
COMP 8920 PROJECT PROPOSAL: WHATSAPP DATA ANALYSIS

Akshay Shah

shahlbz@uwindsor.ca

Srihari Jayachandran

jayachal@uwindsor.ca

Manil Patel

patel13h@uwindsor.ca

1 OVERVIEW

1.1 THE MOTIVATION OF THIS PROJECT

We felt the need to analyze the conversations between a group of people and how the people within a group feels and thinks about the opinions of others. Since most of the people, schools, and organizations have started using WhatsApp for communication, we decided to analyze the group chats in the app to better understand a group of people by incorporating both text and sentiment analysis.

1.2 SHORT DESCRIPTION

The sudden outburst of the Internet has made it easier for people around the world to connect with each other. Social network applications like Twitter, Facebook, WhatsApp, etc. have gained its importance. Huge amounts of data are generated each day in the form of texts, emojis, etc. This data abundance has attracted many researchers and organizations to analyze the communication between people(5). So, in this project, we have tried to implement both text and sentimental analysis to capture the intentions, emotions, and opinions among a group of people which on further research will be beneficial for companies and organizations to easily target/identify a group of people for their research or for selling their products, etc.

1.3 THE POTENTIAL CUSTOMERS

The customers of this project would be the people who are using WhatsApp on a daily basis for communication with their family, friends, colleagues, etc.

1.4 WHAT THE PROJECT WILL DELIVER. IS IT A NEW PRODUCT OR AN EXTENSION OF AN EXISTING ONE?

Our project is an extension of the existing one(7). The major difference would be the combination of both text and sentimental analysis to understand people in a better way. Our project will deliver the graphical visualizations for message count, emojis, frequency of the texts and emojis, lexical diversity, wordcloud and sentiment score(11).

1.5 APPLICATIONS

WhatsApp is considered to be one of the most popular applications for chatting purposes. In fact, The research found that it has monthly active users more than 700 million. Its popularity has made it as the most essential app among smartphone users to connect with their friends and family. Even

Companies and Organizations have started using WhatsApp for communication across various departments. Since we will be incorporating both text and sentimental analysis into the WhatsApp data, it will be of use for educational institutions, companies, organizations, etc. to understand and gain insights about the people with whom they are having a conversation. The text and sentimental analysis also help to avoid conflicts that may arise due to misunderstanding(10).

1.6 METHODOLOGY

To begin the analysis work, data is required. So, we are extracting the chat history from WhatsApp by exporting the data to our mail. Due to the end-to-end encryption feature of WhatsApp, the chat histories are exported to our mail. After getting the required amount of data, we are loading the data into the R-programming platform for cleaning the data. The cleaning process involves the removal of numbers, punctuation, special characters, etc. and stemming the words(10). After the cleaning phase, we implement a Naive Bayes classifier approach for text classification and the Natural language processing approach for sentimental analysis. Finally, with the help of inbuilt functions and libraries, we build the visualizations of our analyzed data(7).

1.7 LITERATURE REVIEW

In this literature review, we are going to explain some prior work on sentiment analysis and text analysis. Sentimental Analysis is a strategy to explore whether a gathered content is in a positive, negative or neutral state(2). In general, there are two main approaches to perform sentiment analysis one is a machine learning technique and another is the lexical based approach(2). Our primary focus on the machine learning approach because it involves learning. Nowadays, there are various applications or techniques available for sentiment analysis on twitter dataset(1). In this paper(1), the author uses an open-source approach which includes the collection of tweets from Twitter API and then it requires some pre-processing, then it analyzes data and visualizes the data. The above project is able to distinguish eight distinct classifications of feeling (disgust, fear, anger, anticipation, sadness, trust, surprise) and two unique sentiments (positive and negative)(1).

We are going to develop a Whatsapp text analysis because it's the most popular messaging service. In January 2015, WhatsApp was the most globally popular messaging app with more than 600 million active users(3). In April 2015, WhatsApp reached 800 million active users(3). In order to get chat data, we will use the export chat method. As chat histories are stored in an encrypted database on the mobile device and messages are transmitted over the Internet with end-to-end encryption, this is currently the only option to access the chat data(4). This project(4) involves data suffix to analyze the WhatsApp chats properties. In addition to that, it also provides a little visualization of data(4).

2 PROJECT GOAL

The main objectives of the project are to capture the intent, behaviour, opinion, and feelings of people. In order to meet our project goals, we will be using both text/sentimental analysis and will be dividing into specific modules to carry out the analysis processes. The following are the modules:

1. Message count module: This will contain the number of messages that were recorded in a day and the number of messages created by each person, along with their visualizations.
2. Emoji module: This will contain the emojis used by each person along with its visualization.
3. Frequency module: This will contain the most often used words and emojis with its visualization.
4. Lexical diversity module: This will contain the count of the unique words used by each person and its visualization.
5. Sentiment module: This will contain wordcloud and sentiment score of the text messages with its visualization. These are the different modules that we will be implementing to meet our objectives(11).

