A logo for college computing

Description automatically generated

|  |  |
| --- | --- |
| Student Full Name: Jeisson Steven Preciado Wilches |  |
| Student Number: 2024326 |  |
| Module Title: Strategic Thinking |  |
| Assessment Title: CA 1 |  |
| Assessment Due Date: 18/05/25 |  |
| Date of Submission: 18/05/25 |  |

**Declaration**

By submitting this assessment, I confirm that I have read the CCT policy on academic misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source.

I declare it to be my own work and that all material from third parties has been appropriately referenced.

I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

Table of Contents

[Introduction 3](#_Toc198046177)

[Objectives 4](#_Toc198046178)

[Problem Definition 4](#_Toc198046179)

[Business Description & Methods 5](#_Toc198046180)

[Technologies & Accomplishment 5](#_Toc198046181)

[Challenges encountered 7](#_Toc198046182)

[Results & Analysis 7](#_Toc198046183)

[Data Base 8](#_Toc198046184)

[Time Series Analysis & ARIMA 10](#_Toc198046185)

[Linear Regression 14](#_Toc198046186)

[K-Means Clustering 16](#_Toc198046187)

[K-Nearest Neighbours 17](#_Toc198046188)

[Simple Exponential Smoothing 19](#_Toc198046189)

[Conclusion 21](#_Toc198046190)

[References 22](#_Toc198046191)

# Introduction

This project leverages machine learning and financial analysis from an Irish company, after looking a database to work it, I have selected their stock prices where I applied models such as Lineal Regression, K-Means Clustering, KNN, etc. The company that I have analysed is called Medtronic, a global enterprise in medical technology, services, and solutions, dedicated to improving healthcare for customers around the world. Medtronic has pioneered innovations in areas like cardiovascular therapies, diabetes, neurological problems and minimally invasive procedures. Essentially, the mission of the company is responsible for resolves pain, restoring health focusing on welfare by using biomedical engineering.

In this analysis, we are going to examine Medtronic’s stock performance using data science techniques as I mentioned before. The dataset comprises historical stock price for the last 5 years, capturing key features such as opening and closing prices, highs and lows, and so on. Using these methods, my aim is to find patterns that help me to have a better understanding of the stock prices, create forecasting and through different techniques get multiple angles to take a decision that benefit the company and stakeholders.

Finaly, examining this company stocks will help us not only to predict market behaviour but also to have a contribution of a deeper understanding of the company stocks during the giving period. This study will serve as a case study applying different machine learning models and time series analysis, also explaining answers and point of views, data visualization will be essential for a better comprehension of the dataset to present the findings.

# Objectives

Present the use of machine learning models to predict stocks prices based on historical trends, struggling with volatile markets, and identifies different market regimes. On the other hand, analyse time series to identifies trends & seasonality, handle with no seasonality trends, and the use of ARIMA due is a statistical model useful for forecasting of the variables from the data that contains, in this case, the Medtronic’ stock prices.

* General Objectives
  + Understand the dataset and the use of models on it.
  + The use of this techniques in real life.
  + Examine the stock prices of an Irish company.
  + Inspect the results obtained and give a deep explanation.
* Specific Objectives
  + Analyse the results through model performance metrics, robustness checks, and comparative analysis.
  + Clean and normalize the dataset and perform it through visualizations.
  + Determine optimal parameters to execute before evaluating the models.
  + Describe the results and their applicability in similar cases, particularly in relation to forecasting, smooth noisy data and identifying correlations.

# Problem Definition

Context: By selecting an Irish company, I have investigated its activities, products, services, and mission to create a general context. This involves a thorough analysis of its business operations, the market it serves, and the unique value propositions it offers. For this project, we then obtain a comprehensive database of the company's market price over the past five years. This data will be utilized for various machine learning model applications, enabling us to predict future market trends, analyse stock performance, and identify potential investment opportunities. The goal is to leverage historical data to inform decision-making processes and drive strategic initiatives.

