# VISVESVARAYA TECHNOLOGICAL UNIVERSITY

## BELAGAVI – 590018, Karnataka INTERNSHIP REPORT

#### ON

“SENTIMENT ANALYSIS USING NEWS-HEADLINES”

***Submitted in partial fulfilment for the award of degree(18CSI85)***

## BACHELOR OF ENGINEERING IN

## INFORMATION SCIENCE AND ENGINEERING

***Submitted by:***

**G SANJANA REDDY**

#### USN

**1JS20IS036**



Conducted at

**VARCONS TECHNOLOGIES PVT LTD**



# JSS ACADEMY OF TECHNICAL EDUCATION

**Department of Information Science and Engineering**

**Accredited by NBA, New Delhi**

**JSSATE-B Campus, Dr.Vishnuvardhan Rd Uttarahalli-Kengeri Main Road, JSS Campus Rd, Srinivaspura, Bengaluru, Karnataka 560060**

# JSS ACADEMY OF TECHNICAL EDUCATION

**Department of Information Science and Engineering**

**Accredited by NBA, New Delhi**

**JSSATE-B Campus, Dr.Vishnuvardhan Rd Uttarahalli-Kengeri Main Road, JSS Campus Rd, Srinivaspura, Bengaluru, Karnataka 560060**



**CERTIFICATE**

This is to certify that the Internship titled **“Sentiment Analysis using News Headlines”** carried out by **Miss. G Sanjana Reddy-(1JS20IS036),** a bonafide student of JSS Academy of Technical Education, in partial fulfillment for the award of **Bachelor of Engineering**, in **BRANCH** under Visvesvaraya Technological University, Belagavi, during the year 2023-2024. It is certified that all corrections/suggestions indicated have been incorporated in the report.

The project report has been approved as it satisfies the academic requirements in respect of Internship prescribed for the course Internship / Professional Practice (18CSI85)

#### Signature of Guide Signature of HOD Signature of Principal

**External Viva:**

Name of the Examiner Signature with Date

1)

2)

# D E C L A R A T I O N

I, **G Sanjana Reddy**, final year student of Branch, College Name - 560 082, declare that the Internship has been successfully completed, in Varcons Technologies Pvt Ltd. This report is submitted in partial fulfillment of the requirements for award of Bachelor Degree in Branch name, during the academic year 2023-2024.

Date : **20-09-2023** :

Place : **Bengaluru**

USN : **1JS20IS036**

NAME : **G Sanjana Reddy**

**OFFER LETTER**



# A C K N O W L E D G E M E N T

This Internship is a result of accumulated guidance, direction and support of several important persons. We take this opportunity to express our gratitude to all who have helped us to complete the Internship.

We express our sincere thanks to our Principal, for providing us adequate facilities to undertake this Internship.

We would like to thank our Head of Dept – branch code, for providing us an opportunity to carry out Internship and for his valuable guidance and support.

We would like to thank our (Lab assistant name) Software Services for guiding us during the period of internship.

We express our deep and profound gratitude to our guide, Guide name, Assistant/Associate Prof, for her keen interest and encouragement at every step in completing the Internship.

We would like to thank all the faculty members of our department for the support extended during the course of Internship.

We would like to thank the non-teaching members of our dept, forhelping us during the Internship.

Last but not the least, we would like to thank our parents and friends without whose constant help, the completion of Internship would have not been possible.

**G SANJANA REDDY**

**1JS20IS036**

# ABSTRACT

In today's information-driven world, staying attuned to global events is of paramount importance. News headlines, as concise summaries of news articles, offer a rich source for extracting sentiment and gaining insights into public perceptions. This research project leverages Natural Language Processing (NLP) techniques and several machine learning algorithms to conduct sentiment analysis on news headlines. We conducted an in-depth investigation into the efficacy of various classification algorithms, including K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Naive Bayes, and Random Forest, in determining sentiment polarity, such as positive, negative, or neutral.

The research involved a multi-stage process. Initially, we collected a diverse dataset of news headlines spanning different domains and time periods. We then preprocessed the text data by tokenization, stop-word removal, and stemming to ensure uniformity and enhance model performance. For feature extraction, we employed the TF-IDF (Term Frequency-Inverse Document Frequency) method to convert the textual data into numerical vectors suitable for machine learning models.

The core of our study revolved around the comparative analysis of classification algorithms. We implemented and fine-tuned each algorithm, considering various hyperparameters to optimize their performance. To evaluate the models, we employed metrics such as accuracy, precision, recall, and F1-score, and conducted cross-validation to ensure the robustness of our findings.

Our study carries implications beyond the realm of sentiment analysis. It contributes to the growing body of knowledge in NLP and machine learning applied to media analysis. Such analysis has applications in various fields, including finance, politics, and public opinion research. For example, financial analysts can employ sentiment analysis to gauge market sentiment and make informed investment decisions. Political researchers can track public sentiment towards political figures or policies to anticipate electoral outcomes and develop effective communication strategies.

Understanding public sentiment towards news also has broader societal implications. It can help media organizations improve content recommendation systems, tailoring news articles to readers' preferences and beliefs. Moreover, policymakers and media regulators can use sentiment analysis to monitor media influence and polarization, facilitating evidence-based decision-making.

# Table of Contents

|  |  |  |
| --- | --- | --- |
| **Sl no** | **Description** | **Page no** |
| 1 | Company Profile |  |
| 2 | About the Company |  |
| [3](https://1.bp.blogspot.com/-dODuK8N5h1Q/Wlnyb3V9HFI/AAAAAAAACL4/WxQtCJ1pM5wccDABg4wIrTBUB0vlikXQQCLcBGAs/s1600/poly1.jpg) | Introduction |  |
| [4](https://1.bp.blogspot.com/-dODuK8N5h1Q/Wlnyb3V9HFI/AAAAAAAACL4/WxQtCJ1pM5wccDABg4wIrTBUB0vlikXQQCLcBGAs/s1600/poly1.jpg) | System Analysis |  |
| [5](https://1.bp.blogspot.com/-dODuK8N5h1Q/Wlnyb3V9HFI/AAAAAAAACL4/WxQtCJ1pM5wccDABg4wIrTBUB0vlikXQQCLcBGAs/s1600/poly1.jpg) | Requirement Analysis |  |
| 6 | Design [Analysis](https://4.bp.blogspot.com/-IOOxgPaXMVc/Wlj3LWvcnjI/AAAAAAAACKE/UeTFYvAxDmUDel5UBjdifeWaApB3-dXVgCLcBGAs/s1600/img1.jpg) |  |
| 7 | [Implementation](https://4.bp.blogspot.com/-IOOxgPaXMVc/Wlj3LWvcnjI/AAAAAAAACKE/UeTFYvAxDmUDel5UBjdifeWaApB3-dXVgCLcBGAs/s1600/img1.jpg) |  |
| 8 | Snapshots |  |
| 9 | Conclusion |  |
| 10 | References |  |

