Reliability of Mobile Application Analysis Based on The User Reviews of Health and Fitness Category Which are Available from Google Play Store

Mr. Donbosco Prashanth
169326G

Dissertation submitted to the Faculty of Information Technology, University of Moratuwa, Sri Lanka for the partial fulfillment of the requirements of Degree of Master of Science in Information Technology

March 2019

Declaration

We declare that this thesis is our own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Donbosco Prashanth

Signature of Student

Date:

Supervised by

Mr. Saminda Premaratne

Senior Lecturer

Faculty of Information Technology

University of Moratuwa

Signature of Supervisor

Date:

Acknowledgments

Completion of a project dissertation has a challenging task which requires an intense effort and dedication which must be endured for a considerable period of time, continually. The assistance, guidance, and familiarization of the author for unfamiliar concepts and knowledge by external but relevant parties would be imperative in addition to the aforementioned qualities have been imperative for successful completion of such a high-level assignment.

First and foremost, I would like to extend the sincere thanks and appreciation to my supervisor Mr. S. C. Premaratne, Department of Information Technology, Faculty of Information Technology, University of Moratuwa. His guidance and expertise had helped for the successful completion of this important assignment which has been the final frontier to conquer in the MSc Program. He has been the driving force for this author to select this topic and through the final project.

I wish to express my heartfelt gratitude to all academic staff of the Faculty of Information and technology, in particular for the opportunity given to be a student of such a valuable course of MSc program and for providing assistance in numerous ways by obliging many requests of the author for this successful completion.

The associates are remarkable in assisting me in my examinations in numerous ways. I would like to extend the heartiest gratitude for the great contribution and moral support from my sister and siblings and would like to mention the great bunch of batch mates who had been a wonderful motivational factor that had made the learning process of this MSc program highly fruitful, entertaining and knowledgeable.

Abstract

The current world everything embedded in the mobile phone and it has many uses in the field of life. Google apps store is a very important role to playing the mobile application development. This research proposed the reliable Health and fitness mobile application based on the Apps reviews which are given by end-user.

An end-user of mobile application review has to be analyzed in various methods because all the users do not have the same knowledge and same requirements but their review or feedback should depend on their knowledge, requirement and user experience.

In this research mainly we are targeting non-technical attributes which are gathered from Google play store to analyze the reliability of health and fitness mobile application based on user review, using the Natural Language Process techniques and classification models. In the current world has live technology improvements and day to day different need of user requirements so this will reflect the mobile application world trend, and it will make live changes in google play app store also.

Based on user review we can to find out proper path to analysis the google play store data to get an effective result in this research because that apps store has more complicated data in a different perspective of Health and Fitness area.

Research has tended to focus for Google play apps store customer has easy to find out the efficient application of Health and Fitness and comparing other mobile apps which are placed in the same apps store. Accordingly to that concept of our research final data will be produced in a structured database. In future idea from this updated database link with the mobile application to upload the google play store for the user access and user can easily predict the which application is efficient of their appropriate requirement.

Table of Content

ABBREVIATIONS

NLP	Natural Language Processing
WEKA	Waikato Environment for Knowledge Analysis
GUI	Graphical User Interface
TFIDE	Term Frequency Inverse Document Frequency
LSA	Latent Semantic Analysis
NLTK	Natural Language Tool Kit
API	Application Programming Interface

List of Figure

List of Tables

Chapter 1

Introduction

1.1 Prolegomena

In the contemporary world, every human being depends on the mobile phone and its services because of the trend of the current generation. We have been witnessing for the past five years an explosion in the reputation of mobile devices and mobile applications [1]. In fact, recent market research shows that the centralized application market for the Google platform (Android) has more than 1.5 million applications [2], under 42 types of categories.

As the smartphone usage rate increases, the consumer often uses a lot of mobile applications to improve their quality of life. A difficult area in this general user steps how to find the proper application of their requirement of 1.5 million applications.

Few types of research have addressed an issue of finding out the useful Mobile application from the Google play store has grown in importance in the light of the recent mobile application world. This study aims to explore and predict efficient mobile application in play store for general users. Before download the mobile application they can predict and provide analysis report in the usage of particular application based on the previous feedback of users and total download, total users, version support and product details.

However, previous work focused only on the problem of future trends in software engineering research for mobile applications [3], but this research only helps general software engineers understand the detailed report.

This paper propose a new approach to find out the suitable application for the user based on the previous customer feedback. Most studies has only focused the technical attributes of the Google API but in this research mainly concern with customer reviews and analysis thus to sentimentally using text miming techniques. Thereafter, create some summary database using data mining techniques then applying the proper efficient algorithm to find a suitable application.

Finally our research can predict and come to a conclusion in the various form of usage of a mobile application which is placed on the Google play store. Based on this outcome of the results to expand as a structured database published online. In future idea from this updated database link with the mobile application to upload the Google play store for the user access and user can easily predict which application is efficient of their appropriate requirement.

1.2 Problem Statement

In the current trend, mobile devices have surpassed fixed Internet access, mobile applications and distribution platforms have become more important. Although the Google apps store lists thousands of mobile phone health apps, it's not always clear that these apps are supported by trusted sources for end users.

The high usage and trust in this platform and their applications make their reliability a critically important goal to achieve. Some apps store application has minacious errors that affect end-user experience and their mobile phone physical resources.

App stores allow users to search, purchase, and install mobile apps, and provide reviews in the form of reviews and ratings. As a result, user reviews are a valuable part of analyzing the reliability of mobile apps. They are available on the Google Apps Store in the Health & Fitness category. The reliability of Google Apps store applications will play an important role in maintaining trust in end-user healthcare.

In this research proposed to analyze the reliability of health and fitness applications based on user reviews which are available from Google Apps Store.

1.3 Natural Language Processing Approach

Natural Language Process approach has various such as Rule based, Traditional Machine learning and Neural Network methods and these approaches are current world trend to make decision making from a large set of textual data. In this research, we used only small features with efficient way.

1.4 Goal of Research

The goal of research has analysis the reliability of mobile application based on user reviews which are available from the Google Play Store under the Health and Fitness category. We selected the top 35 mobile application for this research. For this analysis, we apply the sentiment and topic wise analysis using NLP methods to find out the user opinion and content of the features has in particular mobile application.

1.5 Aim and Specific Objectives

The main aim of this research is to analysis reliable health and fitness application based on the user reviews which are available from Google Apps store.

1. Data sets will be extracted from the Google apps store.
2. Identify suitable methods for NLP.
3. Data pre-processing
4. Implement the NLP models
5. Classifying appropriate techniques
6. Analyze result and discovering trend of the user reviews which accordingly.

1.6 Overview of the report

In this report, we write the progress of the research carried on. The first chapter includes a summary of an overview of the research where the second chapter includes the literature review survey done based on the topic. The third chapter has summarized the technologies into four major areas, and fourth chapter describes our approach. The fifth chapter includes the design of the research and analysis which is still carrying on. Chapter 06 covers the implementation details where chapter 07 evaluates the methods used in the implementation. Finally, chapter 08 discuss the result, limitations and future development for the solution.

1.7 Summary

This chapter introduced the research problem and the solution for Analysis reliable health and fitness application based on the user reviews which are available from Google Apps store. In the next chapter, we review some researches carried out related to our research problem.

Chapter 2

Literature Review

2.1 Introduction

In this chapter, we discuss some of the finding recently done in user reviews of mobile application analysis. We have selected 14 research papers based on our main goal. All the technical reports are published after 2008. A large variety of techniques and algorithms are considered when selecting research papers for the literature review.

The most widely used. The most common text mining approach is to represent text based on keywords. A keyword-based methodology can be combined with other statistical elements (such as machine learning and pattern recognition techniques) to discover the relationships between different elements of a text by recognizing the repetitive patterns present in the text content.

Mobile app developer markets are not just a place where developers can upload their app, and the user can download it. To create a sufficiently large dataset, we removed Google Play's popular apps reviews. Android-market-API [7] Application markets also have a scoring and revision system in place that allows users of these applications to describe their opinion of it as free text [3]. There is a huge number of applications in Google app store that can group data in a text format and offer to use text mining approaches such as the term-based method (TBM), the method based on sentence (PBM), the concept-based method (CBM) and model taxonomy method (TPM) [8]. In this case, when we use a sentence-based approach, its information is more semantic, which creates a lot of redundant noise. Some users typing advisories quickly do not use the appropriate expressions. At this stage, the concept-based approach can be applied because this approach can extract the most useful information and knowledge hidden in the text content and improve the overall analysis.

The corpus approach is the most commonly used way to obtain information by putting emotional words into a large corpus and gaining the emotional score. For example, research done in the past has put the word "happy" and "sad" in the corpus to evaluate the happiness factor of blogs [9]. When analyzing user feedbacks to easily identify these emotions then only we can reach the user's conclusion as this app or not as according to share their reviews on the platform for other users after using the product.

One of the risks in pursuing the above lines of research is that we may have reached the limits of NLP by analyzing poorly user-written reviews. Another risk is that users may prefer the features that are provided to them before requesting them, and when the user complains about these features, it is already too late [5]. The only solution might be to create an updated review system for application stores that provide a better mechanism for feature requests from users.

There are a number of empirical and exploratory studies on the importance of app's reviews in the app development process. In [8], Vasa et al. made an exploratory study about how users input their reviews on app stores and what could affect the way they write reviews. Later, Hoon et al. QT[7] have analyzed nearly 8 million reviews on the Google App Store to discover several statistical characteristics to suggest developers constantly monitor the changing user expectations of mobile apps.

Based on this assumption, MARK ranks keywords based on their association with negative reviews. It measures this association using a metric called contrast score [10]. k-means becomes an excellent resolution for pre-grouping, leaving space in separate and smaller subspaces where others

Clustering algorithms can be applied. In this study needed to analyse the user reviews of the mobile application must be categories and can be divided into several sub-categories there after we can apply the proper, efficient algorithm to analysis the dataset via k-means methods. It can produce reasonable answers for prediction of the good mobile application in a Google apps store. The following formula explains the user reviews counts and rating counts of the mobile application. It will produce effective answers.

$$\varphi = \frac{n}{p} \times (n - p)$$

In this formula, n and p correspond accordingly to negative (1-star and 2-star) and positive (4-star and 5-star) reviews containing the interesting keyword.

Shabtai et al. [11] extracted feature information from the manifest, XML files, API calls and methods used from a set of 2,285 Google Play apps. This large number of methods they can find out with technical attributes as non-technical attributes. This study is applying many various methodologies to analyse the data from the Google play store such as security, usability, fault allocation range and user reviews. In this point of view they more consideration aim with a technical data, but end-user cannot understand the actual problem which is arise in the application so we have to assume the end-user is not a technical person he/she must be a general human. In this study, most authors consider the correlation between technical and non-technical attributes available in the Google Play Store.

J. Xu, D.Ho and L.F Capretz have been identified, the Reliability is an important attribute of software quality in addition to other attributes such as performance, usability and fault prediction [15]. In our view, their findings are said technically analyzed performance, usability and fault prediction, but our study analysis the same attributes from the end-users point of view, based on this statement our research chose the sentiment and concept analysis methods.

Chen et al. [12] compared the maturity ratings of 1,464 equivalent apps between Apple App Store and Google Play, and taking Apple Store ratings as exact ratings; the authors found that 9.7% of Android apps were underestimated and 18.1% were overestimated. The authors also studied a sample of 729,128 reviews from 5,059 Google Play game apps and formed a classifier on user app and comment description sets, as well as ios maturity ratings for automatically check the application maturity ratings. Haet al. (891 users manually reviewed 556 reviews from 59 Google Play apps, categorizing them into topics and subtopics based on content, and found that most of the information in reviews was about the quality of the app. and not security or privacy issues

A recent review of the literature on identifying topics of interest word of the sentence, mostly used method is WordNet. WordNet has a group of synonymous words and they can find out the related word in the relevant topic [16]. It has been suggested the subjective of the sentences are interconnected with the groups of the words and using many to many mapping word can predict the topic.

Technology	Algorithm
Sentiment classification for camera, book, GPS reviews	Artificial neural network, SVM
Polarity analysis for the sentiment of User review	Naïve Bayes classification algorithm
Sentiment classification of user reviews	Naïve Bayes, SVM
Topic modeling	Latent Analysis
Product reviews of Sentiment analysis	Semantic analysis and machine learning

Table 2.1 Summary of the literature survey

2.1 Summary

This chapter discussed the findings of our literature study and summary of the literature reviews and that indicated in table 2.1. It shows how different techniques are used to analysis the user reviews and how applied the specific algorithm to achieve different targets.

Chapter 3

Technology Adapted in Mobile Application Analysis

3.1 Introduction

In the previous chapter, we discussed different findings in the area of mobile application sentiment analysis. This chapter presents different technologies and methodologies used in an analysis of mobile application user reviews.

3.2 Method for extract the features from user reviews text

Major problem of extraction of the features from the text we have to identify the unnecessary words such as digits, punctuation, symbols, stop words and whitespace. NLP has some techniques to remove those types of characters but the problem is some time user expressed their emotion via symbols like Emojis. So when we have to check the user reviews, we must identify the important emotional symbols also. If we analysis those emotions it will add some more accuracy of the result to enhance our predicted level of sentiment analysis of user reviews.

3.3 Method of identifying sentiment polarity classification of user reviews

The major problem of mobile application user reviews analysis; each customer review is composed of textual feedback of the customer's experience from that textual review we want to identify that review has positive, negative. Some kind of reviews does not contain the positive or negative this area, we have to predict it is neutral.

