Quality Measurement of Android Messaging Application Based on User Experience in Microblog

Rakhmat Arianto[1], Ford Lumban Gaol[1],
Edi Abdurachman[1], Yaya Heryadi[1], Harco Leslie Hendric Spits Warnars[1], Benfano Soewito[2]
[1]Doctor of Computer Science, [2]Graduate Computer Science
Bina Nusantara University
Jakarta – Indonesia
arianto@sttpln.ac.id, rakhmat.arianto@binus.ac.id

Horacio Perez – Sanchez
Bioinformatics and High Performance Computing Research Group (BIO-HPC), Computer Engineering Department
Universidad Católica San Antonio de Murcia (UCAM), Spain
hperez@ucam.edu

***Abstract*—There are many options of android messaging application which give opportunity to user in order to choose which one as best or famous android messaging application and make it become suitable for them. Usually, people used to look at the information about best or famous android messaging application by texting in search engine such as google and get some link information from user/blogger reviews, and based on that reviews they will make decisions which one as suitable for them. We proposed the other way how to measure the quality of each android messaging application based on user experience which they text in Microblog such as Twitter. The unstructured data in the Microblog will be processed with 2 operators for sentiment analysis method in RapidMiner such as AYLIEN and ROSETTE. AYLIEN sentiment analysis has 3 categories such as positive, negative, and neutral, whilst ROSETTE sentiment analysis has 2 categories such as positive and negative sentiments. Finally, the finding sentiment analysis with these 2 operators will be compared with PlayStore review.**

Keywords— Unstructured Data, Natural Language Processing, Sentiment Analysis, RapidMiner, Microblog

I. Introduction

The Android operating system has many types and application categories which can be used by users, and one of the most commonly used applications are messaging app where this program serves to facilitate the users in communicating with others through unstructured multimedia content such as text, voice, video, or picture. However, regarding with many different options of messaging apps, the user will face the confusing to determine what suitable messaging app for them, even though there are user reviews which described specification and the advantages of each application. Moreover, Users usually often to search for other user reviews information that review the application via media social or in their websites that specifically discuss the messaging application.

When users search for additional information through social media or blogs that discuss about an application, the user will be faced with many reviews of various parties which will make new problem for users due to make decision to choose suitable apps for them. Some of the user reviews are nothing in common with one another depending on preferences or aspects of the subjectivity of each reviewer and shared writing is not necessarily a form of review to determine whether or not an application desired by the user. These circumstances will enhance the difficulty of users in order to determine which suitable application will be installed in their gadget.

Comments or opinions of user review can be used as a reference for other users who want to install the suitbale applications into their smartphones. In addition, users' comments and opinions can be used to measure the quality of a software and moreover the opinions or comments of users can mention anything regarding with the software, including some attributes of software quality measurement such as security, reliability and user-friendliness[1]. For example, when Windows wants to know whether Windows 10 demanded by the user by removing the beta, and provide facilities to users who want to send the impression while using Windows 10. Based on user's review and comments, then Windows will fix and improve their Windows10.

In measuring the quality of software, we will refer to how well the design used by the software and how many expectations that can be met by the software to the customer [1]. Meanwhile, a quality of software which based on the opinions or reviews of the end-user, can be done by hand or manually, but it will take much time. One of solution for automation of software quality based user's review and comments can be measured with sentiment analysis.

The measurement of software quality based on sentiment analysis which is user's review and comments, will be done by extraction of textual opinion on Android messaging apps that are on Twitter in order to get the quality attributes of a software. The measurement of software quality based on sentiment analysis will be done in 2 steps. Firstly, Twitter's messages related to chosen keyword will be gathered. Secondly, all of the data gathering from 1st activity will be analyzed using RapidMiner which implement Sentiment Analysis either with AYLIEN [2] or ROSETTE [3] Operators. Sentiment analysis using AYLIEN will produce polarities in three categories such as positive, negative, and neutral, whilst ROSETTE generates polarity in two categories between positive and negative. The contribution of this paper will show

978-1-5090-4048-3/17/$31.00 ©2017 IEEE

which one between AYLIEN and ROSETTE operators will have better measurement compared to user review in PlayStore.

II. Related Works

Pagano and Maalej [4], has conducted exploration studies by analyzing more than a million reviews of the Apple AppStore. The researchers investigated how and when users provide feedback, examined the contents of the feedback, and analyzed their impact on communities of users. The researchers found that most of the feedback is provided shortly after the new release, with fast frequency decreased over time. Reviews usually contain several topics, such as user experience, bug reports, and feature requests. Quality and constructive varied, ranging from helpful advice and innovative ideas to words that leads to humiliation. feedback content has an impact on the number Download: positive messages usually leads to better rankings and vice versa. negative feedback as deficiencies are usually destructive and not in accordance with the details of the context and the user experience. The researchers discuss our findings and their impact on a team of software and engineering requirements.

