

PROFESSIONAL TRAINING REPORT

at

Sathyabama Institute of Science and Technology

(Deemed to be University)

Submitted in partial fulfillment of the requirements for the award

of Bachelor of Technology Degree in

Information Technology

By

K.NITHISH

Reg.No. 38120038



DEPARTMENT OF INFORMATION TECHNOLOGY

SCHOOL OF COMPUTING

SATHYABAMA

**INSTITUTE OF SCIENCE AND
TECHNOLOGY (DEEMED TO BE
UNIVERSITY)**

Accredited with Grade “A” by NAAC

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600119

NOVEMBER 2021



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with "A" Grade by NAAC

Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai - 600 119.

Phone: 044 - 2450 3150 / 51 / 52 / 54 / 55 Fax: 044 - 2450 2344

www.sathyabama.ac.in



SCHOOL OF COMPUTING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **K.NITHISH(38120038)** who carried out the project entitled "**SENTIMENT ANALYSIS OF LOCKDOWN IN INDIA DURING COVID 19**" under our supervision from June 2021 to November 2021.

INTERNAL GUIDE

Dr.G.MATHIVANAN, M.E., Ph.D.,

HEAD OF THE DEPARTMENT

Dr. SUBHASHINI, M.E., Ph.D.,

Submitted for Viva voce Examination held on _____

Internal Examiner

External Examiner

DECLARATION

I **K.NITHISH** hereby declare that the Project Report entitled “**SENTIMENT ANALYSIS OF LOCKDOWN IN INDIA DURING COVID 19**” done by me under the guidance of **Dr.G.MATHIVANAN, M.E.,Ph.D.**, is submitted in partial fulfillment of the requirements for the award of Bachelor of Technology degree in Information Technology.

DATE:

PLACE:

SIGNATURE OF THECANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management of SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey our thanks to **Dr. T. SASIKALA M.E., Ph.D.**, Dean, School of Computing and **Dr. R.SUBHASHINI, M.E.,Ph.D.**, Head of the Department, Department of Computer Science and Engineering for providing us the necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to our Project Guide **Dr.G.MATHIVANAN, M.E.,Ph.D.**, for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the Department of **INFORMATION TECHNOLOGY** who were helpful in many ways for the completion of the project

SENTIMENT ANALYSIS OF LOCKDOWN IN INDIA DURING COVID-19: A CASE STUDY ON TWITTER

ABSTRACT

Corona Virus or COVID-19 first appeared in December, 2019 in Wuhan, China. People tweeted aggressively on twitter at that time. This paper analysed the tweets regarding COVID-19 from November, 2019 to May, 2020 in India and its affect. All tweets are categorized into 3 categories (Positive, Negative and Neutral). Multiple datasets are created State-wise, Month-wise, combined of all states to analyze the people's reactions towards Lockdown in June, 2020 and about everything related to COVID-19. Most people started having Negative tweets but with increasing time people shifted towards positive and neutral comments. In April, 2020 most comments were Positive and about winning against Corona virus.

CHAPTER No.	TITLE	PAGE No
	ABSTRACT	i
	LIST OF FIGURES	ii
1	INTRODUCTION	1
	1.1 INTRODUCTION TO PYTHON	1
	1.2 LITERATURE REVIEW	2
2.	AIM AND SCOPE OF THE PRESENT INVESTIGATION	3
	2.1 OUTLINE OF THE PROJECT	3
	2.2 PROBLEM STATEMENT	3
3.	EXPERIMENTAL OR MATERIALS AND METHODS; ALGORITHMS USED	5
	3.1 GENERAL	5
	3.2 OVERVIEW	5
	3.3 MODULE DESIGN	7
4.	RESULTS AND DISCUSSION, PERFORMANCE ANALYSIS	10
	4.1 RESULTS	10
5	SUMMARY AND CONCLUSIONS	11
	REFERENCES	12
	APPENDIX	13
	A.SCREENSHOTS	13
	B. SOURCE CODE	15

CHAPTER 1

INTRODUCTION

Corona viruses are a gathering of related RNA infections that cause illnesses in warm-blooded animals and feathered creatures. In people, these infections cause respiratory plot contamination that can go from mellow to deadly. Mellow diseases incorporate a few instances of the basic cold (which is additionally brought about by different infections, overwhelmingly rhinoviruses), while progressively deadly assortments can cause SARS, MERS, and COVID-19. Side effects in different species change: in chickens, they cause an upper respiratory plot illness, while in dairy animals and pigs they cause loose bowels. Antibodies or antiviral medications to forestall or treat human coronavirus diseases are not found yet. Coronavirus spreads from humans to humans in several manners. It causes respiratory problems and causes breathing problems. It has a low fatality rate compared to SARS or MERS. However, as there is no vaccine available and due to the nature of transmission and spread of virus, countries around the world taken lockdown and isolation as the only preventive steps. As a result people's movement being restricted they spent a good amount of time in home or place of stay. This creates a ideal scenario for people to express their views on social networking sites including Twitter. Like in other parts of the world people of India also express their views about COVID-19 on twitter. In this paper, those tweets are used to analyze the behavior and sentiment of people from December 2019 to May 2020. People were very negative about COVID in the early days. Many were not confident to win the fight against COVID. But as the Lockdown happened, people became optimistic about it. Positive tweets are tremendously high in the amount in the month of April and May. This work provides new insights on COVID-19 and people's thoughts about it. Social media is today's way to read people. There is a good number of papers appeared in the literature since the outbreak causes countries to go in Lockdown state. Sentiment analysis and data visualization of World wide Twitter data on Covid 19 has been presented.

