|  |  |
| --- | --- |
| **Electricity generation Forecast in Canada from 2000 to 2007** | CAPSTONE PROJECT    Vannia Hnatiuk  Brainstation Data Science Winter 2022 Cohort |

# Introduction

Energy is one of the basic human needs. A country needs electricity for manufacturing, economy, and information technology. For this reason, it will be essential to unveil how much electricity Canada will produce and sell to other countries.

**Problem:** What is the cost of consuming electricity? How much will cost to produce electricity? Industry and Businesses must know how much capital to allocate for energy consumption for their production. Equally important, they can determine the cost of their products or services.

Electric Power is recorded monthly over the years. Therefore, a time series was conducted with the intent to generate a forecast on a period of years based on the trends from past decades. In this work, the electricity generation from the time frame 1950-2000 was simulated and then forecasted from 2000 to 2007. Hence, two different models were approached: ARIMA in four different configurations and Facebook Prophet.

# About the data

The .csv file was obtained from the Canada Open Data Portal. This table contains data for the years 1950 - 2007 and is no longer being released. This table contains data described by provinces and territories (15 items: Canada, Newfoundland and Labrador, Prince Edward Island, Nova Scotia; etc.). Additionally, it covers source type such as production by combustion and clean energy. In the Exploratory Data Analysis was filtered on a National Basis and only the variables of interest: Date and Power.

# ExPLORATORY DATA ANALYSIS

During this initial step, it was noted a steady increase throughout the decades in Canada. The first record in January 1950, 4.08 millions of Megawatt hours were produced. Fifty seven years later the electricity production reached 57.33 millions of mwh. Likewise, the human nature consists of to repeat pattern affected by physiological needs or by the weather. The latter was crucial for the demand of electric energy. Consequently, the power generation increases in winter months with the goal to fulfill lightning and heating appliances. Similarly, a small demand in the hottest summer months occurs when people use air conditioning, which consumes high loads of power.

In the purpose to grasp a deep understanding of the Electricity generation, the original pattern was decomposed to obtain the main components that characterize it: the trend, seasonal component, and residual. As a result, the upward trend is confirmed, whereas the seasonal is strongly presented on a monthly basis and the residuals were substantially reduced after a multiplicative decompose. Thus, the relation between these components is given in a multiplicative manner.

In order to generate a forecast: a time series needs to be stationary. Stationarity entails independence of the mean, variance and covariance from time. Accordingly, it was performed two statistic tests to evaluate the stationarity in the time series: Augmented Dickey Fuller test and Kwiatkowski-Phillips-Schmidt-Shin (“KPSS”) test which their null hypothesis need to oppose to ensure stationarity in the system. Then, these tests were auxiliar in the customization of a function to indicate the number of differencing steps that the time series would require to transform into stationary. From this function it was determined that only one differencing process is enough to apply to the time series and set parameter **d** to the SARIMAX model.

Equally important, parameters **p** and **q** were set based on the statistical significance of the lags shown in the Autocorrelation Function and Partial Autocorrelation Function. These parameters were different according of the shape of the time series: differenced, raw data and logarithmic transformed. For a detailed description of the models applied in this project, below it is a table that summarize the configuration of each model and their respective scores.

# Methodology

SARIMAX was set according to SARIMAX(p,d, q) x (P,D,Q)s . Where (p,d,q) is the order of the model for the number of Autoregressive parameters, differences, and Moving Average parameters. While (P,D,Q,s) are the same parameters but with periodicity **s**, set to 12 for monthly data.

