Android Application using Review Mining: Ceylon Tour Guide

M.Nishanthan[1], S.Kavitha[2], S.Mosharani[3], S.Amuthapriya[4] and V.N.Vithana[5]
Faculty of Computing,
Sri Lanka Institute of Information Technology,
Sri Lanka.
[1]n.m.nishanthan@gmail.com, [2]suthaharankavitha@gmail.com, [3]mosharu19@gmail.com,
[4]amu.shun1992@gmail.com [5]nipunika.v@sliit.lk

Abstract— **In the 21st century, Sri Lanka is one of the most visited tourist countries in Asia. Though, the rate of tourists visiting Sri Lanka is rapidly increasing, most of these tourists do not have a clear idea of where to visit. Several web sites such as Agoda, Trip Advisor let tourists see comments and rate places. There are very few applications which assess these public reviews and provide a comprehensive rating specially tailored to Sri Lanka. Hence tourists end up spending a lot of time and effort in looking for locations of their interest. To overcome this problem, our system, Ceylon Tour Guide presents a Tour guide application that provides a comprehensive rating of tourist locations in Sri Lanka using text mining techniques. Public reviews from well-established tourist web sites such as Agoda are collected and stored in a database. The stored details are mined with the text mining algorithms. With the aid of algorithms, mined data will be categorized into positive, neutral and negative reviews. Using the Android application user can search and see the ratings of the important locations. After ending 30 years of civil war, the tourism sector has a vast development. This application will help popularize the important locations in Sri Lanka including north and thus, help increase the economic growth of the country.**

Keywords: Text mining, Sentiment Analysis, Tour Guide, Android application, Sri Lanka, Tourism

I. INTRODUCTION

Sri Lanka has a growing tourism industry. Since gaining independence from the British in 1947, Sri Lanka has continued to attract foreign investors and tourists to the island. The country's important placement also enables it to attract transit visitors into the island [1]. As the rate of tourism increases the tourist should know more about the places to visit.

There are popular tourist web sites that give reviews on various hotels and locations from various tourists that visited the place. Now it has become a practice for websites, to facilitate the expression of opinions by guests and visitors on presented places. Also, the expansion of social networking facilitated tourist posting opinions on the web site. Though Sri Lanka has high tourist turn ups currently there are no applications that give detailed and reliable opinions on Sri Lankan places to visit. Most of the time tourists leave a comment on the places they have visited on popular web sites. By processing these comments we can get a valuable and reliable opinion on the various locations. Tourist will find these overall opinions very valuable when deciding what places to visit.

"Ceylon Tour guide" is an android based application which retrieves reviews on places visited from popular web sites, classify the reviews to positive or negative or neutral using text mining and display a comprehensive rating in a user friendly and efficient way. Classification of reviews in both positive and negative classes is done based on naive Bayes algorithm. As training data, a collection of sentences (pre-classified in positive and negative) was taken from the reviews.

Tourist place guiding system using text mining will be beneficial and helpful to the tourist who is willing to come to Sri Lanka for their vacation. The next sections of this paper discuss the background, methodology, results and discussion, conclusion and future work for this system.

II. BACKGROUND

The research team has gone through some of the existing researches, applications, books and etc. with the purpose of getting some knowledge of the literature review for the research "Ceylon Tour Guide: Android application using reviews". For this proposed system there is no review based application in the tourist sectors. The tourist or people who are visiting a place for the first time in

Sri Lanka, find it difficult to choose the best place to visit.

In this era reviews, based applications are increasing rapidly. In every application or websites, there is a separate part for the reviews. In that area, the reviews of particular things are collected by the application. With that reviews, the system will do an analysis of the place or product whether that is good or not. Here are some researches discussed the review done with the text mining, text mining algorithms used in mining the data that collected and how this text mining will help with the review mining etc.

The researches have been conducted early helped to gain the knowledge about the text mining and able to come with the best solution for this proposed research system. Some of those researches done earlier are briefed below.

