A REPORT

ON

ARTIFICIAL INTELLIGENCE

Submitted by

Mr. Stalin A - 20231CSE0522

Under the guidance of,

Dr.R.Balakrishnan
Mr.S.Sakthi

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

AT

PRESIDENCY UNIVERSITY

BENGALURU

OCTOBER 2025

PRESIDENCY UNIVERSITY

PRESIDENCY SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the report of **CSE7000 – Internship** on **ARTIFICAL INTELLIGENCE** being submitted by **STALIN A** bearing roll number **20231CSE0522** in partial fulfillment of the requirement for the award of the degree of Bachelor of Technology in **Computer Science and Engineering** is a Bonafide work carried out under our supervision.

Dr.R.BalaKrishnan Assistant Professor PSCS Presidency University	**Mr S. Sakthi** Assistant Professor PSCS Presidency University	**Mr S. Sakthi** Assistant Professor PSCS Presidency University
Dr. Asif Mohamed H.B HOD PSCS Presidency University	**Dr. Shakkeera L** Associate Dean PSCS Presidency University	**Dr. Duraipandian N** Dean PSCS & PSIS Presidency University

PRESIDENCY UNIVERSITY

PRESIDENCY SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

DECLARATION

I hereby declare that the work presented in the report entitled **"ARTIFICIAL INTELLIGENCE"**, in partial fulfillment for the award of the Degree of **Bachelor of Technology in Computer Science and Engineering**, is a record of my own investigations carried out under the guidance of **Dr. R. BalaKrishnan**, Assistant Professor, and **Mr. S. Sakthi**, Assistant Professor, Presidency School of Computer Science and Engineering, **Presidency University, Bengaluru**.

I have not submitted the matter presented in this report anywhere else for the award of any other degree.

Name : Stalin A
Roll No: 20231CSE0522
Signature of Student: .

INTERNSHIP COMPLETION CERTIFICATE


SPARKIIT™

INTERNSHIP CERTIFICATE

THIS CERTIFICATE IS AWARDED TO

Stalin A

has successfully completed Artificial Intelligence Internship for 60 days, from 05 Jun, 2025 to 04 Aug, 2025.

Throughout his Internship, the individual exhibited exceptional perseverance, dedication, diligence, and a thirst for knowledge.

20 Sep, 2025

DATE OF ISSUE

APPROVED BY

wipro dice id

MINISTRY OF CORPORATE AFFAIRS

GOVERNMENT OF INDIA

SPARKIIT EDTECH LLP GUWAHATI

Abhishake Dubey

GENERAL INSTRUCTOR

ACKNOWLEDGEMENT

First of all, I am indebted to the **GOD ALMIGHTY** for giving me an opportunity to excel in my efforts to complete this internship on time.I express sincere thanks to our respected **Dean, Dr. Duraipandian N**, Presidency School of Computer Science and Engineering & Presidency School of Information Science, Presidency University, for getting us permission to undergo the internship.

I express heartfelt gratitude to our beloved **Associate Dean, Dr. Shakkeera L**, Presidency School of Computer Science and Engineering, Presidency University, and **Dr. Asif Mohamed H.B, Head of the Department**, Presidency School of Computer Science and Engineering, Presidency University, for rendering timely help in completing this internship successfully.

I am greatly indebted to my reviewers, **Dr. R. BalaKrishnan, Assistant Professor**, and **Mr. S. Sakthi, Assistant Professor**, Presidency School of Computer Science and Engineering, Presidency University, for their inspirational guidance, valuable suggestions, and for providing a chance to express technical capabilities in every respect for the completion of the internship work.

I would like to convey gratitude and heartfelt thanks to the **Internship Coordinator, Dr. Md Ziaur Rahman**, and **Program Internship Coordinator, Mr. S. Sakthi**. I also thank my **family and friends** for the strong support and inspiration they have provided us in bringing out this internship.

Name : Stalin A

Roll No: 20231CSE0522

ABSTRACT

This internship involved applying AI and Machine Learning techniques to solve four real-world problems, covering the full data science pipeline from data collection to model development and evaluation.

The first project developed an **AI Chatbot using NLP techniques** like text preprocessing, tokenization, and intent classification with transformer-based models. This automated system improved customer engagement and reduced support workload.