[**CHAPTER**](https://1.bp.blogspot.com/-dODuK8N5h1Q/Wlnyb3V9HFI/AAAAAAAACL4/WxQtCJ1pM5wccDABg4wIrTBUB0vlikXQQCLcBGAs/s1600/poly1.jpg) **1** **COMPANY PROFILE**

# COMPANY PROFILE

## A Brief History of Company

Company, was incorporated with a goal ”To provide high quality and optimal Technological Solutions to business requirements of our clients”. Every business is a different and has a unique business model and so are the technological requirements. They understand this and hence the solutions provided to these requirements are different as well. They focus on clients requirements and provide them with tailor made technological solutions. They also understand that Reach of their Product to its targeted market or the automation of the existing process into e-client and simple process are the key features that our clients desire from Technological Solution they are looking for and these are the features that we focus on while designing the solutions for their clients.

Company is a Technology Organization providing solutions for all web design and development, MYSQL, PYTHON Programming, HTML, CSS, ASP.NET and LINQ. Meeting the ever increasing automation requirements, Sarvamoola Software Services. specialize in ERP, Connectivity, SEO Services, Conference Management, effective web promotion and tailor-made software products, designing solutions best suiting clients requirements.

we strive to be the front runner in creativity and innovation in software development through their well-researched expertise and establish it as an out of the box software development company in Bangalore, India. As a software development company, they translate this software development expertise into value for their customers through their professional solutions.

They understand that the best desired output can be achieved only by understanding the clients demand better. At our Company we work with them clients and help them to defiine their exact solution requirement. Sometimes even they wonder that they have completely redefined their solution or new application requirement during the brainstorming session, and here they position themselves as an IT solutions consulting group comprising of high caliber consultants.

They believe that Technology when used properly can help any business to scale and achieve new heights of success. It helps Improve its efficiency, profitability, reliability; to put it in one sentence ” Technology helps you to Delight your Customers” and that is what we want to achieve.

# [CHAPTER](https://1.bp.blogspot.com/-dODuK8N5h1Q/Wlnyb3V9HFI/AAAAAAAACL4/WxQtCJ1pM5wccDABg4wIrTBUB0vlikXQQCLcBGAs/s1600/poly1.jpg) 2 ABOUT THE COMPANY

1. **ABOUT THE COMPANY**

We are a Technology Organization providing solutions for all web design and development, Researching and Publishing Papers to ensure the quality of most used ML Models, MYSQL, PYTHON Programming, HTML, CSS, ASP.NET and LINQ. Meeting the ever increasing automation requirements, Varcons Technologies specialize in ERP, Connectivity, SEO Services, Conference Management, effective web promotion and tailor-made software products, designing solutions best suiting clients requirements. The organization where they have a right mix of professionals as a stakeholders to help us serve our clients with best of our capability and with at par industry standards. They have young, enthusiastic, passionate and creative Professionals to develop technological innovations in the field of Mobile technologies, Web applications as well as Business and Enterprise solution. Motto of our organization is to “Collaborate with our clients to provide them with best Technological solution hence creating Good Present and Better Future for our client which will bring a cascading a positive effect in their business shape as well”. Providing a Complete suite of technical solutions is not just our tag line, it is Our Vision for Our Clients and for Us, We strive hard to achieve it.

## Services provided by Varcons Technologies.

* Core Java and Advanced Java
* Research and Development/Improvise of ML Models
* Web services and development
* Dot Net Framework
* Python
* Selenium Testing
* Conference / Event Management Service
* Academic Project Guidance
* On The Job Training
* Software Training

# [CHAPTER](https://1.bp.blogspot.com/-dODuK8N5h1Q/Wlnyb3V9HFI/AAAAAAAACL4/WxQtCJ1pM5wccDABg4wIrTBUB0vlikXQQCLcBGAs/s1600/poly1.jpg) 3 INTRODUCTION

1. **INTRODUCTION**

## Introduction to ML

Machine learning is a subfield of artificial intelligence (AI) that focuses on creating computer systems capable of learning and improving from data without explicit programming. It involves developing algorithms and models that enable machines to recognize patterns, make predictions, and automate decision-making processes based on data.

At its core, machine learning relies on statistical techniques to enable computers to learn and adapt. It starts with a dataset containing input-output pairs, allowing the machine to identify relationships and patterns within the data. Machine learning algorithms use this information to build models that can generalize and make predictions on new, unseen data.

Multinomial Naive Bayes (MNB) is a classification algorithm often used in natural language processing (NLP) tasks, such as text classification and sentiment analysis. In the context of stock prediction, MNB is not typically used as the primary algorithm for making stock price forecasts. However, it can be employed in a related way to analyze and classify news headlines or textual information that might influence stock prices.

Support Vector Machine (SVM), in the context of stock prediction or any other classification task, is a machine learning algorithm used for binary and multiclass classification. SVM works by finding a hyper plane (a decision boundary) in the feature space that best separates different classes of data points.

K-Nearest Neighbors (KNN) is a supervised machine learning algorithm used for both classification and regression tasks. It is a simple and intuitive algorithm that relies on the idea that similar data points are close to each other in the feature space.

Logistic Regression is a classical statistical model used for binary and multi-class classification. Despite its name, it is not used for regression tasks but for predicting the probability of an instance belonging to a particular class.

The Random Forest Classifier is an ensemble learning method that combines the power of multiple decision trees to make accurate and robust predictions for classification tasks. It is a part of the broader family of decision tree-based algorithms.

