National College of Ireland

Configuration Manual

MSc Research Project
Data Analytics

Vaishnavi Gaikwad
Student ID: x23109203

School of Computing
National College of Ireland

Supervisor: Abubakr Siddig

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Vaishnavi Gaikwad

Student ID: x23109203

Programme: M.Sc. Data Analytics **Year:** 2023-2024.

Module: Research Project

Lecturer: Abubakr Siddig

Submission Due Date: 12/08/2024

Project Title: Sentiment Analysis and Growth Forecast of Cashless Payment Applications in India: Pre-COVID, During-COVID, and Post-COVID Perspectives

Word Count: **6000** **Page Count: 15**

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Vaishnavi Gaikwad

Date: 12/08/2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Sentiment Analysis and Growth Forecast of Cashless Payment Applications in India: Pre-COVID, During-COVID, and Post-COVID Perspectives

Vaishnavi Gaikwad
Student ID:x23109203

1. Introduction

This research consists of an implementation procedure that is used in this study. This configuration manual is a part of my research project which describes the steps of hardware, software and technology implementation. This configuration manual aims to guide the practitioners and researchers reproduce things for future research.

2. System Specifications

Hardware Requirements:

Processor AMD Ryzen 5 5500U with Radeon Graphics 2.10 GHz
Installed RAM 16.0 GB (15.3 GB usable)
System type 64-bit operating system, x64-based processor
Edition Windows 11 Home Single Language
Version 23H2

Software Requirements:

A Jupyter IDE for implementation of machine learning and deep learning models and for analysis purpose using Python.

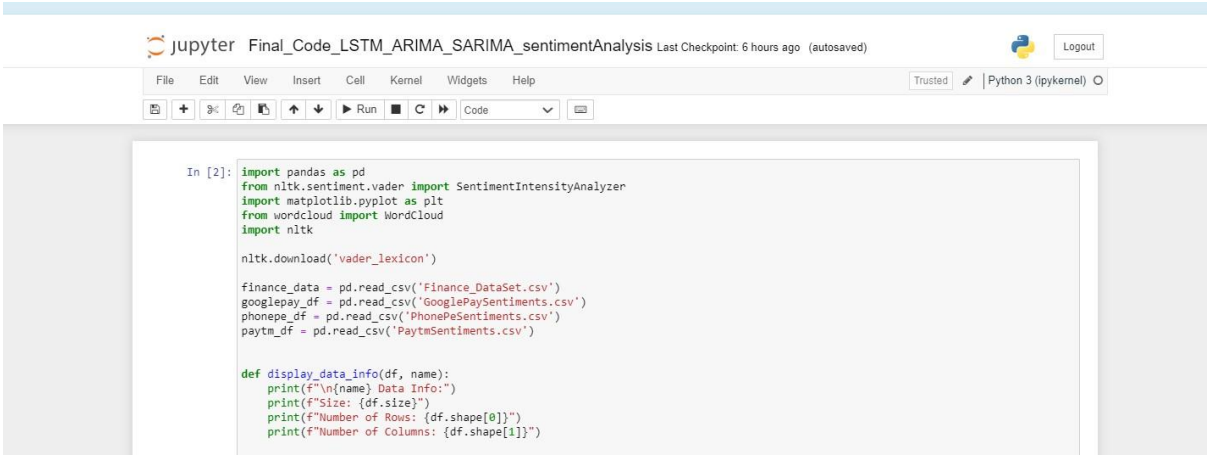

Fig1: Jupyter IDE

MicrosoftExcel 2021:microsoft excel is used to save the extracted data into csv for further analysis.

3. Sentiment Analysis

For sentiment analysis the data is extracted from the apps from Apple App store, to perform the sentiment analysis on the reviews of cashless payment apps like Google Pay phone pay and Paytm the app scraping is performed.

The python library app_store_scraper is used to scrape reviews from apps.

Below is the apps link,

Phone pe: https://apps.apple.com/in/app/phonepe-secure-payments-app/id1170055821

Paytm: https://apps.apple.com/in/app/paytm-secure-upi-payments/id473941634

Google Pay: https://apps.apple.com/in/app/google-pay-save-pay-manage/id1193357041

```
In [2]: !pip install app_store_scraper

Requirement already satisfied: app_store_scraper in c:\users\vaish\anaconda3\lib\site-packages (0.3.5)
Requirement already satisfied: requests==2.23.0 in c:\users\vaish\anaconda3\lib\site-packages (from app_store_scraper) (2.23.0)
Requirement already satisfied: chardet<4,>=3.0.2 in c:\users\vaish\anaconda3\lib\site-packages (from requests==2.23.0->app_stor
e_scraper) (3.0.4)
Requirement already satisfied: idna<3,>=2.5 in c:\users\vaish\anaconda3\lib\site-packages (from requests==2.23.0->app_store_scr
aper) (2.10)
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in c:\users\vaish\anaconda3\lib\site-packages (from requ
ests==2.23.0->app_store_scraper) (1.25.11)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\vaish\anaconda3\lib\site-packages (from requests==2.23.0->app_sto
re_scraper) (2023.11.17)
```