Galvis and Kritina [5], has conducted the research by analyzing feedback from users to third-party mobile applications to detect changes or renewal of the needs of the user. The main problem of using the data feedback from users is the amount of data used and requires a lot of time to the process. Researchers processing comments from the users to take the main topic of the comments thus obtained connectedness comments by the application used. This study is based on research that says that comments from users may be used by software developers as one factor supporting the care process and software development [6].

Guzman and Maalej [7], proposed an automated approach that helps the software developer in filtering, aggregating and analyzing user reviews applications from the App Store. Researchers used natural language processing to obtain information from user reviews. Then extraction sentiment about the features that users are identified and provide an assessment in all reviews. The final part of the process, topic modeling techniques used for fine-grained group features into more meaningful high-level features. This study uses data collected user reviews from 7 apps in the Apple App Store and Google Play Store. This study is based on a previous study [5], [8] and [4] Showed that app store reviews include information that is useful to analysts and app designers, such as user requirements, bug reports, feature requests, and documentation of user experiences with app specific features. This feedback can represent a "voice of the users" and be used to drive the development effort and improve forthcoming releases [9], [10].

Liang et al [11], conducted a study that aimed to examine the effect of a textual review of consumers on mobile application sales. The researchers examined how the sentiment of different topics in an online review affect the sales application. The researchers developed a multifacet sentiment analysis (MFSA) approach to measure dimensions in consumer reviews. In particular, the researchers focused on comments related to the product quality and service quality of an application. Employing a set of real-world seventy-nine paid applications and seventy free applications from the app store iOS, the researchers found that even though consumers' opinion about the quality of the products occupy a larger portion of the consumer reviews, their comments about the quality of service has the effect of a unit strong in sales rank.

In marketing literature, WOM (Word of Mouth) has been well recognized as influencing consumers' purchasing behavior [12]. Cunningham [13] pointed out that consumers are likely to generate conversations related to products and to request information from friends and relatives if they are not sure about a purchase. Bone [14] found that WOM influences short-term and long-term product judgments, especially when a customer faces uncertainties.

Many scholars consider eWOM (Online Word of Mouth) as a determinant of product success [15]–[18] that is moderated by the characteristics of products [18] and consumers [19]. External WOM sources have been found to have a significant effect on retail sales. Recent studies have also analyzed the interplay between online consumer reviews and recommender systems in consumers' decision making [20] and the formation of helpfulness of online product reviews [21].

III. Experimental Details

In this study, we used a tool to assist in the analysis is to use RapidMiner [22] where this tool can help in the preprocessing stage up to the stage to show the test results. his paper focuses on the RapidMiner software package to preprocess and analyze the data and mine diabetes a diabetes prediction models. In RapidMiner we can determine the process model in accordance with what we want. The process consists of several operators. The focus of this study is the use of operator Sentiment Analysis owned by AYLIEN with Sentiment Analysis operator owned by ROSETTE.

The composition of the stages in research experiments as shown in Figure 1. In general, the stages to be carried out in the experiment are in four stages and they are:

1. Gathering Tweets, at this stage we will be looking for Tweets related to the messaging app on Android using Search Twitter been provided in RapidMiner as well as screening the form of tweets retweet and tweet a link.
2. Analyzing Tweets for Sentiment, used to determine the sentiment contained in Tweets that have been collected in the previous stage. The results of this phase, we can know the kind of sentiment contained in tweets is divided into three parts, namely positive, negative, or neutral and the level of subjectivity or objectivity of the tweet.
3. Tweet categorizing, this stage will process the categorization of each Tweet so we can figure out a tweet that has been collected is included in any category.

4. Visualizing the results, this phase we will show you the overall sentiment on Tweet detection results with the results of categorization has been done.

A. Gathering Tweets

In this first step is done by creating a new Process in RapidMiner and add a Search Twitter Operator. We build desired search as we would use the Twitter search API with the keyword "WhatsApp". We've cleaned up our search a little by removing retweets (-rt) and links (-http). We've also restricted the number of tweets to collect to 50 and decided we only want to see English tweets by adding "en" in the language parameter. We've also indicated that we want only recent or popular tweets to be returned using the Result type parameter. An example result of the gathering tweets presented in Table 1.

