Coca Cola (KO) Stock Price Prediction Modelling Project

10th February 2022

1 Introduction

In this report I elaborate on my problem formulation for predicting the stock price of the Coca Cola Stock (stock market ticker: 'KO').

The Coca Cola Company is a multinational beverage corporation involved in manufacturing, retailing, and marketing of the aforementioned. It has a history of more than a century in which it has managed to become a household brand in many regions of the world.

However, the specific niche of beverages The Coca Cola Company is involved in is heavily reliant on two ingredients: sugar and caffeine (derived from coffee beans and coffee's stock market ticker: 'KC=F'). From the lens of Finance, there exists a hint at the relation between commodity stock price of Sugar (stock market ticker: 'SB=F') and KO as stated in an article by Stevenson (2017), however, little literature is found on the impact of commodity (sugar and coffee) prices on the KO stock.

2 Problem Formulation

The problem I intend to advance in is: given historical stock data of the KO stock, can the next day's closing price be predicted accurately? If so, do average price of sugar and coffee improve the stock price prediction of KO?

The **datapoints** of the problem are the stock market data on a single day. The stock market data will be having the following instances/characteristics: KO's Opening Price ($), KO's Highest Price ($), KO's Lowest Price ($), KC=F's Average Price ($), SB=F's Average Price ($). One datapoint has the characteristics of the stock market data as its **features**. The KC=F's Average Price ($) and SB=F's Average Price ($) feature will be calculated using their historical stock market data by implementing the formula: (opening price + closing price)/2. The resulting average price will be used in as a feature.

The value we are interested to procure is KO's Closing Price ($), which is the **label** of a datapoint.

2.1 Problem Formulation: Visual Summary

Datapoints	**Features**					**Label**
Date	KO: Opening	KO: High	KO: Low	KC=F: Average	SB=F: Average	KO: Closing
1.1.10	60.00	61.12	59.88	250.74	18.45	60.60

2.2 Problem Formulation: Sources of Data

The range of the set of datapoints we will use is from 1st of January 2010 to 1st of January 2020 (approx. 2517 datapoints). Our sources of data are as follows:

- KO's Historical Stock Market Data: Yahoo! Finance https://finance.yahoo.com/quote/KO/history?period1=1262304000&period2=1577836800&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true
- SB=F's Historical Stock Market Data: Yahoo! Finance https://finance.yahoo.com/quote/SB%3DF/history?period1=1262304000&period2=1577836800&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true
- KC=F's Historical Stock Market Data: Yahoo! Finance https://finance.yahoo.com/quote/KO/history?period1=1262304000&period2=1577836800&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true