Fig2: App scraping library configurations

Above are the configurations of app scraping,once the app scraping is done fore all 3 apps the reviews are saved into csv files named as : GooglePaySentiments.csv, PhonePeSentiments.csv, 'PaytmSentiments.csv' The sentiment analysis is performed using VADER lexicon library from nltk library that is used to analyse sentiments on text based data that are expressed onany social media platform.

```
In [2]: import pandas as pd
        from nltk.sentiment.vader import SentimentIntensityAnalyzer
        import matplotlib.pyplot as plt
        from wordcloud import WordCloud
        import nltk
```

```
[nltk_data] Downloading package vader_lexicon to
[nltk_data]     C:\Users\vaish\AppData\Roaming\nltk_data...
[nltk_data]   Package vader_lexicon is already up-to-date!
```

Fig3: libraries that is used to analyse sentiments

The Data Preprocessing and Data manipulation technique is performed using the panda's library, in data preprocessing the unnecessary columns are deleted, the convergence of date columns is performed.
The word cloud is used to visual the reviews of app using python word cloud library.
Using Matplotlib the analysed sentiments are plotted for all three apps according to period.
After sentiment analysis is performed to determine the sentiment of people regarding cashless payment app reviews. The Finance data is used to forecast the growth of cashless payment using UPI apps.
The Finance data is extracted from NPCI website that is Indian government national payment website of UPI. dataset (https://www.npci.org.in/what-we-do/upi/upi-ecosystem-statistics#innerTabTwoJun24).
The data is extracted for GOOGLE PAY, PHONE PE, PAYTM.
The remaining data is synthetically generated for finance data using python and saved in csv file for remaining year and the data is merged manually into a single csv file.FinanceDataset.csv

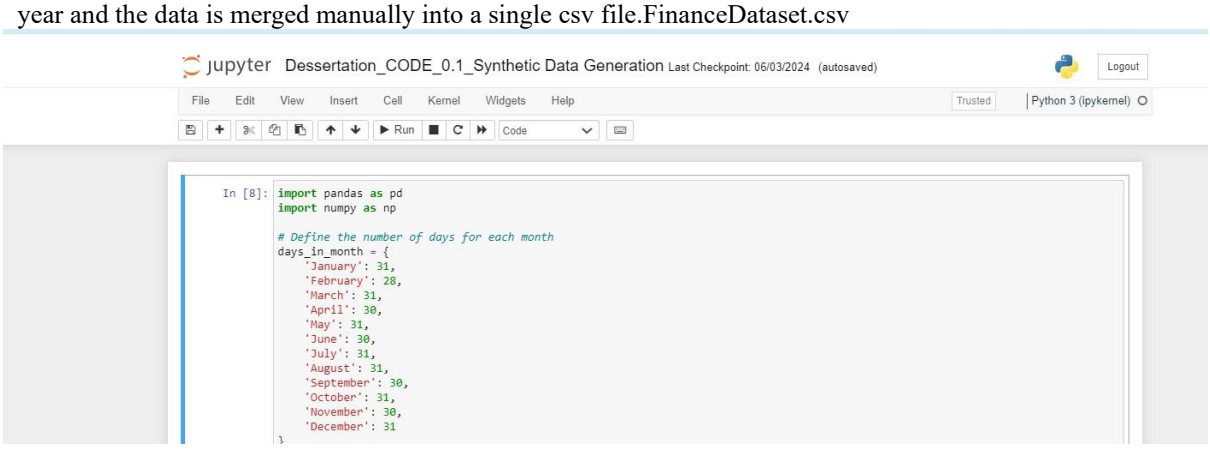

Fig4.Synthetic Data Generation

To understand the correlation between finance data and sentiment reviews the data is merged in single csv and correlation is tested also the heatmap is generated to understand the correlation between finance data and sentiments. This is performed using pandas, seaborn and matplotlib libraries.

```
In [10]: import pandas as pd
         import seaborn as sns
         import matplotlib.pyplot as plt

         # Load the merged data
         finance_data = pd.read_csv('Merged_Finance_Data_with_Sentiment.csv')

         # Convert 'Date' column to datetime format
         finance_data['Date'] = pd.to_datetime(finance_data['Date'])

         # Define date ranges for categorization
         periods = {
             'PRE COVID': ((2017, 1, 1), (2019, 12, 31)),
             'DURING COVID': ((2020, 1, 1), (2022, 12, 31)),
```