The second project focused on forecastin**g Air Quality Index (AQI)** using multi-year pollutant data. Time-series models such as ARIMA, LSTM, and Prophet were tested to predict air quality trends, aiding environmental monitoring and public health.

The third project created a **Credit Card Fraud Detection** system using imbalanced financial data. Methods like SMOTE and class-weight tuning were applied, and models including Logistic Regression, Random Forest, and XGBoost were evaluated for effectiveness.

The final project predicted the number of orders for an e-commerce platform. Historical sales data was analyzed for seasonal and promotional patterns, and regression models like XGBoost and Random Forest were used for demand forecasting, supporting inventory and logistics planning.

This internship enhanced my skills in data preparation, model selection, evaluation, and solving industry-relevant problems.

ARTIFICAL INTELEGENCE INTERNSHIP FROM SPARKIIT

1.INTRODUCTION

This portfolio highlights the practical use of Artificial Intelligence across four impactful projects, demonstrating strong proficiency in machine learning and deep learning for solving real-world challenges. Each project applies the data science lifecycle—from data preparation to model evaluation—to deliver efficient, data-driven solutions across diverse domains.

The **AI Chatbot using NLP** project automates digital customer support through intelligent query understanding and intent classification. Using techniques like TF-IDF, Bag-of-Words, and pre-trained language models, it improves customer engagement and reduces human workload by offering instant, accurate responses.

The **Credit Card Fraud Detection Syste**m focuses on financial security by identifying fraudulent transactions within highly imbalanced data. By applying SMOTE, under-sampling, and advanced algorithms such as Random Forest and XGBoost, the model achieves high recall rates, enabling reliable fraud prevention in real-world financial operations.

The **Air Quality Index (AQI) Analysis & Forecasting** – Indian Cities project leverages time series models like ARIMA and Facebook Prophet to predict air pollution levels. Based on multi-year pollutant data, it delivers accurate AQI forecasts and identifies seasonal trends, helping policymakers design effective environmental interventions.

The **Number of Orders Prediction** project supports e-commerce businesses in optimizing inventory and logistics through accurate demand forecasting. Using extensive feature engineering and regression models such as Linear Regression and XGBoost Regressor, it provides reliable order volume predictions to enhance supply chain efficiency and resource planning

2.OBJECTIVE

This internship involved practical implementation of four AI and Machine Learning projects, addressing challenges in finance, customer service,business planning, and environmental health.

Air Quality Index (AQI) Forecasting: Analyzed multi-year pollution datasets (PM2.5,) from major Indian cities. Applied time-series forecasting models like ARIMA, LSTM, and Prophet to predict future AQI. This helped in understanding pollution trends and supporting public health decision-making.

Number of Orders Prediction: Examined historical sales and order data to uncover demand patterns influenced by seasonality and promotions. Developed regression models with XGBoost to accurately forecast future order volumes. This improved inventory management and reduced excess stock or shortages.

Credit Card Fraud Detection: Worked with highly imbalanced finance transaction data. Used SMOTE for generating synthetic samples of minority (fraud) class and tuned class weights. Evaluated models including XGBoost, Random Forest, and Logistic Regression for their ability to detect fraud with high precision and recall, enhancing transaction security.

AI Chatbot using NLP: Built a conversational agent capable of understanding and responding to user queries automatically. Implemented text preprocessing, intent classification, and response generation using transformer-based deep learning models like BERT. The chatbot improved customer engagement and reduced manual support requirements.

3.Work Progress

• **Air Quality Index (AQI) Analysis and Forecasting** – Indian Cities in this project I have performed analysis historical air quality data across major Indian cities and build a forecasting model (such as ARIMA or Prophet) to

predict future pollution levels (e.g., PM2.5, PM10). The purpose is to identify seasonal trends, understand pollution patterns, and provide meaningful insights that can assist policymakers in decision-making.Central Pollution Control Board (CPCB) and publicly available environmental data portals.The dataset contains AQI records from 2015-01-01 to 2024-12-31, allowing for long-term trend analysis and forecasting.