TABLE 1. EXAMPLE RESULT OF GATHERING TWEETS

ID	Text
7.6621363844527309E17	Great times for Jamie Vardy as Mahrez signs the new contract and rejoins the Whatsapp group... https://t.co/7GqOGobAvR
7.6599426784837222E17	2016 internet minute: 150 million emails 21m whatsapp 2.8m youtube 701K facebook 347K twitter @valaafshar. https://t.co/vGs8jnP0RK
7.6633761518442086E17	Hope @Pvsindhu1 phone is a mile away from her. No gushing calls from 'friends', no WhatsApp & no Twitter. Let her express herself on court.
7.6653909757377331E17	Stall booking for Corporates - P2,000 and Small Businesses - P300. Pls call or whatsapp - 77175107. https://t.co/G9MncsrnOy
7.6653908679017677E17	!!! "@Sirkastiq: Someone needs to do a whatsapp BC to all parents saying "STOP FORWARDING BROADCASTS TO YOUR KIDS!""

B. Analyzing Tweets for Sentiment

In this second step, after the first step we have a collection of 50 tweets stored in an ExampleSet that are ready to be further analyzed. Then, we're going to do from an analysis point of view is, try and determine what the Sentiment of each tweet is, i.e. whether they are Positive, Negative or Neutral. We do this by adding the Analyze Sentiment Operator to our Process and selecting "text" as our "Input attribute". An example result of the analyzing tweets for sentiment presented in Table 2.

TABLE 2. EXAMPLE RESULT OF ANALYZING TWEETS FOR SENTIMENT

ID	Polarity	Subjectivity Confidence	Polarity	Subjectivity
7.6621363844527309E17	0.8263487219810486	0.99938361856067	neutral	subjective
7.6599426784837222E17	0.9908341765403748	0.999999999999996	neutral	objective
7.6633761518442086E17	0.7654720544815063	1.0	negative	subjective
7.6653909757377331E17	0.9718964099884033	0.9999999920902486	neutral	objective
7.6653908679017677E17	0.8124754428863525	0.5171833302407525	negative	subjective

C. Categorizing Tweets

In this third step, we add a Categorize Operator which will basically classify our text based on a particular taxonomy, in this case we were using the IAB QAG taxonomy, which is a standard used in the digital advertising industry for categorizing content. An example result of the categorizing tweets presented in Table 3.

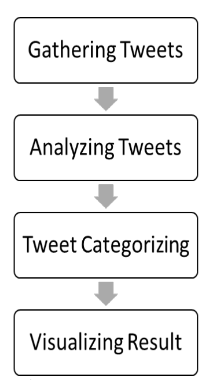

Figure 1. The Stages of Research

TABLE 3. EXAMPLE RESULT OF CATEGORIZING TWEETS

ID	Category	Category Score
7.66213638445 27309E17	Unmoderated UGC	0.20752050056838686
7.65994267848 37222E17	Copyright Infringement	0.3432388367947967
7.66337615184 42086E17	Weddings	0.2712489044013795
7.66545885081 67782E17	Extreme Graphic/Explicit Violence	0.19777008355614484
7.66545774087 83974E17	Options	0.18889991714373625

D. *Visualizing Result*

At this last stage, we will display the results of the phase that has been done before in the form of Pie Chart like on figure 2 below. Figure 2 shows that the result of sentiment analysis is done using the Tweet with the keyword AYLIEN WhatsApp discount negative sentiment that is greater than the positive sentiment.

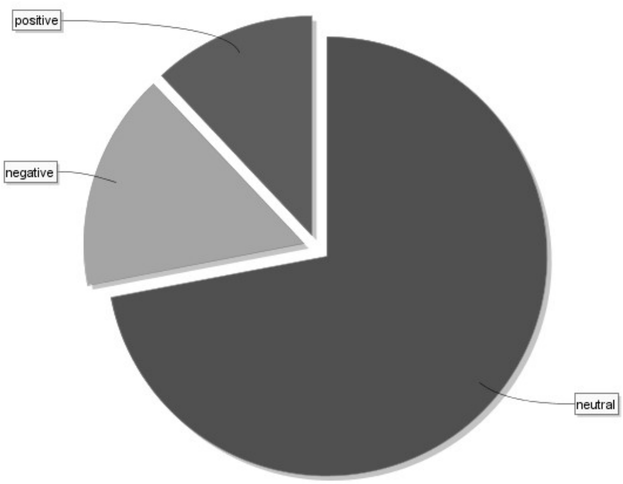

Figure 2. Visualizing The Sentiment Analysis of WhatsApp Result

The result of positive sentiment analysis means that in the tweets collected, many contain positive sentiment words. Conversely, if the results of sentiment analysis showed a negative result means that Tweet about the product contains many negative sentiments. However, if the results show neutral, it is possible that the tweet that discusses the product there is no positive or negative sentiment also has the possibility of tweet is not included in a sentiment or just an opinion.

IV. CONCLUSION

The use of different libraries get different results where Rosette and AYLIEN is a library that can be used on RapidMiner. Rosette will divide the results of the analysis into three categories sentiment and sentiment analysis results AYLIEN split into two categories. Differences in the number of categories of impact on the detection sentences containing sentiment or just opinions. An opinion can not be classified in user sentiment toward a product of messaging.