LITERATURE REVIEW

Sentiment analysis from social media data is one of the highly emerging research fields. It could play a critical role in case of medical emergencies like the COVID-19 pandemic, and hence it is more crucial. Though a lot of research from various angles on sentiment classification and NLP is still in progress, some of the completed works are as follows.

Wu et al. used data from December 31, 2019, to January 28, 2020, on the count of infectious persons exported from Wuhan to deduct the number of cases in Wuhan from December 1, 2019, to January 25, 2020. Cases exported intra-country were then predicted. They predicted the COVID cases across the country by using the flight booking data and COVID positive persons travelled through flight and thus prognosticated the national and international spread of COVID-19 after calculating the impact of the metropolitanwide quarantine of Wuhan and surrounding cities which was started in China from Jan 23–24.

Medford et al. Fetched tweets related to COVID-19 and calculated the frequency of keywords subject to infection prevention practices, inoculation, and racial partiality. They executed a sentiment analysis to observe sentimental valence and dominant emotions. They performed topic modeling to extract and explore hot discussion topics over time. They extracted 126 049 tweets posted by 53 196 different users. The frequency of COVID-19-related tweets suddenly increased from January 21, 2020 onward. The sentiments are classified from approximately half (49.5%) of all posts expressed fear to about 30% expressed surprise. The number of racial posts closely matched the frequency of new cases of COVID-19 positives. The financial and political effects of the COVID-19 were the most common topics in discussion.

Li et al. collected and analyzed the Weibo posts from 17 865 active Weibo users using online ecological recognition (OER) based on some machine learning prognostic models. They evaluated word frequency, sentiment indicator scores

(e.g., depression, anxiety, indignation, and happiness), and cognitive indicators (e.g., social risk assessment, and life gratification) from the extracted posts. To evaluate the differences in the same group they have executed the opinion mining and the paired sample t-test before and after the confirmation of COVID-19 on January 20, 2020. The results evaluated by them reflect that the sensitivity to social risks and the negative sentiments increased while the scores of positive sentiments and life gratification have been decreased.

Pandey et al. addressed the gap between the information and risk of misinformation by developing a life-long learning model that provides genuine information in Hindi, the most commonly used local language in India. They matched the sources of authentic and genuine info, such as the news provided by WHO by using machine learning and NLP. They observed a Cohen's Kappa of 0.54 with the best performing combination and was deployed in their application.

Kayes et al. collected 100 000 tweets with the keyword #coronavirus within Australia. Among these 1 lakh tweets, 3076 contain the keyword "social distancing" and #socialdistancing. They use 8000 tweets for training & validation and 2000 tweets for testing the model. They achieved an accuracy of 83.70% and an F1-Score of 81.62% on the test data. They applied the trained model on the 3076 tweets that contain the keyword "social distancing." They observed that 80% plus tweets discussed "social distancing" have a positive opinion, as illustrated. They concluded that people in Australia supported social distancing as well as accepted it.

Alhajji et al. analyzed a total of 53 127 tweets from the Saudi citizens regarding COVID-19 and found that the numbers of positive tweets are greater than negative tweets for almost all the measures. They found that the most positive sentiments were present in the religious practices-related steps. They concluded that Saudi Twitter users have positive sentiments and support toward the infection control steps in fighting with COVID-19, and this supportive attitude of Saudi citizens results in the overarching confidence of the Saudi government. According to them, at times of pandemic, religious beliefs may also play an important role in preparing

believers. They collected separate tweets on various steps taken by the Saudi government. They extracted 9924 tweets after the announcement of the Grand Mosque closure and found that 76.72% of the tweets were positive. Similarly, they collected tweets for Qatif lockdown, closure of schools and universities, shopping malls, parks, and restaurant closure measures, sports competition suspension hashtags, for the congregational and weekly Friday prayers suspension measure, and finally for nationwide curfew measure.

Abd-Alrazaq et al. included approximately 2.8 million tweets for their study. Out of them, 167 073 tweets from 160 829 different users met the inclusion criteria. They analyzed the tweets on 12 topics and then grouped them into four main themes: sources of the virus; its origin; its effect on population, countries, and the economy; and methods of extenuating the danger of infection. They found the positive mean sentiment for all except two topics: one is causalities caused by COVID-19, and second is increased racism. They observed the minimum mean of 2722 tweets for increased racism and a maximum of 13 413 for economic losses. They also found the highest mean for likes of 15.4 for economic losses and the lowest for travel bans and warnings of 3.94.

