UNIVERSITI PUTRA MALAYSIA

MOBILE APPLICATIONS QUALITY EVALUATION ACCORDING TO ISO 9126 QUALITY MODEL USING NAÏVE BAYES

JABIR ABDULLAHI ALI

FSKTM 2018 60

MOBILE APPLICATIONS QUALITY EVALUATION ACCORDING TO

ISO 9126 QUALITY MODEL USING NAÏVE BAYES

BY

JABIR ABDULLAHI ALI

GS44897

This thesis was submitted to the Faculty of Computer Science and Information Technology, University Putra Malaysia, in Fulfilment of the Requirements for the Master of Computer Science.

MAY, 2018

SUPERVISOR CONFIRMATION

MOBILE APPLICATIONS QUALITY EVALUATION ACCORDING TO
ISO 9126 QUALITY MODEL USING USER REVIEWS AND STAR

Had been prepared by:

JABIR ABDULLAHI ALI

GS44897

Has been checked and approved by:

Dr. Koh Teing Wei

Checked and approved by:

_____

(Dr. Koh Teing Wei)

Project Supervisor

Department of Software Engineering and information System

Faculty of Computer Science and Information Technology

University Putra Malaysia

Date:_____

DECLARATION

I hereby declare that this thesis, submitted to University Putra Malaysia as a fulfilment of the requirements for the Masters of Computer Science has not been submitted previously or concurrently as an exercise for a degree at UPM or any other university. I also certify that the work described here is entirely my own except for excerpts and summaries whose sources are appropriately cited in the references.

Student's Signature,

_____

(JABIR ABDULLAHI ALI)

Date: June 2018

ACKNOWLEDGEMENT

All thanks be to Almighty Allah for making it possible for me to conclude this program. I am also thankful to my beloved parents Abdullahi Ali and Anbara Adan Farah for their love, support and prayer throughout my life.

I specially wish to express my propound gratitude to my supervisor Dr. Koh Teing Wei for his patience, support, guidance and supervision throughout the research period. His meaningful and resourceful contributions have made this work exceptional.

This acknowledgement will not be complete without extending my gratitude to my friends and colleagues within and outside the department who stood by me to ensure that we succeed. Friends I thank you very much for the company and support.

Finally, I have to thank all my family and friends at home whose understanding, support and word of encouragement in one way or the other helped immensely toward the completion of my program. Indeed, your kinship and friendship really contributed toward making this program a success.

ABSTRACT

In this study a tool to evaluate the quality of mobile applications from the user provided reviews in Google Play app store is developed. ISO 9126 quality in use part is used as the quality model. We used a supervised Machine learning technique to undertake this project. Specifically, Weka Naïve Bayes classifier is employed. We downloaded 2000 reviews from Google play, used 70% for training the classifying model and 30% for testing the model. The outcome of the manual labelling of the reviews is that 96% of the reviews are categorized as either satisfaction or effectiveness. This suggests that users tend to talk more about how they like or dislike a mobile app or complain about the ineffectiveness of it. Due to this skewed nature of the data, the classifying model testing part of the study yielded the expected outcome. The model precision is high for both Satisfaction and effectiveness quality characteristics of the ISO 9126 quality in use part, since both of them got large training data sets. However, due to the minimal training reviews received by the safety and productivity categories their precision lags. Therefore, the study suggests that users are more concerned about the effectiveness of an app and how satisfied the use of the app and its features makes them.

ABSTRAK

TABLE OF CONTENTS

CHAPTER 3 METHODOLOGY

CHAPTER 4 RESULTS AND ANALYSIS

CHAPTER 5 CONCLUSION AND FUTURE WORK

LIST OF TABLES

LIST OF FIGURES

xi

CHAPTER 1

INTTRODUCTION

1.1 BACKGROUND

Ever since the advent of the internet has spawned the creation of dynamic websites, large amounts of data are being created every second. These data range from social network interactions among users, news articles and blog posts. One major property of this dynamicity is the ability of users to interact with and deliver their opinions regarding these online posts. These comments contain valuable information that can be utilized for understanding the opinions of one's audience. One major example is the use of tweets to analyse social events, political movements and monitoring reputation of organizations (Bogdanov, 2013). These tweets are used to predict the election outcomes in countries such as the US (Lei Shi, Neeraj Agarwal, Ankur Agrawal, no date). Even though this analysis of user feedback seems promising, the standing question is how is all this related to Software engineering and how we as software engineers can capitalize on this area.

