Twitter Stream Data Mine System

D. P. Brandiwatta
Bachelor of Science (B.Sc.) in Information Technology (Undergraduate)
Sri Lanka Institute of Information Technology Computing Pvt Ltd

A. H. L. D. C. Kaushalya
Bachelor of Science (B.Sc.) in Information Technology (Undergraduate)
Sri Lank Institute of Information Technology Computing Pvt Ltd

K. V. D. L. Sulochana
Bachelor of Science (B.Sc.) in Information Technology (Undergraduate)
Sri Lanka Institute of Information Technology Computing Pvt Ltd

Nipunika Vithana
Lecture
Sri Lanka Institute of Information Technology Computing Pvt Ltd

ABSTRACT
Twitter is a popular social network where users can share short SMS-like messages called tweets. Users share thoughts, links and pictures on twitter, journalists comment on live events, companies promote products and engage with customers. In this research paper introducing a system to users a simple way to look for the tweets and system will give a overall percentage weather post is a positive or negative or neutral . When user search the person who have twitter account, User can see the first 100 tweets and it'll give the percentage with positive, negative and neutral with the emojis. This project based on naive based theory.

Keywords
Naive based, Social network, Twitter

1. INTRODUCTION
Twitter is a popular social network. Most of the people in the world are now using twitter. People using twitter because it's a easy way to connect with people in the world. This project will give a way to see what every person in twitter can search and look the comments tweets of the person. In modern days it's really hard to see the every comment.

Technology is the most common rapidly renewable and changing factor in our world. We have to move according to the technology day by day. Getting more information and to know the world and to do our works more easily and efficiently we use computers. There are several kind of things need to look at What type of technique need to use,

How to get these data from twitter, what type of language need to use for this project, need to developed the project to look over the mobile, what type of tools need to developed the project, what is the data mining algorithm to be used.

These problems need to solve to continue this project. And need a way to handle this data from mobile phone. It's easy to see over the mobile phone. Because nowadays most of the people in the world are using smart phones. So It's very easy to look from mobile phones. Because all applications are now can download over the phone. People are using social networks from their smart phones.

2. BACKGROUND AND RELATED WORKS
By using this system the users will be able to time manage by just read the comments users. This system users can use their twitter account from the mobile phone. using this system users can just search the people who the user want to search. using this system users will use twitter than the other social medias. By using this system the people who don't use twitter people will use twitter. Because it's really easy to use this software and it'll will addict to twitter. Because by using this system it'll give beautiful appearance on the system and it'll easy to search the people who want. Users who don't know about data mining and social medias by this system people will get a good knowledge. Because the system is using naive biased theory. Users will know about the theory by using the system. People who don't use social medias will get a good idea by using this system. And they will addict to use social medias. Users can connect the ip address through mobile phone, then users can easily through the mobile phones. When user try to find which person's comment that the user needs. By using the developed project users can identify by the emoji.To provide a easy way to read the comments.

By using the developed system the users can get solutions such as which users comment need to read. As a example user need to look post and comments of the president of the country, the user just need to type of the name of president Like that the user can search whatever person's comment by this system.

Table I. Comparison between similar Approaches.

Research Name / Proposed System	Sentiment Analysis of twitter data	Data mining in education	Twitter data mining using naive biased	Empirical data mining	Data mining in social network
Percentage With emojis	✗	✗	✗	✗	✗
Display positive, negative, neutral	✓	✓	✓	✗	✓
Used Naive Biased	✗	✗	✓	✗	✗

2.1 Sentiment Analysis of twitter data[1]
This research paper tries to develop sentimental analysis on twitter. And the system using in this research paper two types of methods. Positive and negative classes and positive negative and natural classes. Such popular micro blog called twitter and build models for classifying tweets into positive, negative and neutral sentiment. Build models for two classification tasks a binary task of classifying sentiment into

positive and negative classes and 3-way task of classifying sentiment into positive, negative and neutral classes.

As a advantage this research paper using two types of methods. but No need to use both methods single method will be easy.

2.2 Data mining in education: Data classification and decision tree approach [2]

This research paper aimed to develop a faith on data mining techniques. so that present education and business system may adopt this as a management tool. LIBSVM with Radial basis kernel has been taken as a best choice for data classification.

In this research work support vector machines are established as a best classifier with maximum accuracy and minimum root mean square error. The study also includes a comparative analysis of all support vector machine kernel types and in this the radial basis kernel is identified as a best choice for support vector machine.

2.3 Twitter data mining using naive Bayes multi-label classifier[3]

This research paper discussed how to input twitter data manually. All the data are saved in a data base. Combine with social semantic analysis and natural language processing.

