Capitec stock prices modeling using regression methods

Sifiso Rimana
rimanasifiso@gmail.com

November, 2024

Abstract

In this paper, we

Introduction

Stock price prediction involves estimating the future value of a company's stock or the overall stock market. This process is crucial for investors and financial institutions as it aids in making informed investment decisions, managing risks, and ensuring financial stability. Accurate stock predictions can help investors maximize returns and minimize losses [1].

The methods used for stock price prediction include fundamental analysis, technical analysis, machine learning, and quantitative trading. Fundamental analysis evaluates a company's financial health by analyzing financial statements, management quality, industry conditions, and economic factors to determine the stock's intrinsic value. Technical analysis studies historical price movements and trading volumes to identify patterns and trends that might indicate future price movements. Machine learning and quantitative trading use advanced algorithms to analyze large financial datasets, including historical prices, financial news, and social media sentiment, to predict stock movements [2].

Capitec Business Overview

Company: Capitec Bank Holdings Limited
Ticker: CPI (Johannesburg Stock Exchange)
Current Stock Price: 326.858 ZAR
Market Capitalization: 376.19 billion (ZAR)
Beta: 1.03, indicating that the stock's volatility is like the overall market.

Capitec Bank, established in 2001, is a prominent South African retail bank headquartered in Stellenbosch. It has rapidly grown to become the largest retail bank in the country by customer numbers, serving over 23 million clients as of October 2024 [3]. The Bank offers a wide range of financial services including transactional banking, savings & investments, credit facilities, and insurance products [4].

In 2019, Capitec expanded into business banking by acquiring Mercantile Bank, enhancing its services to small and medium-sized enterprises (SMEs). This acquisition enabled Capitec to offer tailored business banking solutions, including transactional accounts, credit facilities, and merchant services, supporting the growth and development of businesses in South Africa.

Capitec has demonstrated robust financial performance. In the six months leading up to August 31, 2024, the bank reported a 36% increase in interim profit, reaching 6.394 billion rand. This growth was attributed to reduced loan losses and an increase in transactions and commission income. This can be evidenced by the continued rise in its stock price. put a line chart from July to Nov

Capitec's success is largely due to its focus on customer needs, offering user-friendly digital platforms and maintaining extended branch hours to enhance accessibility. The bank's commitment to innovation and efficiency has solidified its position as a leader in the South African banking sector [5].

Data and Methods

In this paper, we attempt to estimate Capitec's closing stock price for the day using the classical linear regression method. We will also compare results obtained using linear regression with those obtained using a random forest regressor. The models will be judged by their train and test metrics, namely R^2

and root mean squared error (RMSE). The data was obtained from the package `yfinance`, which is free and open source, and the data is used as per Yahoo Finance's terms [6]. The data downloaded ranges from January 1, 2023, to November 18, 2024. The below table explains the dataset returned by the library:

Variable	Explanation
date	The date in the format of Year-Month-Date. It tells us about the date for which the opening price, closing price, highest price, lowest price, and volume were recorded.
close	The closing price of Capitec stock at a particular date, measured in South African Rand (ZAR). It represents the value of a share at the end of the day when the stock exchange closes.
open	The opening price of Capitec stock on a particular day, measured in ZAR. It is the value of a share when the stock exchange opens for trade and may differ from the previous day's closing price.
high	The highest price of a Capitec share on a particular day, measured in ZAR. It is the maximum price at which the stock was traded during the day and helps assess the general movement of the stock.
low	The lowest price of a Capitec share on a particular day, measured in ZAR. It is the minimum price at which the stock was traded on that day and is also used to understand the stock's overall movement.
volume	The volume of Capitec stock traded on a particular day, measured in the number of shares traded per day. It represents the total number of shares traded during the day and is a crucial metric for assessing market activity.
adjusted close	The adjusted closing price of Capitec stock, measured in ZAR. This value accounts for events such as stock splits, dividends, and rights offerings that might affect the stock's price. It provides a more accurate reflection of the stock's value over time compared to the regular closing price. We will not be using this variable for simplicity.

