Capstone project at BrainStation

Forecasting import and export trends in Canada

Goals of project

Explaining Canada's market to those who don't know this market very well and find the trend for the past few years and forecast 2020 trend and amount.

Dataset of project

All the data sets have gathered from the United Nations website, which is the reference for all the countries' trades.

https://comtrade.un.org/

https://www.trademap.org/

Why is this a good problem / subject area to apply data science techniques?

Predicting is always interesting in every aspect, especially in business. So, every company needs to know about future sales and revenue otherwise they can't plan for the future. Therefore, one of the data scientist's task is to gather the previous data, clean and categorize them for using them to predict a future trend.

How has it been addressed in the past?

Of course, every company has its own predicting model, but I couldn't find any prediction for the big picture of Canada's trade in the websites. So, this capstone has dedicated to this subject. Therefore, there are many time-series models and I had to try some of them and pick the best one.

Details on the source of the data and the dataset itself (including data format, structure and schema, etc.)

At first, datasets were 13 excel files, which includes different types of time periods, different categories, and different countries' trade amount with Canada, both for import and export (6 files for import, 6 for export and one for the market balance). Totally, it was more than 10,000 rows and 400 columns.

But for forecasting, it wasn't necessary to use all of them. Only total amounts for each month was sufficient for predicting the next year's total amount of each month.

A summary of the preprocessing, feature engineering and any other data cleaning/transformation, and exploratory data analysis (EDA) performed and the motivation and reasoning behind it

For EDA and data cleaning in this project, all the columns and files have investigated for selecting the best features and concatenate all the useful data in one file as the default format of time series and trying different models on it.

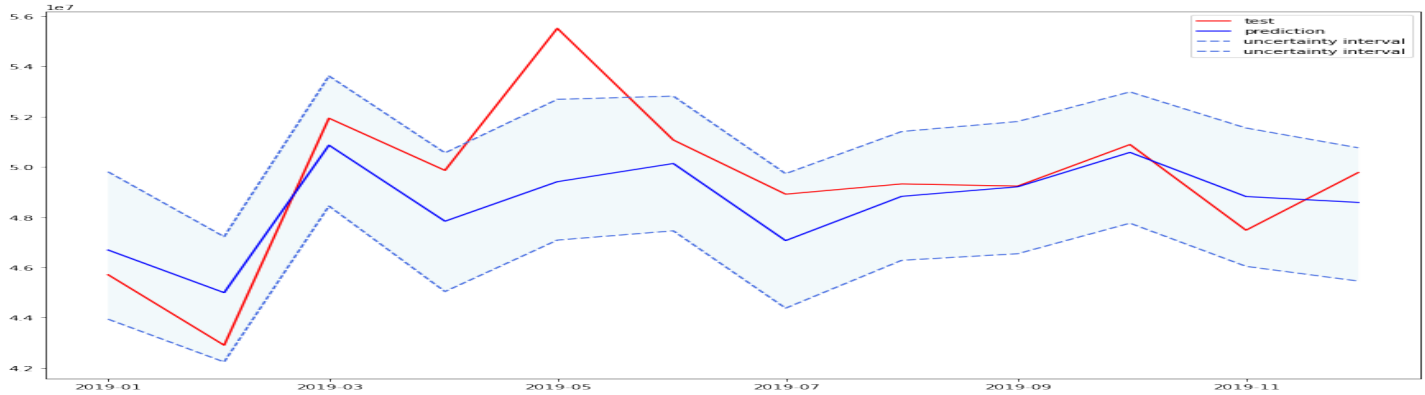
Also, in EDA I found the countries share in Canada's import and export market and using Tableau for making a visualization dashboard. Finally, I investigate the crises impacts on the trend also checked the five top categories and top merchandise for both import and export and plot the graphs to find the trends, turning points, and seasonality for each of them.

A summary of all the modelling completed including the process of model evaluation, selection, and results

For modeling, first of all, I tried the ARIMA model which was explained in the time series' lecture. The prediction for 2019 as the test part wasn't satisfactory. It couldn't illustrate the trend fluctuation very well and almost it was near to a flat line. Also tried Auto Regression (AR) model which is simple than ARIMA, but the result wasn't better than the previous one. Then I decided to try some simple models which usually predict a flat line, such as NAÏVE and Exponential smoothing. The goal for this selection was to find if a smooth line could predict better than the previous fluctuations! However, it wasn't!
While searching the exponential smoothing model, I was found another type of this model which is Holt-winter. It could predict the import and export trend fluctuation better than the others. Also, the RMSE was less than all previous models. Actually, I should mention it's one of the best models for predicting in my dataset (import & export).

Finally, Facebook Prophet! After searching a little about this model, it could draw my attention so much. One of the reasons is that it could make an interval band of prediction which is very important in business. For instance, it can show a range of sales, amount, profit, etc. in the prediction. This ability is very important in the business to know the probability of being in a reasonable range because in business predicting an exact point is almost impossible. So, the Facebook Prophet has two ability: first, shows the exact amount of forecast, and second, shows how much is possible to be in a determined percent up or down of that exact amount.
The trend prediction and RMSE were better than all the others, except for the RMSE of the export dataset in the holt-winter model. But because of the second ability that I mentioned, I preferred to use FBP in my data set for forecasting 2020 and decline the RMSE difference in this case!

Holt-winter model for export data, RMSE= 0.1599



Facebook prophet for export data, RMSE= 0.2153

Findings and conclusions based on all analysis and modelling of the data - how do your results compare against your initial goals & hypotheses?

Finally, I found the trend for 2020 and predict the amount for total import and export for each month till the end of 2020. I could reach my prediction goal and learned so many about time series and forecasting in business which was really enjoyable!

A final summary of the business applications of the project as well as potential next steps and future directions

Forecasting in time series can be applied in every company and it's very important for all projects in different fields. Also, it can be used in all the other fields in which they would like to predict the future according to their past data.
So, the next step would be applying these models for forecasting different categories and merchandise's import, export or sales, as well as predicting any other subjects!

Reza Navidi

March 2020