Impact: This Project demonstrates the use of machine learning tools and the analysis of results, which are crucial for companies interested in investments. By leveraging data analysis and machine learning techniques, businesses can gain a broader understanding of stock trend, including identify long-term, short-term, detect anomalies, and make more informed decisions.

Additionally, the impact of this database, combined with the models applied in the project, lies in enhancing skills in machine learning and time series analysis for specific dataset. By understanding these models and their application to real world data, the skills acquired during this project will provide a deeper context for the data we currently handle and for the future as well, enabling more sophisticated and data that will drive me to take better decision making.

Importance: This Project highlights not only the tools we will use and describe but also the importance of the applicability in the business world. Many companies rely on data that is useful to decision making, particularly when working with financial market price data, and require advanced analytical insights before acting.

The implementation of machine learning can be transformative for financial institutions such as investment funds, banks, and consulting firms, as well as other industries that can benefit from data and insights. By leveraging predictive models and trend analysis, organizations can optimize strategies, mitigate risk, and uncover valuable opportunities hidden within their dataset.

# Business Description & Methods

In this initial step, the objective is to convert raw stock market data into a clean, structured format suitable for modelling. Despite, the original data did not have missing values, there were some observations in the data that was changed from excel file. On the other hand, I have created a simple description of the data to have a clear idea of the data’s observation and feature. A forecasting condition was applied where a time-based split was coded in Python before using a machine learning model. In this split, 80% of the data is allocated for training and 20% for testing.

In this part of the project, metrics such as Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE), along with cross-validation techniques like the Davies-Bouldin Index and Silhouette Score Method, were applied. Additionally, Time Series validation was utilized to assess performance across multiple time windows, ensuring stability against market regime changes. After analysing patterns, I monitored the prediction intervals in this project to quantify the uncertainty in the data set of Medtronic's stock prices. This is crucial as a goal if I am thinking to work in the financial sector as it allows for rebalancing based on predicted returns, optimizing client portfolios, and managing risk assessment scenarios based on predictions.

# Technologies & Accomplishment

This paper presents a detailed description of an Irish company specializing in healthcare products, as well as its performance in the stock market. After downloading the database containing stock prices for the past five years, along with information on the stock's open, close, adjusted close, high, low, and volume, a time series analysis is performed.

This approach is essential for managing databases containing information that may be stationary or contain noise, observing the behaviours of trends over the weeks, searching for repeating scenarios, and creating a pattern that allows us to determine when it is common to sell or buy. In addition to obtaining forecasts using the ARIMA model, this is relevant in the professional field because it facilitates informed decision-making in the financial sector, managing risks, and optimizing strategies.

Furthermore, this project aims to demonstrate the application of machine learning, an extremely efficient tool for handling large volumes of data. One of the machine learning models implemented in this project was linear regression. This model was used because it helps us identify ups and downs in trends over the period. It also attempts to predict stock prices based on the trends shown in the data. Finally, if we apply these models to other databases containing market prices for multiple companies, it will not help us understand the risk or return contribution to the portfolio.

Another machine learning model used was K-Means Clustering, which facilitated the grouping of the company's stocks under study, simplifying the analysis of price movements and volatility. This approach made it possible to identify which stocks exhibited similar behaviours. It is also an effective model for detecting anomalies or outliers, which are easily identifiable in the charts.

Finally, the K-nearest neighbours’ model was used. This model allows us to identify patterns in market price movements, which facilitates the identification of buying or selling opportunities. It also helps make forecasts based on the information obtained by detecting these patterns. Additionally, the model, when analysing the data, can detect anomalies or outliers. These deviations can be due to external factors, such as international trade decisions that directly or indirectly affect the company. Likewise, there may be internal events within the company that cause stock fluctuations. In this way, the model allows us to identify and understand the stock's behaviours during those specific periods.

As an extra, I applied "simple exponential smoothing," where the idea is that this model can study the behaviours of Medtronic's stock price. The reason for applying this model is that, according to the graph showing the stock price movement over the last five years, it shows a time series without seasonality. The reason it was applied to this database is to be able to make short-term forecasts of this company's market price movement, especially when there is no clear trend. Another very important aspect is that this model helps us eliminate noise that doesn't contribute to our forecasting objective. In other words, it smooths the data, considering the most recent data to identify behaviours and then make forecasts. This undoubtedly has valuable applications in the professional world, such as finance, retail, operations, and economics.

