A PROJECT REPORT ON

ROBOTIC PROCESS AUTOMATION WITH INCREASING PRODUCTIVITY USING MACHINE LEARNING

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE
IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE

OF

BACHELOR OF ENGINEERING
(COMPUTER ENGINEERING)

SUBMITTED BY

Mr. Rutik Govind Kulkarni	**B190424323**
Mr. Ayush Mishra	**B190424224**
Mr. Govind Sambhaji Bhujbal	**B190424277**
Mr. Sauraw Arun Diwase	**B190424257**

Sinhgad Institutes

DEPARTMENT OF COMPUTER ENGINEERING

STES'S SINHGAD INSTITUTE OF TECHNOLOGY

KUSGAON (BK), LONAVALA, PUNE- 410401

SAVITRIBAI PHULE PUNE UNIVERSITY

2022-2023

Sinhgad Institutes

CERTIFICATE

This is to certify that the project report entitles

ROBOTIC PROCESS AUTOMATION WITH INCREASING PRODUCTIVITY USING MACHINE LEARNING

Submitted by

Mr. Rutik Govind Kulkarni	**B190424323**
Mr. Ayush Mishra	**B190424224**
Mr. Govind Sambhaji Bhujbal	**B190424277**
Mr. Sauraw Arun Diwase	**B190424257**

is a bonafide student at this institute and the work has been carried out by him/her under the supervision of **Prof. Rupali Shishupal** and it is approved for the partial fulfillment of the requirement of Savitribai Phule Pune University, for the award of the degree of **Bachelor of Engineering** (Computer Engineering).

(Prof. Rupali Shishupal)
Guide
Department of Computer Engineering

(**External Examiner)**

(Dr. S. D. Babar)
Head
Department of Computer Engineering
Lonavala, Pune – 410401

(Dr. M.S. Gaikwad)
Principal
Sinhgad Institute of Technology,
Lonavala, Pune

Place: Lonavala

Date:

ACKNOWLEDGEMENT

We express our sense of gratitude towards our project guide **Prof. Rupali Shishupal** for her valuable guidance and suggestions at every step of study of this project group stage one also her contribution for solution of every problem at each stage.

We are thankful to **Dr. S.D. Babar** head of the department of computer engineering and all the staff members who extended the preparatory steps of this project we are very much thankful to respected principal **Dr. M.S. Gaikwad** for his support and providing all facilities to complete the project.

Finally, we want to thank to all our friends for the support and suggestions at every stage of project.

PROJECT ASSOCIATES:

Student Name	Seat No.	Signature
Mr. Rutik Govind Kulkarni	**B190424323**	
Mr. Ayush Mishra	**B190424224**	
Mr. Govind Sambhaji Bhujbal	**B190424277**	
Mr. Sauraw Arun Diwase	**B190424257**	

ABSTRACT

Robotic Process Automation (RPA) has received growing attention within the digital transformation as this cutting-edge technology automates human behavior and promises high potentials. However, the adoption in purchasing and supply management (PSM) is still in its infancy and has hardly been explored, particularly in the public sector. Based on a multiple case study including 19 organizations of the public and private sector, this paper narrows that gap and presents comprehensive insights into potentials, barriers, suitable processes, and best practices and components for RPA implementation.The findings indicate that adoption depends on the organizations' digital procurement readiness and maturity. Application areas of RPA enlarge with increasing experience and range from transactional and operative tasks within the procure-to-pay process to more strategic use cases in sourcing and supply relationship management. Potentials mainly comprise employee reliefs, cost savings, and increased operational efficiency and quality.

Keywords: RPA, Machine Learning, Purchasing and supply management (PSM), Digital Readiness, Automation.

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION

1.1 Overview

RPA(Robotic Process Automation)

- RPA is defined as an art of using software robots to interact with Software-as-a-Service applications and IT systems to automate the rule-based manual jobs associated with repetitive and transactional processes.
- The robot mimics the interactions of an employee with a system's user interface. The RPA services provide data security, enhanced business efficiency and effectiveness across various business applications without modifying available system and infrastructure.
- By using RPA we will get huge amount of data in DB from any website and by using ML algorithm we will increases its productivity.However, the adoption of RPA and other advanced digital technol-ogies in the PSM function impacts its processes, capabilities, and pro-fessionals.
- Guiding the digitalization of procurement, Srai and Lorentz(2019) developed a framework that combines multiple grand theorieswith digitalization- and PSM-relevant perspectives. The concept links Industry 4.0 technologies with internal and external PSM value drivers,e.g., transaction management, process improvement and innovation,supplier-capability assessment, and relationship management. However, the framework does not yet comprise RPA.
- While emerging tech-nologies, such as blockchain, big data have received increasing attention from scientists, academic studies relatingRPA to supply chain management (SCM) and PSM have remained scarce to date. In academia and practice, the adoption of RPA for procurement still lags behind other business functions such as accounting or human resource management.
- Initial projects in the industry show a failure rate of up to 50 percent, while the application of RPA in the public sector is still in its infancy. Relevant work predominantly examined how the PSM function of private companies could benefit from RPA by emphasizing potentials and initial application areas for operational purchasing, e.g., the procure-to-pay (P2P) process.

- Along this line, most studies shed little light on more strategic tasks and barriers that could hamper adoption and prevent procurement organizations from applying RPA. We are unaware of any scientific publication examining the implementation of RPA in public procurement, although, due to specific legal regulations and workflows, public procurement departments handle many manual processes that could be suitable candidates for RPA

Machine Learning

- Machine learning is programming computers to optimize a performance criterion using example data or past experience . We have a model defined up to some parameters, and learning is the execution of a computer program to optimize the parameters of the model using the training data or past experience. The model may be predictive to make predictions in the future, or descriptive to gain knowledge from data.
- The field of study known as machine learning is concerned with the question of how to construct computer programs that automatically improve with experience.

Microsoft Power BI

- Microsoft Power BI is a business intelligence platform that provides nontechnical business users with tools for aggregating, analyzing, visualizing and sharing data. Power BI's user interface is fairly simple for users familiar with Excel and its deep integration with other Microsoft products makes it a very versatile self-service tool.

- Microsoft Power BI is used to find insights within an organization's data. Power BI can help connect disparate data sets, transform and clean the data into a data model and create charts or graphs to provide visuals of the data. All of this can be shared with other Power BI users within the organization.

1.2 MOTIVATION

How the combination of the two technologies would normally work is as follows. The role of RPA would be to structure and manage workflow for a process, extract needed information from multiple information systems, and perhaps make some simple decisions using rules. The RPA system would typically interact with a previously-created ML scoring application through the use of an API. The machine learning model, which would have been trained on data for which the outcomes are known, would use data that the RPA system pulled from transactional systems. Over time, the ML model could be retrained with new data.

We received rich and comprehensive empirical data by examining eleven RPA adopters (including the consultancies) and eight non-adopters (including those with adoption plans). Based on the triangulated primary and secondary data, we categorize the participating private and public organizations regarding their RPA experience, digital procurement readiness, and maturity level of PSM digitalization (Fig. 1). RPA experience comprises the aggregated information on RPA maturity, i.e., the number and scope of completed and planned projects, the number of deployed bots in the procurement department, and the extent of automated PSM transactions.

1.3 PROBLEM DEFINITION AND OBJECTIVES

Robotic process automation with increasing productivity using machine learning.

In the light of growing competition and cost pressure, the digitalization of purchasing and supply management (PSM) constitutes a major investment priority for organizations as this area tends to lag behind other business functions . For this reason, the advanced digital technology Robotic Process Automation (RPA) is increasingly applied to redesign, optimize, and automate procurement processes.

Fetch data using robotic process automation and using that data we can do the operations to predict the sales report which is then visualised using various techniques through Microsoft Power BI.

1.4 PROJECT SCOPE & LIMITATIONS

Project Scope

Using RPA we can fetch the various Data from various websites and Data is useful for various business purposes.To analyse the sells data for Big Manufacturer.We can predict the sells using our model.
To analyse the Business Intelligence using power BI we can visualize sells prediction on various visualization technique.

In this article, we will discuss the future scope and job options of RPA and understand why it is an important factor in automation marketing.

RPA has provided an excellent solution for organizations to replace repetitive, mundane, rule-based processes with software bots. It is now helping organizations who were looking to increase their workflow accuracy and efficiency. First, RPA was widely adopted in the IT sector. It amazed many big organizations as well as small and medium enterprises with outstanding results. Later, it was adopted in other sectors like Finance, Accounting, Banking, etc.

Limitations :

a) **Long-term sustainability.**

RPA can become a serious decoy from the necessary long-term work needed to digitize and make processes and administrative work more efficient. There is a risk that you may focus on quick fixes rather than doing things the correct way from the start.

b) **Implementation**.

RPA might give you good value in stand-alone tasks, but it takes time and money to set up. It's a bad idea to cut corners on RPA, too, since a miscalibration will lead to errors. An inefficient process won't be transformed simply because you add some "smartness" at the top. You should know that almost half of all RPA systems fail when first rolled out, so the implementation might take more resources than you initially thought.

c) **Error magnification**.

RPA robots can't detect some obvious errors that a human would be able to immediately point out. If your data has problems with it, RPA robots will not call it out, but pass it on, magnifying an error that might have otherwise been caught.

d) **Overall risk**.

Some problems aren't a good fit for RPA, especially when the stakes are high. For example, if you need to handle your purchase invoices, it's likely a better idea to use software that is able to understand and manage the data correctly from the start.

e) **Maintenance**.

Most RPA solutions have to be custom-made to fit your business. It likely won't be worth it to invest in such a system if the way your business runs could change drastically in the future. Even minor changes in your setup can create significant disruption for your RPA robots.

2. LITERATURE SURVEY

1. Robotic Process Automation Through Advance Process Analysis Model

Author : Devansh Hiren Timbadia, Parin Jigishu Shah, Supriya Agrawal, Sughosh Sudhanvan

Robotics Process Automation is an advanced technology that builds an intelligent software robot that can emulate human interactions with a business process. RPA is an efficient automated method where software agents interact through a graphical user interface in a human-like manner. RPA has various applications in most industries like banking and finance, human resources, healthcare, etc. In this paper, we have proposed our RPA process analysis model and have compared it with a traditional model using various comparison parameters like frequency of change, degree of complexity, time is taken, screen usage and volume of transactions and found out that proposed method gave efficient results over the traditional method. Keywords—Robotic Process Automation (RPA), Business Process Analysis, Software Agent, Process Model, Automation Model, Software Process Automation.

2. Robotic Process Automation in purchasing and supply management: A multiple case study on potentials, barriers, and implementation

Author : Christian Flechsig * , Franziska Anslinger, Rainer Lasch

Robotic Process Automation (RPA) has received growing attention within the digital transformation as this cutting-edge technology automates human behavior and promises high potentials. However, the adoption in purchasing and supply management (PSM) is still in its infancy and has hardly been explored, particularly in the public sector. Based on a multiple case study including 19 organizations of the public and private sector, this paper narrows that gap and presents comprehensive insights into potentials, barriers, suitable processes, and best practices and components for RPA implementation. The findings indicate that adoption depends on the organizations' digital procurement readiness and maturity. Application areas of RPA enlarge with increasing experience and range from transactional and operative tasks within the procure-to-pay process to more strategic use cases in sourcing and supply relationship management. Potentials mainly comprise employee reliefs, cost savings, and increased operational efficiency and quality. We uncover multiple technical, organizational, and environmental barriers related to IT infrastructure and human resources, internal communication, financial resources, top management support, organizational structures,

supplier-related issues, and government regulations. Furthermore, our study indicates several differences between the private and public sectors for RPA implementation. We outline implications for the emerging research on RPA and pivotal directions for organizational practice.

