

KULLIYYAH OF INFORMATION AND COMMUNICATION TECHNOLOGY

CSC2301-INTELLIGENT SYSTEMS

**Suicide Rates Prediction Using Regression**

|  |  |
| --- | --- |
| AHMAD ZAIDAN BIN ADNAN | 1718733 |
| MUHAMMAD AMMAR ZAKWAN BIN AMRAN | 1712711 |
| MUHAMMAD HARITH BIN KHAIROLANUAR | 1816349 |

# 

# SEMESTER 1 2019/2020

LECTURER: SURIANI BT. SULAIMAN

# 

# Table of Contents

1. Introduction
2. Problem Statement
3. Objectives
4. Proposed Methodology / Technique
5. References

# Introduction

Machine Learning is an application of artificial intelligence that allows the machine to learn from data and sample and experience without being programmed. So instead of us writing the code, what we can do is we feed data to the machine learning algorithm, and the algorithm or the machine builds the logic based on the given data and can predict the outcome for the future data. Our group selects Suicide Rates Overview from 1985 to 2016 as a dataset. This project we will study the suicide rates from year to year and do the prediction using the Microsoft Azure machine learning.

# Problem Statement

Every year almost 800,000 people died because of suicide, and there are many more who attempt suicide. World Health Organization states that 79% of global suicides occur in low- and middle-income countries. It also mentions that the third leading cause of death in 15-19 years olds is suicide. Suicide is a severe mental health problem that can prevent and suicide is not a shortcut to solve the problem.

# Objectives

1. To collect the data about suicide rates from 1985 to 2016.
2. To study the suicide rates from 1985 to 2016
3. To predict the suicide rates for the upcoming year.

# Data Descriptions

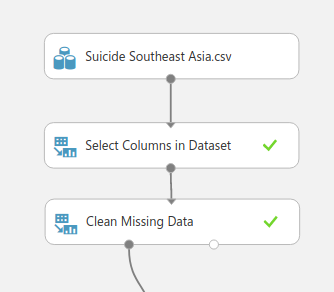
Our dataset is about suicide in Southeast Asia. It consists of 886 number of observations from the year 1985 to 2015 from over Singapore, Thailand, and the Philippines. There is a total of twelve columns as part of the attributes. They include the country year, sex, age, suicides\_no, population, suicides/100k pop, country-year, HDI for the year, gdp\_for\_year ($), gdp\_per\_capita ($), and generation. The type of the dataset is supervised learning because the data of the class is known and provided in the training phase as well as the desired output.

# Proposed Methodology / Technique

The methodology we use in our project is a linear regression algorithm. Linear regression is a supervised machine learning algorithm. Regression is a prediction of a target that is numeric. Because our project is to predict the number of suicide rate, we use linear regression as linear regression predict the number.

# Experimental Setup Data Preprocessing

Data preprocessing is important to make sure that our data is not redundant and have incomplete data as it can affect the quality of the results. The figure below shows the steps taken for data preprocessing.



Firstly, we exclude the unimportant column using the Select Columns in Dataset module. Below is the Column of our visualize dataset. We are excluding suicides\_no, country-year, HDI for year, gdp\_for\_year($), gdp\_per\_capita($), and generation.

Before Run Select Columns in Dataset:



After Run Select Columns in Dataset:



Next, we use Clean Missing Data module to clean the data by removing rows that have any missing values.

Before Run Clean Missing Data:

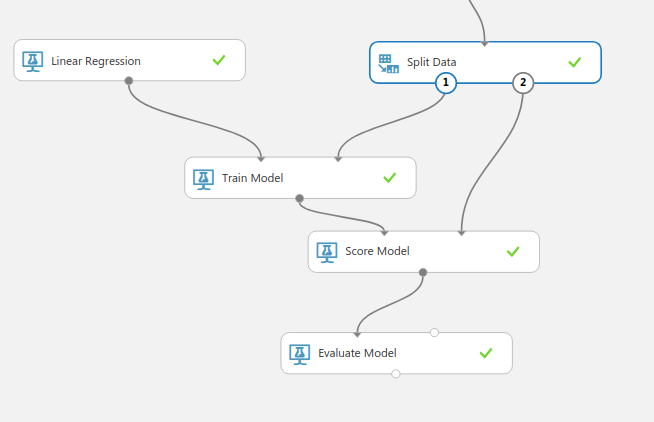


After Run Clean Missing Data:

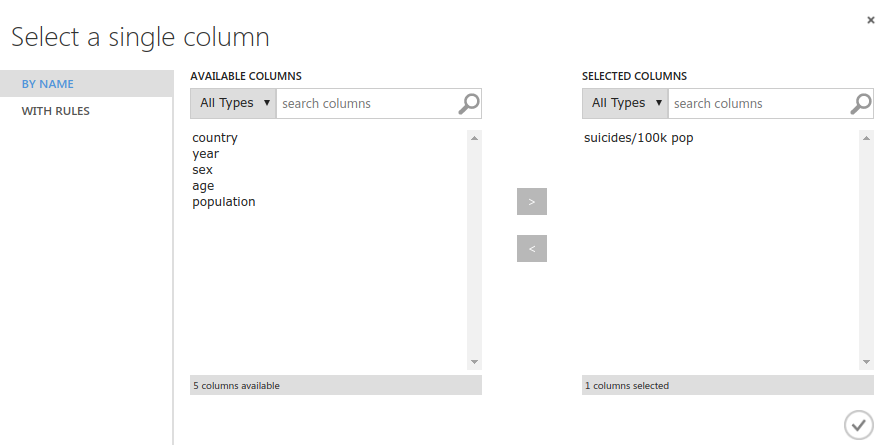


# Data Transformation

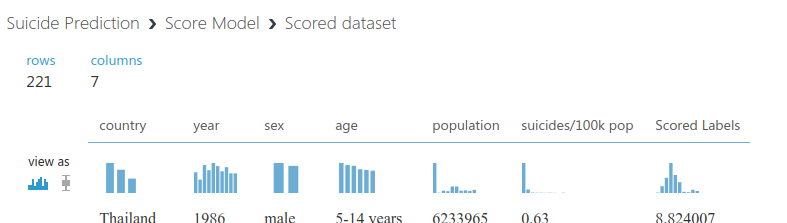
After finish data preprocessing, we proceed to data transformation, which the data will split into a training set and testing set. The figure below shows the data transformation for our project.



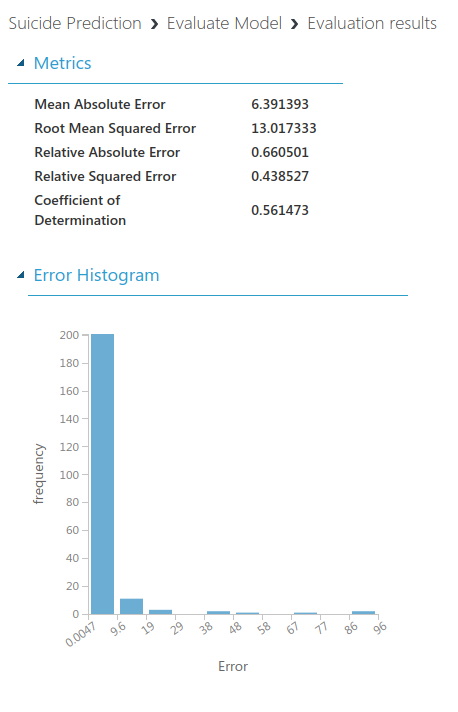
We split the data set it to 0.75. This way, we'll use 75 per cent of the data to train the model, and hold back 25 per cent for testing. Splitting the data for testing and training is essential to have an accurate prediction of data.