Table 1: Explanation of the dataset variables used in the analysis.

Why?

We will be predicting the closing price of Capitec for a given date. All the variables will be treated as independent (also called predictors), and the closing price will be treated as the dependent variable (also called the response).

Mathematical Formulation of Linear Regression

Linear regression is a statistical model that estimates a linear relationship between a real valued response and one or more explanatory variables. The explanatory variables are independent variables while the response variable is dependent. The response variable is also known as the predicted variable, and the explanatory variables are also know as the predictors. If the predictor is a single variable, then we have a simple linear regression. If the predictors are two or more independent variables, then we have a multiple linear regression problem. Linear regression falls into the Supervised Learning category of learning algorithms since it 'learns' from the labeled datasets. That is, given the labeled dataset, it learns the linear relationship that exists between explanatory variables and the response variable, if there is such a meaningful relationship.

We now discuss the mathematical formulation of linear regression. Given a dataset $\left\{\left(\mathbf{x}^{(i)}, y^{(i)}\right)\right\}_{i=1}^{m}$ of m observations, where $\mathbf{x} \in \mathbb{R}^n$ and $y \in \mathbb{R}$, under the assumption of linearity, we want to find a linear (not necessarily deterministic) function $h : \mathbb{R}^n \to \mathbb{R}$, such that

$$h_{\mathbf{w}}(\mathbf{x}^{(i)}) = w_0 \mathbf{x}_0^{(i)} + w_1 \mathbf{x}_1^{(i)} + \cdots + w_n \mathbf{x}_n^{(i)} \quad \text{for } i = 1, \ldots, m \tag{1}$$

where, if we let our feature matrix to be $\mathbf{X} \in \mathbb{R}^{m \times n}$ matrix containing m observations of n independent explanatory variables used to predict $y \in \mathbb{R}$, then $\mathbf{x}^{(i)}$ is the i^{th} observation vector; $\mathbf{w} \in \mathbb{R}^n$ is our

parameter vector (also called weights). So, compactly, equation (1) becomes

$$h_{\mathbf{w}}(\mathbf{X}) = \mathbf{w}^T\mathbf{X} = \mathbf{X}\mathbf{w} \tag{2}$$

The cost function $J(\mathbf{w})$

Now, given the dataset, how do we find or learn the parameters $\mathbf{w}$? We of course want to choose $\mathbf{w}$ so that equation 2 is as close to the true observed y as possible. A **cost function** is a function that measures how close the hypothesis function is to the corresponding true y's. Here it is below:

$$J(\mathbf{w}) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_{\mathbf{w}}(\mathbf{x}^{(i)}) - y^{(i)}\right)^2 \tag{3}$$

or in vector form

$$J(\mathbf{w}) = \frac{1}{2m}\left[\mathbf{w}^T\mathbf{X} - \mathbf{y}\right]^T\left[\mathbf{w}^T\mathbf{X} - \mathbf{y}\right] \tag{4}$$

Learning Methods

The primary goal in most machine learning problems is to minimize the objective function at hand, by this, the algorithm learns the patterns hidden in the dataset. There are numerous ways or methods for our machine algorithm to learn the patterns in our data. In this paper, we discuss the Gradient Descent, Newton-Raphson's Method, and the normal equations.