Pota et al. applied the sentiment analysis on political tweets using a neural-network-based approach. They used the SemEval-2017 data set, which contains 20 633 tweets for training and 12 284 tweets for testing the trained model. Their method represents the text by dense vectors, including information of subwords to observe similarities between words by using morphology and semantics both. Then, they used the convolutional neural network technique and trained the model based on the labeled data set. They applied that model on a collection of tweets gathered during the tenure of some days before U.K. General Elections. According to observed results, they concluded that the CNN approach is better as compared to lexicon-based approaches for classification of sentences into positive and negative polarities.

Luo used a susceptible-infected-recovered (SIR) model to predict the life cycle of COVID-19 in many countries and the world. To predict the COVID-19 life cycle, he

took daily updated data about COVID-19 from “Our World in Data” website and retrogress the mathematical SIR model using publically available codes from the Milan Batista website. He ran regression for countries individually and modified it daily with the newer data. The model then formed is used to predict the life cycle of a full pandemic and then construct the life cycle curve. He fitted the data to plot an initial segment of the curve, and the remaining segment is estimated. He predicted 97%, 99%, and completely ending dates of individual countries and the whole world. They concluded that this pandemic might end by the end of November 2020 from the world.

Tomar et al. performed opinion mining on GST using data from Twitter to find out the public opinion about GST. They construct the data set by manually annotating 1000 tweets plus 2000 reviews from the Internet Movie DataBase (IMDB). They train and test the classifier by applying the data set in three different ways. Model-1 used the IMDB data set for training and testing with the K-fold validation process using $K = 10$, Model-2 used IMDB data set for training and manually annotated data set for testing. Lastly, Model-3 used a full data set for training and testing with the K-fold validation process using $K = 10$. They achieved an accuracy of 74.75%, 67.1%, and 72.3% for Model-1, Model-2, and Model-3, respectively. Observing the results, they stated that if a training data set is constructed using different sources instead of single-source, better accuracy can be achieved.

Ding et al. constructed an entity-level opinion analysis tool named SentiSW for issue comments consisting of sentiment classification and entity recognition. The objective of the developed tool is to classify issued comments into three polarities: positive, negative, and neutral, and to recognize the entity of comments. They build data set manually by annotating 3000 comments selected from 231 732 issued comments taken from different GitHub projects. They used a tenfold cross-validation technique to evaluate the SentiSW tool and received 63.98% average recall, 68.71% average precision, and 77.19% accuracy. They manually label 660 comments by “Person” and “Project” entity to evaluate entity recognition and achieved an overall 88.73% recall, 76.58% precision, and 75.15% accuracy.

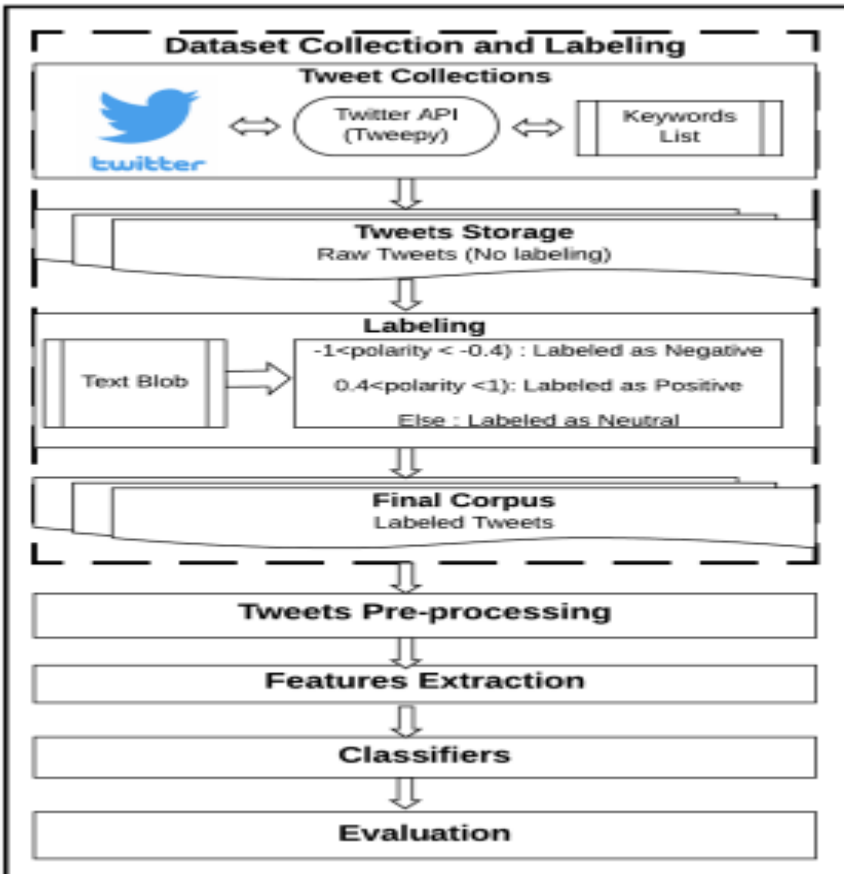
EXISTING SYSTEM

Burnap et al. constructed models to estimate the information flow size and survival using data retrieved from the famous microblogging site Twitter by following the terrorist event in Woolwich, London in 2013. They explained data flow as the propagation over time of info posted to Twitter via retweets. They used zero-truncated negative binomial and Cox proportional hazards regression methods to calculate the estimated value of social, content, and temporal elements of the tweet.