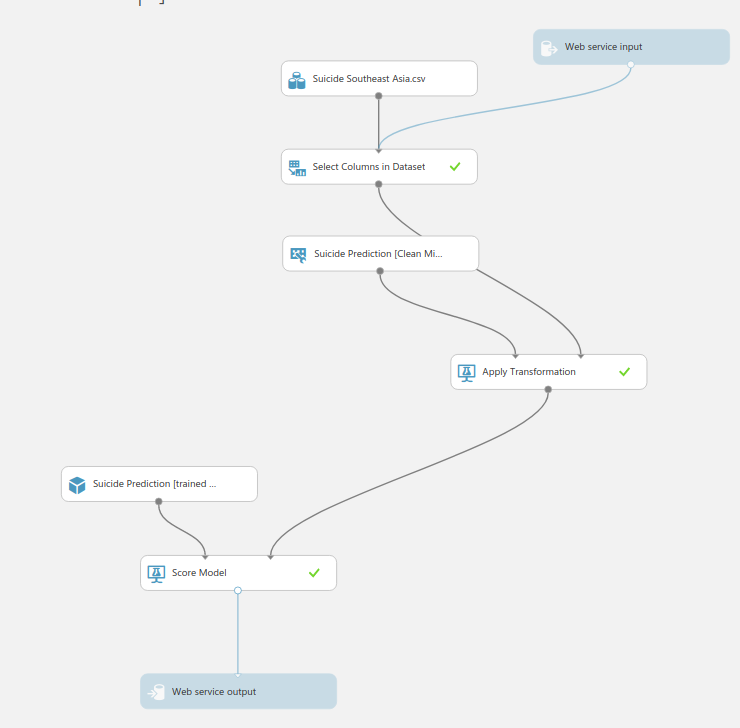
Then, we use Train Model module to train the 75% of the dataset. We then connect Linear Regression module to the Train Model module because we want to use linear regression to predict the outcome. In the Train Model module, we chose suicides/100k pop column as the column to be trained. Therefore, suicides/100k pop is the column we want to predict.



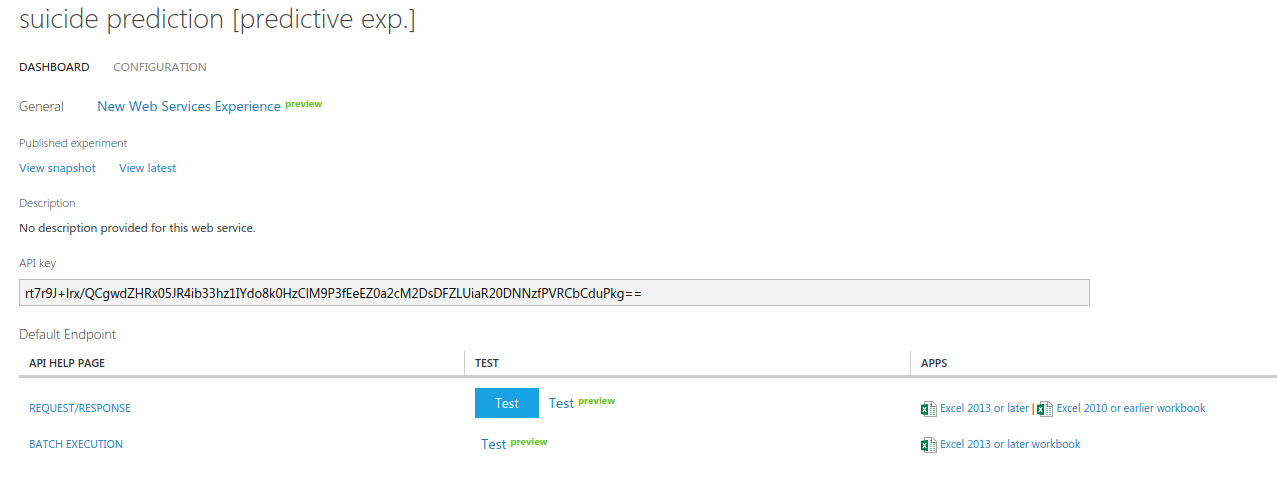
Next, we use Score Model module. Train Model module and Split Model module are connected to the Score Model module. In the Score Model module, the output shows the predicted values for suicides/100k pop which is the scored labels in the diagram and the known values from the test data which is the suicides/100k pop column in the diagram.



