SMS SPAM DETECTOR

Final Year Project

Naveed Hakim

UoB Number: 15026393

Supervisor

Dr. Malik Jahan Khan

Department of Computer Science

Namal College Mianwali

An Associate College of University of Bradford

Declaration

I , Naveed Hakim , hereby testify that I have never been involved in any plagiarism activity during process of my degree. I am aware of plagiarism policy of the university and I understand that properly.

I solemnly say that reported work is my own that is done under supervision of Dr. Malik Jahan Khan . Proper Credits have been given to the work of other people that was used for guidance or implementation purposes by using referencing.

Naveed Hakim

Acknowledgements

First of all , I am highly grateful to Allah Almighty who made me able enough to achieve this important milestone of my life.After The Almighty, I would like to thank my supervisor Dr.Malik Jahan Khan. He remained very supportive during the whole journey that kept me motivated. Without his support this would not have been possible.

In the end , I would like to acknowledge support of my family who stood against all odds for me during my whole educational carrier . I would also like to thank my friends who voluntarily installed my application for testing and gave me access to thier private text messaging data.

Naveed Hakim

Abstract

SMS spam detector is an android application that is developed for detecting spam text messages. After real time spam detection, user will get that message in spam inbox rather than general inbox because in a bulk of spam messages, people usually miss important messages. Significance of this application is that people need a text message classifier similar to email spam classifier. This application is designed by keeping needs of users in view.

Purpose of writing this document is to give a detailed and clear description of a system. It will give a complete overview of background, project flow , proposed solution methodology and implementation of that solution. After that, performance of the system is analysed very carefully by using different testing techniques. Finally , a discussion and description of results is given in last section. This document is to help reviewers of the project to get a detailed overview of work done in version 0.

Contents

1 Introduction

1.1 Project Overview

Text messaging service started in 1992 (O'Mahony 2012) and became popular in short span because people found sending short messages more convenient than other services. In that era, different telecom companies were starting their carrier in market that created environment of competition that led to rates reduction for short text messages. In this way, text messages became fast growing network of communication.

Due to its popularity and low rates, people started using this service for their marketing. Shortly, text message caught attention of spammers who started using text message for spreading their content. This type of text message are called spam messages. Just like spam emails , Spam messages contains unwanted content for the users ,for example advertisement of a product, mobile packages offers, adult content, hate remarks for certain political party and fraud or scam messages about lottery wining etc (Lota & Hossain 2017).

It has been observed that percentage of spam messages sent on daily basis varies from country to country and according to survey conducted by turbine express, Pakistan is ranked in top 12 spammer countries. A report published in India suggested that spam message sent in India are more than 100 million per day that is a huge numberAgarwal et al. (2016). This causes lots of harm to telecommunication companies as well as to users. If we talk about statistics of Asia, more than 30% messages sent daily are Spam messagesYadav et al. (2011). China is another country that is effected by spam messages.

These spam messages not only cause unease for users but it also causes economic lose. People are shifting from text messaging to other applications and telecome companies are unable to provide customer care in order to limit spam messages. Although they have taken important steps for example , limiting number of messages sent to a particular number per day but they are unable to stop spam messages. Spammers keep on changing their techniques and algorithms that work on one technique may not be able to work on another technique Xu et al.

(2012).

There are some websites who are giving free text messaging service. By using these websites , people do their marketing at very low cost. They become able to spread their product by messaging it to more than 10,000 people at once (Yadav et al. 2011). This factor has increased number of spam messages significantly.

Spam messages causes' annoyance to user and in particular cases, SMS spammers send fraud messages in which they ask to call at a certain number for getting a prize and take important details (home address, CNIC number etc.) or money sometimes from user by crimping. Users may be entrapped and suffer if not warned by system on right time.

It may also happens that in bulk of spam messages, recipient misses important text messages. Due to these reasons, it is necessary to handle spam messages in best possible ways. Detecting spam text message and marking that as spam is real need of the hour.

To address this problem , an SMS spam detector has been designed with the motive of classifying text message on basis of certain criteria.Targets are set in order to identify objectives of this research that are given in following section.

1.2 Objectives and Aims

Ultimate objective of this project is to:

- Present a research on spam SMS classification that may evaluate performance of system under multiple criteria.
- Make an classification algorithm in order to solve this problem.
- Devise a technique that may handle evolving techniques of spammers as spammers change their techniques very often and it is necessary to cope with that.
- Make this algorithm language independent is necessary. It should work on every language that can be written in roman English in the same way as it does on English.

- Give the research a final touch by developing a mobile application based on research results that may notify user on arrival of important message and sent junk messages into separate folder.

1.3 Report organization

This report is consist of six chapters that are organized in given way. First chapter named as 'Introduction' consist of introduction to problem and objectives and rationale of developing this application. Chapter two contains background information of SMS Spam Detection and describes basics involved in SMS Spam Detection.Chapter three deals with a systematic literature review of work that is already done on SMS spam detection topic.Systematic Literature Review is different from common literature review. After performing literature review , design of system is proposed in chapter four that deals with overall flow of design of this product. Fifth chapter named as implementation that is basically changing of the proposed approach into applied approach.Results and Analysis is sixth chapter that describes results on different settings of classifier. Discussion over results with a way forward is also given in last chapter.

2 SMS Spam Detection: Back Ground

Word spam means unwanted content in text messaging. To do SMS Spam Detection classification of spam messages is done into spam and non spam. There are a range of classifiers that are used for detection of spam messages.These classifiers work a flow that feature set is given to them. They train themselves on that feature set. This feature set can be any important attributes of data. After training of algorithm , it is given a test feature set and it results result prediction for it. It is necessary to cover theoretical part of these two before going towards implementation of them.

2.1 Naive Bayes Classifier

Naive Bayes classifier is considered to be one of the best classifiers for spam content filtration. It is also light weighted and therefore , a comparatively fast algorithm. Naive Bayes algorithm takes a training data set in beginning and trains itself on that (Mahender & Korde 2012).

Basic principle of Bayesian filter is calculating probability of a text in spam or non spam message . For example , if we have three messages in our database , "contact me on +9230866666666" , "contact me urgently" and "you have won a prize of 100000$. To collect this contact us on +999999999" . Let suppose first two messages were non spam and third one was a spam message. Now , probability will be calculated for each word's occurrence in spam or non spam. let take word contact , for word contact 2/3 chances are that this is a non spam word and 1/3 chances are that it is a spam word. If a new message arrives that have word contact , it will check its probability of being spam or ham words and will check this for all words in same way (Yadav et al. 2011).

