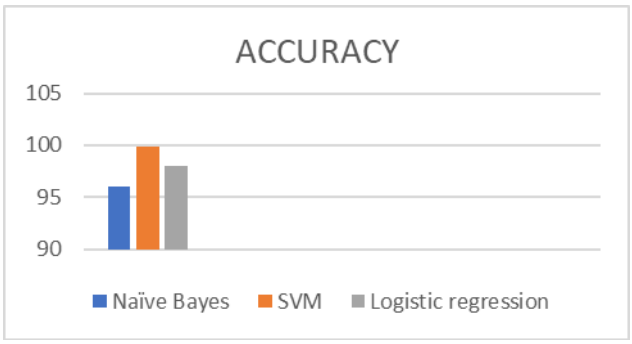

Fig5.1 Accuracy of different machine learning algorithms using proposed model.

Classifier	Accuracy	Precision	Time(sec)	Recall
Naïve Bayesian	96%	0.95	1.2	0.87
Support Vector Machine	99.87%	0.98	0.12	0.96
Logistic Regression	98%	0.97	0.20	0.88

Table5.1 Indicates the machine learning attributes for the proposed system.

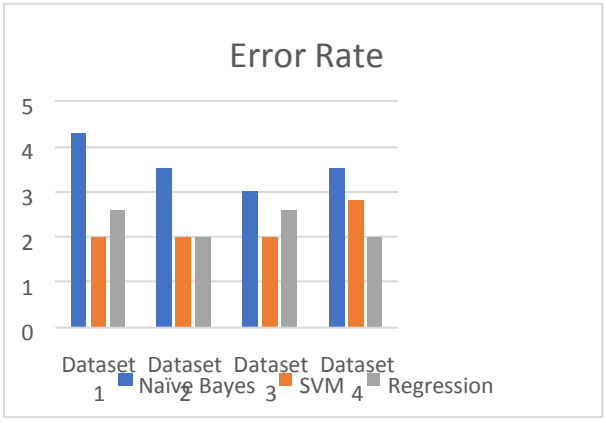

Fig5.2 Error Rate for different algorithms using different datasets.

6. CONCLUSIONS

We have used different types of machine learning algorithms like non-probabilistic, regression and Support vector machine for classifying the junk messages as well as Insincere questions asked on the Quora web application. We get good accuracy with in the less span of time. We achieved a less error rate for all the algorithms. Now our proposed system able classify the both junk messages and Quora queries at a time with good accuracy of 99.87%.

In Future we try to Integrate a decentralized application for classifying the messages and Quora queries using Block Chain technology.

7. REFERENCES

1. K.Chae, A.Alsadoon, P.W.C.Prasad and S.Sreedharan, "Spam filtering email classification (SFECM) using gain and graph mining algorithm," 2017 2nd International Conference on AntiCyber Crimes (ICACC), 2017, pp. 217-222,doi: 10.1109/AntiCybercrime.2017.7905294.

2. M.RAZA, D.Jayasinghe and M.M.A Muslam "A Comprehensive Review on Email Spam Classification using Machine Learning Algorithms," 2021 International Conference on Informaton Networking (ICOIN), 2021,pp-327-332, doi: 10.1109/ICOIN50884.2021.9334020.

3. A.A.Alurkar et al., "A proposed data science approach for email spam classification using machine learning techniques," 2017 Internet of Things Business Models, Users, and Networks, 2017,pp.15, doi:10.1109/CTTE.2017.8260935.

4. V.S.Vinitha and D.K.Renuka, "Performance Analysis of EMail Spam Classification using different Machine Learning Techniques," 2019 International Conference on Advances in Computing and Communication Engineering (ICACCE), 2019, pp.15,doi: 10.1109/ICACCE46606.2019.9080000.

5. N.L.Octaviani, E.Hari Rachmawanto, C.A.Sari and I.M.S. De Rosal, "Comparison of Multinomial Naïve Bayes Classifier, Support Vector Machine, and Recurrent Neural Network to Classify Email Spams," 2020 International Seminar on Application for Technology of Information and Communication(iSemantic), 2020, pp. 17-21, doi: 10.1109/iSemantic50169.2020.9234296.

6. P.U.Anitha, C.V.G.Rao and S.Babu, "Email spam classification using neighbor probability based Naïve Bayes algorithm," 2017 7th International Conference on Communication Systems and Network Technologies (CSNT), 2017, p.350-355, doi: 0.1109/CSNT.2017. 8418565.

7. H. Priyambowo and M. Adriani, "Insincere Question Classification on Question Answering Forum," 2019 International Conference on Electrical Engineering and Informatics (ICEEI), 2019, pp. 390-394,doi: 10.1109/ICEEI47359.2019.8988798.

8. A. S. Nagdeve and M. M. Ambekar, "Spam Detection by designing Machine Learning approach in Twitter Stream," 2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC), 2020, pp. 126-130, doi: 10.1109/ICSIDEMPC49020.2020.9299607.

9. N. Jatana and K. Sharma, "Bayesian spam classification: Time efficient radix encoded fragmented database approach," 2014 International Conference on Computing for Sustainable Global Development (INDIACom), 2014, pp. 939-942, doi: 10.1109/IndiaCom.2014.6828102.