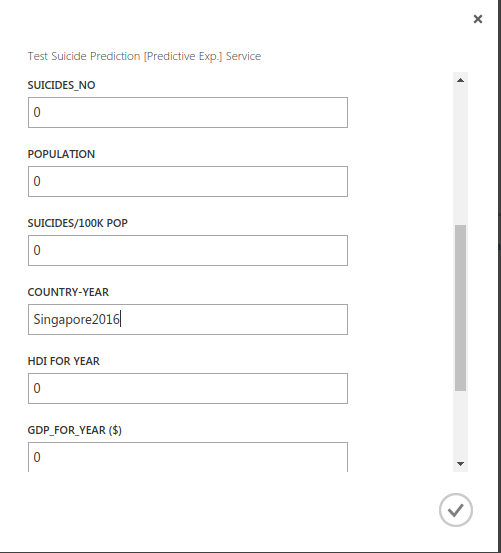
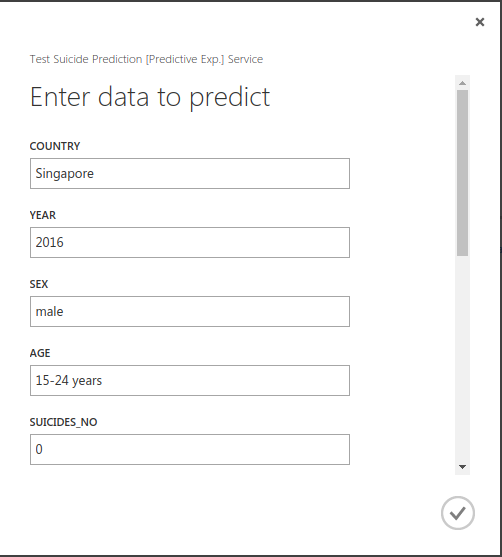
We then evaluate the model using the Evaluate Model module connected from the Score Model module. In order to evaluate, it is preferable for each of the error statistics to have a smaller value. For coefficient of determination, the closer its value is to one (1.0), the better the predictions.



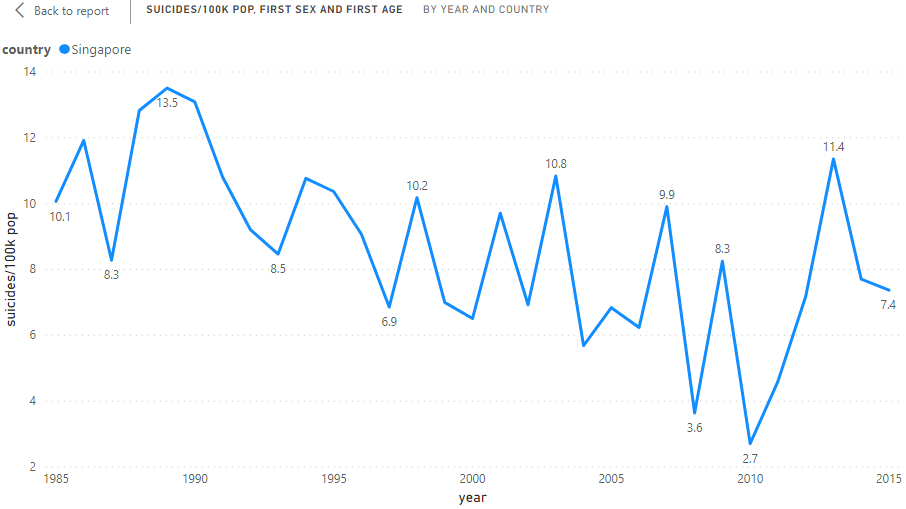
To execute predictive experiment, we deploy the web service by choosing the predictive option. The experiment will automatically execute the process until the experiment transformed into the predictive experiment in the above diagram.

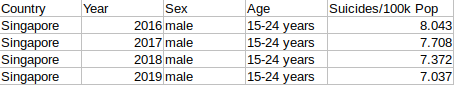


To predict the values, we click Deploy Web Service to get the output above. To begin testing, we choose the Test option for prediction of suicide rate.



Then, we fill in the information needed to predict the suicide rate based on each year. We continued the test until a particular year is reached.