## Problem Statement

"In today's information-saturated environment, assessing public sentiment towards news headlines has become a crucial task for a variety of stakeholders, from media outlets to financial analysts. The objective of this project is to develop an effective sentiment analysis system that accurately classifies news headlines into two categories: positive or negative.

The challenge lies in designing a robust sentiment classification model that can discern the nuanced emotional tone embedded within news headlines, thereby helping users quickly gauge the general sentiment conveyed by a headline. This project will utilize Natural Language Processing (NLP) techniques and machine learning algorithms to accomplish this task.

# [CHAPTER](https://1.bp.blogspot.com/-dODuK8N5h1Q/Wlnyb3V9HFI/AAAAAAAACL4/WxQtCJ1pM5wccDABg4wIrTBUB0vlikXQQCLcBGAs/s1600/poly1.jpg) 4 SYSTEM ANALYSIS

**4. SYSTEM ANALYSIS**

## Existing System

The existing sentiment analysis system that uses K-Nearest Neighbors (KNN) on a limited dataset for predicting sentiments in news headlines may have several defects or shortcomings, including:

**Limited Data Records:** One of the primary defects of the existing system is the use of a small dataset. KNN typically requires a substantial amount of data to make accurate predictions. Using a limited dataset can lead to poor generalization and may result in inaccurate sentiment predictions.

**Lack of Model Complexity:** KNN is a simple and intuitive algorithm, but it may not capture complex patterns and nuances present in news headlines. More sophisticated machine learning models like neural networks or ensemble methods (e.g., Random Forest or Gradient Boosting) often outperform KNN when dealing with textual data and sentiment analysis.

**Scalability Issues:** KNN can be computationally expensive, especially when applied to a large dataset. This could lead to scalability issues, making it less suitable for real-time or high-throughput sentiment analysis applications.

**Distance Metric Sensitivity:** KNN's performance heavily relies on the choice of distance metric. Inadequate selection of the distance metric can impact the quality of sentiment predictions. Identifying the optimal distance metric can be challenging.

**Imbalanced Data:** If the dataset used for training and testing the KNN model is imbalanced (i.e., contains significantly more instances of one sentiment class than another), the model may exhibit bias towards the majority class, leading to inaccurate predictions for the minority class.

## Proposed System

The enhanced approach provides a more substantial dataset of records and incorporates various machine learning models for sentiment analysis. By implementing tokenization, feature extraction, and stop word removal, it aims to improve accuracy in sentiment prediction for news headlines.

**Larger Dataset:** The expanded dataset ensures a more comprehensive representation of sentiments in news headlines, addressing the limitation of the previous system's smaller dataset.

**Multiple Machine Learning Models:** Utilizing various machine learning models allows for selecting the most suitable one, increasing adaptability and potentially outperforming the single KNN model in the previous system.

**Advanced Text Preprocessing:** Tokenization, feature extraction, and stop word removal enhance the system's ability to handle linguistic intricacies in news headlines.

**Model Evaluation:** Rigorous model evaluation with different metrics and cross-validation ensures the performance of chosen models is thoroughly assessed, enhancing reliability.

## Objective of the System

The objective of the system is to conduct sentiment analysis on news headlines. Specifically, it aims to accurately classify news headlines into two categories: positive or negative. This sentiment analysis serves several purposes:

**Understanding Public Perception:** The system seeks to provide insights into how the general public perceives news topics and events. By determining whether news headlines are associated with positive or negative sentiments, it helps gauge the prevailing sentiment trends.

**Enhancing Media Analysis:** Media organizations can use the system to gain a better understanding of how their headlines are received by the audience. It allows them to evaluate the impact of their news reporting on public sentiment.

**Supporting Decision-Making:** Financial analysts and investors can leverage sentiment analysis to gauge market sentiment. Positive news sentiment may correlate with bullish markets, while negative sentiment may indicate bearish trends.

**Improving Content Recommendation:** The system's sentiment analysis results can be integrated into content recommendation systems, helping tailor news articles and recommendations to readers' preferences and emotions.

**Media Influence Assessment:** Policymakers, media regulators, and researchers can use sentiment analysis to assess the influence of media on public opinion. This information can support evidence-based decision-making and policy development.

In summary, the system's primary objective is to provide accurate sentiment analysis of news headlines, allowing various stakeholders to gain insights into public sentiment, make informed decisions, and enhance media-related processes and applications.

# [CHAPTER](https://1.bp.blogspot.com/-dODuK8N5h1Q/Wlnyb3V9HFI/AAAAAAAACL4/WxQtCJ1pM5wccDABg4wIrTBUB0vlikXQQCLcBGAs/s1600/poly1.jpg) 5 REQUIREMENT ANALYSIS

**5. REQUIREMENT ANALYSIS**

## Hardware Requirement Specification

**Processor (CPU):**

A multi-core processor (e.g., Intel Core i5 or higher, AMD Ryzen) is recommended for parallel processing, especially if dealing with a large dataset or complex models.

For more extensive tasks and large-scale analysis, consider using server-grade CPUs with multiple cores or clusters of CPUs in a distributed computing environment.

**Memory (RAM):**

For larger datasets and more memory-intensive models, 16 GB or more of RAM is advisable.

Extremely large datasets may require even more substantial RAM, and high-performance computing environments may benefit from terabytes of RAM.

**Graphics Processing Unit (GPU) (optional but highly recommended for deep learning):**

If using deep learning models (e.g., neural networks), consider a dedicated GPU with CUDA support (e.g., NVIDIA GeForce or Quadro series).

GPUs significantly accelerate training and inference processes in deep learning tasks.

**Storage:**

Fast storage, such as SSDs (Solid State Drives), is recommended for quick data access and model training.

The storage capacity depends on the dataset size, but at least 256 GB of SSD storage is advisable for most use cases.

For handling large datasets, consider multiple SSDs or network-attached storage (NAS) solutions.

**Network Connectivity:**

A high-speed internet connection is essential for downloading and updating datasets, model libraries, and potential cloud-based services.

## Software Requirement Specification

**Python**:

Python is a popular language for machine learning and natural language processing tasks. You will need Python installed on your system.