# Challenges encountered

During the execution of the project, one of the main challenges was finding a database that met the requirements necessary to apply machine learning models and time analysis. There are numerous databases on the internet that present observations and characteristics on various topics related to finance, marketing, and others. For this reason, I selected a completely numerical database that would be easy to use for applying models and, at the same time, would be useful not only for this project but also in the workplace.

Another significant challenge was ensuring that the database was up-to-date and met the necessary standards to obtain accurate and relevant results. When analysing the different available options, we considered not only ease of use but also data quality and applicability in different work scenarios. In the end, the chosen database offered a balance between recent data and its application, allowing for the efficient implementation of various machine learning models and techniques, resulting in deeper and more useful analyses in practical contexts. On the other hand, I also faced problems with running the machine learning models because some of the code was taken directly from other classes I took as a senior in college. These codes presented some specific conditions for a particular database, which meant that my database in this project didn't work. However, some code was changed to meet the code conditions, so I could ultimately solve them without any problems.

These problems weren't easy to solve. There were times when I worried that the solutions I had implemented weren't sufficient and that I would need to start from scratch. First, I had to carefully analyse each line of code to understand its functionality and how it related to the original database. Then, I made modifications to some of the code to meet the necessary conditions, adjusting it to work properly with my own database. Additionally, I had to ensure that all dependencies and libraries were up-to-date and compatible. The process required a lot of trial and error, as well as time spent researching solutions online and using database examples from other university classes. I worked diligently and exhaustively to identify errors and adjust the algorithms as needed. Finally, after several attempts and revisions, I got the machine learning models running smoothly. Obtaining accurate results was a great relief, and I put the perseverance into this project.

# Results & Analysis

Here I will present the results of the machine learning models implemented in this project using Python on Jupyter Notebook. The main objective of the assessment is analysed and predict patterns from a comprehensive dataset by employing various ML techniques, including linear regression, K-means and KNN.

Each model was rigorously trained and evaluated using multiple performance metric to ensure accuracy and reliability. I have evaluated model such as MAE, MSE and R2 scores, while regression model was measured by precision and accuracy. To enhance interpretation, I created a detailed visualization using Matplotlib and Seaborn, including:

* Trend analysis
* Feature correlation
* Model performance comparison

Lineal regression will help me in this project to understand the relationship between stock prices and identify key drivers of stock returns, however, there are some external factors that will have an influence in stock prices for this company such as interest rates, inflation, earning report from the company causing an effect both negative or positive, and sector performance. k-means will group similar stock based on price movements, volatility and correlation, helping me to manage a diversification of my portfolio and create strategies. Next, you will find K-Nearest Neighbours that will help to do process like classification, find any sign to buy, hold or sell, taking in consideration historical patterns that detect breakouts or reveals.

Nevertheless, like I was analysing a dataset about stock price from an Irish company I decided to apply other models study their trends, short-term forecasting and noise. The first model I applies was time series analysis, this will help to study how the past prices influence future movements, also, is a good tool that is capable to identify trends, seasonality, cycles and volatility, all in relation to the dataset. The other model named autoregressive integrated moving average or also known as ARIMA, is functional to capture autocorrelations, like the previous one, handle with no stationary data, and give us an understanding of short to medium term forecasting. Last but not least, simple exponential smoothing, that is useful to smooth out noise and highlight underlying trends, also as the previous one to make short term predictions and if we are working in a financial company is used in dashboards or to present a trading strategy.

These results demonstrate how combining different modelling approaches can yield both high accuracy and actionable insights that make data an important resource and lead any project.

## Data Base

The following graph illustrates the historical performance of Medtronic’ stock price over the past five years, providing investors with valuable insights into the company’s market behaviour. This detailed visualization captures all significant price movements, including periods of substantial growth, relative stability, and temporary declines. By examining these periods trend, market participants can identify recurring patterns, seasonal fluctuations, and potential turning points in the stock market valuation.

