CHAPTER 1

INTRODUCTION

1.1Project Introduction

Sentiment analysis is the task of finding the opinions and affinity of people towards specific topics of interest. Be it a product or a movie, opinions of people matter, and it affects the decision-making process of people. The first thing a person does when he or she wants to buy a product on-line, is to see the kind of reviews and opinions that people have written. Social media such as Facebook, blogs, twitter have become a place where people post their opinions on certain topics. The sentiment of the tweets of a particular subject has multiple usage, including stock market analysis of a company, movie reviews, in psychology to analyze the mood of people that has a variety of applications, and so on.

What is twitter sentiment analysis?

Twitter has emerged as a major micro-blogging website, having over 100 million users generating over 500 million tweets every day. With such large audience, Twitter has consistently attracted users to convey their opinions and perspective about any issue, brand, company or any other topic of interest. Due to this reason, Twitter is used as an informative source by many organizations, institutions and companies.

In order to extract sentiment from tweets, sentiment analysis is used. The results from this can be used in many areas like analyzing and monitoring changes of sentiment with an event, sentiments regarding a particular brand or release of a particular product, analyzing public view of government policies etc.

A **Twitter sentiment analysis** identifies negative, positive, or neutral emotions within the text of a tweet. It is a text analysis using natural language processing (NLP) and machine learning. It identifies and extracts subjective information from original data, providing a company with a better understanding of the social sentiment of its brand, product, or service. At the same time, analyze the online conversations of customers.

Sentiment analysis is frequently used to analyze customer feedback, survey responses, and product reviews.

Some relevant areas of sentiment analysis are monitoring the activities in existing communities within social networks, reputation management, and customer experience.

A Twitter sentiment analysis is the process of determining the emotional tone behind a series of words, specifically on Twitter. A sentiment analysis tool is an automated technique that extracts meaningful customer information related to their attitudes, emotions, and opinions.

A Twitter sentiment analysis determines negative, positive, or neutral emotions within the text of a tweet using NLP and ML models. Sentiment analysis or opinion mining refers to identifying as well as classifying the sentiments that are expressed in the text source. Tweets are often useful in generating a vast amount of sentiment data upon analysis. These data are useful in understanding the opinion of people on social media for a variety of topics.

Twitter sentiment analysis analyzes the sentiment or emotion of tweets. It uses natural language processing and machine learning algorithms to classify tweets automatically as positive, negative, or neutral based on their content. It can be done for individual tweets or a larger dataset related to a particular topic or event.

In this project Twitter sentiment analysis we are using covid vaccination data to check the sentiments of users when they take the vaccination dose and find out the mood of user at covid period. We are using three models two for testing the accuracy of the model and one is a predefined model to check the sentiments.

1.2 Motivation

Twitter is a relatively new tool for measuring social studies. We are talking about millions of people voluntarily writing their opinions on any topic. It is an organic data source for data collection and research. Sociological, political, economic, and analytical analyses are currently being carried out, especially for science and business. Twitter is possibly the best place to get a sampling of public opinion.
Studying individual and group reactions to a specific topic is doing sentiment analysis. Although polls are still a phenomenal tool for data collection, the supply of particular tweets on the subject of interest is functional on Twitter. This allows calculating the behavior behind the sentiment in each tweet and then aggregating the results. Companies use this information to gather public opinion about their products and services, avoid risk, and make data-driven decisions. Until recently, analyzing tweets about a brand, article, or benefit was a manual, complex and

monotonous task; it required a person to go through the tweets on the topic one by one, determine which ones were relevant, carefully examine and classify them according to their criteria. It is not only unsustainable but also expensive and time-consuming. It also makes it vulnerable to human error.

- Voice of Customer (VoC) Programs

Voice of Customer (VoC) Programs is the feedback gathered better to understand customers' feelings and concerns about a brand. This is crucial for the improvement of customer experience. This data will provide us with risk prevention and develop strategies for better fixing issues concerning the company's products or services.
Evaluating how customers feel and their opinions can assist you in distinguishing and addressing rising issues. Net Promoter Score (NPS) studies are the best way to conduct these evaluations. Assumption investigation turns the content into the drivers of NPS.

- Customer Service Experience

An extraordinary client benefit involvement can make or break a company. Estimation investigation and content examination can both be connected to client back discussions. Assumption examination can diminish handling times and increase productivity by coordinating questions with the correct people. Eventually, clients get distant better; a much better a stronger bolster involvement, and you'll be able to diminish churn rates.

- Product Experience

Sentiment analysis can identify how your clients feel about the highlights and benefits of your products or services. This may offer assistance and reveal areas of opportunity that may not have been mindful of before. You may mine online item audits for criticism on a particular item category over all competitors in this showcase. You'll be able, at that point, to apply estimation examination to uncover subjects that your clients feel contrarily approximately. This might uncover openings or joint issues.

- Brand Sentiment Analysis

The sentiments surrounding a brand are one of the most important factors to consider for a good customer experience. Depending on the brand sentiment, sales increase or decrease. This is also reflected in brand loyalty, where positive sentiments result in good reviews and

recommendations, while negative sentiments increase customer churn rates. Sentiment analysis provides brands with tools to monitor how their customers feel about them.

Studying the communities in forums and social media platforms is advisable to be aware of your brand's reputation. Companies must also track their brand, product names, and competitor mentions to understand the brand image in the big picture. This helps companies assess how a public relations campaign or new product launch has impacted overall brand sentiment.

- Social Media Sentiment Analysis

Social media can be one of the most capable ways to reach potential clients and keep existing ones. Great client reviews and posts on social media empower other clients to buy from a company. On the other hand, bad reviews and comments can be one of the most harmful advertising out there.

1.3 Objective And Scope

The objective of Twitter sentiment analysis on COVID vaccination using the RoBERTa model is to analyze and understand the sentiment expressed by Twitter users regarding COVID vaccination. The RoBERTa model, a variant of the Transformer model, is a powerful language model that can learn the contextual meaning of words and sentences.

By applying the RoBERTa model to Twitter data related to COVID vaccination, the goal is to classify tweets into different sentiment categories, such as positive, negative, or neutral. This analysis can provide valuable insights into public opinions, concerns, and attitudes towards COVID vaccination.

The specific objectives of this sentiment analysis can include:

1. Sentiment Classification: Determine whether a tweet expresses a positive, negative, or neutral sentiment towards COVID vaccination. This classification helps in understanding the overall sentiment distribution and identifying trends or patterns.

2. Opinion Mining: Extract and analyze specific opinions, concerns, or issues expressed by Twitter users regarding COVID vaccination. This can provide valuable insights into the reasons behind positive or negative sentiments.

