|  |
| --- |
| Solent University |
| Faculty of Business, Law, and Digital Technologies |
|  |

|  |
| --- |
| Supervisor: Darren Cunningham  Date of Submission: May 2022 |

A picture containing text

Description automatically generatedIcon

Description automatically generated with medium confidence

Software Engineering: 2021/2022

COM616: Dissertation

‘Intelligent Income and Expenditure System’

Rahul Shrestha

Q13671294

Table of Contents

[1. Introduction 1](#_Toc101360957)

[2. Literature Review 1](#_Toc101360958)

[3. Methodology 2](#_Toc101360959)

[3.1 Regression 2](#_Toc101360960)

[3.1.1 Regression: Exploratory Data Analysis 2](#_Toc101360961)

[3.1.2 Regression: Data Preparation & Cleaning 2](#_Toc101360962)

[3.1.3 Regression Model: Long Short-Term Memory (LSTM) 2](#_Toc101360963)

[3.1.4 Regression Model: ARIMA 3](#_Toc101360964)

[3.1.5 Regression Model: Facebook Prpophet 3](#_Toc101360965)

[3. 2 Classification 3](#_Toc101360966)

[3.2.1 Classification: Exploratory Data Analysis 10](#_Toc101360967)

[3.2.2 Classification: Data Preparation & Cleaning 10](#_Toc101360968)

[3.2.3 Classification Model: Logistic Regression 10](#_Toc101360969)

[3.2.4 Classification Model: Decision Tree 10](#_Toc101360970)

[3.2.5 Classification Model: Random Forest 10](#_Toc101360971)

[Results 11](#_Toc101360972)

[Limitations 11](#_Toc101360973)

[References 11](#_Toc101360974)

[Appendix 11](#_Toc101360975)

## 1. Introduction

Debt is a term that no one wants to hear when it comes to their personal finance. Debt can be defined as owing something, usually money, to someone (Chen, 2022). Research from Close Brothers (2019) has found that almost 94% of UK employees are suffering from money worries and almost 77% of employees have said that it has affected them at work. Additionally, The Money Charity has stated that as of November 2021, the average total household debt in the UK is £63,122.

There are many factors as to why someone can get into debt; Zhen (2022) has stated one of the most common causes was due to poor money management. This can take form in several different ways such as impulsive buying, using overdraft and simply spending more than you are earning. Whistl (2017) have found that 91% of the nation have admitted to making impulsive purchases every month and on top of that, Hall (2018) stated in a news article that an average UK adult will spend over £144,000 on impulsive buying during their lifetime. This could be due to the advancement of technology over the years which has allowed the rapid growth of e-commerce and in turn have amplified impulsive buying behaviours. Additionally, due to COVID-19 pandemic, the UK’s top retailers have stated that their online traffic has increased by 52% (Jobling, 2021).

Furthermore, the rise of contactless payments and mobile wallets have also been seen to contribute overspending. A study conducted by Xu et al. (2019) found that using mobile wallets can lead to people spending more money, more frequently. Other common causes of debt include spending future money and having no savings, such as an emergency fund.

One of the best ways to manage your expenses and control your spending is to utilise a budgeting app. Using one can help you keep track of your spending, bills, and generally allows you to become more aware of your finances (Lake & Foreman 2021).

One of the approaches taken to save money is to store it away in a savings account, however, Barclays (2021) have stated in terms of accumulating more wealth for the future, it is better to invest into the stock markets rather than leave it in a savings account. Likewise, Money Helper (2022) have stated that leaving your money in a savings account is not the best option as the interest rate in the saving account is nearly always lower than the rate of inflation. Furthermore, Inman (2022) has reported that inflation in the UK has risen to its highest levels in 30 years, currently around 5.4%. The rise of inflation refers to an increase in prices and the decrease of purchasing power. This means that consumers can purchase less goods and services compared to before (Davies 2022). Moreover, Clark (2020) has reported that the average saving account interest rates have fallen to their lowest levels on record at 0.64% in 2020, meaning that the return on your money will be nonexistent, especially when factoring in the increase of inflation rates. In contrast, the S&P 500, which is defined as a stock market index that tracks the US 500 large-cap companies (Amadeo, 2022), has reported an average annual return of around 10.5% since its inception in 1957, beating the inflation rate and any other savings account (Maverick, 2022).

On the other hand, investing is a lot riskier than storing your money in a savings account, therefore it is not advised for short-term goals, such as anything less than 5 years (Barclays, 2021). Stock markets are volatile, meaning the values of stocks can fluctuate and even drop in value drastically. For this reason, it is advised to aim to invest for at least 5 years as a longer time frame will allow your investments to recover over time (HSBC, n.d). Additionally, HSBC (n.d) have advised that before participating in saving or investing any money, it is important to have an emergency fund in case of any unexpected expense.