Bayesian algorithm is used since 1960 and it is most mature algorithm in SMS Spam Classification but it is language dependant (Wang et al. 2013).

2.2 Neural Network

Neural network is one of the widely used approach in classification problems. It also takes seed data for training and is able to classify incoming data on bases of that. People have used this approach hybrid because of its limitation of being fully dependant on data set.

This machine-learning technique inspired by idea of working of human nervous system for autonomic computing. This paradigm is built upon a basic unit named as neuron or perception. Natural neurons are consist of dendrites (input), axon (output) and a cellular nucleus (processor) . It goes for artificial neural network as well in which neurons are then joined together in a specific way (explained below) to produce desired results. Important and interesting thing about a neuron is that it can learn any specified function to reach to desired output Neural Network consist of three basic units, each unit may have a certain number of neurons in them.

- Input layer that is consist of number of neurons equal to number of inputs of a specific problem.
- Hidden Layer is processing unit. Number of neurons in hidden layer as well as number of hidden layers depend upon nature of data and problem.
- Output layer that have number of neurons equal to number of required outputs.

Network works on a simple formula... Take dot product of input and weights, add biases and activate. Through this simple neural network, almost everything of world can be represented in terms of neural network model. One neural impulse is generated by summing weighted inputs and biases that makes equation of a line and is able to place on cut in input state to classify data. (Classified data image) But if data is complex as given in Fig and cannot be separated by one single line , placing hidden layers become necessary to have a cut in another space over the data and classify data beautifully. Outputs are passed to an activation function to confined outputs in a certain range (Hammerstrom 1993).

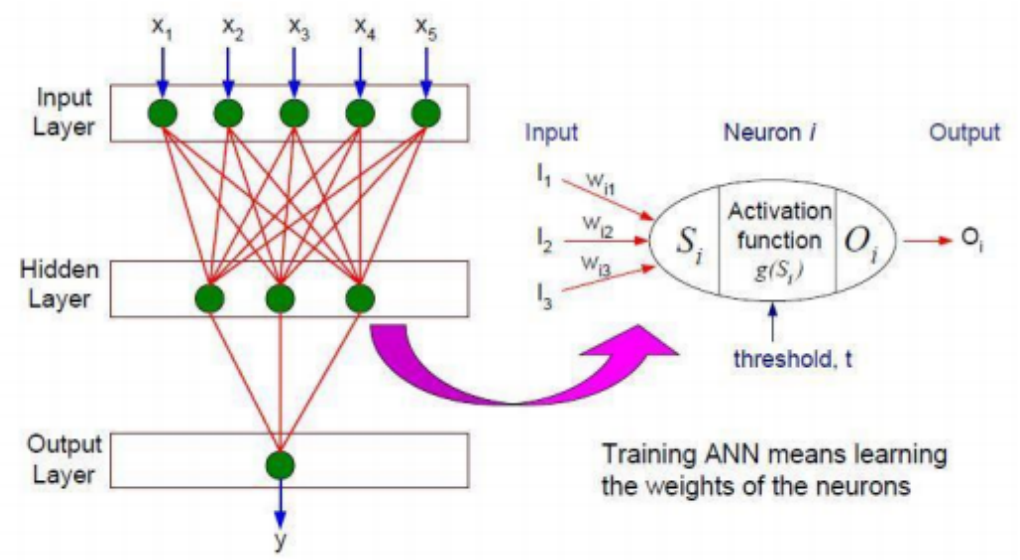

Figure 1: Neural Network Working

In first iteration of a neural network is done by giving it random weights. If result does not come as expected, we have to updates weights that is done using back propagation technique that states to take derivative of error with respect to each error to know how much error each weight is causing and then using that derivative to update weights (Panchal & Panchal 2014). This is repeated until networks gain correct weights that may minimize error in classification.

3 Existing SMS spam detection approaches: A Systematic Review

A systematic literature review is performed to give an overview of maximum possible literature that is available over the internet about SPAM text message classification. Main aim of this section is to provide a top down approach for reviewing existing data about spam detection. A literature review performed systematically ensures that maximum possible literature has been discovered on related topic and minimizes human error because it works on a predefined strategy. Guidelines given by Barbara Kitchenham were used for conducting the review (Kitchenham & Charters 2007)

3.1 The need of a systematic review

Developers and researchers are trying to give methods to handle evolving spam content generation techniques since 20 years Therefore, a lot of work has been done on email spam detection, and social media spam content detection and message spam detection. It is necessary to first find and review already done work in a comprehensive way.

In this section, a systematic review is presented to provide a top down comprehensive approach for reviewing existing data on spam detection. Researcher who perform systematic review must make every effort to find and report existing data. It ensures that none of important data will escape from eye of researcher.

3.2 Research questions identification

Questions	Motivation for investigation
RQ1. What are currently existing approaches for message spam detection?	Aim of this question is to find methodologies that are used for SMS spam detection.
RQ2. What are policies for measuring spam?	Answer will be in form of features identification.
RQ3. What are strengths and limitations of these systems?	Answer of this question will be in form of SWOT analysis that will describe strengths, weakness, opportunities and threats of current systems.

3.3 Search process

To collect papers on SMS spam detection, a search is performed using search strings given below:

- SMS spam detection.
- SMS spam filtration OR email spam filtration.
- Content based text filtering.
- Naive Bayes algorithm for spam message.
- Systematic literature review of SMS spam detection.

This search provide number of articles on spam filtration. To find only important and data of use an inclusion and exclusion criteria was defined.

3.4 Inclusion and exclusion criteria

Inclusion and exclusion criteria was defined by taking guidance from a mapping study on spam detection (Kamoru 2017).

- Irrelevant studies were removed.
- Duplicate publications available on different resources for example ACM, IEEE, Elsevier and Scopus has been removed.

- Articles with high bench marking and most citations were given more considerations.
- Both SMS spam detection and email spam detection data was included.
- Specific attention was given to new work that has been done after 2002 because computer science industry is fast growing industry.
- Systematic literature reviews on spam content detection were included.

