SUICIDE ANALYSIS AND PREDICTION

1.0 Introduction

This iteration is a deep dive into the suicide dataset for knowing the reasons behind suicides each year happening around the world. Though, several studies like John et al had been done in the past. This study is going to make new findings that are useful for govt. entities to make policies that could reduce the mortality rate in the future. I will be looking at different aspects of suicides and predicting how much more suicides are going to happen in the coming years in different countries.

Time Series and Statistics are used for this project. The purpose of this study is to understand the reasons for suicide. Suicide rate in many countries is higher than the total mortality rate. In order to make a change to this scenario, we need to study different patterns and clusters in the data and understand what factors are triggering the tendency for someone to make such decisions. whilst a web-based system will be created that can dynamically create useful visualizations on the python dashboard. Every step will be well documented and updated from time to time on GitHub and this repo will be linked to the mocha host server.

2.0 Background Problem

As we it's been always so heartbreaking listening to each time, we listen to some news about the number of people committing suicide every day in different part of the world. I strongly believe there is need for help from the society to take care of these people who are vulnerable to death. In Ireland, as per the CSO statistical release of 2011, 554 people suicide in 2011 alone. I assume there are number of circumstances that affect people to commit suicide. Here Ireland is just an example among many other countries like Russia and Ukraine their suicide rate is comparatively very high compared to other countries. Have you ever thought what are the possible reasons behind these figures? I was not sure about the answer for this question until I started working on the suicide analysis and prediction. Of course, there could be thousands of reasons behind every death, but I assume there are also some common factors which leads people to commit suicide in every country.

3.0 Technologies and methods

3.1 Data Preparation

In every data analysis, about seventy percentage total time is spent on preparing the data, make it ready for doing analysis. Initially I had to explore the dataset using describe () method. Also using visualizations and some statistical analysis I have cleaned the dataset. Imputation was carefully done based on the time, context, and importance of the variable. I have chosen a dataset which was simple and aggregated. But, later on thinking about the complexity and wide range of the reasons behind committing suicide I did a thorough research about how much additional information I can incorporate into the existing dataset. There have been several variables like continent missing in the dataset. So, I have added additional columns for continent names. Also, I have received another dataset which is similar to the suicide master sheet I have previously received contained much more information. The main reason behind taking this dataset into account is that those variables were very meaningful with respect to the context I am working with, for example, I assume there could be come relation between suicide rates and unemployment or number of internet users and suicides in any country.

