VISVESVARAYA TECHNOLOGICAL UNIVERSITY
"Jnana Sangama", Belgavi- 590018

AN INTERNSHIP REPORT ON

"TWITTER SENTIMENT ANALYSIS"

Submitted in partial fulfillment of the requirement for the award of the degree of

BACHELOR OF ENGINEERING
In

COMPUTER SCIENCE & ENGINEERING

Submitted by

GAURAV KUMAR BHATT (1RR19CS037)

Internship carried out at
VARCONS TECHNOLOGIES Pvt Ltd

Internal Guide	**External Guide**
Prof . Jayashree Badiger	**Mr. Aditya S K**
Professor	CTO
Dept of CSE, RRCE	Varcons, Bangalore

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

RAJARAJESWARI COLLEGE OF ENGINEERING
MYSORE ROAD, BENGALURU-560074
(An ISO 9001:2008 Certified Institute)
(2022-2023)

RAJARAJESWARI COLLEGE OF ENGINEERING
MYSORE ROAD, BENGALURU-560074
(An ISO 9001:2008 Certified Institute)
(Affiliated to Visvesvaraya Technological University, Belgavi)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

CERTIFICATE

Certified that Internship entitled

"TWITTER SENTIMENT ANALYSIS"

Carried out by

GAURAV KUMAR BHATT (1RR19CS037)

The student of **"Rajarajeswari College of Engineering"** in partial fulfillment for the award of the degree of **Bachelor Of Engineering** in **Computer Science & Engineering** of the Visvesvaraya Technological University, Belgaum during the year **2022–2023.** It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report deposited in the departmental library. The Internship report has been approved as it satisfies the academic requirements in respect of Internship work prescribed for the Eighth semester.

……………................   ………………….....   ……………………….

Signature of guide   **Signature of HOD**   **Signature of Principal**

[Prof.Jayashree Badiger]   **[Dr. S Usha Sakthivel]**   **[Dr. R. BALAKRISHNA]**
 Prof., Dept. of CSE   Prof. & HOD, Dept. of CSE   Principal
RRCE, Bangalore   RRCE, Bangalore   RRCE, Bangalore

External Viva-Voce

Examiners: *Signature:*

   1. _____
   2. _____

DECLARATION

I, Gaurav Kumar Bhatt (**1RR19CS037**), student of B.E in Computer Science and Engineering, **Rajarajeswari College of Engineering,** Bengaluru, hereby declare that the internship work entitled **"Twitter Sentiment Analysis"** submitted to the **Visvesvaraya Technological University** during the academic year 2022-2023 is record of an original work done by us under the guidance of **Prof.Jayashree Badiger**, Associate Prof., Department of Computer Science and Engineering, Rajarajeswari College of Engineering, Bengaluru. This internship work is submitted in partial fulfillment of the requirements for the award of the degree of **Bachelor of Engineering** in **Computer Science and Engineering.** The results embodied in this report have not been submitted to any other university or institute for any degree.

(**GAURAV KUMAR BHATT**)
(**1RR19CS037**)

ACKNOWLEDGEMENT

COMPANY CERTIFICATE

Varcons Technologies Pvt Ltd
Communicate. Collaborate. Create

CERTIFICATE OF INTERNSHIP

This is to certify that **Gaurav Kumar Bhatt** whose USN is **1RR19CS037**, has completed their **Machine Learning With Python(Research Based)** Internship organised and handled by **Varcons Technlogoies Pvt Ltd** from **23rd August, 2022** to **27th September, 2022**.

The person to whom this certificate is addressed to has worked on a project titled Stockport | Predictive Sentiment Analysis, As part of the project, They designed the Machine Learning Model, Demonstrated and tested the working of the Model, Prepared a report highligting its flaws by understanding the design briefs and client Specifications that were provided in the Proposal.

During the course of the internship, they demonstrated good design skills with a self-motivated attitude to learning new things. Their performance exceeded expectations and was able to complete the project successfully on time.

To verify this certificate, CLICK HERE ✅

Spoorthi C
DIRECTOR
VARCONS TECHNLOGOIES PVT LTD
AUTHORIZED SIGNATORY

18 M G Road, Ulsoor, Bangalore-560001

www.varconstech.com
contact@varconstech.com

This certificate was generated using CERTIEFY.COM

CONTENTS

LIST OF FIGURES

Chapter 1

Company Profile

1.1 Introduction

The following report describes the activities carried out during the four weeks of full-time internship at Varcons Technologies Private Limited. The document contains information regarding the organization and the responsibilities performed throughout July-August 2018.

The first part of the report tells about the organization, followed by the working plan initially agreed upon by the federation and approved by the University of VTU as a suitable internship. Following, it proceeds to describe in some details about the department and the most relevant task performed. Finally, the report wraps up with the outcome.

Build solutions to fulfill societal needs by integrating contemporary technologies by devoting all our resources by leveraging skills, processes, and resource efficiencies through meritocracy and thereby delighting customers and stakeholders.

Vision

The customer's first choice as a technology provider. Envision is a leader in the information technology servicesarena. Growth as a leader in IT services shall be guided by organizational values and customer needs.

Mission

Collaborate with our clients to provide them with the best Technological solution hence creating Good Present andBetter Future for our clients which will bring a cascading positive effect in their business shape as well.

1.2 Overview of the Organization

- Varcons is a global software application development and IT services company.

- Providing services to several "happy customers" across the globe.

- Varcons Technologies, strive to be the front runner in creativity and innovation in software development through its well-researched expertise and establish it as an out-of-the-box software development company in Bangalore, India

- Deep industry and functional expertise powered by leading technology practices. Contemporary global delivery model to help address a customer need and enhance their business performance, productivity, and, profitability.

