THE NEXT RECESSION

An approaching doom?

OCTUBER 18, 2019

AKANSHA GUPTA
29460069

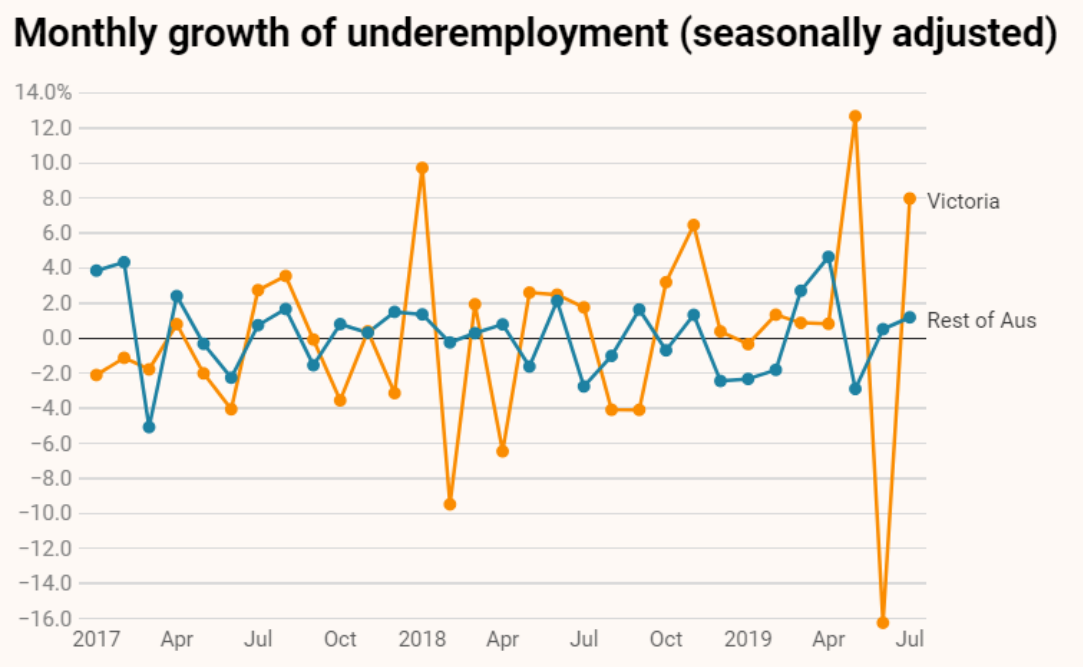
PROJECT DESCRIPTION

THE BACKGROUND

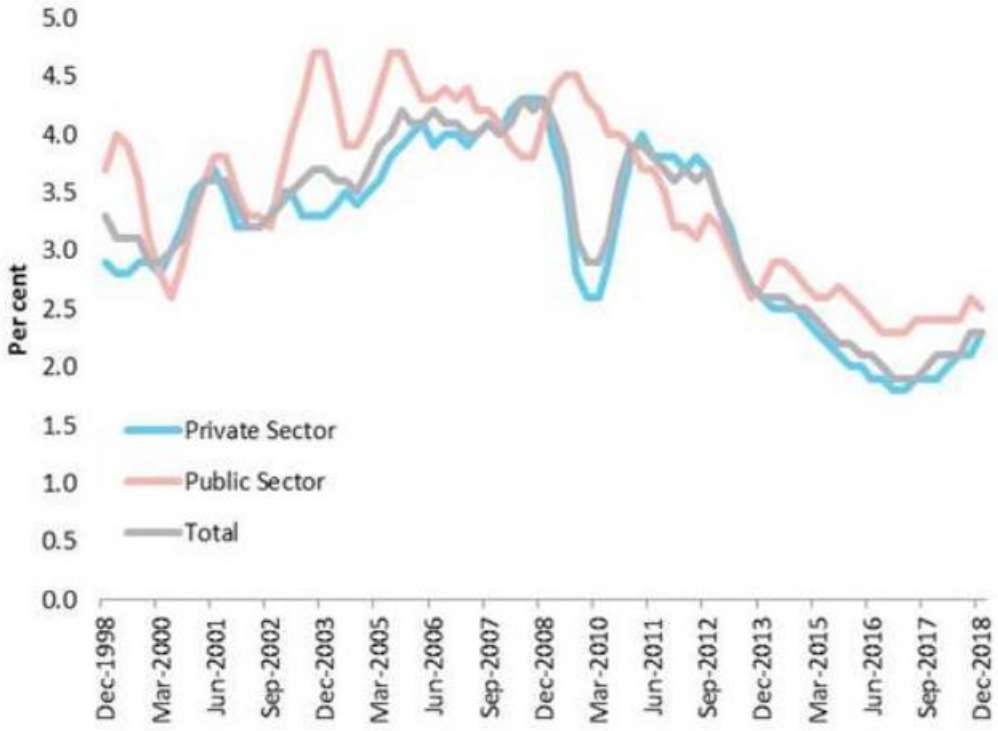
For more than two centuries, Australia has not endured the ignominy of recession; not since the notorious "depression, we had to have" during the early 1990s. As good management coupled with excellent fortune enabled us to escape financial bullets otherwise it might have led us to a lengthy downturn. The recent figures and studies are dragging us towards the upcoming economy that shows some indications that Australia is moving towards a recession So, why we had never considered such issues?

