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.

Introduction

Table of Contents

[Introduction 3](#_Toc197121379)

# 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: A través de una búsqueda y selección de una empresa irlandesa, se investiga la compañía para generar un contexto general de lo que hace la empresa, sus productos y/o servicios y misión. Luego, para ámbito de este proyecto, se consigue una base de datos la cual sean aptas para las aplicaciones de modelos de machine learning, que en esta ocasión será el precio de mercado de la empresa en los últimos cinco años.

Impact: Este proyecto da conocer el uso de las herramientas de machine learning y los análisis de los resultados que son importantes para las empresas interesadas en inversiones, ya que el impacto mediante el uso de herramientas de análisis de datos y de machine learning, les va a permitir tener un contexto mas amplio sobre las acciones y determinar patrones a lo largo de los años, anomalías, etc. Por otro lado, el impacto que esta base de datos juntos con los modelos que se aplicaron en el proyecto es que mejorara las habilidades en el uso de machine learning y análisis de tiempo para bases de datos especifica, no obstante, al tener conocimiento de los modelos y la aplicación en algunas bases de datos, estas habilidades adquiridas durante el desarrollo del trabajo nos permitirán tener un contexto mas amplio de los datos que manejamos y llegaremos a manejar.

Importance: La importancia de este proyecto es tener en cuenta las herramientas que utilizaremos y describiremos no solo como parte de este proyecto, sino que también, como un herramienta que podemos llegar a utilizar en las empresas, esto se debe a que hay compañías que trabajan con base de datos, en este caso de precio en el mercado, que necesitan saber que hacer antes de tomar una decisión, es por esto por lo que la aplicación de machine learning podría ser fundamental en instituciones financieras como fondo de inversiones, bancos, consultoras, y otro tipos de sectores que se puede beneficiar de los resultados del estudio de los datos.

# Methods & Tecniques

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 in the financial industry as it allows for rebalancing based on predicted returns, optimizing client portfolios, and managing risk assessment scenarios based on predictions.

# Technologies & Accomplishment

En este trabajo se presenta una descripción detallada de una empresa irlandesa especializada en productos para la industria de la salud, así como su comportamiento en el mercado bursátil. Tras descargar la base de datos que contiene los precios de las acciones de los últimos cinco años, junto con la información sobre apertura, cierre, cierre ajustado, máximo, mínimo y volumen de las acciones, se procede a realizar un análisis mediante series temporales.

Este enfoque es fundamental para manejar bases de datos que contienen información que puede llegar a ser estacionaria, que contenga ruido, mirar el comportamiento de las tendencias a lo largo de las semanas buscando escenarios que se repitan, creando un patrón que permita saber cuándo es común vender o comprar. Además de obtener pronósticos mediante el uso del modelo ARIMA, esto es relevante en el ámbito profesional porque facilita la toma de decisiones informadas en el sector financiero, gestionando riesgos y optimizando estrategias.

Por otro lado, este proyecto tiene como objetivo demostrar la aplicación del machine learning, una herramienta sumamente eficiente para el manejo de grandes volúmenes de datos. Uno de los modelos de machine learning implementados en este proyecto fue la "regresión lineal". Se uso este modelo ya que nos ayuda a identificar las subidas y bajadas de las tendencias al largo del periodo, también nos intenta pronosticar el precio de las acciones basado en las tendencias que se muestran en los datos, y finalmente, si aplicamos estos modelos a otras bases de datos que contengan los precios de mercado de varias empresas, no ayudara a entender a la contribución de riesgo o rendimiento en el portafolio.

Otro modelo de machine learning utilizado fue el "K-Means Clustering", el cual facilitó la agrupación de las acciones de la empresa estudiada, simplificando el análisis del movimiento de los precios y la volatilidad. Este enfoque permitió identificar qué acciones presentaban comportamientos similares. Además, es un modelo eficaz para detectar anomalías o valores atípicos, los cuales son fácilmente identificables en las gráficas.

