COPPER PRICE PREDICTION

CAPSTONE PROJECT PUBLIC SUMMARY

Salga Benjamin

Submitted to

Central European University

Department of Economics and Business

In partial fulfilment of the requirements for the degree of

Master of Science in Business Analytics

Supervisor: Mihály Orsós

Budapest, Hungary

2022

Objective

The capstone project focuses on solving a warehouse capacity problem using only data analysis and prediction. In most cases, when there is a shortage of warehouse capacity (not enough storage space to store raw materials for production in the long term), a building designer is hired and a new warehouse building is designed or a company is found from which the necessary storage space can be rented. The present project has taken a completely new direction and is simply predicting from publicly available data and visualising the results to solve the problem.

Similar prediction projects in 2022 can have complication by different external factors. The war in Ukraine, various fast spreading dangerous viruses from abroad can bring unpredictability to the world market. This makes it difficult for prediction models to work reliably.

There are three participants in the project: an IT company responsible for the implementation (later referred to as the IT company), a manufacturing company where the problem is located (later referred to as the Client company) and the Central European University.

The Problem

The client company receives a large volume of production orders. They do not have their own products, they only produce parts for other companies' products. It takes several months for the company to complete the production of these parts and the orders for these features are paid for by the customer mainly before production. This type of operation is quite unique in the manufacturing sector. The client company prefers and insists on this type of operation, because their production machines are made sure to be used on the long term, so it gives a stability. At the same time, the payment at the beginning of production ensures that the customer does not cancel the order at the end of the lengthy production, or if they cancel, the main part of the expenses have already been paid for. There have been several examples of the former throughout their history.

However, the downside of this operation is that the Client company sells the raw materials at the beginning of production at a given price, but if the price of the raw materials increases over time, they burn their own profit. At the same time, the customer's warehouse is not big enough to store the raw materials needed to cover the entire order (they produce for several customers in parallel).

The Project

The solution to this problem is to develop a prediction algorithm that predicts expected copper prices in the long and short term from historical copper price data and other publicly available data. This data would be visualized in a BI (Business Intelligence) software and delivered to the Client company. From the predicted and the production planning data (expected raw material consumption + x% scrap material), the workers can manage their warehouse capacity much better.

One of their most important raw materials is copper. They use a lot of it and the Client company thinks they are losing the most money on this material due to its rising price. Therefore, to solve this problem, copper price prediction is the aim of the project. The acceptance criteria for the project are not specific, the Client company wants to decide after 1 month of use whether the device is useful for them. This requires that the prediction algorithm is as close as possible to reality, but more importantly that the trends are well tracked by the predicted data. This would be needed, so that the Client is able to change the amount of copper they buy based on the predictions to maximise profits.

The project consists of four parts: data collection, correlation analysis, prediction and visualization with a BI software.

Correlation Analysis

As a first step, we looked at how publicly available data might correlate with copper and had discussions between the Client and the IT company. The results of these discussions were mainly other metals such as steel rebar, silver, gold and others. At the same time, there was also a correlation analysis done between the data previously collected by the IT company (used for previous projects) and the price of copper. For the correlation, the data of the last 100 days was used and for each day, a separate correlation result was obtained. The reason for this is to test the unpredictability caused by war and other external variables. The correlation analysis revealed that copper is associated with many unexpected data. These unexpected relationships came from previously collected data and copper correlation. These are Bitcoin (BTC) and Ethereum (ETH). Both have a very high correlation with copper price. Rebar is the other data with which copper is highly correlated. For all three data, the correlation is over 91%. However, Platinum is another case. There is not such a high correlation with Platinum, but there is an improving trend between the data. So, the two data moves in the same direction together, therefore, taking Platinum into account can help the prediction. The US Dollar Index serves the same purpose, but in the other direction, since it has an inverse correlation with the copper prices. Furthermore, because of the war, this correlation is getting less and less significant. These correlation figures show nicely that before the Ukrainian War, the correlation was -52.5% and since then, it decreased to -33%. So, when the war ends, the correlation should improve and then it could help in forecasting (negative numbers represent the inverse correlation and the closer the number is to 0 the weaker the correlation). The US Dollar Index and the EUR-USD exchange rate have been included in this analysis, because most of the copper mines are located in the US and in the Chile.