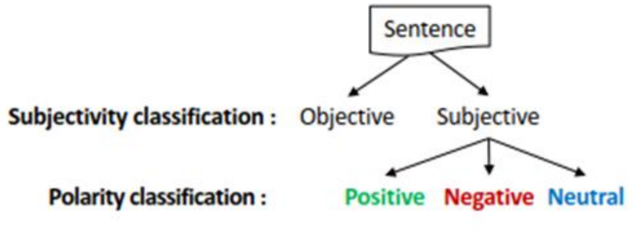


Figure 3.1 Sentence analysis

An example of an objective sentence is "This application for health" is health, while an example of a subjective sentence is "The application is awesome." Polarity classification is the task that distinguishes phrases expressing negative, positive or neutral polarities. Note that a subjective sentence cannot express any positive or negative sentiment (e.g., "I guess he has arrived"). For this purpose, it should be classified as "neutral". The sentence should be divided as a word, and each word has tokenized then tokenized words meaning to match with a dictionary to find out proper opinion, then classification those words express any sentiment value through this method can be used to enhance the result of user reviews analysis.

3.4 Method for identifying topics in a user review

This problem can be outlined topics to separate in the different language words has the same meaning or when we are modeling the topic same word include different topic. We have to map the word in a relevant topic of interest. In this case, the data has to organize and each word check with another word for finding similarity then clusters those words and organizing large blocks of a textual database on feature selection. This model can be summarized into the following model given in figure 3.2

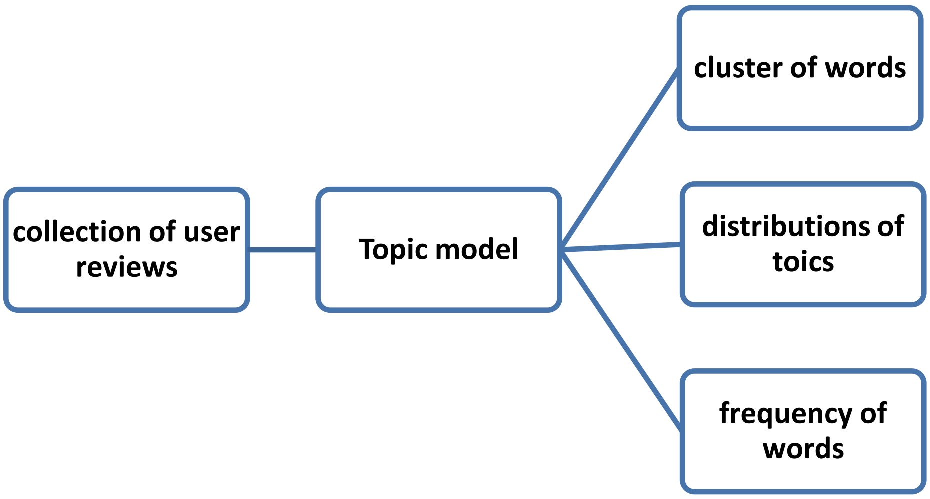


Figure 3.2 Topic model

Each topic, we identify a list of representative reviews summarizing public opinion on this topic. This is achieved by ranking all relevant reviews for a subject according to certain metrics and choosing the top 6 can be used for NLP techniques such as Wordnet. This method can be enriching our research analyzes to map the subject with the word interested.

3.5 Summary

This chapter summarized various technologies used in the user review analysis. To predict the reliability of the mobile application. NLP process is a sufficient task for this research the above descripted technology supports the different purpose which is providing the most efficient result of the output of this analysis. Next chapter describes our approach to solve the current problems related to the user reviews analysis of the mobile application.

Chapter 4

A Novel Approach for Mobile Application Analysis

4.1 Introduction

In the previous chapter, we briefly discussed the methods used in different aspects, the especially focusing area of user reviews analysis: extract features from textual analysis, methods of sentiment analysis and topic modeling for finding the reliability of the mobile application. This chapter describes the selected approach for analyzing the health care mobile related application.

4.2 Proposed model

The development of the mobile application analyzer has five steps, namely, information of reviews, extraction from Google play store, data cleaning, an NLP model such as sentiment analysis and topic modeling, classification and discovering knowledge. This model can be summarized in the following model given in figure 4.1.

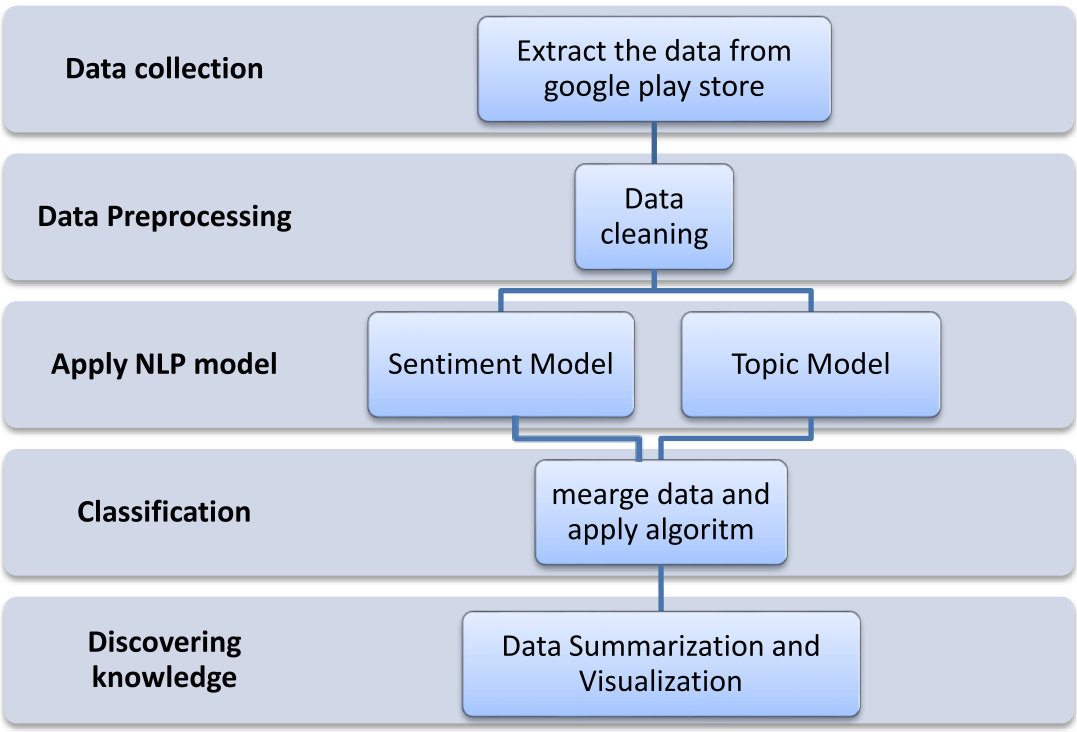


Figure 4.1 Proposed model

4.3 System overview

In this research, the main objective is to analyze the reliability of the mobile health and fitness app based on user feedback, which is the core of NLP approaches to research. Our proposed system is to first gather user feedback and apply NLP approaches, such as sentiment analysis and subject modeling, and apply the relevant process through these two main processes, allowing us to determine the algorithmic statistical value of the post-comparison of the reliability of the mobile application.

After this process, we must perform the user test with the selected mobile application as a result of the output result. This analysis of mobile apps targeting mobile users, so we finally created the mobile app to view this analytics for the end user.

4.4 Data Collection from the Google play store.

The following methods were chosen because it is one of the most practical ways to gather big data from the web site. So we created the small tool will be created using C# to scraping the data from the Google play store. To achieve our web scraping we used HtmlAgilityPack.dll to import our tool and scraping the data efficient manner. This method allows you to browse the site and extract several types of data: text, tables, images, links, etc.

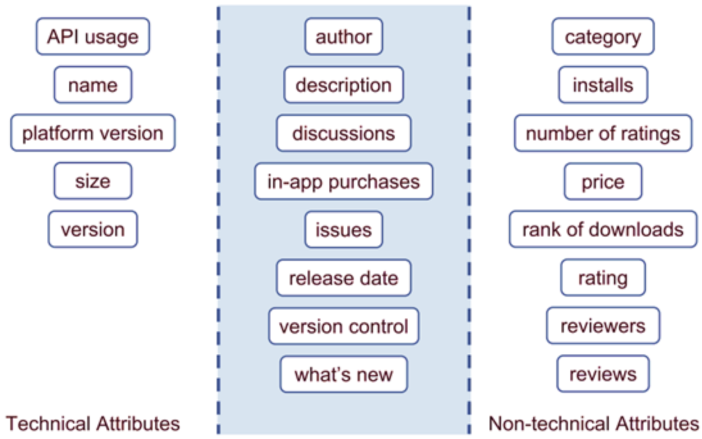


Figure 4.2 Overview of the Google play store API structure

According to the API structure our study capture only for the content of the non-technical attributes and common attributes because technical attributes not related to our study and other point technical data cannot get easily because there is some restriction also. An Overview of the Google play store API structure shown Figure 4.2.

4.5 Preprocessing data

This stage we can use the Weka in order to mine data or in order to analyze your data for machine learning or data mining activities, so there are several advantages of using these tools. The key important thing about choosing Weka explore is that when you are having your data set, size small to medium, this was useful. The above reason can be used for the preprocessing data on this project.

4.6 Identify the suitable methods for NLP

The method is consistent with a variety of many techniques developed to solve the problem of text mining, which is nothing more than the retrieval of relevant information according to the needs of the user. According to information retrieval, four methods are used. Term-based method, phrase-based method, concept-based method and model taxonomy method [8]. Depending on our needs, we plan to use only the first 3 methods. Because of the model taxonomy method, documents are analyzed on a model basis. Models can be structured in taxonomy using an ‘is’ or ‘a’ relation. This useful technique is therefore known as a misinterpretation of models and leads to inefficient performance.

4.7 Discovering Knowledge

According to the previous approach in order to identify the efficient result of the good reviews mobile application in Google play store. Figure 4.3 shown the data mining method of clustering via key word analyzer.

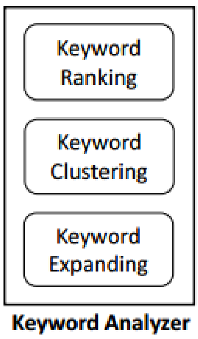


Figure 4.3 Clustering method of keyword analyzer

4.8 Summary

We discussed how we adopt technology to solve the problem identified in this chapter. One of the main challenges of the analysis is finding a reliable mobile app based on the user review. We discuss the analysis and design of our solution in the next chapter.

Chapter 5

Analysis and Design of the Proposed Solution

5.1 Introduction

In the previous chapter, we briefly discussed the technology used in our approach to solve the identified problem. In this chapter describes the system design of the mobile application of health care application review analysis. Furthermore, describes methods and the technology used.

5.2 System design

We have identified 4 major modules for design for our proposed solution.

1. Data collection module
2. NLP model for sentiment analysis
3. NLP model for topic analysis
4. Data visualizing and analyzing

5.2.1 Data collection module

In this research, we must obtain feedback from Google Play Store users in the health and fitness category for data collection. We design a small tool to extract comments from Play Store. We use C # has a special module called HtmlAgilityPack.dll to read the syntax of the website on the website and easy to extract data from the website. In our research, the main task of data collection is to analyze the reliability of the health and fitness apps available on the Google Play Store. Using C #, we create a tool to collect Google Play Store user reviews and collect limited data attributes to analyze the actual result of this research.

5.2.2 NLP model for sentiment analysis

The NLP model we use is a high-level popular programming language for data science. In this research, we primarily analyze user feedback to get sentiment or opinion as an outlet for predicting the reliability of mobile healthcare and finesse applications. From this point of view, Python has a large number of tools and a library specially designed for natural language processing and data science.

Spyder is the Python development environment that provides advanced editing, analysis, debugging, and profiling competences for a complete development tool, as well as exploration, interactive execution, and auditing capabilities. In-depth and attractive visualization of a scientific software package. The Natural Language Toolkit is a python package that allows Python programs to run and exploit data in human language. It consists of a varied collection of libraries for performing operations such as classification, linking, token creation, markup, analysis, and so on. It consists of many datasets and built-in lexical resources. It's a free and easy-to-use open source package.

The text extension adds all the necessary operators for statistical analysis of text and natural language processing (NLP). We can load text from various different data sources, transform them into a huge set of filtering techniques, and finally analyze the data from user reviews. Text Extensions support multiple text formats, including plain text, HTML, or PDF, as well as other data sources. It provides standard filters for token creation, stemming, stop word filtering, n-gram generation to provide everything needed for text preparation and analysis.

The python Vader module supports the analysis of sentiment to automatically detect the polarity of user opinions on a specific argument. To perform this task, a supervised algorithm for assigning tags to user notices indicating whether the opinion is positive, negative or neutral has been adopted. The exact technique used is the algorithm of naïve Bayes classification.

In machine learning, naïve Bayes classification algorithms are a family of probabilistic classifiers based on the Bayes theorem. In the specific case of sentiment analysis, the goal is to classify text or documents in the positive, negative, or neutral class. The implementation of the classifier is characterized by three phases:

1. **The phase of Training:** In this phase the system is trained with the use of a training set constituted by previously classified texts to obtain the probability that a word is positive, negative, or neutral given the class assigned to the text.

2. **The phase of Testing:** This algorithm tested to measure the accuracy of the method
3. **Phase of classification:** the entry of the algorithm is a set of unclassified texts, and it is determined whether they are positive, negative or neutral.

A part of pre-processing precedes the steps described. In this phase, spaces, URLs, tags, and numbers are deleted, all the words are set to lowercase, and the emoticons are replaced by placeholders so they can be used in the classification.

No	Library files name	Purpose of use
01	NLTK	Text analysis, text classification, stemming, tokenization, tagging, parsing
02	nltk.sentiment.vader import SentimentIntensityAnalyzer	To analyze the user review based on sentiment polarity based on the algorithmic way.
03	import gensim.models import Word2Vec	To check the accuracy of most similar words thorough the word vectorization.
04	import TfidfVectorizer, import TruncatedSVD	Sklearn for feature extraction of user reviews For decomposition of user reviews and term analysis

Table 5.1 The Major python libraries used for the solution

5.2.3 NLP model for topic analysis TFIDE and LSA

In this stage, we design the TFIDE model and LSA model to generate the topic for the term of concepts. Both can use our research to produce a more accurate result.