1.3 About the Company

Varcons Technologies 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, Varcons Technologies specialize in ERP, Connectivity, SEO Services, Conference Management, effective web promotion, and tailor-made software products, designing solution best suiting client's requirements. The organization where they have the right mix of professionals as a stakeholder to help us serve our clients to the best of our capability and with 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 solutions. The motto of our organization is to "Collaborate with our clients to provide them with the best Technological solution hence creating Good Present and Better Future for our clients which will bringa cascading 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, and we strive hard to achieve it.

1.4 Services by Varcons Technologies Private Limited

1.4.1 Web development

1.4.2 Core Java and Advanced Java

1.4.3 Python

1.4.4 Dot Net Framework

1.4.5 Selenium Testing

1.4.6 Software Training

1.5 Search Engine Optimization

Search engine optimization is the process of increasing the quality and quantity of website traffic by increasing the visibility of a website or a web page to users of a web search engine.

SEO refers to the improvement of *unpaid* results and excludes direct traffic/visitors and the purchase of paidplacement. SEO may target different kinds of searches, including image search, video search, academic search, news search, and industry-specific vertical search engines.

Optimizing a website may involve editing its content, adding content, and modifying HTML and associated coding to both increase its relevance to specific keywords and remove barriers to the indexing activities of search engines like Google, Yahoo, etc. Promoting a site to increase the number of backlinks, or inbound links is another SEO tactic. By May 2015, mobile search had surpassed desktop search.

As an Internet Marketing Strategy, SEO considers how search engines work, the computer-programmed algorithms that dictate search engine behavior, and what people search for the actual search terms or keywords typed into search engines that are preferred by their targeted audience. SEO is performed because a website ranks in the searchengine results page (SERP). These visitors can then be converted into customers. SEO differs from local search engine optimization in that the latter is focused on optimizing a business's online presence so that its web pages will be displayed by search engines when a user enters a local search for its products or services. The former instead is more focused on national or international searches.

1.6 Application Development

Drive efficiency and effectiveness with agile application solutions from this company. The proven methodology ensures you get products and solutions faster, at a reduced cost, and best quality assurance.

It is the process of creating a computer program or a set of programs to perform the different tasks that a business requires. Every app-building process follows the same steps: gathering requirements, designing prototypes, testing, implementing, and integrating.

1.7 Company Contact

Website

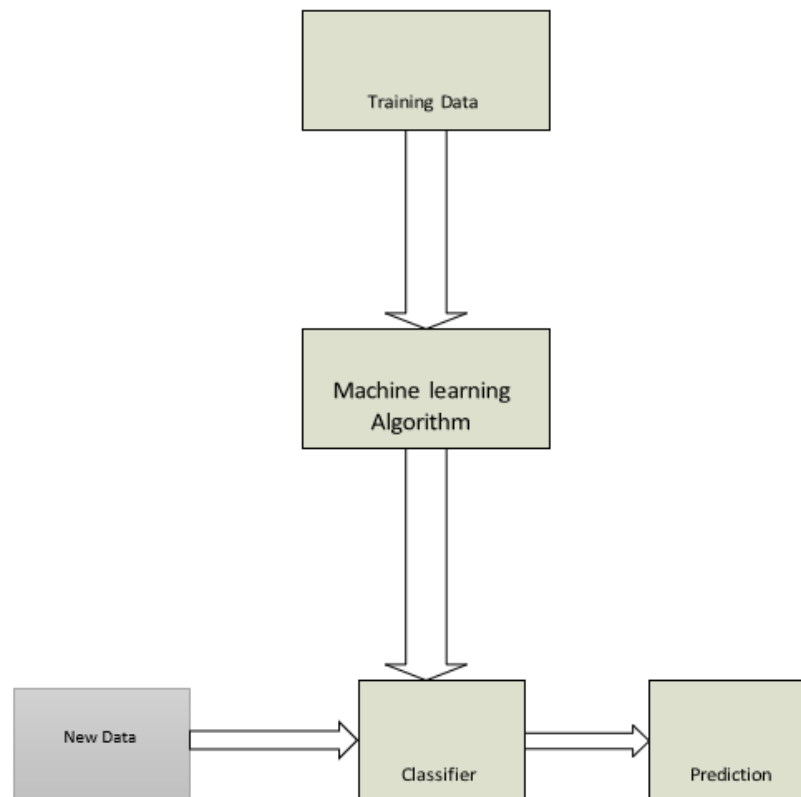https://www.varconstech.com

Address:

.

Twitter Sentiment Analysis

Chapter 2
Tasks Performed

2.1 System Architecture:
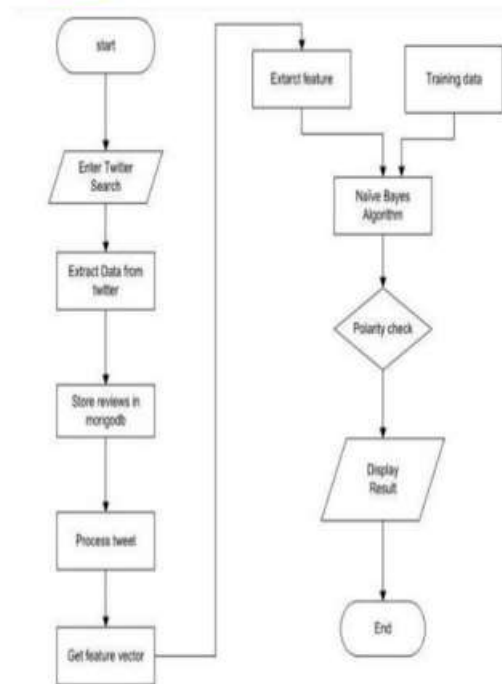
2. SYSTEM FLOW DIAGRAM:-

5.2(a) System Flow Diagram for Sentiment Analysis using Twitter API

Figure: 2.1

UML Diagrams

Flow Chart:-

Flow Chart for Sentiment Analysis using Twitter Sentiment.