3.5 Database for research

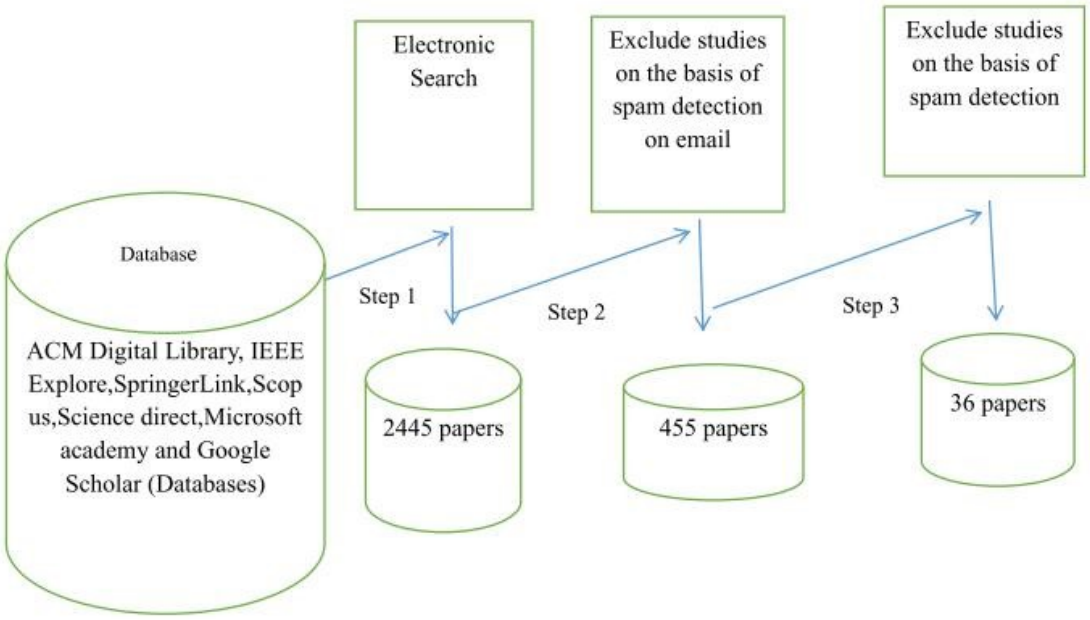

Figure 2: Database for SLR

By following above given process, 36 journals were reviewed in narrowing down research.

Validation of study

To analyze validation of this study, a SWOT analysis is performed.

Strengths

This study follows a proper method of analyzing all data available that makes it more authentic than other studies.

Weakness

This study takes much more time as compared to simple literature review.

Threats

The only threat to validity of this study is human error as human understanding is limited. Efforts are made to minimize this error by taking peer reviews.

Opportunities

Upcoming studies can be added in this review by extending it as systematic review is an extendable methodology.

3.6 Summary of Reviewed literature

Spam content refers to the unwanted content that people send for their promotion but is annoying for readers and users. SMS spam generation arose when text message became popular for communication. Efforts are being made to identify spam content and notify user about them on time. According to statistics of today 93% messages of total sent messages are spam that is a huge amount.

In literature spam detection is divided into two parts.

Content based SMS spam filtration and non-content based SMS spam filtration (Sajedi et al. 2016).

Content based filtration

Content-based SMS spam filtration is concerned with a piece of text that comes from another user. Burden on developer or analyzer is to judge this piece of writing whether it is spam or not on basis of different features. This approach is quite famous and is nearly 70% of SMS spam detection work is done on basis of content-based analysis.

Bases of content based filtering is text , words , organization of words, context of words, formal or informal language , length of text , parts of speech distribution and inclusion of URLS.

Ghulam Mujtaba and Majid Yasin conducted a research in 2014 on content-based spam detection using online database for text messages. They suggested that making a spam words dictionary using database and applying Naive Bayes classification produces best classification results.

Naive Bayes is an old but sound method for detecting text message classification (Mujtaba & Yasin 2014). Rushdi Shams and Robert E. Mercer highlighted that along with SMS spam dictionary, error features also play an important role in spam detection. A message that is difficult in reading is most probably to be a ham message because of informal language. After feature engineering they applied Short Vector Mechanism and Naive Bayes algorithm and reported accuracy of 93% (Karami & Zhou 2014*a*).

Text message may have the content that is not needed in spam detection. Sunil B. Rathod suggested that removing that unnecessary content in very start would help in improving accuracy. Articles, preposition, conjunction and high frequency words according to dataset were removed from all incoming messages. After making bag of words from remaining words a Bysian classifier was applied that touched 95% of maximum frequency (B. Rathod & M. Pattewar 2015).

Non-content based filtration

When a message arrives, along with its content there is metadata that may also help in detection of spam text. Non-content based SMS filtration is consist of following important perimeters that are gathered from reviewed literature.

1. Whitelist/ Blacklist Sender: it is most likely to receive a spam message from the person who is not in contact list rather than ones who are in contact list. Moreover, black list refers to blocked users. Their messages are certainly spam messages for specific Suggested by Yadav et al. (2011).

2. Temporal features: time of text message arrival, reading time and replying time adds to value of ham message. Moreover, number of conversations with certain person in week or day. Suggested by (Sajedi et al. 2016)

3. Network features: It is important to check whether a text message is coming from a network or a website. As spammers user auto text messaging to spread their content. Sugggested by (Xu et al. 2012)

Machine learning algorithms

After taking out content and non-content based features of a text message, generally a learning algorithm is applied on these features. These machine-learning algorithms involve Naive Bayes classification, SVM, Multi layer Perception, K'th nearest neighbor and hybrid of any two methods. Naive Bayes classifier is old but most reliable amongst these approaches because of the fact that it works on bag of words model that describe spam and ham messages on ground of spam or ham words presence. Detailed analysis of all methodologies of features selection and machine learning algorithms is given below in form of SWOT analysis.

3.7 SWOT Analysis

Algorithm	Strengths	Weakness	Opportunities	References
Content Based Filtration	Content-based filtration is powerful as it evaluates content that is actually bases of spam.	It is language dependent. It cannot handle evolving techniques used by spammers because of rigid words	Content based filtering can serve as bases of not only SMS spam detection but also tweeter , Facebook and other media spam detection	(Kamoru 2017) ,(Alsmadi & Alhami 2015)
Non Content Based Filtration	Language independence is biggest strength of non-content based featuring.	It is not complete in itself as content based is.	Can be used along with content based featuring for text message app development.	(Choudhary & Jain 2017) ,(Yadav et al. 2011)

Algorithm	Strengths	Weakness	Opportunities	References
Naive Bayes filter	Simple , accurate Most tested	It is based on a wrong assumption on testing and training data to be in same language.	It can be used hybrid to produce best results.	Shahi & Yadav 2014 ,(Alsmadi & Alhami 2015)
Multi-layer perceptron	Simple , fast , compassionate	Best on linear classification.	Can be basis of hybrid algorithem	(Cormack 2008)
Support Vector Mechanism	Is able to achieve desired accuracy	Slow as compared to others	- Lota & Hossain (2017),Karami & Zhou (2014*b*)	

SMS and email spam detection difference

People have started working on this though email spam detection but still there is a lot more to do in message domain because of short content of message , different and non-dictionary languages and new spam techniques that spammers use to deceive developed algorithms. Furthermore, text message does not contain images lots of time that may help in detecting whether it is spam or not (Delany et al. 2012),(Ali & Maqsood 2018).

