



# **Contact Tracing Of Covid-19 Patients Using Tweets**

**K D S Perera**

MS20902834

M.Sc. in IT

Specialized in Information Technology

Supervisor: Dr Jeewanee Bamunusinghe

November 2021

18672

**Department of Information Technology  
Faculty of Graduate Studies and Research  
Sri Lanka Institute of Information Technology**

# Table of Contents

Table of Contents.....	2
List of Figures .....	4
List of Tables .....	5
Abstract.....	6
Declaration.....	7
Acknowledgement .....	8
Chapter 1 Introduction .....	9
1.1 Why corona contact tracing is important .....	9
1.2 Social Media .....	11
1.3 Problem Statement.....	14
1.4 Main Objective .....	15
1.5 Sub Objective .....	15
1.6 Research Questions .....	16
1.7 Outline of the thesis.....	16
1.7.1 Chapter 1.....	16
1.7.2 Chapter 2.....	16
1.7.3 Chapter 3.....	16
1.7.4 Chapter 4.....	17
1.7.5 Chapter 5.....	17
1.7.6 Chapter 6.....	17
1.7.7 Chapter 7.....	17
1.7.1 Chapter 8.....	17
Chapter 2 Literature Survey .....	18
2.1 Tweets hashtag suggestions and linking.....	18
2.2 Influence detection using tweets.....	21
2.3 Mining Twitter Feeds for Information .....	23
2.4 Misleading covid19 information detection in Tweeter .....	25
2.5 Use Tweets to Track Relevant Topics.....	27
2.6 Word Recognition and Tweet Segmentation for User Tracking .....	29
2.7 COVID-19 Sentiment Analysis in Specific Tweets .....	30
2.8 Machine Learning approach for Twitter Analysis to Prevent Spreading Covid-19.....	32
Chapter 3 Methodology.....	34
3.1 Proposed System.....	35
3.2 Twitter Data .....	36
	2

3.2.1 Twitter API Authentication .....	37
3.2.2 Twitter API Authentication – OAuth 1.0 .....	37
3.2.3 Twitter API Authentication – OAuth 2.0 Bearer Token.....	39
3.2.4 Twitter API Authentication – Basic Authentication .....	39
3.2.5 API .....	39
3.2.6 Twitter ID's .....	40
3.2.7 Twitter Security .....	40
3.2.8 Twitter Search API.....	42
3.3 Data Pre-Processing .....	43
3.3.1 Twitter scraping .....	43
3.4 Sentimental Analysis .....	48
3.4.1 Sentimental Analysis for preprocessed data .....	50
3.5 Individual tweets.....	57
3.6 Natural Language Processing .....	58
3.6.1 Tools and approaches .....	59
3.6.2 NLP use cases .....	60
3.6.3 Name Entity Recognition .....	61
3.6.4 Name Entity Recognition for the twitter dataset .....	62
Chapter 4 Results and Discussion .....	65
Chapter 5 Conclusion and Future Work.....	66
5.1 Conclusion.....	66
5.2 Future work.....	67
Chapter 6 Feasibility.....	68
Chapter 7 References.....	69
Chapter 8 Appendix .....	71

## List of Figures

Figure 1: Survey report of social media .....	12
Figure 2: Twitter Usage .....	13
Figure 3: Number of bugs, requirement, and miscellaneous tweets .....	24
Figure 4: Proposed Framework Misleading vocid19 information diction .....	26
Figure 5: Proposed System Flow .....	36
Figure 6: API Flow .....	39
Figure 7: Raw tweet .....	46
Figure 8: Plot data to check null fields .....	47
Figure 9: Tweets Source .....	51
Figure 10: Cleaned tweets .....	51
Figure 11: Sentimental analysis labeling .....	52
Figure 12: Sentiment Data into a plot .....	52
Figure 13: Most Used keywords .....	53
Figure 14: Positive word count .....	53
Figure 15: Negative word Count .....	54
Figure 16: Neutral word count .....	54
Figure 17: Overall word count .....	55
Figure 18: Positive Users .....	55
Figure 19: Positive User View .....	56
Figure 20: Data Extraction of Each Tweet .....	57
Figure 21: Processing Steps .....	60
Figure 23: Key words extraction .....	62
Figure 24: Extract Information from tweets .....	63
Figure 25: Network diagram of close contacts .....	64

## List of Tables

Table 1 Propose framework experiment results 1 and result 2 .....	19
Table 2 Proposed framework results with experiment 4 and 5 .....	20
Table 3 Modified Models Result's Comparison .....	26

## Abstract

COVID-19 has had an impact on everyone's lives. In recent years, In the field of healthcare, the use of social media as a tool for professional education & communication has expanded; The benefits and drawbacks of these networks have been widely discussed with different results. During COVID-19 disease outbreak usage of social media has exploded to the point where it has become an essential part of modern healthcare systems. It is simple and innovative to create content for Twitter and other social media sites. Such data has become strategically relevant for companies seeking input on their products, brand endorsements, and merchandising, among other things. People have started to have mixed feelings about the situation due to the rapid rise in infection and death rate. Contact with an infectious person who sneezes or speaks less than one meter away can spread COVID-19. Keeping a safe distance from people who have Covid 19 is one way to avoid transmission. The problem is that many sufferers do not exhibit any symptoms, which means that people must be more aware of those they come into contact with physically.

Social networking divides people into various groups, such as rural areas or residential subdivisions. Because it's most generally connected with the internet, communication can also take place in person, which is a little-known reality; this provides the idea that we spend a substantial amount of time in front of computers. Today's youth are becoming more reliant on their cell phones and social media sites to stay 'up to date. Social networking is one of the media that can be used to socialize. One of the most widely used media for disseminating information about the Covid-19 case, as well as appeals to stay at home and maintain a healthy distance in order to avoid contracting the coronavirus, is Twitter. Despite the fact that socialization has taken place, many people remain dissatisfied with the government's initiative, making it difficult to control the spread of Covid-19. As a result, it's necessary to compare the answer to the stay-at-home campaign with the spread of Covid-19.

Due to the void-19 pandemic, the biggest problem that the government is facing is identifying the first contacts of the covid-19 patients. Most of the people are avoiding telling the truth for the government. But people use social media day to day life. Most of the people post/ share their life experience. This is a common way for Twitter users to explain their day to day disease development after the first set of symptoms. This pattern may have gained traction when celebrities began to tweet about their symptoms.

## Declaration

I hereby declare that this thesis represents my own work which has been done after registration for the masters of science in information technology (M.Sc. in IT) at Sri Lankan Institute of Information Technology (SLIIT), and has not been previously included in a thesis or dissertation submitted to this or any other institution for a degree, diploma or other qualifications.

I have attempted to identify all the risks related to this research that may arise in conducting this research, obtained the relevant ethical and/or safety approval (where applicable), and acknowledged my obligations and the rights of the participants.

Signature:

Date:

## **Acknowledgement**

I shall begin by first acknowledging my thesis supervisor Dr. Jeewanee Bamunusinghe. She approved me with constant support and guidance whenever I ran into queries, and always welcomed any problems or questions that I had regarding the research and writing process of my thesis. Without her guided correct direction, my thesis would not have taken this shape.

Finally, I would convey my heartfelt thankfulness for the constant support and encouragement that I received from my parents and my sister for the consistent support throughout the years of my study as well as when developing my thesis. The process of writing my thesis would not have been enjoyable without their unconditional support and love.

Thank you,

Damitha Perera