PROPOSED SYSTEM

This study also analyzes views concerning COVID-19 by focusing on people who interact and share social media on Twitter. As a platform for our experiments, we present a new large-scale sentiment data set COVIDSENTI, which consists of 90 000 COVID-19-related tweets collected in the early stages of the pandemic, from February to March 2020. The tweets have been labeled into positive, negative, and neutral sentiment classes. We analyzed the collected tweets for sentiment classification using different sets of features and classifiers. Negative opinion played an important role in conditioning public sentiment, for instance, we observed that people favored lockdown earlier in the pandemic; however, as expected, sentiment shifted by midMarch. Our study supports the view that there is a need to develop a proactive and agile public health presence to combat the spread of negative sentiment on social media following a pandemic.

SYSTEM ARCHITECTURE



SYSTEM REQUIREMENTS:

HARDWARE REQUIREMENTS:

- System - Pentium-IV
- Speed - 2.4GHZ
- Hard disk - 40GB
- RAM - 512MB

SOFTWARE REQUIREMENTS:

- Operating System - Windows XP
- Coding language - Java
- IDE - NetBeans

INTRODUCTION

Java is one of the world's most important and widely used computer languages, and it has held this distinction for many years. Unlike some other computer languages whose influence has weared with passage of time, while Java's has grown.

APPLICATION OF JAVA

Java is widely used in every corner of world and of human life. Java is not only used in softwares but is also widely used in designing hardware controlling software components. There are more than 930 million JRE downloads each year and 3 billion mobile phones run java.

Following are some other usage of Java:

1. Developing Desktop Applications
2. Web Applications like Linkedin.com, Snapdeal.com etc
3. Mobile Operating System like Android
4. Embedded Systems
5. Robotics and games etc.

FEATURES OF JAVA

The prime reason behind creation of Java was to bring portability and security feature into a computer language. Beside these two major features, there were many other features that played an important role in moulding out the final form of this outstanding language. Those features are;

1) Simple

Java is easy to learn and its syntax is quite simple, clean and easy to understand. The confusing and ambiguous concepts of C++ are either left out in Java or they have been re-implemented in a cleaner way.

Eg: Pointers and Operator Overloading are not there in java but were an important part of C++.

2) Object Oriented

In java everything is Object which has some data and behaviour. Java can be easily extended as it is based on Object Model.

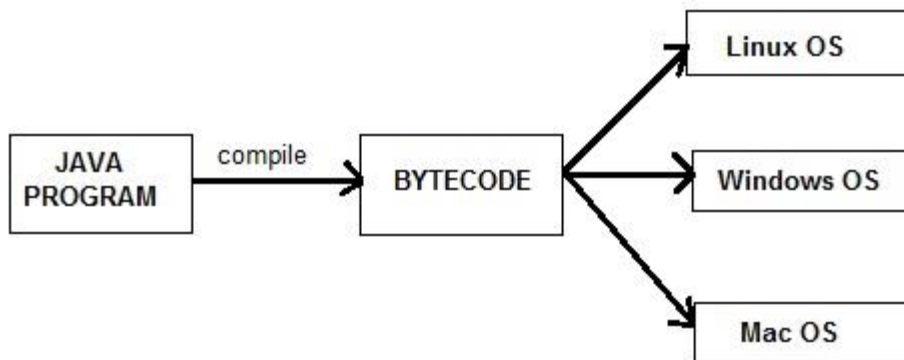
3) Robust

Java makes an effort to eliminate error prone codes by emphasizing mainly on compile time error checking and runtime checking. But the main areas which Java improved were Memory Management and mishandled Exceptions by introducing automatic Garbage Collector and Exception Handling.

4) Platform Independent

Unlike other programming languages such as C, C++ etc. which are compiled into platform specific machines. Java is guaranteed to be write-once, run-anywhere language.

On compilation Java program is compiled into byte code. This byte code is platform independent and can be run on any machine, plus this byte code format also provide security. Any machine with Java Runtime Environment can run Java Programs.



5) Secure

When it comes to security, Java is always the first choice. With java secure features it enable us to develop virus free, temper free system. Java program always runs in Java runtime environment with almost null interaction with system OS, hence it is more secure.

6) Multi-Threading

Java multithreading feature makes it possible to write program that can do many tasks simultaneously. Benefit of multithreading is that it utilizes same memory and other resources to execute multiple threads at the same time, like While typing, grammatical errors are checked along.

7) Architectural Neutral

Compiler generates byte codes, which have nothing to do with a particular computer architecture, hence a Java program is easy to interpret on any machine.