Text message classification has gained importance over email filtration because its use is increasing day by day. It can be suggested to use email filtration technique on text message but that is not applicable because of the fact that text message classification is significantly separate area of

research. Therefore, The need of the hour is to give a product that may help user in identifying spam messages as well as it should help them by marking spam and placing that in separate folder to avoid text message over flooding.

Moreover, Motivation to do this project is to play role in developing a real

time and efficient product that may deal with problems mentioned above and are not dealt before. Furthermore, mobile companies may integrate this component in their built-in SMS app.

Conclusion

This systematic literature review suggests that although a lot of work in relevant domain has been done but there does not exist any product in Pakistan that may classify text message and notify user instantly due to language barriers. In light of this review, system design is presented in next chapter.

4 Proposed SMS Spam Detection Approach

SMS spam detection is a domain in which there is already lots of work is done . After studying and analysing that work , it has been realised that this product should be able to perform following necessary action.

- Research carried should fill the gaps left in previous researches that are documented in Weaknesses and opportunities part of SWOT analysis.
- Research should be able to be give a product that is language independent.
- Carried research should have a growing database that may help in dealing with spammers evolving techniques.
- Application should be full fledged android application. Full fledged application will help both user and developer in two ways. it will give developer a real time testing data that may help him in getting a real analysis of performance of the product as well as he will get real time data to test his product. This will be a benefit for user as well , as he will be having an application that may bring ease into his life.

By keeping all these needs in view , following is the proposed approach to carry out this research.

4.1 Flow Chart

This system known as SMS spam detection will be able to detect an upcoming message and categories it into spam or non-spam. Flow Chart diagram is given below:

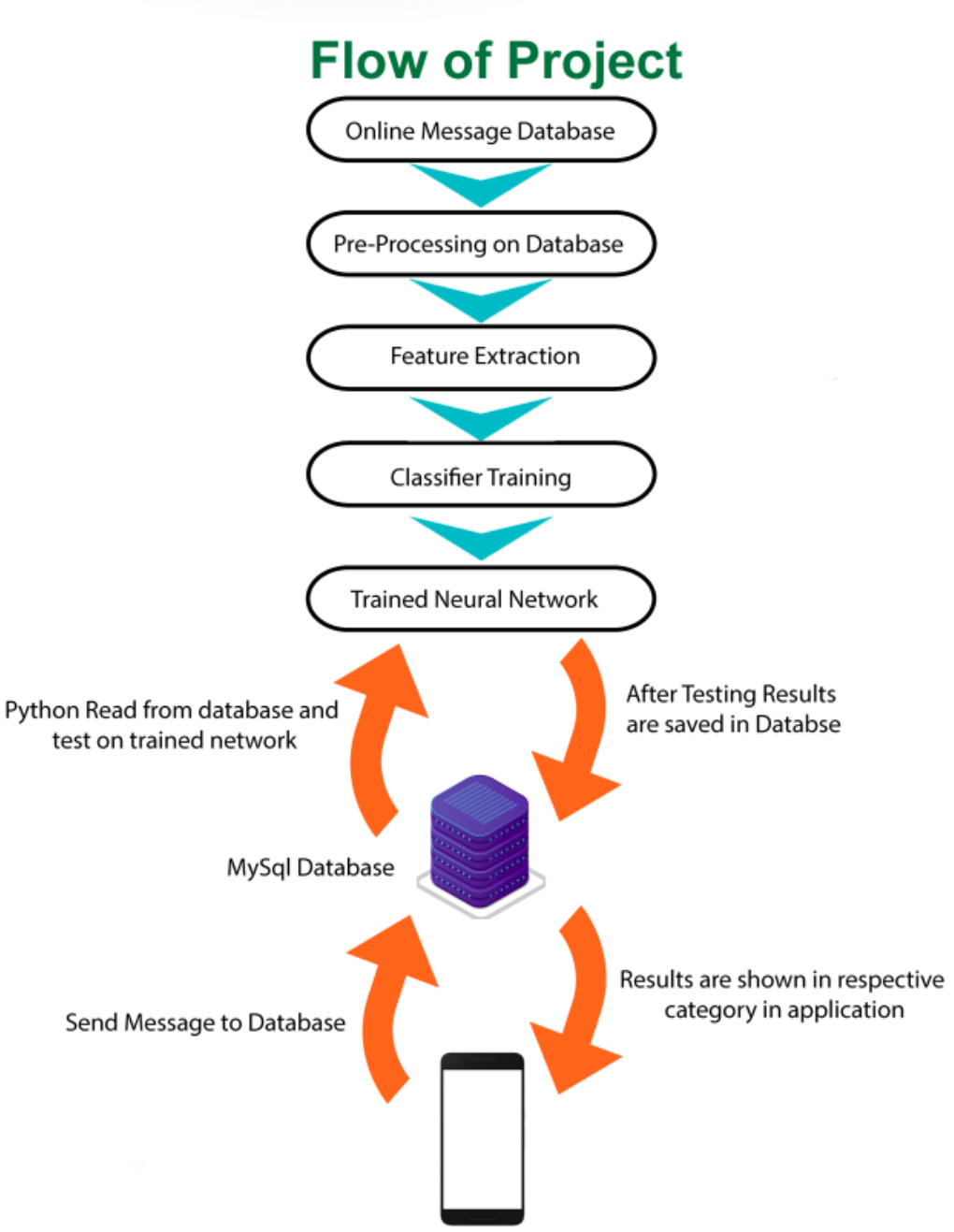

Figure 3: Flow Chart

4.2 Proposed pseudo code for spam detection

Input: Incoming real time text message in android application
Output: Classified text message
Algorithm
Training

- Python server reads from training data from database, training data includes text of messages, category and some non-content based features.
- Server performs prepossessing on textual data
- Server extracts more features using natural language processing and Naive bayes filter.
- Neural network is tuned and trained.

For each incoming text message in android:

- Non content based features are extracted including manual or non-manual text message check, white list/blacklist check and number is in contact list or not check.
- Features and text of message is saved in database.
- Python reads from database, performs prepossessing and extracts same features that were extracted in training data.
- Trained classifier is tested on that incoming message.
- On basis of results, text message is classified in spam or non spam and sent in respective tabs of application.

Note: For each classified database is maintained that is included in training data
for next time training.