5.2.3.1 TFIDE

The term analysis is very important to analyze the reliability of mobile application used in end-user opinion. The early research studies using Bag of word model some situation, it worked, but from this research, we use the TF - IDE model to find the term analysis of reviews because the major drawbacks of Bag of Words model provide all words have the same importance and no semantic information preserved.

Based on the user reviews sentence of some semantic information is preserved as uncommon words are given more importance than common words. This model has a frequent term to separate value after that occurrence calculated with each document with IDF value.

TF = Term Frequency (Each word and document value)

IDF = Inverse Document Frequency (Each word has separated IDF value)

TF-TDF = TF * IDF

$$\mathbf{log}\,\frac{(\textbf{Number of occurrences of a word in a document})}{(\boldsymbol{Number\ of\ words\ in\ that\ document})}$$

Figure 5.1 Solution TFIDE Model

This model contains the more efficient value to the huge corpus of data higher value and generating a more specific word from the reviews.

5.2.3.2 Latent Semantic Analysis

Latent semantic also another method to find out the topic from the huge data collection. This method used for the resolution of verification. We have to compare both result and choose the more reliable topic based on mobile application reviews. This semantic analysis is a technique for analyzing the relationships between a set of documents and the terms that contain by producing a set of concepts related to the documents and terms. Description of the algorithm shows Figure 5.2.3.1

$$\boldsymbol{A}_{\ [\boldsymbol{m}\times\boldsymbol{n}]} = \boldsymbol{U}_{\ [\boldsymbol{m}\times\boldsymbol{r}]} * \boldsymbol{S}_{\ [\boldsymbol{r}\times\boldsymbol{r}]} * (\boldsymbol{V}_{\ [\boldsymbol{n}\times\boldsymbol{r}]})^{\boldsymbol{T}}$$

A : Input Data Matrix

m x n matrix (m = number of documents, n = number of words/features)

U : Left Singular matrix

m x r matrix (m = number of documents, r = number of concepts)

S : Rank Matrix

r x r matrix (r = rank of A)

V : Right Singular Matrix

n x r matrix (n = number of words/features, r = number of concepts)

Figure 5.2 Solution LSA Model

5.2.4 Data testing and visualizing

In this phase, we used some visualizing libraries from the python and weka for analysis our result set. This module helps to predict the result as the visual output of the analysis studies. Because some python library has good visualization effect to better understand of the user (Appendix B).

5.3 Summary

This chapter explained the design of the mobile application analysis and especially scientific approach for the NLP techniques which is important to this research. The next chapter gives more details of the implementation of the design based on this analysis and design.

Chapter 6

Implementation of the Solution

6.1 Introduction

This chapter provides implementation details of each of the four modules mentioned in the previous chapter. Moreover, this presents programming language and algorithms used in each module with the sample outputs.

6.2 Implementation of data scraping tool

In this research one of the biggest challenges that we faced on the implementation part of the data collection often, researchers build tools that aren't supported in the current Google App store because of access restriction. At this point, we are developing a separate tool and scraping data from the app store. It's a huge process within our research period Google App store has changed its web structure 2 times. While it is not a new problem, it still remains a challenge to collect large data which are under related to health and fitness category. The GUI of the scraping tool shows Figure 6.1.

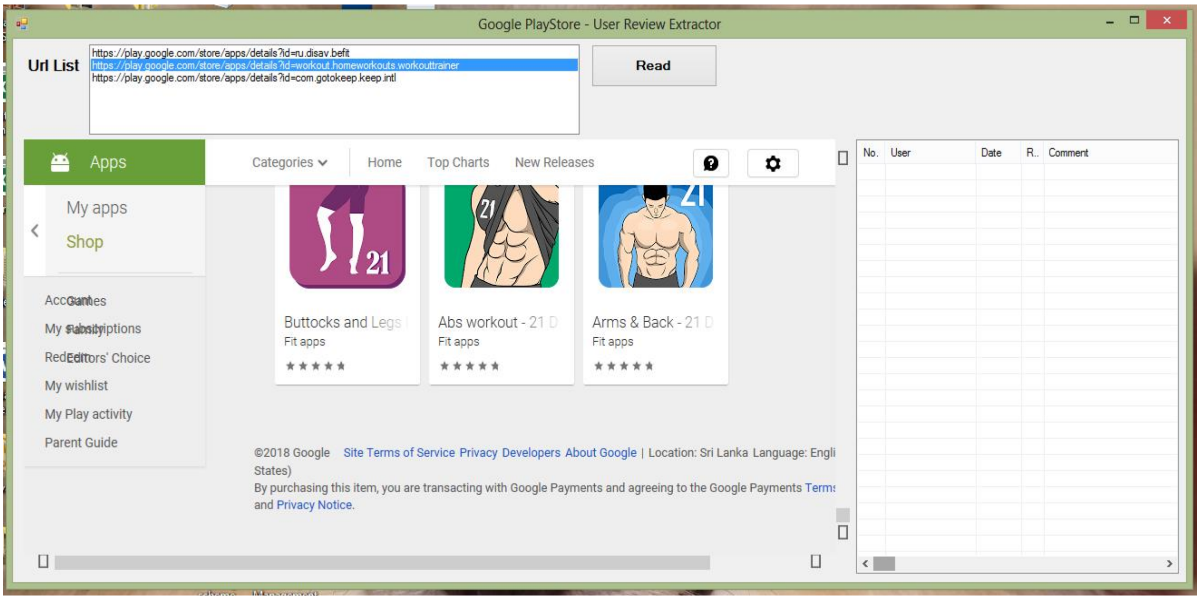


Figure 6.1 User review scarping tool

During the development process of the scraping tool google has changed their comments structural design in this point redesigned the tool. That will shows in Figure 6.2.

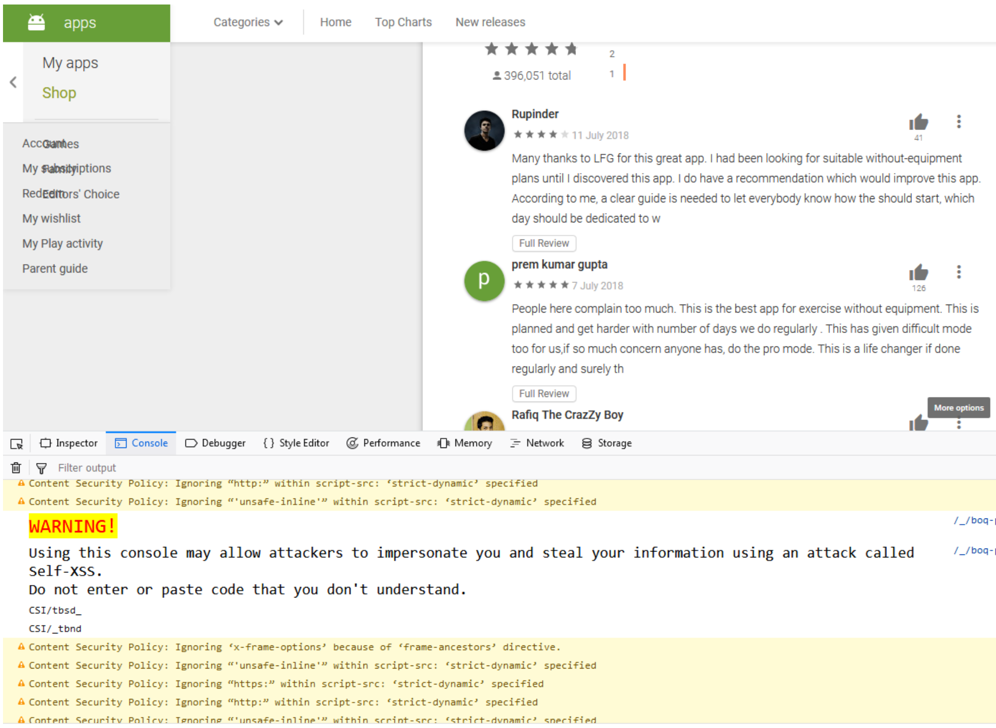


Figure 6.2 Google play Apps store structure change

	B	C	D
1	User name	Rate	User reviews
2	Rupinder	4	Many thanks to LFG for this great app. I had been looking for suitable without-equipment plans until I discovered this app. I do have a recommendation which w
3	prem kumar gupta	5	People here complain too much. This is the best app for exercise without equipment. This is planned and get harder with number of days we do regularly . This
4	Edward Courtney	5	OMG this is the best workout app out there in my opinion
5	RadRhett 8	5	The app is really great
6	Devaansh Patel	5	The best app for home workouts ever! You don't even need weights for MOST of the exercises. I highly recommend it to people who due to any reasons cannot g
7	Adieb Affhan	5	Its what the name of the apps says.. A home workout apps that you can do anywhere and at anytime. But it would be so great if you guys could make an apps fo
8	Pierrot Jr Raymond	5	The app has various workout plans. No weights needed and if you stick to it you will see results. Plus is I can play my own music playlist in the background while
9	Chris van den Berg	5	I honestly love this app. I play a lot of sports and I'm still in school so I don't get a lot of time to go to the gym
10	Tushar Majumdar	4	This is a great app for home workout especially if you don't have access to gym. You will be in good shape with these workout for beginners
11	A Wightman	5	Love this program. I have seen results and I feel it pushes me. The 28 day programs are great and builds up your strength and endurance
12	Xuxan Xstha	5	Thanks to the creater of this app..it helps a lot.Within few months I have created my abs
13	HARI KAMATH	5	Starts out to be tough....eventually seeing results....hustle....don't give up....guys hitting the gym...get in shape...first....lose fat...this is a good option for losing
14	Awale mohamed muse	5	You guys are truly amazing. I hate going to the gym and now I have my own private gym right here in my room. Thanks alot.
15	Soumya Mishra	5	It helped me gain six pack abs As you all know that everyone wants a perfect body so this app is best for Gain a spectacular body ðŸ˜‰ðŸ˜‰ðŸ'ªðŸ'ªðŸ'ŒðŸ'Œ
16	Chuck Zayn	5	If you have a hectic days and dont have much time to go to the gym or park
17	A Google user	5	It has a very fabulous levelwise exercises and core workout toooo This app is amazing I can't express my gratitude towards this app Thnx for this app
18	A Google user	5	It's cool not so bad if keep doing your training everyday with full body work out without Wight just body Wight
19	A Google user	4	I think they need to rdeduce the reps because most people are beginners and they are just too much and straining
20	Thomas Bolde	5	I like this app. The workouts are great
21	Daniel Barlocker	5	I love how its has different stage workouts for easy progression. I also love that I can do all the exercises without equipment.
22	Ernesto Abais	5	Help!!! What do I do after the 30 day Challenge? I usually just follow that because I don't know what workout to do next
23	Lakshmi J	5	Really easy and helpful now I am having 8 packs but I am 14 years old only
24	Osamuyi Okpame	5	This is super amazing. Technology is helping me achieve my goal faster than I thought.
25	J Rima	5	I love this app and it is amazing I have only been doing it for 2 days and I already am way healthier and fitter
26	My kids have taken over	5	This is a brilliant app to exercise people that complain only do so because they find it too difficult
27	Daniel Freire	5	Excelente app para ejercitarse en casa. Las indicaciones son claras y las rutinas son ajustables. Espero conseguir las metas propuestas
28	Anugrah Bahtiar	5	I love this app. It helps me sweating everyday. I feel fit and get better shape day by dayðŸ˜
29	A Google user	5	Awesome but i am like 14 and do these exercise but I often heard doing this effect our height so this is my growing age tell me will this effect my height should
30	A Google user	5	This app is the most important app in my life I can't take a break I had burned 4365in one month you should download the app because you will have a good hea
31	A Google user	5	Great app
32	Abhishek verma	5	Just follow what it says and see your results happen..
33	Nate Strickland	5	I live in a place where there is no gym so these workout help me accomplish my workout goals using my own body weight. Get fit!
34	Ajay kumar singh	5	à¤§à¤¨à¥à¤¯à¤µà¤¾à¤¦ à¤‡à¤¸ à¤¶à¤¾à¤¨à¤¦à¤¾à¤° à¤à¤ª à¤•à¥‡à¤²à¤¿à¤ à¤¯à¤¹ à¤¹à¤®à¤¾à¤°à¥‡ à¤œà¥ˆà¤¸à¥‡ à¤²à¥‹à¤—à¥‹à¤‚ à¤•à¥‡ à¤²à¤¿à¤ à¤²
35	Himanshu Sharma	4	Good for peoples who is suffering from the problem of lack of money due to which they cannot afford gym.
36	SMART GAMER	5	Sir i am traying to make my six paks i hope i will make my abs in month its my dream ðŸ˜.