User reviews are pervasive in ecommerce sites, as well as mobile app stores such as apple's apple store, Google's playstore and Microsoft store. Users religiously provide their feedback as reviews. These reviews include a wide range of topics including user experience, feature request, bug report as well as simple praise (Nabil and Stanik, 2016). This feedback can be used by software engineers, developers and analysts to enhance their product and fulfil the needs of their users and customers. However, the magnitude and number of these reviews make it quite challenging for software developers to take a full advantage of it. Developers resort to manually read each single review and analyse it. This task can be less demanding if the number of reviews are minimal, but imagine if there are thousands or even millions of them. Users on the other hand rely on reviews previously provided by users in order to download an app or buy it. This shows the importance of user reviews to both developers and users.

It has been confirmed that reviews of mobile apps have a major impact on the successof an app (Of *et al.*, 2010; Kim, Lee and Son, 2011; Harman, Jia and Zhang, 2012). Userreviews contain information that could help developers improve the quality of their apps, and increase their revenue. Kim et al. conducted interviews of app buyers and discovered that reviews are one of the key determinants in the user's purchase of an app(Kim, Lee and Son, 2011).

1

Similarly, Mudambi and Schuff showed that user reviews have a major impact on the sales of onlineproducts(Of *et al.*, 2010). Harman et al. have shown a strong correlation between app ratings and thetotal downloads of an app(Harman, Jia and Zhang, 2012).

User reviews contain valuable information which may help developers better understand user needs and complaints during software maintenance and evolution. Making use of user reviews to mine valuable information for improving Apps is critical for retaining the existing users and attracting new users. Studies show that over one third of users changed their ratings following a developer response, and the median rating change is a one-star increase out of five (S. McIlroy, W. Shang, N. Ali, 2015).

1.2 PROBLEM STATEMENT

User reviews contain valuable insights that can be used to improve the quality of mobile applications. Users post the difficulties they have faced while using the app, the improvements they would like to see in the next versions as well as their general experience with the app. The developers, however, face significant challenges in compiling those reviews as they have to go through thousands or sometimes even millions of user's feedback in order to harness its potential. This method proved to be tricky since it can potentially waste a lot of valuable time and is inefficient. This problem calls for the establishment of an automatic technique that can identify the quality of an app based on software quality models that are available.

1.3 OBJECTIVE OF THE STUDY

The objective of this study is to implement a prototype tool that would automatically evaluate the quality of an app based on the user provided reviews. This study will analyse user reviews and based on those reviews evaluate the quality of the app under hand. In this study a prototype tool that would automatically identify the quality of an app based on the ISO 9126 quality model will be developed. Various techniques are going to be implemented which are discussed later.

1.4SCOPE OF THE STUDY

There is various way to go about classifying user reviews. However, this study will use Naïve Bayes algorithm technique to undertake this study and the quality models that will be used in this study is ISO 9126. The emphasis of this study is to produce a tool that will help

developers in the software evolution process and enhance future versions of their app by taking advantage of the large amounts of feedback users submit to app stores. To be specific, the app storethat will be investigatedis Google's Playstore. This work is supposed to be accomplished in 14 weeks.

1.5 EXPECTED OUTCOME

This study and its resultant tool will improve the application evolution process as well as significantly minimizing the time software developers and analysts would spend on analysing and classifying user reviews.

1.6 ORGANIZATION OF THE THESIS

There are 4 chapters in this report. Chapter 1 is the introduction chapter where a brief overview of what the study is provided. In this chapter the problem statement, objectives and the expected results of the study are discussed. In chapter 2, literature review is conducted. How different quality gurus define software quality and also look at various software quality models are explored. Classification techniques are also discussed. And lastly studies related to this one are shown. In chapter 3, the methodology used in this study is discussed. And chapter 4 is about the result and analysis of the study.

REFERENCES

Bhardwaj, A. and Kapoor, A. (2014) 'Analysis of software quality model', 1(6), pp. 853–856.

Bird, S., Klein, E., Loper, E., Bird, S., Klein, E., Loper, E., Bird, C. S., Klein, E. and Loper, E. (2009) *Natural Language Processing with Python*. O'Reilly Media, Inc.

Boehm, Barry W., Brown, J. R, and Lipow, M. (1976) 'Quantitative evaluation of software quality', in *2nd international conference on Software engineering*.

Bogdanov, J. W. M. and P. (2013) 'Introduction — Topic models : What they are and why they matter', *Poetics*, 41, pp. 545–569. doi: 10.1016/j.poetic.2013.10.001.

Crosby, B. P. B. (1989) *Let ' s Talk Quality : 96 Questions You Always Wanted to Ask Phil Crosby (Plume)*. New York: McGraw-Hill.