4.3 Proposed steps for solution

4.3.1 Data Collection

Due to a lot of work done in spam detection domain , there is a large number of data that is available for spam detection but keeping context of SMS spam detection in view , publicly available data is very limited. There is not even a single SMS spam detection data set in URDU language over the internet up till now (2019). However , some amount of data is available in English language as well as that data set has only content of messages and does not have metadata of messages.

Data set that is used currently is taken from two sources. One is Kaggle data set (*Spam Dataset* 2017) , UCI machine learning repository data set.This data set is consist of two columns named as text and category. Text column has sample English language text messages in it and category represents whether a message is ham or spam.

This data set is consist of 5547 instances out of which 70 % of messages are ham messages and remaining are spam messages.

As for as non content based features data set is concerned , that is not available over the internet and dummy data is used for version 0 of this system.

4.3.2 Pre-processing Phase

Pre-processing is the phase in which data is prepared for feature extraction. As data set contains raw data that have lots of unwanted things that may create noise. Noise is unnecessary data that will be of no help in detection of spam message as well as that may cause significant lose in accuracy (Aski & Sourati 2016).

For this data following are the steps that are performed to pre-process data.

- All blank rows are removed as they are of no use.
- Stop words are removed.
- All words are changed into stem words. For example trying, tried and try will be considered as single word ‘tri’ to avoid wrong interpretation.

- All spam entries are mapped on 0 and all ham are mapped on 1 because of the fact that neural network can get trained on numbers easily.
- Further columns were added in order to make space for features against each number. Details and exact number of these features are given in next section.
- Punctuation's usually don't help in spam classification so that were also removed.

4.3.3 Feature Extraction

Feature extraction is most important phase of spam detection because of the fact that it has a direct impact on classifiers training and accuracy achieved. After lot of research and study over the topic , following were features that were used in this project.

Along with research some of the features were identified by looking at common patterns in spam and non spam messages. Following are two important types of features extracted. Extracted features can increase as well as decrease accuracy of detection as it has direct impact on output (Uysal et al. 2013).

Language Dependent features

Language dependant features are the features that are dependant on language , for example if training data is in English then Urdu text message will not be classified by using the classifier trained over that.

Following are important language dependant features used :

Profanity Level

Profanity level means that whether a textual string is consist of profane text or not. The text that contains abusive or adult content is called profane text.

Naive Bayes Filter

As explained in section , Naive Bayes filter works on textual model . It can classify text message as spam or non spam. Adding this output to neural network's feature set was an experiment done.

Language Independent features

The features that does not depend on language are called language independent features. Purpose of these features is to make algorithm work independently of language.

List of language independent features is given below.

Length of text

length of a text message is important because of the fact that text messages are 160 characters long at max. They cannot be longer than 160 characters. It has been noted on bases of training data set that usually short text messages are not spam messages as people usually do not use Short Message Service for long text messages.

Longer text messages are mostly spam messages because spammers try to include information about them in one single text. They try to give maximum information about them to cause an impact on user's mind. However , this is not fixed trend. Some spammers may use short messages for spreading there content in order to deceive spam detection algorithm.

Length of a string is calculated on basis of space characters and so becomes language independent features (Zhang et al. 2016).

Url Check

Usually messages that are spam have a URL in them so that user may click over the URL and they may get their benefits from that. Apart from being spam ,these links are sometimes dangerous for user and may cause him loss. People can encrypt someone's phone as well as they may steel important information from one click of user.

As URLs have no specific language so this feature is also independent of feature and will help in making algorithm language independent.

Mathematical Character Check

This feature checks if text message string contains a mathematical character or not. Mathematical character means numeric characters and mathematical op-

erators. Dates are also in numeric so they are also included in this feature.

Mathematical character check is also language independent feature. If training data set is in English and contains mathematical character , same mathematical character can also be identified in an Urdu string.

Spam messages usually contain mathematical characters in which they describe there offers , important dates , prices and wining amount etc.

Phone Number Check

Spammers add phone numbers to contact on in order to do frauds or promotions .To check this suspect this feature is included and phone number is checked in each text message string in order to give this feature value of 1 or 0.

Emoji Check

Emoji 's are most important way to convey emotions during text message chat in this era. During informal chat and conversation , people usually use a lot of smiles and symbols for showing different moods to other person. However , this is not the case with spam content. Spam content is formal and does not contain any smile most of the time apart from a few very rare exceptions. Therefore , this is an extremely important feature in spam detection.

Metadata Features

Meta data features are also language in dependant features. Following are the important metadata features that helped in classifying a text message in spam or non spam.

Number of Sender

As suggested by (Xu et al. 2012) number of sender is also important in getting a spam text message identified. Number of sender can either be in contact list or not . If a number is in contact list there is very rare chance that text message will be a spam message . This feature is also language independent feature.

Manual or auto sent text message

People who use text messaging service for promotions or marketing ,do not send a text message from manual mobile phone . They use auto texting or text messaging website to send messages in bulk so that becomes a key factor to know if a text message is spam or not spam Zhang et al. (2016).

4.4 Classifier Training and Testing

After features extraction , main part of this system is to build a classifier on extracted features that may classify incoming text message as spam or non spam message . There were number of algorithms available to solve classification problems , However , we chose neural network and Naive Bayes Classifier.

Naive Bayes Classifier is mature and stable algorithm for detection of spam text. Its basic rule is distinguishing a text on basis of bag of worlds model.

Neural Network is another paradigm in field of artificial intelligence that is trained and tested on a specific data. Reason for choosing neural network as main classifier is because it can take results of other algorithms as a feature vector too. Due to this quality of neural network this approach is followed.

Idea is to build a neural network that will take all features and output of Naive Bayes filter that should already been trained over same data set as that of neural network to enhance effect of both of the algorithms . This trained neural network will then be tested on real time data incoming from android application.

4.5 Growing Database

System is meant to keep with growing database that will help in future spam detection because spammers change there techniques with time .With growing database , probability of getting deceived by growing database is less .

After proposing a methodology for classifier , it was also necessary to give an overview of non function requirements for android application that are given below :

4.6 Nonfunctional Requirements

Nonfunctional requirements specify that how should an ideal system work. Each best mobile application should have nonfunctional requirements fulfilled for best performance. Following nonfunctional requirements are specified for this system. Moreover, guidance to identify nonfunctional requirements were taken from Khalique et al. (2017).

Portability:

Portability reefers to ability of system to be translated from on system to another. This system will be able to translate itself in different android versions.

Performance:

Response to request should be given by the system in less possible time refers to best performance of application. This application specifically needs to be fast because user will not be interested in knowing status of a message after passage of certain time . This application is time concerned and factor of time can not be neglected.