This chart will serve as an essential analytical tool for both short term traders and long term investors, enabling them to assess risk factors, time their entry and exit points, and make data better for investment decisions. Furthermore, the graphical representation helps contextualize how external events such as earnings reports, product launches or macroeconomic changes have had an influence in the stock market of this company, creating a comprehensive framework for future price predictions and portfolio strategy development.

A graph showing a line graph

AI-generated content may be incorrect.

Now, let’s examine the key features of this dataset obtained from Yahoo Finance, which provides comprehensive market information for each trading day:

* Date: Records the specific trading days when market operations occurred in the U.S. stock exchange.
* Open: Indicates the opening price of the stock at the beginning of the trading session.
* High: Captures the highest Price point the stock reached during that particular trading day.
* Low: Represents the lowest trading price the stock hit during the daily session.
* Close: Shows the final trading price when the market closed for the day.
* Adjusted Close: Reflects the closing price after accounting for corporate actions including:
  + Dividend payments to shareholders
  + Stock splits
  + Capital distributions

This adjustment provides a more accurate reflection of the stock’s historical value.

* Volume: Quantifies the total of shares traded during the session, serving as an important liquidity indicator.

These metric or features will help us to provide as traders and analyst with the essential data points needed to work, with technical analysis, historical performance evaluation, and a correctly application of machine learning to get information, understand and modelling these features and, of course, take better decisions.

A screenshot of a computer code

AI-generated content may be incorrect.

## Time Series Analysis & ARIMA

Continuing with the results of the models applied to the database, I performed the time series analysis, but not before training and testing the database, which has six features and 1,257 observations. Here, I considered using the Baselines model, which allows me to measure minimum performance before applying the ARIMA model, along with MAPE (Mean Absolute Percentage Error). These are the four predictions I have performed:

1. Historical
2. Last Year
3. Last Know Value
4. Naive Seasonal Forecast

A graph with blue rectangular bars

AI-generated content may be incorrect.

The result was that three of them were under 10%, which is good because that mean that the TSM in our database has less volatility or complex patterns, being “last\_year\_mean” the best baseline model with less than 6.5%. So, after getting this result the next step is starting to apply ARIMA model to reduce noisy in the data, get tendencies and identify some patterns, having a better understanding and precision than just the Baselines model.

Once I have applied ARIMA model, this was the first view.

A graph of a bar graph

AI-generated content may be incorrect.

Consequently, based on the “Autocorrelation” image, it is evident that the series is not stationary. Furthermore, this conclusion is corroborated by the P-Value and ADF Statistic, with values of 0.669774 and -1.208857, respectively.

In this part, now we must transform into a stationary station using (1,1,0), which is for trended data and is workable for data about stock prices, like this project.

A graph of different values

AI-generated content may be incorrect.

After applying ARIMA model (1,1,0), this is the result from the Medtronic stocks.

A comparison of a graph

AI-generated content may be incorrect.

In this case, according from the “Autocorrelation” image we see that now our series is stationary. I have calculated the p-value and ADF statistics again to confirm that is stationary and the result was 0.000 and -19.70 respectively. Next, I have coded the Akaike information criterion (AIC) on python to compare the quality of different ARIMA models in terms of best balance of fit and determine which one is the best, and the result from the smallest value was (1,1,1) 4421.30.

Below, I have compared the “actual” price market with the ARIMA (1,1,1) forecast.

A graph with red and blue lines

AI-generated content may be incorrect.

Afterwards, I have used the ARIMA model but with confidence intervals that are particularly valuable for risk assessment and decision-making. The 80% interval might be suitable for routine planning where some risk is acceptable, while the 99% interval would be appropriate for conservative scenarios where being wrong could have severe consequences.

A graph showing the growth of the stock market

AI-generated content may be incorrect.

