Mini-Project – 2B Web based on ML (ITM 601)

END-TERM EVALUATION

SMS Spam Detector

T. E. Information Technology

By

Raj Jaiswal 57
Ashish Yadav 58
Allan Rodrigues 59
Jonathan Sardinha 60

Under Guidance of
Dr Minal Lopes

Department of Information Technology
St. Francis Institute of Technology
(Engineering College)

University of Mumbai
2021-2022

CERTIFICATE

This is to certify that the project entitled **"SMS Spam Detector"** is a bonafide work of **"Raj Jaiswal, Ashish Yadav, Allan Rodrigues, Jonathan Sardinha" 57, 58, 59, 60** submitted to the University of Mumbai towards completion of mini project work for the subject of **Mini Project – 2B Web Based on ML (ITM 601).**

Dr. Minal Lopes
Supervisor/Guide

Dr. Joanne Gomes
HOD-IT

Examiners

1.----------

2.----------

Date:

DECLARATION

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

---------- ----------

(Name of student and Signature) (Name of student and Signature)

---------- ----------

(Name of student and Signature) (Name of student and Signature)

ABSTRACT

Over recent years, as the popularity of mobile phone devices has increased, Short Message Service (SMS) has grown into a multi-billion dollar industry. At the same time, reduction in the cost of messaging services has resulted in growth in unsolicited commercial advertisements (spams) being sent to mobile phones. In parts of Asia, up to 30% of text messages were spam in 2012. Lack of real databases for SMS spam, short length of messages and limited features, and their informal language are the factors that may cause the established email filtering algorithms to underperform in their classification.

Many SMS spam detection methods already exist and different classifiers such as Support Vector machine, Naïve Bayes, Linear Regression, Decision Tree, Random Forest and many other machine learning algorithms were used. In this project, we present an approach that can detect and filter the spam messages using Naive Bayes Classifier. The SMS spam collection data set is used for testing the method. The dataset is split into two categories for training and testing.

After calculating the performance metrics we observed that we achieved an accuracy score of 98.83% and a precision score of 95.27% which shows that this model gives accurate results while detecting ham or spam messages.

Index

List of Abbreviations

Sr. No.	Abbreviation	Full Form
1	SVM	Support Vector Machine
2	LR	Linear Regression
3	RF	Random Forest
4	MCC	Matthews correlation coefficient

List of Figures

Chapter 1
Introduction

1.1 Introduction

The popularity of mobile devices is increasing day by day as they provide a large variety of services by reducing the cost of services. Short Message Service (SMS) is considered one of the widely used communication services. However, this has led to an increase in mobile devices attacks like SMS Spam. SMS spams are one of the concerns and many people do not like to receive them since they are annoying. Spam emails and messages are unwanted for receivers which are sent to the users without their prior permission. It contains different forms such as adult content, selling items or services, and so on. The spam has increased these days due to more mobile devices deployed in the environment for e-mail and message communication. Currently, 85% of mails and messages received by mobile users are spam. Due to these spam mails and messages, the valuable e-mails and messages are affected because each user has limited Internet services, short time, and memory.

1.2 Scope

Many SMS spam detection methods already exist and different classifiers were used, such classifiers depended on Support Vector machine, Naïve Bayes, Linear Regression, Decision Tree, Random Forest and many other machine learning algorithms. In response, SMS spammers have been adapting their strategies in increasingly innovative ways. Consequently,more effective approaches are needed to detect and filter SMS spam automatically and accurately.

In this project, we present an approach that can detect and filter the spam messages using Naive Bayes Classifier. The SMS spam collection data set is used for testing the method. The dataset is split into two categories for training and testing.

1.3 Objectives and Problem Statement

With digital currency many people are prone to smishing attacks in which the fraudster sends fake SMS messages pretending to be a reputable company in order to get individuals to reveal their personal information, passwords or credit card details.

A spam email often contains some indicative keywords, such as “free” or “awards,” or unusual distribution of punctuation marks and capital letters,such as “BUY!!” or “MONEY,” such that these keywords become important features that a machine-learning-based classification
algorithm can use.

- Our SMS Spam Detector checks the SMS entered across various messages in a dataset and gives a “Clear” message if the SMS is original and sent from a genuine organization and it will give a warning “Alert” message if the SMS is sent from a fraudster pretending to be a genuine organization

- This will help our users to differentiate between spam and actual messages thus saving them from fraud.

Chapter 2
Literature Review

Reference Paper [1]: SMS Spam Detection Using Non Content Features
Algorithm: In this project spam detection was done using features that include temporal and graph topology information but exclude content, thus addressing user privacy issues. Focus is on identifying professional spammers on the basis of overall message-sending patterns. Finding SMS spam is done on the server side, given that client-side detection requires mostly content-and ID-based solutions. Algorithm used is support vector machine (SVM) and k -nearest neighbor (k -NN) algorithms

Review: Experiment results show that by using only a set of static features to train the classifier , a performance of AUC of 88.3% is achieved. If temporal and network features are used individually, the AUC can get additional 7 and 8 percent improvements, respectively.
These results indicate that, compared with static features, network properties and temporal information in an SMS communication network can indeed help achieve better performance

Reference Paper [2]: SMS Spam Detection using Machine Learning Approach

Algorithm: The highest accuracy achieved is from Multinomial NB with the accuracy of 98.88%.The second highest accuracy from SVM is 98.86% close to Multinomial NB. The other three algorithms were k-nearest neighbor, random forests and adaboost with decision trees with accuracy of 97.47%, 98.57% and 98.59%

Review: From simulation results, multinomial naive Bayes with laplace smoothing and SVM with linear kernel are among the best classifiers for SMS spam detection.The best classifier is the one utilizing SVM as the learning algorithm,which yields overall accuracy of 97.64%. Adding meaningful features such as the length of messages in number of characters, adding certain thresholds for the length, and analyzing the learning curves and misclassified data have been the factors that contributed to this improvement in results

Reference Paper [3]: SMS Spam Detection using H2O Framework

Algorithm: The H20 framework consists of a set of processes:
The first one is the selection of the dataset, then the features will be selected and extracted from the dataset. In the next process, the classification methods will be determined; this framework will use three classifiers: random forest, deep learning and naïve bayes moreover all the experiments will be made in H2O platform. After that, the evaluation metrics will be specified and comparison can be made between classifiers.