• **Credit Card Fraud Detection System** in this project i have developed a high-performance machine learning model that accurately identifies fraudulent credit card transactions. The focus is on minimizeing false negatives, while effectively handling the severe class imbalance problem using oversampling techniques such as SMOTE.The Data set taken from Kaggle Credit Card Fraud Detection Dataset.

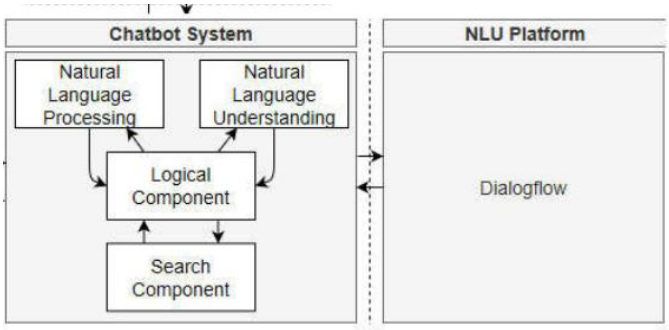
• **Number of Orders Prediction** in this project, I developed a machine learning model to predict daily or weekly customer order volumes using historical sales data. The model was trained using the Brazilian E-Commerce Public Dataset by Olist (from Kaggle), which contains transaction records from 2017 and 2018.

• **AI Chatbot using Natural Language Processing (NLP)** in this project i have designed design and develop an AI-powered chatbot capable of understanding and responding to human queries using NLP techniques.

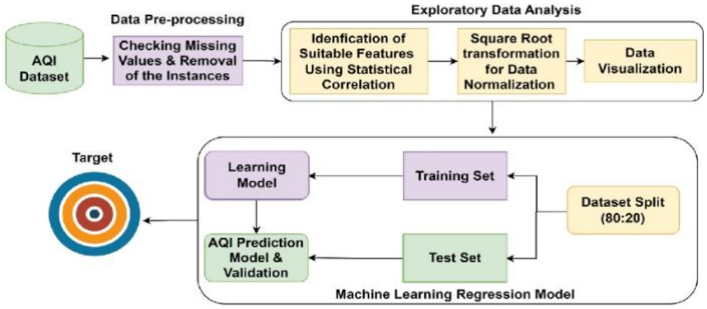
4. PARAMETERS USED

• The **AI Chatbot** project involved preparing and cleaning text data so that the system could understand user queries accurately. Techniques such as Tokenization (splitting text into words) and Lemmatization (converting words to their base form) were used during preprocessing.

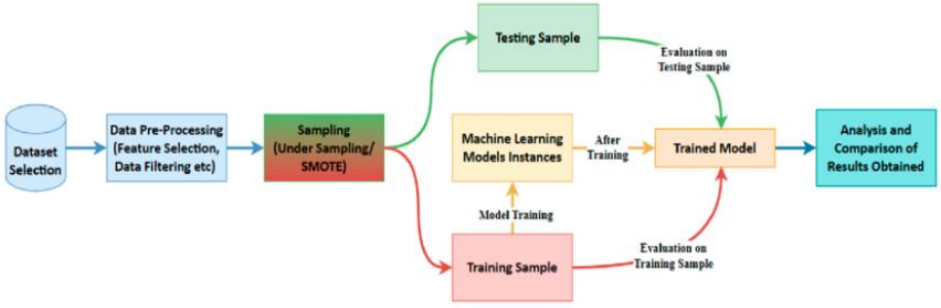
• A pre-trained language model like BERT or GPT was then applied to improve the chatbot's ability to interpret user input and generate meaningful responses.



• **Air Quality Index** Forecasting project, pollution measurements such as PM2.5 were used as the main input data. Time Series Forecasting methods, including ARIMA and Prophet, were applied to analyze patterns over days, seasons, and years. This helped in predicting future air quality values.



• **Credit Card Fraud Detection** project primarily focused on the issue of imbalance between normal and fraudulent transactions. Since fraud cases were very rare, a balancing method called SMOTE was used to ensure the model learned to identify fraud effectively.



• **Number of Orders Prediction** project, included using previous day order counts and weekly averages to help the model understand demand patterns. The XGBoost Regressor model was then used for prediction, allowing accurate estimation of future order numbers.