Reliability:

System should not crash. It should be flexible enough to handle exceptions .

Availability:

This app should be available all the time because of the fact that message can come anytime.As classification is server dependant , SMS receiving is managed to be available using local storage.

Security requirements:

As app hold very private data of user, it should take responsibility of security.When things like SMS data comes into an application , security of data is the key.

5 Implementation For Results

To implement the design , this system is divided into seven main modules. Each module is interlinked and contributes towards main design. This product is an android application that is connected with python back end.

Modules are described one by one below :

5.1 Database Collection

Initially , data was collected from Kaggle and UCI Machine learning repository . However , after sometime , application starts to maintain users data as data set and on scheduled training new data is included afterwards.

5.2 Python Pre-Processor

Training data set is pre-processed by python script.

5.3 Python Feature Extractor

This module is responsible for extracting all content based features for training of neural network.

5.4 Classifier Trainer

Classifier is mixture of neural network and Naive Bayes filter. Initial settings are given in diagram below :

After training of classifier it is now ready to perform testing on each incoming message.

5.5 Android Text Message Receiver

Android text message receiver is responsible for receiving text message in android. This android text receiver is built in android studio. It works for android 5 and above. A broadcast receiver is built for this purpose.

5.6 Broadcast Receivers

An android phone can send or receive a broadcast .Text messages are also received as a broadcast. To receive incoming text message a service is implemented in android that continuously runs in background and waits for a text message to arrive . As soon as text message arrives broadcast receiver receives the message and sends to database.

5.7 Android Feature Extractor

This module is responsible for extracting metadata features of text message. These two features are whether number is in contact list or not and whether the text sent is manual or auto sent text message. Output of these features is also sent to database.

5.8 SQL Lite database:

SQL Lite database is a local relational database that can store data in form of tables. SQL Lite database is consist of one table with five columns :

- Text of the message
- Contact number from where the message came
- Result that shows whether message is spam or non spam in order to retrieve spam messages in spam box and non spam messages in non spam box.

This is android local storage that keeps these two things in order to show user messages even if internet is not available.

5.9 SQL Database

SQL database is a global database that is synchronized with SQL Lite database as well as it stores both of the meta data features that are extracted.It also stores other python features in order to utelise this data afterwards Android uses rest API to store data On SQL server.

Table structure | Relation view

#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra	Action
1	id	int(11)			No	None		AUTO_INCREMENT	Change Drop More
2	Mac	int(11)			No	None			Change Drop More
3	SMS	longtext	utf8_general_ci		No	None			Change Drop More
4	Number	bigint(20)			No	None			Change Drop More
5	Iscontact	tinyint(1)			No	None			Change Drop More
6	Count	int(11)			No	None			Change Drop More
7	Profanity_level	float			No	None			Change Drop More
8	CountOfMathematicalChar	int(11)			No	None			Change Drop More
9	UrlCheck	tinyint(1)			No	None			Change Drop More
10	HavePhoneNum	tinyint(1)			No	None			Change Drop More
11	HaveSimilies	tinyint(1)			No	None			Change Drop More
12	SpamWords	float			No	None			Change Drop More
13	Spam	int(11)			No	None			Change Drop More

Figure 4: Structure of Database

5.10 Python Result Analyser

Python result analyser runs constantly with a gap of 1 second and checks SQL database to see if a new message has arrived or not. As soon as it gets the message and meta data features associated with that in database , it performs pre-processing and feature extraction on that specific text , sends all features to neural network and gets SMS classified as spam or non spam.

Text message is associated with an IMI number that is of that specific phone from which message is coming and to which result has to be sent. This step is done in order to avoid mixing of messages if two devices get connected with database.

5.11 Android Front end

Result is sent back to android application according to its IMI and mobile number of sender . Android app shows results to user in form of a toast message and throws spam and non spam messages in separate box's as shown below :

Surprisingly , this whole process does not take more than two seconds to happen

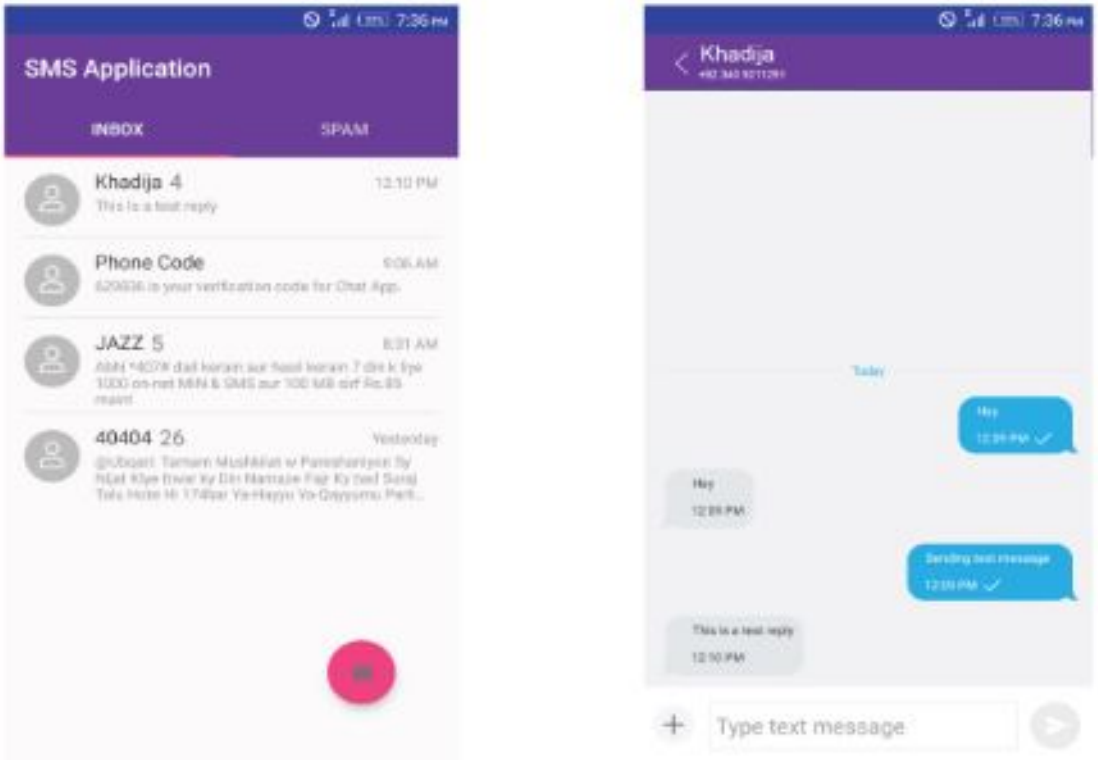

Figure 5: Android Application for Spam Detection

6 Results and Analysis

The most important part of carrying out a research is to document results and perform an analysis on that results . after initial settings of classifier , further experimentation was done with changing multiple factors that will be discussed in this chapter.