3. Public Perception Analysis: Identify the prevailing sentiment towards COVID vaccination on Twitter over time. This analysis can help monitor shifts in public opinion, track the effectiveness of vaccination campaigns, and identify potential areas of concern or misinformation.

4. Identifying Influencers: Identify influential users or accounts on Twitter who are actively discussing COVID vaccination and analyze their sentiments. This can help understand the impact of influential voices in shaping public opinion.

Overall, the objective of Twitter sentiment analysis on COVID vaccination using the RoBERTa model is to gain a deeper understanding of the sentiment landscape, public perceptions, and emerging trends related to COVID vaccination on the Twitter platform. This analysis can assist public health organizations, policymakers, and researchers in making data-driven decisions and designing effective communication strategies related to COVID vaccination.

1.4 Hardware Requirements:

1. Central Processing Unit (CPU) — Intel Core i5 6th Generation processor or higher. An AMD equivalent processor will also be optimal.
2. RAM — 8 GB minimum, 16 GB or higher is recommended.
3. Operating System — Ubuntu or Microsoft Windows 10. I recommend updating Windows 10 to the latest version before proceeding forward.

1.5 Software Requirements:

- Programming Language: Python
- Integrated Development Environment (IDE): Googlecolab
- Machine Learning Libraries/Frameworks: Python: Textblob,sklearn,wordcloud etc
- Data Manipulation and Analysis Libraries: Python: NumPy, Pandas.
- Supporting Tools and Libraries: Additional tools or libraries to preprocess data, handle text or image data, or perform specific tasks such as NLTK (Natural Language Toolkit) for text processing .
- Testing and Evaluation Frameworks: It's crucial to evaluate your ML models' performance.
- Documentation and Collaboration Tools: googlecolab .

CHAPTER 2

SYSTEM ANALYSIS & REQUIREMENTS SPECIFICATIONS

2.1 System Analysis

Data Collection: The system should be able to connect to the Twitter API using Tweepy to collect tweets in real-time or from a specific time period. The collected data should include relevant information such as tweet text, timestamp, user information, and other metadata.

Preprocessing: The system should perform preprocessing on the collected tweets to clean the data and make it suitable for sentiment analysis. This may include removing stop words, punctuation, and special characters, as well as normalizing the text by converting it to lowercase.

Sentiment Analysis: The system should apply a sentiment analysis algorithm to determine the sentiment of each tweet. This can be achieved using machine learning techniques such as classification algorithms (e.g., Naive Bayes, Support Vector Machines) or pre-trained models.

Sentiment Classification: The system should categorize each tweet into sentiment categories such as positive, negative, or neutral based on the sentiment analysis results. This can be done by defining a threshold for sentiment polarity scores or using predefined sentiment lexicons.

Visualization: The system should provide visualizations to display the sentiment analysis results. This can include generating charts, graphs, or word clouds to represent the overall sentiment distribution or sentiment trends over time.

2.2 Requirements Specifications:

Data Collection: The system should provide options to collect tweets based on user-defined search queries, hashtags, user handles, or geographical locations. It should also support filtering options such as language filtering or retweet filtering.

Preprocessing: The system should perform preprocessing tasks such as removing stop words, punctuation, and special characters, as well as normalizing the text by converting it to lowercase.

Sentiment Analysis Algorithm: The system should implement a sentiment analysis algorithm or integrate pre-trained sentiment analysis models. The algorithm should be capable of determining the sentiment polarity of a given tweet.

Sentiment Classification: The system should categorize tweets into sentiment categories (positive, negative, neutral) based on the sentiment analysis results. It should allow users to define custom sentiment thresholds or use default settings.

Visualization: The system should generate visualizations such as charts, graphs, or word clouds to represent the sentiment analysis results. The visualizations should be intuitive, interactive, and provide insights into the sentiment distribution and trends.

Export and Reporting: The system should allow users to export the sentiment analysis results and visualizations in various formats, such as CSV, Excel, or PDF. It should also provide options for generating reports summarizing the sentiment analysis findings.

Error Handling: The system should handle errors gracefully and provide meaningful error messages to users in case of issues with data collection, preprocessing, sentiment analysis, or visualization.

Scalability: The system should be designed to handle a large volume of tweets efficiently, ensuring scalability and performance even during peak loads.

2.3 DFD :-

DFD is the abbreviation for Data Flow Diagram. The flow of data of a system or a process is represented by DFD. It also gives insight into the inputs and outputs of each entity and the process itself. DFD does not have control flow and no loops or decision rules are present. Specific operations depending on the type of data can be explained by a flowchart.
It is a graphical tool, useful for communicating with users ,managers and other personnel. it is useful for analyzing existing as well as proposed system.

It provides an overview of

- What data is system processes.
- What transformation are performed.
- What data are stored.
- What results are produced , etc.

Data Flow Diagram can be represented in several ways. The DFD belongs to structuredanalysis modeling tools. Data Flow diagrams are very popular because they help us to visualize the major steps and data involved in software-system processes.

Components of DFD

The Data Flow Diagram has 4 components:

- Process Input to output transformation in a system takes place because of process function. The symbols of a process are rectangular with rounded corners, oval, rectangle or a circle. The process is named a short sentence, in one word or a phrase to express its essence
- Data Flow Data flow describes the information transferring between different parts of the systems. The arrow symbol is the symbol of data flow. A relatable name should be given to the flow to determine the information which is being moved. Data flow also represents material along with information that is being moved. Material shifts are modeled in systems that are not merely informative. A given flow should only transfer a single type of information. The direction of flow is represented by the arrow which can also be bidirectional.
- Warehouse The data is stored in the warehouse for later use. Two horizontal lines represent the symbol of the store. The warehouse is simply not restricted to being a data file rather it can be anything like a folder with documents, an optical disc, a filing cabinet. The data

warehouse can be viewed independent of its implementation. When the data flow from the warehouse it is considered as data reading and when data flows to the warehouse it is called data entry or data updating.

- Terminator The Terminator is an external entity that stands outside of the system and communicates with the system. It can be, for example, organizations like banks, groups of people like customers or different departments of the same organization, which is not a part of the model system and is an external entity. Modeled systems also communicate with terminator.