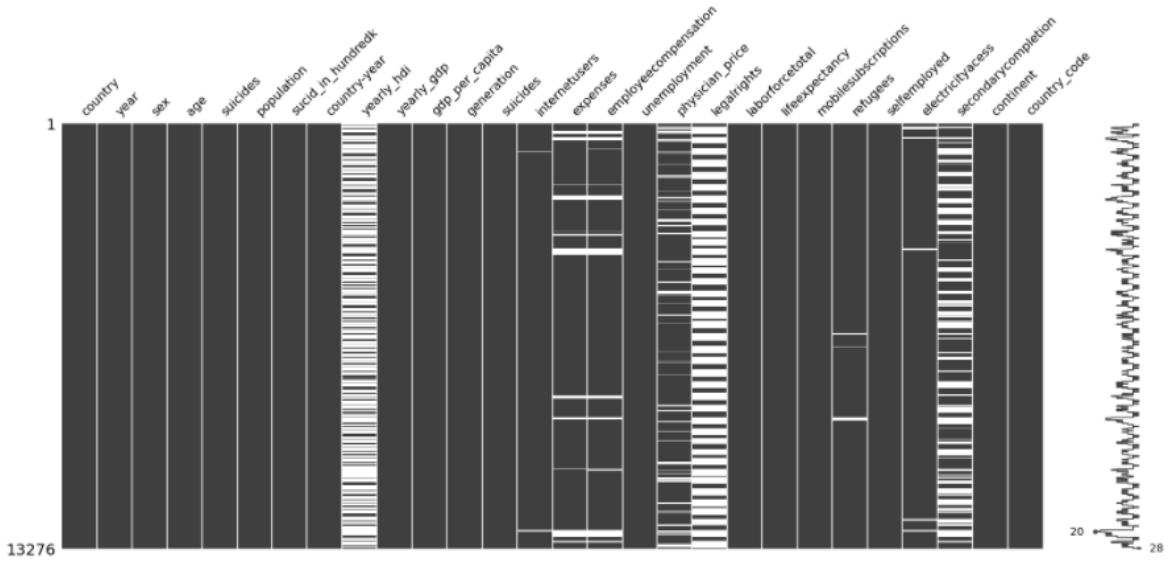
```
df.describe(include='all').T
```

✓ 0.2s

	count	unique	top	freq	mean	std	min	25%	50%	75%
country	27820	101	Netherlands	382	NaN	NaN	NaN	NaN	NaN	NaN
year	27820.0	NaN	NaN	NaN	2001.258375	8.469055	1985.0	1995.0	2002.0	2008.0
sex	27820	2	female	13910	NaN	NaN	NaN	NaN	NaN	NaN
age	27820	6	15-24 years	4642	NaN	NaN	NaN	NaN	NaN	NaN
suicides	27820.0	NaN	NaN	NaN	242.574407	902.047917	0.0	3.0	25.0	131.0
population	27820.0	NaN	NaN	NaN	1844793.617398	3911779.441756	278.0	97498.5	430150.0	1486143.25
sucid_in_hundredk	27820.0	NaN	NaN	NaN	12.816097	18.961511	0.0	0.92	5.99	16.62
country-year	27820	2321	Belarus2007	12	NaN	NaN	NaN	NaN	NaN	NaN
yearly_hdi	8364.0	NaN	NaN	NaN	0.776601	0.093367	0.483	0.713	0.779	0.855
yearly_gdp	27820.0	NaN	NaN	NaN	445597926548.398254	1453907394884.571777	46919625.0	8985352832.0	48114688201.0	260000000000.0
gdp_per_capita	27820.0	NaN	NaN	NaN	16866.464414	18887.576472	251.0	3447.0	9372.0	24874.0
generation	27820	6	Generation X	6408	NaN	NaN	NaN	NaN	NaN	NaN

There were many such variables making my research firm on the ground in terms of working with useful and meaningful information for machine learning modeling. Second Main reason is that Data Visualization is a major part of my final project. If I had a greater number of variables in the dataset, I would get more opportunities of making more visualizations. Outliers in the data are a one main thing we need to carefully do. Replacing the outliers without thinking why they

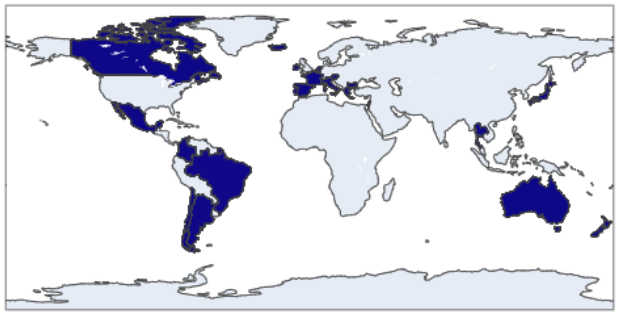
occur is a dangerous practice.



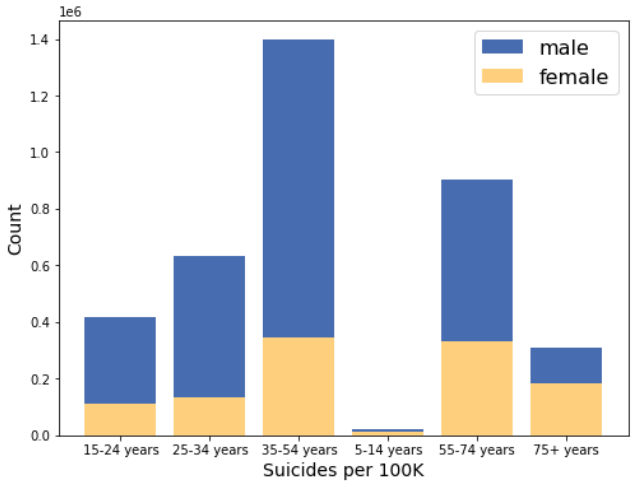
Above figure gave me a clear idea of how much data is missing in each column of the dataset. Also used visualizations like boxplot and histogram, I have explored the dataset for data preparation.