PROJECT DESCRIPTION

The slowdown in Australia's construction industries has intensified the occurrence of recession across all the sectors. Moreover, overall renovation work decreased by 3.8% in three months, making a weak market prediction with a 1% decrease. (Letts, 2019). As the growing economy is directly proportional to the spending of the consumer. Which can only occur if more people are earning or being employed? But the below graph shows us the increasing underemployment rate in different sections of Australia.



Apart from that, the reports released by the Parliament of Australia which mentions a decline in the average wage rate of both public and private sector. The Australian Bureau of Statistics (ABS) along with Wage Price Index (WPI), measured the wage growth at 2.3% per annum (see Figure 2) as well as averaged development of 2.2% per annum over the five years up to December 2018. Using this measure, this contrasts with an average annual salary development of 3.3% over the past five years to December 2013. (Gilfillan, G)



Wage growth in the public and private sector 1998-2018

This implies that the industry believes that interest rates will be smaller in five years ' time than in two years ' time, and this is usually a sign of slowing the economy and upcoming recession. Hence, this project leads us towards the indicators that point towards the recession after the **data wrangling** and **analysis**.

Data accommodation and processes from different sources by data engineer for collecting knowledge.

- Expert knowledge and predictive analysis of given data by data **researcher** as well as **statisticians** will visualize the heatmap and corrplot generated**.**
- **Business** and **System Analyst** improves the system by studying current practices, designing modification, recommends control by identifying the pace of occurring problem and approach required to solve those problems by using ROC plot and come up to predict the model.
- **Financial Analyst** determines the cost of operations by establishing standard costs; collecting operational data. Improves financial status by analysing results. These data scientist helps to analyse the finance market and interest rate of banks.
- **Economists** often study historical trends and use them to make forecasts. They research and analyse data using a variety of software programs, including spreadsheets, statistical analysis, and database management programs. To study the past recession for analysing any specific pattern which might resembles the current scenario.

All the above data scientist can help in proactively collecting the information from various sources and analyses it for a better understanding of how the business operates as well as for the creation of AI tools that can automate certain processes within the organization.

BUSINESS MODELLING:

As recession is calculated by two successive quarterly periods of falling GDP.(O'Brien, 2018) which is evaluated by the Australian Bureau of Statistics (ABS). After the data is been collected from different sources, which then further involves a good amount of data wrangling in the process. After which it is analysed and presented as reports on both websites respectively i.e. by the Australian Bureau of Statistic as well as National Bureau of Economic Research. This information helps spreads awareness among people as almost every sector gets affected its impact. So, who gets affected?

In the below section, I have covered how the knowledge of upcoming economic downturn can benefit different stakeholders by tightening their belts.

Government: Our government leaders can concentrate more on controlling its increasing rate than battling with it. This is done by limiting expenses, resisting tax cuts, and reining in growth in the money supply

Large business: It can save their industries from a decline in sales revenue as well as profit, which can lead to shorting of staff and freezing the hiring procedure.

This includes changes in the **manufacturing** procedure, by pausing the new products rollouts as well as decreasing expenditure on **marketing** and advertising.

Stock Market**:** As the falling revenue directly affects the **stocks market**, which causes **dividends** to suffer or entirely vanish. This information can further cause to an institutional investor, to reinvest for the better stocks.

Central Bank: Will alert bank before slower revenue drop of the valuation of corporation's debt, finances or its credit rating.