Review: Results showed that RF is the best algorithm in terms of precision, recall, f-measure and accuracy with 95%, 85%, 0.89% 0.97% values respectively, however RF is not the best in terms of time. On the other hand, NB is the best classifier in terms of runtime with 0.6 seconds.
Nevertheless, NB is the worst according to precision, recall, f-measure and accuracy

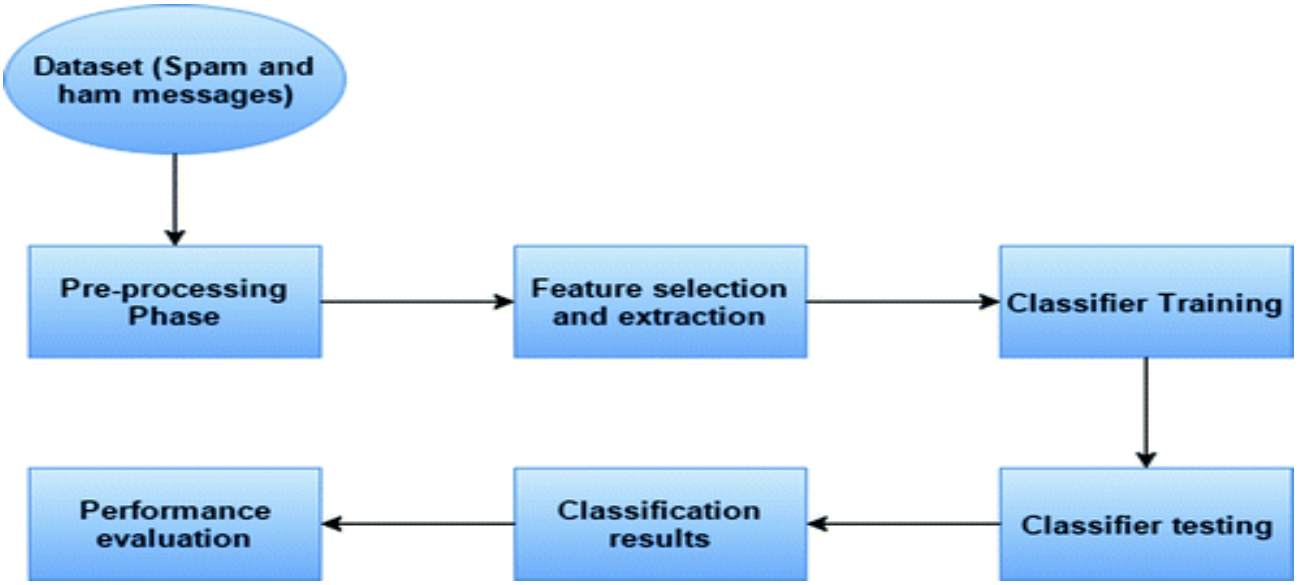
Reference Paper [4]: Spam Detection Approach for Secure Mobile Message Communication Using Machine Learning Algorithm

Algorithm: A spam detection method using machine learning algorithms such as Linear Regression, k-nearest neighbor, and decision tree for classification of ham and spam messages. The SMS spam collection dataset was considered for testing of the current research. The dataset was divided into two categories: 30% for testing and 70% for training purposes for the predictive models.

Review: The Linear Regression at hyperparameter C = 1 achieved 99% accuracy, 100% specificity, sensitivity 86%, and MCC, 93% and the processing time is 0.494 seconds. The classifier decision tree obtained 98% accuracy, 94% specificity, sensitivity 86%, and MCC 95%, and the processing time is 46.032 seconds. Similarly, the k-nearest neighbor classifier achieved 95% accuracy, 100% specificity, sensitivity 60%, and MCC 80%, and the processing time is 0.630 seconds. Experimental results suggest that LR is the best classifier for the classifications of ham and spam messages

Chapter 3
Proposed Work

3.1 System Diagram



3.1 System Working

3.2 System Implementation

Data Collection: Data Collection includes collecting data from various sources such as RDBMS, Data Warehouse which is relevant. In this phase we have used a dataset which has spam as well as ham messages from Kaggle repository.

Data Cleaning: In this phase we have cleansed all the data which was taken into consideration. We have removed all the white-spaces, lowered the alphabet so that words like Equal and equal become the same, remove the remaining punctuation, tokenized each message to represent the message as a list of words and done stemming, dropped null values and converted all the words to their root word, like floor, floor to floor.

Generating Testing and Training Data Sets: We split the dataset into training and testing. 70% of the dataset was used for training purposes and 30% of the dataset was used for testing purposes.

Applying Naïve Bayes Algorithm : After splitting the dataset into training and testing we have applied the Naïve Bayes Algorithm for training the Data. On this basis, the spam feature will be classified.

Prediction: After applying the Naive Bayes algorithm we tested how well our model predicts the ham or spam messages. We have given input messages to the model to check whether the message is SPAM or HAM.

Performance Evaluation: We observed the results obtained and calculated the performance metrics of our model. We calculated accuracy score, precision score, recall score and F1 score. Based on these calculations we could conclude if the prediction accuracy of our model is strong or weak.

Chapter 4
Implementation Details

4.1 Dataset Details

We have portrayed the work done by scientists in the field of distinguishing Texts highlights by concentrating exclusively on the field of SPAM recognizable proof. We have used a Machine Learning Algorithm in recognizing the SPAM and HAM subsequent to gathering the dataset from Kaggle. The dataset had three rows: the 1st row had the count of messages, the 2nd had the indication whether it was a spam or a ham message and the 3rd row had the message. This spam dataset has a total ham percentage of 87% and a total spam percentage of 13% and it contains roughly 5169 unique values.

4.2 Algorithm Details

Naive Bayes classifier: Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other.

Bayes' Theorem finds the probability of an event occurring given the probability of another event that has already occurred. Bayes' theorem is stated mathematically as the following equation:

$$P(A/B) = \frac{P(B/A)\,P(A)}{P(B)}$$

where A and B are events and P(B) ≠ 0.

- We are trying to find the probability of event A, given the event B is true. Event B is also termed as evidence.
- P(A) is the **priori** of A (the prior probability, i.e. Probability of event before evidence is seen). The evidence is an attribute value of an unknown instance(here, it is event B).
- P(A|B) is a posteriori probability of B, i.e. probability of event after evidence is seen.

NB algorithm is applied to the final extracted features. The speed and simplicity along with high accuracy of this algorithm makes it a desirable classifier for spam detection problems.