8) Portable

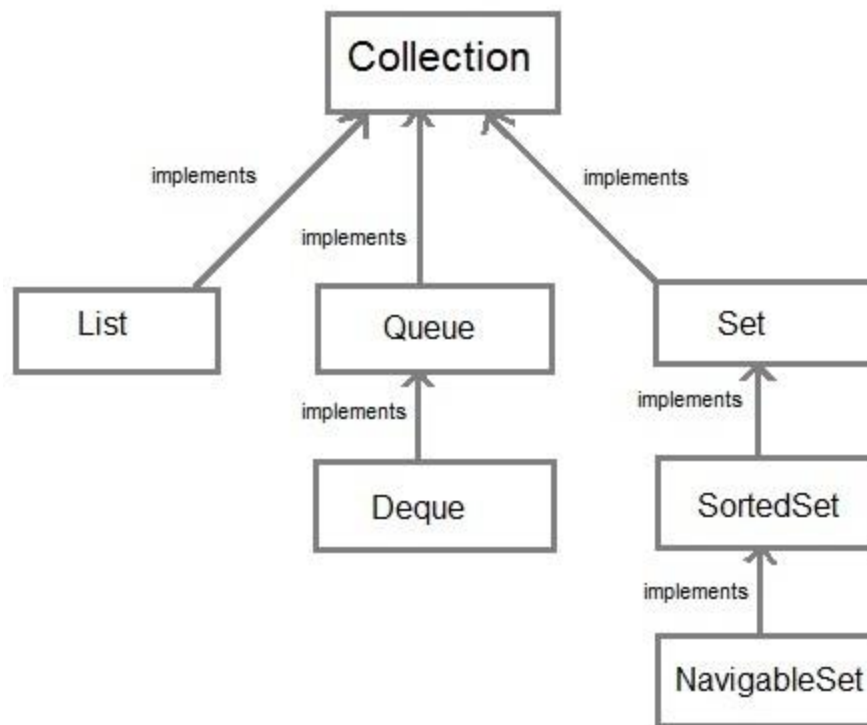
Java Byte code can be carried to any platform. No implementation dependent features. Everything related to storage is predefined, example: size of primitive data types

10) High Performance

Java is an interpreted language, so it will never be as fast as a compiled language like C or C++. But, Java enables high performance with the use of just-in-time compiler.

COLLECTION FRAMEWORK

Collection framework was not part of original Java release. Collections was added to J2SE 1.2. Prior to Java 2, Java provided adhoc classes such as Dictionary, Vector, Stack and Properties to store and manipulate groups of objects. Collection framework provides many important classes and interfaces to collect and organize group of alike objects.



MYSQL

MySQL, officially, but also called "My Sequel" is the world's most widely used open-source relational database management system (RDBMS) that runs as a server providing multi-user access to a number of databases, though SQLite probably has more total embedded deployments. The SQL phrase stands for Structured Query Language.

The MySQL development project has made its source code available under the terms of the GNU General Public License, as well as under a variety of

proprietary agreements. MySQL was owned and sponsored by a single for-profit firm, the Swedish company MySQL AB, now owned by Oracle Corporation. MySQL is a popular choice of database for use in web applications, and is a central component of the widely used LAMP open source web application software stack (and other 'AMP' stacks).

LAMP is an acronym for "Linux, Apache, MySQL, Perl/PHP/Python." Freesoftware-open source projects that require a full-featured database management system often use MySQL.

For commercial use, several paid editions are available, and offer additional functionality. Applications which use MySQL databases include: TYPO3, MODx, Joomla, WordPress, phpBB, MyBB, Drupal and other software. MySQL is also used in many high-profile, large-scale websites, including Wikipedia, Google (though not for searches), Facebook, Twitter, Flickr and YouTube.

FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

The feasibility study investigates the problem and the information needs of the stakeholders. It seeks to determine the resources required to provide an information systems solution, the cost and benefits of such a solution, and the feasibility of such a solution.

The goal of the feasibility study is to consider alternative information systems solutions, evaluate their feasibility, and propose the alternative most suitable to the organization. The feasibility of a proposed solution is evaluated in terms of its components.

ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system

efficiently. The user must not feel threatened by the system, instead must accept it as a necessity.

MODULES

- Data Collection Module
- Pre-Processing Module
- Feature Extraction Module
- Evaluation Module

MODULES DESCRIPTION

Data collection Module

Out of million tweets crawled from February to March 2020, we included 90 000 unique tweets from 70 000 users that met the selection criteria. Our analysis identified some topics, such as quarantine, lockdown, and stay home. COVIDSENTI was divided equally into three subdata sets of COVIDSENTI-A, COVIDSENTI-B, and COVIDSENTI-C, respectively, of positive, negative, or neutral sentiments. The data set (tweets) has been collected by using Tweepy.

Pre-Processing Module

As raw tweets are often short, unstructured, informal, and noisy, the first step of sentiment analysis is to preprocess the data. To do so, the following series of techniques are applied in the given order to improve the text.