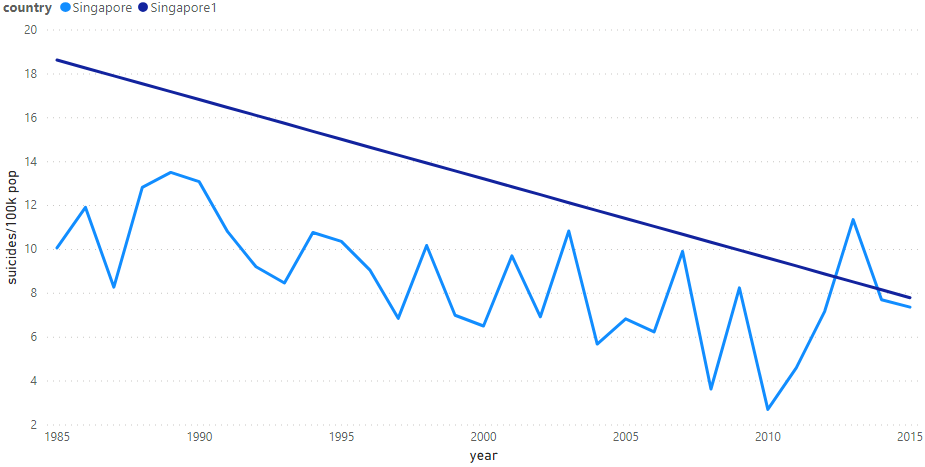


Finally, we received the outcome of the predictive experiment. The predicted value is the suicides/100k pop column.