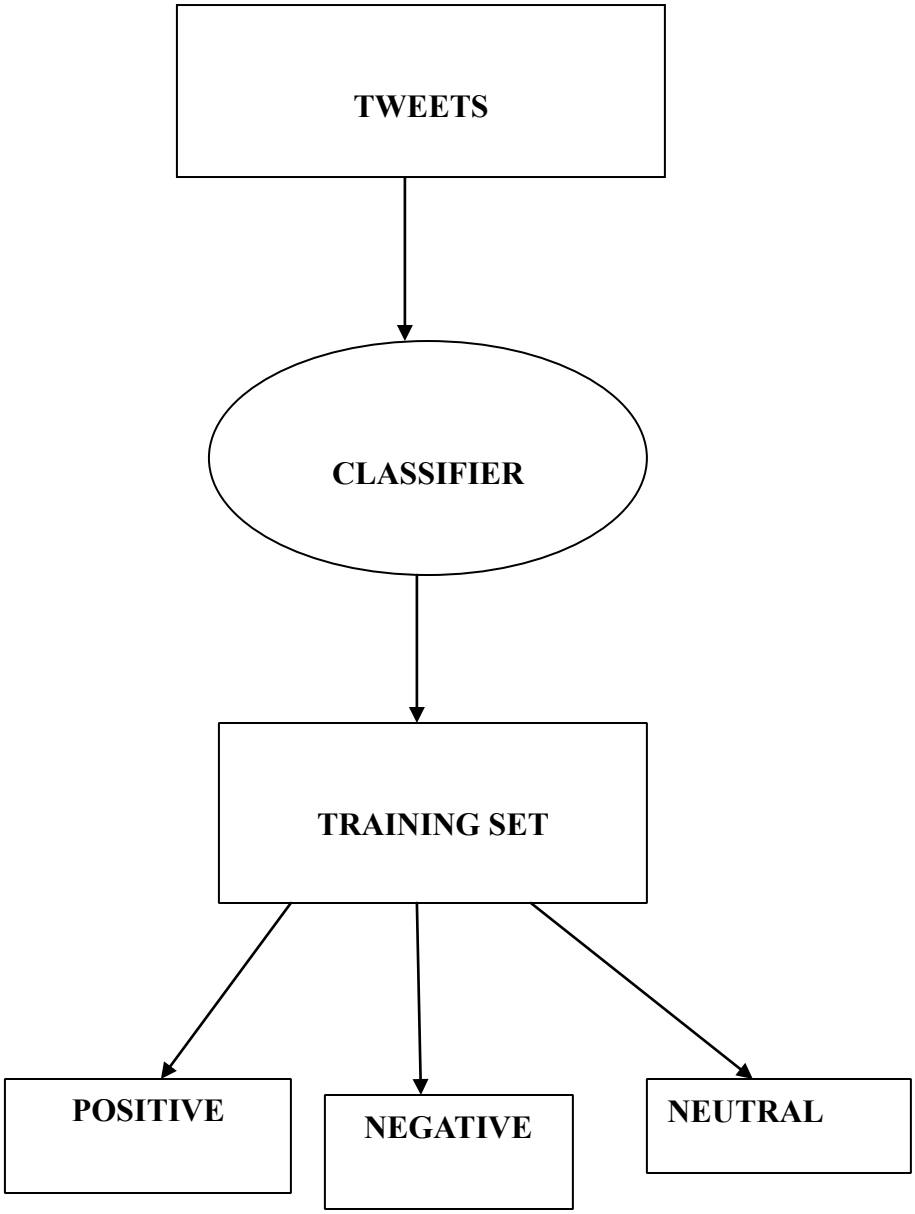
Rules for creating DFD

- The name of the entity should be easy and understandable without any extra assistance(like comments).
- The processes should be numbered or put in ordered list to be referred easily.
- The DFD should maintain consistency across all the DFD levels.
- A single DFD can have a maximum of nine processes and a minimum of three processes.

Symbols Used in DFD

- Square Box: A square box defines source or destination of the system. It is also called entity. It is represented by rectangle.
- Arrow or Line: An arrow identifies the data flow i.e. it gives information to the data that is in motion.
- Circle or bubble chart: It represents as a process that gives us information. It is also called processing box.
- Open Rectangle: An open rectangle is a data store. In this data is store either temporary or permanently.

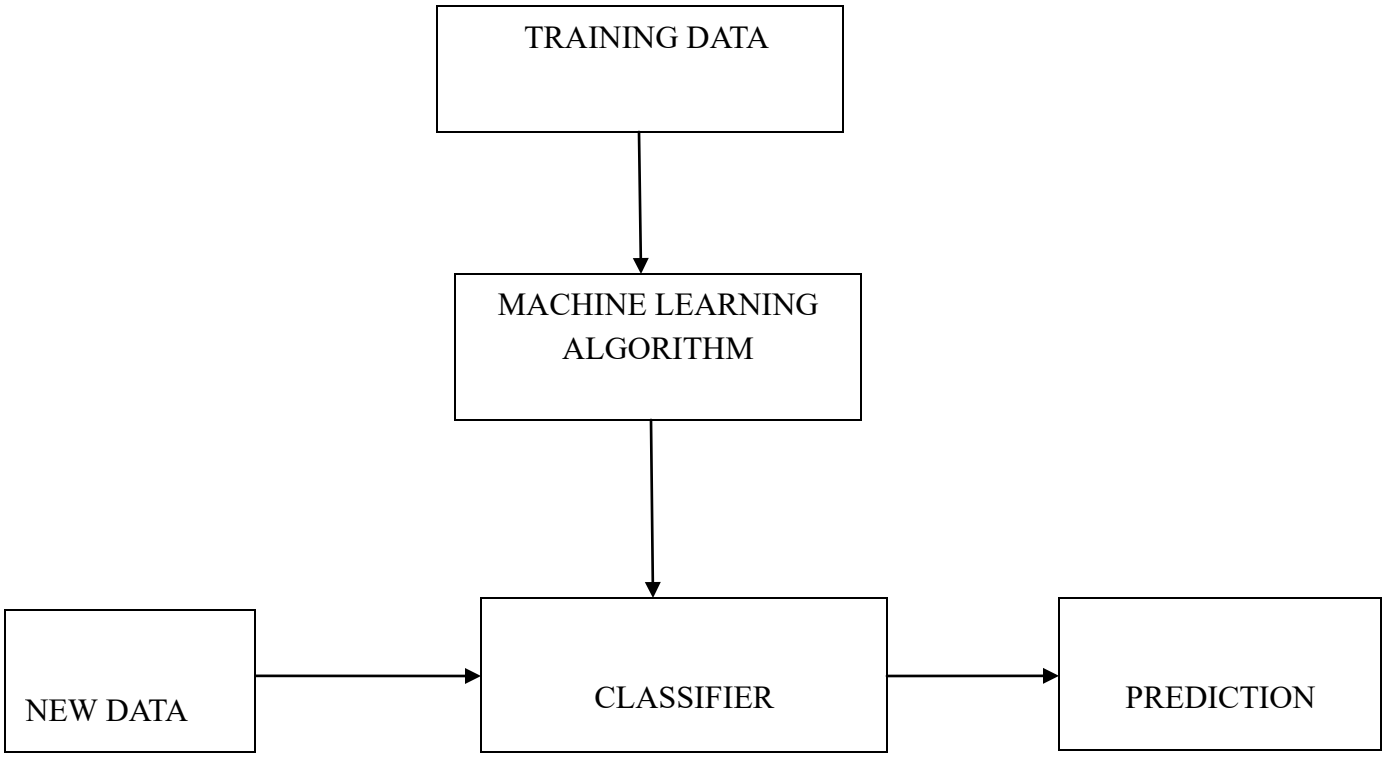
DFD :-


TWEETS
CLASSIFIER
TRAINING SET
POSITIVE
NEGATIVE
NEUTRAL

2.4System flow Diagram:

System flow diagrams are the diagram type that shows you the flow of data and how decisions can affect the events surrounding it. Like other types of flowcharts, system flowcharts consist of start/end terminals, processes, and decisions, all connected by arrows showing the flow and how data moves in the flow. It is one of the graphical representations of the flow of data in a system in software engineering. Different symbols are strung together to show data flow, including what happens to data and where it goes. Input is the data that the system receives, while output is the data that the system sends out.

In a system, the process usually involves feeding the system with a source of input. Then, that input will be processed and modified in some way to produce the output that you will get in the end.



CHAPTER – 3

IMPLEMENTATION

The goal of this project is to analyze the sentiment (positive, negative, or neutral) of tweets on Twitter based on covid vaccination data onto public opinion and attitudes towards the subject of interest.

In this project tweets are extracted using a python libraries. determine the sentiment of the tweets.

Steps for implementation :

Setup:

- Install recommended library and other necessary dependencies.

Data Collection:

- Using the csv file data that is present in Kaggle.com.
- Defining the criteria for collecting tweets, such as search keywords, hashtags, or user timelines.

Data Preprocessing:

- Cleaning the collected tweet data by removing special characters, URLs, and unnecessary symbols.
- Tokenizing the text into individual words or phrases.
- Performing text normalization techniques like stemming or lemmatization.
- Removing stop words or irrelevant terms that do not contribute to sentiment analysis.

Sentiment Analysis:

- Choosing a sentiment analysis approach, such as using pre-trained models or training custom models.
- Here logistic,SVM algorithm is chosen for training the data for sentiment analysis.
- Using Roberta model for testing the data.

- Applying the selected sentiment analysis technique to determine the sentiment polarity (positive, negative, neutral) of each tweet.
- Assigning a sentiment score or label to each tweet based on the determined polarity.

3.1 SVM (Support Vector Machine)

SVM model is applied for prediction in this project. Support Vector Machines (SVM) is a popular machine learning algorithm that can be used for sentiment analysis in Twitter data. SVM is a supervised learning algorithm that can classify data into different categories by finding an optimal hyperplane that maximally separates the data points of different classes.

Data Preprocessing: Clean the raw tweet data by removing special characters, URLs, and unnecessary symbols.

Tokenize the text into individual words or phrases.Perform text normalization techniques like stemming or lemmatization.Remove stop words or irrelevant terms that do not contribute to sentiment analysis.

Labeling: Assign sentiment labels to the tweets based on the sentiment categories (positive, negative, neutral). Manually label a subset of tweets as training data or use pre-labeled datasets.

Training Data Preparation: Split the labeled dataset into training and testing sets. Typically, use a majority portion for training (e.g., 80%) and the remaining for testing and evaluation.

Model Training: Train an SVM model using the labeled training data and their corresponding feature vectors. During training, the SVM algorithm finds the optimal hyperplane that maximizes the margin between different sentiment classes.
Model Evaluation: Evaluate the trained SVM model using the labeled testing data.Calculate evaluation metrics such as accuracy, precision, recall, and F1-score to assess the model's performance.

Prediction: Use the trained SVM model to predict the sentiment of new, unlabeled tweets.

Extract their feature vectors using the same feature extraction techniques applied during training.

Post-processing: Analyze the predicted sentiment labels to gain insights into the sentiment distribution .

3.2 Logistic regression

The steps involved in training a logistic regression model for data are as follows:

1. Data Preparation: Prepare your dataset by cleaning and preprocessing the data. This step involves handling missing values, encoding categorical variables, and normalizing or scaling numerical features.

2. Splitting the Data: Divide your dataset into two subsets: a training set and a test set. The training set will be used to train the logistic regression model, while the test set will be used to evaluate its performance.

3. Feature Selection: Select the relevant features or predictors from your dataset that will be used to train the model. This step involves analyzing the importance and correlation of different features and choosing the ones that have the most significant impact on the target variable.

4. Model Initialization: Initialize the logistic regression model. In Python, you can use libraries like scikit-learn to create and initialize the logistic regression object.

5. Model Training: Train the logistic regression model using the training dataset. During training, the model learns the coefficients or weights for each feature to fit the data and minimize the logistic loss or cross-entropy loss function.

6. Model Evaluation: Evaluate the performance of the trained logistic regression model using the test dataset. Calculate various metrics such as accuracy, precision, recall, and F1 score to assess how well the model generalizes to unseen data.

7. Model Optimization (Optional): If the model performance is not satisfactory, you can try optimizing the model by adjusting hyperparameters. Hyperparameters like learning rate, regularization parameter, and number of iterations can be tuned using techniques like grid search or random search.

8. Model Deployment: Once you are satisfied with the model's performance, you can deploy it to make predictions on new, unseen data. The logistic regression model can classify new instances into different classes based on the learned decision boundary.

Remember that logistic regression is a binary classification algorithm, primarily used when the target variable is binary (two classes). If you have a multiclass classification problem, you can

extend logistic regression to handle multiple classes using techniques like one-vs-rest or multinomial logistic regression.

3.3 Roberta model

RoBERTa (Robustly Optimized BERT Pretraining Approach) is a state-of-the-art transformer-based neural network model for natural language processing (NLP). It is an extension of the popular BERT (Bidirectional Encoder Representations from Transformers) model developed by Google.

Steps for model working:

1. Data Preprocessing: Prepare the new data for testing by performing the same preprocessing steps that were applied during training. This may include cleaning the text, encoding categorical variables, and normalizing or scaling numerical features.

2. Tokenization: Tokenize the input text data into tokens that the RoBERTa model understands. RoBERTa requires tokenization specific to its pretraining, such as using WordPiece or SentencePiece tokenizers.

3. Encoding: Convert the tokenized text into numerical representations that can be fed into the RoBERTa model. This step involves mapping the tokens to their corresponding token IDs and generating attention masks to indicate which tokens are important for the model to pay attention to.