## Linear Regression

Here, I have started to work with the applying of machine learning models to the data base that I have been working on. This is because, this model will allow me to predict trends in short periods of time, also, will show me the best baseline model using other complex models

After made some examples finding which would be the best option for being a dependent and independent variable, I have chosen as an independent variable (x) “Open” and as a dependent variable (y) “Close”. The reason why I have selected these two variables is because I want to focus on intraday prediction from close price from the open price from this data I am working on.

We can see that there is a positive correlation between these two variables. After training and testing these values, I have got the CV mean and the standard deviation, which values are 0.995 and 0.0009 respectively. CV mean has a high value which could mean overfitting and leakage because of my dataset, this is a financial data where the markets are in constantly changing, so that is the reason for the outcome’s values.

A graph with blue dots

AI-generated content may be incorrect.

First, cross-validation must be checked, so the data should be trained and tested. Linear regression has been specified for the cross-validation process. This method ensures that the model's performance is evaluated accurately by splitting the dataset into training and testing sets. Additionally, k-fold cross-validation with k=10 was applied to split the dataset to reduce the risk of overfitting. This technique divides the data into 10 subsets and uses each subset as a test set while the remaining ones serve as the training set. This will help to provide a robust estimate of the model's performance.

However, for datasets involving stock prices, an alternative cross-validation method is required due to the temporal dependency of the data points. In this case, time series split is ideal for the project. Time-series split maintains the chronological order of observations, ensuring that future data points are never used to predict past values. This preserves the integrity of the time-related information and provides a more accurate evaluation of the model's predictive power.

## K-Means Clustering

This is the next machines learning model that I have decided to develop for my project. This will help to identify same groups in stocks with similar behaviours. It is important if we are talking about risk assessment, trading strategies or portfolio diversification using the centroids to take a decision.

A graph with red and blue dots

AI-generated content may be incorrect.

After obtaining the result, we can understand the positive distribution and centroids, which indicate each stock's group. For instance, centroids can serve as equilibrium points for risk categories. If a price is near the centroid, I might hold or ignore it, but if it is far, I should start selling or stay alert. Opportunities may arise to sell when prices approach centroids and buy when stocks are oversold (far from centroids), anticipating a bounce.

To elaborate further, understanding the positive distribution helps in identifying patterns and trends within the stock market. The centroids act as benchmark levels where the stock prices tend to change over time. This change pull towards the centroids can be leveraged to make informed trading decisions. When stock prices hover around the centroids, it often signifies a balanced state where supply meets demand. Hence, holding onto such stocks might be a prudent decision.

On the other hand, if the stock prices stray significantly from these centroids, it indicates either an overbought or an oversold condition. For example, if a stock operation is too high above the centroid, it may present a good opportunity to sell before a potential price correction takes place. Conversely, if a stock fall well below the centroid, this could signal an oversold condition, making it an attractive time to purchase, expecting a rebound towards the centroid.

On the other hand, both cross validation and hyperparameter tuning I have used Silhouette Score and Davies-Bouldin Index. A graph of different types of data

AI-generated content may be incorrect.

Davies suggested taking the smallest value, which is two, while Silhouette recommended taking the highest value, which is also two. Therefore, for this data set, it is advised to use 2 as the number of clusters.

Selecting the optimal number of clusters is often crucial in ensuring the validity and reliability of this kind of dataset that I am using for this project. By using two clusters based on both Davies' and Silhouette's suggestions, I can balance between the smallest and highest values, potentially providing a more robust and consistent outcome. This approach helps me in capturing the best structure of the data and may lead to better insights and more accurate interpretations.

## K-Nearest Neighbours

I used the K-Nearest Neighbours model on my dataset to identify patterns for short-term forecasting. It helps analyse repeating trends and reactions to similar market conditions and captures nonlinear relationships without complex modelling. This model is particularly useful because classifies data points based on their proximity to other points in the dataset, Allowing robust pattern recognition even in volatile or unpredictable scenarios. By comparing new data against historical records, the K-Nearest Neighbours model can predict future movements with significant accuracy. Additionally, it doesn't require the assumptions of linearity, making it valuable for diverse and complex datasets where traditional models might fail.