5. TOOLS AND TECHNOLOGIES

- •NLP(Natural Language Processing)
- •Pandas
- •NumPy
- •Transformers
- •Matplotlib
- •ARIMA
- •Handling Class Imbalance-SMOTE
- •Logistic Regression, Random Forest, XGBoost
- •Linear Regression, Random Forest Regressor, XGBoost Regressor

6. SKILLS GAINED

This internship provided practical, end-to-end experience in the data science workflow, including data preprocessing, model development, evaluation, and interpretation.

I worked with machine learning algorithms such as Random Forest and XGBoost for classification and regression tasks, handled challenges like class imbalance using SMOTE. I also gained hands-on exposure to Natural Language Processing for building intelligent chatbot systems.

Through these projects, I strengthened my proficiency in Python (Pandas, Scikit-learn) and enhanced my analytical and problem-solving skills to convert data insights into meaningful business decisions.

My practical experience also included working with Time Series Analysis and NLP. For forecasting tasks, I applied models like ARIMA and Prophet to predict trends such as air quality and daily order volumes, supported by feature engineering techniques like lag features and moving averages. In NLP, I used pre-trained models like BERT/GPT to achieve accurate intent classification in the chatbot system.

Throughout these projects, I solidified an advanced technical skillset anchored in the **Python ecosystem**

7. RESULTS

• Air quality Index

```
dataframes = {
    'df_stations': df_stations,
    'df_station_day': df_station_day,
    'df_city_hour': df_city_hour,
    'df_station_hour': df_station_hour,
    'df_city_day': df_city_day
}

for name, df in dataframes.items():
    print(f"--- Cleaning and exploring {name} ---")

    # Check for missing values
    print("Missing values per column:")
    print(df.isnull().sum())

    # Check for duplicate rows
    print("\nNumber of duplicate rows:")
    print(df.duplicated().sum())

    # Inspect data types
    print("\nData types:")
    df.info()

    # Convert 'Datetime' column to datetime objects if it exists
    if 'Datetime' in df.columns:
        df[
            'Datetime'
        ] = pd.to_datetime(df['Datetime'])
        print("\n'Datetime' column converted to datetime objects.")

    print("\n" + "="*40 + "\n")
```