Fig5.Correlation between finance data and sentiments data

4. Cashless Payment evolution via apps:

Forecasting the growth of cashless payment using UPI apps like Google Pay phone pe and Paytm. Based on finance data the machine learning and deep learning models are you used to forecast the actual values predicted values for the finance data along with forecasting the future of cashless payment app evolution. we have also predicted the app that will be most used in future by using machine learning and deep learning models. 1]LSTM model let us see configurations for LSTM model: libraries used:

```
In [11]: import pandas as pd
         import numpy as np
         import tensorflow as tf
         import random
         from sklearn.preprocessing import MinMaxScaler
         from keras.models import Sequential
         from keras.layers import LSTM, Dense, Dropout
         import matplotlib.pyplot as plt
```

```
C:\Users\vaish\AppData\Local\Temp\ipykernel_4024\657942017.py:31: FutureWarning: DataFrame.fillna with 'method' is deprecate
d and will raise in a future version. Use obj.ffill() or obj.bfill() instead.
  bank_df.fillna(method='ffill', inplace=True)
C:\Users\vaish\AppData\Local\Temp\ipykernel_4024\657942017.py:32: FutureWarning: DataFrame.fillna with 'method' is deprecate
d and will raise in a future version. Use obj.ffill() or obj.bfill() instead.
  bank_df.fillna(method='bfill', inplace=True)
C:\Users\vaish\AppData\Local\Temp\ipykernel_4024\657942017.py:31: FutureWarning: DataFrame.fillna with 'method' is deprecate
d and will raise in a future version. Use obj.ffill() or obj.bfill() instead.
  bank_df.fillna(method='ffill', inplace=True)
C:\Users\vaish\AppData\Local\Temp\ipykernel_4024\657942017.py:32: FutureWarning: DataFrame.fillna with 'method' is deprecate
d and will raise in a future version. Use obj.ffill() or obj.bfill() instead.
  bank_df.fillna(method='bfill', inplace=True)
C:\Users\vaish\AppData\Local\Temp\ipykernel_4024\657942017.py:31: FutureWarning: DataFrame.fillna with 'method' is deprecate
d and will raise in a future version. Use obj.ffill() or obj.bfill() instead.
  bank_df.fillna(method='ffill', inplace=True)
C:\Users\vaish\AppData\Local\Temp\ipykernel_4024\657942017.py:32: FutureWarning: DataFrame.fillna with 'method' is deprecate
d and will raise in a future version. Use obj.ffill() or obj.bfill() instead.
  bank_df.fillna(method='bfill', inplace=True)
C:\Users\vaish\anaconda3\Lib\site-packages\keras\src\layers\rnn\rnn.py:204: UserWarning: Do not pass an `input shape`/`input
```

Fig6. LSTM model configurations

for data preprocessing and data manipulation the pandas NumPy and for data scaling process the min Max scaler is used.

From Keras library the LSTM model is extracted and used to fit in for the finance data. The forecasting for cashless payment app evolution is performed using matplotlib. b) ARIMA Model:

Here we have forecasted the evolution of cashless payments via apps by using the ARIMA model.

let us see the libraries used to define the ARIMA model for finance data:

```
n [13]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        from statsmodels.tsa.arima.model import ARIMA
```

```
C:\Users\vaish\AppData\Local\Temp\ipykernel_4024\1660307804.py:28: FutureWarning: DataFrame.fillna with 'method' is deprecated
and will raise in a future version. Use obj.ffill() or obj.bfill() instead.
  bank_df.fillna(method='ffill', inplace=True)
C:\Users\vaish\AppData\Local\Temp\ipykernel_4024\1660307804.py:29: FutureWarning: DataFrame.fillna with 'method' is deprecated
and will raise in a future version. Use obj.ffill() or obj.bfill() instead.
  bank_df.fillna(method='bfill', inplace=True)
C:\Users\vaish\anaconda3\Lib\site-packages\statsmodels\tsa\base\tsa_model.py:473: ValueWarning: A date index has been provided,
but it has no associated frequency information and so will be ignored when e.g. forecasting.
  self._init_dates(dates, freq)
C:\Users\vaish\anaconda3\Lib\site-packages\statsmodels\tsa\base\tsa_model.py:473: ValueWarning: A date index has been provided,
but it has no associated frequency information and so will be ignored when e.g. forecasting.
  self._init_dates(dates, freq)
C:\Users\vaish\anaconda3\Lib\site-packages\statsmodels\tsa\base\tsa_model.py:473: ValueWarning: A date index has been provided,
but it has no associated frequency information and so will be ignored when e.g. forecasting.
  self._init_dates(dates, freq)
C:\Users\vaish\anaconda3\Lib\site-packages\statsmodels\tsa\base\tsa_model.py:836: ValueWarning: No supported index is availabl
e. Prediction results will be given with an integer index beginning at `start`.
  return get_prediction_index(
C:\Users\vaish\anaconda3\Lib\site-packages\statsmodels\tsa\base\tsa_model.py:836: FutureWarning: No supported index is availabl
e. In the next version, calling this method in a model without a supported index will result in an exception.
  return get_prediction_index(
C:\Users\vaish\AppData\Local\Temp\ipykernel_4024\1660307804.py:28: FutureWarning: DataFrame.fillna with 'method' is deprecated
and will raise in a future version. Use obj.ffill() or obj.bfill() instead.
  bank_df.fillna(method='ffill', inplace=True)
C:\Users\vaish\AppData\Local\Temp\ipykernel_4024\1660307804.py:29: FutureWarning: DataFrame.fillna with 'method' is deprecated
and will raise in a future version. Use obj.ffill() or obj.bfill() instead.
  bank_df.fillna(method='bfill', inplace=True)
C:\Users\vaish\anaconda3\Lib\site-packages\statsmodels\tsa\base\tsa_model.py:473: ValueWarning: A date index has been provided,
but it has no associated frequency information and so will be ignored when e.g. forecasting.
  self._init_dates(dates, freq)
C:\Users\vaish\anaconda3\Lib\site-packages\statsmodels\tsa\base\tsa_model.py:473: ValueWarning: A date index has been provided,
but it has no associated frequency information and so will be ignored when e.g. forecasting.
  self._init_dates(dates, freq)
C:\Users\vaish\anaconda3\Lib\site-packages\statsmodels\tsa\base\tsa_model.py:473: ValueWarning: A date index has been provided,
```