Figure 6.3 Data pre-processing

6.3 Implementation of the NLP model for sentiment analysis

In this stage implement through the python call the NLTK and Vader Sentiment library class for analysis the sentiment polarity of review, and it will produce the mathematical score. NLTK comes with a built-in sentiment analysis module, nltk.sentiment.vader, that can analyze a reviews element and classify sentences as positive, negative and neutral based on polarity score and snippet of code explaining how this could be done is presented below Figure 6.4:

1. **positive sentiment:** `compound` score >= 0.05
2. **neutral sentiment:** (`compound` score > -0.05) and (`compound` score < 0.05)
3. **negative sentiment:** `compound` score <= -0.05

Figure 6.4 Vader polarity value score for sentiment analysis

Implementation of NLP model for sentiment analysis we added some efficient code for to delete duplicate data using downloaded CSV file, removed unwanted symbols, tokenize all the comment using nltk, removing stop word through word tokenizer of nltk it is showing **Figure 6.5.**

```python
#remove unwanted symbols
for i in range(len(example)):
   example[i] = re.sub(r"\W"," ",example[i])
   example[i] = re.sub(r"\d"," ",example[i])
   example[i] = re.sub(r"\s+[a-z]\s+"," ",example[i],flags=re.I)
   example[i] = re.sub(r"\s+"," ",example[i])
   example[i] = re.sub(r"^\s","",example[i])
   example[i] = re.sub(r"\s$","",example[i])
   example[i] = example[i].lower()

#remove stop words
for i in range(len(example)):
    words = nltk.word_tokenize(example[i])
    words = [word for word in words if word not in stopwords.words('english')]
    example[i] = ' '.join(words)
```

Figure 6.5 Tokenizing code for stop word and unwonted symbols

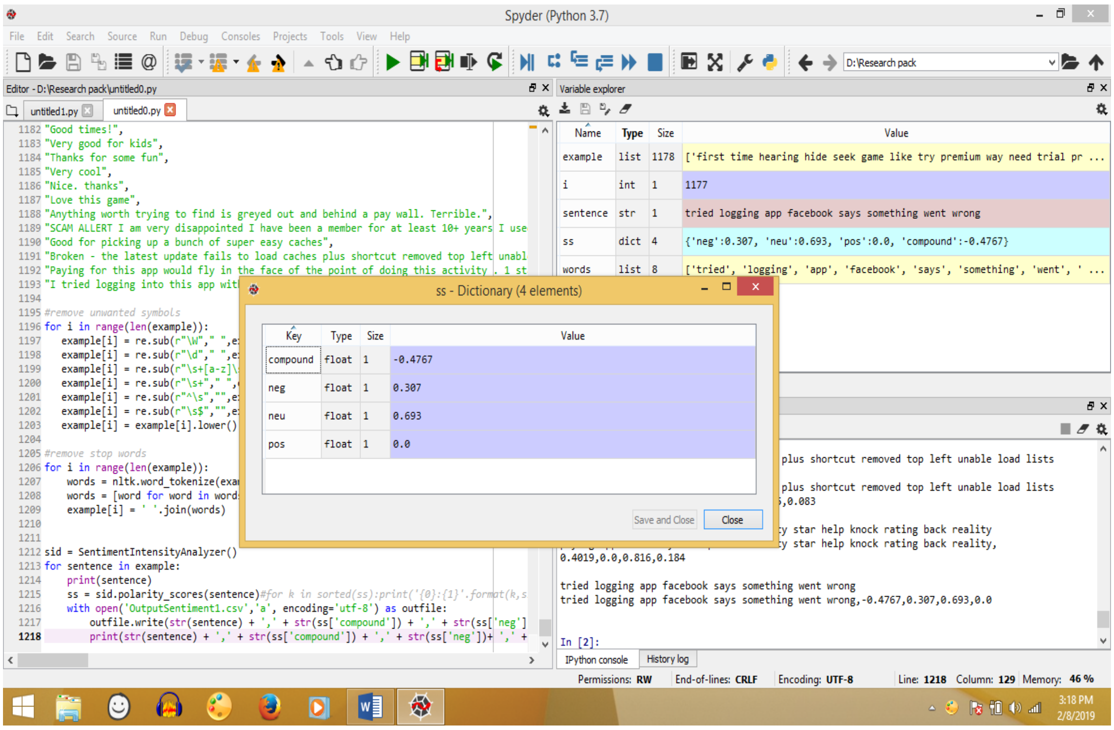


Figure 6.6 Polarity analysis score

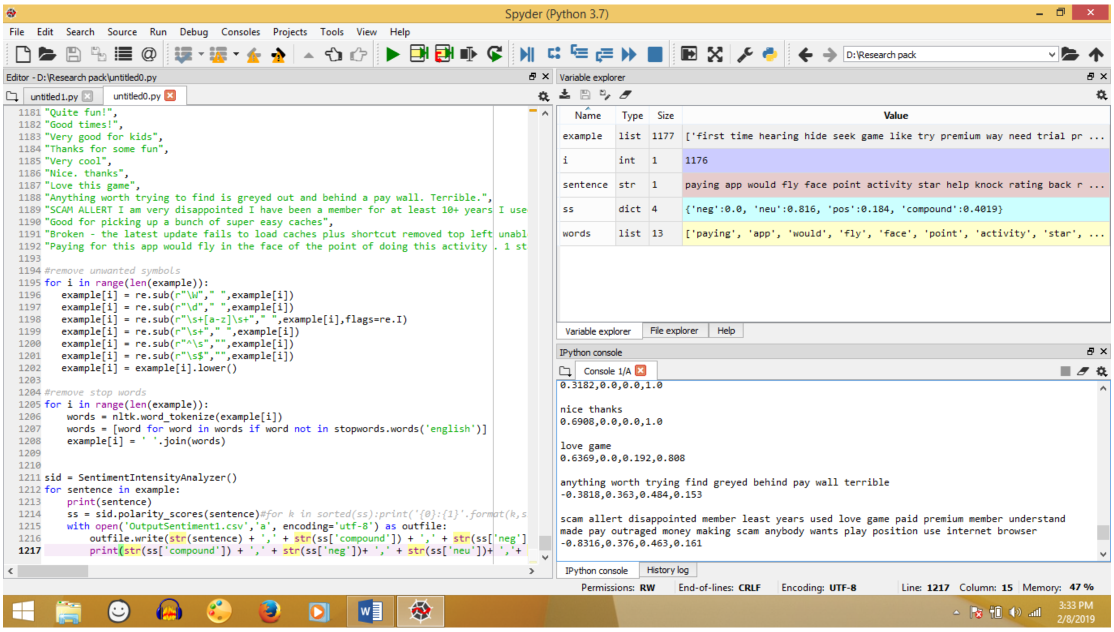


Figure 6.7 Successful execution of sentiment analysis

6.4 Implementation of NLP model for topic analysis

TF-TDF model implemented for the topic analysis of the mobile application analyzer. The figure 6.8 shows the code of topic analysis. (Appendix C)

untitled1.py | untitled0.py | final try concept.py | NLP Part 18 - Latent Semantic Analysis Part 2.py

```python


# Creating the SVD
lsa = TruncatedSVD(n_components = 4, n_iter = 100)
lsa.fit(X)


# First Column of V
row1 = lsa.components_[3]

# Word Concept Dictionary Creation
concept_words = {}

# Visualizing the concepts
terms = vectorizer.get_feature_names()
for i,comp in enumerate(lsa.components_):
    componentTerms = zip(terms,comp)
    sortedTerms = sorted(componentTerms,key=lambda x:x[1],reverse=True)
    sortedTerms = sortedTerms[:10]
    print("\nConcept",i,":")
    for term in sortedTerms:
        print(term)

# Sentence Concepts
for key in concept_words.keys():
    sentence_scores = []
    for sentence in dataset:
        words = nltk.word_tokenize(sentence)
        score = 0
        for word in words:
            for word_with_score in concept_words[key]:
                if word == word_with_score[0]:
                    score += word_with_score[1]
        sentence_scores.append(score)
    print("\n"+key+":")
    for sentence_score in sentence_scores:
        print(sentence_score)
```

Figure 6.8 Implementation of topic analysis module

6.5 Implementation of Mobile Application

This phase, produce the mobile analysis as an output for the mobile user. This research target on the end-user of a mobile application to find reliable mobile application health and fitness purpose. This application includes for this research for the final outcome should reach to the mobile user for visualizing the data for their easy understanding.

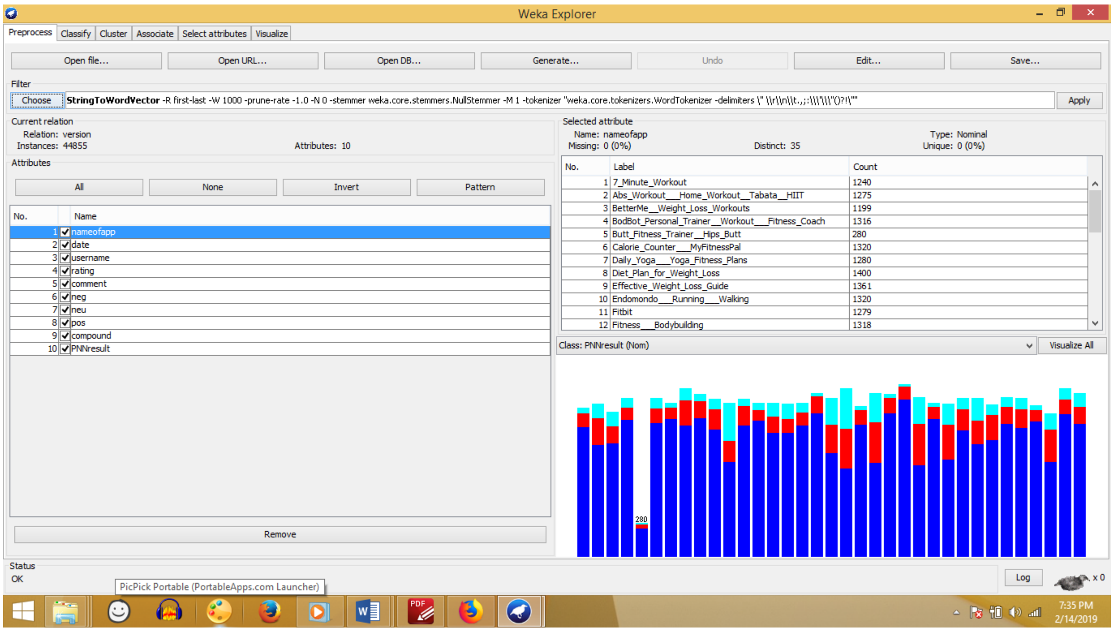


Figure 6.9 Weka data visualization

6.6 Summary

In this chapter discussed the implementation of the system, it contains most of the code and tools which are we used to implement the proposed system. In next chapter describes an evaluation of the implemented system.

Chapter 7

Evaluation

7.1 Introduction

The previous chapter discussed the details on the implementation of the modules of the proposed solution for the mobile application analysis. This chapter justifies and evaluates the results of the proposed system.

7.2 Data collection module

Data scraping tool work properly and downloaded user reviews of top 30 mobile applications from the google play store with non-technical attributes. Figure 7.1 shows the total number of user review and name of the application.

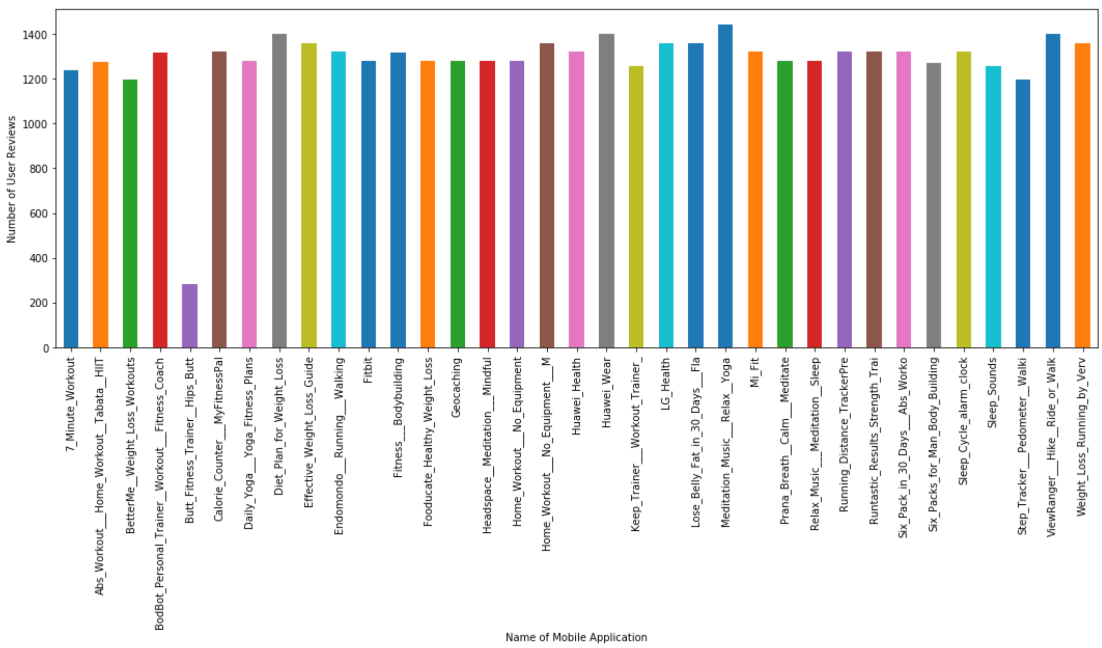


Figure 7.1 Extracted user reviews of top 35 application from the google play store

The reviews of top 35 mobile application are successfully downloaded. Total number of reviews 44,857 for the analysis of the average 35 application have more than 1,150 user reviews.

7.3 NLP model for sentiment analysis

In this module successfully executed and produce the sentiment polarity score. Our experiments confirm the success rate of accuracy also. The figure 7.2 clearly displays the output of the CSV file. This result has 4 values negative, neutral, positive and compound. The compound value showing the overall score of the sentiment analysis.

	A	B	C	D	E
1	comment	compound	negative	Neutral	Positive
2	first time hearing hide seek game like try premium way need trial program noob looks fun get outside	0.5994	0.128	0.617	0.256
3	upgraded premium unable access premium geocaches good time basic geocaches	0.4404	0	0.756	0.244
4	need pay caches difficulty incredibly merchant like	-0.0056	0.359	0.378	0.263
5	works time hate english translation	-0.5719	0.481	0.519	0
6	love app gives opportunity go explore places never also works google maps	0.7906	0	0.588	0.412
7	goede app	0	0	1	0
8	google play unavailable try buy premium	0.34	0	0.676	0.324
9	would nice track travel bugs directly app	0.4215	0	0.682	0.318
10	love finding geocache	0.6369	0	0.323	0.677
11	overly expensive access cached	0	0	1	0
12	free version unusable	0.5106	0	0.377	0.623
13	great fun app	0.8126	0	0.119	0.881
14	addicted first day	0	0	1	0
15	love way choose map style super easy use	0.9001	0	0.312	0.688
16	enjoying app several years best geocaching app available	0.8225	0	0.441	0.559
17	love kids really enjoy	0.8268	0	0.206	0.794
18	wish wasnt membership	0.4019	0	0.426	0.574
19	love wish every geocach hint	0.7845	0	0.303	0.697
20	fun game hide seek treasure love	0.8402	0.127	0.149	0.724
21	fun ages	0.5106	0	0.233	0.767
22	lots fun whole family	0.5106	0	0.476	0.524

Figure 7.2 Sentiment analysis polarity score of user reviews

So it will help with the prediction accuracy with minimal optimization of the dataset. The reliable techniques which are mentioned the above research methodology will be produce the good prediction result and based on the result we can recommend improve the algorithm to the next level. Sometime our result not producing efficient prediction we will move to change the K means values and reproduce the result again and again when our study found the proper mechanism of collecting user reviews on Google play store. The Figure 7.1 shows final summary of Sentiment Analysis from 35 Applications with the rank order. It is clearly indicated the Meditation Music Relax Yoga application user reviews produces a higher rate of positive sentiments.