REFERENCES

[1] R. Dehkharghani and C. Yilmaz, “Automatically identifying a software product’s quality attributes through sentiment analysis of tweets,” in 2013 1st International Workshop on Natural Language Analysis in Software Engineering, NaturaLiSE 2013 - Proceedings, 2013, pp. 25–30.

[2] “AYLIEN | We Bring Intelligence to the Web,” AYLIEN. [Online]. Available: http://aylien.com. [Accessed: 18-Aug-2016].

[3] “Sentiment Analysis,” Rosette, 04-Feb-2016. [Online]. Available: https://www.rosette.com/function/sentiment-analysis/. [Accessed: 18-Aug-2016].

[4] D. Pagano and W. Maalej, “User feedback in the appstore: An empirical study,” in 2013 21st IEEE International Requirements Engineering Conference, RE 2013 - Proceedings, 2013, pp. 125–134.

[5] L. V. Galvis Carreño and K. Winbladh, “Analysis of user comments: an approach for software requirements evolution,” in Proceedings of the 2013 International Conference on Software Engineering, 2013, pp. 582–591.

[6] K. Schneider, “Focusing spontaneous feedback to support system evolution,” in 2011 IEEE 19th International Requirements Engineering Conference, 2011, pp. 165–174.

[7] E. Guzman, “Do Users Like This Feature? A Fine Grained Sentiment Analysis of App Reviews,” in IEEE 22nd International Requirements Engineering Conference (RE), 2015, pp. 153–162.

[8] M. Harman, Y. Jia, and Z. Yuanyuan, “App Store Mining and Analysis: MSR for App Stores,” in Proceedings of the 19th International Symposium on Software Testing and Analysis, Zurich, Switzerland, 2012, pp. 108–111.

[9] W. Maalej and D. Pagano, “On the Socialness of Software,” presented at the Ninth IEEE International Conference on Dependable, Autonomic and Secure Computing, Sydney, NSW, 2011, pp. 864–871.

[10] N. Seyff, F. Graf, and N. Maiden, “Using Mobile RE Tools to Give End-Users Their Own Voice,” presented at the 2010 18th IEEE International Requirements Engineering Conference, Sydney, NSW, 2010, pp. 37–46.

[11] T.-P. Liang, X. Li, C.-T. Yang, and M. Wang, “What in Consumer Reviews Affects the Sales of Mobile Apps: A Multifacet Sentiment Analysis Approach,” Int. J. Electron. Commer., vol. 20, no. 2, pp. 236–260, Dec. 2015.

[12] M. L. Richins and T. Root-Shaffer, “The role of evolvement and opinion leadership in consumer word-of-mouth: An implicit model made explicit,” NA-Adv. Consum. Res., vol. 15, 1988.

[13] S. M. Cunningham, “Perceived risk as a factor in the diffusion of new product information,” Raymond M Haas Ed, pp. 698–721, 1966.

[14] P. F. Bone, “Word-of-mouth effects on short-term and long-term product judgments,” J. Bus. Res., vol. 32, no. 3, pp. 213–223, 1995.

[15] C. Dellarocas, “The digitization of word of mouth: Promise and challenges of online feedback mechanisms,” Manag. Sci., vol. 49, no. 10, pp. 1407–1424, 2003.

[16] D. Godes, D. Mayzlin, Y. Chen, S. Das, C. Dellarocas, B. Pfeiffer, B. Libai, S. Sen, M. Shi, and P. Verlegh, “The firm’s management of social interactions,” Mark. Lett., vol. 16, no. 3–4, pp. 415–428, 2005.

[17] L. J. Harrison-Walker, “The measurement of word-of-mouth communication and an investigation of service quality and customer commitment as potential antecedents,” J. Serv. Res., vol. 4, no. 1, pp. 60–75, 2001.

[18] F. Zhu and X. Zhang, “Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics,” J. Mark., vol. 74, no. 2, pp. 133–148, 2010.

[19] J. Q. Zhang, G. Craciun, and D. Shin, “When does electronic word-of-mouth matter? A study of consumer product reviews,” J. Bus. Res., vol. 63, no. 12, pp. 1336–1341, 2010.

[20] D. Baum and M. Spann, “The interplay between online consumer reviews and recommender systems: An experimental analysis,” Int. J. Electron. Commer., vol. 19, no. 1, pp. 129–162, 2014.

[21] M. Li, L. Huang, C.-H. Tan, and K.-K. Wei, "Helpfulness of online product reviews as seen by consumers: Source and content features," Int. J. Electron. Commer., vol. 17, no. 4, pp. 101–136, 2013.

[22] M. Hofmann and R. Klinkenberg, RapidMiner: Data mining use cases and business analytics applications. CRC Press, 2013.