4.3 Performance Metric Details

Applying naive Bayes with multinomial event model gave the following results:

1. **Accuracy Score:** It is a ratio of a ratio of correctly predicted observations to the total number of observations. We achieved an accuracy score of 98.83% which is a good accuracy.
2. **Precision Score:** Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. The ideal precision score is close to 1 and we achieved a precision score of 95.27%.
3. **Recall Score:** Recall is a metric that quantifies the number of correct positive predictions made out of all positive predictions that could have been made. It is the measure of completeness. We achieved a recall score of 95.91%.
4. **F1 score:** F1 score is defined as the harmonic mean between precision and recall. It is used as a statistical measure to rate performance. We achieved a recall score of 95.59%.

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP+FN}$$

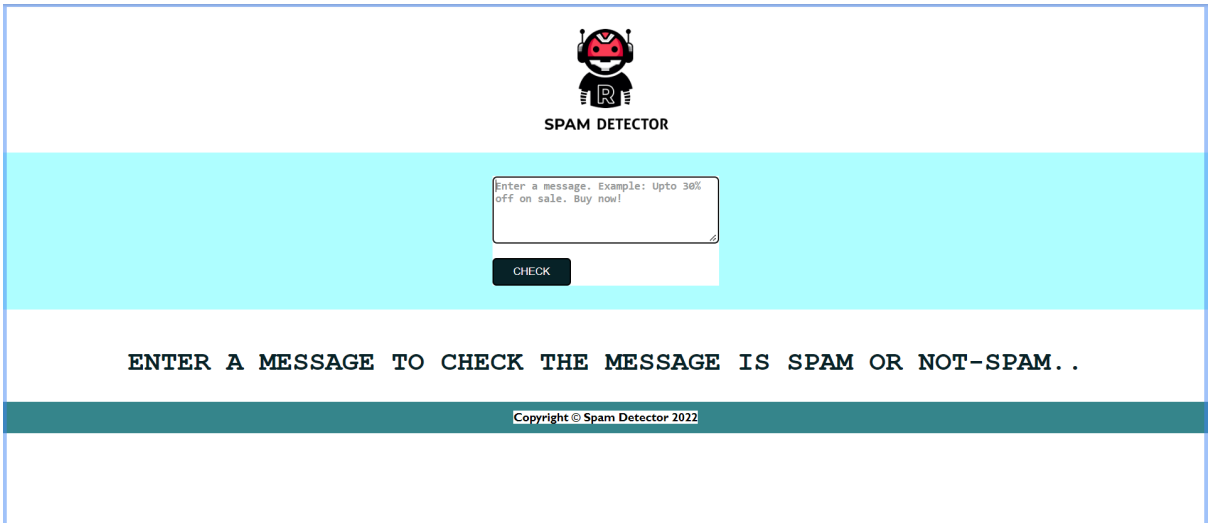
$$F1 = \frac{2\ X\ precision\ X\ recall}{precision + recall}$$

$$accuracy = \frac{TP + TN}{TP + FN + TN + FP}$$

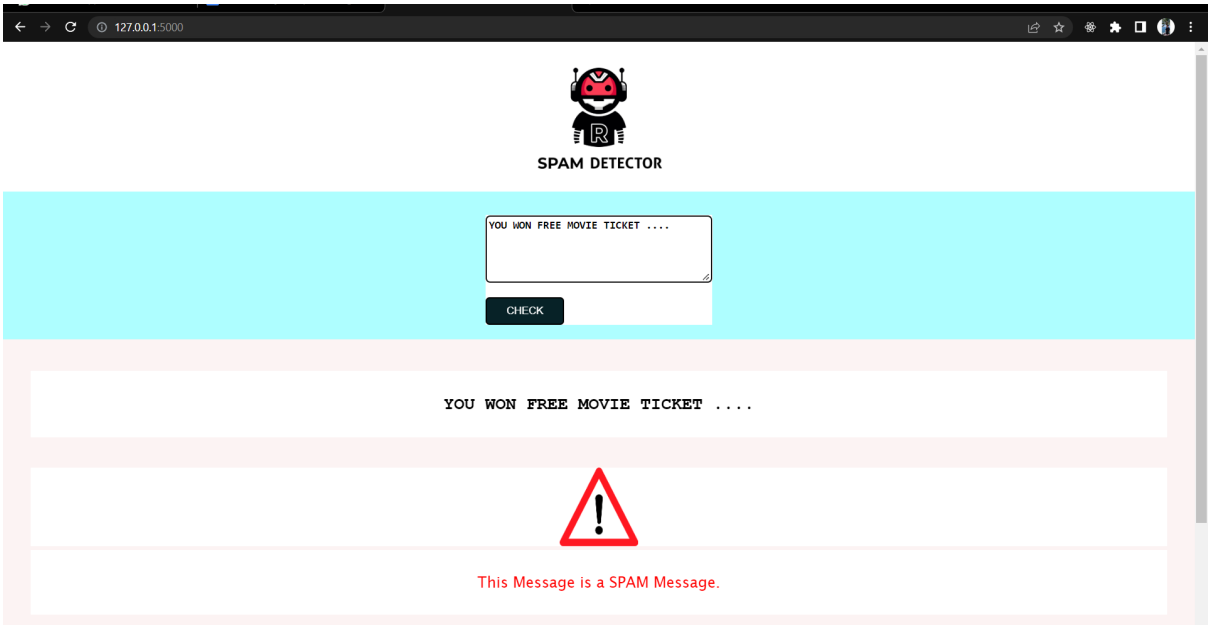
where

- TP = True Positive
- TN = True Negative
- FP = False Positive
- FN = False Negative