## 2. Literature Review

A scoping review conducted by Harper et al. (2021) found that people involved in the criminal justice system are disproportionally indebt compared to the average person. They suggested that reducing debt in this population can improve re-entry outcomes and quality of life. Furthermore, Van Beek et al. (2021) systematic review found debt to be a risk factor for criminal behaviour. Thus, utilisation of a financial app could be particularly beneficial for those who have a criminal background, manage their debt. This could result in fewer crimes.

Additionally, A review conducted Swanton and Gainsbury (2020) found that debt problems led people to take part in gambling addiction which in turn resulted in bigger mental health problems. It was also stated gambling-related debt problem increased the likelihood of psychological distress, substance use, crime, and suicidality. Plus, findings from research conducted by Franzen and Bradaric (2018) showed that there is a gap of knowledge when it comes to managing money and being financially aware, especially in college students. It was also stated that due to the poor money management skills, students had increased stress levels and were not performing well in their academics. It also led to some students dropping out of school. Additionally, they suggested that utilisation of budgeting apps could lead to student maintaining and attaining financial wellness.

In a study by Ong et al. (2019), it was founded that when those in debt where given debt relief and they had experienced significant improvements in their cognitive function and reports of less anxiety. Furthermore, French et al. (2020) founded that utilisation of a finance app significantly improved financial knowledge which translated to improved financial behaviours.

## 3. Methodology

The stock market can be explored using two methods known as technical analysis and fundamental analysis. Fundamental analysis is defined as a method to determine the real (intrinsic) value of a stock by examining economic and financial factors of the company (Segal, 2021). Investors and traders that use fundamental analysis believe that the market does not accurately estimate the value of stocks and therefore they try and find a true worth of a company (The Street, 2022). They find and invest in stocks; they believe are undervalued by the market and hope the stock’s value increases over time.

On the other hand, technical analysis is defined as using historical market data to evaluate the price trends and patterns, to predict future markets behaviour (Chen, 2021). Saravanan (2019) has stated that fundamental analysis is more theoretical and that using technical analysis is seen to be more practical as it uses more factual, concrete data. Additionally, The Street (2022) has claimed that trading decisions are best made from technical analysis using trend evaluation and pattern recognition as they believe that stocks are accurately valued, thus fundamental analysis is unnecessary.

Technical Indicators fall into the realm of technical analysis, and Chen (2021) defined it as mathematical calculations and patterns derived from historical data. There are many technical indicators available out there and they can be classed into five categories: trend, momentum, relative strength, mean reversion, and volume (Barone, 2022). Folger (2022) has advised that when developing a trading strategy, it is recommended not to use different indicators from the same category as this can result in multicollinearity but as this project is aimed towards beginners, I have chosen easy to understand and beginner-friendly indicators which goes against Folgers’ advice.

The dataset used to train and test the models will be historical stock data, in this case Apple’s (AAPL) stock data was used; scarped from Yahoo Finance using the python library ‘yfinance’. Additionally, extra columns will be added to the dataset which will consist of for the four technical indicators data and for the classification models, an extra five columns will be added which will include trading signals from each indicator and a target variable , which will be an overall recommendation for the trading signal. This project takes on a supervised learning approach; Petersson (2021) defines supervised learning as models that are trained on input data labelled to specific output. This allows the model to learn and detect underlying patterns and relationship between the input and output data so that it can accurately predict on unseen input data. The aim of this project is to utilise technical analysis to predict future stock prices and trading signals of a stock such as buy, hold, or sell based on its historical data and the technical indicators data.

The technical indicators that are used and added to the dataset are:

|  |  |
| --- | --- |
| COLUMN | EXPLANATION |
| Stochastic Oscillator | Stochastic Oscillator (SO), which was developed by George Lane in the 1950’s, is a popular technical indicator when it comes to generating oversold and overbought signals (Hayes, 2021). Anderson (2022) defines SO to describe the relationship between the stock price, relative to its high and low prices over a predetermined period (14 days being the popular period). Additionally, Anderson (2022) has stated that SO has a good history of being accurate when it comes generating buy and sell signals.  SO has two components that work together in building a trading signal, the fast line denoted as ‘%K’ and the slow line denoted as ‘%D’ (West, n.d). Both signals produce a value that ranges between 0 to 100, typically values below 20 are seen as oversold which infers a buy signal and values over 80 are seen as overbought which infers a sell signal (West, n.d).  K% is calculated by = 100 \* ((14 Day Closing Price – 14 Day Lowest Price ) – (14 Highest Price – 14 Day Lowest Price))  D% is calculated by = moving average of %K over 3 days.  (For Clasifcation Model Only)  For this project, the SO indicator will follow the traditional rules when producing a trading signal such that :  A ‘buy’ signal will be created when:   * The %K value/line is below 20 * The %D value/line is below 20   A sell signal will be created when:   * The %K value/line is above 80 * The %D value/line is above 80   Chart, histogram  Description automatically generatedHere is a graph displaying the SO indicator based on the past 6 months of the Apple (AAPL) stock: |
| Relative Strength Index (RSI) | The Relative Strength Index (RSI), which was developed by J. Welles Winder in 1970, is also a momentum indicator like the stochastic oscillator that is used by traders to identify whether the market is an overbought or oversold state. Gumparthi (2017) describes RSI to measure the speed and change of price movements over a previous trading period.  The RSI also produces a value ranging from 0 to 100 but unlike the SO, values over 70 are seen as overbought and values under 30 are seen as oversold, according to Fernando (2022).  Even though, the RSI and SO are both momentum indicators, they both have different underlying methods and theories. Ross (2021) has stated the RSI is more useful in trending markets whereas SO is more useful when the market is trading in consistent ranges.  A study conducted by Gumparthi (2017) to the test validity of RSI signals in trading strategies found that the RSI to be an effective indicator, that was able to produce an accurate buy and sell signals for both short-term and long-term investments. It was also discovered that it successfully predicted future trends in the market.  Fernando (2022) described the RSI to be calculated using the following formulas:   1. Avg Loss = Sum of Losses over the past 14 periods / 14 2. Avg Gain = Sum of Gains over the past 14 periods / 14 3. RS = Average Gain / Average Loss 4. RSI = 100 – 100 ( 1 + RS).   (For Clasifcation Model Only)  For this project, the traditional boundaries will be used to create a trading signal for the RSI; such that values under 30 will be seen as buy signals and values over 70 will be seen as sell signals.  Graphical user interface, chart  Description automatically generated |
| Moving Average Convergence Divergence (MACD) | The Moving Average Convergence Divergence (MACD) was developed by Gerald Appel in 1979 and it used as trend-following momentum indicator (Schlossberg, 2022). Silberstein (2022) defined MACD to describe the relationship between two moving averages of a stock and it is calculated by subtracting the 26-period exponential moving average (EMA) from the 12-period EMA, this is referred to as the MACD line. Additionally, there is another component referred to as the signal line, that works with the MACD line to come up with a trading signal. The signal line is calculated by finding out the 9-period EMA of the MACD. Mathematically written as:   * MACD = 12D EMA – 26DEMA * Signa = 9D EMA of MACD   Here is a diagram displaying the MACD line and the Signal line for the past 6 months of the Apple (AAPL) stock:  Chart, line chart, histogram  Description automatically generated  (For Clasifcation Model Only)  For this project, MACD indicator will produce a buy signal when the MACD line crosses **above** the signal line thus the sell signal will be created when the MACD line crosses **below** the signal line. |
| Bollinger Bands | Bollinger Bands (BB) was created by John Bollinger in the 1980’s and it has been described to offer numerous insights into price and volatility, such as monitoring breakouts, following trends and determining overbought and oversold levels (Mitchell, 2022).  BB consist of three components that work together to highlight how prices are distributed around an average value. Binance Academy (2018) described the components to be calculated using the following formulas:   * Middle Band= 20-day simple moving average (SMA) * Upper Band = Middle Band + (2 x 20-day stand deviation) * Lower Band = Middle Band – (2 x 20-day stand deviation)   Here is a diagram displaying the BB for the past 6 months of the Apple (AAPL) stock:  A picture containing diagram  Description automatically generated  (For Clasifcation Model Only)  For this project, the BB indicator will be used to determine overbought and oversold level to create buy and sell signals. Buy signals will be created when the price crosses below the lower band and alternatively, sell signals when the price cross above the upper band. |
| Recommender  (Target Variable for Classification Model) | (For Clasifcation Model Only)  The Recommender column (dependant variable) contains an overall recommendation in whether to buy, sell, or hold the stock based on the signals from the other indicators.  Upon further inspection, the function I created to derive trading signals the MACD indicators were producing inaccurate signals so therefore they have not taken in consideration when creating the overall signals, however the MACD line and the signal line will still be used when training the models.  To ensure signals were as accurate as possible I followed the following steps:  A simple if-else function, where if all three of the indicators stated the same signal, the value would be declared as that signal or if at least two out of three indicators stated the same signal, it was declared as that signal. Everything else that did not fit into the above statements were labelled as ‘Unclassed.’ This is the table outlining the above function:   |  |  |  |  | | --- | --- | --- | --- | | RSI | SO | BB | Recommender | | Buy | Buy | Buy | Buy | | Sell | Sell | Sell | Sell | | Hold | Hold | Hold | Hold | | Buy | Buy | ? | Buy | | ? | Buy | Buy | Buy | | Buy | ? | Buy | Buy | | Sell | Sell | ? | Sell | | ? | Sell | Sell | Sell | | Sell | ? | Sell | Sell | | Hold | Hold | ? | Hold | | ? | Hold | Hold | Hold | | Hold | ? | Hold | Hold | |