3.2 Data Visualization

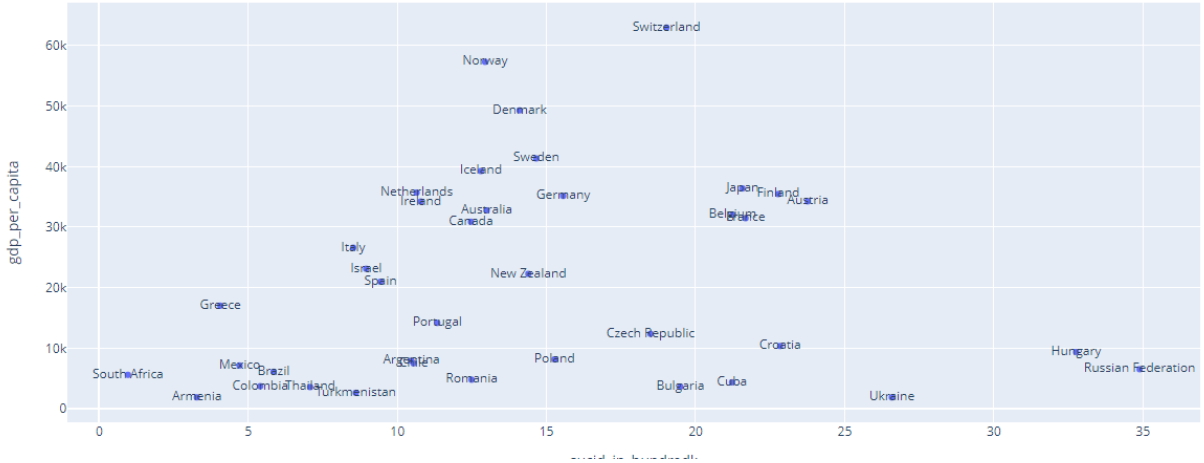
Firstly, I want to talk about python dashboard. It's always wonderful to see how we are able to make models and interpret them. But it is also important to note, recently there are number of concerns about how well we are able to make modifications to the existing model and maintain them. So, our model has to work dynamic and make prediction based on the available data. In recent years programmers used use VueJS or web-based languages for making dashboards, we now have most advanced packaged like Streamlit has made these process easier and more efficient. I am going to use some of the python packages like plotly to make interactive dashboard and make models that can make great predictions.



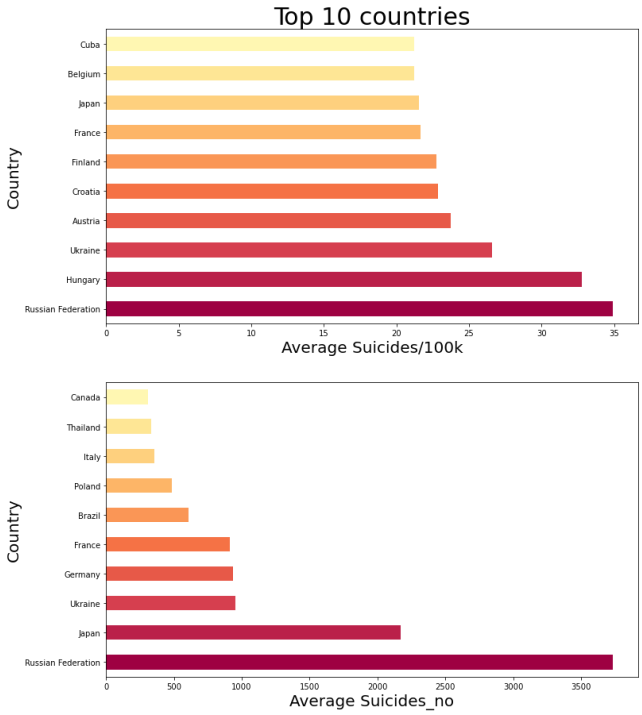
Global suicide rate is visualized using plotly. It is an animation frame page visitors are able to see the information based on the year. Colored regions represents the rate of suicides per hundred thousand.