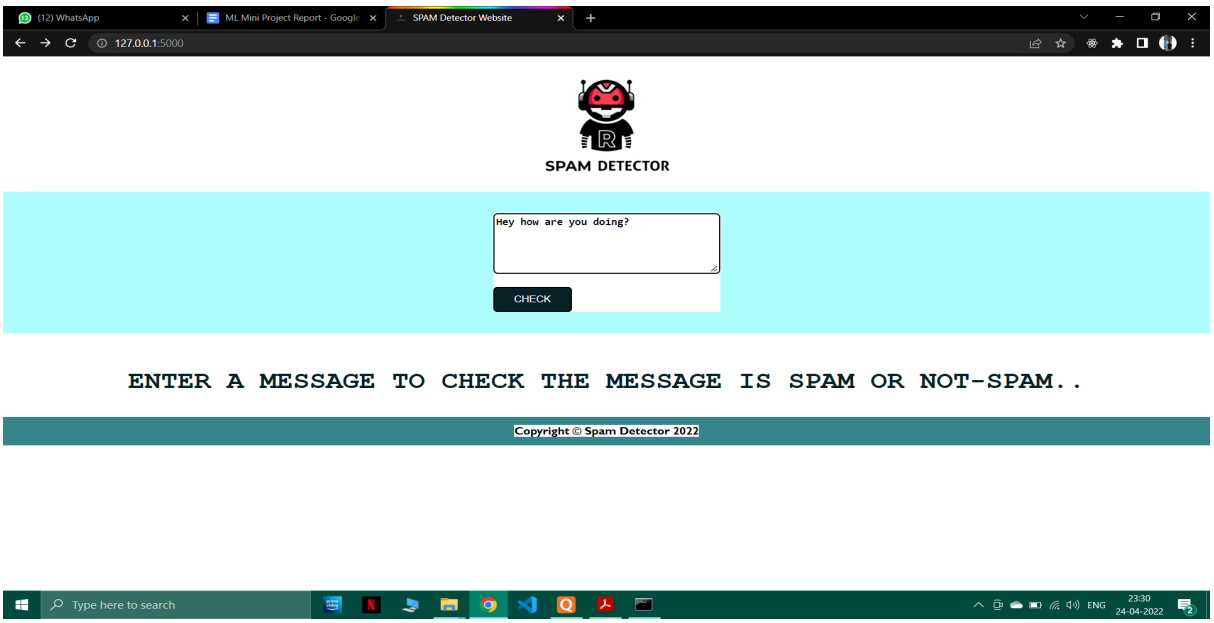
Chapter 5
Results



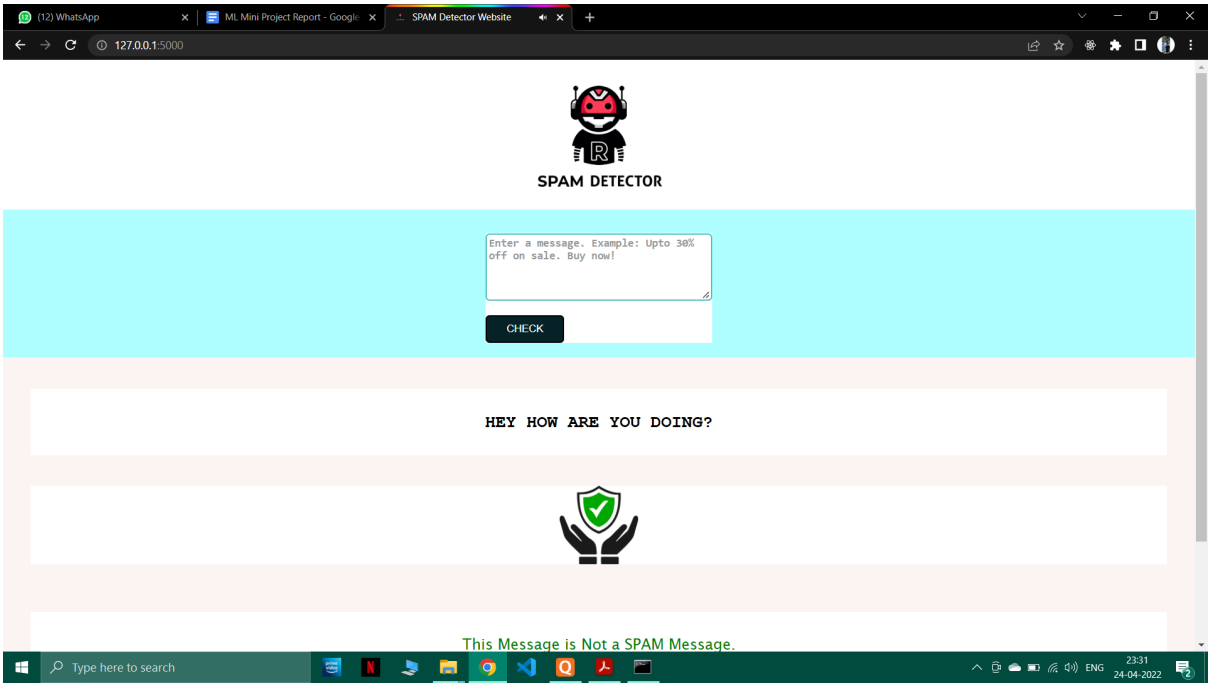
5.1 Spam Detector Home Page



5.2 Spam message



5.3 Ham message



Chapter 6
Conclusion

Thus we successfully built a model that can help our users to differentiate between spam SMS messages and actually genuine SMS messages thus saving them from online frauds and smishing attacks.
From the above discussion and experimentation we can conclude that machine learning algorithms can play a vital role in identifying SMS Spam. After calculating the performance metrics such as accuracy, precision, recall and F1 score we observed that the percentage of all these metrics was above 95% for Naives Bayes classifier with the ideal (strong) value being close to 1. Hence we can conclude that the Naive Bayes Classifier will give us quite accurate results and is a more than reliable machine learning algorithm to detect Spam messages.

Future Scope
There are many machine learning algorithms which can be used for SMS spam detection. Even though the results obtained by using Naive Bayes Classifier are quite reliable, we can obtain better or similar results using SVM (Support Vector Machine) Classifier. It is one of the most popularly used algorithms for classification as well as regression problems. Random Forest, Linear Regression and K Nearest Neighbor(KNN) are also viable options. Each algorithm has their own strong and weak points, some give better accuracy, precision while some give better runtime. Thus in the future this SMS Spam Detection model can be improved by using a more accurate classifier which yields even better results than our model.

References

[1] Xu, Qian, et al. "Sms spam detection using noncontent features." IEEE Intelligent Systems 27.6 (2012): 44-51.

[2] Shirani-Mehr, Houshmand. "SMS spam detection using machine learning approach." unpublished) http://cs229. stanford. edu/proj2013/Shir aniMeh r-SMSSpamDetectionUsingMachineLearningApproach. pdf (2013).

[3] Suleiman, Dima, and Ghazi Al-Naymat. "SMS spam detection using H2O framework." Procedia computer science 113 (2017): 154-161.

[4] GuangJun, Luo, et al. "Spam detection approach for secure mobile message communication using machine learning algorithms." Security and Communication Networks 2020 (2020).