Por último, se empleó el modelo "K-nearest neighbors". Este modelo permite identificar patrones en los movimientos de precios del mercado, lo cual facilita la identificación de oportunidades de compra o venta. Asimismo, ayuda a realizar pronósticos basados en la información obtenida al detectar dichos patrones. Adicionalmente, el modelo al analizar los datos es capaz de detectar anomalías o valores atípicos. Estas desviaciones pueden deberse a factores externos, como decisiones en el comercio internacional que afecten directa o indirectamente a la empresa. Asimismo, puede haber eventos internos en la empresa que provoquen fluctuaciones en las acciones. De esta manera, el modelo permite identificar y entender el comportamiento de las acciones durante esos periodos específicos.

Como extra he aplique “simple exponential smoothing” en donde la idea es que este modelo pueda estudiar el comportamiento del precio de la acción de la empresa Medtronic. La razón para la aplicación de este modelo es que según la gráfica que nos muestra el movimiento del precio de las acciones a lo largo de los últimos 5 años, es una gráfica que nos muestra series temporales sin estacionalidad. La razón por la cual se aplico para esta base de datos es para poder pronosticas a corto plazo el movimiento del precio en el mercado de esta empresa, especialmente cuando no hay una tendencia clara. Otro aspecto muy importante es que este modelo nos ayuda a eliminar ruido que no le aporten a nuestro objetivo de realizar pronósticos, en otras palabras, suaviza los datos lo que hace que tenga en cuenta los datos mas recientes para identificar el comportamiento y luego poder hacer los pronósticos. Esto sin duda alguna, tiene una aplicación valiosa en el mundo profesional como lo puede ser en finanzas, retail, operaciones y en la economía.

# Challenges encountered

Durante la ejecución del trabajo, uno de los principales desafíos fue encontrar una base de datos que cumpliera con los requisitos necesarios para aplicar modelos de machine learning y análisis de tiempo. En internet existen numerosas bases de datos que presentan observaciones y características sobre diversos temas relacionados con finanzas, mercadeo, entre otros. Por esta razón, seleccioné una base de datos completamente numérica que fuera fácil de utilizar para la aplicación de modelos y que, al mismo tiempo, resultara útil no solo para este proyecto, sino también en el ámbito laboral.

Además, otro desafío considerable fue asegurarse de que la base de datos estuviera actualizada y cumpliera con los estándares necesarios para obtener resultados precisos y relevantes. Al analizar las diferentes opciones disponibles, se tomó en cuenta no solo la facilidad de uso sino también la calidad de los datos y su aplicabilidad en diferentes escenarios laborales. Al final, la base de datos elegida ofrecía un equilibrio entre datos recientes y la aplicación que le podía dar, permitiendo implementar diversos modelos y técnicas de machine learning eficientemente, lo cual resultó en análisis más profundos y útiles en contextos prácticos.

Por otro lado, también enfrente problemas con los modelos de machine learning al momento de ejecutarlos ya que algunos de los códigos fueron tomados directamente de otras clases tome como estudiante en el último año universitario. Estos códigos presentaban algunas condiciones específicas para una base de datos determinada, lo que hace que para mi base de datos en este proyecto no funcionara. No obstante, se cambiaron algunos códigos con el fin de que cumplieran con las condiciones de los códigos, para que finalmente pudiera dar respuesta sin ningún problema.

Estos problemas no fueron fáciles de resolver. Hubo momentos en los que me preocupaba que las soluciones que había implementado no fueran suficientes y que necesitaría comenzar desde cero. Primero, tuve que analizar detenidamente cada línea de código para entender sus funcionalidades y cómo estaban relacionadas con la base de datos original. Luego, realicé modificaciones en algunos códigos para cumplir con las condiciones necesarias, ajustándolos para que funcionaran correctamente con mi propia base de datos. Además, tuve que asegurarme de que todas las dependencias y bibliotecas estuvieran actualizadas y compatibles.

El proceso requería mucha prueba y error, así como tiempo dedicado a investigar soluciones en foros y documentación técnica. Trabajé de manera diligente para identificar los errores y ajustar los algoritmos según fuese necesario. Finalmente, después de varios intentos y revisiones, logré que los modelos de machine learning funcionaran sin inconvenientes. Obtener resultados precisos y fiables fue un gran alivio y un testimonio del esfuerzo y la perseverancia que puse en este proyecto.