2.2 Data acquisition and cleaning:

Data acquisition and cleaning are two essential steps in the data pre-processing phase of a machine learning project. Let's take a closer look at what each of these steps involves:

Data Acquisition:

Data acquisition is the process of collecting raw data from various sources such as databases, APIs, websites, or other data repositories. The quality of the data is critical as it can significantly impact the accuracy and performance of the machine learning model. Here are some steps involved in data acquisition:

- Identify Data Sources: Identify the data sources that are relevant to the problem you are trying to solve. This could be data from internal databases, public datasets, or data obtained through web scraping.
- Collect Data: Collect data from the identified sources using appropriate tools and techniques.
- Data Integration: Integrate the collected data from different sources into a single dataset.
- Data Transformation: Transform the data into a suitable format that can be analyzed and modeled.

Data Cleaning:

Data cleaning is the process of identifying and correcting errors, inconsistencies, and inaccuracies in the collected dataset. It is essential to perform data cleaning to ensure that the model is trained on high-quality data. Here are some steps involved in data cleaning:

- Identify Missing Values: Identify missing data points in the dataset and determine the best way to handle them. This could involve imputing missing values using mean, median, or mode, or removing the entire row or column with missing values.
- Identify Outliers: Identify outliers in the data and determine if they need to be removed or kept in the dataset.
- Handle Duplicate Data: Identify and remove duplicate data points in the dataset.
- Handle Inconsistencies: Identify inconsistencies in the dataset, such as typos or misspellings, and correct them.

- Handle Irrelevant Data: Identify irrelevant data points that do not contribute to the problem being solved, and remove them from the dataset.

2.3 Data Visualization:

Data visualization refers to the use of visual elements such as charts, graphs, maps, and other graphical representations to communicate complex data in an easy-to-understand manner. The main goal of data visualization is to enable people to easily and quickly interpret and comprehend large amounts of data, patterns, and relationships within the data.

Effective data visualization uses various design techniques to display data in a clear and intuitive manner. For example, the use of colors, shapes, sizes, and positioning of visual elements can help emphasize important points or trends in the data. Good data visualization also takes into account the target audience, purpose of the visualization, and the type of data being presented.

Data visualization has many practical applications, including business intelligence, scientific research, finance, marketing, and many other fields. It helps decision-makers gain insights and identify patterns and trends that might otherwise be hidden in large volumes of data.

2.4 Modelling:

Data modeling is the process of creating a conceptual representation of data and its relationships, in order to facilitate understanding, communication, and organization of that data. The primary purpose of data modeling is to design a blueprint of the data that an organization or system needs to operate effectively. It is used to identify the data requirements, business rules, and constraints that must be considered when developing a database or information system.

Data modeling typically involves several steps, including:

- Requirements gathering: understanding the business objectives, identifying the data entities, and determining the relationships between them.
- Conceptual modeling: creating a high-level overview of the data and its relationships, often using graphical representations such as entity-relationship diagrams (ERDs).
- Logical modeling: refining the conceptual model to create a more detailed representation of the data, including the attributes of each entity and the relationships between them.

- Physical Modeling: mapping the logical model to the physical storage mechanisms that will be used to store and retrieve the data.

- Implementation: building the database or information system based on the model, and verifying that it meets the requirements of the organization or system.

Data modeling is important because it helps to ensure that data is organized, consistent, and accurate, which makes it easier to manage and analyze. It also helps to facilitate communication between stakeholders, and can improve the accuracy and efficiency of software development projects.

2.5 Testing:

Testing is an important process that ensures the accuracy, performance, and reliability of the machine learning models. The main goal of testing is to verify that the models are producing the expected outputs and that they can handle different scenarios and edge cases.

There are different types of testing that can be applied in ML projects, including:

- Unit Testing: This involves testing individual components of the ML system, such as the data pre-processing pipeline, feature extraction, and model training. Unit testing can be automated using tools like pytest or unittest.

- Integration Testing: This involves testing the interaction between different components of the ML system, such as the data pipeline, the model, and the output generation. Integration testing can be done manually or automated using tools like Jenkins or Travis CI.

- Performance Testing: This involves testing the performance of the ML model under different conditions, such as different data sizes, different hardware configurations, and different hyper parameters. Performance testing can be automated using tools like PyTorch Profiler or TensorFlow Benchmark.

Overall, testing is essential to ensure the quality and reliability of the machine learning models, and it should be an integral part of the development process. Proper testing can help to identify and fix errors early, improve the overall performance of the model, and increase the confidence in the results produced by the model.

2.6 Comparison and Measurement:

Comparison and measurement refer to the process of evaluating the performance of different models or algorithms that are trained to solve a particular problem.

Comparison involves comparing the performance of multiple models on a specific task or dataset. This comparison could involve measuring metrics such as accuracy, precision, recall, F1 score, or area under the curve (AUC) to determine which model performs best.

Measurement involves quantifying the performance of a single model or algorithm on a specific task or dataset. This measurement could involve calculating the same metrics mentioned above to understand how well the model is performing.

Both comparison and measurement are critical components of any machine learning project, as they help to assess the effectiveness of the model and identify areas for improvement. By conducting thorough comparisons and measurements, machine learning practitioners can determine which models or algorithms are best suited for a given task and optimize their performance to achieve the desired outcomes.

The measurement of a property may be categorized by the following criteria: type, magnitude, unit, and uncertainty.[citation needed] They enable unambiguous comparisons between measurements. The level of measurement is a taxonomy for the methodological character of a comparison. For example, two states of a property may be compared by ratio, difference, or ordinal preference. The type is commonly not explicitly expressed, but implicit in the definition of a measurement procedure. The magnitude is the numerical value of the characterization, usually obtained with a suitably chosen measuring instrument. A unit assigns a mathematical weighting factor to the magnitude that is derived as a ratio to the property of an artifact used as standard or a natural physical quantity. An uncertainty represents the random and systemic errors of the measurement procedure; it indicates a confidence level in the measurement. Errors are evaluated by methodically repeating measurements and considering the accuracy and precision of the measuring instrument.