Paper Publication

SMS Spam Detection Using Machine Learning

Ashish Yadav[1], **Allan Rodrigues**[2], **Jonathan Sardinha**[3],**Raj Jaiswal**[4]

[1,2]Department of Information Technology, ST Francis Institute of Technology, Borivali , 400103,Mumbai, India

Email:Sfit.ac.in

Abstract.

In the advanced reality where digitization is all over the place, SMS has become perhaps the most fundamental type of correspondence, not at all like other visiting based informing frameworks like Facebook, WhatsApp and so on, SMS doesn't need dynamic web association by any means. As we as a whole know that Hackers/Spammer attempts to barge in Mobile Computing Device, and SMS support for cell phones had become defenseless, as assailant attempts to interfere to the framework by sending undesirable connection, with which on clicking those interface the aggressor can acquire remote access over the versatile registering gadget. Thus, to distinguish those messages Authors have fostered a framework which will recognize such noxious messages and will distinguish regardless of whether the message is SPAM or HAM

Keywords: SMS, SPAM and HAM, Machine Learning

1. Introduction

SMS is one of the best types of correspondence. It depends on cell correspondence frameworks, simply the cell should be in the organization inclusion region to send or get the message. Nearly everybody is involving this help for correspondence. Different associations manage SMS for speaking with their clients/clients, banks and other government associations likewise use SMS for correspondence. Likewise, numerous business associations utilize this assistance for publicizing. Subsequently, SMS is assuming an indispensable part, as dynamic web association isn't needed by any stretch of the imagination in this system. So because of the enormous use of SMS, it has become perhaps the most loved place for programmers and spammers. It hushes up simple for a programmer to think twice about one's phone just by passing or communicating Malicious connect to end client, the cell phone will consequently be compromised assuming end client click on the connection or message being sent by

programmer/spammer, and we can know the rest how a programmer can take advantage of the framework on the off chance that he deals with the framework. So it has become particularly vital to confine the substance which the end client is getting. So there should be a framework which could see the end client regardless of whether the got message is SPAM, Non SPAM message is known as HAM. So by recognizing the previously mentioned issues and issues, we have fostered a framework which can distinguish whether a Message is SPAM or HAM in view of the substance of the message utilizing Machine Learning method. In this part we have given a short outline of Machine Learning; different kinds of Machine Learning and the methods we have utilized for fostering the Machine Learning Model Naive Bayes.

ML is comprehensively arranged into two classes: a) Supervised Learning3, b) Unsupervised Learning4. Fundamental Categories of Machine Learning: Supervised Learning: Supervised advancing otherwise called prescient displaying, is the method involved with making expectations utilizing information. Instances of Supervised learning are Classification5 and Regression6. An administered picking up Training informational index is pre named for grouping issues or capacity values are known if there should arise an occurrence of relapse. Subsequent preparation is done and the model has a base expense work for the preparation informational collection, later switched for scoring where we can anticipate values for new information.

Characterization: It recognizes bunch enrollment. That is assuming we have various occasions described by input boundaries, which can be named in an unexpected way, and we need our framework to foresee which name ought to be utilized.

Regression: Regression is a combination of multi-dimensional power supply and function interpolation. The regression problem is used to find the approximation of the function with a minimum error deviation or a cost function. In other words, the regression technique simply tries to predict numeric dependence, a function value, for example, of a data set. Figure 1. Diagrammatically shows how supervised learning is to solve problems

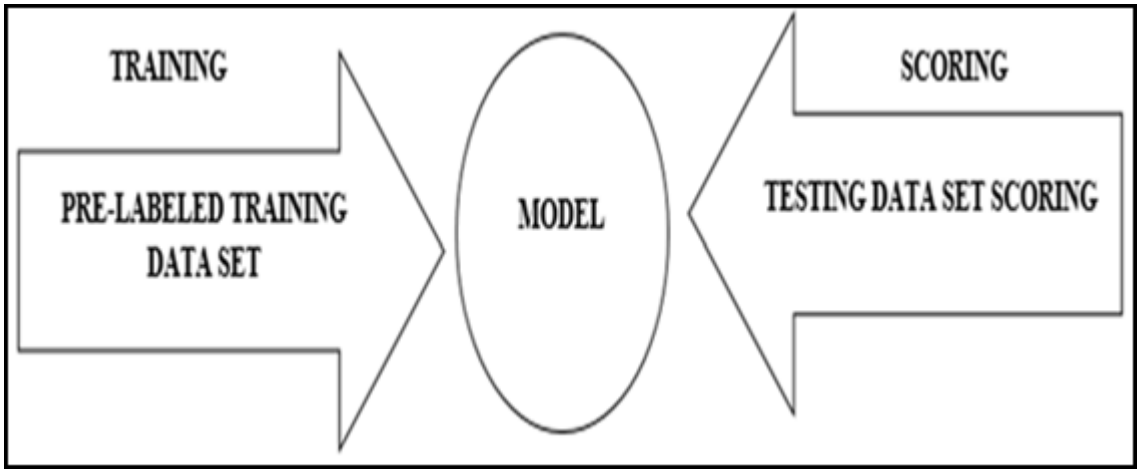


Fig. 1. Supervised Learning

Example of supervised learning, if a system has a data set which is a series of email messages, supervised learning task is to predict whether each email message is spam or non-spam(ham). This is supervised learning because there is a specific outcome namely spam or ham.

Unsupervised Learning: Unsupervised Learning is the process of extracting structure from data or how to best represent data. Examples of Unsupervised Learning are Clustering[7] (is partitioning a data set into meaningful similar sub classes called cluster) and Association[8] (method for discovering relations between existing attributes within a data set or data base). In an unsupervised learning situation, where the algorithm detects data features automatically, this depends on the purpose of the algorithm as well as the assumptions

made on what the properties and observed values are. Figure 2. Describes how unsupervised learning is used to solve problems.