3. Microsoft Power BI Desktop for Data Analytics

Author : João Garrott Marques Negreirosa, João Alexandre Lobo Marquesb

This article reviews MS-Power BI Desktop that empowers managers to see and understand their data. By drag and drop, it is possible to discover trends and outliers from interactive dashboards, including combining data from multiple sources (from spreadsheets and data warehouses to big cloud data) at a glance. If asking questions is a key role of any manager (either using a browser, a tablet or a phone), business intelligence (BI) is the tool that allows the change managerial perspective of the problem and reveal deeper meaning of the data. Inevitable, basic technical jargons such as data mining or on-line analytical processing (OLAP) are introduced too. Three practical cases are disclosed while its layouts are discussed on a managerial and academic perspective.

4. Human Resource Analytics using Power Bi Visualization Tool

Author : Mohammed Ameer, Simhadri Prem Rahul, Dr.Suneetha Manne

Human resource(HR) management is a subject of vast knowledge in which predictive analytics is one of its main components which includes employee turnover analysis, employee work performance analysis, and training requirements analysis as results. The main purpose of Human Resource management is to measure the work achievement of employees and their role in the services or business which acts as benefits to the company and to analyze employee period in the company. The main motto of Human Resource analytics is to identify skilled individuals who strive extremely for the return of investment for the organization by considering several factors which help for a better understanding of the individual by predictive analysis. Employee churn is considered a major problem for many organizations. It is one of the crucial problems to identify because it affects sustainability and also the organization’s planning and enhancing work culture harmony. Therefore, the Human Resource department in every organization is striving hard and paying attention to identify the underlying improvements.

5. Predictive Analysis for Big Mart Sales Using Machine Learning Algorithms

Author : Ranjitha P1,Spandana M2

Abstract— Currently, supermarket run-centres, Big Marts keep track of each individual item's sales data in order to anticipate potential consumer demand and update inventory management. Anomalies and general trends are often discovered by mining the data warehouse's data store. For retailers like Big Mart, the resulting data can be used to forecast future sales volume using various machine learning techniques like big mart. A predictive model was developed using Xgboost, Linear regression, Polynomial regression, and Ridge regression techniques for forecasting the sales of a business such as Big -Mart, and it was discovered that the model outperforms existing models.

6. Business Forecasting System using Machine Learning Approach

Authors : Md. Anisur Rahman Mia*, Rupon Ghosh*

Abstract—Forecasting system can help any business to predict its future sale, profit and loss. So it is very useful while launching a new product or manufacturing existing products. In this work, a business forecasting system is designed and implemented to forecast the amount of future sale of products using Machine Learning (ML). ML algorithms build a pattern from input variables then make decision. Use of big data technology in researches also increasing. By combining big data and ML together powerful predictive systems can be designed. So, big data processing technology has been used in this work to prepare data for training purpose of the proposed system. Prediction accuracy of the proposed system varied 99% to 75% for different products and mean absolute percentage error (MAPE) is 7.32%. So the proposed system is very much efficient to predict future sale.

Sr. No.	Title of the Paper	Year of publication	Method /algorithm used	Features/advantages/results	Limitations
1	Robotic Process Automation Through Advance Process Analysis Model	2020	Integraed Defination Group of Modelling Method, Artificial Intelligence Method	Classify the process and provide suitability of the process for RPA	Factors like cost estimation, paybacks, and benefits
2	Robotic Process Automation in purchasing and supply management: A multiple case study on potentials, barriers, and implementation	2020	RPA adoption in PSM)	RPA has high connectivity to multiple software applications, thus complementing existing ERP	First, as with much qualitative research, our empirical study is restricted regarding the number and variety of examined organizations
3	Microsoft Power BI Desktop for Data Analytics	2019	Power BI (Visualisation Tools)	Classify the result for visualization	MS-Power BI cannot be considered one single standalone software.

4	Human Resource Analytics using Power Bi Visualization Tool	2020	Logistic Regression, Random Forest Classifier	Its implementation can boost up the economic growth of a company as it uses simple dashboards for performance levels.	Increase in complexity leads to lack in awareness
5	Predictive Analysis for Big Mart Sales Using Machine Learning Algorithms	2020	Holte's Linear Trend Model, Holt winter Seasonal Method	The Method of forecasting for predicting sells value is very beneficial towards ERP system.	Unforeseen cash flow and unmanaged production, staff and financing.
6	Business Forecasting System using Machine Learning Approach	2021	Deep Learning Network (DNN), Feed Forward Neural Network (FFNN)	ML technology is implemented to develop a predictionsystem and to forecast sale using the system. ML finds specific patterns or relationship from the training data.	Training error was found maximum for the commodity that has much price deviation over time.

3. SOFTWARE REQUIREMNETS SPECIFICATION

3.1 DATABASE REQUIREMENTS

Microsoft Excel File or .csv File is required for the sake of storing data which we have fetched from flipkart, Amazon websites.

3.2 SOFTWARE REQUIREMENTS

- Operating system: Windows 8 or newer.
- Microsoft Power BI Desktop.
- Python 3.6.
- Browser: Chrome/Firefox.
- VS Code
- Automation A360
- Power Bi desktop
- Jupyter Notebook
-

3.3 HARDWARE REQUIREMENTS

- CPU: 1 gigahertz (GHz) 64-bit (x64) processor or better recommended.
- Memory (RAM): At least 2 GB available, 4 GB or more recommended.
- Hard Disk Drive: Min 32 GB, 64 GB or more recommended.
- Peripheral Devices: Monitor, Mouse and Keyboard.

4. SYSTEM DESIGN

4.1 PROPOSED SYSTEM ARCHITECTURE

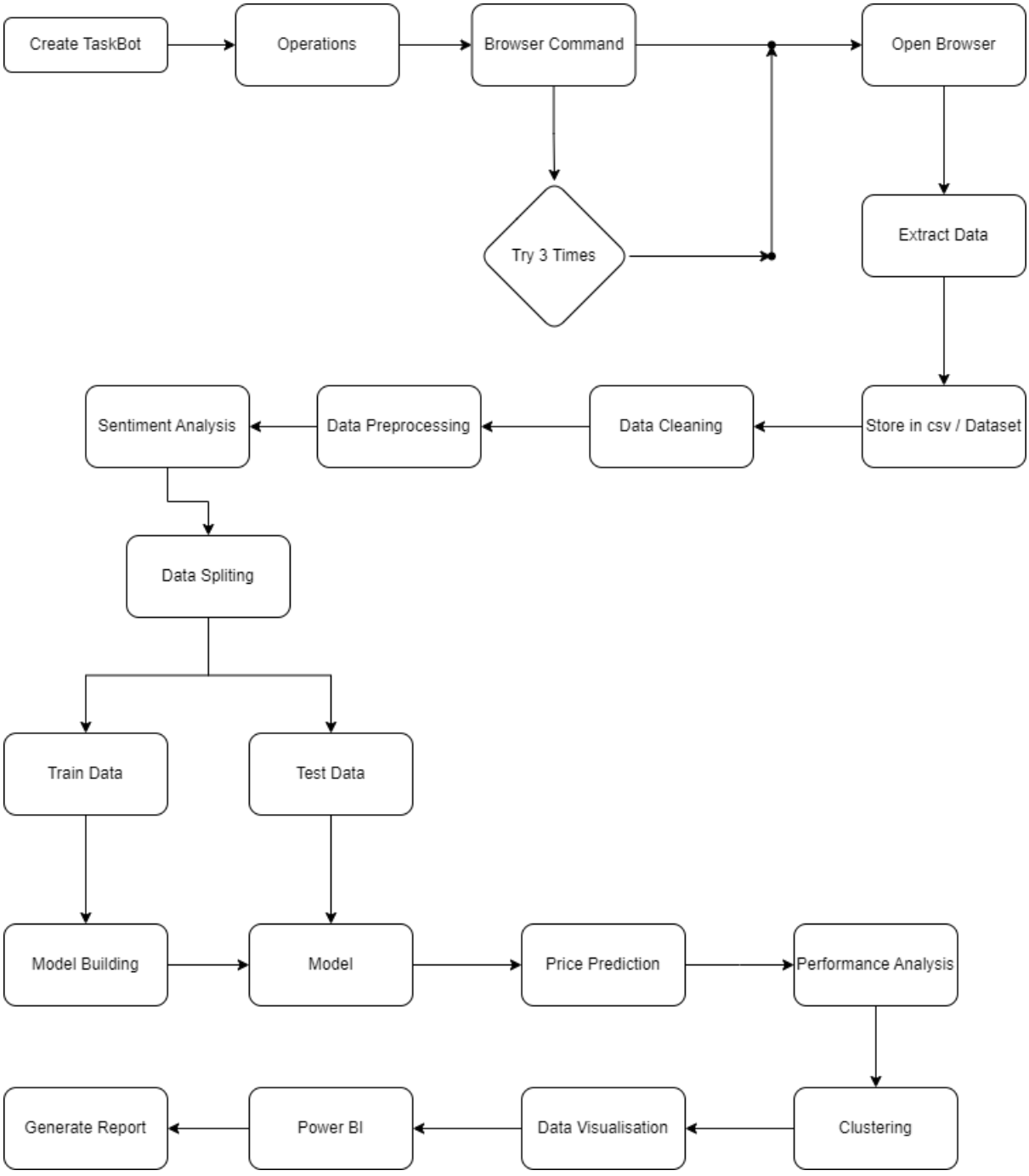


Figure No. 4.1

System Architecture

4.2 MATHEMATICAL MODEL

Sentiment Analysis:

- Let `reviews` be the list of text reviews.
- For each review in `reviews`:
  - Create a TextBlob object `text_blob` for the review.
  - Calculate the polarity of the review using `polarity = text_blob.sentiment.polarity`.
  - Map the polarity value to a numerical scale from 1 to 5 using `sentiment_score = round((polarity + 1) * 2.5)`.
  - Append the `sentiment_score` to the `sentiment_scores` list.
- Add the `sentiment_scores` as a new column 'Sentiment' to the DataFrame `df`.

Price Prediction:

- Let `X_train` be the training set features, `y_train` be the training set target variable, `X_test` be the testing set features, and `y_test` be the testing set target variable.
- Create a linear regression model `model`.
- Train the model using `model.fit(X_train, y_train)`.
- Make predictions on the testing set using `y_pred = model.predict(X_test)`.
- Calculate the mean squared error between `y_test` and `y_pred` using `mse = mean_squared_error(y_test, y_pred)`.