**Integrated Development Environment (IDE):**

While not strictly required, using an IDE like Jupyter Notebook, PyCharm, or VSCode can make the development process more efficient.

**Libraries and Packages:**

Pandas: Pandas is essential for data manipulation and analysis. You can use it to load, manipulate, and clean your dataset.

Scikit-Learn: Scikit-Learn provides a wide range of machine learning algorithms and tools for tasks like model training, evaluation, and preprocessing.

NLTK (Natural Language Toolkit): NLTK is a powerful library for natural language processing tasks, including text tokenization, stop word removal, and lemmatization.

Text Vectorization Libraries: You'll likely need libraries like CountVectorizer or TfidfVectorizer from Scikit-Learn to convert text data into numerical vectors for machine learning models.

Visualization Libraries: Libraries like Matplotlib and Seaborn are useful for data visualization, allowing you to create charts and plots to visualize your results.

**Machine Learning Frameworks (if using deep learning):**

TensorFlow or PyTorch: These deep learning frameworks are essential if you plan to work with deep neural networks for sentiment analysis.

Additional Libraries and Tools:

Numpy: Numpy is useful for numerical operations and array manipulation in Python.

**Operating System Compatibility**:

Ensure that the software you choose is compatible with your operating system (Windows, macOS, Linux).

**Dependencies and Package Management:**

Use a package manager like pip or conda to install and manage the required libraries and packages efficiently.

**Cloud Services (optional):**

Consider cloud platforms like AWS, Google Cloud, or Microsoft Azure if you need scalable computing resources for large-scale sentiment analysis projects.

# [CHAPTER](https://1.bp.blogspot.com/-dODuK8N5h1Q/Wlnyb3V9HFI/AAAAAAAACL4/WxQtCJ1pM5wccDABg4wIrTBUB0vlikXQQCLcBGAs/s1600/poly1.jpg) 6 DESIGN ANALYSIS

1. **DESIGN & ANALYSIS**

Sentiment analysis, a subset of Natural Language Processing (NLP), is a powerful technique that allows us to automatically determine the sentiment or emotional tone expressed in text data, such as news headlines. In this design analysis, we will delve into the key aspects of a sentiment analysis project focused on news headlines.

**Project Overview:**

The primary objective of this project is to develop a robust sentiment analysis system capable of accurately classifying news headlines into either positive or negative sentiments. This system will have practical applications in fields such as media analysis, financial forecasting, and public opinion research.

**Project Design:**

1. **Data Collection and Preparation:**

* Data Sources: The first step in designing a sentiment analysis project is to gather a diverse and representative dataset of news headlines. These headlines can be obtained from various sources, including news websites, RSS feeds, or datasets available online.
* Data Preprocessing: Raw text data often requires extensive preprocessing. This includes tasks such as text cleaning, lowercasing, tokenization (breaking text into words or phrases), stop word removal, and stemming or lemmatization to reduce words to their root form. These steps help ensure consistency and improve the quality of the text data.

1. **Feature Extraction:**

* Vectorization: To make the text data suitable for machine learning models, we use techniques like Count Vectorization or TF-IDF (Term Frequency-Inverse Document Frequency) Vectorization. This process converts text into numerical vectors, where each word or phrase becomes a feature with associated values.

1. **Model Selection:**

* Machine Learning Models: Choosing the appropriate machine learning models is a critical aspect of project design. Common models for sentiment analysis include Multinomial Naive Bayes, Support Vector Machines (SVM), Random Forest, and deep learning models like recurrent neural networks (RNNs) or convolutional neural networks (CNNs).
* Model Evaluation: The selected models must be evaluated using relevant metrics such as accuracy, precision, recall, and F1-score. Cross-validation techniques help ensure the models generalize well to unseen data.

1. **Training and Testing:**

* Dataset Splitting: The dataset is typically divided into training, validation, and testing sets. The training set is used to train the model, the validation set helps fine-tune hyperparameters, and the testing set assesses the model's performance on unseen data.
* Model Training: The training process involves feeding the vectorized data into the chosen machine learning models. The models learn to identify patterns and associations between features and sentiment labels during this phase.

1. **Model Optimization:**

* Hyperparameter Tuning: The performance of machine learning models can often be improved by fine-tuning hyperparameters. Techniques like grid search or random search can be used to find optimal hyperparameter values.
* Feature Selection: Feature importance analysis helps identify which words or phrases are most influential in sentiment prediction. This analysis guides feature selection and dimensionality reduction, improving model efficiency.

1. **Results Analysis:**

* Confusion Matrix: Analyzing the confusion matrix helps understand the model's performance, including true positives, true negatives, false positives, and false negatives.
* Feature Importance: Understanding which features contribute most to sentiment predictions provides valuable insights into why certain headlines are classified as positive or negative.

1. **Integration and Deployment:**

* Application Integration: The sentiment analysis model can be integrated into various applications, such as content recommendation systems, financial analysis tools, or media monitoring dashboards.
* API Development: Creating an API (Application Programming Interface) allows other software systems to interact with the sentiment analysis service, enabling real-time sentiment classification.

1. **Maintenance and Monitoring:**

* Continuous Learning: To maintain the model's accuracy over time, periodic updates may be necessary to adapt to changes in language use and emerging sentiments.
* Monitoring: Implementing monitoring mechanisms to track the model's performance in production ensures that it continues to provide reliable results.

# [CHAPTE](https://1.bp.blogspot.com/-dODuK8N5h1Q/Wlnyb3V9HFI/AAAAAAAACL4/WxQtCJ1pM5wccDABg4wIrTBUB0vlikXQQCLcBGAs/s1600/poly1.jpg)R 7 IMPLEMENTATION

1. **IMPLEMENTATION**

Implementation is the stage where the theoretical design is turned into a working system. The most crucial stage in achieving a new successful system and in giving confidence on the new system for the users that it will work efficiently and effectively.

The system can be implemented only after thorough testing is done and if it is found to work according to the specification. It involves careful planning, investigation of the current system and it constraints on implementation, design of methods to achieve the change over and an evaluation of change over methods a part from planning.