[2] proposed a text mining system customized to improve laboratories efficiency by taking the electronic commends as input for that proposed mining system. Inhere used the basic text mining algorithm, will almost be modified by adding new steps, modify some steps by some customizing Natural Language Processing, data mining techniques and building the document database.

[3] develop a decision-support component of Tsunami early warning system by using twitter. This project approaches using focused twitter crawling, trustworthiness analysis, reparsing, and multilingual tweet classification in the context of how they could be used for monitoring crises and verified in different experiments based on human annotated gold standard corpus.

[4] proposed a text mining and sentiment analysis system to state or emotion of writer's expression in words. The proposed approach is coined hierarchical classification and the effects of the approach in a different combination of classification algorithms and filtering schemes are discussed over three sets of controversial online news articles where binary and multi-class classifications are applied.

The text mining techniques to review the movie reviews and customer review for the product are discussed here. They give a basic idea about the mining the review and how it is categorized in to tables with different options. With that categorized table they generate the class based summary of the reviews are discussed. Feature based summarization is evaluated with the perspective of feature extraction, opinion sentence extraction and they check the accuracy of orientation prediction of opinion sentence [5], [6].

[7] discusses the techniques used in the text mining, and the roles played in the text mining. They have illustrated the each technique and described separately. The algorithms used in each technique are discussed here.

In [8] their aim to generate a reliable classification approach of customer reviews based on an existing domain-specific corpus by applying a lexicon-based sentiment analysis. They build a lexicon of those text components with a semantic orientation .They apply a sentiment analysis based on the lexicon order to generate a classification. Their classification results are evaluated against a set of reviews with quantitative ratings.

[9] they implemented the proposed algorithm and they tested its performance for movie user review. Algorithm classified words without lexical content (prepositions and pronouns).Positive and negative classes are done based on a naive Bayes algorithm. They removed insignificant words and introduced in classification groups of words (ngrams).

The main objective behind that system is that tourist can easily extract subjective and useful information. Their system also reduces the time required for searching and easily supports the growth of tourism. The objective of this project is to process the reviews, of tourist places, collected from a website based on sentiment analysis. They provide suggested places to the user based on their requirements and show top most review of a particular place [10].

They begin to investigate this question. Another challenge of micro blogging is the incredible breadth of topic that is covered. It is not an exaggeration to say that people tweet about anything and everything. Therefore, to be able to build systems to mine Twitter sentiment about any given topic, they need a method for quickly identifying data that can be used for training. In this paper, they explore one method for building such data: using Twitter hash tags (e.g., #best feeling, #epic fail, #news) to identify positive, negative, and neutral tweets to use for training three way sentiment classifiers. [11].

[12] states in the context of social networks, the simplest way to compute user similarities are by means of accepted similarity metrics. Finding similar users and modeling their profile and social relationships is a key issue in several research fields. Discussions among users of social networks are potential of great interest. Millions of users, in fact, share through social networks information, news, and general events. This user discusses and shares posts with other Facebook users (who) related to various topics (what). They may express positive, negative or neutral opinions. It is a text analysis software program designed to process written text by looking for a dictionary match for each word composing the text. If the target word matches the Dictionary

word, the appropriate word category scale for that word is increased. The sentiment analysis phase does not require any external dictionary of polarized words to catch the sentiment polarity in everyday colloquial texts. The approach makes it possible to map colloquial expressions with new words, slangs, and errors.

The destination image is one of the topics that are widely studied in the literature. This is because it is considered a powerful tool for destination marketers, in order to obtain a high competitive advantage in the tourist markets The image formation prior to the visit, has been highly influenced by the information provided by travelers' worth of mouth (WOM).This is encouraged customers to interact and share experiences through an increasing number of reviews, based not only on product quality but also on service experiences. Trust and previous experience are also key mechanisms to reduce the uncertainty from users' reviews although the user posting the reviews can also score the product or service under reviews [13].