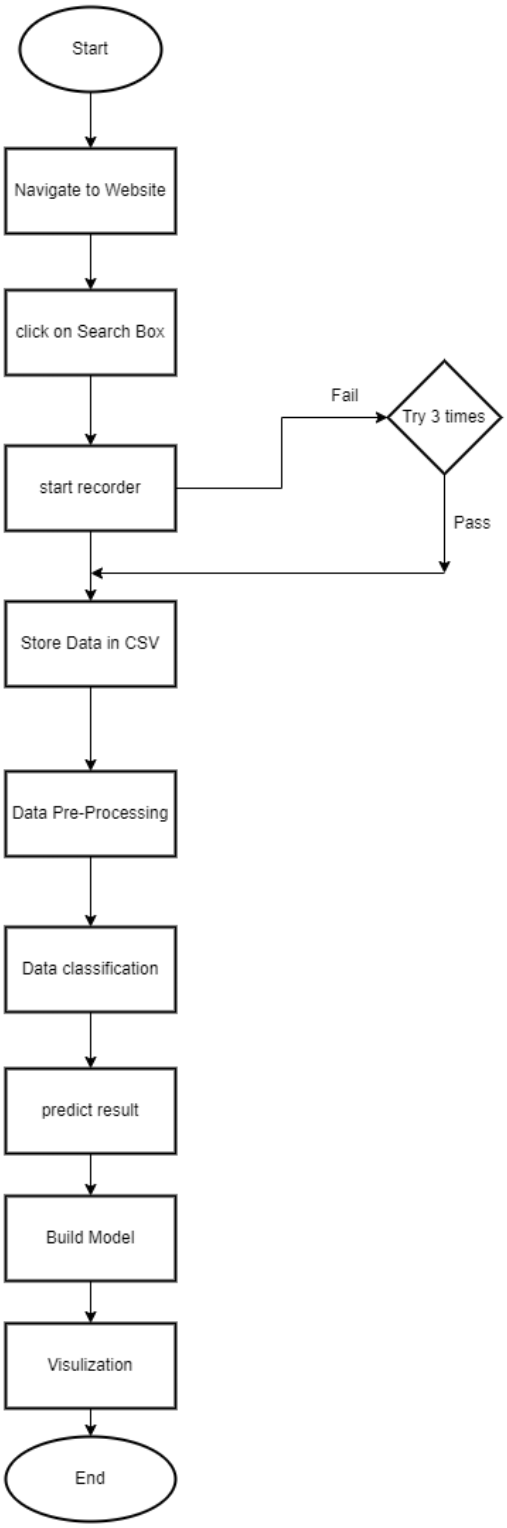
Clustering:

- Let `dataset` be the dataset to be clustered.
- Let `columns_for_clustering` be the list of columns used for clustering.
- For each column `col` in `columns_for_clustering`:
  - Prepare the data for clustering by selecting the data for the current column.
  - Perform one-hot encoding for categorical variables using `column_encoded = pd.get_dummies(column_data)`.
  - Impute missing values using `column_imputed = imputer.fit_transform(column_encoded)` (where `imputer` is a SimpleImputer object).
  - Scale the data using `column_scaled = scaler.fit_transform(column_imputed)` (where `scaler` is a StandardScaler object).

- Create a K-means clustering model `model` with `k` clusters.
- Fit the model to the scaled data using `model.fit(column_scaled)`.
- Get the cluster labels using `cluster_labels = model.labels_`.
- Add the cluster labels as a new column in the `dataset` DataFrame using `dataset[col + '_Cluster'] = cluster_labels.astype(str)`.

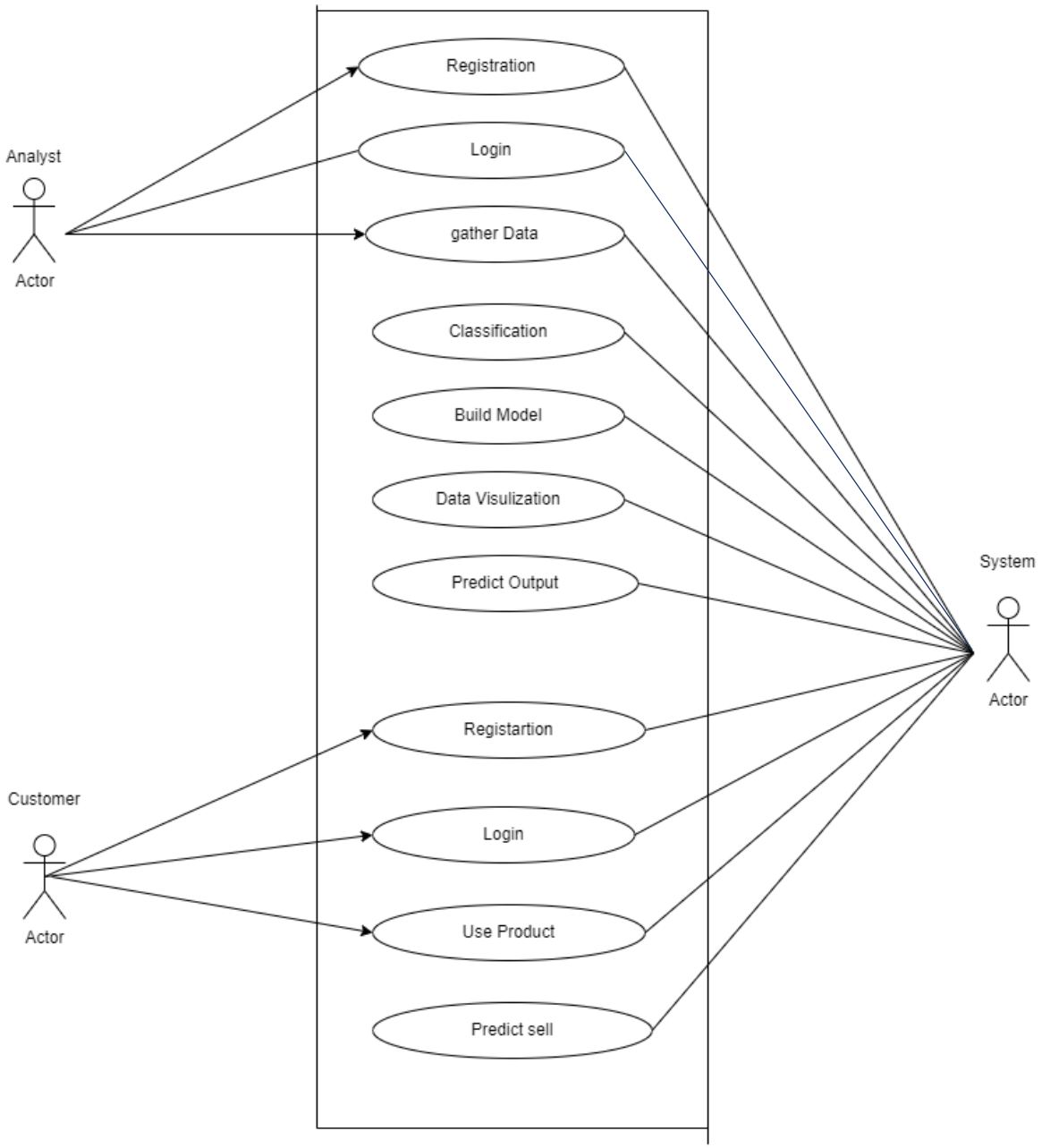
- Visualize the clusters for each column using scatter plots.

