Comparative Analysis of Semi/Non-Parametric Regression Models on Bitcoin Price Prediction

Aviejay Paul, M.Sc. Computer Science (Data Science)

University of Dublin, Trinity College, 2021

Supervisor: Dr. Bahman Honari

Bitcoin is a cryptocurrency which uses peer-to- peer technology for secure transactions based on block-chain technology. It is one of the most widely accepted crypto-currency in the world today. In this paper, we present a thorough comparative analysis of the performance of different predictive regression models (primarily non-parametric and semi-parametric) on Bitcoin price predictions. The models that we have used to perform the comparative analysis are K-Nearest Neighbours, Random Forest Regressor, Kernel Ridge Regression, Cubic Spline and Gaussian Process Regressor. We have also attempted to check which model is able to accurately predict the sudden rise in the bitcoin price using endogenous features related to price, followed by predicting the sudden slumps. We primarily make use of two columns from the dataset, which are date and closing price (USD). At the same time, we have also introduced two new features in the training data set. These new features are yesterday's Bitcoin price and the difference between yesterday's price and the price day before yesterday. Training of the model has been performed using data from the year 2013 to 2020 and the test set is the data from the year 2021. We find that the Gaussian Process, when used with an appropriate kernel, and Kernel Ridge are the most effective of the studied models while predicting Bitcoin price using strictly endogenous features related to price.

Note: This work focusses on the prediction performance and analysis of different models rather than on optimization strategies used by the different models.

Keywords: Bitcoin, K-Nearest Neighbours, Random Forest Regressor, Kernel Ridge Regression, Smoothing Spline, Gaussian Process Regressor.