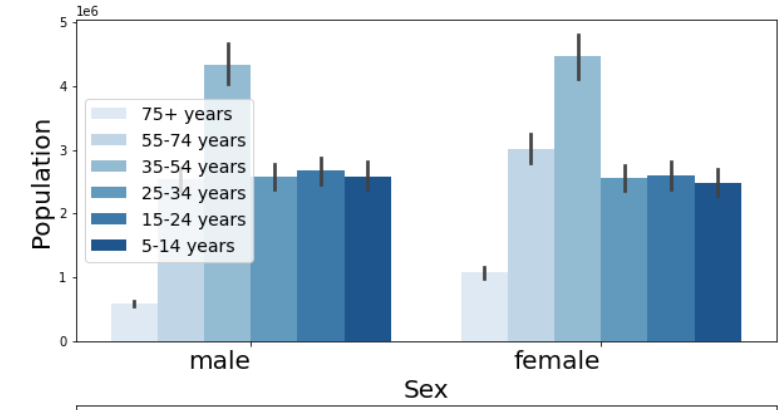
As per the fig. It's very evident that most of the suicides are happening between the age of 35 and 54. And out of them majority are Males. In all the age groups females are less affected groups. Also, we can see from the age of five to fourteen children are less likely to commit suicide.

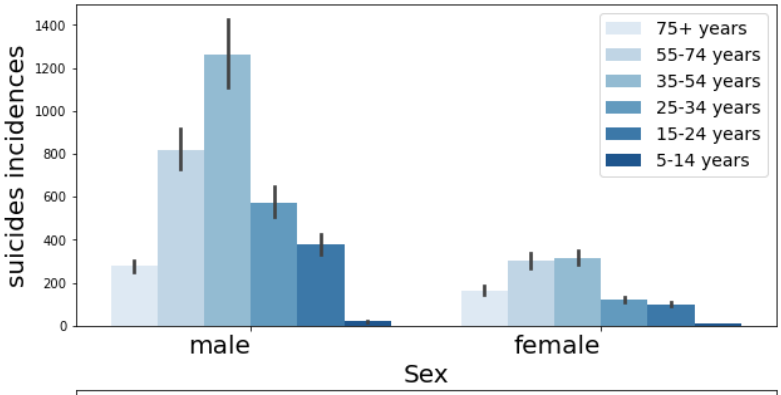


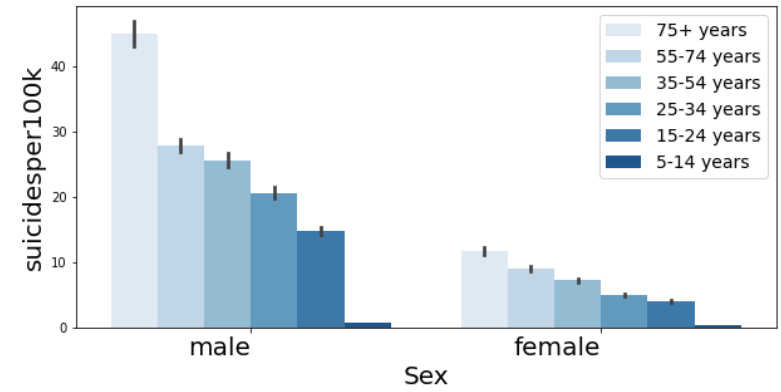
As per the fig we can see there are countries like Russia, Ukraine and Hungary having some sort of relation in terms of number of suicides per hundred k. As per the report from BBC News, the reasons for suicides in Ukraine is an after math of the conflicts between Russia and Ukraine and continuous war. Our data clearly shown the one of the most affected countries in terms of suicide as Ukraine.