4.3 DATA FLOW DIAGRAM

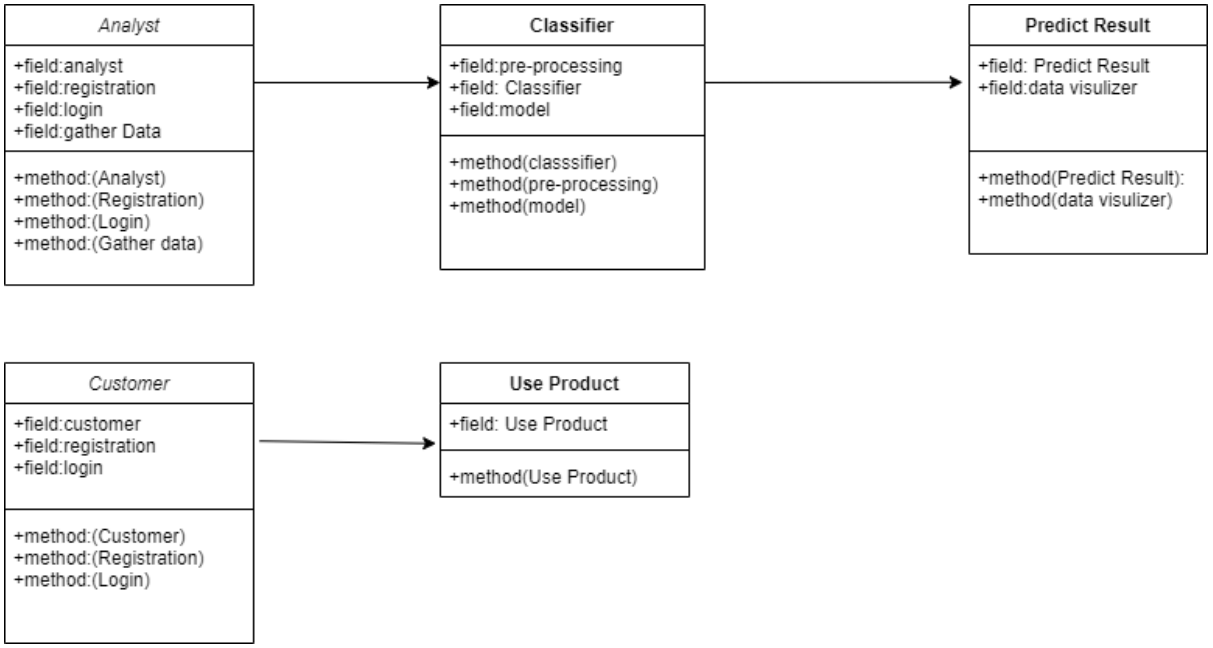


4.4 UML DIAGRAMS

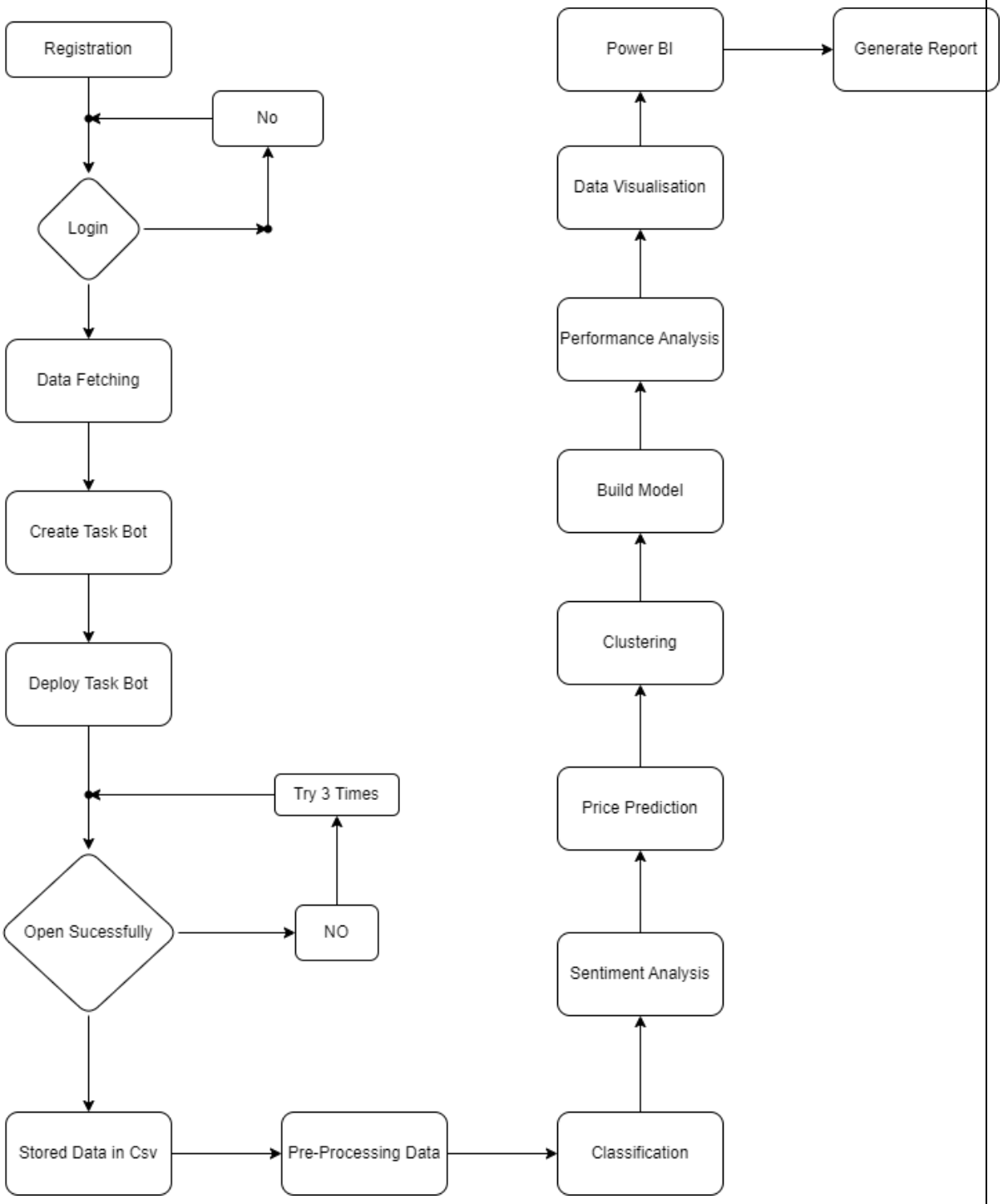
4.4.1 Use Case Diagram



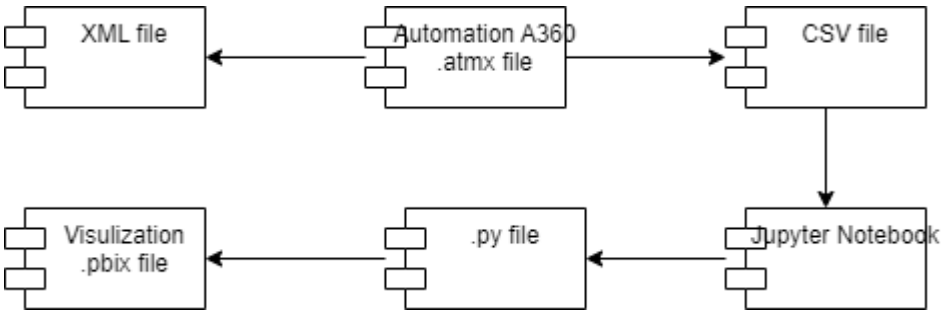
4.4.2 Class Diagram



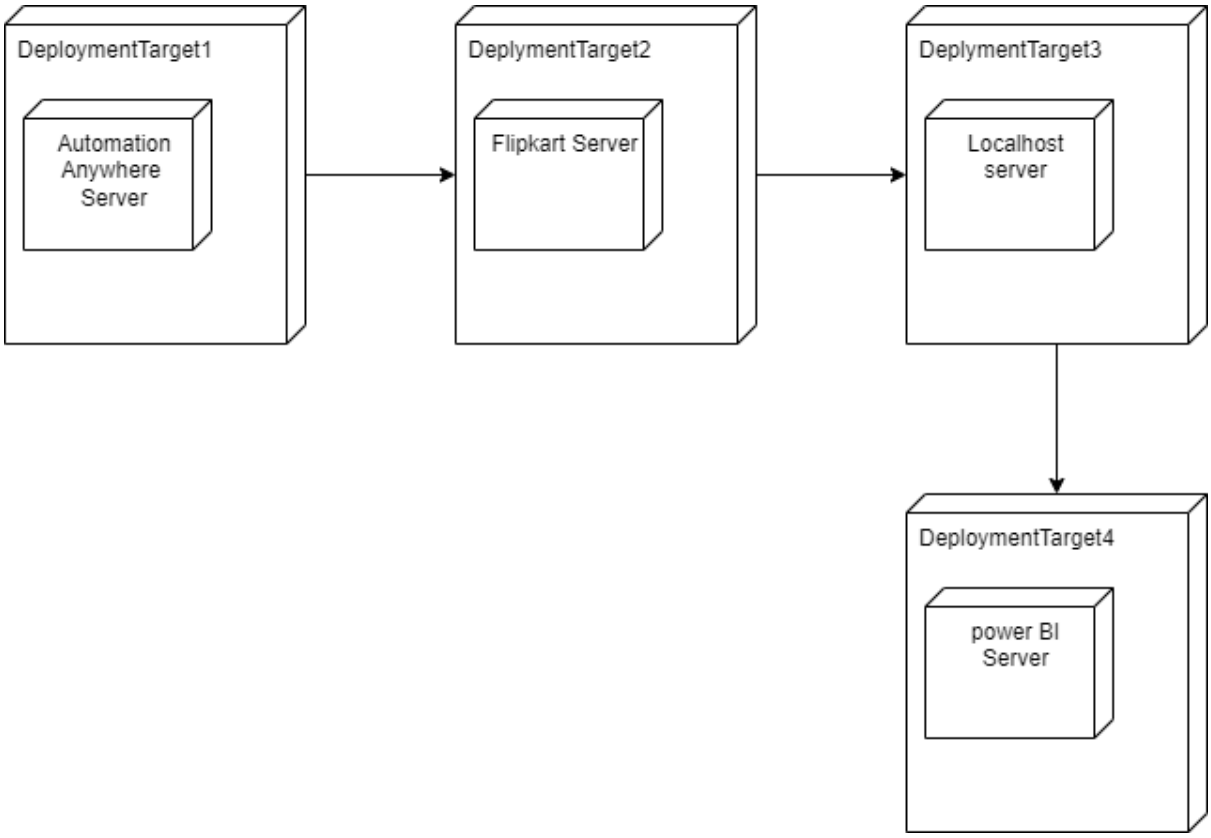
4.4.3 Activity Diagram



4.4.4 COMPONENT DIAGRAM



4.4.5 DEPLOYMENT DIAGRAM



5 PROJECT PLAN

5.1 Project Estimate

1. RPA Development:

-Duration: The duration for scraping the mobiles from the Flipkart website can be estimated by multiplying the number of mobiles (3114) by an average scraping time per mobile (e.g., 5 seconds). This will give you the total scraping time. You can then convert it to the appropriate time unit (e.g., hours or days) based on the expected execution time for each mobile.\

- Effort: The effort required for RPA development depends on the complexity of the scraping process and the expertise of the team members involved. It is recommended to assess the scraping requirements in detail to determine the level of effort required.

2. Data Pre-processing and Analysis:

- Duration: The provided code for data pre-processing and analysis seems to be relatively straightforward and efficient. The execution time will mainly depend on the size of the dataset. Given that the dataset has 3114 entries, the execution time should be minimal.

- Effort: Data pre-processing and analysis typically involve handling missing values, performing basic data exploration, and calculating summary statistics. The effort required for this stage is generally moderate.

3. Data Visualization:

- Duration: The provided code for data visualization utilizes matplotlib and seaborn libraries, which are commonly used for creating plots. The execution time for visualization will depend on the number and complexity of the visualizations required. It is recommended to identify the specific visualizations needed and estimate their individual execution times.

- Effort: The effort required for data visualization is generally moderate. It involves creating scatter plots, bar charts, pair plots, and heatmaps based on the dataset columns. The complexity and number of visualizations will influence the overall effort required.

4. Sentiment Analysis:

- Duration: The provided code for sentiment analysis utilizes the TextBlob library for performing sentiment analysis on the "Review" column. The execution time will depend on the length and number of reviews in the dataset. It is recommended to assess the average review length and estimate the sentiment analysis time accordingly.

- Effort: The effort required for sentiment analysis is generally moderate. It involves performing sentiment analysis on the text data and assigning sentiment scores to each review. Consider the expertise of team members in natural language processing (NLP) tasks.

5. Price Prediction (Machine Learning):

- Duration: The provided code for price prediction uses linear regression to predict the selling price of mobiles based on the dataset. The execution time will depend on the size of the dataset and the complexity of the linear regression model. Larger datasets or more complex models may require additional time.

- Effort: The effort required for price prediction using machine learning is moderate. It involves tasks such as splitting the dataset into training and testing sets, handling missing values, creating a linear regression model, training the model, making predictions, and evaluating the model's performance. Consider the expertise of team members in machine learning and regression techniques.

6. Clustering (Machine Learning):

- Duration: The provided code for clustering utilizes the K-means clustering algorithm to cluster the data based on different columns. The duration will depend on the size of the dataset and the complexity of the clustering process. Larger datasets or higher-dimensional data may increase the execution time.

- Effort: The effort required for clustering using machine learning is moderate. It involves preparing the data for clustering, applying one-hot encoding or label encoding if necessary, scaling the data, creating the clustering model, fitting the model, and visualizing the clusters. Consider the expertise of team members in machine learning and clustering techniques.

Please note that these estimates provide a general outline of the project effort and duration based on the provided code and information. It is crucial to perform a detailed analysis of your specific project requirements and consult with the relevant

5.1.1 Project Resources

1. RPA Development:

- Development team: Assign one or more developers proficient in web scraping techniques and experienced with RPA tools such as UiPath, Automation Anywhere, or Blue Prism.

- RPA software: Provide the selected RPA software tool along with any necessary licenses or subscriptions.

- Hardware: Ensure that the development team has access to computers or servers with sufficient processing power and memory to handle the scraping tasks efficiently.

2. Data Pre-processing and Analysis:

- Data analysts: Assign data analysts who are skilled in data cleaning, handling missing values, and performing exploratory data analysis (EDA).

- Software tools: Provide the necessary software tools for data pre-processing and analysis, such as Python with libraries like Pandas, NumPy, and Scikit-learn. Ensure that the required versions of the libraries are installed.

3. Data Visualization:

- Data visualization experts: Assign individuals with expertise in data visualization techniques and tools like matplotlib and seaborn.

- Software tools: Provide the required data visualization software, such as Python libraries or tools like Tableau or Power BI. Ensure that the necessary software licenses or subscriptions are available.

4. Sentiment Analysis:

- Natural Language Processing (NLP) experts: Assign individuals with expertise in NLP and sentiment analysis techniques.

- Software tools: Provide software tools or libraries for sentiment analysis, such as TextBlob or NLTK in Python. Ensure that the required versions of the libraries are installed.

5. Price Prediction (Machine Learning):

- Machine learning experts: Assign individuals with knowledge and experience in machine learning algorithms, particularly regression techniques like linear regression.

- Software tools: Provide the necessary software tools for machine learning, such as Python with libraries like Scikit-learn or TensorFlow. Ensure that the required versions of the libraries are installed.

6. Clustering (Machine Learning):

- Machine learning experts: Assign individuals with knowledge and experience in clustering algorithms, specifically K-means clustering.

- Software tools: Provide software tools for machine learning and clustering, such as Python libraries like Scikit-learn. Ensure that the required versions of the libraries are installed.

Additionally, consider assigning a project manager or team lead who can coordinate the efforts of the various team members, ensure timely completion of tasks, and handle any project-related issues. The project manager should have good communication and leadership skills to effectively manage the resources and ensure project success.

5.2 Risk Management

1. Technical Risks:

- Data source changes: Regularly monitor the data source and implement error handling mechanisms to handle changes in the website structure or format. Keep the scraping code flexible and adaptable to accommodate such changes.

- Tool limitations: Stay updated with software releases, follow best practices, and have contingency plans in place to switch to alternative tools if the current RPA software or libraries present limitations or encounter bugs.