1) Almost every social media platform uses hashtags to represent topics, i.e., #COVID-19, #StayHome, #StaySafe, and #Coronavirus. In most cases, hashtags are unnecessary to sentiment and can affect the performance. Thus, in our first step, we performed basic cleaning of the text by removing unnecessary hashtags, just the hashtag character not the hashtag text.

2) The second step is to case-fold the text. To avoid recognizing the same word as a different word due to capitalization, we fold all capitalized letters to lower case.

Feature Extraction Module

In this experiment, vectorization techniques and word embeddings are used for feature extraction. Term frequency-inverse document frequency (TF-IDF) has been used for vectorization. Similarly, for word embeddings, pretrained Word2Vec, GloVe, and fastText embeddings trained on Common Crawl and Wikipedia are used and have 300-D vectors. In addition, we used hybrid models, such as hybrid ranking (HyRank) and Improved Word Vector (IWV) that incorporate sentiment and context of tweets for Twitter sentiment analysis.

Evaluation Module

This study also analyzes views concerning COVID-19 by focusing on people who interact and share social media on Twitter. As a platform for our experiments, we present a new large-scale sentiment data set COVIDSENTI, which consists of 90 000 COVID-19-related tweets collected in the early stages of the pandemic, from February to March 2020. The tweets have been labeled into positive, negative, and neutral sentiment classes. We analyzed the collected tweets for sentiment classification using different sets of features and classifiers.

SYSTEM DESIGN

INTRODUCTION

Design is a multi- step that focuses on data structure software architecture, procedural details, algorithm etc... and interface between modules. The design process also translate the requirements into presentation of software that can be accessed for quality before coding begins. Computer software design change continuously as new methods; better analysis and border understanding evolved. Software design is at relatively early stage in its revolution.

Therefore, software design methodology lacks the depth, flexibility and quantitative nature that are normally associated with more classical engineering disciplines. However techniques for software designs do exist, criteria for design qualities are available and design notation can be applied.

DESIGN STRUCTURE

INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.

- Methods for preparing input validations and steps to follow when error occur.

OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

The output form of an information system should accomplish one or more of the following objectives.

- ❖ Convey information about past activities, current status or projections of the
- ❖ Future.
- ❖ Signal important events, opportunities, problems, or warnings.
- ❖ Trigger an action.
- ❖ Confirm an action.

SYSTEM TESTING

TESTING PROCESS

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product it is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

TYPES OF TESTS

Unit Testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program input produces valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

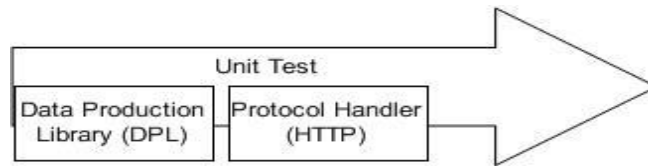


Fig. Unit Testing

Integration Testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

Functional Testing

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation and user manuals.

Functional testing is centered on the following items:

Valid Input is used to identified classes of valid input must be accepted.

Invalid Input is used to identified classes of invalid input must be rejected.

Functions is used to identified functions must be exercised.

Output is used to identify classes of application outputs.

Systems/Procedures is used to interfacing systems or procedures must be invoked. Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive Processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

System Testing

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

White Box Testing

White Box Testing is a testing in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is used to test areas that cannot be reached from a black box level.

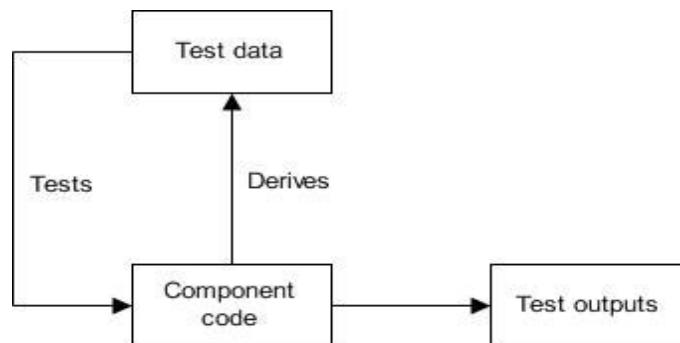


Fig. White box Testing

Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements

document. It is a testing in which the software under test is treated, as a black box .you **cannot “see” into it. The test provides inputs and responds to outputs without**

Considering how the software works.

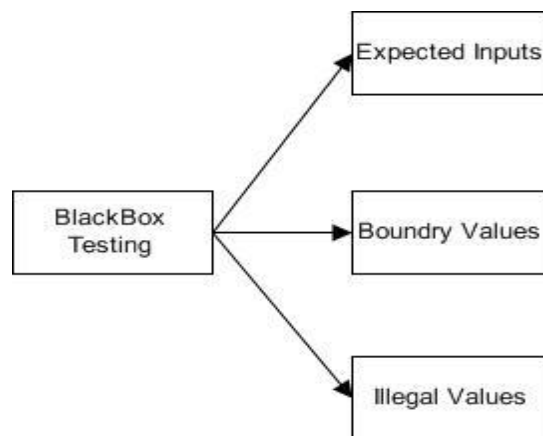


Fig. Black box Testing

TEST STRATEGY AND APPROACH

Field testing will be performed manually and functional tests will be written in detail.