[14] tell about Customer reviews are recognized as fruitful information sources for monitoring and enhancing customer satisfaction levels, particularly as they convey the real voices of actual customers expressing relatively unambiguous opinions. A customer review is utilized frequently to monitor and, so enhance customer satisfaction with mobile services. Sentiment analysis is a technique for identifying the ways in which sentiments are expressed in text and for determining whether they represent positive or negative feelings toward a specific product or service. For mobile service providers, customer reviews can be a useful and relatively low-effort means of understanding the minds of customers. Knowledge thus gained, furthermore, is quite reliable, since customers write reviews voluntarily.

III.   METHODOLOGY

This system was developed using the prototype methodology, this methodology was selected because there may be many complex situations as we were new to the technologies. As undergraduate students engaged in the final year research project, the team had to learn about many areas such as android application development, text mining algorithms etc. The project team decided on the project objectives and planned to develop the system in 10 months with a team of four members

To develop this system the team prepared questionnaires and analyzed tourist requirements as the primary data gathering. And also done the major literature review as the secondary data

gathering to clarify more about the mining algorithms and process to the text mining. The project team used object oriented approach to doing the design. Architecture design in figure 1 illustrates the infrastructure of hardware, software and communication components to be considered in implementing the proposed system.
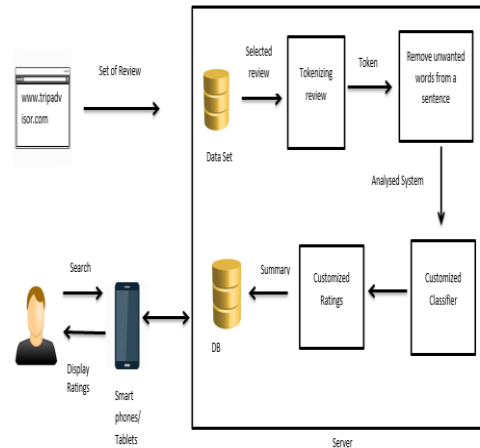

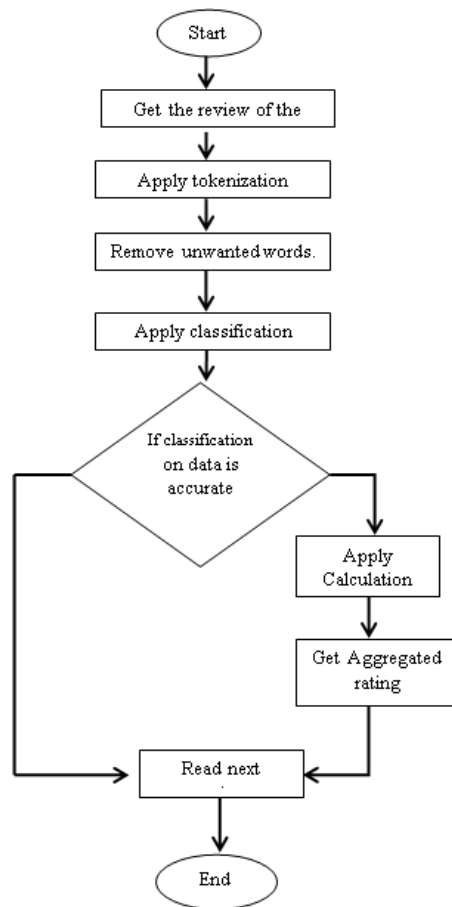Figure 1. System Architecture diagram

Figure 2. Flow chart

The system uses the website to take the reviews for the analysis, as there are many issues in taking their reviews, the team has created a new website (dummy system) for the research purpose. In the website reviews for the tourist places in Sri Lanka are reviewed and stored in the database. To do the text mining the system has the server side. In the server, every review is mined using text mining algorithms. The algorithms used to mine the data are tokenization, removing undesired words, customized classifier and customized rating. With the aid of the algorithms, reviews are mined and the final rating is stored in the database. Finally, the server and the android application are connected to display the rating of a place. The tourist can view the description of the place and final aggregated rating of that place.