4. Model Inference: Pass the preprocessed and encoded data through the trained RoBERTa model to obtain predictions. The RoBERTa model will generate a probability distribution over the possible classes for each input instance.

5. Post-processing: Depending on the task at hand, you may need to post-process the model's predictions. For example, if you are performing sentiment analysis, you might threshold the probabilities to assign a positive or negative label.

6. Evaluation: Compare the model's predictions with the ground truth labels to evaluate its performance. Calculate relevant evaluation metrics such as accuracy, precision, recall, and F1 score to assess how well the model performs on the test data.

CHAPTER 4

4.1 Project Management

Project management of a sentiment analysis project involves organizing and overseeing the various tasks and resources to ensure the successful execution and completion of the project.

Project management for a sentiment analysis project involves effectively planning, organizing, and executing various tasks to ensure the successful completion of the project. Clearly define the objectives and goals of the project. Identify the specific outcomes you aim to achieve through the sentiment analysis of Twitter data. For example, it could be monitoring brand sentiment, understanding public opinion on a particular topic, or analyzing sentiment trends over time.Develop a detailed project plan that outlines the tasks, timelines, and resources required to complete the project. Define milestones and deliverables, and allocate resources accordingly. Consider factors like data collection, preprocessing, analysis techniques, model selection, evaluation, and reporting.Assemble a team with the necessary skills and expertise for the project.

Identify team members responsible for data collection, preprocessing, sentiment analysis, visualization, and reporting. Assign roles and responsibilities, and establish clear communication channels within the team.Determine the data sources and methods for collecting Twitter data. Define the search queries, hashtags, or user profiles to target. Ensure data privacy and compliance with relevant regulations. Establish procedures for data preprocessing, including text cleaning, tokenization, and feature extraction.Research and select appropriate sentiment analysis techniques for the project. Consider both traditional machine learning approaches (e.g., SVM, Naive Bayes) and modern deep learning methods (e.g., recurrent neural networks, transformers). Evaluate the pros and cons of each technique based on project requirements. Develop and train the sentiment analysis models using suitable datasets. Fine-tune the models based on labeled data specific to your project. Optimize hyperparameters, assess model performance, and ensure the models align with project objectives. Document the process for reproducibility.Integrate the developed models into the sentiment analysis pipeline. Perform extensive testing and validation to ensure the accuracy and reliability of sentiment predictions. Monitor the system's performance, evaluate any limitations or biases, and iteratively refine the models as needed.Analyze the sentiment analysis results to extract meaningful insights. Generate

visualizations such as sentiment distribution charts, word clouds, or sentiment trends over time. Interpret the findings to gain insights into public opinion, brand reputation, or other relevant aspects related to the project objectives.Regularly assess the project's progress and performance against the defined objectives and milestones. Conduct periodic reviews, gather feedback from stakeholders, and make necessary adjustments to ensure the project stays on track. Document lessons learned for future improvements.Prepare comprehensive reports and presentations summarizing the project methodology, findings, and insights. Clearly communicate the results to stakeholders, clients, or project sponsors. Tailor the presentation to the intended audience, emphasizing the impact and relevance of the sentiment analysis results.Document the project workflow, code, and any relevant procedures, including data sources and preprocessing steps. Maintain proper documentation for future reference, replication, or extension of the project. Establish a system for ongoing maintenance and support as needed.Conduct a thorough project review to evaluate the overall success and lessons learned. Archive project artifacts and finalize documentation. Celebrate achievements, recognize team members' contributions, and transition any ongoing responsibilities or maintenance to the appropriate stakeholders. Celebrate the successful completion of the project with the team. Perform a final review to ensure all project deliverables are met.Archive project documentation and code for future reference.

Throughout the project management process, it's crucial to maintain open communication with the team, stakeholders, and clients. Regularly track progress, address challenges proactively, and adapt the project plan if needed. Flexibility, collaboration, and effective project management techniques will contribute to the successful completion of a Twitter sentiment analysis project.

We should remember that effective project management involves regular communication, collaboration among team members, and adaptability to handle unforeseen challenges or changes throughout the project lifecycle.

4.2 Process Planning and Scheduling :

Project Planning and Scheduling', though separate, are two sides of the same coin in project management. Fundamentally, 'Project planning' is all about choosing and designing effective policies and methodologies to attain project objectives. While 'Project scheduling' is a procedure of assigning tasks to get them completed by allocating appropriate resources within an estimated budget and time-frame.

The basis of project planning is the entire project. Unlikely, project scheduling focuses only on the project-related tasks, the project start/end dates and project dependencies. Thus, a ‘project plan’ is a comprehensive document that contains the project aims, scope, costing, risks, and schedule. And a project schedule includes the estimated dates and sequential project tasks to be executed.

Project Planning : The project planning phase refers to:

- Developing a project to make it ready for investment
- Determines the jobs/tasks required to attain project objectives

Stages of Project Planning

The project planning stages are enlisted below:

1. Identifying the key project sponsors and stakeholders, to determine the basis of project scope, budget, and time-frame for project execution.
2. Upon enlisting the stake-holder requirements, prioritizing/setting project objectives.
3. Identifying the project deliverables required to attain the project objectives.
4. Creating the project schedule.
5. Identifying the project risks, if any, and develop suitable mitigation plans.
6. Communicating and presenting the project plan to stakeholders.

Benefits of Project Planning

- Route-Map: The project plan offers a road-way that gives direction to the project from start to end.
- Documentation of Customer Requirements: A well-articulated project plan enables the record of the requirements of the customers in a documented form. This provides a precise direction instead of relying on assumptions, which could be incorrect and may lead to project errors.
- Task Autonomy: Planning enables one to assign tasks to specific team members and gives autonomy. The team feels a sense of responsibility and ownership of the success or failure of a project. Consequently, it urges them to work better or encourages them to bring inconsistent results.

- Resource Estimation: Planning is vital as in a way, it enables us to estimate resources, costing and time. It gives a judgment of any delays if several members are working on various projects at a time.
- Mitigation Plan: The project plan gives a way to forecast risks, if any, and plan for mitigation strategies accordingly.
- Identification of Employee Capabilities: The planning phase enables to identify employees with certain skill-sets or expertise. And as the tasks get assigned, team members get trained on a lacking skill-sets or either upgraded on the ones they possess.
- Strengths and Short-Comings of Previous Projects: Project plans also help to analyze and improve or learn from the previous project records and facilitate decision-making.