Employees: Aware of the **labour** market before the sales decline, the manufacturer will cutbacks the hiring which can cause more work for a fewer **employee** with longer working hours.

Manufacturer**: C**an help the **manufacturer** to produce the product with lower quantity affecting the loyalty of **consumer** towards the brand.

Consumer: It can prepare the **organisation** to spend less on advertising and marketing, it equally declines the customer confidence and spending on the product.

Smaller business*:* Without significant money reserves and big capital assets as collateral, along with greater difficulty in obtaining extra funding, so smaller companies will be prepared before it comes.

As per our model and its benefitted stakeholders, instead of buying forever, one can lease software on a subscription basis. This facility of subscription will make be easier for the stakeholders as instead of paying for a collective upfront software you are charged on monthly, quarterly and annually basis.

This project can alert the above stakeholder to be aware and to identify an upcoming business trend.

Challenges

- The collection of raw data of GDP rate of Australia.
- Cleaning the data for insights and building a model from it.
- The rate of which data is growing with every passing day.
- Requires to be a domain expert by blending my skills with its background.
- When there is a huge set of data, the possibility of privacy and security comes with it.
- **Data wrangling** for inconsistent and unstructured data.
- Identifying the better model for data analysis.
- Refracting code along with validating resources and changes to be documented with version control.

- As the GDP rate is quarterly updated by the Australian Government, our predictive model will be run quarterly and will provide the output four times a year.

Values

- Triggers respondent companies to be aware of the future economic trend that can help their industries and the market to create new business strategies
- Helping the organisation who is planning to operate overseas.
- Government to implement a fiscal policy that involves tax-cutting and increase in spending.
- Australian public finances to decide and put constraint on public sector.
- Preparing smaller and larger business to make a primary-defensive move than their rivals.
- Alarming labour market to create a financial cushion.
- To make the government aware of the upcoming challenges & building new strategies based on that.
- Focusing larger organization to aim low by keeping both innovation and cost-cutting incremental.
- Make companies diversify the products from a different manufacturer.
- Central Bank to lower the interest rate for economic recovery.

DATA MODELLING AND DATA SOURCE

This section will look at the types of data our model will use that will help to predict the upcoming recessions in Australia. Along with hardware and software resource required to run the finished product. The data will require for predicting the upcoming recession is as below: -

DATA CHARACTERISTIC

Data Source	Historical GDP rate and other housing market value related data are centralised and widely available on data.gov.au. While data for currency devaluation can be retrieved from Currency Devaluation document
Data Type	Mainly numerical. Can be obtained in different format e.g. .csv, .txt, .pdf, .flat files etc. These data might be available in both a structured and semi-structured format.
Volume	Data available in both data.gov.au or are in terabyte which is reasonable as we are calculating the net rate of economic growth in Australia. The maximum sample size of files will be around 3000KB. So, we can download the file or as a series split into 57 smaller WARC file, or even a sample file.
Velocity	Though data is been uploaded and updated over data.gov.au and data.oecd.org every quarter for the purpose of development, research or real-time analysis. But for our modelling, we will be needing data to be continuously updated to give the market a responsive alert for any upcoming doom.
Variety	Multiple data sources are required for whole analysis like GDP rate of Australia, Housing Market, the Borrowing capacity of the economy, Currency devaluation, unemployment rate from different data sources.
Variability	Not all documents and data sources will be updated and may improve over time. Quality of data available, that might be structured or unstructured.
Veracity	Possible issues include: - Data can be of varying quality, so the available columns might not be accurate. The correct interpretation of .csv file available by the Australian Government over data.gov.au

	Wrangling is required before the interpretation of data in the model.
Storage	Data available by the Australian Government and data.oecd.org. is in a structured and semi-structured format. Such a huge volume of data from different data sources needed to be accumulated and stored in a single database i.e. My SQL Server.
Processing	When we are dealing with a huge dataset, we need to process it all together after fetching the data from different sources and cleaning it on **Parabola** platform data or outliers due to error in manual entering of data as there might be a chance of missing, which all together in a single .csv file. After this, we can easily import the file over our jupyter notebook such to implement our code for prediction analysis. Batch Processing for model building. Batch loading and processing of the initial files Interactive processing for visualising and evaluating models Once analysed, the high volume of new data available every quarterly which will be needed for visualizing and wrangling and will be stored to come up with a conclusion.