No	Name of Application	Total comments	Negative	Neutral	Positive	Rank
1	Meditation Music Relax Yoga	1442	1.60	7.63	90.78	1
2	Daily Yoga Yoga Fitness Plans	1280	3.13	7.34	89.53	2
3	Sleep Sounds	1258	3.02	7.63	89.35	3
4	Prana Breath Calm Meditate	1280	2.89	7.81	89.30	4
5	Fooducate Healthy Weight Loss	1280	4.92	6.64	88.44	5
6	Lose Belly Fat in 30 Days Fla	1358	2.72	9.57	87.70	6
7	Home Workout No Equipment M	1360	2.06	10.44	87.50	7
8	Keep Trainer Workout Trainer	1258	2.70	9.86	87.44	8
9	7 Minute Workout	1240	3.71	9.35	86.94	9
10	BodBot Personal Trainer Workout Fitness Coach	1316	6.23	7.45	86.32	10
11	Home Workout No Equipment	1279	6.33	8.21	85.46	11
12	Effective Weight Loss Guide	1361	4.70	10.43	84.86	12
13	Butt Fitness Trainer Hips Butt	280	4.29	11.07	84.64	13
14	ViewRanger Hike Ride or Walk	1400	6.79	8.57	84.64	13
15	Calorie Counter MyFitnessPal	1320	6.74	9.02	84.24	15
16	Six Pack in 30 Days Abs Worko	1324	5.97	10.57	83.46	16
17	Fitness Bodybuilding	1318	4.70	12.52	82.78	17
18	Sleep Cycle alarm clock	1320	6.97	11.89	81.14	18
19	Weight Loss Running by Verv	1361	8.60	10.36	81.04	19
20	Geocaching	1280	8.75	10.55	80.70	20
21	Headspace Meditation Mindful	1280	10.23	9.14	80.63	21
22	Endomondo Running Walking	1320	6.67	12.95	80.38	22
23	Running Distance TrackerPre	1320	7.05	13.26	79.70	23
24	BetterMe Weight Loss Workouts	1199	9.92	11.68	78.40	24
25	Diet Plan for Weight Loss	1400	7.21	14.86	77.93	25
26	Six Packs for Man Body Building	1272	7.00	16.67	76.34	26
27	Abs Workout Home Workout Tabata HIIT	1275	9.57	17.49	72.94	27
28	Runtastic Results Strength Trai	1320	14.09	15.00	70.91	28
29	Step Tracker Pedometer Walki	1195	11.21	22.68	66.11	29
30	Huawei Health	1320	16.82	17.95	65.23	30
31	Relax Music Meditation Sleep	1280	13.75	22.89	63.36	31
32	Fitbit	1279	24.71	13.45	61.85	32
33	Mi Fit	1320	16.59	25.83	57.58	33
34	LG Health	1360	17.79	24.71	57.50	34
35	Huawei Wear	1400	24.21	23.50	52.29	35

Table 7.1 Final summary of Sentiment Analysis from 35 Applications

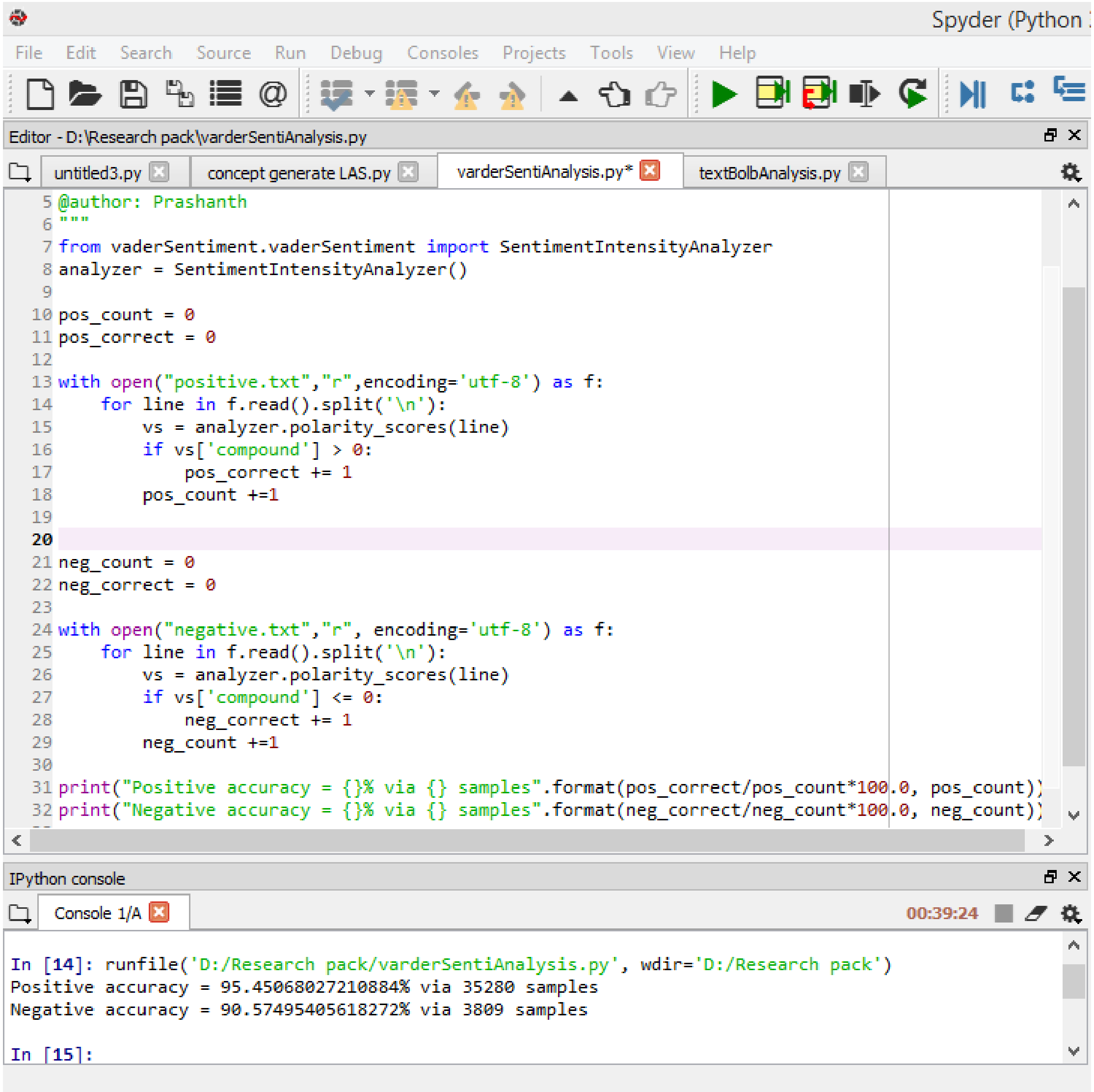


Figure 7.3 Output of accuracy using Varder sentiment analysis

The accuracy level generated by using varder sentiment polarity class. It clearly indicates the positive and negative rate of accuracy with a sample rate. But the traditional system of textblob class shows 100% of accuracy that not proper analysis result. The Output of the varder accuracy level indicating Figure 7.3 and textblob method indicating Figure 7.4. The Comparison of output textblob method and latest vadermethod is shown in the Figure 7.5.

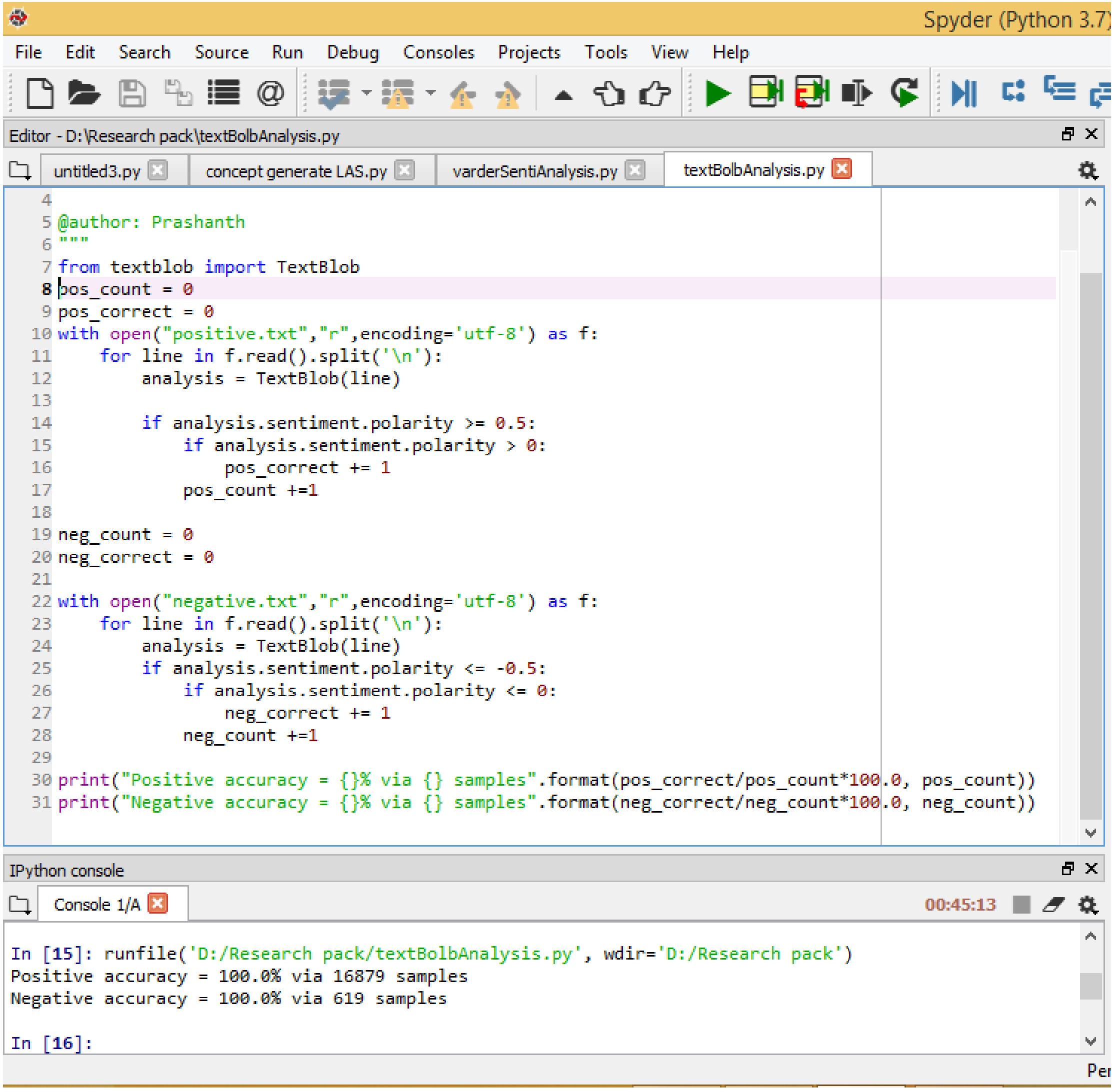


Figure 7.4 Output of accuracy using traditional textblob sentiment analysis

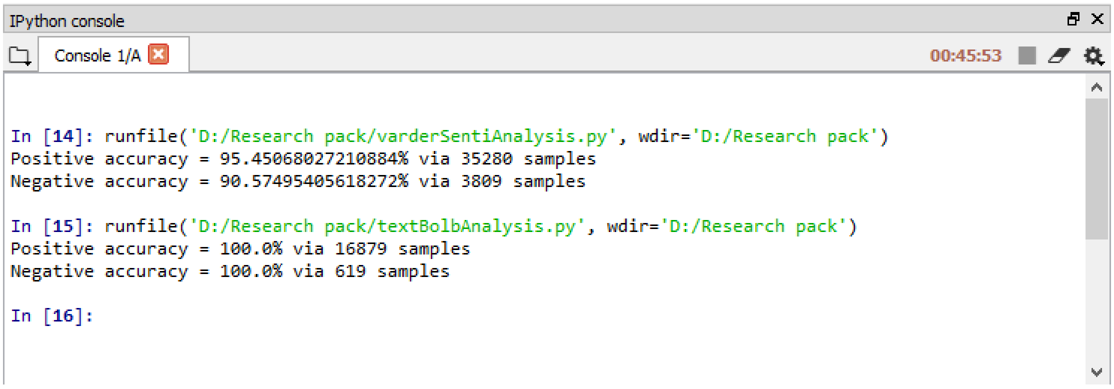


Figure 7.5 Comparison of output traditional method and latest vadermethod

7.4 NLP model for topic analysis

The evaluation of this module has not produced the sufficient output but has some kind of analysis result that will help to improve our modeling. So our investigation of this evaluation topic modeling working properly, but our dataset not contain the average amount of content for topic analysis. Figure 7.6 shows the concept generated output and it is contains the rank of value.