Chapter 3

Technical and Non-Technical Activities

3.1 Non-technical activities:

- **Research:** First and foremost we tried to learn more about Parkinson's disease, its causes, symptoms, and treatments. This knowledge helped us to better understand the dataset we were working with and provide context for our analysis. This knowledge also helped u s analyze the Parkinson's dataset more effectively and identify key insights that could inform future research.

- **Data visualization:** Create visually appealing and easy-to-understand graphs and charts to present your findings. This can be an effective way to communicate complex information to non-technical audiences. Creating clear and concise data visualizations can help you communicate complex information more effectively. We have used tools like charts, graphs, and info graphics to present your findings in a visually appealing and easy-to-understand way. This can be particularly helpful when presenting our results to non-technical audiences.

- **Communication:** Practice your communication skills by presenting your findings to your colleagues or supervisors. This will not only help you improve your presentation skills but also allow you to receive feedback and suggestions for future research.

- **Collaboration:** Work with other interns or colleagues to create interdisciplinary projects that combine technical and non-technical skills. For example, you could collaborate with a graphic designer to create data visualizations, or work with a writer to create compelling content about Parkinson's disease.

- **Oral Communication Skills:** Making a speech or presentation before a large group of people such as manager, customer, and your peers. The need to effectively communicate technical information about our particular discipline to others and the need to maintain a friendly work environment. We must have a good communication skill in any organization

today. My internship taught me of how to communicate effectively both within the team and outside the team.

- **How to Work with Co-Workers:** Everyone is different and getting along with co-workers isn't always easy. We often come across unusual characters at work. My internship helped me learn how to interact with different types of people in a professional setting, this can take some getting used to but will help me to settle in at new jobs faster.

- **Time Management Skills:** To be an effective employee and successful we must be able to manage our time efficiently. Most people have fairly good time management skills from school or university experience. In the workplace however, it can be quite different, often with a much higher tempo of tasks needing to be completed and usually unforeseen variations during the day. My internship taught me techniques of how to manage time better.

3.2 Technical Activities:

- **Data preprocessing:** This involves cleaning, organizing, and transforming the raw Parkinson's dataset to make it more suitable for machine learning analysis. This may include tasks such as removing duplicates, handling missing values, and scaling or normalizing the data.

- **Feature selection:** This involves identifying which features (i.e., variables or columns) in the Parkinson's dataset are most relevant for predicting the outcome of interest (e.g., Parkinson's diagnosis). This may involve using statistical techniques such as correlation analysis or machine learning algorithms such as decision trees.

- **Model selection and training**: This involves selecting a machine learning algorithm that is appropriate for the Parkinson's dataset and training it on the data to create a predictive model. This may involve using techniques such as cross-validation to evaluate the performance of different algorithms and hyperparameter tuning to optimize the model's accuracy.

- **Evaluation and testing:** This involves evaluating the performance of the trained model using a separate test dataset. This may involve using metrics such as accuracy, precision, recall, or F1 score to evaluate the model's performance, as well as using techniques such as confusion matrices or ROC curves to visualize the model's performance.

- **Documentation:** This involves documenting the entire machine learning project, including the data preprocessing steps, feature selection process, model selection and training process, evaluation and testing results, and deployment process. This may involve creating detailed reports, diagrams, and code documentation to ensure that the project can be replicated and maintained by others in the future.

- Technology is the application of knowledge for achieving practical goals in a reproducible way. The word technology can also mean the products resulting from such efforts, including both tangible tools such as utensils or machines, and intangible ones such as software.

- Technology plays a critical role in science, engineering, and everyday life. Technological advancements have led to significant changes in society. The earliest known technology is the stone tool, used during prehistoric times, followed by the control of fire, which contributed to the growth of the human brain and the development of language during the Ice Age.

- The invention of the wheel in the Bronze Age allowed greater travel and the creation of more complex machines. More recent technological inventions, including the printing press, telephone, and the Internet, have lowered barriers to communication and ushered in the knowledge economy.

- While technology contributes to economic development and improves human prosperity, it can also have negative impacts like pollution and resource depletion, and can cause social harms like technological unemployment resulting from automation. As a result, there are ongoing philosophical and political debates about the role and use of technology, the ethics of technology, and ways to mitigate its downsides

Chapter 4

OUTCOME OF INTERNSHIP

4.1 Learning from Internship Program:

The Internship program was beneficial for us. It helped us in improving our various technical skills and enhanced my knowledge in new areas.

- We gained knowledge in the domain of AIML, the various issues involved and mechanisms in the systems etc.
- By studying Machine leaning, we also learnt that how machine learning functions and what are the various issues one need to be aware of while analyzing data.
- We worked on a data analysis project for the first time we got introduced to this important area of data analysis and the importance of the same.
- We brushed up our knowledge on Python as a coding language and also the required libraries for the purpose of analyzing data and representing the same by using python libraries like Matplotlib.