Figure 1. 1

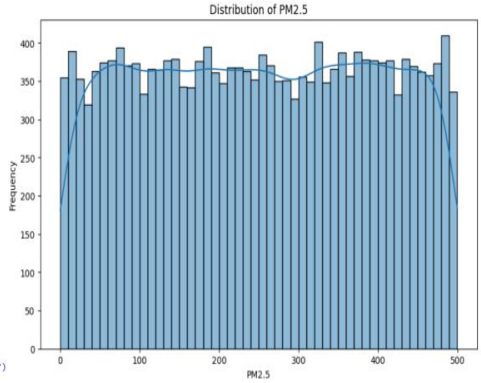


Figure 1.2

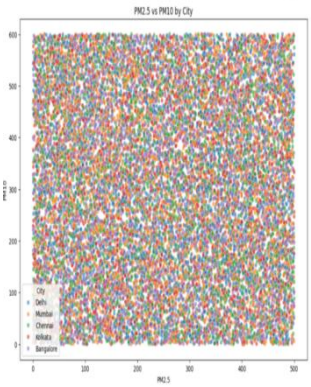


Figure1.3

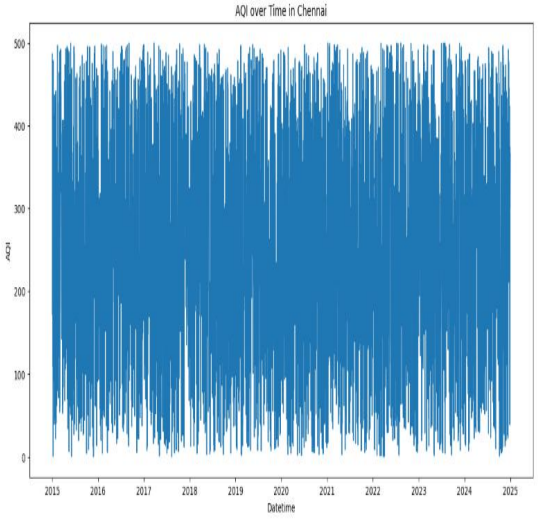


Figure 1.4

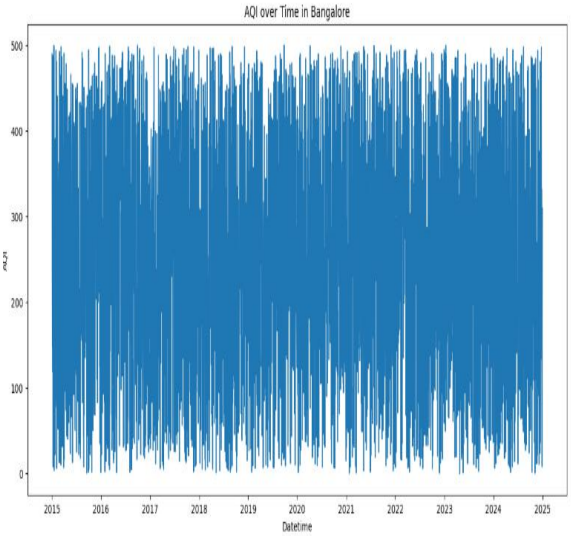


Figure 1. 5

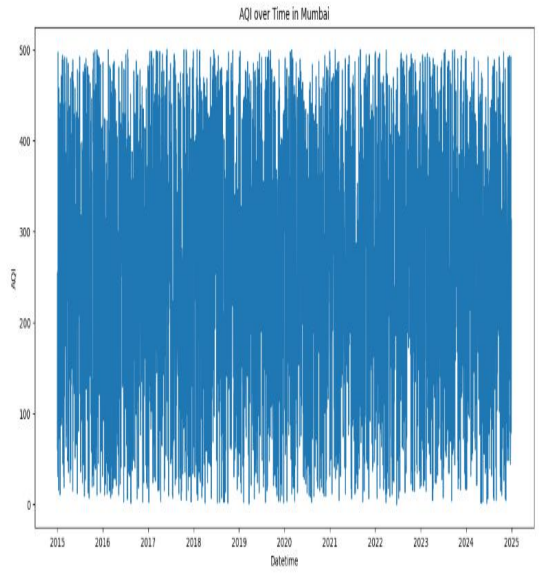


Figure 1.6

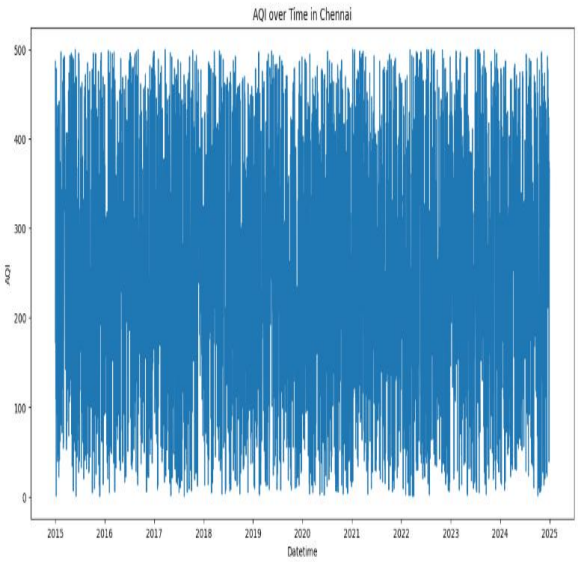


Figure 1. 7

• Credit Card Fraud Detection

```
df = pd.read_csv("creditcard.csv")
print("Shape:", df.shape)
print(df.head())
```

```
Shape: (107046, 31)
   Time        V1        V2        V3        V4        V5        V6        V7  \
0     0 -1.359807 -0.072781  2.536347  1.378155 -0.338321  0.462388  0.239599
1     0  1.191857  0.266151  0.166480  0.448154  0.060018 -0.082361 -0.078803
2     1 -1.358354 -1.340163  1.773209  0.379780 -0.503198  1.800499  0.791461
3     1 -0.966272 -0.185226  1.792993 -0.863291 -0.010309  1.247203  0.237609
4     2 -1.158233  0.877737  1.548718  0.403034 -0.407193  0.095921  0.592941

         V8        V9  ...       V21       V22       V23       V24       V25  \
0  0.098698  0.363787  ... -0.018307  0.277838 -0.110474  0.066928  0.128539