Crosby, P. B. (1979) *Quality is free the art of making quality certain .pdf*. New York: McGraw-Hill.

Crosby, P. B., Deming, W. E., Juran, J. M. and Suarez, J. G. (1992) 'Three Experts on Quality Management : by', (92).

Deming, W. E. (1989) 'Foundation for management of quality in the Western World', in *Institute of Management and Sciences*. Oska.

Fatima, J. and Arora, D. (2018) 'Classification Approach to Extract Strongly Liked and Disliked Features Through Online User Opinions', pp. 123–135.

Harman, M., Jia, Y. and Zhang, Y. (2012) 'App Store Mining and Analysis : MSR for App Stores', pp. 108–111.

Hoyer, R. W. and Hoyer, B. B. Y. (2001) 'What Is Quality ?', *Quality Progress*, 7, pp. 52–62.

ISO, I. O. for S. (2001) 'ISO 9126-1:2001, Software engineering –Product quality, Part 1: Quality model'.

Juran, J. M., & Gryna, F. M. (Eds.). (1988) *Quality Control Handbook*. New York: McGraw-Hill.

Kim, H., Lee, H. L. and Son, J. E. (2011) 'An exploratory study on the determinants of smartphone app purchase', in *The 11th International DSI and the 16th APDSI Joint Meeting*. Taipei, Taiwan.

L, T. (2010) *Data mining with R: learning with case studies*. Chapman & Hall/ CRC, Boca Raton.

Lei Shi, Neeraj Agarwal, Ankur Agrawal, R. G. (no date) *Predicting US Primary Elections with Twitter*.

Lu, M. (2017) 'Automatic Classification of Non-Functional Requirements from Augmented App User Reviews'.

McCall, J. A., Richards, P. K., and Walters, G. F. (1977) 'Factors in Software Quality', *Nat'l Tech.Information Service*, I,2,3(November).

Mcilroy, S., Ali, N., Khalid, H. and Hassan, A. E. (2016) 'Analyzing and automatically labelling the types of user issues that are raised in mobile app reviews', *Empirical Software Engineering*. Empirical Software Engineering, pp. 1067–1106. doi: 10.1007/s10664-015-9375-7.

Nabil, H. and Stanik, C. (2016) 'On the automatic classification of app reviews', *Requirement engineering*, pp. 311–331. doi: 10.1007/s00766-016-0251-9.

Ning Chen, Jialiu Lin, Steven HOI, Xiaokui Xiao, B. Z. (2014) *AR-Miner : Mining Informative Reviews for Developers from Mobile App Marketplace*.

Number of apps available in leading app stores as of March 2017 (no date). Available at: https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/.

Of, A. S. T., Eviews, C. U. R., Mazon, O. N. A., Mudambi, B. S. M., Hall, A., Schuff, D. and Hall, S. (2010) 'What makes a helpful online review? A study of customer reviews on amazon.com', 34(1), pp. 185–200.

Pang, B. and Lee, L. (2008) 'Opinion Mining and Sentiment Analysis', 2, pp. 1–135. doi: 10.1561/1500000001.

Pang, B., Lee, L., Rd, H. and Jose, S. (1988) 'Thumbs up ? Sentiment Classification using Machine Learning Techniques'.

Robson, C. (2002) *Real World Research A resource for Social Scientists and Practitioner-Researchers .pdf*. Blackwell Publisher Ltd.

S. McIlroy, W. Shang, N. Ali, and A. H. (2015) 'Is it worth responding to reviews? A case study of the top free apps in the Google Play store.pdf', *IEEE Software*. doi: 10.1109/MS.2015.149.

Warit Leopairote, Athasit Surarerks, N. P. (2013) 'Evaluating Software Quality in Use using User Reviews Mining', in *10th International Conference on Computer Science and Software Engineering (JCSSE).*, pp. 257–262.

Zhang, L., Hua, K., Wang, H., Qian, G. and Zhang, L. (2014) 'Sentiment Analysis on Reviews of Mobile Users', *Procedia - Procedia Computer Science*. Elsevier Masson SAS, 34, pp. 458–465. doi: 10.1016/j.procs.2014.07.013.

Zhang, W., Yang, Y., Wang, Q. and Shu, F. (no date) 'An Empirical Study on Classification of Non-Functional Requirements'.

Zheng, W. and Ye, Q. (2009) 'Sentiment Classification of Chinese Traveler Reviews by Support Vector Machine Algorithm'. doi: 10.1109/IITA.2009.457.