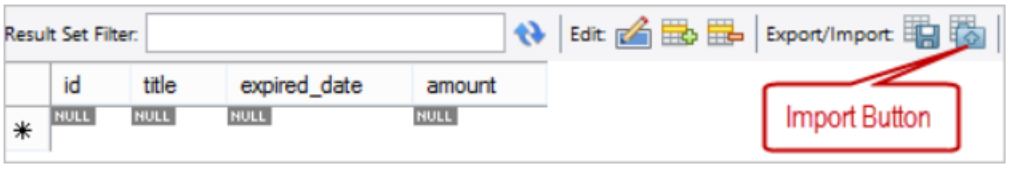
DATA RESOURCE

Based on the collective data downloaded from different sources like monthly GDP and unemployment rate of Australia can be fetched from the Australian Government website, although these data are updated quarterly. Apart from that housing value, interest rate and currency devaluation can be obtained from data.oecd.org.

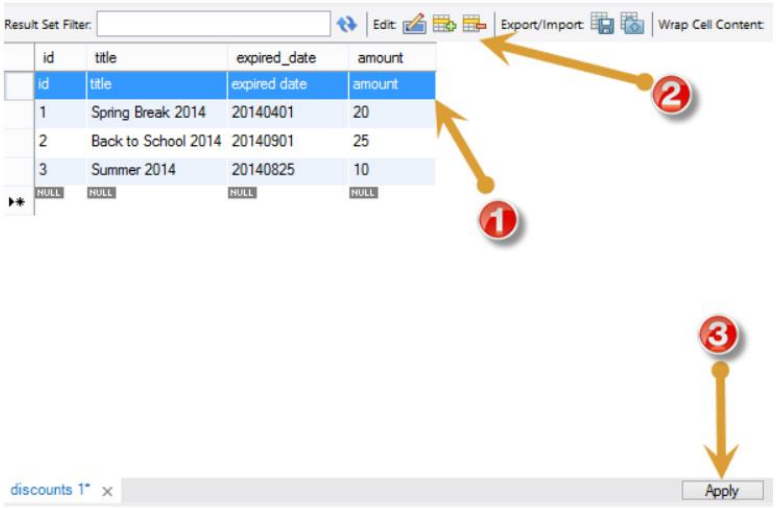
As per our data, we might have a high volume of data for which we may require **Parabola** Server which can create channels for your data that will eliminate duplicates, merge, break, match, apply logic, extract, validate, and more. It's a sandbox atmosphere, so your potential uses are limitless.

Once you've set up a stream as you like, you can have it run on a schedule automatically if you want to. After this, we can export the data and store it in **My SQL Server** which has the capacity of 10 million rows. Apart from that, it has a file and inserts privileges as well as a tool to import data into a table which also allows editing data before making changes.

Open table to which the data is loaded.



Once we are done making changes, we do apply the following method



After the data is been properly stored, we can begin data processing for further analysis of our data over the modelling tool like R studio or Jupyter notebook. The reason for choosing R or Jupyter notebook is as we are predicting the upcoming recession, we will be using R packages like

data sets- for building a regression model.

ggplot2 – this tool helps to build a plot on data visualization

GGally – an extension of ggplot2 and can be useful in creating a plot matrix as a part of initial exploratory data visualization

scatterplot3d – for complex linear regression as well as for multiple predictors.

DATA ANALYSIS

In this model, when we retrieve the input from different data sources, we can initiate our analysis for the prediction of upcoming recession in Australia.

What we need

The Anaconda toolkit and R studio

What we going to do

We are going to build a machine learning pipeline to predict an upcoming recession in Australia. We will work on the dataset after 2000. This includes the historic changes observed in terms of monetary policy design, execution and performance of recently available data, the 2-quarter decline in GDP rate, housing value, the unemployment rate of Australia. Once we receive a structured data , these will be indexed for the predictive recession analysis.

To achieve this, we will follow the below steps: -

- Now, we need to merge the data from all the sources and create a dataset with 6 specific columns which will help us to determine our analysis.

These columns will be as below in the dataset (.csv)

Month_year	GDP rate(trillion)	Housing value (average %)	Unemployment rate (%)	Interest rate (%)	Currency Devaluation (%)
2019M01	2.4	22	6.1	.75	26

Each column in the dataset makes an impact on the model. Once we collect our model, we can apply it to our machine learning tool “logistic regression”. The reason for choosing this model is that it’s relatively simple, uni-variate in nature and market rhetoric implies a linear decision boundary between curve inversion formation and recession.