Figure 2 shows the flow chart of the system. Android and C# were used for the programming. The MS SQL is used for the development of the database. Every module in the System was tested with Unit Testing and System Testing.

IV.   RESULTS

Results of the system are shown in the screenshots of the interfaces and also shows the final outcomes from the android application and the server implementation. In this evidence section, it briefly discusses the output with the performance, reliability and the accuracy of the system.

A.   Android Interfaces

Figure 3 shows the interface user can search for the places in Sri Lanka or can select a place which is listed in the interface.
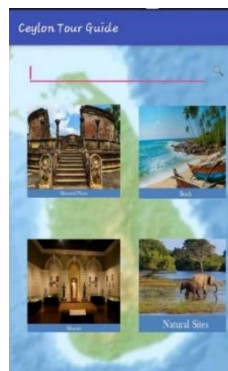

Figure 3. Android interface to search a place

Figure 4 shows the interface that displays the place under the historical sector. The tourist can choose their desired place. By selecting the major place of Sri Lanka the user can see the sub places in Anuradhapura, if the user knows the place they can simply select the place.
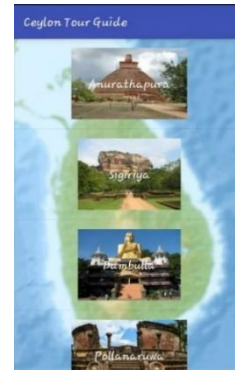

Figure 4. Android interface to display the place in Sri Lanka

Figure 5. Android interface to display the place in Anuradhapura

Figure 5 shows the interface displays the place in Anuradhapura. The tourist can choose their desired place.
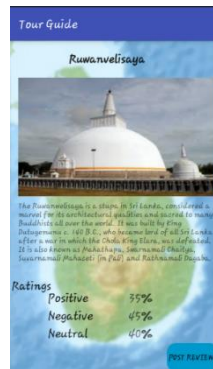

Figure 6. Android interface to display the rating of Ruwanvelisaya

Figure 7. Android interface to insert new review

Figure 6 shows the interface the user can see the description of a particular place and the ratings of the place. If that user wants to add a review the user can add new reviews also. Figure 7 shows the interface the user can enter the reviews of a particular place. The user can simply type their thought and like about the place through this interface.

V.   DISCUSSION

In this tourist era, the existing applications are not very user friendly and efficient to reduce the time to search a place. The user has to face various problems while searching for a good place to visit during their short period of vacation. This research "Ceylon Tour Guide", is designed to solve some of these issues faced by the tourists. The system is thoroughly tested to ensure accuracy and reliability.

During the development process, the research team faced many problems such as connecting the two different databases, connecting android with the database, issues with the security, providing aggregate rating and Lack of good reviews in the website to mine.

The team used the following techniques to solve the above mentioned issues in order to complete the system in a successful manner

a) Used the modified algorithms to do the proper ratings.
b) Used web service to connect android with the database.
c) Used password oriented to login to the system.

When designing the "Ceylon Tour Guide" system there were some assumptions perceived. There are listed below

a) Users know how to search for information and he/she can understand the flow of the system.
b) Users have basic knowledge of the using smart phone as well as working with Android applications.
c) The connection between server and android should not be lost.
d) The user account has user name and password.

The algorithms that we use to mine the data will give the accurate results for the rating in the form of positive negative and neutral. The accuracy is gained by the process of the system that is, set of reviews are gathered from the website and the reviews are mine together for a particular place and the customized algorithms are used to divide among the rating groups (positive, negative, neutral) and the rating is done with aid of the Bayes rule and it gives the 80% ratings for positive negative and neutral. Because some reviews are in a sarcastic manner it includes some words which look like positive but the real meaning will give you the neutral meaning. Because of that, it will give the 80% accuracy.