1  0.085102 -0.255425  ... -0.225775 -0.638672  0.101288 -0.339846  0.167170
2  0.247676 -1.514654  ...  0.247998  0.771679  0.909412 -0.689281 -0.327642
3  0.377436 -1.387024  ... -0.108300  0.005274 -0.190321 -1.175575  0.647376
4 -0.270533  0.817739  ... -0.009431  0.798278 -0.137458  0.141267 -0.206010

        V26       V27       V28  Amount  Class
0 -0.189115  0.133558 -0.021053  149.62    0.0
1  0.125895 -0.008983  0.014724    2.69    0.0
2 -0.139097 -0.055353 -0.059752  378.66    0.0
3 -0.221929  0.062723  0.061458  123.50    0.0
4  0.502292  0.219422  0.215153   69.99    0.0

[5 rows x 31 columns]
```

Figure 2. 1

```
# Compare results
results_df = pd.DataFrame(results).T
print("\nFinal Results Comparison:\n")
print(results_df)

results_df.plot(kind="bar", figsize=(8,5))
plt.title("Model Comparison (SMOTE)")
plt.ylabel("Score")
plt.show()
```

Figure 2. 2

```
Final Results Comparison:

                     ROC-AUC     PR-AUC
Logistic Regression  0.983754   0.741838
Random Forest        0.983768   0.918902
XGBoost              0.980912   0.933124
LightGBM             0.975278   0.918103
```