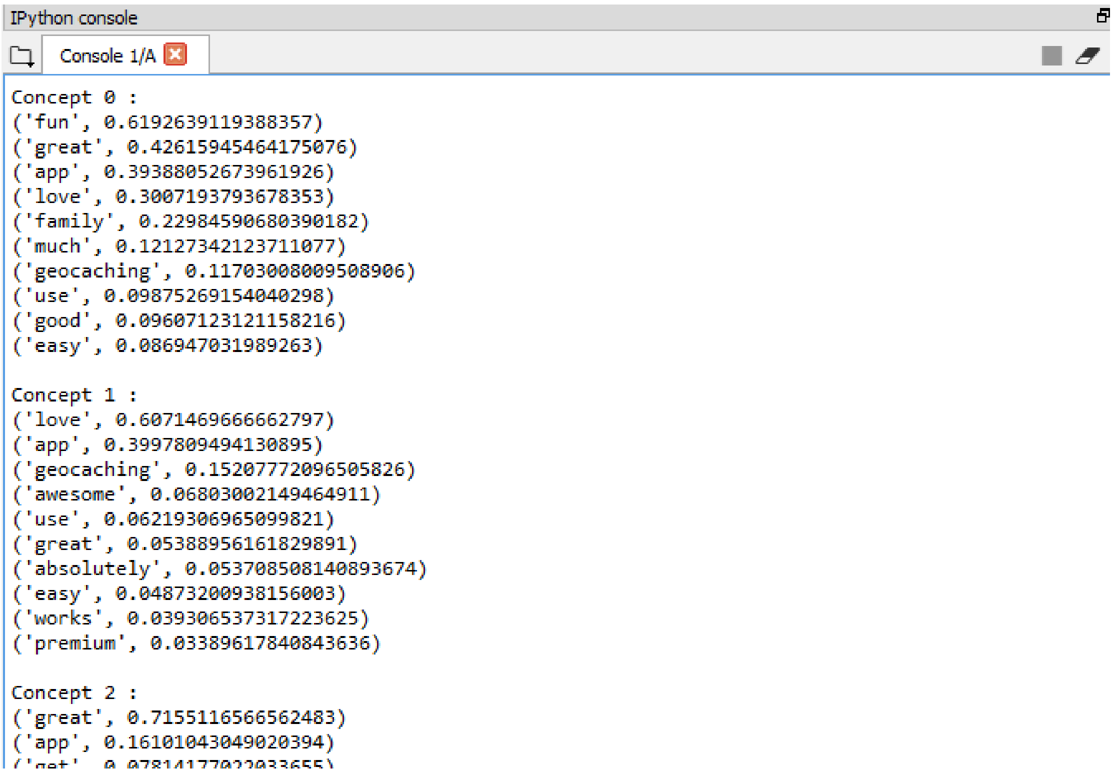


Figure 7.6 Output result of topic analysis

The two main results we obtained during this research are the sentiment of the user reviews and topics based on the features of user reviews. Table 7.2 shows the analysis of the topic, which are sorted and finalize based on the topic rank value through the features of the related words. The mention table has analyzed topic and contains related words also. This research first collected all the words and using topic analyzer and selected related words are selected through the concept based ranking values. The fitness category contains 12055 user reviews, health care contains 8047, money category contains 3670 user reviews, quality category contains 5946, performance based category contains 9389 and Error category contains 5082. But overall user reviews 44855. (Appendix F) The percentage of total user reviews based on topic shows in the figure 7.7.

Topic Analysis	Related words
Error	Fix, try, problem, issue, useless, issues, uninstall, wrong, disappointed, waste, problems, worst, uninstalled, unable, frustrating, error, difficult, geocaching, inaccurate, crashing, crashes, pain, bug, caches, slow, sucks, missing, lose, bad, hate
Health	Food, calories, nutrition, breathing, diet, fat, heart, healthy, fitbit, foods, mind, eat, calorie, kg, meal, water, sleep, fit
Performance	Update, Data, Version, New, Tracking, Screen, Option, results, download, battery, connect, gps, map, feature, device, android, speed, videos, program, interface, session, upgrade, Bluetooth, Connection, recording, memory, fixed,
Quality	Perfect, Excellent, Amazing, recommend, effective, accurate, fantastic, wonderful, brilliant, recommended, perfectly, information, quality, service, rating, product, design, customize
Fitness	Exercise, workouts, exercises, steps, working, weight, exercise, fitness, body, running, tried, calories, training, meditation, run, yoga, diet, life, walking, count, play, breathing, shape, walk, stress, muscles, practice, breath, hiking
Money	Free, Premium, Money, Pretty, worth, paid, buy, trial, pounds, expensive, refund, monthly, purchase

Table 7.2 Topics and related word

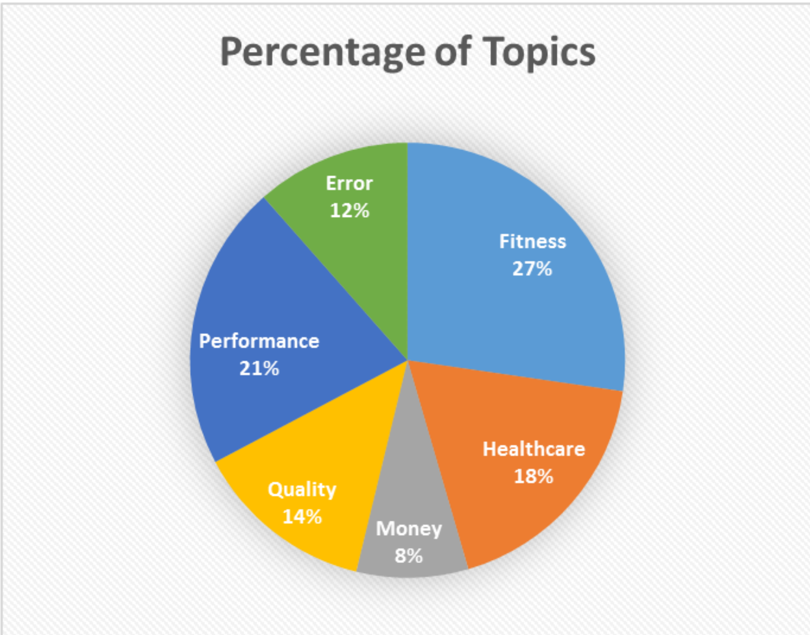


Figure 7.7 Percentage of total user reviews based on topics

7.5 Data visualizing and analyzing

Evaluation of data analysis working properly and it contains, the more details, visual representation of the result. So it will help with the prediction accuracy with minimal optimization of the dataset. The reliable techniques which are mentioned the above research methodology will be produce the good prediction result and based on the result we can recommend improve the algorithm to the next level.

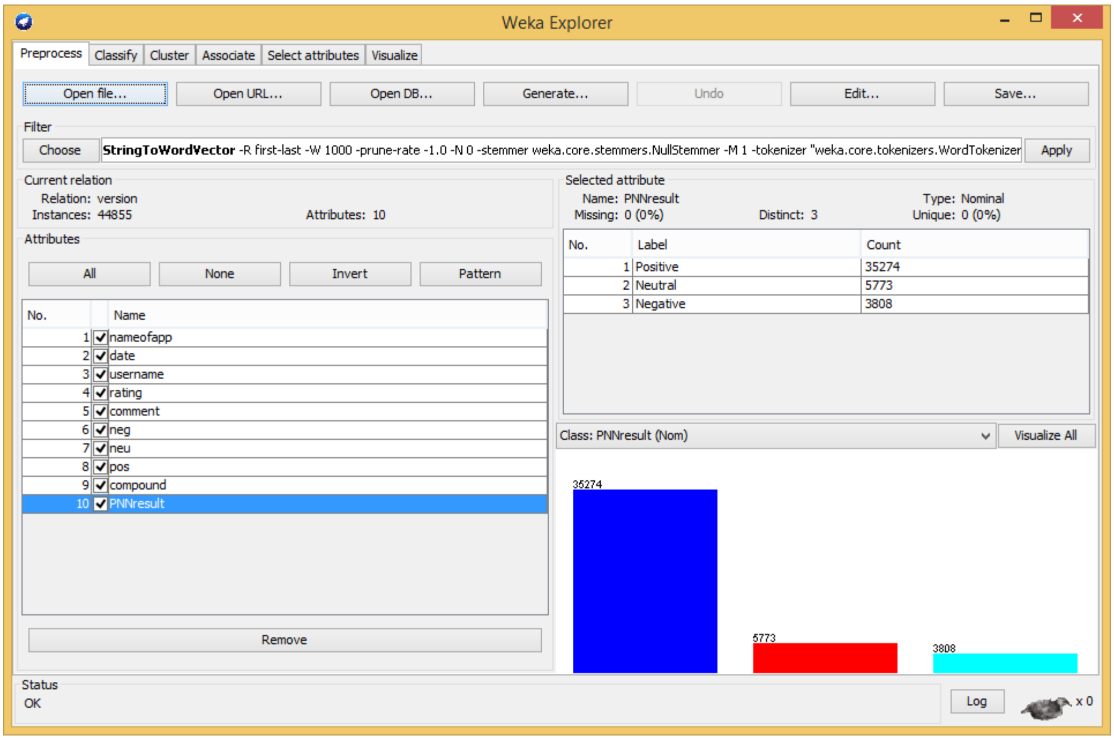


Figure 7.8 Visually represent the summary of sentiment analysis

Evaluation of the combine results of sentiment and topic, less number of the user reviews are contained the features and those have most probably indicated the positive level of human expression. The figure 7.8 shows the number of user reviews contain the number of positive or negative user reviews based on the topics.

The correlation between sentiment and topics is interesting because overall positive user reviews are showing less number of negative user reviews and more user reviews are contain the multiple topics.

Application Name	Total comment	Fitness		Money		Quality		Performance		Error		Healthcare	
		Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos
7_Minute_Workout	1240	17	394	5	53	1	222	6	127	13	38	3	197
Abs_Workout_Home_Workout_Tabata_HIIT	1275	18	172	12	46	2	101	46	142	54	89	5	43
BetterMe_Weight_Loss_Workouts	1199	25	262	34	133	2	112	20	92	48	71	12	109
BodBot_Personal_Trainer_Workout_Fitness_Coach	1316	30	531	32	210	7	219	19	292	41	132	14	249
Butt_Fitness_Trainer_Hips_Butt	280	4	47	0	5	0	38	1	14	4	3	2	17
Calorie_Counter_MyFitnessPal	1320	32	486	16	127	3	143	28	306	47	106	22	641
Daily_Yoga_Yoga_Fitness_Plans	1280	18	406	15	146	3	174	24	259	17	60	2	84
Diet_Plan_for_Weight_Loss	1400	40	192	3	19	3	152	10	47	48	42	35	185
Effective_Weight_Loss_Guide	1361	34	216	6	30	4	151	10	58	31	41	10	111
Endomondo_Running_Walking	1320	33	243	19	90	5	118	51	250	37	62	5	89
Fitbit	1279	94	280	52	94	10	92	233	424	209	266	146	409
Fitness_Bodybuilding	1318	22	271	11	63	1	129	15	145	27	30	5	108
Fooducate_Healthy_Weight_Loss	1280	22	295	15	122	7	169	16	234	29	78	28	628
Geocaching	1280	13	87	35	135	2	99	37	186	63	228	1	7
Headspace_Meditation_Mindful	1280	51	393	18	201	7	271	32	237	56	134	14	155
Home_Workout_No_Equipment	1279	37	596	5	87	5	254	14	217	29	78	23	333
Home_Workout_No_Equipment_M	1360	4	270	2	33	2	168	3	73	7	23	1	135
Huawei_Health	1320	95	323	9	42	14	113	145	438	103	168	64	296
Huawei_Wear	1400	69	165	25	37	9	66	213	415	158	144	78	181
Keep_Trainer_Workout_Trainer	1258	9	379	0	34	0	162	7	65	17	28	7	230
LG_Health	1360	118	290	6	30	22	117	81	197	111	101	31	153
Lose_Belly_Fat_in_30_Days_Fla	1358	15	303	5	38	1	195	5	90	12	64	6	162
Meditation_Music_Relax_Yoga	1442	9	297	0	22	0	222	3	104	4	33	2	219
Mi_Fit	1320	48	165	18	19	6	82	108	300	118	113	33	154
Prana_Breath_Calm_Meditate	1280	17	479	6	87	1	283	10	200	8	49	8	295
Relax_Music_Meditation_Sleep	1280	26	62	11	44	1	84	65	83	81	53	16	106
Running_Distance_TrackerPre	1320	51	350	9	84	5	201	39	237	49	65	1	40
Runtastic_Results_Strength_Trai	1320	27	335	77	214	2	141	35	200	74	55	9	131
Six_Pack_in_30_Days_Abs_Worko	1324	42	400	4	65	3	202	19	168	41	119	23	222
Six_Packs_for_Man_Body_Building	1272	24	129	4	35	0	69	10	48	35	23	7	47
Sleep_Cycle_alarm_clock	1320	12	64	13	134	11	247	23	190	19	48	35	386
Sleep_Sounds	1258	11	92	0	39	3	229	5	140	9	37	22	500
Step_Tracker_Pedometer_Walki	1195	49	205	5	38	8	76	32	62	54	53	9	76
ViewRanger_Hike_Ride_or_Walk	1400	18	369	32	198	4	382	69	661	41	120	0	26
Weight_Loss_Running_by_Verv	1361	47	401	41	171	3	141	35	214	55	104	5	85
Grand Total	**44855**	**1181**	**9949**	**545**	**2925**	**157**	**5624**	**1469**	**6915**	**1749**	**2858**	**684**	**6809**

Figure 7.9 No of user reviews are contain no of positive & negative user reviews

Figure 7.9 shows the analysis of user reviews based on a particular topic and contains a rate of sentiment values among selected 35 applications. In this case, sentiment of positive and negative of user reviews corresponds to the topics and given the reliability mobile application, based on user reviews. The topic of money related positive user reviews has less no of errors, but performance level is very high (e.g., “BodBot-Personal-Trainer-Workout-Fitness-Coach application”). Based on this result, we have obtained accurate comparison of the mobile application through the sentiment and concept.

The other interning part of this module is the creation of the mobile application to access the result in the form of a small mobile application with statistical and visual representation of the 35 most important applications. (Appendix D) Figure 7.10 shows the interface and displays the value of our result.