We need adjustment before proceeding with the data.

Firstly, we need to reset the dataset from quarterly terms to monthly and store it in a column as a primary key for the whole dataset. We can also assume within recession quarters growth of all monthly data is negative.

Secondly, for predictive analysis of two to ten years, we need to shift the data one quarterback to get analysis based on its historic data

Processing

Once data is being analysed stored and cleaned, we must check the correlation between every column in the dataset by using corrplot by using below **libraries**

```
from string import ascii_letters
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

For computing correlation matrix

```
sns.set(style="white")

# Generate a large random dataset
rs = np.random.RandomState(33)
d = pd.DataFrame(data=rs.normal(size=(100, 26)),
                 columns=list(ascii_letters[26:]))

# Compute the correlation matrix
corr = d.corr()
```

For creating a heatmap: -

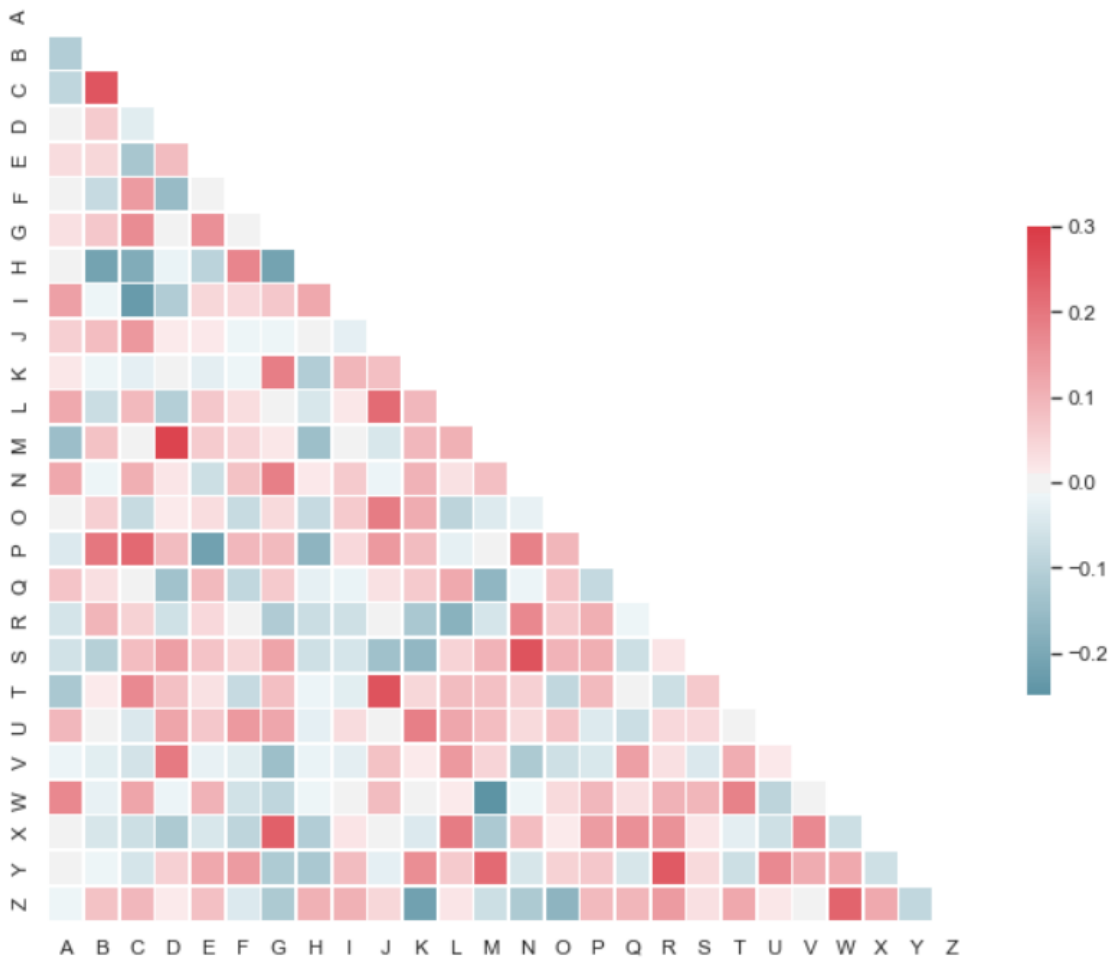
```
# Generate a mask for the upper triangle
mask = np.zeros_like(corr, dtype=np.bool)
mask[np.triu_indices_from(mask)] = True