Work Experience

Our Internship was quite satisfactory in terms of work environment. The team with which we worked with was very friendly and helped us in solving our problems. New experiences include An internship is a period of work experience offered by an organization for a limited period of time.Once confined to medical graduates, internship is used practice for a wide range of placements in businesses, non-profit organizations and government agencies. They are typically undertaken by students and graduates looking to gain relevant skills and experience in a particular field. Employers benefit from these placements because they often recruit employees from their best interns, who have known capabilities, thus saving time and money in the long run. Internships are usually arranged by third-party organizations that recruit interns on behalf of industry groups. Rules vary from country to country about when interns should be regarded as employees. The system can be open to exploitation by unscrupulous employers

Internships exist in a wide variety of industries and settings. An internship can be paid, unpaid, or partially paid (in the form of a stipend). Internships may be part-time or full-time and are usually flexible with students' schedules. A typical internship lasts between one and four months, but can be shorter or longer, depending on the organization involved. The act of job shadowing may also constitute interning. Insights: Many large corporations, particularly investment banks, have "insights" programs that serve as a pre-internship event numbering a day to a week, either in person or virtually. Paid internships are common in professional fields including medicine, architecture, science, engineering, law, business (especially accounting and finance), technology, and advertising. [citation needed] Work experience internships usually occur during the second or third year of schooling. This type of internship is to expand an intern's knowledge both in their school studies and also at the company. The intern is expected to bring ideas and knowledge from school into the company. Work research, virtual research (graduation) or dissertation: This is mostly done by students who are in their final year of school. With this kind of internship, a student does research for a particular company. The company can have something that they feel they need to improve, or the student can choose a topic in the company themselves. The results of the research study will be put in a report and often will have to be presented.

 Unpaid internships are typically through non-profit charities and think tanks which often have unpaid or volunteer positions. State law and state enforcement agencies may impose requirements on unpaid internship programs under Minimum Wage Act. A program must meet criteria to be properly classified as an unpaid internship. Part of this requirement is proving that the intern is the primary beneficiary of the relationship. Unpaid interns perform work that is not routine and work that company doesn't depend upon. Partially-paid internships is when students are paid in the form of a stipend. Stipends are typically a fixed amount of money that is paid out on a regular basis. Usually, interns that are paid with stipends are paid on a set schedule associated with the organization. Virtual Internship are internships that are done remotely on email, phone, and web communication. This offers flexibility as physical presence isn't required. It still provides the capacity to gain job experience without the conventional requirement of being physically present in an office. Virtual interns generally have the opportunity to work at their own pace. International Internships are internships done in a country other than the one that the country of residence.

These internships can either be in person or done remotely. Van Mol analyzed employer perspectives on study abroad versus international internships in 31 European countries, finding that employers value international internships more than international study, while Predovic, Dennis and Jones found that international internships developed cognitive skills like how new information is learned and the motivation to learn.

Returnship are internships for experienced workers who are looking to return to the workforce after taking time away to care for parents or children. Internship for a fee Companies in search of interns often find and place students in mostly unpaid internships, for a fee. These companies charge students to assist with research, promising to refund the fee if no internship is found.

The programs vary and aim to provide internship placements at reputable companies. Some companies may also provide controlled housing in a new city, mentorship, support, networking, weekend activities or academic credit. Some programs offer extra add-ons such as language classes, networking events, local excursions, and other academic options. Some companies specifically fund scholarships and grants for low-income applicants. Critics of internships criticize the practice of requiring certain college credits to be obtained only through unpaid internships. Depending on the cost of the school, this is often seen as an unethical practice, as it requires students to exchange paid-for and often limited tuition credits to work an uncompensated job. Paying for academic credits is a way to ensure students complete the duration of the internship, since they can be held accountable by their academic institution. For example, a student may be awarded academic credit only after their university receives a positive review from the intern's supervisor at the sponsoring organization. Secondary level work experience Work experience in England was established in the 1970s by Jack Pidcock, Principal Careers Officer of Manchester Careers Service.

The Service organized two weeks work experience for all Year 10 pupils in Manchester Local Education Authority schools, including those for pupils with special educational needs. Ironically, it was initially resisted by trade unions, and at first he had a job convincing schools, until eventually he persuaded the L.E.A. and councilors to go ahead. It became highly valued by pupils, teachers,

- **Team work**

In this project 1-2 people work together thus providing enough opportunity for team work and coordination. This was a good experience for us as the team was very cooperative and understanding.

- **Responsibility and Keeping Commitments**

The importance of keeping commitments and the time of others was an important thing, which I learnt as an intern.

4.2 Technical Outcomes:

The technical outcomes of our internship are as listed below:

- **Improved data management and organization:**

  We learnt how to collect, clean, organize, and store data more efficiently using tools such as Python, R, or SQL.

- **Data visualization:**

  We learnt how to create meaningful and insightful visualizations of data using tools such as Tableau, Matplot-lib, or ggplot2. This may involve creating graphs, charts, and interactive dashboards to help communicate insights and trends.

- **Statistical analysis:**

  We learnt how to perform statistical analysis on data using techniques such as hypothesis testing, regression analysis, or cluster analysis. This may involve using tools such as pandas, NumPy, or scikit-learn in Python.

- **Machine learning:**

  We learnt how to apply machine learning techniques to analysis and classify data. This may involve using algorithms such as decision trees, random forests, or neural networks to identify patterns and make predictions.

- **Data modeling and forecasting:**

  We learnt how to use data modeling techniques to make predictions about future trends or outcomes based on historical data. This may involve using tools such as ARIMA, exponential smoothing, or regression analysis.

- **Optimization:**

  We learnt how to use optimization techniques to find the best solutions to complex problems. This may involve using tools such as linear programming, dynamic programming, or Meta heuristics.

- **Data-driven decision making:**

  We learnt how to use data analysis to make informed decisions based on empirical evidence. This may involve working with stakeholders to identify key performance indicators (KPIs) and using data to track progress and identify opportunities for improvement.

4.2.1 Applying Knowledge to the Task

In knowledge intensive environment, task in an organization is typically performed by a group of people who have task related knowledge and expertise. Each group may require task-related knowledge on different topic domains and documents to accomplish its tasks. Document recommendation methods are very useful to resolve the information overload problem and proactively support knowledge workers in the performance of tasks by recommending appropriate documents to meet their information needs.