The huge information source for users as a result we can see the increasing use of social media.

This paper presents systems to connect tweets

3. METHODOLOGY

The methodology section described the steps we were going to follow during the life cycle of the research project. This section mainly described requirement gathering, the process model to follow, the design and implementation techniques, resources and project management plan. When we consider the development pattern of our research we decided that the iterative waterfall model was suitable for the development of the system. In practice, defects were introduced during every phase of the software life cycle. Hence feedback paths must be added to the classical waterfall model. The resulting Iterative Waterfall Model was one of the most widely used process models. Defects were often detected late in the software life cycle, for example, a missing requirement might go unnoticed

until the system testing phase. Once a defect was detected, the team must return to the phase in which the defect was introduced; redosome of the work from that phase, and all subsequent phases. Ideally, errors should be detected during the phase in which they were introduced. A problem that was introduced and detected in the design phase can be fixed with relative ease. The same problem will be much harder to fix if it was not detected until system testing.

The aim of the Requirement Analysis and Specification phase was to identify the exact requirements of the system and to document them properly. Therefore our team analyzed the main functionalities of the system and finalized requirements.

After analyzing gathered requirements next step was to start the designing phase. We planned to use the Microsoft visual studio software to draw some diagrams like DFD, ER, Activity diagrams and Use Case Diagrams to get the idea about the system. We were going to implement a Desktop application with more functionalities than the existing systems. Before developing the real system we were going to design a prototype.

In this phase we were going to translate the software design into working system. We used PHP as our programming.

We will implement a Desktop Application with Multi-tasking facilities.

During this phase we were going to integrate different modules in a well-planned manner. As a result our three group members had separate functions to develop, Integration was normally carried out through a three steps.

- Unit Testing
- Integration Testing
- System Testing

We expected to do system testing to ensure that the developed system was working correctly. And also finally we planned to do acceptance testing to ensure that the client satisfies with the system.

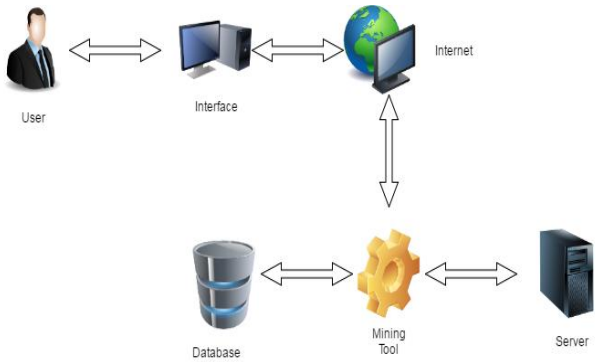

Figure 1 illustrates top level architecture of the proposed solution.

This research project discussed lot of researching in data mining for figure outing what are the most efficient and accurate ways of using the novel techniques to provide a better solution for read the tweets. We started our research based project with a master plan. So according to our plan we divided the whole project in to below major parts,

- Search the user.
- Get the data from twitter.
- Get the result as positive, negative and neutral.

4. DESIGN OF THE SYSTEM

Our proposed expert system mainly consists of following component; Main part of the design process is the data mining. Naive based used as the algorithm of the system. We search the person who is in twitter. Can search whatever person who has a twitter account.

Fist user need to register to use the system. Need to provide user name password and the mobile number to access the system. After register user can use the system. in the register interface. In the login interface user can login to the system. By providing the mobile number user will get a message to the phone as successfully registered. After user log in to the system, user can see the main page. In the main page user can find the search bar to search the person who has twitter account. user can search whatever person who has a twitter account. After search the person it'll appear the first 100 tweets of that person. User can easily see the tweets with the link. If user wants to go to that page user can easily go the page by clicking the link. User can read the tweets one by one easily. Our system providing the percentage so it's easy to

user think whether to read or not. And the percentage will divided into three parts. Positive, negative, neutral. Percentage will give with the emojis. If the tweet has a good percentage more than 50% the emoji will appear with the smiling face. If the tweet is less than 50% it will appear sad emoji. If it's a neutral one it'll appear a normal face. So it is very easy to user to think whether to read the tweet or ignore the tweet.

For the developing we used php and my sql as the data base. So for t he developing first we need to get the data from twitter. So it's a very hard one. First we need to register as a developer in the twitter to get the data. After register as a developer twitter will give the customer key, customer secret, access token and the access token secret. By providing these details user can get the data from twitter. Three arrays used for the positive negative and for the neutral data. . Mobile platform is leading technologies in modern day. Also with the usage and the demand of the system we will expand the updates of twitter which are to be recognized by the system in to a considerable amount. Most people have very busy schedule so this kind of system very useful for the future. And also to attract and help local community more we are planning to enhance the local language support for the system with Sinhala language. That kind of improvement increase user friendliness for the system. Those kind improvements want to be done in future to our system.