# Set up the matplotlib figure
f, ax = plt.subplots(figsize=(11, 9))

# Generate a custom diverging colormap
cmap = sns.diverging_palette(220, 10, as_cmap=True)

# Draw the heatmap with the mask and correct aspect ratio
sns.heatmap(corr, mask=mask, cmap=cmap, vmax=.3, center=0,
            square=True, linewidths=.5, cbar_kws={"shrink": .5})
```

Output: -



Based on the output of above, we can observe a correlation between the attributes by the strength of its colour and its range between -1 to 1. If predictor variables have a higher correlation among

themselves then one of the predictors must be included in the model as including both in the model will lead to biasness.

As the predictor variables are categorical, the dependent variable is binary that have value as 1(success) or 0 (failure) and predicts P(Y=1) as a function of X. We will be using below libraries and packages for our Logistic Model.

```
import pandas as pd
import numpy as np
from sklearn import preprocessing
import matplotlib.pyplot as plt
plt.rc("font", size=14)
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
import seaborn as sns
sns.set(style="white")
sns.set(style="whitegrid", color_codes=True)
```

Input variables:-

Month_year- Primary variable with the value of year-month like 2018M01.(alphanumeric)

GDP rate- GDP rate of Australia for each month. (numeric)

Housing value- Rate of housing value for each month. (numeric)

Unemployment- Rate of employment of labour market for each month. (numeric)

Currency Devaluation- Rate of which the currency rate falls or increases. (numeric)

Predict variables (desired target)

y- did the predictor variables causes a recession (binary: 1 and 0)

Logistic Regression Model Fitting

```
from sklearn.linear_model import LogisticRegression
from sklearn import metrics

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=0)
logreg = LogisticRegression()
logreg.fit(X_train, y_train)
```

We need to randomize the data by splitting them into training (70%) and test set (30%).

Predicting the results by calculating the accuracy of the test set

```
y_pred = logreg.predict(X_test)
print('Accuracy of logistic regression classifier on test set: {:.2f}'.format(logreg.score(X_test, y_tes
t)))
```

```
Accuracy of logistic regression classifier on test set: 0.74
```

Confusion Matrix

```
from sklearn.metrics import confusion_matrix
confusion_matrix = confusion_matrix(y_test, y_pred)
print(confusion_matrix)
```

```
from sklearn.metrics import classification_report
print(classification_report(y_test, y_pred))
```

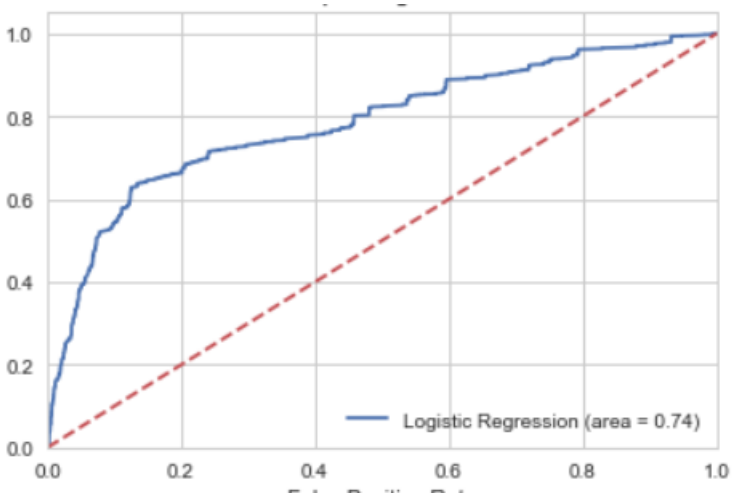
```
             precision    recall  f1-score   support