A worker's document referencing behavior can be model as a knowledge flow (KF) to represent the evolution of his information needs over time. However, the information needs of workers and groups may change over time. Additionally, most traditional recommendation methods which provide personalized recommendation do not consider worker's KFs, or the information needs of majority of workers in a group to recommend task knowledge. In this work, I, integrate the KF mining method and propose group-based recommendation method, including group based collaborative filtering (GCF) and group content- based filtering (GCBF), to actively provide the task related documents for group. Experimental results show that the proposed methods have better performance than the personalized recommendation methods in recommending the needed documents for groups. Thus, the recommended documents can fulfill the groups task needs and facilitate knowledge sharing among groups.

4.3 Non-Technical Outcomes

4.3.1 Verbal and Written Communication Skills

We have demonstrated and learnt our strong communication skills, both verbal and written, throughout the internship, while working for training and development.

4.3.2 Personality Development

Personality development plays an imperative role at workplace as it decides the way an individual interacts with his fellow workers and responds to various situations. How an individual behaves at the workplace depends on his/her personality. Personality development helps in polishing and grooming individuals and makes them better and efficient resources for the organization. Personality development also reduces stress levels and teaches an individual to face even the worst situations with a smile. Personality reflects how one conducts himself/herself in the professional environment. Never carry your personal problems to work. Personality development helps an individual to keep his personal life separate from his professional life.

Differences in opinions and views often lead to conflicts and arguments among employees. Employees with different attitudes and Mindset find it extremely difficult to adjust with each other and work in unison. Personality development sessions motivate an individual to think positively and eventually reduce stress at the workplace. Individuals as a result of personality development tend to behave in a mature way; making the organization a much better place to work. Personality development is essential to bring a change in an individual's attitude, thinking, behavior and mindsets. It also strengthens the relationship among co -workers.

4.3.3 Time Management

Proper time management at work place has a number of positive effects, ranging from making you more focused and valuable employee of reducing the stress of your job. Employers appreciate employees who can get the maximum amount of good works done in minimum amount of time. It also strengthens the relationship among co- workers.

4.3.4 Resource Utilization Skills

Effective management of resources is essential tasks of companies that are managing different projects. It is important for them to efficiently organize allocate personal as well as equipment for different projects, same time avoiding idle resources. Having the information of availability of resource and have those available at the right time for the activities plays the vital role in managing the costs and smoothly executing the project activities.

4.3.5 Time Managing Skills:

To be an effective employee and successful we must be able to manage our time efficiently. Most people have effectively good tome management skill from school or university experience. We had an opportunity to improve our time management skills such that we could apply it both in our professional and personal life. Oftentimes we were tested for time management with tasks to be performed within schedule that helped us to complete the tasks within scheduled time.

4.3.6 Teamwork:

Teamwork is such an important aspect of running a successful company and our internship has thought us how to do this on a business level. Key techniques adopted in our internship for effective team work:

- Communication Process: exchange contact details, establishing rules and regulations and limits on communication forms (Email, Texts, Mobile Facebook), maximum length of time responses, when face-to-face meetings were held.
- Task Analysis and distribution of workload: dividing up questions or parts of questions into equal proportions, going off to work individually to return the day before the task is due to put it all together.
- Meeting Schedule: Meetings were conducted frequently (once in 4 days) to know the status of the task being completed. This allowed us maintain our pace completing the assigned work.
- Open communication to avoid conflicts.
- Effective coordination to avoid confusion and the overstepping of boundaries. Efficient cooperation to perform the tasks in a timely manner and produce the required results, especially in the form of workload sharing.

- High levels of interdependence to maintain high levels of trust, risk-taking, and performance.

4.3.7 DATA SETS

- A data set (or dataset) is a collection of data. In the case of tabular data, a data set corresponds to one or more database tables, where every column of a table represents a particular variable, and each row corresponds to a given record of the data set in question. The data set lists values for each of the variables, such as for example height and weight of an object, for each member of the data set. Data sets can also consist of a collection of documents or files.

- In the open data discipline, data set is the unit to measure the information released in a public open data repository. The European data.europa.eu portal aggregates more than a million data sets.

- Some other issues (real-time data sources, non-relational data sets, etc.) increases the difficulty to reach a consensus about it.

- Several characteristics define a data set's structure and properties. These include the number and types of the attributes or variables, and various statistical measures applicable to them, such as standard deviation and kurtosis.

- The values may be numbers, such as real numbers or integers, for example representing a person's height in centimeters, but may also be nominal data (i.e., not consisting of numerical values), for example representing a person's ethnicity. More generally, values may be of any of the kinds described as a level of measurement. For each variable, the values are normally all of the same kind.

- there may also be missing values, which must be indicated in some way. In statistics, data sets usually come from actual observations obtained by sampling a statistical population, and each row corresponds to the observations on one element of that population. Data sets may further be generated by algorithms for the purpose of testing certain kinds of software. Some modern statistical analysis software such as SPSS still present their data in the classical data set fashion. If data is missing or suspicious an imputation method may be used to complete a data set.

- The use of basic technology is also a feature of non-human animal species. Tool use was once considered a defining characteristic of the genus Homo. This view

was supplanted after discovering evidence of tool use among chimpanzees and other primates, dolphins, and crows. For example, researchers have observed wild chimpanzees using basic foraging tools, pestles, levers, using leaves as sponges, and tree bark or vines as probes to fish termites. West African chimpanzees use stone hammers and anvils for cracking nuts, as do capuchin monkeys of Boa Vista, Brazil. Tool use is not the only form of animal technology use; for example, beaver dams, built with wooden sticks or large stones, are a technology with "dramatic" impacts on river habitats and ecosystems.

- One of the most important things is that writing must be appropriate for others to understand. I have learnt to use the terminology that makes sense to my readers. When we become a professional in a particular field, we have to learn language of that field. As a professional, people hire us to work on their behalf because we have the knowledge and skills that they don't possess. Therefore, we find our self-communicating in writing with people who don't know the language of our field; our writing must make sense to people to read it.