After training and testing, standardize my data, and execute the machine learning model, the result was:

* R2 score: 0.9865  
  This result is close to 1, which means that in this occasion this model captures almost all variability of my data set about stock prices. Nevertheless, this could also mean that there is overfitting.
* MAE: 1.3231 & MSE: 3.314  
  Which the average of error in my model prediction, in this case my model’s prediction is $1.32 away from the actual prices. Meanwhile, MSE means that my model’s squared error average is 3.314, focusing on large errors than small errors.

A graph with blue dots and red lines

AI-generated content may be incorrect.

The blue dots represent the model's predictions, while the red line indicates perfect predictions. In this visualization, each blue dot corresponds to a specific prediction made by the model for a given data point. The closer these blue dots are to the red line, the more accurate the model's predictions are. Deviations from the red line show errors or inaccuracies in the model's predictions, providing insights into areas where the model may need further refinement or improvement. This comparison helps assess the model’s performance and guides adjustments to enhance the predictive accuracy.

On the other hand, as we have been doing through this project, it is time to apply hyperparameter tuning and cross-validation for each machine learning model. In this case for this machine learning model (KNN) I did not apply cross validation, this is because the previous machine learning model explained I coded “Time Series Split”, that means that for this exercise will have the same name “tscv” and it was split into 5, increasing training. However, in the case of hyperparameter I have used another method for k-nearest neighbours, trying to find the best value for the ML model, which is the number of similar data point that the model considers doing any kind of predictions. After analysing the results from the parameter grid, this determine that the optimal parameter is 10. This indicates that using 10 neighbours will achieve a balance that yields predictions with less errors and better generalization for the dataset, allowing the model to consider sufficient data points and avoid outliers.

# Simple Exponential Smoothing

Finally, after applying Time Series Model and ARIMA model to transform into a stationary series I have decided to continue applying a basic series forecasting technique, simple exponential smoothing. This will help me to predict futures values calculating past observations and averaging from my dataset. Like this data set is about stock prices from Medtronic, I consider that is interesting to understand the use of SES in this dataset, in this case, SES will help me to find opportunities in the stock market when prices deviate from the smoothed trend, this smoothed will remove the noise and focus on the short-term trends.

A graph of a stock market

AI-generated content may be incorrect.

After obtaining the results from single exponential smoothing, I decided to continue using three alpha values: 0.2, 0.5, and 0.6. These alpha values allow me to assess how different parameters influence the forecast. For my dataset, the smallest value is 0.2 that focus more on past observations, making the model slow to react to recent changes. This can be useful in stable environments where historical trends are less volatile. On the other hand, the highest alpha value was 0.8 that react quickly to recent changes and are more suitable in dynamic environments, capturing trends that may not be evident with smaller alphas.

At first glance, an alpha of 0.8 appears to follow the stock market trend best, reacting very quickly to the latest data points. However, this high sensitivity may risk overfitting, where the model becomes too responsive to short-term fluctuations and loses its ability to generalize.

Therefore, after careful consideration, I determined that an alpha of 0.6 is optimal for my dataset. It provides a balance between capturing trends and smoothing out noise effectively. This middle ground allows the model to remain responsive to new information while maintaining enough stability to avoid the pitfalls of overfitting. By using 0.6, I aim to achieve reliable and accurate predictions that align well with observed data patterns without being overly influenced by temporary anomalies.

A graph of different numbers

AI-generated content may be incorrect.

# Conclusion

During the project, I have studied the application of models in time series and machine learning models in databases about the price of a company’s stocks over the last five years. Additionally, I have utilized codes recently covered in class, which has helped improve my skills in these tools and their practical application across various sectors such as the industrial sector.