Figure 7.10 Mobile Application Interface

Chapter 8

Discussion

8.1 Introduction

All previous chapters discussed the identified problem and the proposed solution. This chapter presents some limitations and future improvements may be proposed for further work.

8.2 Limitations

Although research has some unavoidable limitations. First, because of the time limit, this research analyzes user reviews for the past 3 years with non-technical attributes available on the Google Apps Store. Second, this research analyzes only the top 35 mobile apps that belong to the Health and Fitness category of the Google Store.

8.3 Further Developments

The work performed in this research responds to the analysis of the mobile application on the basis of downloaded user reviews and analyzes performed using distinct technologies and effective algorithms with various tools. It will produce a reasonable result with the limitation, but the evidence from this study suggests the current trend, mobile application platform growth in speedy, we must design the online sentiment analyzer and the resulting analysis will help the end user of the mobile application.

8.4 Summary

This chapter concluded about the analysis, limitation and extended future work.

References:

[1] William Martin, Federica Sarro, Yue Jia, Yuanyuan Zhang, and Mark Harman, "A Survey of App Store Analysis for Software Engineering" Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955.

[2] Johnnatan Messias, Joao P. Diniz, Elias Soares, Miller Ferreira,Matheus Araujo, Lucas Bastos, Manoel Miranda, Fabricio Benevenuto, Universidade Federal de Minas Gerais (UFMG), Brazil, "Towards Sentiment Analysis for Mobile Devices," in pp. 1039.

[3] Meiyappan Nagappan, Emad Shihab, "Future Trends in Software Engineering Research for Mobile Apps," J. Name Stand. Abbrev., in press.

[4] Min-Yuh Day, Yue-Da Lin, "Deep Learning for Sentiment Analysis on Google Play Consumer Review,"

[5] Tie Qiu, Senior Member, IEEE, Baochao Chen, Arun Kumar Sangaiah, Jianhua Ma, and Runhe Huang, Senior Member, IEEE "A Survey of Mobile Social Networks: Applications, Social Characteristics, and Challenges,",pp4

[6] Berg Insight. The mobile application market. Online: http://www.berginsight.com/ReportPDF/ProductSheet/bi-app1-ps.pdf, Last accessed Oct 2013.

[7] Number of apps available in leading app stores as of july 2015. On-line:http://www.statista.com/statistics/276623/number-of-apps- vailable-in-leading-app-stores/, Last accessed Oct 2015.

[8] Sonali Vijay Gaikwad, Archana Chaugule,Pramod Patil "Text Mining Methods and Techniques"

[9] Min-Yuh Day, Yue-Da Lin, “Deep Learning for Sentiment Analysis on Google Play Consumer Review” 2017 IEEE International Conference on Information Reuse and Integration in pp1,2

[10] Phong Minh Vu, Tam The Nguyen, Hung Viet Pham, Tung Thanh Nguyen “Mining User Opinions in Mobile App Reviews: A Keyword-based Approach” in ppr 1, 3, 4.

[11] A. Shabtai, Y. Fledel, and Y. Elovici, "Automated static code analysis for classifying Android applications using machine learning," in Proc. Int. Conf. Com put. Intell. Secur., 2010, pp. 329-333.

[12] Hutto, C.J. and Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

[13] L. Zhuang, F. Jing and X.-Y. Zhu, "Movie review mining and summarization," in Proceedings of the 15th ACM international conference on Information and knowledge management (CIKM '06), ACM, New York, NY, USA, 2006.

[14] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," Foundations and Trends in information Retrieval, vol. 2, no. 1-2, pp. 1-135, 2008.

[15] J. Xu, D.Ho and L.F Capretz, “An Empirical Validation of Object-Oriented Design Metrics for Fault Prediction,” Journal of Computer Science, Vol. 4, No. 7, pp. 571-577, ANSInet, 2008.

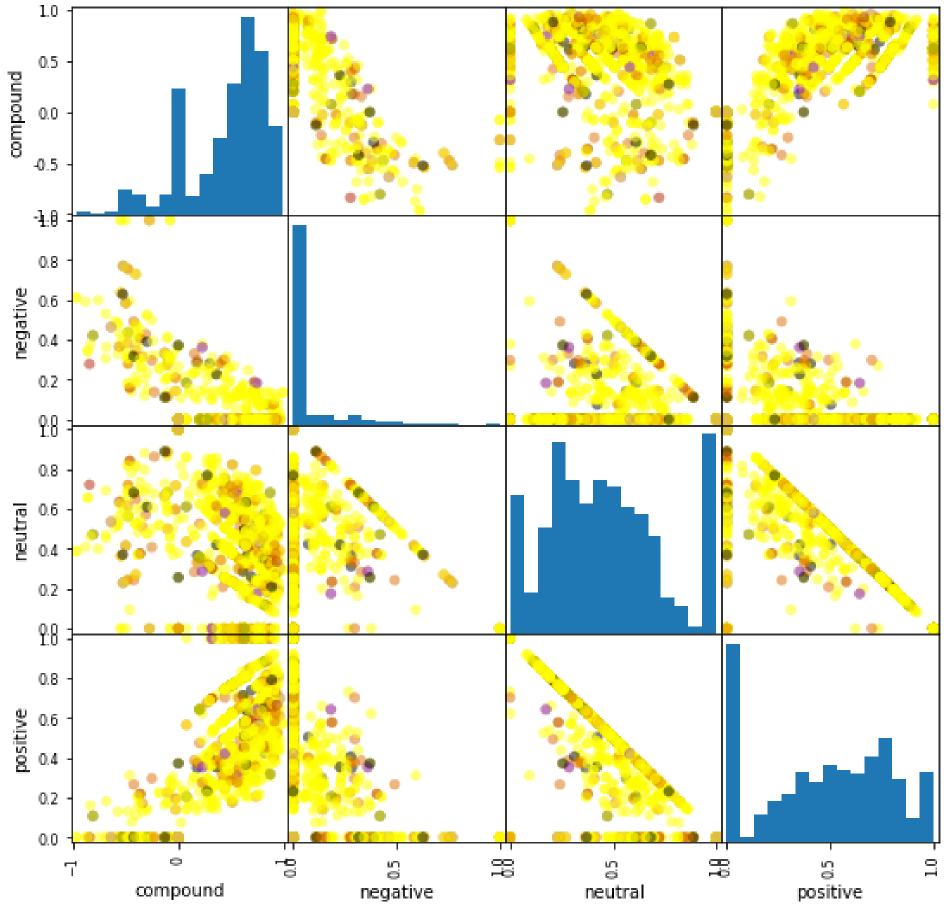
[16] Princeton University, "About WordNet," 2010. [Online] Available: http://wordnet.princeton.edu.

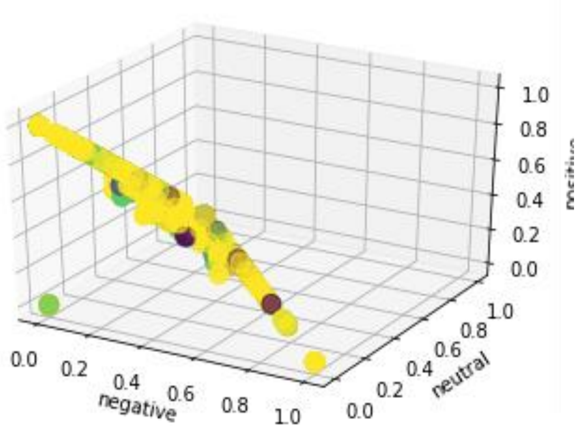
Appendix A

Command Prompt

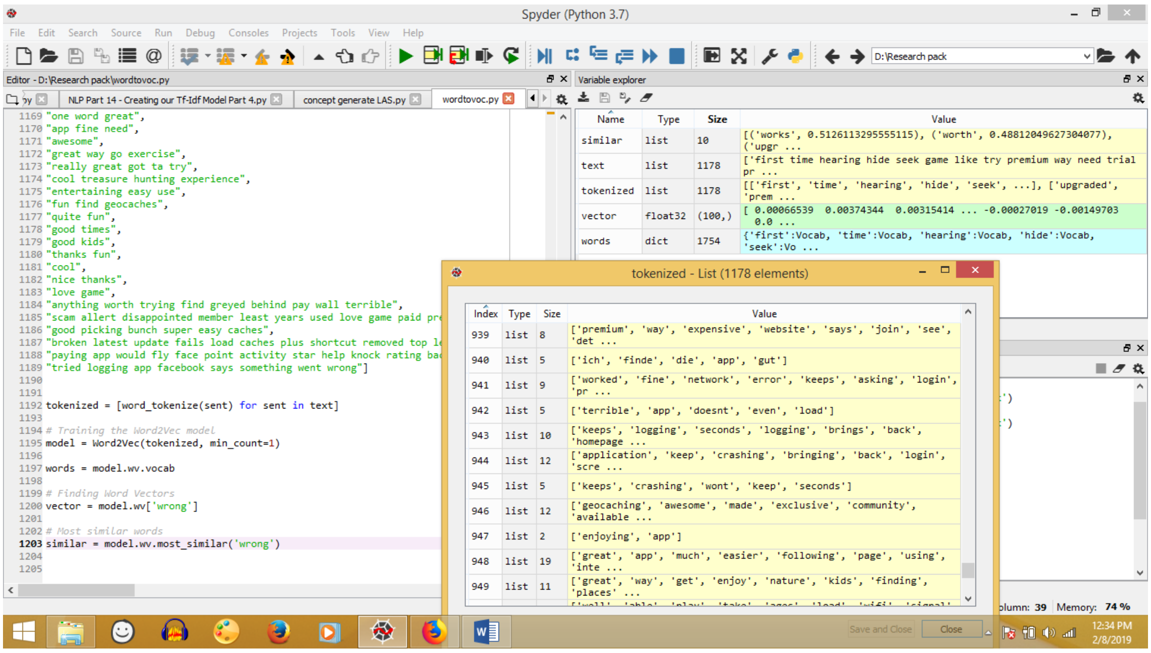
```
C:\>d:

D:\>cd pre

D:\Pre>copy *.csv combine.csv
7_Minute_WorkoutPre.csv
Abs_Workout___Home_Workout__Tabata__HIITPre.csv
BetterMe__Weight_Loss_WorkoutsPre.csv
BodBot_Personal_Trainer__Workout___Fitness_CoachPre.csv
Butt_Fitness_Trainer__Hips_Butt_LegsPre.csv
Calorie_Counter___MyFitnessPalPre.csv
Daily_Yoga___Yoga_Fitness_PlansPre.csv
Diet_Plan_for_Weight_LossPre.csv
Effective_Weight_Loss_GuidePre.csv
Endomondo___Running___WalkingPre.csv
FitbitPre.csv
Fitness___BodybuildingPre.csv
Fooducate_Healthy_Weight_Loss___Calorie_CounterPre.csv
GeocachingPre.csv
Headspace__Meditation___MindfulnessPre.csv
Home_Workout___No_EquipmentPre.csv
Home_Workout___No_Equipment___Meal_Planner.csv
Huawei_HealthPre.csv
Huawei_WearPre.csv
Keep_Trainer___Workout_Trainer___Fitness_CoachPre.csv
LG_HealthPre.csv
Lose_Belly_Fat_in_30_Days___Flat_StomachPre.csv
Meditation_Music___Relax__YogaPre.csv
Mi_FitPre.csv
Prana_Breath__Calm___MeditatePre.csv
Relax_Music___Meditation__Sleep_Music__White_NoisePre.csv
Running_Distance_TrackerPre.csv
Runtastic_Results_Strength_Training___BodyweightPre.csv
Six_Packs_for_Man_Body_Building_with_No_EquipmentPre.csv
Six_Pack_in_30_Days__Abs_WorkoutPre.csv
Sleep_Cycle_alarm_clockPre.csv
Sleep_SoundsPre.csv
Step_Tracker___Pedometer__Walking_for_weight_lossPre.csv
ViewRanger___Hike__Ride_or_WalkPre.csv
Weight_Loss_Running_by_VervPre.csv
        1 file(s) copied.

D:\Pre>
```