4.2.1 Project Development Approach

In Twitter sentiment analysis for covid vaaccination, the choice of project development approach depends on factors such as project complexity, team size, timeline, and available resources. Here are a few common development approaches that can be applied to Twitter sentiment analysis projects:

Waterfall Approach:

The Waterfall approach follows a linear and sequential development process.Each phase (requirements gathering, design, development, testing, deployment) is completed before moving on to the next.This approach works well when project requirements are well-defined and unlikely to change significantly during the development process.However, it may not be suitable for projects where requirements evolve rapidly or require frequent iterations.

Prototype Approach:

The prototype approach involves developing an initial version or prototype of the sentiment analysis system.The prototype focuses on the core functionality and features.Users or stakeholders can provide feedback based on the prototype, guiding further development.This approach allows for early validation of concepts and user requirements, reducing the risk of developing a system that does not meet expectations.

However, careful planning is necessary to ensure the prototype can evolve into a fully functional system without significant rework.

Incremental Approach:

The incremental approach involves dividing the project into increments or modules, each delivering a specific set of features or functionalities.Each increment builds upon the previous increment, adding new functionality or improving existing features.This approach allows for the delivery of working components in stages, providing value early in the development process.It can be particularly useful when there are time constraints or when specific features need to be prioritized.

4.2.2 Project Plan

A project plan—sometimes called a work plan—is a blueprint of the goals, objectives, and tasks your team needs to accomplish for a specific project. Our project plan should include information about our project schedule, scope, due dates, and deliverables for all phases of the project lifecycle. But not all project planning processes are created equal—which leads some teams to underutilize them or skip over them completely. To write an effective project plan, we need to be methodical (follow a series of steps), specific, and clear when it comes to our ideas and execution strategy.

Project planning is the second stage in the project management process, following project initiation and preceding project execution. During the project planning stage, the project manager creates a project plan, which maps out project requirements. The project planning phase typically includes setting project goals, designating project resources, and mapping out the project schedule. Project plans set the stage for the entire project. Without one, we're missing a critical step in the overall project management process. When we launch into a project without defined goals or objectives, it can lead to disorganized work, frustration, and even scope creep. A clear, written project management plan provides a baseline direction to all stakeholders, while also keeping everyone accountable. It confirms that we have the resources you need for the project before it actually begins.

A project plan also allows us, as the person in charge of leading execution, to forecast any potential challenges you could run into while the project is still in the planning stages. That way, you can ensure the project will be achievable—or course-correct if necessary. According to a study conducted by the Project Management Institute, there is a strong correlation between project planning and project success—the better your plan, the better your outcome. So, conquering the planning phase also makes for better project efficiency and results.

The basic outline of any project plan can be summarized in these five steps:

- Define your project's stakeholders, scope, quality baseline, deliverables, milestones, success criteria and requirements. Create a project charter, work breakdown structure (WBS) and a statement of work (SOW).
- Identify risks and assign deliverables to your team members, who will perform the tasks required and monitor the risks associated with them.
- Organize your project team (customers, stakeholders, teams, ad hoc members, and so on), and define their roles and responsibilities.
- List the necessary project resources, such as personnel, equipment, salaries, and materials, then estimate their cost.
- Develop change management procedures and forms.
- Create a communication plan, schedule, budget and other guiding documents for the project.

4.3 Milestones

A project milestone is a project planning tool that's used to mark a point in a project schedule. Project milestones can note the start and finish of a project, mark the completion of a major phase of work or anything that's worth highlighting in a project, such as the production of project deliverables. Milestones help project teams coordinate their efforts by helping everybody understand the objectives of the project and the action steps that must be taken to achieve them.

Project milestones help project teams focus on major progress points in a project, which helps project managers with project planning and scheduling. Just as tasks break a larger project into manageable parts, milestones break down project phases to help project managers plan, schedule and execute them.

Project milestones provide a way to more accurately estimate the time it'll take to complete your project by marking important dates and events, making them essential for precise project planning and scheduling. Because of their versatility, milestones are an important element of project documents such as the project schedule, project charter and project plan.They're also used in scheduling methodologies, such as the critical path method (CPM), or project management tools like Gantt charts, which can determine major scheduling periods. With project milestones, you can better calculate the slack in your project by segmenting the project timeline into intervals, or smaller time frames to control and track progress.

4.4 Deliverables

Projects produce deliverables, which are simply the results of project activities. Project deliverables can be big or small, and their number varies depending on the project. They're agreed upon by the project management team and stakeholders during the project planning phase.Put another way, there are inputs and outputs in any type of project. Inputs are what you put into the project, such as data, resources, etc., and the outcomes are the deliverables. Again, those deliverables vary greatly. For example, a project deliverable can be either a product or service or it can be the documentation that's part of the project closure.

Project management tools such as Gantt charts, kanban boards and project calendars can help us track the progress of our team on the completion of project deliverables. In addition to these project management tools, ProjectManager has one-click reporting that captures data on project variance, time, cost and more. These reports can be shared as PDF attachments or printed depending on the stakeholder's preference.

A deliverable is quantifiable. It's something that was created over time, with resources and effort. A project milestone, while encompassing deliverables, is a marker in time to indicate the transition from one thing to another.

4.5 Risk Management

Risk management is the process of identifying, assessing and controlling threats to an organization's capital and earnings. These risks stem from a variety of sources, including financial uncertainties, legal liabilities, technology issues, strategic management errors, accidents and natural disasters. A successful risk management program helps an organization consider the full range of risks it faces. Risk management also examines the relationship between risks and the cascading impact they could have on an organization's strategic goals.

Risk management involves identifying, analyzing, and accepting or mitigating uncertainty in investment decisions. Put simply, it is the process of monitoring and dealing with the financial risks associated with investing. Risk management essentially occurs when an investor or fund manager analyzes and attempts to quantify the potential for losses in an investment, such as a moral hazard, and then takes the appropriate action (or inaction) to meet their objectives and risk tolerance.Risk management structures are tailored to do more than just point out existing risks. A good risk management structure should also calculate the uncertainties and predict their influence on a business. Consequently, the result is a choice between accepting risks or rejecting them. Acceptance or rejection of risks is dependent on the tolerance levels that a business has already defined for itself.

If a business sets up risk management as a disciplined and continuous process for the purpose of identifying and resolving risks, then the risk management structures can be used to support other risk mitigation systems. They include planning, organization, cost control, and budgeting. In such a case, the business will not usually experience many surprises, because the focus is on proactive risk management.

Response to risks usually takes one of the following forms:

- Avoidance: A business strives to eliminate a particular risk by getting rid of its cause.
- Mitigation: Decreasing the projected financial value associated with a risk by lowering the possibility of the occurrence of the risk.

- Acceptance: In some cases, a business may be forced to accept a risk. This option is possible if a business entity develops contingencies to mitigate the impact of the risk, should it occur.