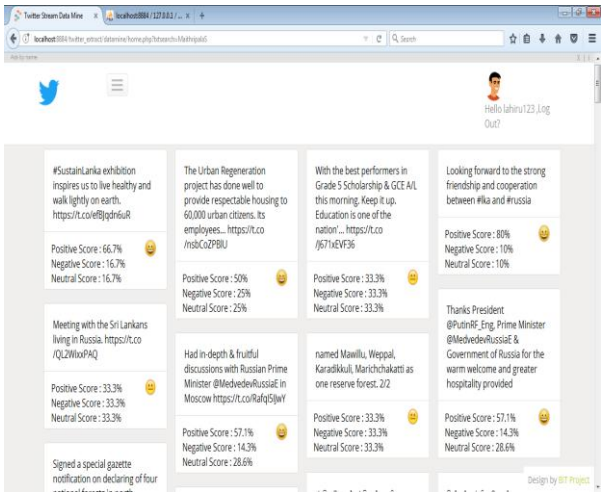

Figure 2 Interfaces of the system

5. 5. CONCLUSION

"Twitter Stream Data Mine" is a web based system which can Be use by every people in Sri Lanka. This research project helps to the people who spend busy life. People can use twitter better than it was. No need to read every comment. Only things to do in this system is first need to register as a user then just need to type the user name and the password. Then system identify the user then directly user can work with the system. After that user can search whatever person in twitter to view the comments.

6. FURTHER WORK

As a further development to the system we are targeting to expand the multi-platform capability through mobile support. Also the computer vision algorithms will be accelerated with. Mobile platform is leading technologies in modern day. So we are targeting release Android mobile platform and IOS compatibility in the near future. Most people have very busy schedule so this kind of system very useful for the future. And also to attract and help local community more we are planning to enhance the local language support for the system with Sinhala and Tamil languages. That kind of improvement increase user friendliness for the system. Those kind improvements want to be done in future to our system.

7. LIMITATIONS

In here we are arrange this application only for 100 tweets. We developed for limited words. We only develop this for windows application so that is not yet develop for smart phones like Android, IOS and etc. For English language and for other ordinary languages likewise Sinhala Tamil that will not support. They are the limitation of our system and we will improve these limitations in the future.

8. ACKNOWLEDGEMENT

This research was supported by Sri Lanka Institute Of Information Technology and Supervisor Ms. Nipunika Vithana. The research group thanks our colleagues from D.P.Brandiwatta A.H.L.D.C.Kaushalaya K.V.D.L.Sulochana who provided insight and expertise that greatly assisted the research.The research group would also like to show our gratitude to the Ms Nipunika Vithana for sharing their pearls of wisdom with us during the course of this research, the group members are also immensely grateful to Ms. Nipunika Vithana for their comments on an earlier version of the manuscript, although any errors are our own and should not tarnish the reputations of these esteemed persons.

9. REFERENCES

[1] "Sentiment Anal//ysis of twitter data",http://www.cs.columbia.edu/~julia/papers/Agarwal etal11.pdf.

[2] "Data mining in education: Data classification and decision tree approach", International Journal of e-Education, e-Business, e-Management and e-Learning, Vol. 2, No. 2, April 2012, Available:http: www.ijeeee.org/Papers/097-Z00080E10038.pdf.

[3] "Twitter data mining using naive Bayes multi-label classifier", International Research Journal of Engineering and Technology (IRJET), Volume: 03 Issue: 06,June 2016.Available: https://www.irjet.net/archives/V3/i6/IRJET-V3I6397.pdf.

[4] "Empirical Analysis of data mining techniques for social network websites", An international journal of advanced computer technology, 3 (2), February-2014 (Volume-III, Issue-II). Available: http://www.ijact.in/index.php/ijact/article/viewFile/261/213.

[5] "Filtering and classification of user based on social media data using memetic and naive bayes methods", International Journal of Computer Science and Information Technologies, Vol.6(5),2015.Available: http://ijcsit.com/docs/Volume%206/vol6issue05/ijcsit20150605138.pdf.

[6] "Data mining in social networks", https://www.cs.purdue.edu/homes/neville/papers/jensen-neville-nas2002.pdf.

[7] "Implementation of data mining in analyzing social media users personality with naive bayes classifier: A case study of instagram social media",International Journal of Computer Science Issues, Volume 13, Issue 4, July 2016,Available:http://www.ijcsi.org/papers/IJCSI-13-4-76-82.pdf.