          0       0.71      0.80      0.75      7666
          1       0.77      0.67      0.72      7675

avg / total       0.74      0.74      0.74     15341
```

Interpretation

```
from sklearn.metrics import roc_auc_score
from sklearn.metrics import roc_curve
logit_roc_auc = roc_auc_score(y_test, logreg.predict(X_test))
fpr, tpr, thresholds = roc_curve(y_test, logreg.predict_proba(X_test)[:,1])
plt.figure()
plt.plot(fpr, tpr, label='Logistic Regression (area = %0.2f)' % logit_roc_auc)
plt.plot([0, 1], [0, 1],'r--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
```



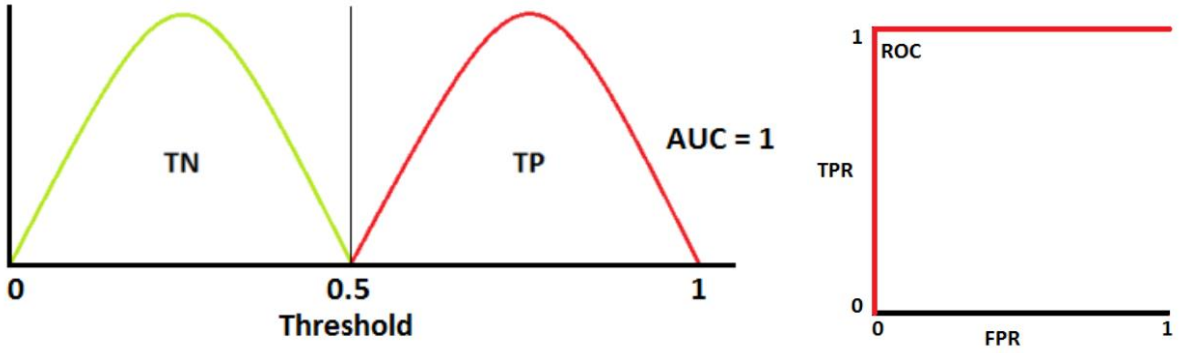
This is another common method used with binary classifiers is the receiver operating characteristic (ROC) curve. The dotted line represents a purely random classifier's ROC curve; a strong classifier stays as far away as possible from that line (to the top-left corner). It has a plot of the false positive rate (x-axis) versus the true positive rate (y-axis) for several different candidate threshold values between 0.0 and 1.0

After performing the above interpretation for our model, we can come on a conclusion whether the recession will occur in the next quarter or not.

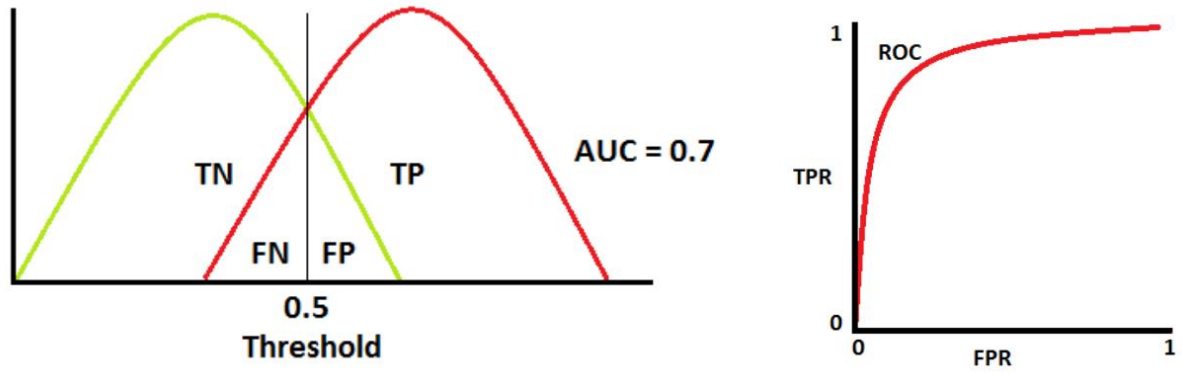
Output

As per the analysis of ROC curve which requires AUROC to range from 0.5 to 1.0. hence, the stakeholder will view the value in a given range to interpret the chances of the occurrence of recession.

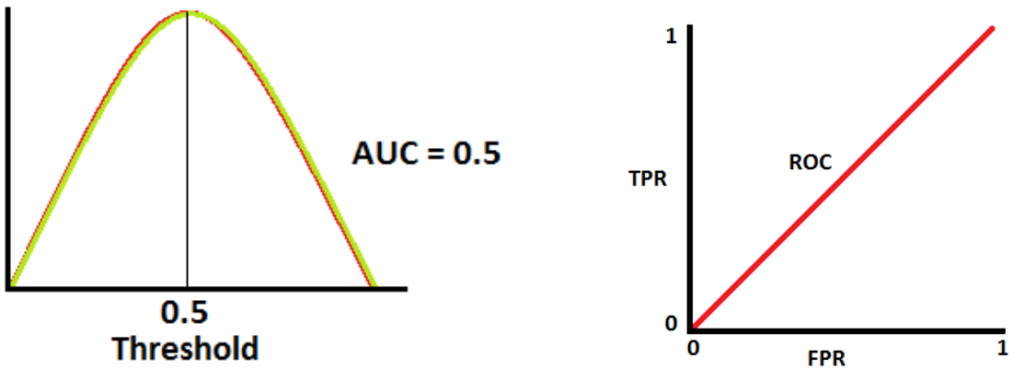
Case 1: If the model AUC value is 1; then it is said to be an ideal model and we can clearly predict that chances of a recession in the upcoming quarter.



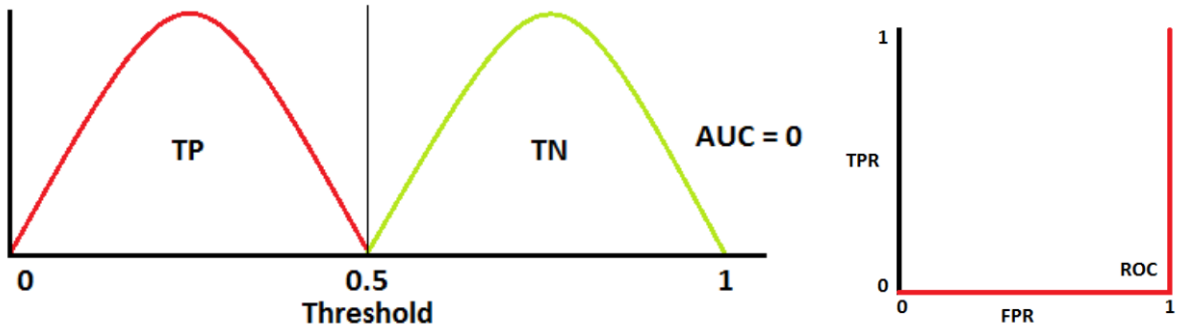
Case 2: If the model AUC value is 0.7 and there is an overlapping between two distribution; then there is 70% of the chances of prediction by the model.



Case 3: If the model AUC value is 0.5 and the distribution is fully overlapped; then there is no chance of prediction by the model.



Case 3: If the model AUC value is 0 and the distribution is not overlapped; then there is a chance of classes been reciprocated.