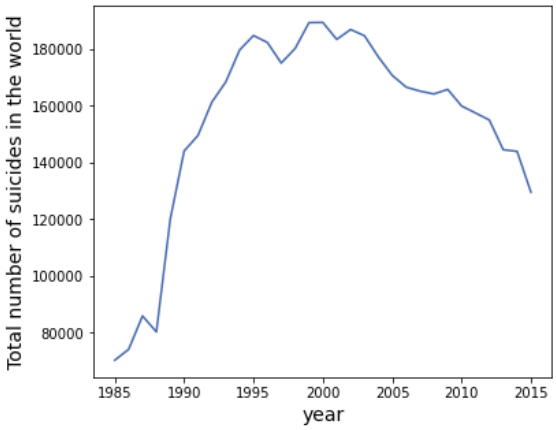


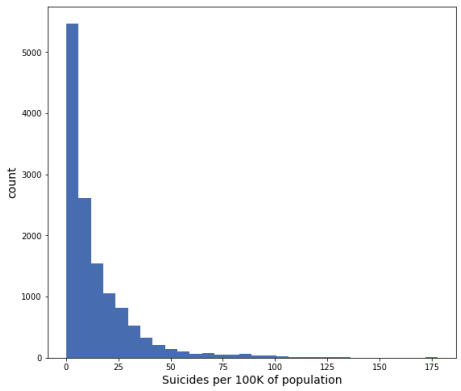
The above figure shows the scale of suicides among different countries. It is very clear to see that Russian federation is showing the most suicide rate among all the other countries. As per the previous studies like Bellman and Namdev study it's very clear that Russia is facing issues with suicidal behaviour from male population and also their drinking habits have significant effect on leading them to commit suicide.

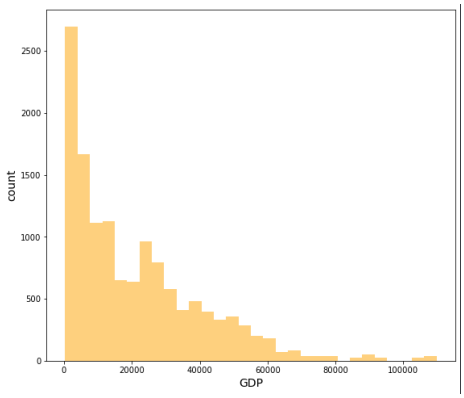

1e6
Population
75+ years
55-74 years
35-54 years
25-34 years
15-24 years
5-14 years
male
female
Sex

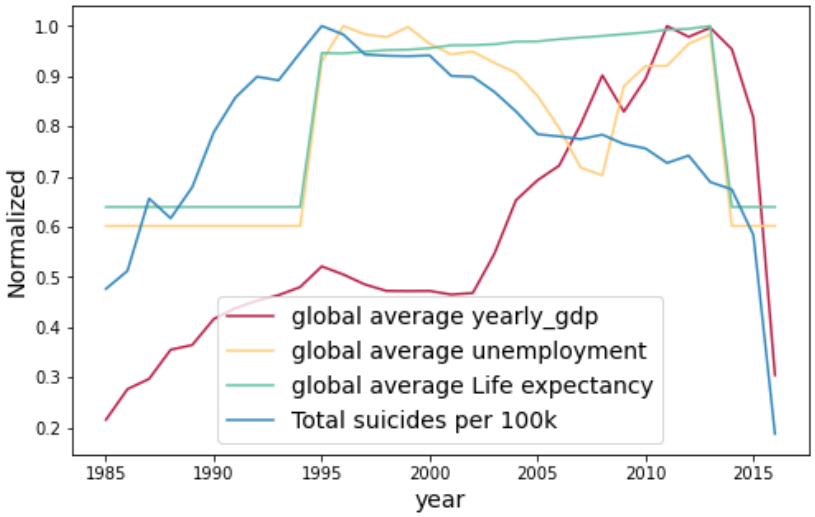

suicides incidences
75+ years
55-74 years
35-54 years
25-34 years
15-24 years
5-14 years
male
female
Sex


suicidesper100k
75+ years
55-74 years
35-54 years
25-34 years
15-24 years
5-14 years
male
female
Sex


Total number of suicides in the world
180000
160000
140000
120000
100000
80000
1985
1990
1995
2000
2005
2010
2015
year