3 PROJECT TEAM

Name	Responsibility	Availability (Hours for week)
Akshay Shah	Literature Review, Text Analysis, Documentation	4
Manil Patel	Schedule & Milestones, Data Collection, Lexical Diversity	4
Srihari Jayachandran	Proposal Documentation, Sentimental analysis, Presentation	4

Table 1 Project Team

4 SCHEDULE AND MILESTONES

Milestones	Description	Milestones Criteria	Planned Date
MO	Start Project+	Topic Identification	2020-01-24
M1	Project Proposal	Identify Project Goals, Requirement, and Challenges	2020-01-31
M2	Project Work Phase 1:	Extract the Data and Download tools	2020-02-12
M3	Project Work Phase 2:	Start learning language and perform an operation on Data	2020-02-28
M4	Project Work Phase 3:	Start working on Text Analysis	2020-03-03
M5	Project Work Phase 4:	Start working on Sentimental Analysis	2020-03-15
M6	Documentation	Prepare a Final Report and PPT on Project	2020-03-28

Table 2 Schedule and Milestones

5 COMMUNICATION AND REPORTING

Communication is a key element of any successful project. In this project, we plan to organize a weekly person meetup to discuss the project in a discussion room of Leddy Library. In addition to that, we are also planning to meet the professor in their office hour regarding further improvement in the project. For the documentation part, we're going to use Microsoft Office 365 and Overleaf Latex Text editor. Moreover, we will use GitHub, uWindsor mail and WhatsApp to share content and code as well as to communicate.

6 DELIVERY PLAN

Social Media provides humongous data for a data scientist to perform various algorithms on data to predict the output. Here, we are going to present a console application based on Text and Sentiment analysis using which we can identify a person's behavior, emotion, intent, etc. We decided to use WhatsApp group chat as a data set to analyze a group of people due to easy retrieval and an abundance of data. Also, we are going to deliver several models with visualization which include Message count, Emoji module, Frequency module, Lexical diversity, and Sentiment module. We are planning to present this project by April first week.

REFERENCES

- [1] Saini, R. Punhani, R. Bathla and V. K. Shukla, "Sentiment Analysis on Twitter Data using R," 2019 International Conference on Automation, Computational and Technology Management (ICACTM), London, United Kingdom, 2019, pp. 68-72. doi: 10.1109/ICACTM.2019.8776685
- [2] Walaa Medhat, Ahmed Hassan, Hoda Korashy, Sentiment analysis algorithms and applications: A survey, Ain Shams Engineering Journal, Volume 5, Issue 4,

-
- 2014, Pages 1093-1113, ISSN 2090-4479, <https://doi.org/10.1016/j.asej.2014.04.011>. (<http://www.sciencedirect.com/science/article/pii/S2090447914000550>)
- [3] A.Al-Omary, W. M. El-Medany and K. J. E. Isa, "The Impact of SNS in Higher Education: A Case Study of Using WhatsApp in the University of Bahrain,"2015 Fifth International Conference on e-Learning (econf), Manama, 2015, pp. 296-300. doi: 10.1109/ECONF.2015.72
URL:<http://ieeexplore.ieee.org.ledproxy2.uwindsor.ca/stamp/stamp.jsp?tp=&arnumber=7478249&isnumber=7478193>
- [4] Schwind and M. Seufert, "WhatsAnalyzer: A Tool for Collecting and Analyzing WhatsApp Mobile Messaging Communication Data,"2018 30th International Teletraffic Congress (ITC 30), Vienna, 2018, pp. 85-88. doi: 10.1109/ITC30.2018.00020
URL:<http://ieeexplore.ieee.org.ledproxy2.uwindsor.ca/stamp/stamp.jsp?tp=/&arnumber=8493058&isnumber=8493038>
- [5] Kaur, R. (2019). Insight to Emotional tones in WhatsApp Through Sentiment Analysis. IJRAR.
- [6] Frey, L., Botan, C., Kreps, G. (1999). Investigating communication: An introduction to research methods. (2nd ed.) Boston: Allyn Bacon
- [7] Rosenfeld, Avi & Sina, Sigal & Sarne, David Avidov, Or Kraus, Sarit. (2016). WhatsApp Usage Patterns and Prediction Models. ICWSM/IUSSP Workshop on Social Media and Demographic Research.
- [8] Kumar, Naveen & Sharma, Sudhansh. (2017). Survey Analysis on the usage and Impact of Whatsapp Messenger. Global Journal of Enterprise Information System. 8. 52. 10.18311/gjeis/2016/15741.
- [9] B.R, "Text Analysis for WhatsApp (Sentimental Analysis)," 5 December 2019. [Online]. Available: <https://medium.com/@karurbalamathiprojects/text-analysis-for-whatsapp-sentimental-analysis-4963fca574b6>.
- [10] "WhatsApp Chat Sentiment Analysis in R," [Online]. Available: <http://www.planetanalytics.in/2017/09/whatsapp-chat-sentiment-analysis-in-r.html>.
- [11] "Text Analysis using WhatsApp data," [Online]. Available: https://cran.r-project.org/web/packages/rwhatsapp/vignettes/Text_Analysis_using_WhatsApp_data.html.