# Assignment Two

## Topic : Electricity Demand Forecasting

**Group Member**
Zhipeng Liu 12764268
Qingyuan Li 13016389

# Code Space

↳ 66 cells hidden

# Github Link

# Introduction

Since the liberalization of the energy market introduced to Australia in late 1990, the power supply market has become more dynamic compared to its past.The dynamic caused an increase in complexity for different parties involved in the electricity supply activity.

For the Australian government, it's role is to maintain the stability of the market transactions and power supply in the long run. For electricity suppliers and retailers, they aim to maximize profits from electricity generating, transmitting, and distribution processes.

There are lots of factors involved in their decision-making strategies. But one of the most important factors is how to predict the electricity demands accurately given a period.

If the government can forecast electricity demands in a specific region correctly, it might reduce the chance of power outages by wisely schedule electricity generators. It protects economic activities.It also helps the government to achieve a dynamic balance between supply and demand.

For electricity generators,there isn't an economical way to store electricity once be generated.Both overproduction and under production could lead to financial losses.They could reschedule their production strategy based on the predicted demands. For power transmission companies and retailers, they can use this information to reallocate their resources to maximize the profits.

Given the information above, we decided to conduct a project to build a model for electricity demand forecasting.

# Exploration

## Dataset

The dataset we used was downloaded from the website of the Australian Energy Market Operator, which is an organization that manages the national electricity market. The dataset consists of 18-month electricity demand trend curve, which is 26208 lines of data. Each row indicates how many electricity has been used for 30 mins in a specific region in Australia with columns describing 'PRICE', 'SETTLEMENTDATE', 'TOTALEDEMAND', 'REGION' 'RRP'. In this dataset, we don't need as many columns as this dataset provides to us, we will process our data later on in the next section.

## Data pre-processing

We have stored our data in the Google drive; so we could pull our data and use it directly in the Google cloud platform to run our code. As we introduced before, our dataset consists of few columns, but the only two we need are our TOTALDEMAND and SETTLEMENTDATE, which means the amount of electricity used at that time point, as we are going to predict the amount of electricity demand for the next moment, and the moment could be quiet flexible to defined by ourselves.

### drop unneeded columns

```
data_18.drop(['REGION','RRP','PERIODTYPE'], axis=1, inplace=True)
data_19.drop(['REGION','RRP','PERIODTYPE'], axis=1, inplace=True)
```

### split the data

```
split_date = pd.Timestamp('2018-10-01 00:00:00')
split_date2 = pd.Timestamp('2018-10-01 00:30:00')
train = data_18.loc[:split_date]
test = data_18.loc[split_date2:]
```

### scale the data into small range

Because our data type range is enormous in the beginning, for example, the date type, this would make our operation and calculation slower. To optimize this thing, we use a scaler to scale all of our data in to the range of -1 to 1.

```
scaler = MinMaxScaler(feature_range=(-1, 1))
train_scaler = scaler.fit_transform(train)
test_scaler = scaler.transform(test)
```

# Methodology

The problem we are trying to solve is to predict electricity demand given historical data. It is a time series analysis problem. And we only used one variable to predict the electricity demand. It narrows our problem to univariate time series forecasting.

At the beginning stage of model selection, we choose a traditional method called ARMA. This method combined two techniques which are autoregression and moving average. We put a lot of effort on data exploration when we tried to build ARMA. But we failed to make it work properly. After that, we turned our focus on machine learning method. It leads us to two models which are Neural Network and Long Short

Term Memory(Deep learning model). And both models performed good results. In this section, we'll briefly introduce the methods we used for electricity demand forecasting.

### Statistical Method

In the beginning, we performed different size of rolling windows to plot the rolling mean and standard deviation. We used four different sizes of window. The result shows, with the increase of the window size, both rolling mean and standard deviation become smooth. We used the Augmented Dickey-Fuller test to prove the time series we got is stationary, which mean we could conduct time series analysis. And then, we introduced an ARMA model from a library called StatsModels. The model got a good result on in-sample data( RMSE = 109.7). But when it comes to predicting out-of-sample, the model failed to work correctly. So, we decided to use the neural network to perform the task.

### Machine Learning Method

We built three different neural networks. They are very similar. The main difference is we used a different number of layers and nodes in hidden layers.

We selected Adam as the optimization algorithm to minimize our cost function and mean squared error to update the parameters of the model. From reults, we found the model's performance didn't see improvements, even the number of learnable parameters increased. We thought it is because of the model has already achieved a high fitting( r2_score > 0.96 ). The historical demands(training data) could well explain the output of the model (predicted data).

## Evaluation

For ARMA model, we used root mean square error to evaluate the model's performance. It represents how far the data point away from the regression line.

$$\sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(y_{\text{test}}^{(i)}-\hat{y}_{\text{test}}^{(i)}\right)^2}=\sqrt{MSE\text{test}}=RMSE_{\text{test}}$$

For neural network and LSTM models, we used R-squared measure since it is good at predicting continuous variables.

$$\mathrm{R}^2=1-\frac{\sum_{i=1}(\hat{y}^{(i)}-y^{(i)})^2}{\sum_{i=1}(\bar{y}-y^{(i)})^2}$$

## Conclusion

We use the previous time point and the next time point to train our model, each point represents 30mins of electricity use. So the model is tailored to predict the next 30mins demand, which is not in line with our common expectation. To predict the next day of demand, we try to make our training process as the previous day to the next day mapping, in this case, we would get our next day prediction. Unfortunately, we only loaded 18 months' data, which means we only have approximately540 samples. It is completely not enough for our training process.

In addition to that, we came up with another method to predict a whole day demand. To achieve that, we simply just predict our prediction again, because our model is designed to predict the next 30 mins, by predicting the result of the previous prediction again, we might be able to get a certain amount of period by doing this repetitively. This is our first thoughts, it turns out we failed, because, if we predict the previous prediction again, it will extend the previous trend irreversibly. As a result, one solution to improving our result is to apply the first method and train our model as much as we can at the same time.

# Ethical

We selected two machine learning models to predict electricity demands. The first metol is the artificial neural network. The second model is For our LSTM model; there are many situations in which it can be applied. For example, in the stock market, we could use this model to estimate the price for the next moment. But the stock market is a game; if a strategy can take effect at the moment, it will be removed quickly in a short period. As a consequence, some people probably would use this model to estimate price and end up losing big money, or earning money; either way, it is not a good practice in our real life.

For neural network model, we may learn something wrong, and a good example is that when it comes to advertisement, it may show some ads to a specific group of users, this may lead to prejudice and discrimination. Another ethical issue that might happen to us is a moral actor; for example, an autonomous car at some emergencies may take choice for us as a moral actor. It's hard to judge whose responsibility it is. Nevertheless, machine learning is our next megatrend; we shall take more social issues into account as we are progressing. In that way, it is not only for our technology progression, but also will benefit our generations.