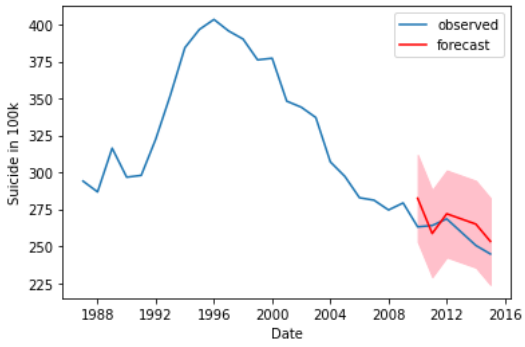

count
5000
4000
3000
2000
1000
0
0
25
50
75
100
125
150
175
Suicides per 100K of population

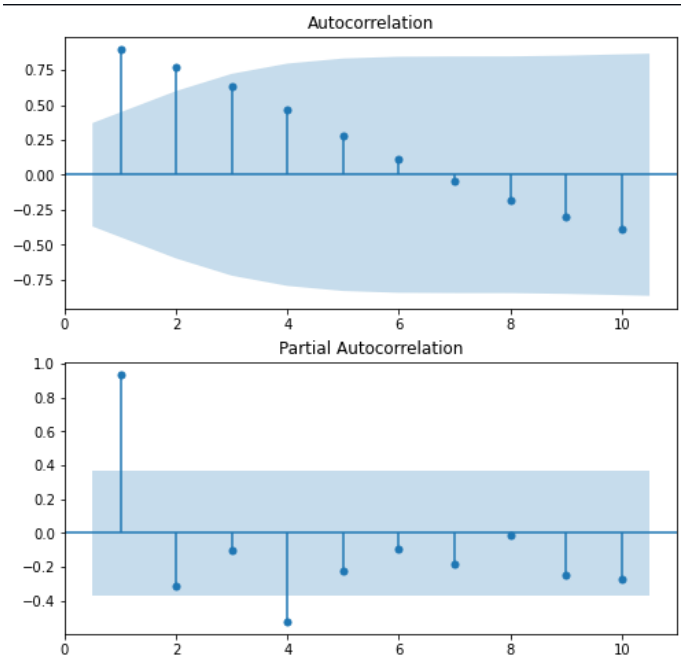

count
2500
2000
1500
1000
500
0
0
20000
40000
60000
80000
100000
GDP


Normalized
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
1985
1990
1995
2000
2005
2010
2015
year
global average yearly_gdp
global average unemployment
global average Life expectancy
Total suicides per 100k

3.3 Modeling and Forecasting

Secondly, working with time series forecasting is a crucial part in my dissertation. I have number of different targets in my dissertation. I have been looking for ways to predict the number of suicides in upcoming years. My interest in time series and ML made me dive deep into sophisticated time series models like SARIMA and VAR to make models on the suicide data and forecast future suicides in different countries. Dealing with numerical and categorical variables I have used decision tree classifier capable of classifying purposes. I could see the Brunello et al has successfully implemented them in their research which really inspired me to adapt the idea in suicide analysis.





AIC and BIC graphs made for checking the order of ARIMA Model. In my research I am trying to work on predictions so, I will be looking at the AIC. From the graph we can see we need to make an AR1 Model.

SARIMAX Results

Dep. Variable:	sucid_in_hundredk	No. Observations:	28
Model:	SARIMAX(1, 0, 0)	Log Likelihood	−120.633
Date:	Tue, 07 Jun 2022	AIC	245.266
Time:	23:35:39	BIC	247.930
Sample:	0	HQIC	246.080
	− 28		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.9982	0.009	109.622	0.000	0.980	1.016
sigma2	264.3394	81.676	3.236	0.001	104.257	424.421

Ljung-Box (L1) (Q):	2.17	Jarque-Bera (JB):	1.12
Prob(Q):	0.14	Prob(JB):	0.57
Heteroskedasticity (H):	0.20	Skew:	0.46
Prob(H) (two-sided):	0.02	Kurtosis:	2.68

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).