Test Objectives

- ☐ All field entries must work properly.
- ☐ Pages must be activated from the identified link.
- ☐ The entry screen, messages and responses must not be delayed.
- ☐ Features to be tested
- ☐ Verify that the entries are of the correct format
- ☐ No duplicate entries should be allowed
- ☐ All links should take the user to the correct page.

Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications.

Acceptance Testing

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

ALPHA TESTING

In software development, alpha test will be a test among the teams to confirm that your product works. Originally, the term alpha test meant the first phase of testing in a software development process. The first phase includes unit testing, component testing, and system testing. It also enables us to test the product on the lowest common denominator machines to make sure download times are acceptable and pre loaders work.

BETA TESTING

In software development, a beta test is the second phase of software testing in which a sampling of the intended audience tries the product out. Beta testing can be considered "pre-release testing." Beta test versions of software are now distributed to curriculum specialists and teachers to give the program a "real world" test.

REQUIREMENT ANALYSIS

Requirement analysis, also called requirement engineering, is the process of determining user expectations for a new modified product. It encompasses the tasks that determine the need for analysing, documenting, validating and managing software or system requirements. The requirements should be documentable, actionable, measurable, testable and traceable related to identified business needs or opportunities and define to a level of detail, sufficient for system design.

FUNCTIONAL REQUIREMENTS

It is a technical specification requirement for the software products. It is the first step in the requirement analysis process which lists the requirements of particular software systems including functional, performance and security requirements. The function of the system depends mainly on the quality hardware used to run the software with given functionality.

Usability

It specifies how easy the system must be use. It is easy to ask queries in any format which is short or long, porter stemming algorithm stimulates the desired response for user.

Robustness

It refers to a program that performs well not only under ordinary conditions but also under unusual conditions. It is the ability of the user to cope with errors for irrelevant queries during execution.

Security

The state of providing protected access to resource is security. The system provides good security and unauthorized users cannot access the system there by providing high security.

Reliability

It is the probability of how often the software fails. The measurement is often expressed in MTBF (Mean Time Between Failures). The requirement is needed in order

to ensure that the processes work correctly and completely without being aborted. It can handle any load and survive and even capable of working around any failure.

Compatibility

It is supported by version above all web browsers. Using any web servers like localhost makes the system real-time experience.

Flexibility

The flexibility of the project is provided in such a way that it has the ability to run on different environments being executed by different users.

Safety

Safety is a measure taken to prevent trouble. Every query is processed in a secured manner without letting others to know one's personal information.

NON- FUNCTIONAL REQUIREMENTS

Portability

It is the usability of the same software in different environments. The project can be run in any operating system.

Performance

These requirements determine the resources required, time interval, throughput and everything that deals with the performance of the system.

Accuracy

The result of the requesting query is very accurate and high speed of retrieving information. The degree of security provided by the system is high and effective.

Maintainability

Project is simple as further updates can be easily done without affecting its stability. Maintainability basically defines that how easy it is to maintain the system. It means that how easy it is to maintain the system, analyse, change and test the application. Maintainability of this project is simple as further updates can be easily done without affecting its stability.

DATA FLOW DIAGRAM

Data Flow Diagram (DFD) is a two-dimensional diagram that describes how data is processed and transmitted in a system. The graphical depiction recognizes each source of data and how it interacts with other data sources to reach a mutual output. In order to draft a data flow diagram one must

