Predicting Electricity Demand and Price

Intro

This report was compiled to examine how well a model can predict both the demand and price of electricity in Queensland and South Australia. Data was collected from <https://www.aemo.com.au/Electricity/National-Electricity-Market-NEM/Data-dashboard#aggregated-data>. A model that could predict demand could be used to ensure that enough electricity is available, which could be used to limit blackouts. A model that could predict price could be used to ensure something the consumes a lot of power could be ran when price for power is at a low.

Data

The data collected was price and demand for electricity in Queensland and South Australia. The data found can be downloaded from here <https://www.aemo.com.au/Electricity/National-Electricity-Market-NEM/Data-dashboard#aggregated-data>. Each month was downloaded and inserted into a single csv file.

January for South Australia has a different date representation, so Queensland dates will be used instead of South Australia’s dates.

Plotting the data

A screenshot of a social media post

Description automatically generated

A screenshot of a social media post

Description automatically generated

From the plots, Queensland demand peaks in summer and dips in winter. Meanwhile South Australia’s summer demand in summer is all over the place and during the winter it peaks compared to the previous and upcoming seasons.

The price plots don’t reveal that much.

Table of Means

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Queensland | | South Australia | |
| Summer | Winter | Summer | Winter |
| Demand | 6670 | 6054 | 1370 | 1391 |
| Price | 76 | 77 | 120 | 95 |

Examining the table, it can be seen that in Queensland the price doesn’t change that much, meanwhile in South Australia there is a massive change. Queensland’s demand increases by around a tenth in the summer and South Australia’s demand change isn’t that much.

Queensland also has a much lower price in both seasons compared to South Australia. Queensland considerate higher demand even when multiplying South Australia’s by 2.76 to get to compensate for QLD’s population.

Modelling the data

When using all the data to create a linear regression model the price and demand the root mean square error for demand in Queensland is 511 and 217 for South Australia. While price error is 18 and 44 for QLD and SA.

When using 50% of the data for training and 50% for testing to create a linear model, the error for predicting the demand and price error it comes to a root mean square error of 512 in demand and 19 for price in QLD. Meanwhile for SA demand error is 217 and 45 for price.

This means when using a model to predict data point it can be somewhat effective.

Uses

I was wondering how well the models could predict the price would be under a threshold and what the true positives, true negatives, false positives and false negatives would be.

When training the model, it used a random 50% of the data points then the model was used to predict the remaining points price.

First, I created a QLD model and tested how well it could predict the price was under a 75 threshold.

|  |  |
| --- | --- |
| True Positive | 3477 |
| True Negative | 2283 |
| False Positive | 911 |
| False Negative | 2089 |

This means a linear model is somewhat good at predicting the price in Queensland.

For South Australia model I wanted prices under 90, which created this table

|  |  |
| --- | --- |
| True Positive | 1704 |
| True Negative | 2926 |
| False Positive | 913 |
| False Negative | 3217 |

A linear regression model is worst when predicting South Australia prices.

Both cases had a lot of false negatives which means it predicts the price wouldn’t be under the threshold when it is under the threshold. This means it wouldn’t be running as often and would take a more downtime than needed. This is bad.

More models

I’m really interested in this and was wondering how other models could predict when the price is under a threshold and what their true rates and false rates are.

Queensland

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Algorithm | True Positive | True Negatives | False Positive | False Negative |
| Gaussian process regression (Squared exponential kernel) | 4499 | 2392 | 801 | 1068 |
| SVM (Linear) | 4554 | 1637 | 1557 | 1012 |
| SVM (Gaussian) | 5028 | 2312 | 881 | 539 |
| High dimensional linear regression | 4513 | 1691 | 1503 | 1053 |
| Tree regression | 4087 | 2515 | 679 | 759 |
| Ensemble of regression | 4687 | 2506 | 688 | 879 |

South Australia

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Algorithm | True Positive | True Negatives | False Positive | False Negative |
| Gaussian process regression (Squared exponential kernel) | 1500 | 4765 | 832 | 1663 |
| SVM (Linear) | 796 | 4941 | 656 | 2367 |
| SVM (Gaussian) | 1524 | 5230 | 367 | 1639 |
| High dimensional linear regression | 798 | 4931 | 666 | 2365 |
| Tree regression | 2240 | 4893 | 704 | 923 |
| Ensemble of regression | 1877 | 3950 | 1647 | 1286 |

Useful

Unlikely to get a random number of data points throughout the year and then must predict the rest. I retrieved data from 2017 and 2016 and created a model with all points from them and then used to predict the values for 2018.

Queensland

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Algorithm | True Positive | True Negatives | False Positive | False Negative |
| Linear Regression | 4415 | 4308 | 2164 | 6633 |
| Gaussian process regression (Squared exponential kernel) | 7599 | 3935 | 2537 | 3449 |
| SVM (Linear) | 9482 | 2008 | 4464 | 1566 |
| SVM (Gaussian) | 8821 | 3208 | 3263 | 2228 |
| High dimensional linear regression | 9459 | 2210 | 4362 | 1589 |
| Tree regression | 8437 | 3344 | 3128 | 2611 |
| Ensemble of regression | 7538 | 3674 | 2798 | 3510 |

South Australia

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Algorithm | True Positive | True Negatives | False Positive | False Negative |
| Linear Regression | 2331 | 8481 | 2689 | 4019 |
| Gaussian process regression (Squared exponential kernel) | 3807 | 7128 | 4043 | 2542 |
| SVM (Linear) | 5247 | 4574 | 6597 | 1102 |
| SVM (Gaussian) | 4386 | 5835 | 5336 | 1963 |
| High dimensional linear regression | 5271 | 4689 | 6483 | 1077 |
| Tree regression | 4164 | 6168 | 5002 | 2186 |
| Ensemble of regression | 3840 | 6654 | 4518 | 2508 |