2. Data Quality Risks:

- Incomplete or inaccurate data: Implement data validation checks, use multiple data sources for cross-referencing, and conduct regular quality checks to ensure the accuracy and completeness of the scraped data.

3. Legal and Ethical Risks:

- Web scraping legality: Review and understand the legal requirements and terms of service of the targeted websites. Ensure compliance with data scraping laws and regulations, and respect the website's terms of use and robots.txt file.

- Ethical considerations: Consider the ethical implications of scraping data, particularly when dealing with personal or sensitive information. Anonymize or aggregate data when necessary to protect privacy rights.

4. Project Management Risks:

- Scope creep: Establish a clear project scope and have a change management process in place to evaluate and prioritize new requests or requirements. Regularly review and communicate the scope to stakeholders to manage expectations.

- Resource constraints: Conduct resource planning and allocate sufficient time and personnel for the project. Regularly review the project schedule and address any resource constraints or competing priorities.

5. Model Performance Risks:

- Prediction accuracy: Continuously monitor and evaluate the performance of machine learning models used for sentiment analysis, price prediction, and clustering. Consider using alternative algorithms, fine-tuning model parameters, and incorporating feedback loops to improve accuracy.

6. Stakeholder Risks:

- Misalignment of expectations: Establish clear communication channels with stakeholders and hold regular progress meetings to ensure everyone is aligned regarding project outcomes and timelines. Provide regular status updates to manage stakeholder expectations effectively.

5.2.1 Risk Identification

- Review project documentation, conduct brainstorming sessions, and analyze historical data to identify potential risks.

- Categorize risks into different types (e.g., technical, operational, legal, financial) for comprehensive coverage.

- Document risks in a risk register or risk log, including descriptions, potential impacts, and likelihood of occurrence.

5.2.2 Risk Analysis

- Assess identified risks based on their potential impact and likelihood of occurrence.

- Prioritize risks by considering severity, potential to affect project objectives, and available resources.

- Assign risk owners or responsible individuals for effective risk management.

5.2.3 Overview of Risk Mitigation, Monitoring, Management

Risk Mitigation:

- Develop risk mitigation strategies and actions for each identified risk to reduce probability or impact.
- Assign responsibilities for implementing mitigation actions and establish clear timelines.
- Utilize risk avoidance, risk transfer, risk reduction, and risk acceptance strategies as appropriate.

Risk Monitoring:

- Regularly monitor identified risks throughout the project lifecycle.
- Track changes in risk likelihood and impact, updating the risk register accordingly.
- Implement mechanisms for early warning and detection of new risks.

Risk Management:

- Establish a risk management plan outlining the overall approach, roles and responsibilities, communication channels, and escalation procedures.
- Conduct regular risk review meetings to discuss the status of identified risks, mitigation actions, and new risks.
- Maintain open communication with stakeholders, ensuring awareness of the project's risk profile and potential impacts.

5.3 Project schedule

The project schedule is a critical component of project management as it provides a roadmap for executing and completing the project within the defined timeframe. Here are the steps to develop a professional project schedule:

1. Identify Project Tasks: Break down the project scope into smaller, manageable tasks. Identify all the activities required to achieve the project objectives. Each task should be defined with a clear description and deliverables.

2. Estimate Task Durations: Assign estimated durations to each task based on factors such as complexity, resources required, and historical data. Consider input from relevant team members to ensure accurate estimations.

3. Define Task Dependencies: Determine the dependencies between tasks. Identify tasks that need to be completed before others can start (predecessors) or tasks that can only start after certain tasks are completed (successors).

4. Assign Resources: Assign responsible individuals or teams to each task. Consider their availability, expertise, and workload to ensure realistic resource allocation. Clearly communicate roles and responsibilities to the assigned resources.

5. Create Task Schedule: Use project management software or tools like Gantt charts to create a visual representation of the project schedule. Incorporate the task list, durations, and dependencies into the schedule.

6. Identify Critical Path: Analyze the task dependencies to determine the critical path. The critical path is the longest sequence of dependent tasks that determines the overall project duration. Tasks on the critical path must be closely monitored as delays can impact the project timeline.

7. Include Milestones: Identify key milestones in the project, which represent significant achievements or deliverables. Milestones act as important checkpoints for tracking project progress and ensuring alignment with the schedule.

8. Account for Contingencies: Factor in contingencies and buffers to accommodate potential delays, risks, or changes. This helps in maintaining schedule flexibility and managing unforeseen circumstances.

9. Regularly Monitor and Update: Continuously monitor the project's progress against the schedule. Track task completion, identify any delays or deviations, and update the schedule as needed. Communicate any schedule changes to the team and stakeholders.

10. Communicate and Coordinate: Regularly communicate the project schedule with the project team, stakeholders, and relevant parties. Ensure that everyone understands the timeline, dependencies, and their respective roles in meeting the schedule.

5.4 Timeline Chart

Sr. No.	Objective	Duration
1	Define Objective & Problem Statement	3 weeks
2	Doing Research on Problem Statement	4 weeks
3	Studying IEEE & various journal paper	3 weeks
4	Deciding Base Paper	1 weeks
5	Deciding Tools & technologies to use for project	2 weeks
6	Building System Design	4 weeks
7	Data gathering for our problem statement using RPA and storing into .csv file	2 weeks
8	Fetching .csv file and fitting various ML algorithms on it.	4 weeks

6 PROJECT IMPLEMENTATION

6.1 Overview of Project Modules

There are 5 modules in our project which are

1) Robotic Process Automation
2) Sentiment Analysis
3) Price Prediction
4) Clustering
5) Power Bi

In these 5 modules we are fetching the required data from websites such as flipkart, Amazon. After fetching data, we are prepressing that data & applying various machine learning algorithms such as text blob for sentiment analysis, Linear Regression for price prediction, K-Means for clustering. After all these process we are uploading whole data to power bi and generating final report.

6.2 Tools and Technologies Used

6.2.1 Tools

1. Automation 360
2. Google Colab / Jupyter
3. Microsoft Office
4. Microsoft Power Bi Desktop

6.2.2 Technologies

1. Robotic Process Automation
2. Python
2. Machine Learning
3. Power Bi

6.3 Algorithm Details

6.3.1 Data Visualization Algorithm

Data visualization using various plots and charts. Here's the algorithm used in the code:

Import the required libraries for data visualization, namely matplotlib.pyplot and seaborn.

Scatter plot: Create a scatter plot using the scatter() function from matplotlib.pyplot. It plots the 'Selling Price' on the x-axis and 'Original Price' on the y-axis. Set the title, x-label, and y-label using the respective functions, and display the plot using plt.show().

Bar chart: Calculate the average rating by brand using the groupby() and mean() functions. Create a bar chart using the bar() function, where the x-axis represents the brand names obtained from avg_rating.index, and the y-axis represents the average rating values obtained from avg_rating.values. Set the title, x-label, and y-label using the respective functions, and display the chart using plt.show().

Pairplot: Create a pairplot using the pairplot() function from seaborn. The pairplot visualizes the relationships between the 'Memory', 'Storage', and 'Rating' variables in the DataFrame df. Display the pairplot using plt.show().

Distribution plot: Create a distribution plot using the displot() function from seaborn. It plots the distribution of the 'Rating' variable in the DataFrame df. Display the distribution plot using plt.show().

Heatmap: Calculate the correlation matrix between the numerical columns ('Memory', 'Storage', 'Rating', 'Selling Price', 'Original Price') using the corr() function. Create a heatmap using the heatmap() function from seaborn. Set the annot parameter to True to display the correlation values on the heatmap. Set the title and display the heatmap using plt.show().

This algorithm combines various plotting and visualization techniques to explore and present data relationships, distributions, and correlations.

6.3.2 Sentiment Analysis

Sentiment analysis on the 'Review' column of a DataFrame using the TextBlob library. Here's the algorithm used in the code:

Import the required library TextBlob for sentiment analysis.

Select the 'Review' column from the DataFrame df and store it in the variable 'reviews'.

Initialize an empty list sentiment_scores to store the sentiment scores.

Iterate through each review in the 'reviews' list.

Create a TextBlob object text_blob for each review.

Calculate the polarity of the review using the sentiment.polarity property of the TextBlob object. The polarity value ranges from -1 to 1, representing negative to positive sentiment.

Map the polarity value to a numerical scale from 1 to 5 by adding 1 to the polarity and multiplying by 2.5. This rescales the polarity to fit the desired range of sentiment scores.

Round the sentiment score using the round() function to obtain a whole number.

Append the sentiment score to the sentiment_scores list.

Add the 'Sentiment' column to the DataFrame df with the sentiment scores obtained.

Print the updated DataFrame df with the added 'Sentiment' column.

This algorithm performs sentiment analysis on each review in the 'Review' column, assigns a sentiment score from 1 to 5, and adds the scores as a new column 'Sentiment' to the DataFrame.

6.3.3 Linear Regression

Steps in Linear Regression: To implement the Linear Regression using Python, we will use the same steps as we have done in previous topics of Regression. Below are the steps:

1. Data Pre-processing step
2. Fitting Linear Regression to the Training set
3. Predicting the test result
4. Test accuracy of the result(Creation of Confusion matrix)
5. Visualizing the test set result.

6.3.4 Clustering

Clustering analysis on a dataset using K-means clustering. Here's the algorithm used in the code:

1. Import the necessary libraries: `sklearn.cluster.KMeans`, `sklearn.preprocessing.StandardScaler`, and `sklearn.impute.SimpleImputer`.

2. Load the dataset from a CSV file into the DataFrame `dataset`.

3. Specify the columns to be used for clustering in the `columns_for_clustering` list.

4. Prepare the data for clustering by selecting the columns specified in `columns_for_clustering` and assigning them to the variable `X`.

5. Iterate over each column in `X` for clustering.

6. Prepare the data for clustering by selecting the data for the current column (`column_data`).

7. Perform one-hot encoding for categorical variables in the `column_data` using `pd.get_dummies()`.

8. Impute missing values in the `column_data` using `SimpleImputer` with the strategy set to 'mean'.

9. Scale the data in the `column_data` using `StandardScaler`.

10. Specify the number of clusters (`k`) and create a K-means clustering model using `KMeans` with the specified number of clusters and a random state of 42.

11. Fit the K-means model to the scaled data.

12. Retrieve the cluster labels assigned to each data point.

13. Add the cluster labels as a new column in the `dataset` DataFrame by concatenating the column name (`col`) and '_Cluster'. Convert the cluster labels to strings.

14. Print the updated `dataset` with the added cluster columns.

15. Visualize the clusters for each column using scatter plots.

16. Iterate over each column in `X` for visualization.

17. Retrieve the unique cluster labels for the current column.

18. Create a colormap based on the number of unique cluster labels.

19. Map each cluster label to a color.

20. Retrieve the colors for each data point based on the cluster labels.

21. Create scatter plots to visualize the clusters for each column, where the x-axis represents the current column, and the y-axis represents 'Selling Price', 'Original Price', 'Memory', or 'Storage', respectively.

22. Set the colors of the data points based on the cluster labels.

23. Set the x-label, y-label, and title of each scatter plot.

24. Display the scatter plots using `plt.show()`.

This algorithm performs K-means clustering on each column specified for clustering, adds the cluster labels as new columns to the dataset, and visualizes the clusters using scatter plots for each column.

