

Aaron Amha

Owen Crenshaw

Zach Nguyen

Ben Xiang

CSCI-4502: Data mining

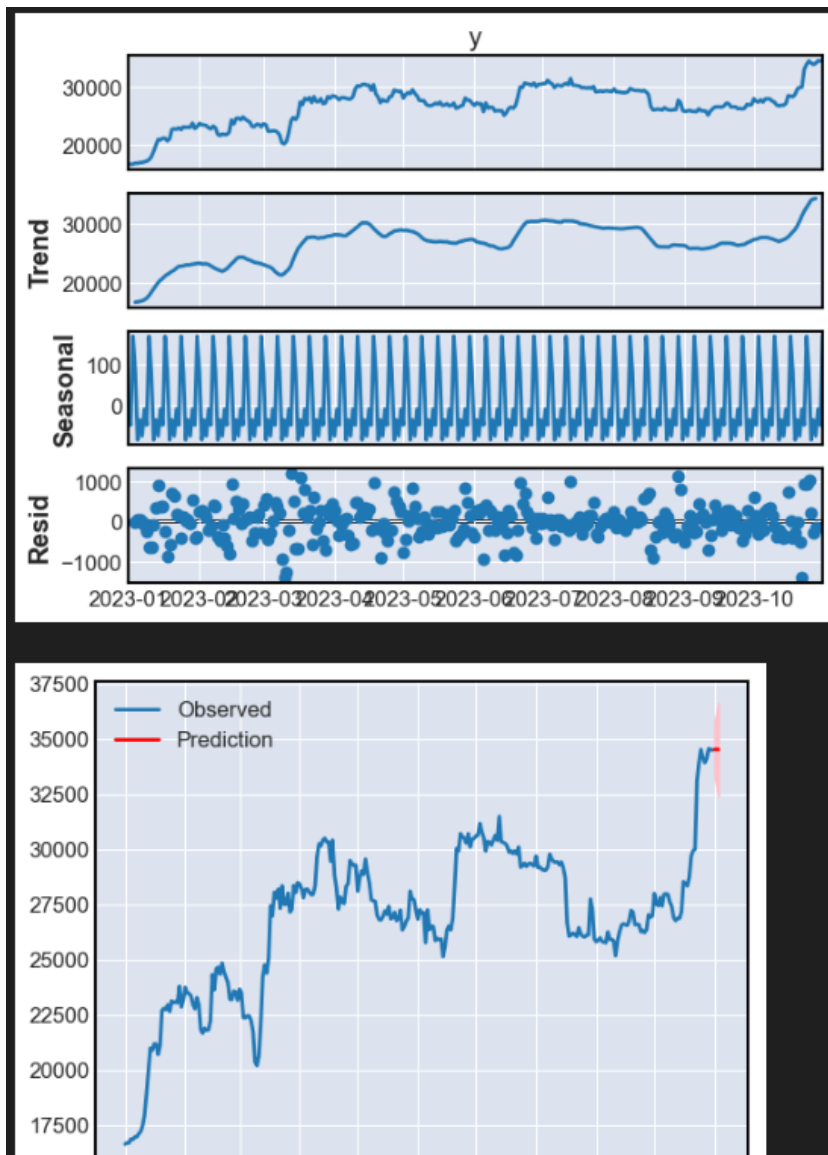
Course Project Update Report

improved, incorporated, or fixed the project proposal

Since our project proposal that we presented in the class, we have made significant changes based on the great feedback we received. You can view the project proposal through this link here: [Project Proposal](#). In the Project Proposal our teammate Aaron Amha has fixed the introduction, literature review, and the proposed work by adding more references. However, this section of the paper is to aim to fix the evaluation portion and explain more in depth of our process that we used or found. First let's address the method for calculating the identified price metric. In our initial project proposal, we did not include how we are going to calculate the the price metric, but since then we have been working on the project together and are currently using RMSE because it is easy to interpret, its predictive accuracy of a model, and it is also widely used for large models like ours. For handling issues like sample size, false positives, and outliers we are still trying to brainstorm about it, but what we have come up with is collecting more data, since we are only using bitcoin data from Jan 1, 2023. And based on the csv file of the bitcoin prices since Jan 1, 2023, the data seemed to be a little too small, thus it might cause a problem. We can also use Bonferroni correction as well, but we will need to figure out how to put that to the test.

work and progress

With our team of talented and motivated individuals, we are hard at working to make this project a reality. Currently, our team has gathered plenty of information about bitcoin prices from Jan 1, 2023 to Oct 29, 2023 and have made significant models and graphs from the data gathered. From there, our team is currently working on a machine learning algorithm to predict future prices of bitcoin. Here are some images:



The first graph is gathering information from the csv and making predictions on the csv information, the second graph is the actual graph that was plotted with a red line to show the prediction over a 3 day period. The machine learning model is not perfect and has flaws that our team is working on fixing before the project due date. In addition, all our work is currently on track, however that might change near the future with midterms, thanksgiving break, etc. coming closer for many of our team members.

challenges you have faced

There were many significant challenges that my teammates had faced. When we initially proposed this project, we did not realize that the api that the kaggle data set was using to grab bitcoin data was blocked in Boulder. When running the api, I kept on getting a restriction error, and upon further investigation we found that the API is blocked by Boulder County. However we were able to use individual parts of the API to gather information, but it was really bothersome. In addition, since all my team members have not ever coded a machine learning algorithm before, we struggle to have a working algorithm that can predict a good prediction model based off of our data that we have collected. The current machine learning model that we have is having some issues where it is only giving us a straight red line (prediction line), no matter how many periods we have it set to. So for now that is the most challenging part of the project we are currently facing and working on to improve,

changes from the original project proposal

So far there is no sign of any significant changes from our original project proposal, we are currently working and learning as much as we can. But if there are any changes that we are making we will be sure to include it on our next project update report.

More detail in your proposed work

In this section, we delve into the core of our project's methodology and models for Bitcoin price prediction. Our methodology centers on the robust principles of time series analysis, utilizing the renowned AutoRegressive Integrated Moving Average (ARIMA) model. To kickstart our process, we meticulously prepare the data by importing historical Bitcoin price records from a CSV file. Subsequently, we employ time series decomposition techniques to disentangle the underlying components, unveiling trends, seasonality, and residuals within the data. Our choice of the ARIMA model, specifically configured as ARIMA(5,1,0), serves as the linchpin for Bitcoin price forecasting. It adeptly incorporates autoregressive, differencing, and moving average elements to craft accurate predictions. These predictions are evaluated rigorously, with the Root Mean Squared Error (RMSE) taking center stage as our assessment metric. RMSE meticulously quantifies the fidelity of our forecasts by assessing the disparities between projected and actual Bitcoin prices. Lower RMSE values are indicative of superior predictive performance. Additionally, our approach encompasses the creation of insightful visualizations that showcase the forecasted values alongside their confidence intervals, offering a holistic view of the model's performance. This multifaceted strategy empowers us to navigate the intricate realm of Bitcoin price prediction with confidence and precision.