- From a developers point of view, these are things like languages and frameworks, data storage mechanisms, services and products to allow us to create platforms and applications for the web to build websites and web applications (a web application is a website that offers functionality, storing data - text, images, etc. - that can be processed and return a useful output to you, or that you can come back to later, such as Facebook or Quora).

- Being able to work with other people in the department and office is key. Interns were encouraged to work together and brainstorm ideas to help the company gain followers and ultimately receive more donation for their cause. Being able to work together and help one another was vital to the department's success. Teamwork is such an important aspect of running successful company and my internships have taught me how to do this on a business level.

Chapter 5

RESULTS


```
In [6]: sns.countplot(x=train_df['Sentiment'])

Out[6]: <AxesSubplot:xlabel='Sentiment', ylabel='count'>
```



```
In [7]: train_df.duplicated().sum()
Out[7]: 2700

In [8]: train_df= train_df.drop_duplicates()

In [9]: train_df.Sentiment.value_counts().plot(kind='pie', autopct='%1.0f%%',figsize=(5,5),colors=["red", "yellow", "green",'blue'])

Out[9]: <AxesSubplot:ylabel='Sentiment'>
```
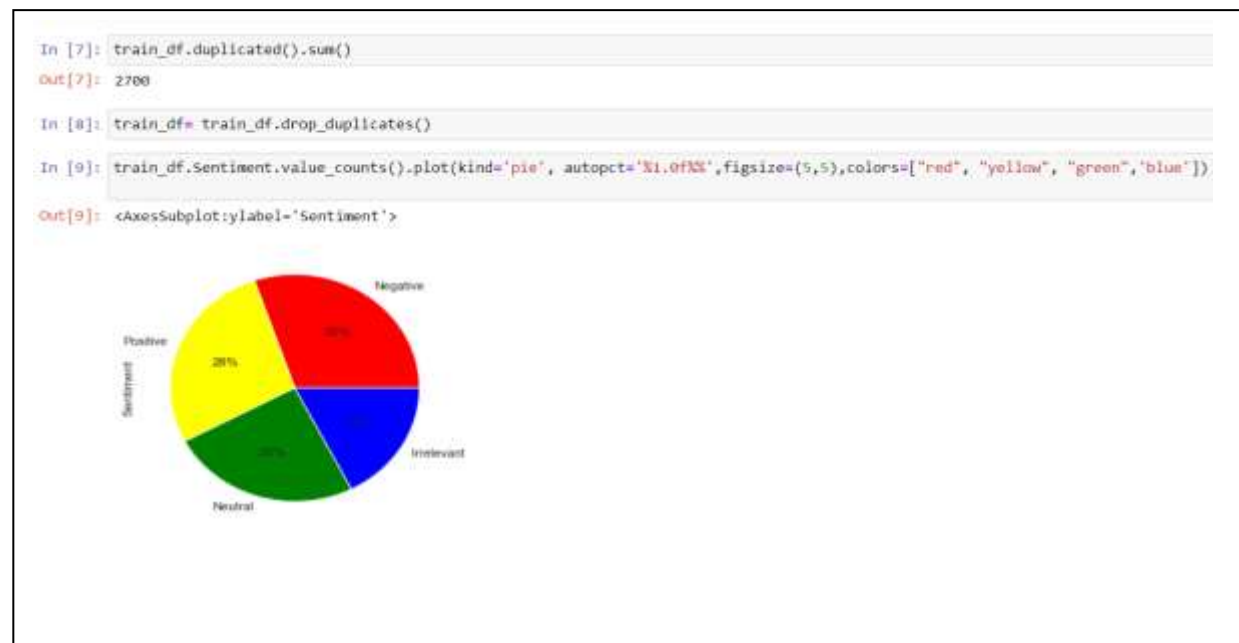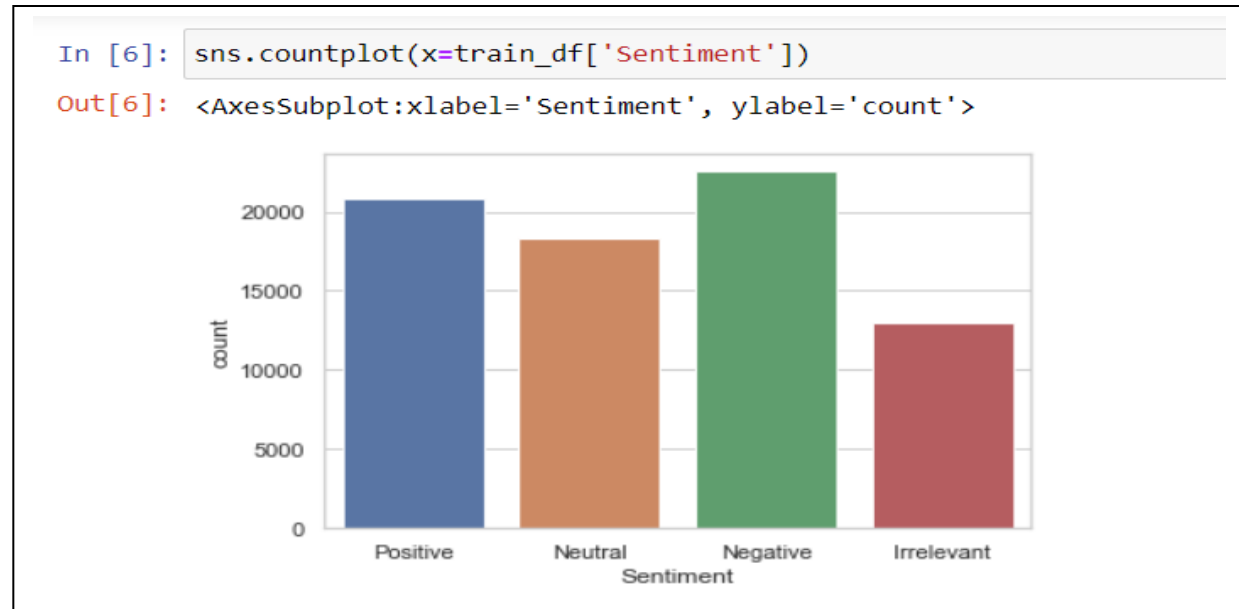
Figure: 5.2

```
step1: remove additional letter such as @

In [15]: REPLACE_WITH_SPACE = re.compile("(@)")
         SPACE = " "

         def preprocess_reviews(reviews):
             reviews = [REPLACE_WITH_SPACE.sub(SPACE, line.lower()) for line in reviews]

             return reviews

         reviews_train_clean = preprocess_reviews(tweet_train)
         reviews_valid_clean = preprocess_reviews(tweet_valid)

step2: remove Stop Words

In [16]: def remove_stop_words(corpus):
             removed_stop_words = []
             for review in corpus:
                 removed_stop_words.append(
                     ' '.join([word for word in review.split() if word not in english_stop_words]))
             return removed_stop_words

         no_stop_words_train = remove_stop_words(reviews_train_clean)
         no_stop_words_valid = remove_stop_words(reviews_valid_clean)
```