Naive Bayes Filter

Naive Bayes filter is the commonly used technique to find out spam content in the text. So i used this to find spam in text messages. This algorithm is content based filter and takes only string of text and classifies it as percentage of spam and non spam.

It will vectorize the given data on basis of spaces in the text. Then , it will calculate possibility of each word of being spam or non spam. On arrival of new text , same will be done with that and its probability of being spam/non spam will be given on basis of presence of spam words in it.

Following pi chart will show accuracy that i got form this technique.

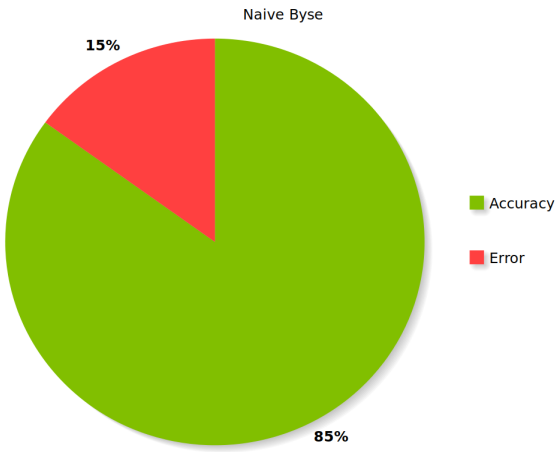

Figure 6: Visual Result of Naive Bayes

Neural Network

6.1 Experiment 1

Data for this experiment was collected from Kaggle and before staring the experiment following parameters was set.

- **Train Data:** 60%
- **Test Data:** 40%
- **Epochs:** 60
- **Activation Function:** relu

6.1.1 Results

When this experiment was run very good accuracy of 94% is obtained and the following pi chart shows the results of accuracy and error

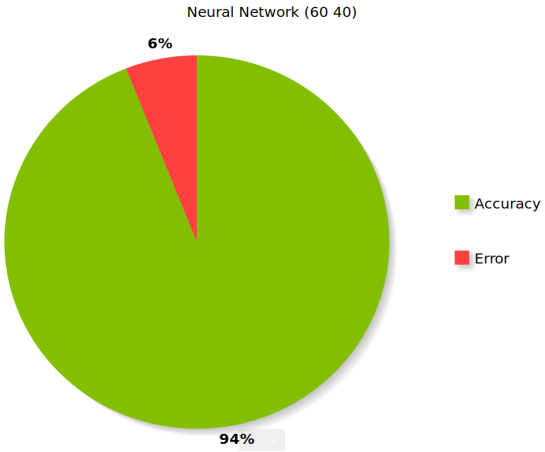

Figure 7: Visual Result of 60% Training Data and 40% Testing Data

6.2 Experiment 2

Data for this experiment was also collected from Kaggle and before staring the experiment fol owing parameters was set.

- **Train Data:** 70%
- **Test Data:** 30%
- **Epochs:** 60
- **Activation Function:** relu

6.2.1 Results

When this experiment was run very good accuracy of 95% is obtained and the following pi chart shows the results of accuracy and error

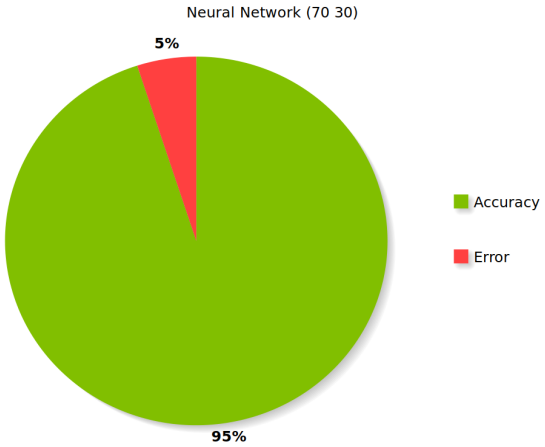

Figure 8: Visual Result of 70% Training Data and 30% Testing Data

6.3 Experiment 3

Data for this experiment was also collected from Kaggle and before staring the experiment fol owing parameters was set.

- **Train Data:** 80%
- **Test Data:** 20%
- **Epochs:** 60
- **Activation Function:** relu

6.3.1 Results

When this experiment was run very good accuracy of 96.05% is obtained and the following pi chart shows the results of accuracy and error

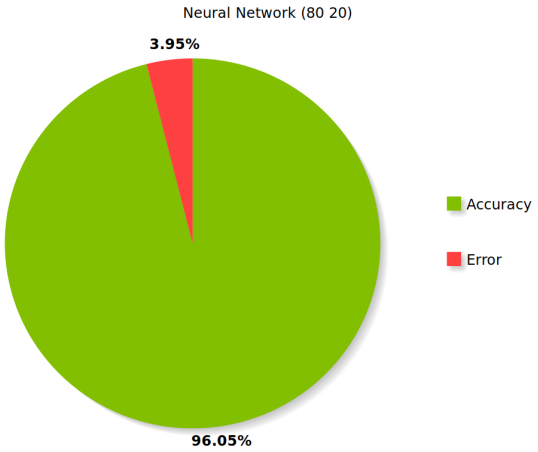

Figure 9: Visual Result of 80% Training Data and 20% Testing Data

6.4 Comparison of Results

Naive Bayes filter gave less accuracy alone as well as accuracy of neural network was also not satisfactory in first iteration this was because of the fact that features involved were not enough for spam detection. Afterwards , multiple features were identified and implemented that helped in increasing accuracy of neural network.

However , on adding meta data features , accuracy dropped two to three percent because neural network is fully dependant on data and data used for these features was dummy data in start. However , on increase of data set application will perform better on real data. As far as Naive Bayes filter is concerned , that is also language dependant and therefore , cannot be a perfect solution alone . Adding naive Bayes results to neural network features , accuracy got improved as that is one of the most mature algorithm in field of spam detection.

In neural networks setting hyper parameters is another important factor that helps in generating better results . Neural network's hidden layers are kept half of sum of output and input layer as suggested by() and accuracy results were as

given below :

6.5 SWOT Analysis of this system

S	W	O	T
Language Independent Algorithm , Database is growing so that even if spammers change their techniques , algorithm will still remain stable	Meta data features are dummy that may limit performance	Application can server as basis of any text messaging application, feature extraction can made more stronger in order to improve performance.	There is no security involved in storing text messages.