Fig7.ARIMA model Configurations

for data preprocessing and data manipulation technique the pandas and NumPy library is used, Arima model is extracted from stats model libraries of Python. And to visualise the forecasting of cashless payment app evolution is performed by using matplotlib plots. c)SARIMA model:

The SARIMA MODEL is used to forecast the cashless payment app evolution by using the stats model SARIMAX library by analysing the seasonal trends in finance data. Using pandas and NumPy for data preprocessing and data manipulation is performed. The visualisation of evolution of cashless payment app evolution is forecasted using python matplotlib library.

```
In [15]: import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
         from statsmodels.tsa.statespace.sarimax import SARIMAX
```

```
C:\Users\vaish\AppData\Local\Temp\ipykernel_4024\635663078.py:28: FutureWarning: DataFrame.fillna with 'method' is deprecated a
nd will raise in a future version. Use obj.ffill() or obj.bfill() instead.
  bank_df.fillna(method='ffill', inplace=True)
C:\Users\vaish\AppData\Local\Temp\ipykernel_4024\635663078.py:29: FutureWarning: DataFrame.fillna with 'method' is deprecated a
nd will raise in a future version. Use obj.ffill() or obj.bfill() instead.
  bank_df.fillna(method='bfill', inplace=True)
C:\Users\vaish\anaconda3\Lib\site-packages\statsmodels\tsa\base\tsa_model.py:473: ValueWarning: A date index has been provided,
but it has no associated frequency information and so will be ignored when e.g. forecasting.
  self._init_dates(dates, freq)
C:\Users\vaish\anaconda3\Lib\site-packages\statsmodels\tsa\base\tsa_model.py:473: ValueWarning: A date index has been provided,
but it has no associated frequency information and so will be ignored when e.g. forecasting.
  self._init_dates(dates, freq)
C:\Users\vaish\anaconda3\Lib\site-packages\statsmodels\tsa\base\tsa_model.py:836: ValueWarning: No supported index is availabl
e. Prediction results will be given with an integer index beginning at `start`.
  return get_prediction_index(
C:\Users\vaish\anaconda3\Lib\site-packages\statsmodels\tsa\base\tsa_model.py:836: FutureWarning: No supported index is availabl
e. In the next version, calling this method in a model without a supported index will result in an exception.
  return get_prediction_index(
C:\Users\vaish\AppData\Local\Temp\ipykernel_4024\635663078.py:28: FutureWarning: DataFrame.fillna with 'method' is deprecated a
nd will raise in a future version. Use obj.ffill() or obj.bfill() instead.
  bank_df.fillna(method='ffill', inplace=True)
C:\Users\vaish\AppData\Local\Temp\ipykernel_4024\635663078.py:29: FutureWarning: DataFrame.fillna with 'method' is deprecated a
nd will raise in a future version. Use obj.ffill() or obj.bfill() instead.
```

Fig 8.SARIMA model configurations

5. CONCLUSION

This Configuration manual represents the study of cashless payment evolution in India and analysing sentiments on the public reviews. This document represents the steps that are used to perform the sentiment analysis the libraries used and install, and the key technologies used for forecasting cashless payment app evolution in India. This manual contains all the required information to replicate this study.