Capstone project proposal

Urmidola Raye

1. What is the problem you want to solve?

I want to create a time series model to forecast/ predict the export trade volume of energy product (crude oil and natural gas) between the United States and Canada during 2015 based on the historical data that I have till 2014 and then compare them with the 2015 published data and check for any anomaly. If any anomaly is present, I would like to discuss about the factors responsible for such anomaly.

2. Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they would't have otherwise?

The Organization of the Petroleum Exporting Countries (OPEC) is the client. Based on my analysis, the OPEC will get an idea about the estimates of values of crude oil trade from separate estimates of crude oil volume and prices between USA and Canada (not members of the OPEC). It will also facilitate systematic comparison between similar data series published by other agencies and detect historical anomalies as well as any future anomaly and their relation with the economic condition of the export market of energy product. Based on this analysis, the OPEC will also get an idea whether USA and Canada are trying to ensure the stabilization of oil markets in order to secure an efficient, economic and regular supply of crude oil to consumers, a steady income to producers and a fair return on capital for those investing in the crude oil industry and will help them to take necessary actions accordingly.

3. What data are you going to use for this? How will you acquire this data?

The data will be downloaded from http://open.canada.ca/data/. The data on export source agency, origin country, destination country, number of megabarrels exported per day and number of days in a month in which export took place will be considered.

4. In brief, outline your approach to solving this problem (knowing that this might change later).

First, the data set will be cleaned and the missing values will be taken care of.

Second, suitable structure will be chosen to store the data.

Third, forecasting will be done using ARIMA or other time series models and comparing the above mentioned data types for each month of a year to the same month of the next year as well as year to year comparison. These values will then be compared with the published values and the socio-economic reasons behind any such anomaly (if determined) will be searched for.

Lastly, any correlation between drop in oil price with export of crude oil will be searched for to explain the economic scenario of oil market.

5. What are your deliverables? Typically this would include code, along with a paper and/or a slide deck.

The final project will come with code and slide deck.