6.6 Discussion

SMS Spam Detection is another world in it self and there is still a lot of room for improvements . We have given a basic structure and flow upon which multiple other great things can be done in order to make improvements in spam classifier as well as in user interface. Although most of design requirements are full filled in this application , however , it is still a beginner's level application and work can be done in order to improve it further. We plan to carry this work on and deploy this application as full fledged application.

Since basic idea was to make a spam classifier that may work independent of language , that has been achieved with satisfactory results . This applications bases are complete and that can be tested on android devices of version android 5 and above.

Due to high computations and limitations of frameworks that are android and python , spam classification is dependant on internet , however , this will not be the case in future because plan is to migrate this application from android to

python to avoid platform barrier. python server will also be terminated then and each user will be able to get his neural network trained in his own mobile phone according to his data only.

However , this system is still able to serve as base of SMS spam detection applications . It can also b used as embedded tool in builtin application of text messages for android. Due to its language Independence , it can also be able to operate in any other language and growing data base helps in making this app mature enough to deal with spammers new techniques.

References

Agarwal, S., Kaur, S. & Garhwal, S. (2016), 'SMS spam detection for Indian messages', *Proceedings on 2015 1st International Conference on Next Generation Computing Technologies, NGCT 2015* (September), 634–638.

Ali, S. S. & Maqsood, J. (2018), '.Net library for SMS spam detection using machine learning: A cross platform solution', *Proceedings of 2018 15th International Bhurban Conference on Applied Sciences and Technology, IBCAST 2018* **2018-Janua**, 470–476.

Alsmadi, I. & Alhami, I. (2015), 'Clustering and classification of email contents', *Journal of King Saud University - Computer and Information Sciences* **27**(1), 46–57.

Aski, A. S. & Sourati, N. K. (2016), 'Proposed efficient algorithm to filter spam using machine learning techniques', *Pacific Science Review A: Natural Science and Engineering* **18**(2), 145–149.

B. Rathod, S. & M. Pattewar, T. (2015), Content based spam detection in email using bayesian classifier, pp. 1257–1261.

Choudhary, N. & Jain, A. K. (2017), 'Towards Filtering of SMS Spam Messages Using Machine Learning Based Technique', pp. 18–30.

Cormack, G. V. (2008), 'Email Spam Filtering: A Systematic Review', *Foundations and Trends® in Information Retrieval* **1**(4), 335–455.

Delany, S. J., Buckley, M. & Greene, D. (2012), 'SMS spam filtering: Methods and data', *Expert Systems with Applications* **39**(10), 9899–9908.

Hammerstrom, D. (1993), 'Working with neural networks', *IEEE Spectrum* **30**(7), 46–53.

Kamoru, B. A. (2017), 'A Mapping Study to Investigate Spam Detection on Social Networks', **11**(11), 16–34.

Karami, A. & Zhou, L. (2014*a*), 'Improving Static SMS Spam Detection by Using New Content-based Features', *Americas Conference on Information Systems* pp. 1–9.

Karami, A. & Zhou, L. (2014*b*), 'Improving Static SMS Spam Detection by Using New Content-based Features', *Americas Conference on Information Systems* pp. 1–9.

Khalique, F., Butt, W. H. & Khan, S. A. (2017), Creating Domain Non-functional Requirements Software Product Line Engineering Using Model Transformations, *in* '2017 International Conference on Frontiers of Information Technology (FIT)', pp. 41–45.

Kitchenham, B. & Charters, S. (2007), 'Guidelines for performing Systematic Literature reviews in Software Engineering Version 2.3', *Engineering* **45**(4ve), 1051.

Lota, L. N. & Hossain, B. M. M. (2017), 'A Systematic Literature Review on SMS Spam Detection Techniques', *International Journal of Information Technology and Computer Science* **9**(7), 42–50.

Mahender, C. N. & Korde, V. (2012), 'Text Classificationandclassifiers: a Survey', *International Journal of Artificial Intelligence & Applications (IJAIA)* **3**(2), 85–99.

Mujtaba, G. & Yasin, M. (2014), 'SMS Spam Detection Using Simple Message Content Features', *Journal of Basic Applied Scientific Research* **4**(4), 275–279.

O'Mahony, J. (2012), 'Text messaging at 20: how SMS changed the world'.
URL: *https://www.telegraph.co.uk/technology/mobile-phones/9718336/Text-messaging-at-20-how-SMS-changed-the-world.html*

Panchal, F. S. & Panchal, M. (2014), 'International Journal of Computer Science and Mobile Computing Review on Methods of Selecting Number of Hidden Nodes in Artificial Neural Network', *International Journal of Computer Science and Mobile Computing* **3**(11), 455–464.

Sajedi, H., Zarghami Parast, G. & Akbari, F. (2016), 'Sms spam filtering using machine learning techniques: A survey', *Machine Learning Research* **1**(1), 1–14.

Shahi, T. B. & Yadav, A. (2014), 'Mobile SMS Spam Filtering for Nepali Text Using Naïve Bayesian and Support Vector Machine', *International Journal of Intelligence Science* **04**(01), 24–28.

Spam Dataset (2017).
URL: *https://www.kaggle.com/ishansoni/sms-spam-collection-dataset*

Uysal, A. K., Gunal, S., Ergin, S. & Gunal, E. S. (2013), 'The impact of feature extraction and selection on SMS spam filtering', *Elektronika ir Elektrotechnika* **19**(5), 67–72.

Wang, C., Danilevsky, M., Desai, N., Zhang, Y., Nguyen, P., Taula, T. & Han, J. (2013), 'A phrase mining framework for recursive construction of a topical hierarchy', *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '13* p. 437.
URL: *http://dl.acm.org/citation.cfm?doid=2487575.2487631*

Xu, Q., Xiang, E. W., Yang, Q., Kong, H., Bay, C. & Kong, H. (2012), 'ieee_intelligent_sys_SMS-Spam_Detection.pdf', (January).

Yadav, K., Kumaraguru, P., Goyal, A., Gupta, A. & Naik, V. (2011), 'SMSAssassin : crowdsourcing driven mobile-based system for SMS Spam filtering SMSAssassin : Crowdsourcing Driven Mobile-based System for SMS Spam Filtering', (November).

Zhang, X., Xiong, G., Hu, Y., Zhu, F., Dong, X. & Nyberg, T. R. (2016), 'A method of SMS spam filtering based on AdaBoost algorithm', *Proceedings of the World Congress on Intelligent Control and Automation (WCICA)* **2016-Septe**(July), 2328–2332.