On the contrary, Facebook prophet had the seasonality mode set to ‘multiplicative’, yearly\_seasonality to 6. In the same fashion, holidays have a considerable effect in time series, in this project the default Canadian festivities were input to the model setting.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | Configuration | | train  MAE | | test  MAE | | train MAPE | test MAPE | Observations |
| ARIMA seasonal difference | SARIMAX(15,1,8) | 500,860 | | 1,781,418 | | 230.40% | | 309.42% | Highest magnitudes of Train and test in MAE and MAPE. Forecast couldn’t hold the trend, it was reduced to the mean of the forecast. |
| ARIMA on original data | SARIMAX(12,1,6) | 579,058 | | 3,912,764 | | 3.11% | | 8.43% | Simulates well past trends and first 11 months of test data. Later starts drifting up and the wave shrinks. |
| ARIMA with seasonal component | SARIMAX(3,1,1)(2,1,2)[12] | 574,952 | | 1,170,547 | | 2.62% | | 8.13% | Simulates well past trends and first 11 months of test data. The prediction could keep original wave and starts to attenuate steadily. |
| ARIMA with logarithmic transformation | SARIMAX(1,1,0) | 1,529,293 | | 9,446,019 | | 5.69% | | 20.40% | Model with stationary series. Nonetheless, the error metrics were not better. Overfitted. |
| Facebook Prophet | seasonality\_mode ='multiplicative', CAN holidays,  yearly\_seasonality=6 | 784,938 | | 1,998,463 | | 4.16% | | 4.18% | Best model, in comparisson with ARIMA. Generated more robust predictions for more periods ahead. |

**Table1:** *Comparison between different configuration models based on Mean Absolute Error and Mean Percentage Error in training and test dataset.*

After comparing the scores between different models and configuration. It is certain that Facebook Prophet performed the best; specifically, because it could generate solid predictions several years later than only 11 months in the best ARIMA models. Provided that, the forecast (‘yhat’) in 2007 generated by Prophet is going to be used to estimate the sale price of electricity in that year.

# Insights

Based on data obtained through Open Data Canada, in 2007 the price of 1 non residential kwh ranged from 96 cents to 106 cents. For this reason, this unitary cost was multiplied by the power generated in that month. Still, this estimate doesn’t cover the cost of Transmission and Distribution and inflationary rates. In 2007, an average of $ 52 billions of Canadian dollars were sold from the Electric companies(depending on each province) to the industry and commercial sector.

|  |  |  |  |
| --- | --- | --- | --- |
| Date | Power(Kwh) | CAD Dollar | National Cost |
| ene-07 | 61,551,516,021 | $ 0.999 | $ 61,489,964,505 |
| feb-07 | 55,346,578,852 | $ 1.006 | $ 55,678,658,325 |
| mar-07 | 56,597,615,362 | $ 1.017 | $ 57,559,774,823 |
| abr-07 | 50,262,422,077 | $ 0.989 | $ 49,709,535,434 |
| may-07 | 47,734,850,477 | $ 0.961 | $ 45,873,191,308 |
| jun-07 | 46,161,730,911 | $ 0.994 | $ 45,884,760,525 |
| jul-07 | 47,979,067,430 | $ 1.011 | $ 48,506,837,172 |
| ago-07 | 48,864,710,476 | $ 1.063 | $ 51,943,187,237 |
| sep-07 | 46,434,590,468 | $ 1.039 | $ 48,245,539,496 |
| oct-07 | 50,697,092,991 | $ 1.029 | $ 52,167,308,688 |
| nov-07 | 53,887,145,491 | $ 1.015 | $ 54,695,452,673 |
| dic-07 | 59,698,562,849 | $ 1.029 | $ 61,429,821,171 |

**Table2:** *Power has been downscaled from Megawatt to Kilowatt hour. The unitary price of kwh was upscaled from cents to 1 dolar, in order to make a calculation in right magnitudes.*

# Conclusions

For a fast and reliable prediction of energy capacity or demand Facebook Prophet can be applied because it’s training and prediction times takes seconds to execute. Additionally, the error metrics such as Mean Absolute Error and Mean Percentage Error between the training dataset and test dataset were very close where it didn´t overfitted. Moreover, it is capable to make predictions for long term. Above all, any analyst with little domain knowledge can use it without preprocessing and intricate parameter setting.

In Economy terms, Electric Energy generation becomes more relevant regarding the current conflict in Europe. Even though, in Canada, the main energy source is hydro; the sanction of buying oil from Russia, can significantly affect the portion of energy generated by fuel combustion.