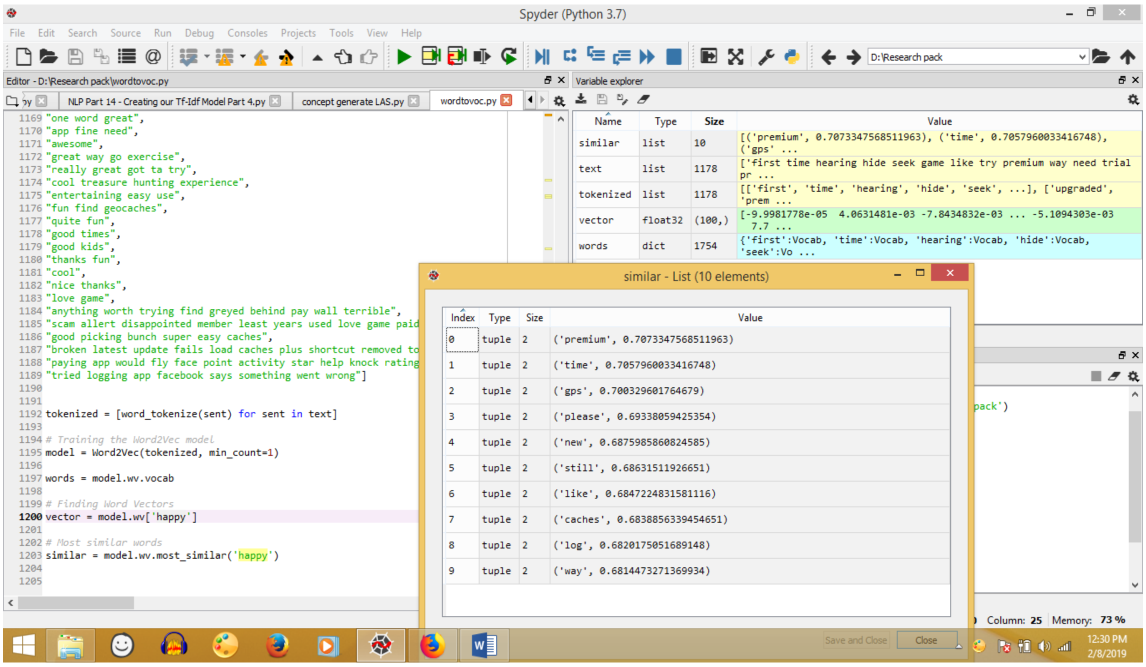
Appendix B



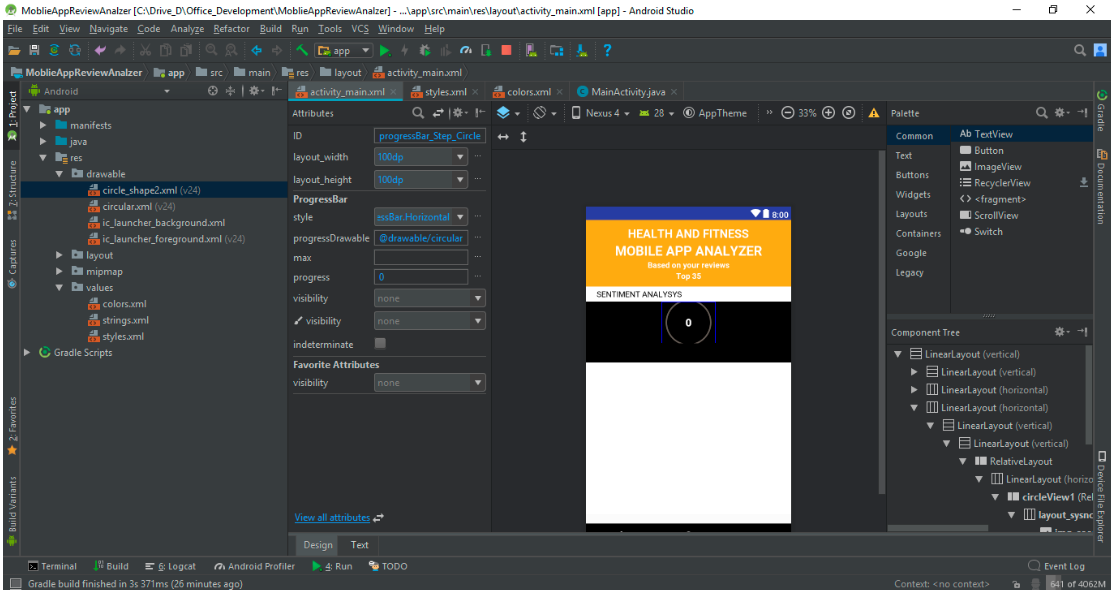


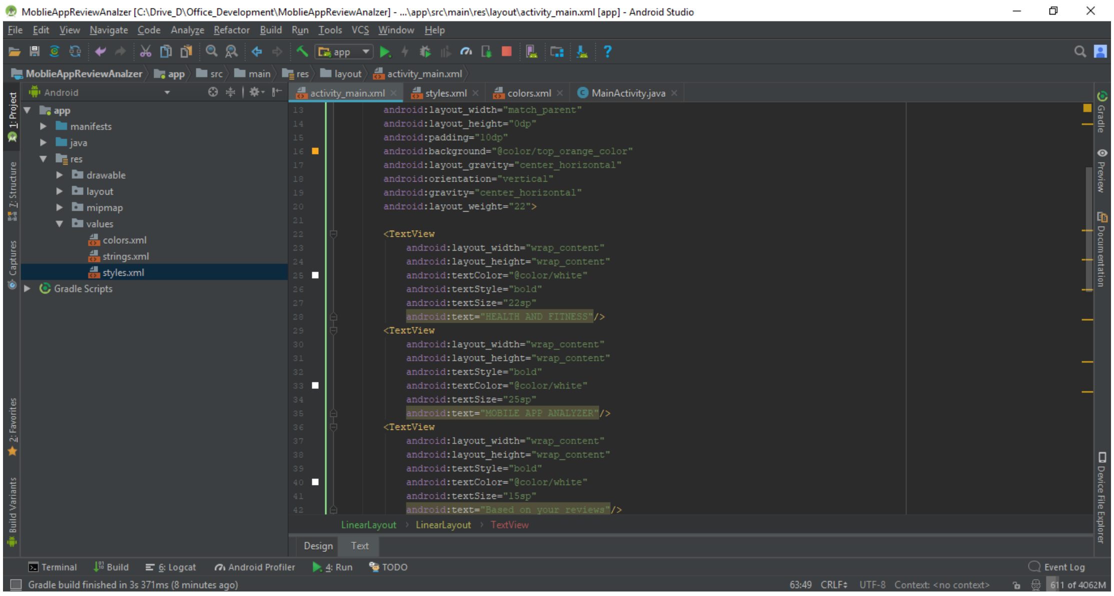
Appendix C


Spyder (Python 3.7)
File Edit Search Source Run Debug Consoles Projects Tools View Help
D:\Research pack
Editor - D:\Research pack\wordtovoc.py
NLP Part 14 - Creating our Tf-Idf Model Part 4.py
concept generate LAS.py
wordtovoc.py
"one word great",
"app fine need",
"awesome",
"great way go exercise",
"really great got ta try",
"cool treasure hunting experience",
"entertaining easy use",
"fun find geocaches",
"quite fun",
"good times",
"good kids",
"thanks fun",
"cool",
"nice thanks",
"love game",
"anything worth trying find greyed behind pay wall terrible",
"tried logging app facebook says something went wrong"]
tokenized = [word_tokenize(sent) for sent in text]
# Training the Word2Vec model
model = Word2Vec(tokenized, min_count=1)
words = model.wv.vocab
# Finding Word Vectors
vector = model.wv['wrong']
# Most similar words
similar = model.wv.most_similar('wrong')
Variable explorer
Name Type Size Value
similar list 10 [('works', 0.5126113295555115), ('worth', 0.48812049627304077), ('upgr ...
text list 1178 ['first time hearing hide seek game like try premium way need trial pr ...
tokenized list 1178 [['first', 'time', 'hearing', 'hide', 'seek', ...], ['upgraded', 'prem ...
vector float32 (100,) [0.00066539 0.00374344 0.00315414 ... -0.00027019 -0.00149703 0.0 ...
words dict 1754 {'first':Vocab, 'time':Vocab, 'hearing':Vocab, 'hide':Vocab, 'seek':Vo ...
tokenized - List (1178 elements)
Index Type Size Value
939 list 8 ['premium', 'way', 'expensive', 'website', 'says', 'join', 'see', 'det ...
940 list 5 ['ich', 'finde', 'die', 'app', 'gut']
941 list 9 ['worked', 'fine', 'network', 'error', 'keeps', 'asking', 'login', 'pr ...
942 list 5 ['terrible', 'app', 'doesnt', 'even', 'load']
943 list 10 ['keeps', 'logging', 'seconds', 'logging', 'brings', 'back', 'homepage ...
944 list 12 ['application', 'keep', 'crashing', 'bringing', 'back', 'login', 'scre ...
945 list 5 ['keeps', 'crashing', 'wont', 'keep', 'seconds']
946 list 12 ['geocaching', 'awesome', 'made', 'exclusive', 'community', 'available ...
947 list 2 ['enjoying', 'app']
948 list 19 ['great', 'app', 'much', 'easier', 'following', 'page', 'using', 'inte ...
949 list 11 ['great', 'way', 'get', 'enjoy', 'nature', 'kids', 'finding', 'places' ...
Save and Close
Close
Memory: 74 %
12:34 PM
2/8/2019


Spyder (Python 3.7)
File Edit Search Source Run Debug Consoles Projects Tools View Help
D:\Research pack
Editor - D:\Research pack\wordtovoc.py
NLP Part 14 - Creating our Tf-Idf Model Part 4.py
concept generate LAS.py
wordtovoc.py
tokenized = [word_tokenize(sent) for sent in text]
# Training the Word2Vec model
model = Word2Vec(tokenized, min_count=1)
words = model.wv.vocab
# Finding Word Vectors
vector = model.wv['happy']
# Most similar words
similar = model.wv.most_similar('happy')
Variable explorer
Name Type Size Value
similar list 10 [('premium', 0.7073347568511963), ('time', 0.7057960033416748), ('gps' ...
text list 1178 ['first time hearing hide seek game like try premium way need trial pr ...
tokenized list 1178 [['first', 'time', 'hearing', 'hide', 'seek', ...], ['upgraded', 'prem ...
vector float32 (100,) [-9.9981778e-05 4.0631481e-03 -7.8434832e-03 ... -5.1094303e-03 7.7 ...
words dict 1754 {'first':Vocab, 'time':Vocab, 'hearing':Vocab, 'hide':Vocab, 'seek':Vo ...
similar - List (10 elements)
Index Type Size Value
0 tuple 2 ('premium', 0.7073347568511963)
1 tuple 2 ('time', 0.7057960033416748)
2 tuple 2 ('gps', 0.700329601764679)
3 tuple 2 ('please', 0.69338059425354)
4 tuple 2 ('new', 0.6875985860824585)
5 tuple 2 ('still', 0.68631511926651)
6 tuple 2 ('like', 0.6847224831581116)
7 tuple 2 ('caches', 0.6838856339454651)
8 tuple 2 ('log', 0.6820175051689148)
9 tuple 2 ('way', 0.6814473271369934)
Save and Close
Close
Column: 25 Memory: 73 %
12:30 PM
2/8/2019

Appendix D


MoblieAppReviewAnalzer [C:\Drive_D\Office_Development\MoblieAppReviewAnalzer] - ...\app\src\main\res\layout\activity_main.xml [app] - Android Studio
HEALTH AND FITNESS
MOBILE APP ANALYZER
SENTIMENT ANALYSYS
Gradle build finished in 3s 371ms (26 minutes ago)


MoblieAppReviewAnalzer [C:\Drive_D\Office_Development\MoblieAppReviewAnalzer] - ...\app\src\main\res\layout\activity_main.xml [app] - Android Studio
android:layout_width="match_parent"
android:layout_height="0dp"
android:padding="10dp"
android:background="@color/top_orange_color"
android:layout_gravity="center_horizontal"
android:orientation="vertical"
android:gravity="center_horizontal"
android:layout_weight="22">
<TextView
android:layout_width="wrap_content"
android:layout_height="wrap_content"
android:textColor="@color/white"
android:textStyle="bold"
android:textSize="22sp"
android:text="HEALTH AND FITNESS"/>
<TextView
android:layout_width="wrap_content"
android:layout_height="wrap_content"
android:textStyle="bold"
android:textColor="@color/white"
android:textSize="25sp"
android:text="MOBILE APP ANALYZER"/>
<TextView
android:layout_width="wrap_content"
android:layout_height="wrap_content"
android:textStyle="bold"
android:textColor="@color/white"
android:textSize="15sp"
android:text="Based on your reviews"/>
Gradle build finished in 3s 371ms (8 minutes ago)

Appendix E

Total number of comments and number of topic related user reviews

No	Name of Mobile App	Total comment	Fitness	Money	Quality	Performance	Error	Health care
1	7-Minute-Workout	1240	429	60	225	143	54	209
2	Abs-Workout-Home-Workout-Tabata-HIIT	1275	206	60	105	228	162	56
3	BetterMe-Weight-Loss-Workouts	1199	305	178	117	127	127	128
4	BodBot-Personal-Trainer-Workout-Fitness-Coach	1316	581	252	227	326	178	266
5	Butt-Fitness-Trainer-Hips-Butt	280	57	5	39	15	8	22
6	Calorie-Counter-MyFitnessPal	1320	550	150	146	357	169	701
7	Daily-Yoga-Yoga-Fitness-Plans	1280	437	167	177	299	81	89
8	Diet-Plan-for-Weight-Loss	1400	258	24	157	65	100	245
9	Effective-Weight-Loss-Guide	1361	268	37	156	73	80	138
10	Endomondo-Running-Walking	1320	310	117	134	337	113	101
11	Fitbit	1279	405	150	110	734	513	607
12	Fitness-Bodybuilding	1318	319	85	132	174	64	121
13	Fooducate-Healthy-Weight-Loss	1280	328	140	178	262	113	679
14	Geocaching	1280	111	191	102	249	327	9
15	Headspace-Meditation-Mindful	1280	461	228	282	285	206	177
16	Home-Workout-No-Equipment	1279	664	92	259	242	113	377
17	Home-Workout-No-Equipment-M	1360	291	36	174	81	31	144
18	Huawei-Health	1320	481	56	138	695	304	409
19	Huawei-Wear	1400	292	65	82	791	339	308
20	Keep-Trainer-Workout-Trainer-	1258	413	34	163	75	48	249
21	LG-Health	1360	523	38	171	332	254	217
22	Lose-Belly-Fat-in-30-Days-Fla	1358	338	44	198	104	78	172
23	Meditation-Music-Relax-Yoga	1442	322	23	225	116	38	231
24	Mi-Fit	1320	277	41	95	519	264	220
25	Prana-Breath-Calm-Meditate	1280	513	93	285	220	61	313
26	Relax-Music-Meditation-Sleep	1280	96	55	85	184	166	145
27	Running-Distance-TrackerPre	1320	439	96	226	300	126	45
28	Runtastic-Results-Strength-Trai	1320	385	333	146	258	142	146
29	Six-Pack-in-30-Days-Abs-Worko	1324	456	71	208	197	165	255
30	Six-Packs-for-Man-Body-Building	1272	171	39	71	69	68	57
31	Sleep-Cycle-alarm-clock	1320	81	152	272	224	74	448
32	Sleep-Sounds	1258	103	39	233	148	49	551
33	Step-Tracker-Pedometer-Walki	1195	315	46	94	129	126	89
34	ViewRanger-Hike-Ride-or-Walk	1400	403	242	389	765	167	28
35	Weight-Loss-Running-by-Verv	1361	467	231	145	266	174	95
	Grand Total	**44855**	**12055**	**3670**	**5946**	**9389**	**5082**	**8047**