Two major tasks of preparing the implementation are education and training of the users and testing of the system. The more complex the system being implemented, the more involved will be the system analysis and design effort required just for implementation.

The implementation phase comprises of several activities. The required hardware and software acquisition is carried out. The system may require some software to be developed. For this, programs are written and tested. The user then changes over to his new fully tested system and the old system is discontinued.

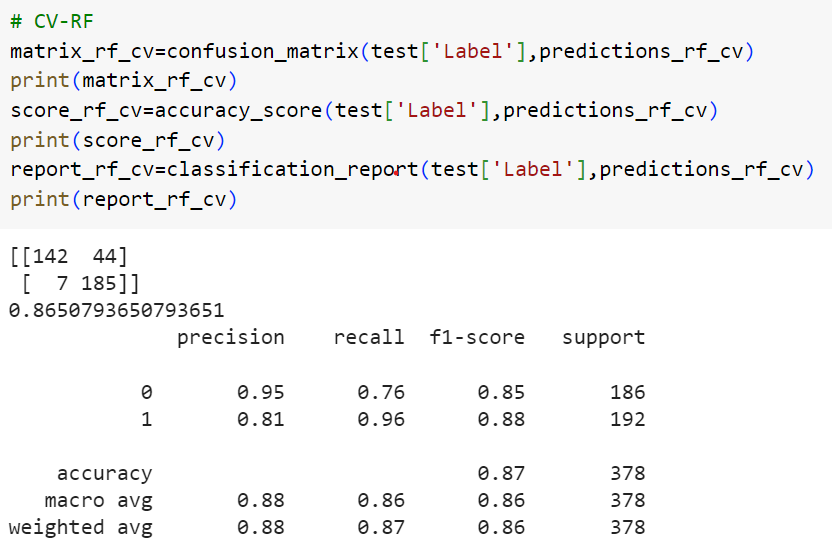
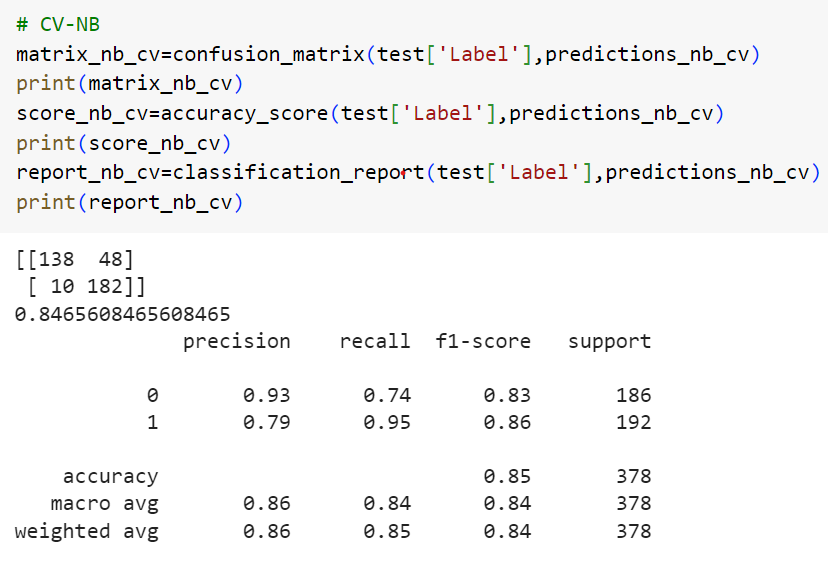
## TESTING

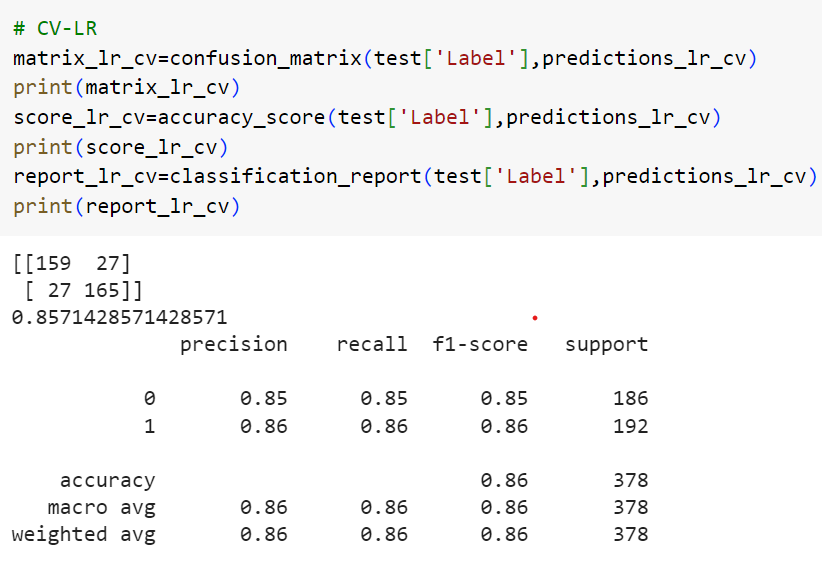
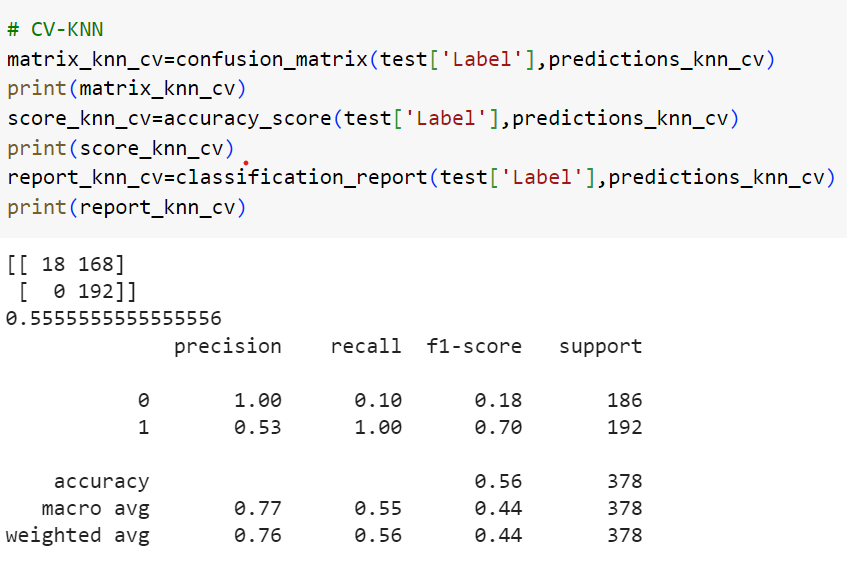
The testing phase is an important part of software development. It is the Information zed system will help in automate process of finding errors and missing operations and also a complete verification to determine whether the objectives are met and the user requirements are satisfied. Software testing is carried out in three steps:

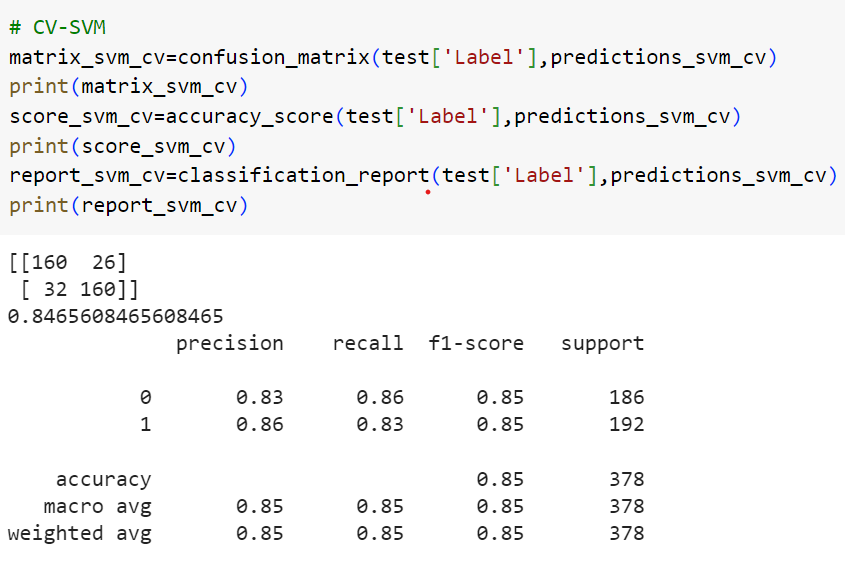
1. The first includes unit testing, where in each module is tested to provide its correctness, validity and also determine any missing operations and to verify whether the objectives have been met. Errors are noted down and corrected immediately.
2. Unit testing is the important and major part of the project. So errors are rectified easily in particular module and program clarity is increased. In this project entire system is divided into several modules and is developed individually. So unit testing is conducted to individual modules.
3. The second step includes Integration testing. It need not be the case, the software whose modules when run individually and showing perfect results, will also show perfect results when run as a whole.

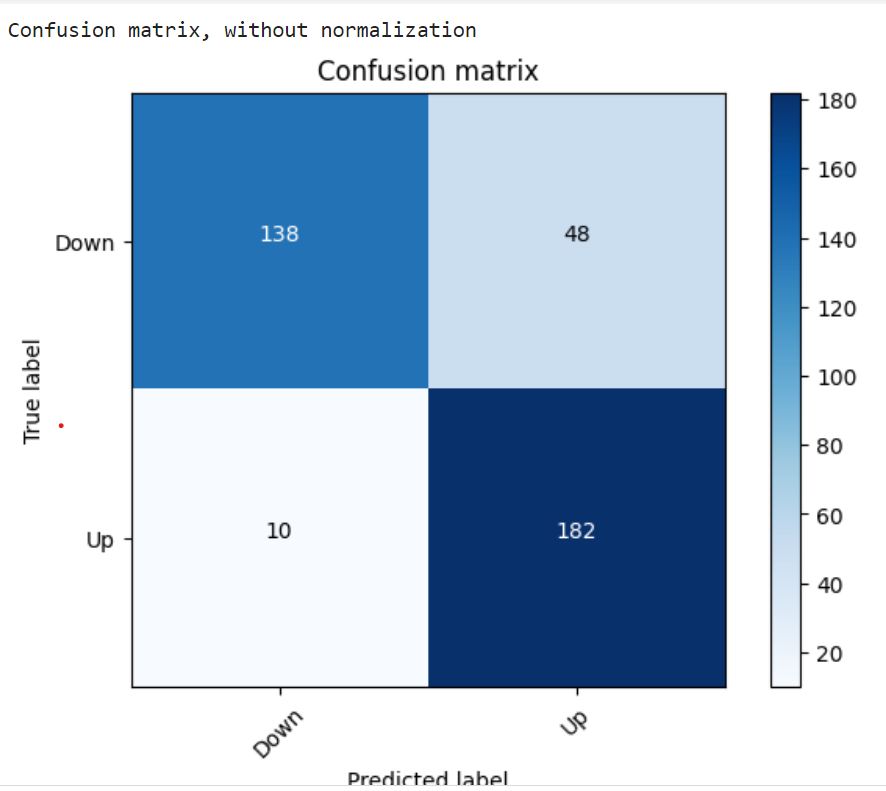
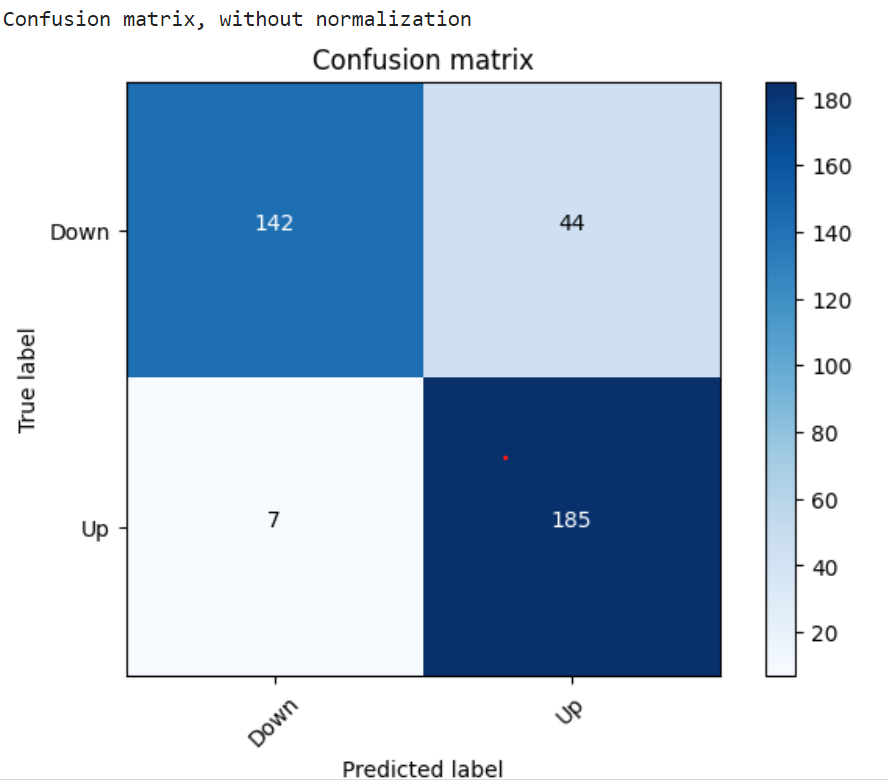
# [CHAPTE](https://1.bp.blogspot.com/-dODuK8N5h1Q/Wlnyb3V9HFI/AAAAAAAACL4/WxQtCJ1pM5wccDABg4wIrTBUB0vlikXQQCLcBGAs/s1600/poly1.jpg)R 8 SNAPSHOTS