Some of the industries are:

* **Finance**: Machine learning models can be used extensively in finance for analysing stock prices, predicting market trends, and managing risks. For example, banks and hedge funds can leverage these models to forecast financial markets and optimize investment strategies.
* **Retail**: In retail, demand forecasting models help businesses predict product demand, manage inventory, and adjust pricing strategies. So, customer segmentation allows me as a retailer to understand consumer behaviours and tailor marketing efforts, accordingly, enhancing customer satisfaction and loyalty.
* **E-commerce**: E-commerce platforms utilize machine learning for Customer Relationship Management (CRM) by predicting customer preferences and behaviours. For example, product recommendation systems suggest items based on user history and preferences, thereby increasing sales and improving user experience.
* **Healthcare**: The healthcare industry benefits from machine learning-assisted diagnosis, which can analyse medical images and patient data to provide accurate diagnoses. Disease prevention models can also predict outbreaks and help in planning preventive measures.
* **Logistics and Transport**: Machine learning will help me to find a route optimization by finding the most efficient paths for delivery vehicles, reducing costs, and improving delivery times. Maintenance prediction models anticipate equipment failures, enabling preventative maintenance and minimizing downtime.
* **Energy**: Energy companies use machine learning to predict electricity consumption patterns, ensuring efficient energy distribution and load management. Fraud detection algorithms identify unusual activities and prevent fraudulent transactions.
* **Agriculture**: In agriculture, crop prediction models forecast produce based on weather patterns, soil conditions, and other factors. Pest detection systems monitor crops and identify pest infestations early, allowing timely interventions.
* **Entertainment**: Content recommendation engines in streaming services suggest movies, shows, or music to users based on their viewing or listening preferences. Audience prediction models help entertainment companies to find potential viewership and plan their content release strategies.
* **Politics**: In politics, machine learning can assist in crime prevention by analysing crime data to identify patterns and predict future incidents. Optimization of public resources involves using predictive models to allocate resources efficiently, improving public services and infrastructure.

Throughout the project, I integrated these concepts with real data, improving not only my technical skills but also my understanding of how these models can drive innovation and efficiency across many fields.

# References

* Auffarth, B. (2021) Machine Learning for Time-Series with Python. Birmingham: Packt Publishing.
* Pal, A. and Prakash, P.K.S. (2017) Practical Time Series Analysis. Birmingham: Packt Publishing.
* Peixeiro, M. (2022) Time Series Forecasting in Python. Shelter Island, NY: Manning Publications.
* Meta Brains (2022) Python for Machine Learning - The Complete Beginner's Course. Packt Publishing.
* Brownlee, J. (2025) Time Series Forecast Uncertainty Using Confidence Intervals in Python [Online]. Available at: <https://machinelearningmastery.com/time-series-forecast-uncertainty-using-confidence-intervals-python/>
* mGalarnyk (2025) Python\_Tutorials [Online]. Available at: <https://github.com/mGalarnyk/Python_Tutorials>
* Analytics Vidhya (2025) Linear Regression in Python with Scikit-Learn [Online]. Available at: <https://medium.com/analytics-vidhya/linear-regression-in-python-with-scikit-learn-e1bb8a059cd2>
* Towards Data Science (2025) K-Nearest Neighbor (KNN) Python [Online]. Available at: <https://towardsdatascience.com/k-nearest-neighbor-python-2fccc47d2a55>
* Scikit-learn (2025) KNeighborsRegressor [Online]. Available at: <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsRegressor.html>
* Scikit-learn (2025) GridSearchCV [Online]. Available at: <https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html>
* Scikit-learn (2025) TimeSeriesSplit [Online]. Available at: <https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.TimeSeriesSplit.html>
* StatsModels (2025) StatsModels Documentation [Online]. Available at: <https://www.statsmodels.org/stable/index.html>
* IBM (2025) Linear Regression [Online]. Available at: <https://www.ibm.com/think/topics/linear-regression>
* IBM (2025) K-Means Clustering [Online]. Available at: <https://www.ibm.com/think/topics/k-means-clustering>
* Elastic (2025) What is KNN? [Online]. Available at: <https://www.elastic.co/what-is/knn>
* Medtronic (2025) Medtronic Official Website [Online]. Available at: <https://www.medtronic.com/>
* Yahoo Finance (2025) Yahoo Finance [Online]. Available at: <https://finance.yahoo.com/>
* DataCamp (2025) Introduction to Machine Learning in Python [Online]. Available at: <https://www.datacamp.com/community/tutorials/introduction-machine-learning-python>