7 SOFTWARE TESTING

7.1 Types of Testing

Testing is a critical component of the project to ensure the quality and reliability of the software or system being developed. Here are some common types of testing:

1. Unit Testing: Focuses on testing individual components or units of the software in isolation. It helps verify the correctness of each unit's functionality.
2. Integration Testing: Verifies the interaction and communication between different components to ensure they work together correctly. It helps identify issues that may arise from integrating different modules or subsystems.
3. System Testing: Tests the entire system as a whole to validate its compliance with functional and non-functional requirements. It ensures that the system operates as intended and meets the specified criteria.
4. Acceptance Testing: Involves testing the system with real users to determine if it meets their needs and requirements. It aims to gain user confidence and acceptance before the system is deployed.
5. Performance Testing: Assesses the system's performance under various conditions, such as load, stress, or scalability testing. It ensures that the system performs well and handles expected workloads.
6. Security Testing: Identifies vulnerabilities and weaknesses in the system's security measures to ensure data protection. It includes testing for authentication, authorization, encryption, and other security aspects.
7. Regression Testing: Re-tests previously tested functionality to ensure that changes or fixes do not introduce new issues. It helps maintain the overall quality of the system as it evolves.
8. User Interface (UI) Testing: Checks the usability and effectiveness of the user interface. It focuses on aspects like layout, responsiveness, and user interactions.
9. Exploratory Testing: Involves ad-hoc testing without predefined test cases to discover defects through exploration. Testers rely on their experience, intuition, and domain knowledge to identify issues.
10. Compatibility Testing: Ensures that the system functions correctly across different devices, browsers, or operating systems. It verifies compatibility to deliver a consistent user experience.

The choice of testing types depends on the project's nature, requirements, and associated risks. A comprehensive testing strategy usually involves a combination of these types to achieve thorough coverage.

7.2 Test cases & Test Results

Test cases and test results play a crucial role in the testing process. Here's an overview of the steps involved:

1. Identify Test Scenarios: Based on the project requirements and testing objectives, identify the key scenarios that need to be tested. This helps prioritize testing efforts.
2. Create Test Cases: For each test scenario, create detailed test cases. A test case typically includes the following elements:
    - Test case ID: A unique identifier for the test case.
    - Test case description: A clear and concise description of the test case.
    - Test steps: Step-by-step instructions to execute the test.
    - Expected results: The expected outcome or behavior of the system.
    - Test data: Any specific data required for the test case.
3. Execute Test Cases: Follow the test steps defined in the test cases and execute the tests on the system or software being tested.
4. Compare Results: Compare the actual results obtained from executing the test cases with the expected results stated in the test cases. Identify any discrepancies or deviations.
5. Defect Reporting: If there is a discrepancy between the actual and expected results, report it as a defect or bug in a designated defect tracking system. Provide clear and detailed information about the issue.
6. Retest and Verify: Once defects are fixed, retest the affected test cases to ensure the issues have been resolved successfully. Verify that the expected results are now achieved.
7. Record Test Results: Document the actual results for each executed test case, including any defects found, resolutions, and other relevant information. Maintain proper documentation for future reference and traceability.

By following this systematic approach, test teams can ensure that the software or system is thoroughly tested, defects are identified and resolved, and the project meets the required quality standards.

8 RESULTS

8.1 Outcomes

For Robotic Process Automation :

Certainly! RPA (Robotic Process Automation) is a technology that allows automation of repetitive tasks by using software robots or bots. In the context of obtaining data for the Flipkart_Data.csv file, RPA can be used to automate the process of extracting information from the Flipkart website.

Here's a general overview of how RPA can be used to obtain data from Flipkart:

1. Identifying the target data : Determine the specific information you want to extract from the Flipkart website. For example, you may be interested in collecting data about mobile phones, including their brand, model, colour, storage, price, etc.

2. Designing the automation workflow : Using an RPA tool or framework, design the workflow for the data extraction process. This typically involves creating a sequence of steps that mimic human interactions with the website.

3. Navigating to the Flipkart website : The automation starts by launching a web browser and navigating to the Flipkart website.

4. Searching for the desired products : Using the search functionality of the Flipkart website, the automation enters relevant search keywords (e.g., "mobile phones") and initiates the search.

5. Iterating through the search results : The automation script iterates through the search results pages, extracting the necessary information for each product listed. This may involve clicking on product links, accessing product details pages, and retrieving the required data such as brand, model, color, storage, price, etc.

6. Storing the extracted data : As the automation extracts the data for each product, it stores the information in memory or a data structure. Once all the desired products have been processed, the data is compiled into a structured format, such as a CSV file.

7. Exporting data to Flipkart_Data.csv : Finally, the automation exports the collected data to a CSV file named Flipkart_Data.csv. The CSV file will have columns corresponding to the extracted data fields, with each row representing a product.

It's important to note that the specific implementation details may vary depending on the RPA tool or framework used, as well as the structure and behavior of the Flipkart website. RPA developers typically utilize techniques like web scraping, data extraction, and automation capabilities provided by the chosen RPA platform to accomplish the task.

For Sentiment Analysis :

The given code performs sentiment analysis on the 'Review' column of a dataset using the TextBlob library. It calculates sentiment scores for each review and maps the sentiment to a numerical scale from 1 to 5. The outcome of running this code on the 'Flipkart_Data.csv' dataset would be the dataset 'df' with an additional column named 'Sentiment' that contains the sentiment scores for each review.

The 'Sentiment' column will be populated with numerical values ranging from 1 to 5, representing the sentiment of each review. The sentiment scores are calculated based on the polarity of the text using the TextBlob library. The sentiment scores are then rounded and scaled to fit within the range of 1 to 5.

The resulting 'df' dataset will have the original columns from the 'Flipkart_Data.csv' file along with the newly added 'Sentiment' column. This column will reflect the sentiment of each review, allowing for further analysis or visualization of the sentiment distribution within the dataset.

For Price Prediction :

The given code performs price prediction using a Linear Regression model on the 'Flipkart_Data.csv' dataset. However, there is an issue with the code because the 'df' variable is not defined before the code snippet.

Assuming that the 'df' variable represents the dataset loaded from the 'Flipkart_Data.csv' file, here is a description of the expected outcome of running this code:

Data Pre-processing :

The code converts the 'Memory' and 'Storage' columns to string type.

It removes specific strings ('MB', 'GB', 'Expandable Upto', 'TB') from the 'Memory' and 'Storage' columns and converts them to float values.

Data Visualization and Preparation:

The code imports necessary libraries for data visualization and model training.

It performs one-hot encoding on categorical variables in the dataset using the get_dummies() function from pandas.

The dataset is split into training and testing sets using the train_test_split() function from scikit-learn.

Data Imputation:

The code uses a SimpleImputer from scikit-learn to impute missing values in the training and testing sets.

The missing values are filled based on the mean of the respective columns.

Model Training and Prediction:

A Linear Regression model is created using the LinearRegression() class from scikit-learn.

The model is trained on the training set using the fit() function.

Predictions are made on the testing set using the trained model, and the predicted values are stored in the 'y_pred' variable.

Model Evaluation:

The code calculates the mean squared error (MSE) between the predicted values ('y_pred') and the actual values ('y_test') using the mean_squared_error() function from scikit-learn.

The MSE is printed as 'Mean Squared Error'.

Data Visualization:

The code creates a scatter plot to visualize the predicted prices ('y_pred') vs. the actual prices ('y_test') using the scatter() function from matplotlib.

The x-axis represents the actual prices, and the y-axis represents the predicted prices.

The plot is displayed using the show() function from matplotlib.

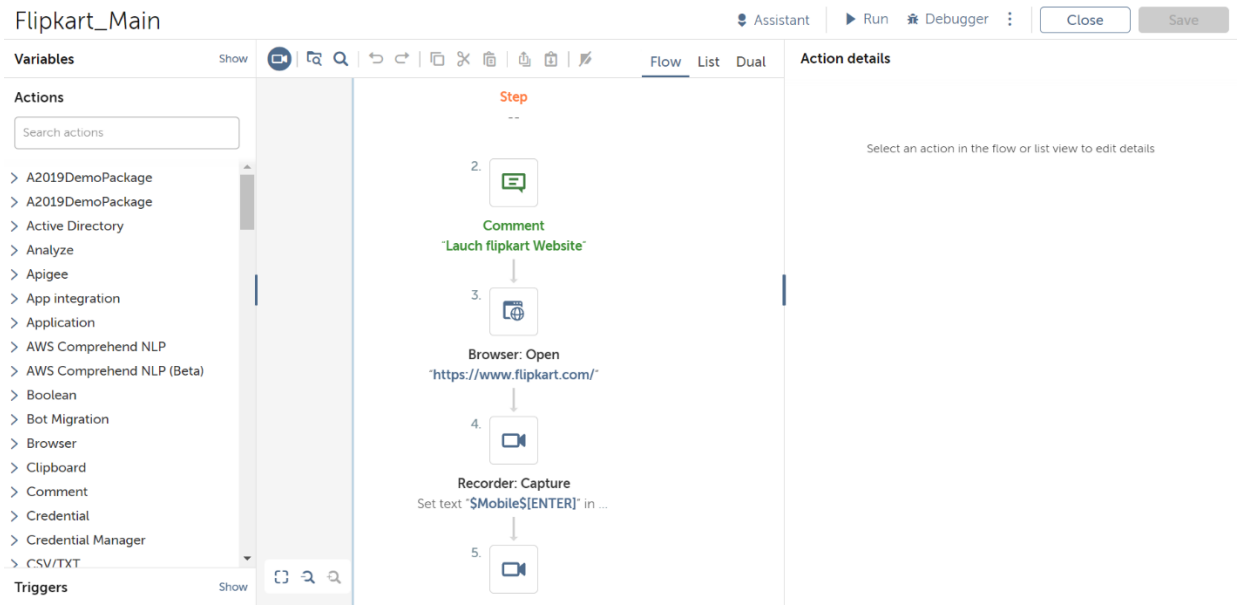
The outcome of running this code would be the printed mean squared error (MSE) value, which indicates the performance of the linear regression model in predicting the prices. Additionally, a scatter plot is displayed to visualize the relationship between the predicted and actual prices, providing an insight into the accuracy of the model's predictions.