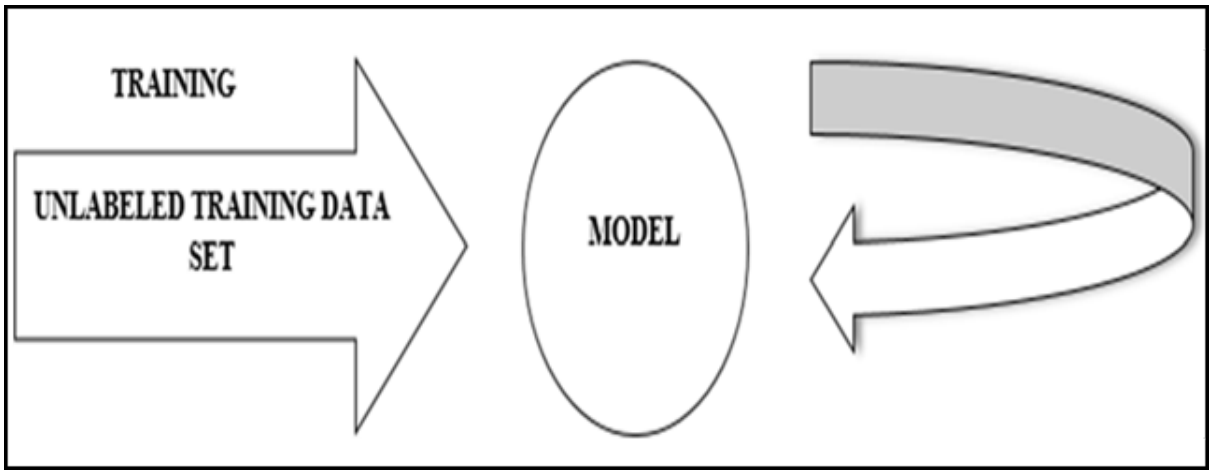


Fig.2. Unsupervised Learning

For example, if any data set was the characteristics and purchasing behavior of shoppers at grocery stores, the unsupervised learning task might be to segment the shoppers into groups or "clusters" that exhibit similar behaviors. Such learning methods might find that college students, parents with young children, and older adults have characteristic shopping behaviors that are similar within each group but dissimilar from the other. This is an unsupervised learning task because there is no right or wrong about how many clusters can be found in the data, which people belong in which cluster, or even how to describe each cluster.

2. Related Work

After the development of Machine Learning calculation and its use in record characterization, a great deal has been researched on distinguishing the elements of text. In this part we have portrayed the work done by scientists in the field of distinguishing Texts highlights by restricting their concentration exclusively on the field of SPAM recognizable proof. We have used different Machine Learning Algorithms in recognizing the SPAM and HAM subsequent to gathering the dataset from Kaggle. The dataset had three rows: the 1st row had the count of messages, the 2nd had the indication whether it is a spam or a ham message and the 3rd row had the message . We loaded the data set on google collab .importing the libraries ,read the data set , Plot Implementation of Bag of Words Approach, Data pre-processing, Implementation of Naïve Bayes Machine Learning Algorithm , Evaluating our SMS Spam Detection Model . After evaluating the model the pickle file was generated named as model.pkl , Since this was a web based project we used html and css for the front end of our web page . A python file was created named as app.py in app.py the code for flask connection was written and model.pkl file . In our web page we created a text box where the user has to paste the text and since the model was trained to detect the spam or a ham message as soon as the user enters the text he receives an acknowledgement whether the message is a spam or a ham since we have used audio tag in our html file the audio gets played according to the result.

3. Methodology

3.1. Workflow:

Data Collection: In this phase we have collected a dataset based on which they have performed the experimentation from Kaggle Repository[14].

Data Cleaning: In this phase the we have cleansed all the data which were taken into consideration. Authors have removed all the white-spaces, lowered the alphabet so that words like Equal and equal become the same, remove the remaining punctuation, like! is not that much important, tokenize each message, to represent the message as a list of words and done stemming, converting all the words to their root word, like floor, floored to floor.

Generating Testing and Training Data Sets: we have created the testing and training data on the converted cleansed datasets.

Applying Naïve Bayes Algorithm : We have used Naïve Bayes Algorithm for Training the Data. On the basis, the spam feature will be classified.

Prediction: We have given input messages to check whether the message is SPAM or HAM.

Figure 4. shows the workflow or architectural layout of how authors have classified the SPAM.

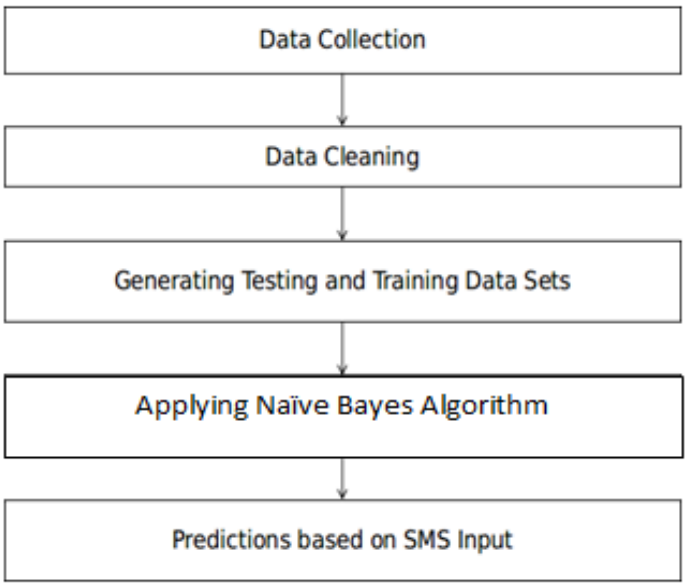


Fig 3. Workflow

4. **Experimentation**

As a part of experimentation, authors after creating the vector set, passed 2 inputs to test whether or not the model (including the word vectors) is able to check whether the message is SPAM or HAM.

Input I: given to the developed system: “Ok lar... Joking wif u oni...”

Input II: given to the developed system: “Free entry in 2 a wkly comp to win FA Cup final tkts 21st May 2005. Text FA to 87121 to receive entry question(std txt rate)T&C's apply 08452810075over18's”

Output of the above inputs are discussed in the Section IV of this literature.

5. Results and Discussions

Technology Stack Used in this research: Python based Flask Platform. Python Module Dependencies: beautifulsoup4==4.6.0, numpy==1.13.1, scikit-learn==0.19.0, scipy==0.19.1, sklearn==0.0, pandas, flask.