Thus, if the value of AUC is below 0.5, all your forecasts will have to be reversed (make zeroes into zeroes). That should correct the curve and be above 0.5 for your AUC.

References

Alesina, Alberto, and Francesco Giavazzi. *Fiscal Policy after the Financial Crisis (National Bureau of Economic Research Conference Report)*. Chicago: University Of Chicago Press, 2013

Hall, M. (2019, 7 14). *Unemployment and Recession—What's the Relation?* Retrieved from Investopedia: https://www.investopedia.com/ask/answers/032515/why-does-unemployment-tend-rise-during-recession.asp

Gilfillan, G. (2019). *The extent and causes of the wage growth slowdown in Australia – Parliament of Australia*. [online] Aph.gov.au. Available at: https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/rp/rp1819/WageSlowdown [Accessed 18 Sep. 2019].

Letts, S. (2019, 8 28). *Construction recession deepens and spreads across all sectors*. Retrieved from ABC: https://www.abc.net.au/news/2019-08-28/construction-work-done-q2-2019/11455886

Li, S. (2019, February 27). Building A Logistic Regression in Python, Step by Step. Retrieved from https://towardsdatascience.com/building-a-logistic-regression-in-python-step-by-step-becd4d56c9c8.

Narkhede, S. (2019). *Understanding AUC - ROC Curve*. [online] Medium. Available at: https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5 [Accessed 16 Oct. 2019].

O'Brien, G. (2018, Dec 10). *27 years and counting since Australia's last recession*. Retrieved from Parliament of Australia: https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/BriefingBook46p/LastRecession

Pettinger, T. (2019). *What happens in a recession? | Economics Help*. [online] Economicshelp.org. Available at: https://www.economicshelp.org/blog/351/business/what-happens-in-a-recession/ [Accessed 19 Sep. 2019].

Plotting a diagonal correlation matrix¶. (n.d.). Retrieved from https://seaborn.pydata.org/examples/many_pairwise_correlations.html.

Sakbani, M. (2010), "The global recession", *Studies in Economics and Finance*, Vol. 27 No. 2, pp. 91-109. https://doi.org/10.1108/10867371011048607