VI.   CONCLUSION AND FUTURE WORK

Sri Lanka is developing the tourism sector. "Ceylon Tour Guide" will be very beneficial for the tourists visiting Sri Lanka. The team has used a set of techniques and algorithms to mine the reviews of a particular tourist place. The objective is to provide an aggregated rating of the tourist reviews for a tourist place. Experimental results show the effectiveness of the proposed approach. In addition, with the proposed approach, it is more efficient and easy to obtain a comprehensive rating. This system will highly impact on the tourism sector as currently, the tourist economy is growing vastly.
Limitations of our systems are stated below,

a) The accuracy of overall rating depends on the public reviews retrieved from the web site.
b) This system works with android platform.

The above mentioned problems in the limitations can be solved as follows.

a) Encourage the users to give correct and usable reviews.
b) Build this application on other platforms such as ios.

In this digitalized era, people rely more on technology for opinions than on people such as tourist guides. Android application "Ceylon Tour Guide', is a timely application that would help improve the tourism in Sri Lanka. The team plans to improve and expand Ceylon Tour Guide in the future to provide the following:

a) Show the path of the tourist place by using the GPS technology
b) With the image processing, the description about the place will be displayed.
c) The image gallery will be created to upload images of the places.

VII.   REFERENCES

[1]   Overview of Tourism Industry in Sri Lanka. (n.d.). Retrieved September 3, 2016, from http://www.sltda.lk/overview_of_tourism_in_sri_lanka

[2]   Tariq, D. (2013). Improving Laboratories Efficiency through Website Using Text Mining. Eng. &Tech. Journal, 31(2).

[3]   Andrea, Z., Stuart E, M., Laurissa N, T., & Xinyue, W. (2013, May). Social Media Text Mining and Network Analysis for Decision Support in Natural Crisis Management. 10th International ISCRAM Conference.

[4]   Li, J., Fong, S., Zhuang, Y., & Khoury, R. (2014). Hierarchical Classification in Text Mining for Sentiment Analysis. 2014 International Conference on Soft Computing and Machine Intelligence. doi:10.1109/iscmi.2014.37

[5]   Zhuang, L., Jing, F., & Zhu, X. (2006). Movie review mining and summarization. Proceedings of the 15th ACM International Conference on Information and

Knowledge Management - CIKM '06. doi:10.1145/1183614.1183625

[6]   Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. Proceedings of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '04. doi:10.1145/1014052.1014073

[7]   Gupta, V., & Lehal, G. S. (2009). A Survey of Text Mining Techniques and Applications. Journal of Emerging Technologies in Web Intelligence, 1(1). doi:10.4304/jetwi.1.1.60-76

[8]   Gräbner, D., Zanker, M., Fliedl, G., & Fuchs, M. (2012). Classification of Customer Reviews based on Sentiment Analysis. Information and Communication Technologies in Tourism 2012, 460-470. Doi: 10.1007/978-3-7091-1142-0_40

[9]   SMEUREANU, I., & BUCUR, C. (2012). Applying Supervised Opinion Mining Techniques on Online User Reviews. Academy of Economic Studies, 12.

[10]   Sharma, D., Kulshreshtha, A., & Paygude, P. (2015). Tourview: Sentiment Based Analysis on Tourist Domain. International Journal of Computer Science and Information Technologies.

[11]   Kouloumpis, E., Wilson, T., & Moore, J. (2011). Twitter Sentiment Analysis: The Good the Bad and the OMG! Association for the Advancement of Artificial Intelligence.

[12]   Kang, D., & Park, Y. (2014). Review-based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach. Expert Systems with Applications, 41(4), 1041-1050. doi:10.1016/j.eswa.2013.07.101

[13]   Gonzalez-Rodriguez, M., Martinez-Torres, M., & Toral, S. (2014). Monitoring Travel-Related Information on Social Media through Sentiment Analysis. 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing. doi:10.1109/ucc.2014.102

[14]   Terrana, D., Augello, A., & Pilato, G. (2014). Facebook Users Relationships Analysis Based on Sentiment Classification. 2014 IEEE International Conference on Semantic Computing. doi:10.1109/icsc.2014.59

65