In this section authors have discussed the output generated by the system on the basis of Inputs given in the system, discussed in Section III of this literature. Figure 5, represents the Output 1 where the system was given Input I as: **“Ok lar... Joking wif u oni...”**

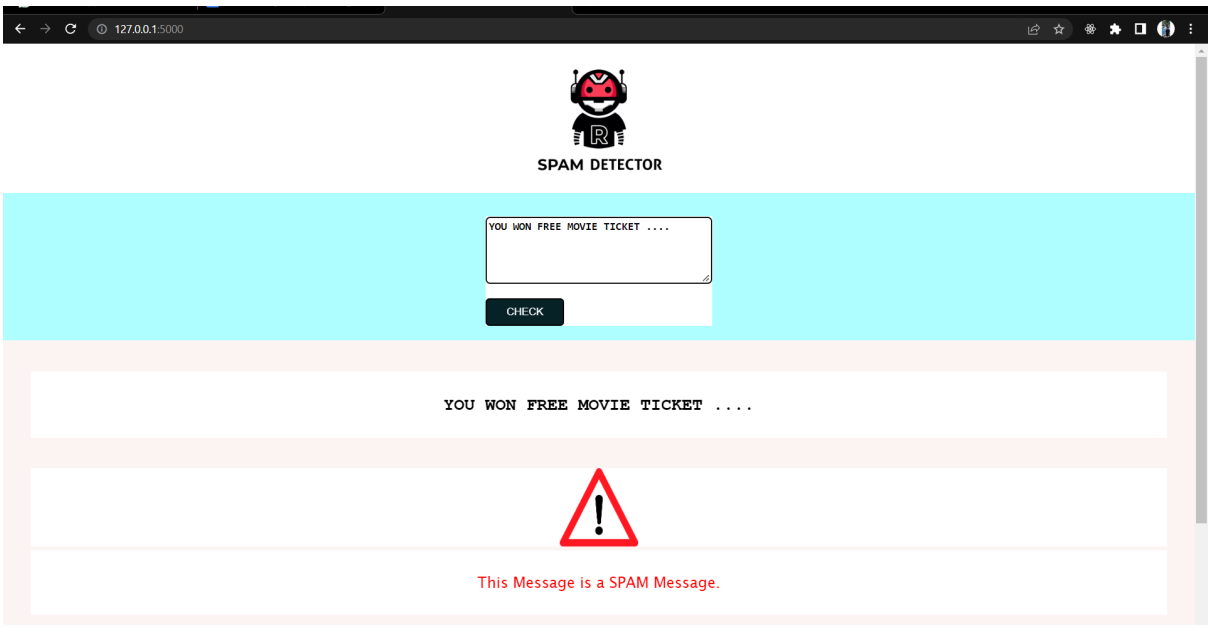


Fig 4. Spam Message detected

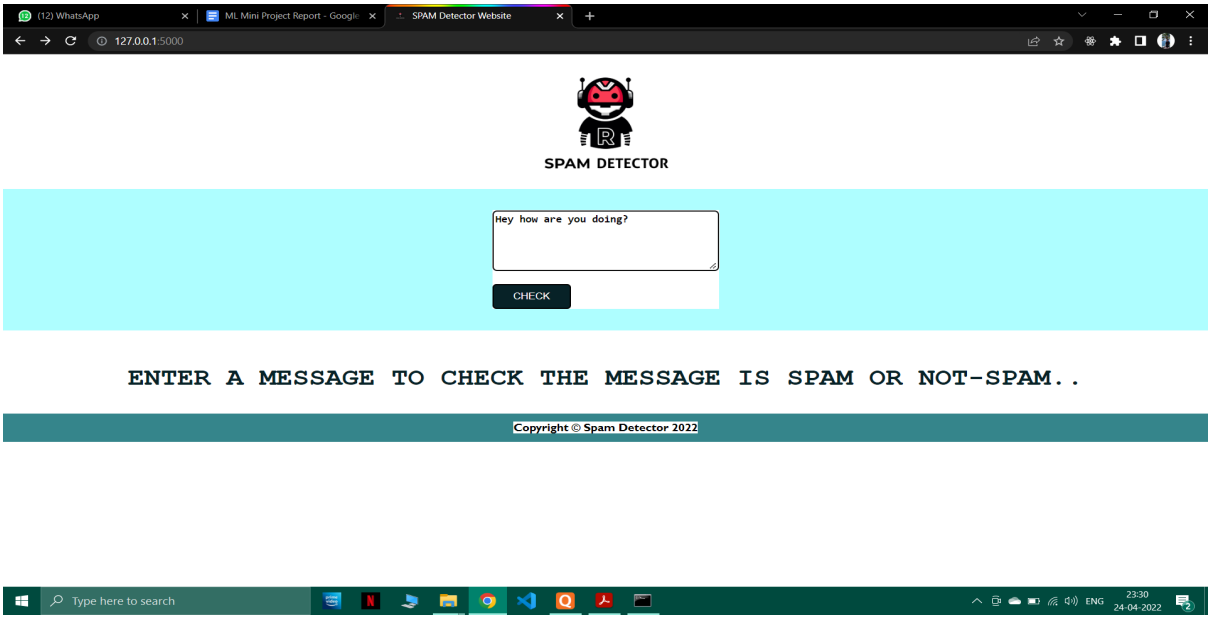
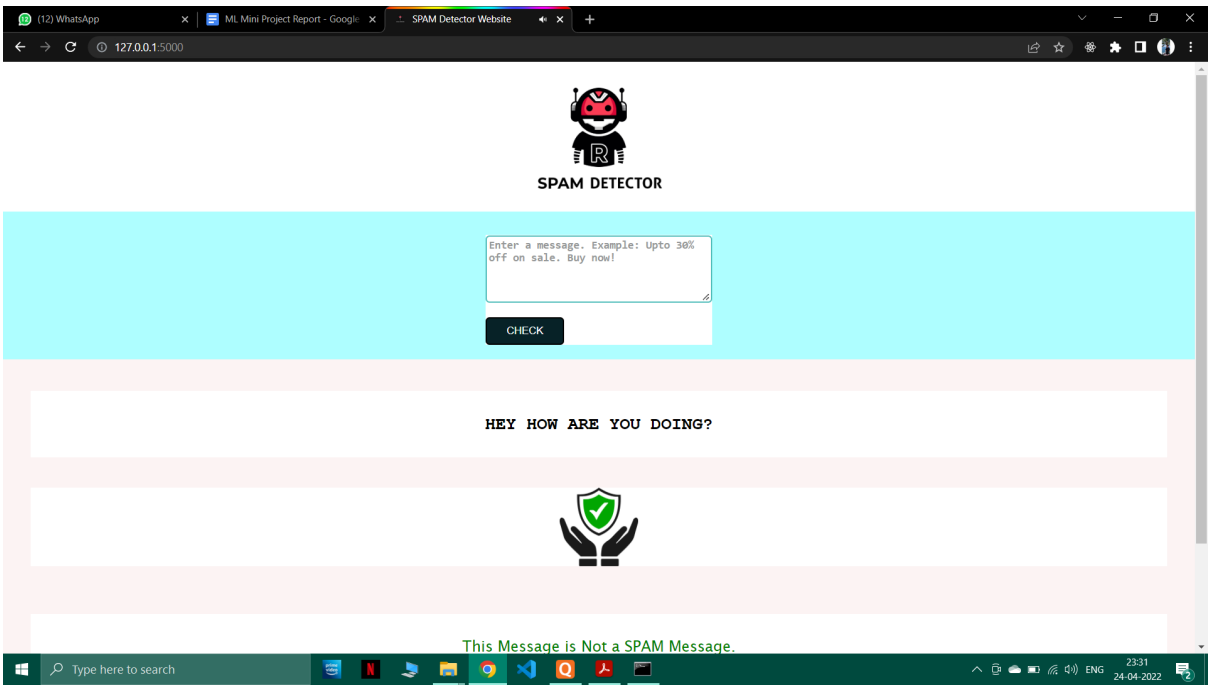