4.5.1 Risk Identification

Risk identification is the process of documenting any risks that could keep an organization or program from reaching its objective. It's the first step in the risk management process, which is designed to help companies understand and plan for potential risks. Examples of risks include theft, business downturns, accidents, lawsuits or data breaches.When you identify risks, look for events that may prevent a project from achieving its goal. The risk's origin can be the project itself or external sources.

There are several situations for which you might need to identify risks, including:

- To support an investment decision
- To assess cost uncertainty or operational costs
- To analyze multiple alternatives
- To test a program before its acquisition

4.5.2 Risk Analysis

Risk analysis is the process of identifying and analyzing potential issues that could negatively impact key business initiatives or projects. This process is done in order to help organizations avoid or mitigate those risks. Performing a risk analysis includes considering the possibility of adverse events caused by either natural processes, like severe storms, earthquakes or floods, or adverse events caused by malicious or inadvertent human activities. An important part of risk analysis is identifying the potential for harm from these events, as well as the likelihood that they will occur.

Enterprises and other organizations use risk analysis to:

- anticipate and reduce the effect of harmful results from adverse events;
- evaluate whether the potential risks of a project are balanced by its benefits to aid in the decision process when evaluating whether to move forward with the project;

- plan responses for technology or equipment failure or loss from adverse events, both natural and human-caused; and
- identify the impact of and prepare for changes in the enterprise environment, including the likelihood of new competitors entering the market or changes to government regulatory policy.

4.5.3 Risk Planning

The project risk plan balances the investment of the mitigation against the benefit for the project. The project team often develops an alternative method for accomplishing a project goal when a risk event has been identified that may frustrate the accomplishment of that goal. These plans are called contingency plans.

Risk Assessment: Assess the identified risks to determine their likelihood of occurrence and potential impact. Evaluate the severity of each risk and prioritize them based on their significance.

Risk Analysis: Analyze the root causes and consequences of identified risks. This step helps in understanding the underlying factors contributing to risks and their potential impact on the project or organization.

Risk Monitoring and Control: Establish a system to monitor identified risks throughout the project or organizational operations. Regularly review and update the risk register, track the effectiveness of risk response strategies, and make adjustments as necessary.

Contingency Planning: Develop contingency plans for high-impact risks that could severely disrupt the project or organization. These plans outline specific actions to be taken if the identified risks materialize, allowing for a prompt response and minimizing the potential damage.

Communication and Stakeholder Engagement: Ensure effective communication and engagement with stakeholders regarding risks, mitigation strategies, and progress. Keep stakeholders informed about potential risks and the actions being taken to address them. Documentation: Maintain a comprehensive record of identified risks, assessments, response plans, and outcomes. This documentation serves as a reference for future projects and helps in learning from past experiences.

4.6 Estimation

4.6.1 Cost Analysis

Cost analysis, also known as cost estimation or cost assessment, is the process of evaluating and determining the expenses associated with a project, activity, or business operation. It involves identifying, quantifying, and analyzing the various costs involved to make informed decisions, allocate resources effectively, and assess the financial feasibility of a particular endeavor.

The purpose of cost analysis is to provide a comprehensive understanding of the expenses involved in a project or operation. It helps in budgeting, planning, and evaluating the financial viability of a venture. Cost analysis involves examining both direct costs, which are directly attributable to the project or activity, and indirect costs, which are shared expenses or overhead costs.

Cost analysis is a crucial aspect of project management, financial planning, and business operations. It helps in making informed decisions, optimizing resource allocation, setting realistic budgets, and assessing the financial performance of an initiative. Effective cost analysis contributes to the overall success and profitability of projects and organizations.

When considering the cost analysis of Twitter sentiment analysis, several factors need to be taken into account.

Data Collection Cost: The first step in sentiment analysis is collecting the relevant data from Twitter. There are multiple ways to gather tweets, such as using the Twitter API or third-party services. Some APIs provide free access to a limited number of tweets, while others may require a subscription or payment based on the volume of data needed. The cost of data collection depends on the method chosen and the scale of the analysis.

Data Preprocessing Cost: Preprocessing involves cleaning and preparing the collected data for analysis. It typically includes removing irrelevant information, handling noise, normalizing text, and filtering out unwanted elements. The cost of data preprocessing depends on the complexity of the data and the tools or resources used for the task.

Sentiment Analysis Algorithm Cost: Implementing sentiment analysis requires selecting or developing an appropriate algorithm. There are various techniques available, such as rule-based approaches, machine learning models, or deep learning architectures. The cost associated with

this step includes the development or acquisition of the algorithm, training or fine-tuning the model, and any necessary computational resources.

Infrastructure and Computing Resources Cost: Sentiment analysis often requires significant computational resources, especially when processing a large volume of tweets in real-time. The cost of infrastructure depends on factors such as the number of tweets to analyze, the complexity of the sentiment analysis algorithm, and the desired response time. Cloud computing platforms like Amazon Web Services (AWS) or Google Cloud Platform (GCP) offer options for scalable and cost-effective infrastructure.

Human Annotation or Ground Truth: In some cases, sentiment analysis models require labeled data for training or evaluation purposes. This data is typically annotated by human annotators who assign sentiment labels to a sample set of tweets. The cost of human annotation depends on the number of tweets, the complexity of the annotation task, and the expertise of the annotators.

Maintenance and Updates: Sentiment analysis models may require periodic updates and maintenance to adapt to changing language patterns or to improve accuracy. These updates may include retraining the model with new data or making modifications to the algorithm. The cost associated with maintenance and updates should be considered for long-term sentiment analysis projects.

Integration and Deployment: Integrating the sentiment analysis system into an existing infrastructure or deploying it as a standalone application can involve additional costs. These costs can include software development, system integration, user interface design, and any necessary licensing fees.

The cost analysis of Twitter sentiment analysis can vary widely depending on the specific requirements of the project, the scale of analysis, and the resources available.

CHAPTER 6

SYSTEM TESTING AND MAINTANENCE

System testing is a type of software testing that evaluates the overall functionality and performance of a complete and fully integrated software solution. It tests if the system meets the specified requirements and if it is suitable for delivery to the end-users. This type of testing is performed after the integration testing and before the acceptance testing.

System Testing is a type of software testing that is performed on a complete integrated system to evaluate the compliance of the system with the corresponding requirements. In system testing, integration testing passed components are taken as input. The goal of integration testing is to detect any irregularity between the units that are integrated together. System testing detects defects within both the integrated units and the whole system. The result of system testing is the observed behavior of a component or a system when it is tested. System Testing is carried out on the whole system in the context of either system requirement specifications or functional requirement specifications or in the context of both. System testing tests the design and behavior of the system and also the expectations of the customer. It is performed to test the system beyond the bounds mentioned in the software requirements specification (SRS). System Testing is basically performed by a testing team that is independent of the development team that helps to test the quality of the system impartial. It has both functional and non-functional testing. System Testing is a black-box testing. System Testing is performed after the integration testing and before the acceptance testing.