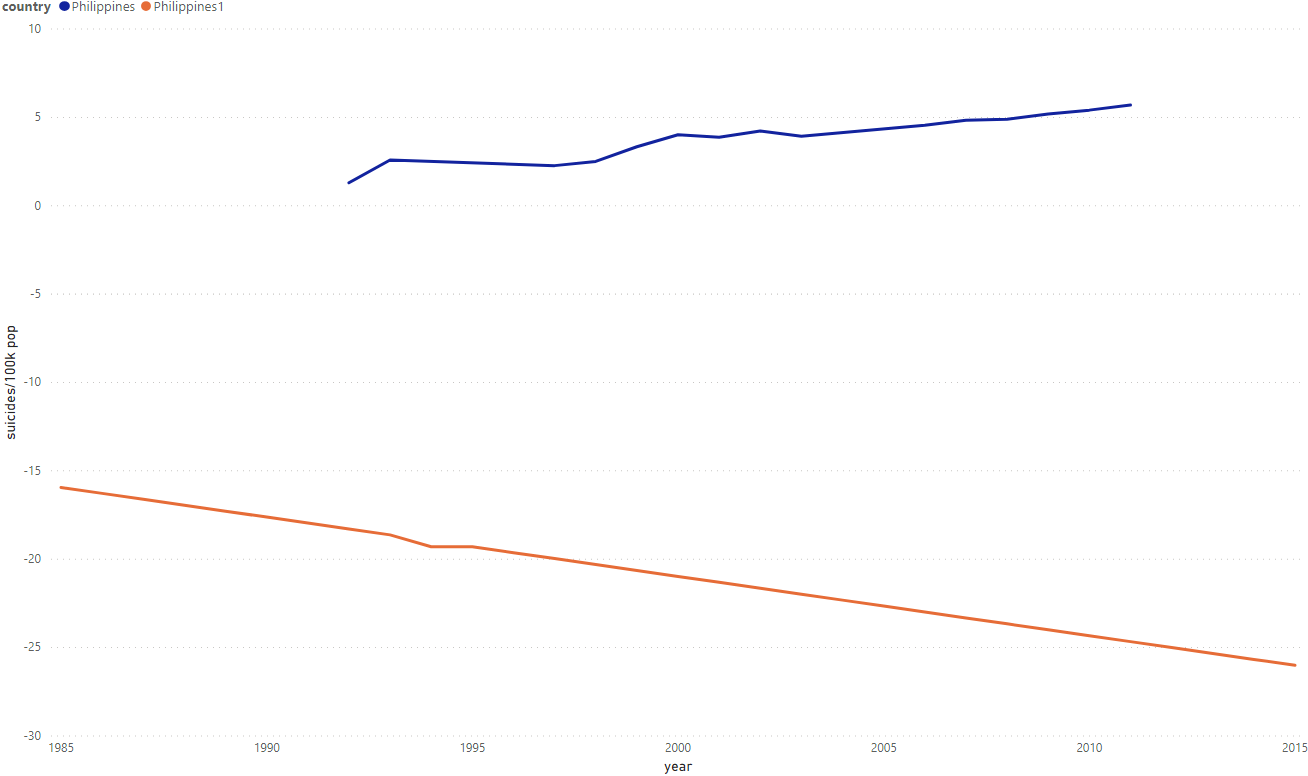
# 

# Predictive Experiment

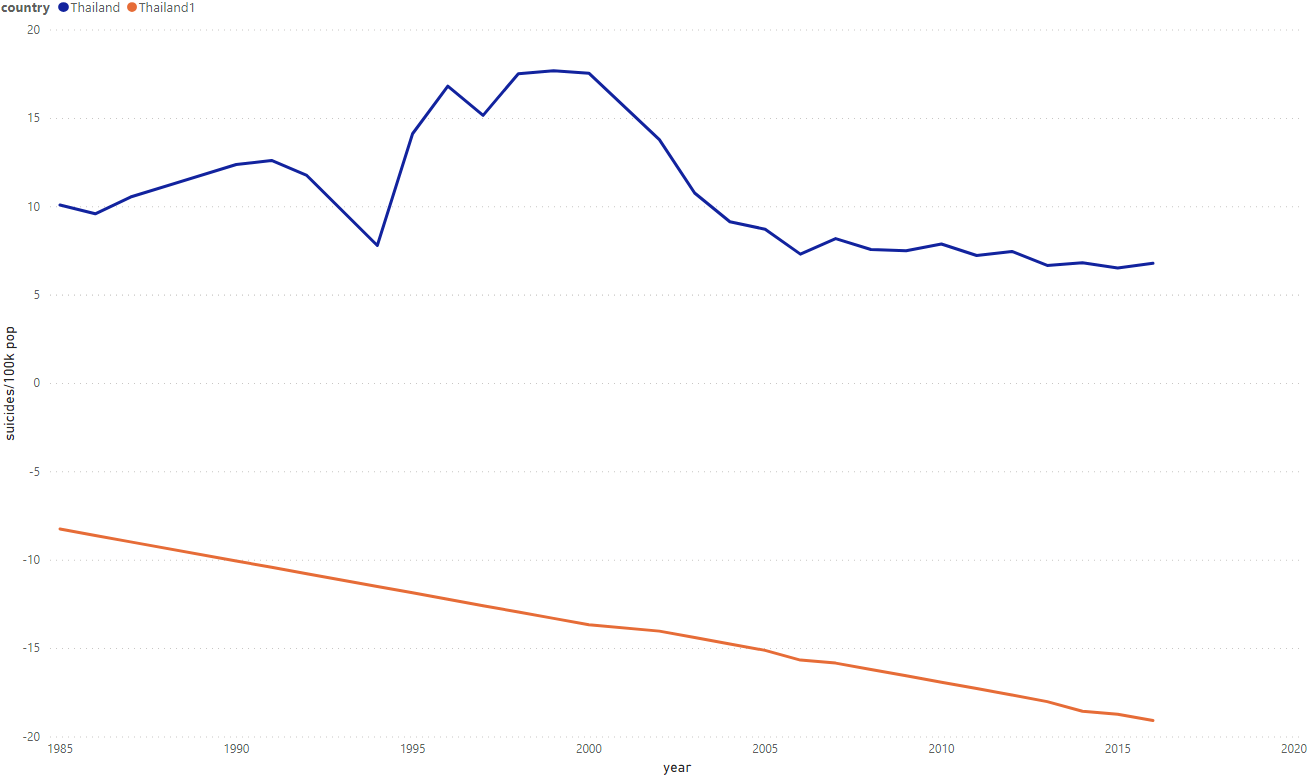
## Prediction in the same year as in the dataset.

Singapore - Original Dataset

Singapore1 - Predictive Dataset

Philippines - Original Dataset

Philippines1 - Predictive Dataset

Thailand - Original Dataset

Thailand1 - Predictive Dataset

## Observation

As we can see, the predictive result shows the same pattern as in another country where it’s negative linear regression.

# References

Rusty. (2018, December 1). Suicide Rates Overview 1985 to 2016. Retrieved from <https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016>.

Suicide. (n.d.). Retrieved from <https://www.who.int/news-room/fact-sheets/detail/suicide>.