Fig 5. Ham message detected



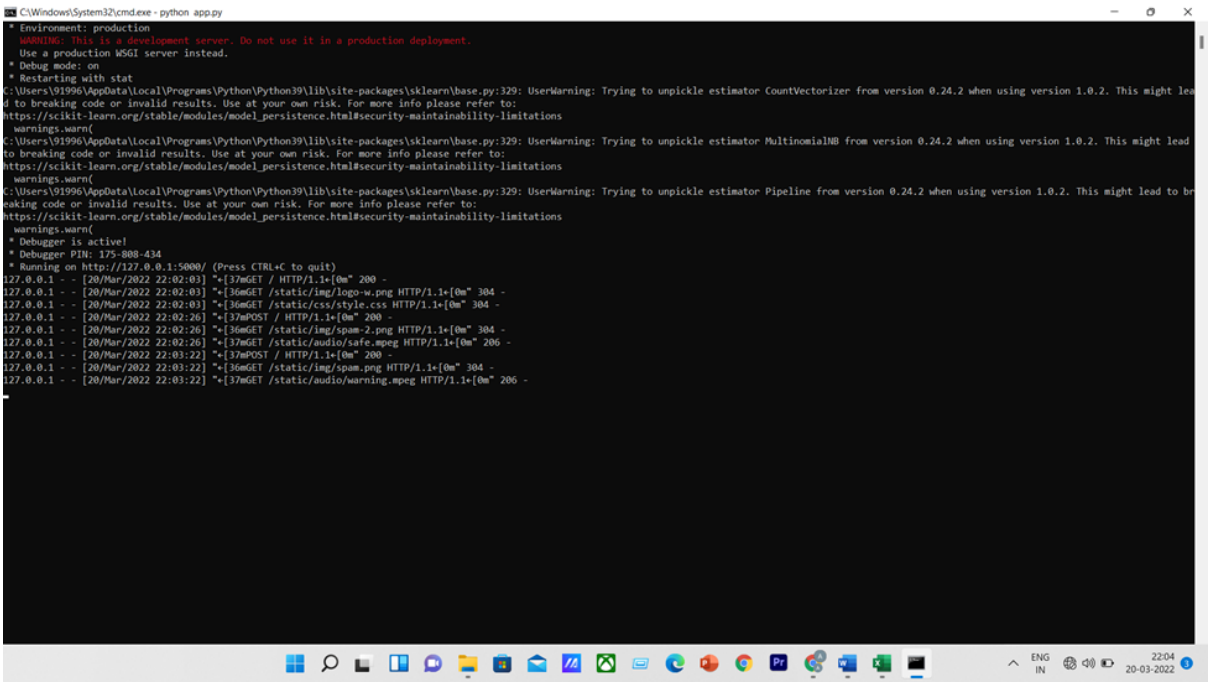


Fig 6. Input received at Python Backend server

6. Conclusion

Thus we successfully built a model that can help our users to differentiate between spam SMS messages and actually genuine SMS messages thus saving them from online frauds and smishing attacks.

From the above discussion and experimentation we can conclude that machine learning algorithms can play a vital role in identifying SMS Spam. After calculating the performance metrics such as accuracy, precision, recall and F1 score we observed that the percentage of all these metrics was above 95% for Naives Bayes classifier with the ideal (strong) value being close to 1. Hence we can conclude that the Naive Bayes Classifier will give us quite accurate results and is a more than reliable machine learning algorithm to detect Spam messages.

Future Scope

There are many machine learning algorithms which can be used for SMS spam detection. Even though the results obtained by using Naive Bayes Classifier are quite reliable, we can obtain better or similar results using SVM (Support Vector Machine) Classifier. It is one of the most popularly used algorithms for classification as well as regression problems. Random Forest, Linear Regression and K Nearest Neighbor(KNN) are also viable options. Each algorithm has their own strong and weak points, some give better accuracy, precision while some give better runtime. Thus in the future this SMS Spam Detection model can be improved by using a more accurate classifier which yields even better results than our model.

References

1. Njoku, Mary Gloria. (2015). The use of short message service in post-secondary education.
2. Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar, “Foundations of Machine Learning”, The MIT Press ISBN 9780262018258, 2012.
3. Sumit Das, Aritra Dey, Akash Pal, Nabamita Roy, “Application of Artificial Intelligence in Machine Learning: Review And Prospect”, International Journal of Computer Applications”, Volume 115, Number 9. 2015.
4. S.B. Kotsiantis. “Supervised Machine Learning: A Review of Classification Techniques” Informatica 31 (2007) 249-268
5. Nanhay Singh, Ram Shringar Raw, Chauhan R.K. ,” Data Mining With Regression Technique”, Journal of Information Systems and Communication ISSN: 0976-8742 & E-ISSN: 0976-8750, Volume 3, Issue 1, 2012, pp.-199-202
6. Amandeep Kaur Mann & Navneet Kaur, “Review Papers on Clustering Techniques”, Global Journal of Computer Science and Technology Software & Data Engineering, Volume 13, Issue 5, Version 1.0, 2013.
7. Ashima Sethi, Prerna Mahajan, “Association Rule Mining: A Review”, The International Journal of Computer Science and Application, Volume 1, No. 9, November 2013.
8. Swati Gupta, “A Regression Modeling Technique on Data Mining”, International Journal of Computer Applications, Volume 116, No. 9, April 2015.