System Testing Process: System Testing is performed in the following steps:

- Test Environment Setup: Create testing environment for the better quality testing.
- Create Test Case: Generate test case for the testing process.
- Create Test Data: Generate the data that is to be tested.
- Execute Test Case: After the generation of the test case and the test data, test cases are executed.
- Defect Reporting: Defects in the system are detected.
- Regression Testing: It is carried out to test the side effects of the testing process.
- Log Defects: Defects are fixed in this step.
- Retest: If the test is not successful then again test is performed.

Types of System Testing:

- Performance Testing: Performance Testing is a type of software testing that is carried out to test the speed, scalability, stability and reliability of the software product or application.
- Load Testing: Load Testing is a type of software Testing which is carried out to determine the behavior of a system or software product under extreme load.
- Stress Testing: Stress Testing is a type of software testing performed to check the robustness of the system under the varying loads.
- Scalability Testing: Scalability Testing is a type of software testing which is carried out to check the performance of a software application or system in terms of its capability to scale up or scale down the number of user request load.

Advantages of System Testing :

- The testers do not require more knowledge of programming to carry out this testing.
- It will test the entire product or software so that we will easily detect the errors or defects which cannot be identified during the unit testing and integration testing.
- The testing environment is similar to that of the real time production or business environment.
- It checks the entire functionality of the system with different test scripts and also it covers the technical and business requirements of clients.

6.1Maintenance

Software maintenance refers to the process of modifying and updating a software system after it has been delivered to the customer. This can include fixing bugs, adding new features, improving performance, or updating the software to work with new hardware or software systems. The goal of software maintenance is to keep the software system working correctly, efficiently, and securely, and to ensure that it continues to meet the needs of the users.

There are several key aspects of software maintenance :

1. Bug fixing: The process of finding and fixing errors and problems in the software.
2. Enhancements: The process of adding new features or improving existing features to meet the evolving needs of the users.

3. Performance optimization: The process of improving the speed, efficiency, and reliability of the software.
4. Porting and migration: The process of adapting the software to run on new hardware or software platforms.
5. Documentation: The process of creating, updating, and maintaining the documentation for the software, including user manuals, technical specifications, and design documents. Software maintenance is a critical part of the software development life cycle and is necessary to ensure that the software continues to meet the needs of the users over time. It is also important to consider the cost and effort required for software maintenance when planning and developing a software system.

Software maintenance is the process of modifying a software system after it has been delivered to the customer. The goal of maintenance is to improve the system's functionality, performance, and reliability and to adapt it to changing requirements and environments.

There are several types of software maintenance, including:

- Corrective maintenance: This involves fixing errors and bugs in the software system.
- Adaptive maintenance: This involves modifying the software system to adapt it to changes in the environment, such as changes in hardware or software, government polices, business rules.
- Perfective maintenance: This involves improving the functionality, performance, reliability and restructuring the software system to improve the changeability.
- Preventive maintenance: This involves taking measures to prevent future problems, such as optimization, updating documentation, reviewing and testing the system, and implementing
- Software maintenance can also involve upgrading the software to a new version or platform. This can be necessary to keep up with changes in technology and to ensure that the software remains compatible with other systems.
- The success of software maintenance depends on effective communication with stakeholders, including users, developers, and management. Regular updates and reports can help to keep stakeholders informed and involved in the maintenance process.

6.1.1 Need for Maintenance –
Software Maintenance must be performed in order to:

- Correct faults.
- Improve the design.
- Implement enhancements.
- Interface with other systems.
- Accommodate programs so that different hardware, software, system features, and telecommunications facilities can be used.
- Migrate legacy software.
- Retire software.
- Requirement of user changes.
- Run the code fast

CHAPTER 7

CONCLUSION

In conclusion, the application of sentiment analysis using COVID vaccination data on the RoBERTa model has proven to be a valuable tool for understanding public sentiment on Twitter. By leveraging the power of RoBERTa's language representation capabilities, we can gain insights into the opinions, attitudes, and concerns expressed by Twitter users regarding COVID vaccination.

Through this analysis, we have been able to classify tweets into different sentiment categories such as positive, negative, or neutral, thereby understanding the overall sentiment distribution surrounding COVID vaccination on Twitter. This has enabled us to identify prevailing trends, monitor shifts in public opinion, and track the effectiveness of vaccination campaigns.

The RoBERTa model's ability to learn the contextual meaning of words and sentences has allowed us to extract specific opinions and concerns expressed by Twitter users, providing deeper insights into the reasons behind positive or negative sentiments. This information can be immensely valuable for public health organizations, policymakers, and researchers in making data-driven decisions, designing effective communication strategies, and addressing concerns or misinformation related to COVID vaccination.

Furthermore, by identifying influential users or accounts on Twitter who actively discuss COVID vaccination and analyzing their sentiments, we have gained a better understanding of the impact of influential voices in shaping public opinion. This knowledge can assist in targeted communication efforts and engagement with key stakeholders.

It is important to note that the accuracy and reliability of sentiment analysis using the RoBERTa model depend on the quality and representativeness of the training data. Ensuring a diverse and unbiased dataset for training is crucial to obtain meaningful and actionable insights.

In summary, sentiment analysis on COVID vaccination using the RoBERTa model has proven to be a powerful approach for analyzing Twitter data. It allows us to understand the sentiment landscape, public perceptions, and emerging trends surrounding COVID vaccination on the Twitter platform. These insights contribute to informed decision-making, effective communication strategies, and the promotion of public health initiatives related to COVID vaccination.

CHAPTER 8

References/Bibliography

1.Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford.
Link: https://cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf

2.kaggel.com for downloading covid vaccination .csv file
www.kaggle.com

3.Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of Twitter data. Proceedings of the Workshop on Language in Social Media (LSM), 30-38. Link: https://www.aclweb.org/anthology/W11-1604.pdf

4.Goel, S., Gupta, S., & Shroff, G. (2020). Twitter Sentiment Analysis: A Comparative Study of Supervised Learning Approaches. International Journal of Computational Intelligence Research, 16(7), 1459-1472.
Link: https://www.ripublication.com/irph/ijcirv16n7spl_07.pdf

5.googleColab laboratory link
https://colab.research.google.com/drive/16JO4UMR422ICxBsH_q1tw4c7JPkK6cPU?usp=sharing