Bitcoin candlestick predictions using lagged features and machine learning algorithm in R

Training various machine learning algorithms to predict the next candlestick of the bitcoin price using various lagged features

Sandoche Adittane

2025-04-16

Abstract

This report explores how to get the best accuracy on predicting the next candlestick of the bitcoin chart using the previous ones. It compares different algorithms: Generalized Linear Model, Decision Tree, Random Forest, KNN and Gradient boosting and different number of lagged features. This project is part of the 'Data Science: Capstone' module of HarvardX PH125.9x from the edx platform.

Contents

1 Overview **3**

1.1 Introduction to Bitcoin 3

1.2 What are candlesticks? 3

1.3 Goal of the study 3

1.4 Applications 4

2 Exploratory data analysis **4**

List of Figures

List of Tables

1 Overview

In this study we will try to predict the direction of the next candlestick of the bitcoin chart. Before starting, it's important to understand what are Bitcoin, candlesticks and what is the goal of this study.

1.1 Introduction to Bitcoin

This last years Bitcoin (BTC) has been gaining attention not only by retail investors but also by institutional investor. In 2025 we've seen the emergence of spot Bitcoin exchange-traded funds (ETF) from institutions such as BlackRock, VanEck, Grayscale. With a market capitalization of about 1.68 billion in dollars at the time of writing, Bitcoin started as a peer-to-peer currency, a free alernative to centralized currencies controlled by central banks. It is now used more as an investment, a store of value and even considered as a strategic reserve assets by some countries.

TODO: Add examples with sources.

Bitcoin ows is decentralization and to it's data structure, the blockchain, a chain of block that contains transaction, and to its consensus, the proof of work. Without going too much into details, it makes a Bitcoin a currency that does not rely on a centralized server. Proof of work is a cryptographic competition where the Bitcoin servers called nodes compete to decide which one is the next block to be added to the blockchain. They go through a process called mining where nodes have to use their computing power to find a number called nonce. This computing power is called the hashrate. The node who succeed at "mining" successfully gets rewarded for that.

TODO: reference to my article

The fact that Bitcoin is defined by its codebase is quite facinating, also having all its ledger visible and publically available gives a lot of data available to analyze. Moreover unlike stocks BTC can be traded any time, there is no opening or closing hours, the bitcoin market never stops and it is very easy for anyone to buy and sell bitcoin. Those are two reasons worth studying bitcoin's candlestick charts instead of other asset.

1.2 What are candlesticks?

Let's talk about the candlestick. The price of assets such as bitcoin is described by a serie of candle stick defined by, an opening price, a close price a high and a low also called OHLC. A candlestick can be "up" / "bullish" if closing price is higher than opening price, or "down" / "bearish" otherwise. You can see this visually with the following figure. " "

https://i0.wp.com/techqualitypedia.com/wp-content/uploads/2024/09/candlestick-components.jpg?w=1491&ssl=1 Source: https://techqualitypedia.com/candlestick-patterns-bullish/

The candle stick chart is defined as a time serie of candles, each candle is defined at a defined time and have a time duration. We will explain more in detail in the exploratory analysis.

1.3 Goal of the study

The goal of the study is to find a model able to predict the direction of a candlestick using N previous candles. This number N will be also part of the research. We will have to not only find N but also find what are the best features to achieve the best accuracy.

1.4 Applications

Why is the direction of a candlestick matter? Because being able to predict the direction of the next candle could enable trader to buy and sell on spot market when the predicted candle is green. Also perpetual futures trader can go both way, they can long when the prediction says “up” and “short” when the predictions says “down”.

TODO: Give some resource to learn about spot vs future.

2 Exploratory data analysis

In this section