3.3.1 grid search ARIMA parameters for time series

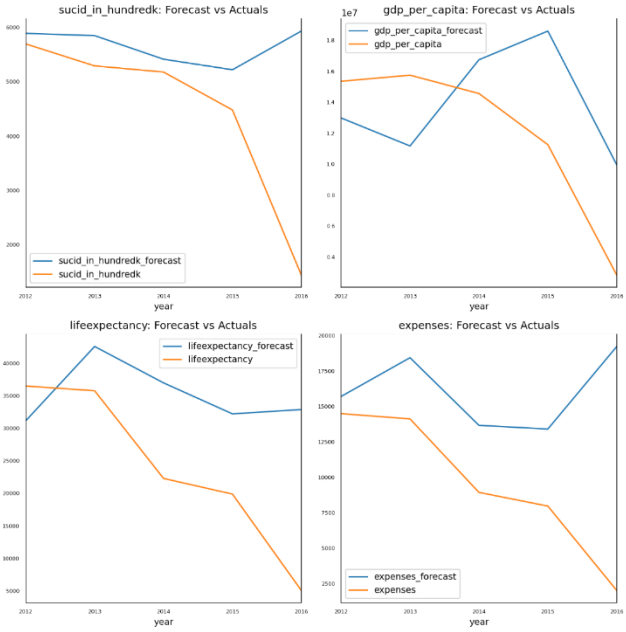
```
ARIMA(0, 0, 0) RMSE=63.575
ARIMA(0, 0, 1) RMSE=33.394
ARIMA(0, 1, 0) RMSE=9.886
ARIMA(0, 1, 1) RMSE=9.792
ARIMA(0, 1, 2) RMSE=9.744
ARIMA(0, 2, 0) RMSE=15.093
ARIMA(0, 2, 1) RMSE=13.221
ARIMA(0, 2, 2) RMSE=13.159
Best ARIMA(0, 1, 2) RMSE=9.744
ARIMA(1, 0, 0) RMSE=11.713
ARIMA(1, 0, 2) RMSE=11.477
ARIMA(1, 1, 0) RMSE=9.746
ARIMA(1, 2, 0) RMSE=11.383
Best ARIMA(0, 1, 2) RMSE=9.744
ARIMA(2, 0, 0) RMSE=12.069
ARIMA(2, 1, 0) RMSE=8.936
ARIMA(2, 1, 1) RMSE=131.442
ARIMA(2, 2, 0) RMSE=11.006
Best ARIMA(2, 1, 0) RMSE=8.936
ARIMA(4, 0, 0) RMSE=13.455
ARIMA(4, 1, 0) RMSE=9.786
ARIMA(4, 2, 0) RMSE=12.836
Best ARIMA(2, 1, 0) RMSE=8.936
ARIMA(6, 1, 0) RMSE=11.287
Best ARIMA(2, 1, 0) RMSE=8.936
ARIMA(8, 1, 0) RMSE=16.211
Best ARIMA(2, 1, 0) RMSE=8.936
Best ARIMA(2, 1, 0) RMSE=8.936
```

3.3.2 Prediction using Vector Auto Regression Models (VAR Model)

Another Model used for the time series data is VAR model. Vector Auto Regression. The reason behind using this model is that it helps in forecasting models based on multiple variables in time series. Usually, we use single variable and sequential time for time series Analysis. But here I was able to include multiple variables in the model as you can see in the figure. Vector Autoregressive Models are one of the best models we could use choose for time series.

year	sucid_in_hundredk_2d	gdp_per_capita_2d	lifeexpectancy_2d	expenses_2d
2012-01-01	529.412795	-4.469357e+06	-5481.612016	1599.709929
2013-01-01	-348.035624	1.086854e+06	16733.380993	1462.733174
2014-01-01	-393.171701	7.379111e+06	-17050.184437	-7497.633351
2015-01-01	240.134938	-3.715248e+06	833.249851	4506.396883
2016-01-01	905.938622	-1.050009e+07	5435.718296	6079.875040

You can see in the above example we have predicted the number of suicides for the year 2016 using VAR Model on the time series sequential data.

