Classification Model

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.

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 necessary.

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 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 this project, the following will be considered as a ‘buy’ signal:   * The %K value/line is below 20 * The %D value/line is below 20   The sell signal when:   * The %K value/line is above 80 * The %D value/line is above 80   Here 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 this project, the traditional boundaries will be used such that values under 30 will be seen as buy signals and values over 70 will be seen as sell signals. |
| 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:  For this project, the buy signal will be created 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:    For this project, the BB 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 | The Recommender column (dependant variable) contains an overall recommendation in whether to buy, sell, or hold the stock.  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 signal 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 | |

## Exploratory Data Analysis

## Machine Learning Model: Logistic Regression

## Machine Learning Model: Decision Tree

## Machine Learning Model: Random Forest

## Limitations

## References