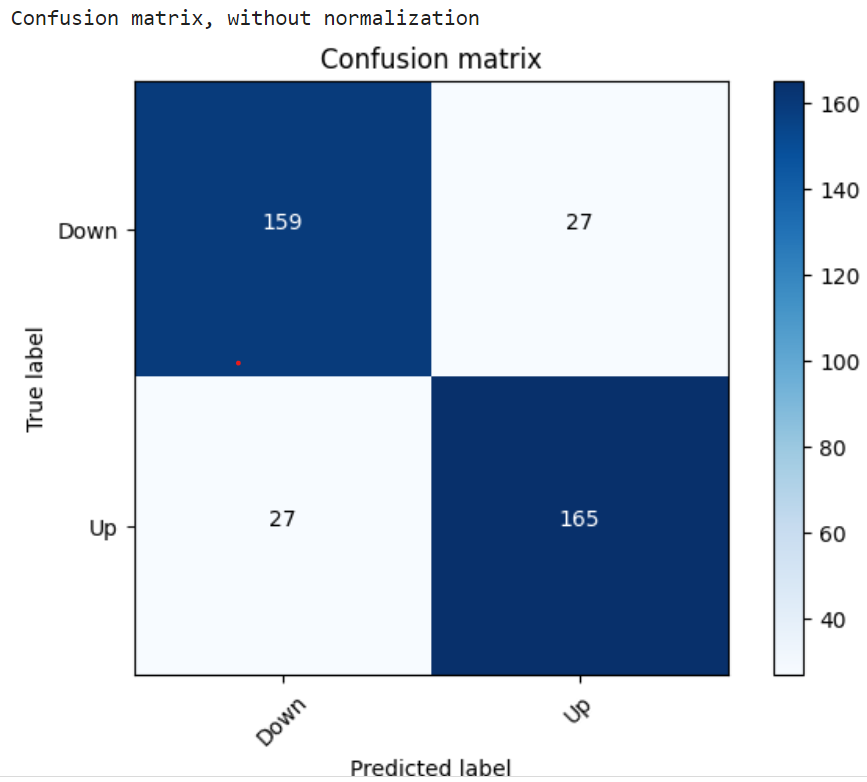
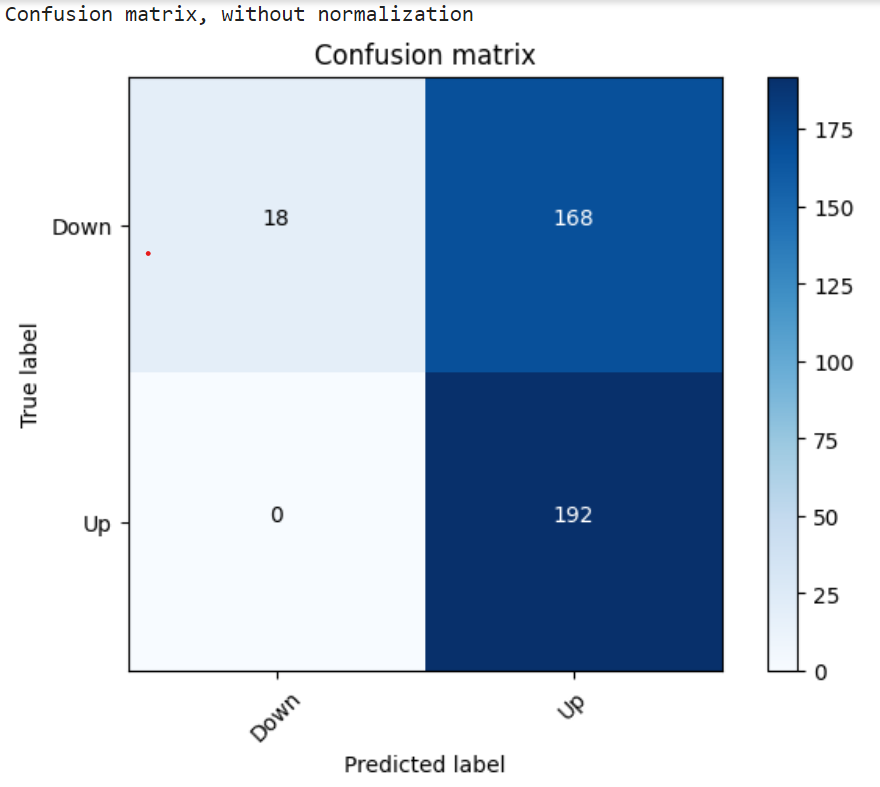
* 1. **SNAPSHOTS**