Gradient Descent Method

Gradient Descent is a method that learns the parameters $\mathbf{w}$ by iteratively updating $\mathbf{w}$ according to this rule

$$\mathbf{w} := \mathbf{w} - \alpha\nabla_{\mathbf{w}}J(\mathbf{w})$$

where α is the learning rate, and $\nabla_{\mathbf{w}}J(\mathbf{w})$ is

$$\frac{\partial}{\partial\mathbf{w}}J(\mathbf{w}) = \nabla_{\mathbf{w}}J(\mathbf{w}) = \frac{1}{m}\sum_{i=1}^{m}\left(h_{\mathbf{w}}(\mathbf{x}^{(i)}) - y^{(i)}\right)\mathbf{x}^{(i)} \tag{5}$$

whose vectorized form is

$$\nabla_{\mathbf{w}}J(\mathbf{w}) = \frac{1}{m}\mathbf{X}^T\left[\mathbf{w}^T\mathbf{X} - \mathbf{Y}\right] \tag{6}$$

The gradient descent algorithm is then summarized below:

Algorithm 4.1 (Gradient Descent)**.** The gradient descent method can be run until convergence (i.e., until the updated $\mathbf{w}$ is negligible): Run until convergence { $\mathbf{w} := \mathbf{w} - \alpha\nabla_{\mathbf{w}}J(\mathbf{w})$ } or we can choose the number of iterations to run the algorithm:

```
Initialize w at random
for i until the number of iterations:
    update w according to the gradient descent updating rule
return w
```

□

One major benefit of gradient descent is its ability to handle large datasets efficiently, and can also converge to a global minimum for convex functions, ensuring optimal solutions for problems like linear regression. Variants of gradient have been developed to improve the original one, and they include the stochastic and mini-batch forms, where only subsets of data are used per iteration, making it scalable and computationally less expensive.

However, there are some drawbacks: the algorithm is sensitive to the learning rate, which requires careful tuning—too high a rate can cause divergence, while too low can result in very slow convergence. Additionally, gradient descent can get stuck in local minima or saddle points in non-convex problems, which makes it less effective for more complex models, such as deep neural networks. Gradient descent may require many iterations to converge, particularly with large datasets, and is computationally expensive if not properly optimized.

Newton-Raphson's Method

Another method for learning is the Newton-Raphson's Method, which has the following updating rule:

$$\mathbf{w} := \mathbf{w} - \mathbf{H}^{-1}\nabla_{\mathbf{w}} J(\mathbf{w})$$

where $\mathbf{H}$ is the Hessian of $J(\mathbf{w})$, which is found to be

$$\mathbf{H} = \nabla^2_{\mathbf{w}} J(\mathbf{w}) = \frac{1}{m}\mathbf{X}^T\mathbf{X} \tag{7}$$

It's algorithm is given below:

Algorithm 4.2 (Newton-Raphson's Method)**.** The Newton-Raphson method can also be run until convergence (i.e., until the updated $\mathbf{w}$ is negligible): Run until convergence $\{ \mathbf{w} := \mathbf{w} - \mathbf{H}^{-1}\nabla_{\mathbf{w}} J(\mathbf{w}) \}$ or we can choose the number of iterations to run the algorithm:

```
Initialize w at random
for i until the number of iterations:
    compute the gradient and Hessian
    update w using Newton-Raphson updating rule
return w
```

□

A significant benefit of Newton-Raphson is its fast convergence rate near an optimal solution, since it makes use of both gradient and second-order information (the Hessian). This means it often converges much faster than gradient descent, especially for well-behaved convex functions.

Additionally, Newton-Raphson can adaptively adjust its step size based on the curvature of the cost function, allowing it to take larger steps in flat areas and smaller ones in steep regions, which enhances its efficiency. However, the drawbacks are notable too: calculating the Hessian matrix is computationally expensive, especially for high-dimensional datasets, making the method less practical for large-scale machine learning problems.

Furthermore, Newton-Raphson may struggle in situations where the Hessian is not invertible or near *saddle points*, which can lead to divergence or unpredictable behavior. Despite its fast convergence in well-defined scenarios, these computational complexities and sensitivities limit its applicability in large-scale or complex machine learning models, where simpler first-order methods often prove to be more practical.
We can run this update rule using the similar approaches discussed in the gradient descent algorithm.

