**Software Requirements Specification**

For

**COVID-19 TWEETS: SENTIMENT ANALYSIS**

Prepared by

|  |  |  |
| --- | --- | --- |
| **NAME** | **SAP ID** | **Branch** |
| Shivam Singh Sikarwar | 500068623 | B Tech CSE CCVT |
| Shivika Maheshwari | 500067093 | B Tech CSE CCVT |
| Shridhar Avinash | 500067870 | B Tech CSE CCVT |



Department of Informatics

School Of Computer Science

UNIVERSITY OF PETROLEUM & ENERGY STUDIES,

DEHRADUN- 248007. Uttarakhand

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Topic** | | **Page No** |
| Table of Content | | 2 |
| Revision History | | 3 |
| 1 | Introduction | 4 |
|  | 1.1 Purpose of the Project | 4 |
|  | 1.2 Target Beneficiary | 4 |
|  | 1.3 Project Scope | 4 |
|  | 1.4 References | 4 |
| 2 | Project Description | 4 |
|  | 2.1 Reference Algorithm | 4 |
|  | 2.2 Data/ Data structure | 4 |
|  | 2.3 SWOT Analysis | 4 |
|  | 2.4 Project Features | 5 |
|  | 2.5 User Classes and Characteristics | 5 |
|  | 2.6 Design and Implementation Constraints | 5 |
|  | 2.7 Design diagrams | 5 |
|  | 2.8 Assumption and Dependencies | 6 |
| 3 | System Requirements | 6 |
|  | 3.1 User Interface | 6 |
|  | 3.2 Software Interface | 6 |
|  | 3.3 Database Interface | 6 |
|  | 3.4 Protocols | 6 |
| 4 | Non-functional Requirements | 6 |
|  | 4.1 Performance requirements | 6 |
|  | 4.2 Security requirements | 6 |
|  | 4.3 Software Quality Attributes | 6 |
| 5 | Other Requirements | 6 |
| Appendix A: Glossary | | 6 |
| Appendix B: Analysis Model | | 6 |
| Appendix C: Issues List | | 6 |

**Revision History**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Change** | **Reason for Changes** | **Mentor Signature** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

General Instructions:

1. Font should be Times New Roman 12
2. The main heading should be All Capital with Times New Roman 14
3. Sub-Heading should be Times new roman 12, Underline
4. The line gap should be 1.15
5. Justified alignment should be used for all text
6. The content inside a table should be Times New Roman 10
7. Caption for both Table and Figure should be Times New Roman 11
8. Add Source for all Images used.

|  |  |  |
| --- | --- | --- |
| 1 | INTRODUCTION | |
|  | 1.1 Purpose of the Project | Modeling and evaluating users’ sentiments towards different topics of a given query. |
|  | 1.2 Target Beneficiary | Organizations that want to acquire statistical analysis on covid-19 |
|  | 1.3 Project Scope | Expandable to multiple social media platforms and applications also it is faster than human analysis. |
|  | 1.4 References | * <https://www.mdpi.com/2078-2489/12/5/204> * https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=3572023 |
| 2 | PROJECT DESCRIPTION | |
|  | 2.1 Reference Algorithm | Naïve Bayes, Logistic Regression, Bag of words, stopwords, Sequence models  data structure: python dictionaries, NumPy arrays, Pandas dataframe |
|  | 2.2 Characteristic of Data | Data was brought in through twitter API, a set of tweets enclosing certain keywords that gauges sentiments related to covid-19  Data source - Twitter API  Sampling technique: Pandas, NLTK, Numpy  Pre-processing methods:  . Stemming, lemmatization, bag of words, stop words, words to vector conversion  . Removing @handles, URL’s, exclamation marks, special characters |
|  | 2.3 SWOT Analysis | Strength   * Model performs better in terms of already pre-processed tweets without irrelevant information * Sentiment analysis outputs are more classification based. * Analysis performed by the model are comparatively better in consideration with manual procedure   Weakness   * Model performs poorly for long sentences. * Model performs poorly for tweets with more amount of irrelevant and contradictory keywords. * No embedded authentication procedure.   Opportunities   * Learn new technologies and algorithms * Opportunity to work with text data and human speech analysis using AI powered solutions * Chance to work on human emotional intelligence * Project can be deployed for social causes   Threats   * No embedded authentication procedure. * Information misconduct |
|  | 2.4 Project Features | * Analyzing tweets from the period of covid-19 to understand people’s sentiment towards the pandemic. * Observing the change in opinions of people over time regarding covid-19 pandemic |
|  | 2.5 User Classes and Characteristics | People who want to understand emotions and sentiments of people worldwide on the perspective of covid-19 pandemic depicted on Social media platform, Twitter. |
|  | 2.6 Design and Implementation Constraints | Minimum Recommended  OS  Windows XP SP2 or higher Windows 7 or higher  Processor  Dual core 2.4GHz Quad Core 2.5GHz  Memory  2048 MB 4096 MB  Storage  2 GB 2 GB  Graphics  Video card with 512MB of VRAM Video card with 1024MB of VRAM |
|  | 2.7 Design diagrams | **Level – 1 DFD:**    **FLOW - CHART** |
|  | 2.8 Assumption and Dependencies | * Fully functional on a python configured environment * Tweets having relevant keywords are preffered |
| 3 | SYSTEM REQUIREMENTS | |
|  | 3.1 Software Interface | Python 3.9 or better  Deep Learning Frameworks or better |
| 4 | NON-FUNCTIONAL REQUIREMENTS | |
|  | 4.1 Performance requirements | * Proper working system * Python configured environment * Pre – Installed dependencies |
|  | 4.3 Software Quality Attributes | It classifies user’s sentiment as either positive, negative or neutral. It is fast and accurate. |
| Appendix A: Glossary | | CPU: Central processing unit  GPU: Graphical processing unit  NLP: Natural Language Processing  BOW: Bag of Words  NLTK: Natural language Toolkit  TF-IDF: Term Frequency – Inverse Document Frequency |
| Appendix B: Analysis Model | | Sentiment Score and classification |
| Appendix C: Issues List | | This is a dynamic list of the open requirements issues. |