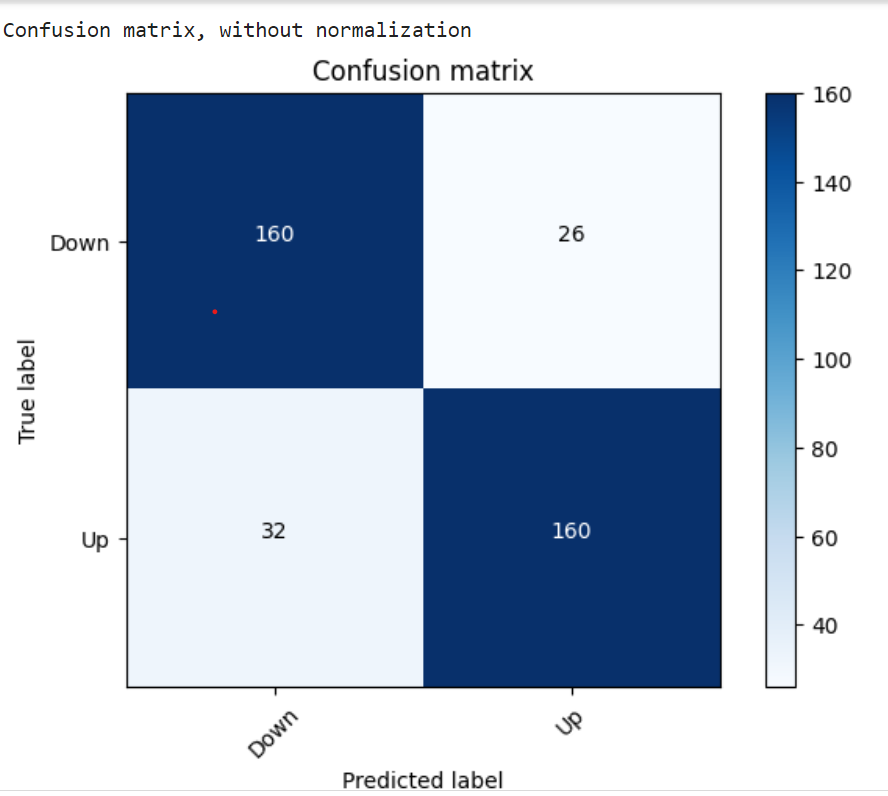
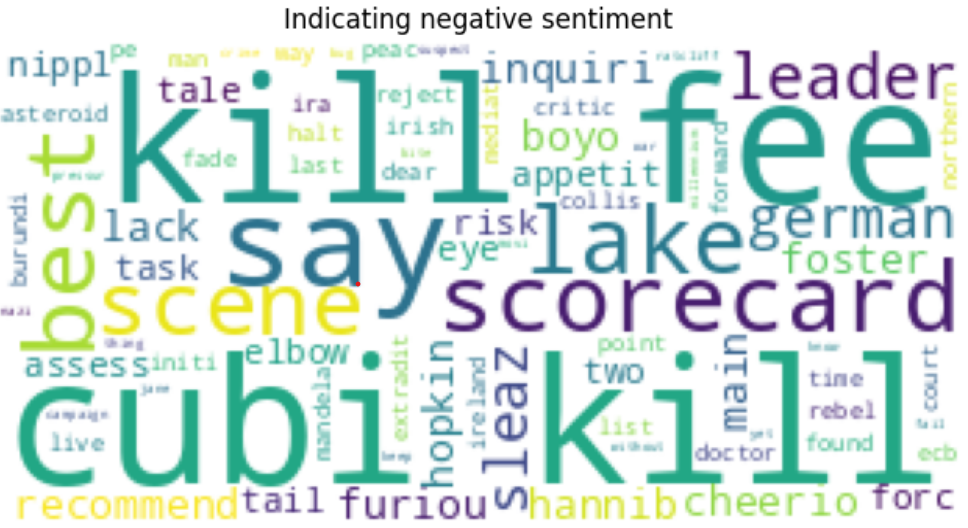
****

****





# [CHAPTE](https://1.bp.blogspot.com/-dODuK8N5h1Q/Wlnyb3V9HFI/AAAAAAAACL4/WxQtCJ1pM5wccDABg4wIrTBUB0vlikXQQCLcBGAs/s1600/poly1.jpg)R 9 CONCLUTION

* 1. **CONCLUTION**

## In conclusion, when conducting sentiment analysis based on news headlines, it is essential to effectively evaluate the performance of sentiment classification model. To accomplish this, you can utilize a confusion matrix in conjunction with various performance metrics such as accuracy and the F1 score. Here are the key takeaways:

## Confusion Matrix: A confusion matrix is a tabular representation that summarizes the performance of a classification model by tallying true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). It provides insight into how well your model is making predictions and identifies areas where errors may occur.

## Accuracy: Accuracy measures the proportion of correctly classified news headlines (both positive and negative) out of the total predictions. It offers an overall assessment of your model's correctness and performance.

## F1 Score: The F1 score is a harmonic mean of precision and recall. It strikes a balance between minimizing false positives (precision) and minimizing false negatives (recall). A high F1 score indicates a model that achieves both high precision and high recall, making it valuable in scenarios where you require a robust performance.

## By effectively analyzing the performance of your sentiment classification model through the use of a confusion matrix, accuracy measurement, and the F1 score, you can gain valuable insights into the effectiveness of your approach to stock sentiment analysis. These metrics provide a comprehensive evaluation of your model's predictive capabilities and help identify areas for improvement.

# REFERENCE

[**https://www.youtube.com/live/HsI5YF6uxGs?si=FP0RAqhWDmWI5HaZ**](https://www.youtube.com/live/HsI5YF6uxGs?si=FP0RAqhWDmWI5HaZ)

[**https://github.com/sree-cola/sentimentanalysisusingnewsheadlines**](https://github.com/sree-cola/sentimentanalysisusingnewsheadlines)

[**https://www.datacamp.com/tutorial/text-analytics-beginners-nltk**](https://www.datacamp.com/tutorial/text-analytics-beginners-nltk)

[**https://www.analyticsvidhya.com/blog/2022/07/sentiment-analysis-using-python/**](https://www.analyticsvidhya.com/blog/2022/07/sentiment-analysis-using-python/)