Data Collection

The correlation analysis was done after the data collection. The correlation used the data collected from the internet with different python codes. For data mining, the BTC and ETH data did not have to be scraped, because this data was already available on the IT company's server and did not need to be collected. I got permission from the IT company to use the data. Copper, US dollar index and Steel Rebar price data were scraped from the internet, from publicly available websites. These websites are the Macrotrends.net and Investing.com. A python code is used to collect only the new prices, but the historical data was downloaded in a csv file and was imported into the server manually. For the up to date database, the Windows Task Scheduler, Windows' built-in program was used, which helps to run the code every day at an exact time. The rest of the data was downloaded via the Oanda online trading website through an API. These data were EUR-USD, Platinum and Silver data.

The downloaded data is stored in a MySQL database in separate tables. These tables are updated each day by the code with new data. To access the server and store the data, the Client company pays a monthly fee to the IT company.

Prediction

The prediction algorithm uses the following variables: 'COPPER', 'EUR_USD', 'PLATINUM', 'REBAR', 'SILVER', 'USD_INDEX', 'ETH', 'BTC', and historical copper data. The historical copper prices are used to produce 5 new variables. Each variable tests the relationship with 1 less day than the previous one (maximum difference of 5 days). When the data is ready, it is saved to a new table and this table is separated into two data tables for the prediction algorithm:

train (75% of the total data table for training) and test (25% of the total data table for testing). The prediction uses the prediction algorithm created by the developers of the IT company, which uses basic functions from the Facebook fbprophet python package. Prophecy prediction functions are constantly being developed and adjusted by programmers. The python code saves the results in a separate table on the IT company's server, the BI software connects to this server and this table and the visualisation is created from it.

The results of the last four months are not good enough. Unfortunately, the gap between reality and predicted figures is too large. The trends are quite similar but always predict copper prices to be much more expensive than reality. The tool is not usable in the current operation.

Data Visualization

The data visualisation was created in the IT company's proprietary business intelligence software. The first step was to create the connection between BI and the server, then the extracted data was used to create a data table from which the views were created.

Two-line graphs were produced. The first one shows the main metal price movements. This view does not include a prediction, but it helps the employees to have an idea of how the prices are moving and helps them form their own prediction guess, which is needed as a second opinion besides the prediction algorithm. Thus, supporting the flow from both sides (machine and human prediction) is expected to allow for a more accurate operation.

The second graph is the copper price prediction. The whole project aims to produce this graph. The graph shows two lines: the blue one is the actual copper price and red is the predicted copper price. I have previously written about the performance of the last four months, however, if we take into account the period before the Ukrainian war, the prediction is very strong. On the other hand, if we look at the pre-Ukrainian war period, we can see that the system has tracked the trends of change quite accurately and with a very narrow margin of error. These predictions from the past provide the motivation to continue and improve, so that in the near future it can work as the customer expects.

Future of the Project

This project has a lot of potential beyond the copper price prediction. First of all, the prices of the other main base materials could be predicted. This gives a variability over almost the total of the stock. Then, from the production planning data in your ERP system, you can calculate the expected raw material requirements for a given machine on a given day. From this data and the raw material price predictions, the algorithm could automatically make purchase quantity recommendations to the workers.

Conclusion

To summarize the project, a manufacturing company tries solve its storage capacity problem with a system that predicts the price of one of its most important raw materials (copper). This allows the manufacturing company to keep some copper in stock because it can see the possible price changes and optimize the stock accordingly. The project was very interesting and I learned a lot during the implementation, especially about the python language. I am sure that I will be able to use what I learned during the implementation in a wide range of cases. Currently, I have written this code to the best of my knowledge, however, I hope that as time passes, I will have the opportunity to improve it.