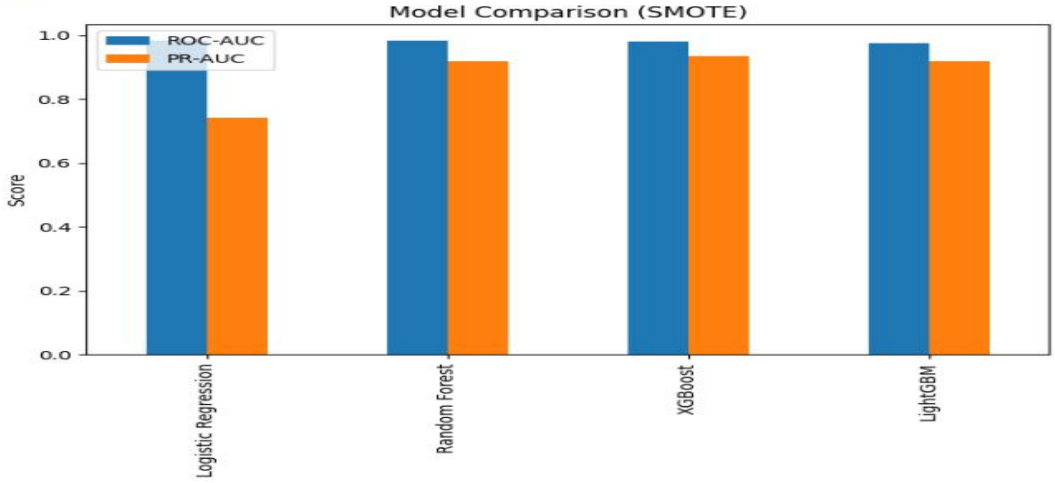


Figure 2. 3

• Number Of Order Prediction

```
import matplotlib.pyplot as plt
import seaborn as sns

# Get the count of customers per state
state_counts = df['customer_state'].value_counts()

# Create a bar plot
plt.figure(figsize=(12, 6))
sns.barplot(x=state_counts.index, y=state_counts.values, palette='viridis')
plt.title('Number of Customers per State')
plt.xlabel('State')
plt.ylabel('Number of Customers')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```

Figure 3. 1

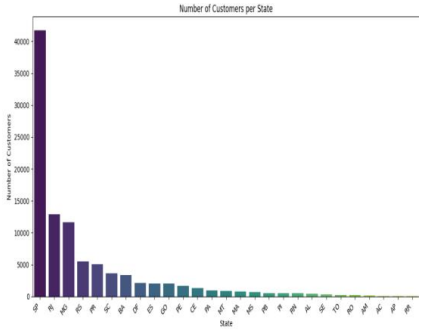


Figure3.2

3. 1

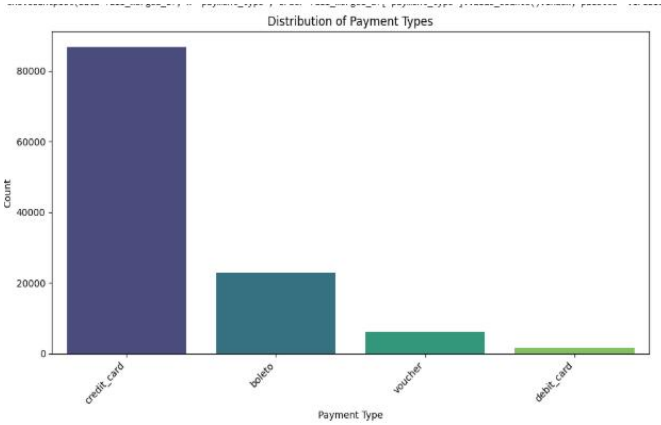


Figure 3. 3

```
plt.figure(figsize=(10, 6))
sns.histplot(full_merged_df['payment_value'], bins=50, kde=True)
plt.title('Distribution of Payment Values')
plt.xlabel('Payment Value')
plt.ylabel('Frequency')
plt.show()
```

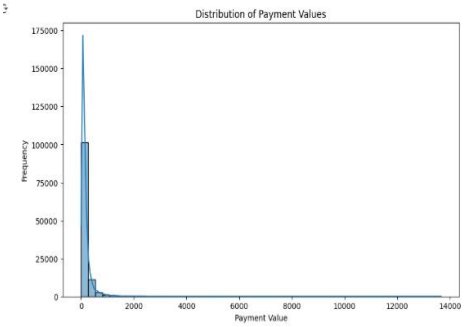


Figure 3. 4

	customer_unique_id	product_category_name_english	purchase_count
0	0000366f3b9a7992bf8c76cfdf3221e2	bed_bath_table	1
1	0000b849f77a49e4a4ce2b2a4ca5be3f	health_beauty	1
2	0000f46a3911fa3c0805444483337064	stationery	1
3	0000f6ccb0745a6a4b88665a16c9f078	telephony	1
4	0004aac84e0df4da2b147fca70cf8255	telephony	1

Figure 3. 5

- **AI chat Box**

```
print("Chatbot is running! Type 'quit' to exit.")

while True:
    user_input = input("You: ")
    if user_input.lower() == 'quit':
        print("Chatbot: Goodbye!")
        break
    response = chatbot_response(user_input)
    print(f"Chatbot: {response}")
```

```
Chatbot is running! Type 'quit' to exit.
You:  hi
Chatbot: Sorry, I didn't understand that. Can you rephrase?
You:  hi
Chatbot: Sorry, I didn't understand that. Can you rephrase?
You:  hey
Chatbot: How can I assist you?
You:  exit
Chatbot: Sorry, I didn't understand that. Can you rephrase?
You:  quit
Chatbot: Goodbye!
```

Figure 4. 1

8. CONCLUSION

The internship provided a transformative learning experience, enabling me to apply Artificial Intelligence (AI) and Machine Learning (ML) techniques to real-world challenges across finance, environmental science, and e-

commerce. Through the four-project portfolio—**Credit Card Fraud Detection**, **Air Quality Index Forecasting, AI Chatbot Development, and Orders Prediction**—I gained hands-on expertise in Python-based data preprocessing, exploratory data analysis, and the implementation of advanced ML algorithms. I also developed structured problem-solving skills in handling issues such as class imbalance using SMOTE and performing Time Series Forecasting using ARIMA and Prophet.

The solutions developed in this internship demonstrated strong predictive and automation capabilities. The models helped reduce fraud risk, improved demand forecasting for business planning, automated customer support using NLP, and supported AQI prediction for environmental insights. Overall, this experience strengthened my technical skills and analytical thinking, preparing me for future roles in Data Science and Machine Learning.

9.Reference

[1] GitHub- https://github.com/stalin112/Artificial-Intellegence
[2] Sparkiit- https://www.sparkiit.net/