Normal equations

In linear regression, we define a matrix known as the **design matrix**, denoted as $\mathbf{X}$. The design matrix $\mathbf{X}$ is an $m \times (n+1)$ matrix (where m is the number of training examples and n is the number of features, plus an additional column for the intercept term). Each row of $\mathbf{X}$ represents a single training example, including its features:

$$\mathbf{X} = \begin{bmatrix} - - (\mathbf{x}^{(1)})^T - - \\ - - (\mathbf{x}^{(2)})^T - - \\ \vdots \\ - - (\mathbf{x}^{(m)})^T - - \end{bmatrix}$$

where $\mathbf{x}^{(i)}$ is a row vector containing the features of the i-th training example. The normal equations are derived to find the optimal parameters $\mathbf{w}$ that minimize the cost function $J(\mathbf{w})$ in linear regression. To minimize $J(\mathbf{w})$ (equation 4), we take the derivative with respect to $\mathbf{w}$ and set it to zero:

$$\nabla_{\mathbf{w}} J(\mathbf{w}) = \mathbf{X}^T\mathbf{X}\mathbf{w} - \mathbf{X}^T\mathbf{y} = 0$$

Solving for $\mathbf{w}$, we obtain the normal equations:

$$\mathbf{w} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

This provides a closed-form solution for finding the optimal parameters $\mathbf{w}$.
The normal equations offer some key benefits and limitations when compared to iterative methods such as gradient descent since they provide a direct solution without requiring iterations. This makes the method easier to implement for small datasets, where computational resources are not a concern. Additionally, the solution obtained from the normal equations is deterministic—it will always produce the same result for a given dataset, as there is no dependency on hyperparameters like learning rate or initialization values.

However, the normal equations also have some significant limitations. Computing $(\mathbf{X}^T\mathbf{X})^{-1}$ has a time complexity of $O(n^3)$, which can become computationally expensive for datasets with a large number of features (n). Furthermore, if the matrix $\mathbf{X}^T\mathbf{X}$ is not invertible, such as when features are *linearly dependent* or *highly correlated*, the normal equations cannot be directly used. This can be partially addressed by using regularization techniques or computing the **pseudo-inverse**, but it complicates the process. In addition, the memory requirements for storing and manipulating $\mathbf{X}^T\mathbf{X}$ can be prohibitive for high-dimensional data. As a result, for larger datasets, iterative methods like gradient descent are often preferred, as they are more scalable and can handle situations where the normal equations are impractical.

Making use of ready made libraries or packages

Implementing algorithms from scratch is intensive, and may result in poor performing algorithms than those that are already built and packaged, in this paper, we utilize ready made packages to streamline the process, in particular, we use `sklearn` and `statsmodels`.

References

[1] Shah, A. et al., 2023. *Identifying Trades Using Technical Analysis*, Mumbai, Maharashtra, India: VJTI College.

[2] Wikipedia, 2024. *Stock market prediction*. [Online] Available at: `https://en.wikipedia.org/wiki/Stock_market_prediction` [Accessed 19 November 2024].

[3] Reuters, 2024. *South Africa's Capitec Bank reports 36% jump in half-year profit*. [Online] Available at: `https://www.reuters.com/business/finance/south-africas-capitec-bank-reports-36-jump-half-year-profit-2024-10-01` [Accessed 19 November 2024].

[4] Bank, C., n.d. *Business banking*. [Online] Available at: `https://www.capitecbank.co.za/business/` [Accessed 17 November 2024].

[5] Capitec, n.d. *About us*. [Online] Available at: `https://www.capitecbank.co.za/about-us/` [Accessed 18 November 2024].

[6] Yahoo, 2017. *Yahoo Terms of Service*. [Online] Available at: `https://legal.yahoo.com/us/en/yahoo/terms/otos/index.html` [Accessed 18 November 2024].

[7] Analysis, S., 2024. *Capitec Bank Holdings Limited (JSE: CPI)*. [Online] Available at: `https://stockanalysis.com/quote/jse/CPI/` [Accessed 19 November 2024].