## 

## 3.1 Regression

Predicting future stock prices can be also classed as a time series problem in which time series forecasting methods can be applied. Tableau (2022) defined time series forecasting as making scientific predictions based on historical timed stamped data. In relation to stock prices, Christie (2020) stated that stock prices should be treated as discrete time series data as stock prices are taken sequentially in time. As mentioned above, AAPL’s historical stock data will be used as the dataset to train and test the regression models, unlike with the classification models, this dataset will contain all AAPL’s stock data available, dating back to when the company first went public on December 12, 1980.

### 3.1.1 Regression: Exploratory Data Analysis

Conducting an exploratory data analysis (EDA) is an important step in which preliminary investigation are led to provide an insight into the data and their interactions (Sonal, 2021). EDA normally consists of using graphical representations and summary of statistics.

Table containing the EDA that was conducted:

|  |  |
| --- | --- |
| Description | Result |
| 1. Gathering historical data for the AAPL stock and storing it in a data frame. Applying ‘.shape’ function returns the dimensions of the dataset. From the result, you can interpret the base dataset to contains 10427 rows and 7 columns. |  |
| 2. Applying ‘.columns’ returns the columns headers in the dataset. From the result, you can see it to contain:   * Open * Close * High * Low * Close * Volume * Dividends * Stock Splits |  |
| 3. |  |
|  |  |
|  |  |
| 3. Applying ‘.info’ function provides you with an overview of the dataset. From the result, you can see the different column headers and its data types. |  |
| 4. Applying ‘.describe’ provides with a statistical summary for the numerical columns in the dataset. |  |
| 5. Applying the ‘.tail’ function returns the last 5 rows in the data frame. From the results, you can see we have the latest stock data available. |  |
| 6. Applying the ‘.head’ function returns the first 5 rows in the data frame. From the result, you can see, we have stock data from far back as ‘1980-12-12’ in the dataset. |  |
| 7. Applying the ‘.isna’ function returns if there are any null values in the dataset. It is very important to check for null values in machine learning as most algorithm does not support missing values. From the result, you can see there are no null values present in the dataset. |  |
|  |  |
| 8. Applying ‘.duplicated’ returns if there are any duplicate records in the dataset. This is also very important to check as having duplicates in the dataset could result in inaccurate predictions from the models. Again, from the results you can see there are no duplicates present in the dataset. |  |
| 9. Heatmap displays the correlation between the different variables on scale from -1 to 1. From the results, you can see this graph isn’t useful as you can see large number of ‘1’ which shows they are positively correlated and interrelated. This could be due to the small difference in between the values. |  |
| 11. Open and Close Price |  |
|  |  |
| 12. High and Low Price |  |
| 13. Volume Traded |  |
|  |  |
| 14. Returns |  |
|  |  |
| 15. PACF |  |
| 16. SO |  |
| 17. RSI |  |
| 18.Bollinger Bands |  |
| 19. MACD |  |

### 3.1.2 Regression: Data Preparation & Cleaning

### 3.1.3 Regression Model: Long Short-Term Memory (LSTM) Univariate

### 3.1.4 Regression Model: Long Short-Term Memory (LSTM) Multivariate

### 3.1.4 Regression Model: ARIMA

### 3.1.5 Regression Model: Facebook Prophet

## 3. 2 Classification

The dataset used to train and test the classification models will be historical stock data, in this case the past 10 years of the Apple’s (AAPL) stock data was used; scarped from Yahoo Finance using the python library ‘yfinance’. Additionally, five extra columns will be added to the dataset which will consist of four technical indicators and a target variable column ‘recommender’. The aim of the classification model will be to accurately predict buy/sell/hold signals based on technical analysis and technical indicators.

### 3.2.1 Classification: Exploratory Data Analysis

### 3.2.2 Classification: Data Preparation & Cleaning

### 3.2.3 Classification Model: Logistic Regression

### 3.2.4 Classification Model: Decision Tree

### 3.2.5 Classification Model: Random Forest

## Results

## Limitations

## References

## Appendix