Figure: 5.3

```
Machine Learning Model

In [24]: #text_classifier = RandomForestClassifier(n_estimators=500, random_state=0)

In [25]: #text_classifier.fit(X_train, y_train)

Out[25]: RandomForestClassifier(n_estimators=500, random_state=0)

In [26]: #y_pred=text_classifier.predict(X_val)

In [27]: #print(classification_report(y_val,y_pred))
```

	precision	recall	f1-score	support
0	0.92	0.64	0.75	8811
1	0.79	0.87	0.83	15166
2	0.77	0.80	0.78	12330
3	0.79	0.82	0.80	13853
accuracy			0.80	50160
macro avg	0.82	0.78	0.79	50160
weighted avg	0.81	0.80	0.80	50160

Figure: 5.4

```
In [36]: X_train, X_val, y_train, y_val = train_test_split(X, target, train_size = 0.5)
         text_classifier = RandomForestClassifier(n_estimators=500, random_state=0)
         text_classifier.fit(X_train, y_train)
         y_pred=text_classifier.predict(X_val)
         print(classification_report(y_val,y_pred))
```

```
               precision    recall  f1-score   support

            0       0.96      0.78      0.86      6272
            1       0.88      0.92      0.90     10836
            2       0.91      0.85      0.88      8959
            3       0.81      0.92      0.86      9761

     accuracy                           0.88     35828
    macro avg       0.89      0.87      0.87     35828
 weighted avg       0.88      0.88      0.88     35828
```

```
In [37]: X_train, X_val, y_train, y_val = train_test_split(X, target, train_size = 0.7)
         text_classifier = RandomForestClassifier(n_estimators=500, random_state=0)
         text_classifier.fit(X_train, y_train)
         y_pred=text_classifier.predict(X_val)
         print(classification_report(y_val,y_pred))
```

```
               precision    recall  f1-score   support

            0       0.97      0.84      0.90      3733
            1       0.92      0.93      0.92      6538
            2       0.93      0.90      0.91      5308
            3       0.85      0.94      0.89      5918

     accuracy                           0.91     21497
    macro avg       0.92      0.90      0.91     21497
 weighted avg       0.91      0.91      0.91     21497
```

Figure: 5.5

CONCLUSION

Twitter sentiment analysis is the process of using natural language processing and machine learning techniques to extract and analyze the sentiment expressed in tweets. The goal is to identify whether a tweet has a positive, negative, or neutral sentiment. Sentiment analysis can be used to understand public opinion about a particular topic, brand, product, or event.

Hence, the project helps to predict a particular price for the stocks and the dimension for a Sentiment. The project also proves that RandomForestClassifier is 91% accurate. Considering the real-time price, the predicted value and the real-time value is around 80-91% accurate.

The future work of this particular project will consider more data values and more data characteristics in the dataset, to analyse the data better and predict accurate price for the given Stocks and Their Sentiments.

Another important aspect of this project was testing. I recognized the importance of testing early on and made sure to write unit tests for each component and feature of the application. This allowed me to catch bugs early and ensure that the application was functioning as intended

One of the most significant takeaways from this project has been the importance of proper planning and documentation. Before starting the actual development work, I spent a significant amount of time researching and defining the project requirements, which allowed me to stay organized and focused throughout the development process. I also made sure to document each step of the development process, which helped me to stay on track and troubleshoot any issues that arose

REFERENCES

[1] H.L. Harter,Method of Least Squares and some alternatives-Part II.International Static Review.1972,43(2) ,pp. 125-190.

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

[3] Lu. Sifei et al,A hybrid regression technique for house prices prediction. In proceedings of IEEE conference on Industrial Engineering and Engineering Management: 2017.

[4] R. Victor,Machine learning project:Predicting Boston house prices with regression in towards datascience.

[5] S. Neelam,G. Kiran,Valuation of house prices using predictive techniques, Internal Journal of Advances in Electronics and Computer Sciences:2018,vol 5

[6] S. Abhishek.:Ridge regression vs Lasso,How these two popular ML Regression techniques work. Analytics India magazine,2018

[7] S.Raheel.Choosing the right encoding method-Label vs One hot encoder. Towards datascience,2018.

[8] Raj, J. S., & Ananthi, J. V. (2019). Recurrent Neural Networks and Nonlinear Prediction in Support Vector Machines. Journal of Soft Computing Paradigm (JSCP), 1(01), 33-40.

[9]   Predicting house prices in Bengaluru(Machine Hackathon)
https://www.machinehack.com/course/predicting-house-prices-inbengaluru/