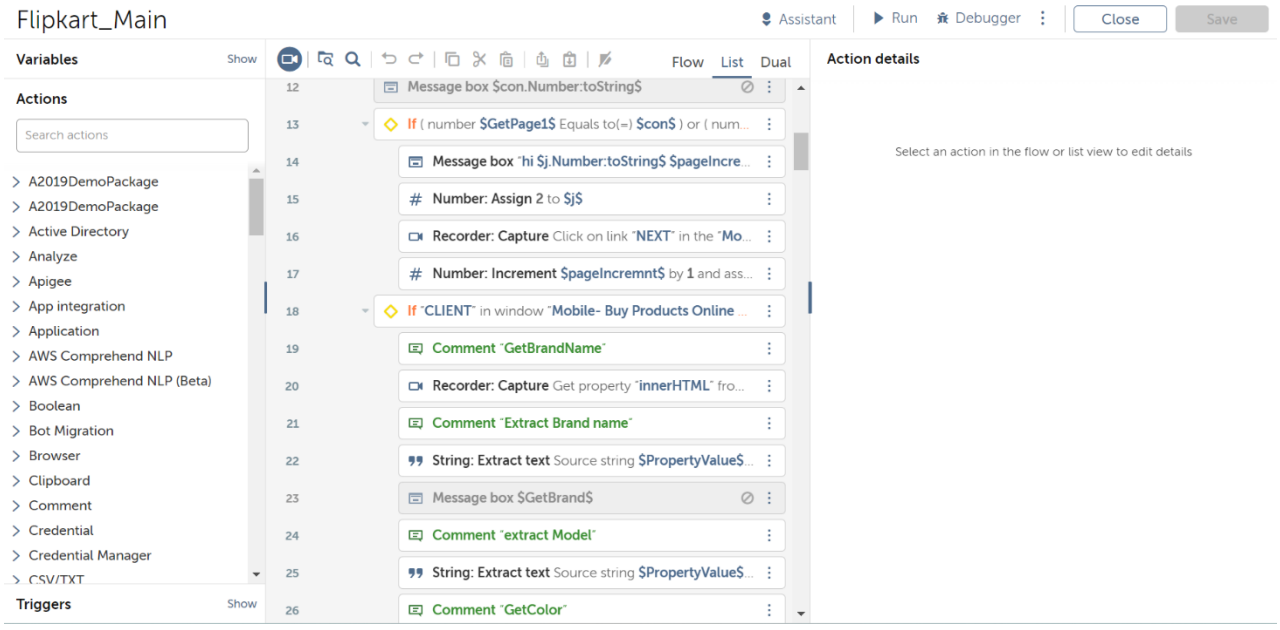
For Clustering :

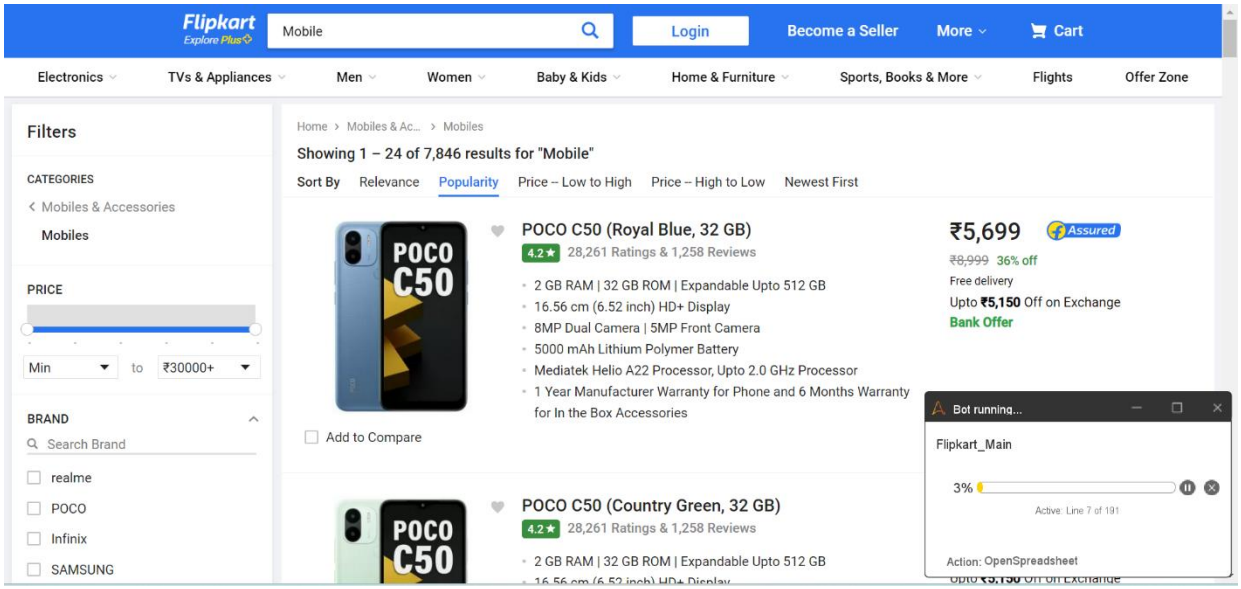
The code provided performs clustering on the "Flipkart_Data.csv" dataset using the K-means algorithm. Here is the outcome of the code:

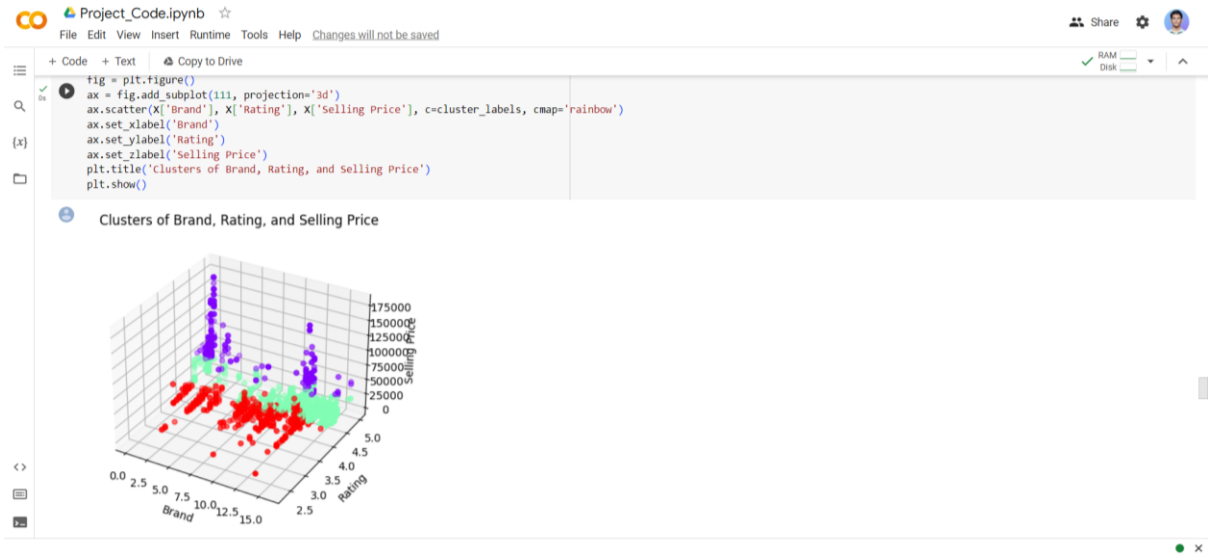
1. The dataset is loaded from the "Flipkart_Data.csv" file.
2. The selected columns for clustering are 'Brand', 'Model', 'Color', 'Memory', 'Storage', and 'Rating'.
3. The code iterates over each column and performs the following steps for each column:
   - Performs one-hot encoding for categorical variables.
   - Imputes missing values using the mean strategy.
   - Scales the data using StandardScaler.
   - Creates a K-means clustering model with 3 clusters.
   - Fits the model to the data.
   - Gets the cluster labels for each data point.
   - Adds the cluster labels as a new column in the dataset.
4. The updated dataset with the added cluster columns is printed.
5. The code then proceeds to visualize the clusters for each column:
   - For each column, it creates a colormap based on the unique cluster labels.
   - It maps each cluster label to a color.
   - It retrieves the colors for each data point based on the cluster labels.
   - It creates scatter plots to visualize the clusters for 'Selling Price' and 'Original Price' against the current column.
   - It creates scatter plots to visualize the clusters for 'Memory' and 'Storage' against the current column.
6. Finally, the code visualizes the clusters in 3D using scatter plots:
   - For the first code snippet, it plots 'Rating', 'Selling Price', and 'Original Price' in a 3D scatter plot.