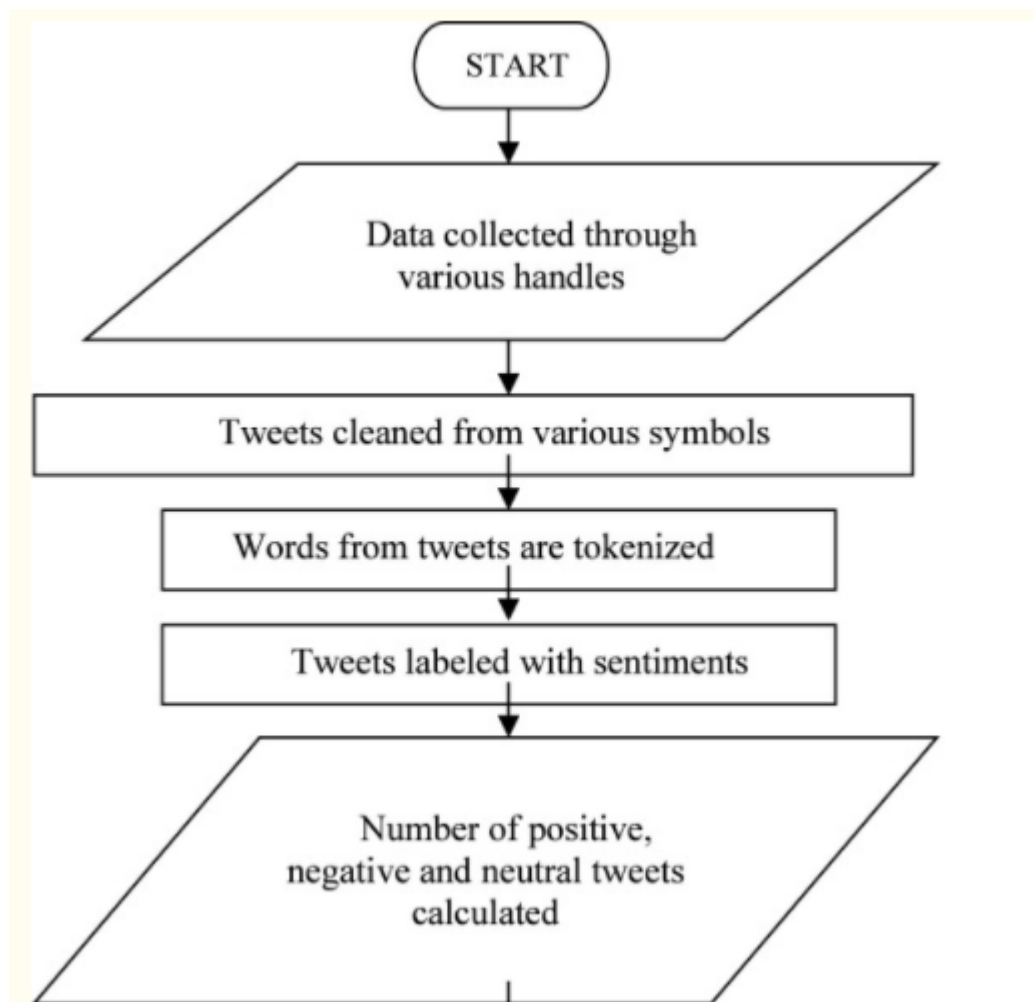
- Identify external inputs and outputs
- Determine how the inputs and outputs relate to each other
- Explain with graphics how these connections relate and what they result in.

Role of DFD:

- It is a documentation support which is understood by both programmers and nonprogrammers. As DFD postulates only what processes are accomplished not how they are performed.

- A physical DFD postulates where the data flows and who processes the data.

- It permits analyst to isolate areas of interest in the organization and study them by examining the data that enter the process and viewing how they are altered when they leave.



CONCLUSION

From the analysis in this paper, it is observed that people in India were mostly expressing their thoughts with positive sentiments. This paper concludes that there was a sudden hike in the tweets on every date when there is an announcement of Lockdown. Since the explosion of COVID-19 conspiracy theories, social media has been widely used both for and against misinformation and misconceptions. In this article, we address the issue of Twitter sentiment on COVID-19-related Twitter posts. We benchmark sentiment analysis methods in the analysis of COVID-19-related sentiment. Our findings indicate that the population favored the lockdown and stay home order in February; however, their opinion shifted by mid-March. While the people were posting tweets with negative sentiments in India, the twitter audience of India seems to reward positive sentiment much more than negative sentiment reversing the overall polarity to positive. This analysis was unique in its way as Lockdown shows peaks and were positive. Most people were criticizing Lockdown on the face but twitter was full of positive tweets.

REFERENCES

- [1] Manguri, Kamaran H, Ramadhan Rebaz N Amin, and Pshko R Mohammed, "Twitter Sentiment Analysis on Worldwide COVID-19 Outbreaks," Kurdistan Journal of Applied Research, 54–65, 2020.
- [2] C. Kaur and A. Sharma, "Twitter Sentiment Analysis on Coronavirus using Textblob," EasyChair 2516-2314, 2020.
- [3] Twitter sentiment analysis: The good the bad and the omg! Kouloumpis, Efthymios, and Wilson, Theresa, and Moore, Johanna, Fifth International AAAI Conference on weblogs and social media, 2011.
- [4] A. D. J. A. a. S. Dubey, "Twitter Sentiment Analysis during COVID19 Outbreak," 2020.
- [5] P. Tyagi and R. J. A. a. S. Tripathi, "A Review towards the Sentiment Analysis Techniques for the Analysis of Twitter Data," 2019.
- [6] Varsha Sahayak, Vijaya Shete and Apashabi Pathan, "Sentiment Analysis on Twitter Data," in International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163 Issue 1, Volume 2 (January 2015).
- [7] Namrata Godbole, Manjunath Srinivasaiah, Steven Skiena, "Large Scale Sentiment Analysis for News and Blogs," Google Inc., New York NY, USA and Dept. of Computer Science, Stony Brook University, Stony Brook, NY 11794-4400, USA.
- [8] N. K. Rajput, B. A. Grover, and V. K. J. a. p. a. Rathi, "Word frequency and sentiment analysis of twitter messages during Coronavirus pandemic," 2020.
- [9] M. Ra, B. Ab, and S. Kc, "COVID-19 Outbreak: Tweet based Analysis and Visualization towards the Influence of Coronavirus in the World," 2020.
- [10] Twitter Sentiment Analysis During Covid-19 Outbreak in Nepal, Pokharel, Bishwo Prakash, Available at SSRN 3624719, 2020.