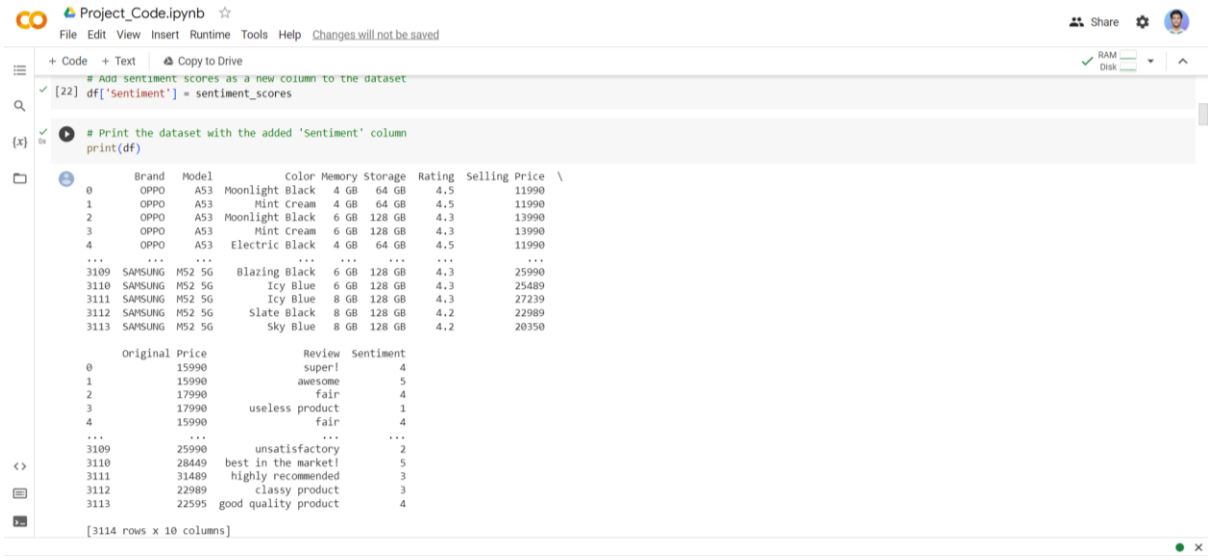
8.2 Screen Shots

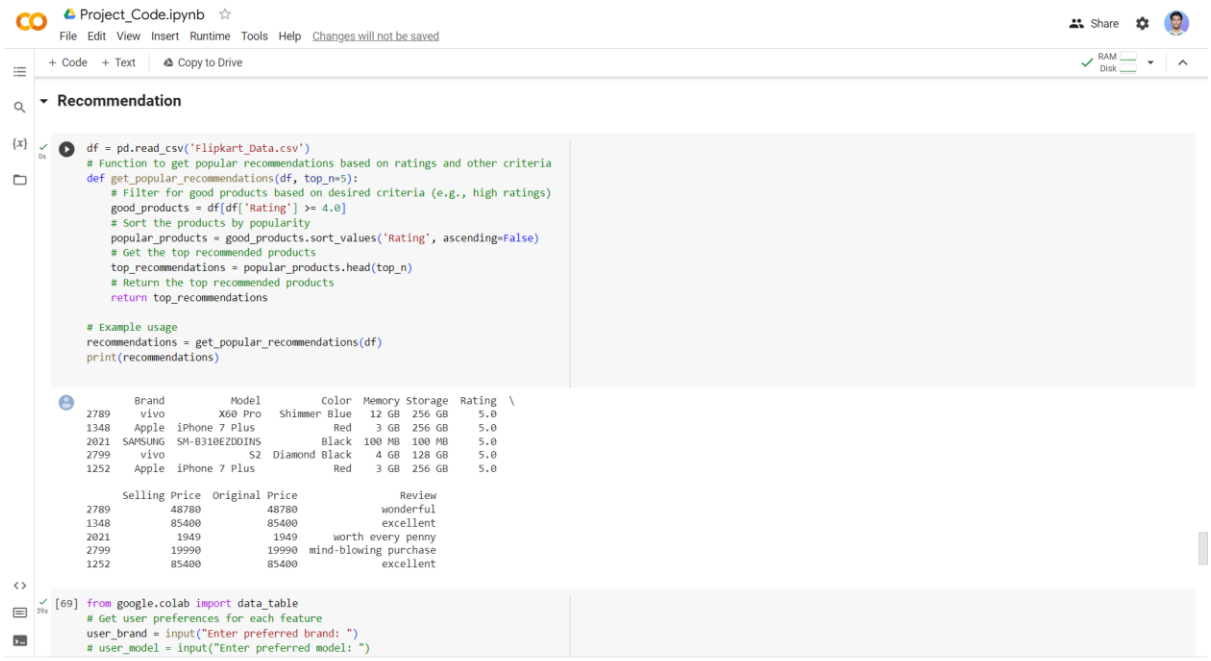


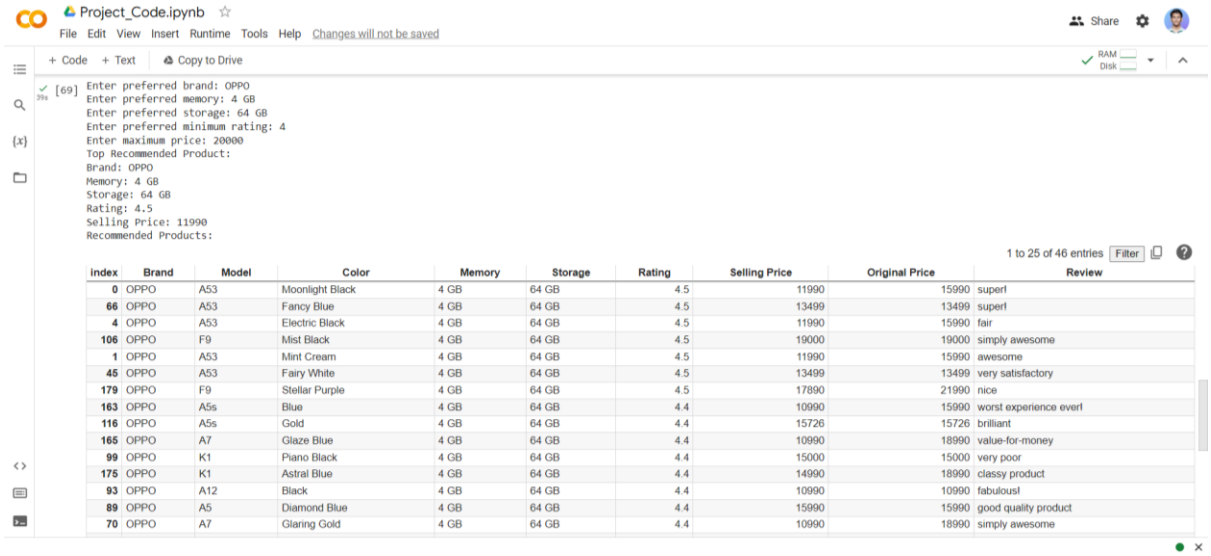



Flipkart
Explore Plus
Mobile
Login
Become a Seller
More
Cart
Electronics
TVs & Appliances
Men
Women
Baby & Kids
Home & Furniture
Sports, Books & More
Flights
Offer Zone
Filters
CATEGORIES
Mobiles & Accessories
Mobiles
PRICE
Min
to
₹30000+
BRAND
Search Brand
realme
POCO
Infinix
SAMSUNG
Home › Mobiles & Ac... › Mobiles
Showing 1 – 24 of 7,846 results for "Mobile"
Sort By
Relevance
Popularity
Price -- Low to High
Price -- High to Low
Newest First
POCO C50 (Royal Blue, 32 GB)
4.2★ 28,261 Ratings & 1,258 Reviews
2 GB RAM | 32 GB ROM | Expandable Upto 512 GB
16.56 cm (6.52 inch) HD+ Display
8MP Dual Camera | 5MP Front Camera
5000 mAh Lithium Polymer Battery
Mediatek Helio A22 Processor, Upto 2.0 GHz Processor
1 Year Manufacturer Warranty for Phone and 6 Months Warranty for In the Box Accessories
Add to Compare
₹5,699
Assured
₹8,999 36% off
Free delivery
Upto ₹5,150 Off on Exchange
Bank Offer
POCO C50 (Country Green, 32 GB)
4.2★ 28,261 Ratings & 1,258 Reviews
2 GB RAM | 32 GB ROM | Expandable Upto 512 GB
Bot running...
Flipkart_Main
3%
Active: Line 7 of 191
Action: OpenSpreadsheet


Project_Code.ipynb
File Edit View Insert Runtime Tools Help Changes will not be saved
Share
+ Code + Text Copy to Drive
RAM
Disk
fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')
ax.scatter(X['Brand'], X['Rating'], X['Selling Price'], c=cluster_labels, cmap='rainbow')
ax.set_xlabel('Brand')
ax.set_ylabel('Rating')
ax.set_zlabel('Selling Price')
plt.title('Clusters of Brand, Rating, and Selling Price')
plt.show()
Clusters of Brand, Rating, and Selling Price
Selling Price
Brand
Rating


Project_Code.ipynb
File Edit View Insert Runtime Tools Help Changes will not be saved
Share
+ Code + Text Copy to Drive
RAM
Disk
Add sentiment scores as a new column to the dataset
df['Sentiment'] = sentiment_scores
# Print the dataset with the added 'Sentiment' column
print(df)
Brand Model Color Memory Storage Rating Selling Price Original Price Review Sentiment
0 OPPO A53 Moonlight Black 4 GB 64 GB 4.5 11990 15990 super! 4
1 OPPO A53 Mint Cream 4 GB 64 GB 4.5 11990 15990 awesome 5
2 OPPO A53 Moonlight Black 6 GB 128 GB 4.3 13990 17990 fair 4
3 OPPO A53 Mint Cream 6 GB 128 GB 4.3 13990 17990 useless product 1
4 OPPO A53 Electric Black 4 GB 64 GB 4.5 11990 15990 fair 4
3109 SAMSUNG M52 5G Blazing Black 6 GB 128 GB 4.3 25990 25990 unsatisfactory 2
3110 SAMSUNG M52 5G Icy Blue 6 GB 128 GB 4.3 25489 28449 best in the market! 5
3111 SAMSUNG M52 5G Icy Blue 8 GB 128 GB 4.3 27239 31489 highly recommended 3
3112 SAMSUNG M52 5G Slate Black 8 GB 128 GB 4.2 22989 22989 classy product 3
3113 SAMSUNG M52 5G Sky Blue 8 GB 128 GB 4.2 20350 22595 good quality product 4
[3114 rows x 10 columns]





index	Brand	Model	Color	Memory	Storage	Rating	Selling Price	Original Price	Review
0	OPPO	A53	Moonlight Black	4 GB	64 GB	4.5	11990	15990	superl
66	OPPO	A53	Fancy Blue	4 GB	64 GB	4.5	13499	13499	superl
4	OPPO	A53	Electric Black	4 GB	64 GB	4.5	11990	15990	fair
106	OPPO	F9	Mist Black	4 GB	64 GB	4.5	19000	19000	simply awesome
1	OPPO	A53	Mint Cream	4 GB	64 GB	4.5	11990	15990	awesome
45	OPPO	A53	Fairy White	4 GB	64 GB	4.5	13499	13499	very satisfactory
179	OPPO	F9	Stellar Purple	4 GB	64 GB	4.5	17890	21990	nice
163	OPPO	A5s	Blue	4 GB	64 GB	4.4	10990	15990	worst experience ever!
116	OPPO	A5s	Gold	4 GB	64 GB	4.4	15726	15726	brilliant
165	OPPO	A7	Glaze Blue	4 GB	64 GB	4.4	10990	18990	value-for-money
99	OPPO	K1	Piano Black	4 GB	64 GB	4.4	15000	15000	very poor
175	OPPO	K1	Astral Blue	4 GB	64 GB	4.4	14990	18990	classy product
93	OPPO	A12	Black	4 GB	64 GB	4.4	10990	10990	fabulous!
89	OPPO	A5	Diamond Blue	4 GB	64 GB	4.4	15990	15990	good quality product
70	OPPO	A7	Glaring Gold	4 GB	64 GB	4.4	10990	18990	simply awesome

9 CONCLUSIONS

9.1 Conclusion

We have successfully completed the objectives of our project. We have fetched data from flipkart website using RPA. We have applied various Machine Learning algorithms such as text blob for sentiment analysis, Linear Regression for price prediction, K-Means for clustering. After all these processes we are uploading whole data to power bi and generating final report.

9.2 Future Work

RPA, ML are such a powerful technology in today's era which we can use in lot of sectors such as to full fill purchase & supply chain management, E-Commerce websites, production house , healthcare, education, business sectors which study the customer need , almost in every sectors we can use RPA, ML to increase the productivity of their sectors.

9.3 Applications

a) **Customer service.**

RPA helps companies provide better customer service by automating contact center tasks, including verifying e-signatures, uploading scanned documents and verifying information for automatic approvals or rejections.

b) **Accounting.**

Organizations use RPA for general accounting, operational accounting, transactional reporting and budgeting.

c) **Financial services.**

Companies in the financial services industry use RPA for foreign exchange payments, automating account openings and closings, managing audit requests and processing insurance claims.

d) **Healthcare.**

Medical organizations use RPA for handling patient records, claims, customer support, account management, billing, reporting and analytics.

e) **Human resources.**

RPA can automate HR tasks, including onboarding and offboarding, updating employee information and time sheet submission processes.

f) **Supply chain management.**

RPA can be used in supply chain management for procurement, automating order processing and payments, monitoring inventory levels and tracking shipments.

REFERENCES

1. Devansh Hiren Timbadia, Parin Jigishu Shah, Supriya Agrawal, Sughosh Sudhanvan (2020) Robotic Process Automation Through Advance Process Analysis Model
2. Christian Flechsig * , Franziska Anslinger, Rainer Lasch (2020)Robotic Process Automation in purchasing and supply management: A multiple case study on potentials, barriers, and implementation
3. João Garrott Marques Negreirosa, João Alexandre Lobo Marquesb (2019) Microsoft Power BI Desktop for Data Analytics
4. Mohammed Ameer, Simhadri Prem Rahul, Dr.Suneetha Manne (2020) Human Resource Analytics using Power Bi Visualization Tool
5. Ranjitha P1, Spandana M2 (2020) Predictive Analysis for Big Mart Sales Using Machine Learning Algorithms
6. Md. Anisur Rahman Mia*, Rupon Ghosh* (2021) Business Forecasting System using Machine Learning Approach
7. Shilpi Kulshrestha, M. L. Saini (2021) Study for the Prediction of E-Commerce Business Market Growth Using Machine Learning Algorithm
8. Bruno Varela, Jorge Bernardino, Isabel Pedrosa (2020) Twitter Sensitivity Analysis in a Higher School Using Power BI
9. Wei Hu Haiyan Xie, Mantas Nakas, Wei Shi Mengmeng Wang,Xi'an China (2019) Power BI for Impacts Analysis on Cost of Living Caused by Industry Prevalence in Smart Cities
10. Estela Maria Macas Ruiz,Wayner Xavier Bustamante Granda,Diego Andrés Tinitana Orteg,Rommel Vicente Torres Tandazo (2019) Governance of Social Networks in the educational contextanalyzed through Power BI: Case Study Students belonging to the Pio Jaramillo Alvarado Institution.
11. Vijay Krishnan, S Bharanidharan, G Krishnamoorthy (2017) Research Data Analysis with Power BI
12. Ranjitha P1,Spandana M2 (2021) Predictive Analysis for Big Mart Sales Using Machine Learning Algorithms
13. Bernhard Axmann ,Harmoko Harmoko (2020) Robotic Process Automation: An Overview and Comparison to Other Technology in Industry 4.0
14. Vijay Krishnan, S Bharanidharan, G Krishnamoorthy (2017) Research Data Analysis with Power BI.