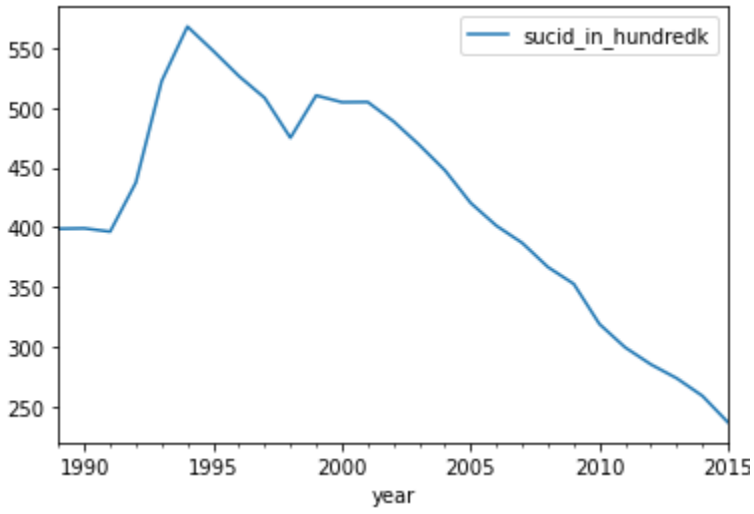



```
Forecast Accuracy of: gdp_per_capita
mape   :  0.8708
me     :  2337090.5241
mae    :  4961066.6444
mpe    :  0.7028
rmse   :  5796573.8089
corr   :  0.1443
minmax :  0.337

Forecast Accuracy of: lifeexpectancy
mape   :  1.3675
me     :  10752.5554
mae    :  13426.9478
mpe    :  1.2942
rmse   :  15143.4315
corr   :  0.3972
minmax :  0.3953

Forecast Accuracy of: sucid_in_hundredk
mape   :  0.6739
me     :  1202.9923
mae    :  1202.9923
mpe    :  0.6739
rmse   :  1996.3575
corr   :  -0.1687
minmax :  0.2109

Forecast Accuracy of: expenses
mape   :  1.8365
me     :  6167.0735
mae    :  6329.6375
mpe    :  1.8257
rmse   :  7927.7401
corr   :  0.1224
minmax :  0.3928
```

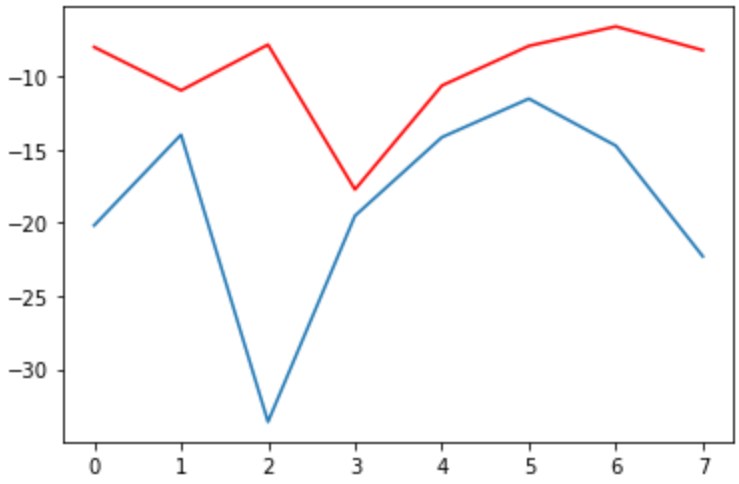
As you can see in the fig, we have calculated the accuracy of each variable. This helps us understand how well our model is performing.

3.3.3 Prediction using Auto Regression Models (AR Model)

Next model I have created is Auto regression model, Train and Test was split int seventy and 30 percentages. Seventy percentage of the data was used for training the model and rest thirty percentage was used for testing. I have got 11.792 Root mean squared error. Also, I could save different models to the local and I was able to load the models later and update them accordingly.



As you can see in the diagram above, the suicides per hundred thousand is distributed throughout the year is shown.



In this fig, the blue line is the test data and red line is the predicted data. I have AR 1 Model with one window.

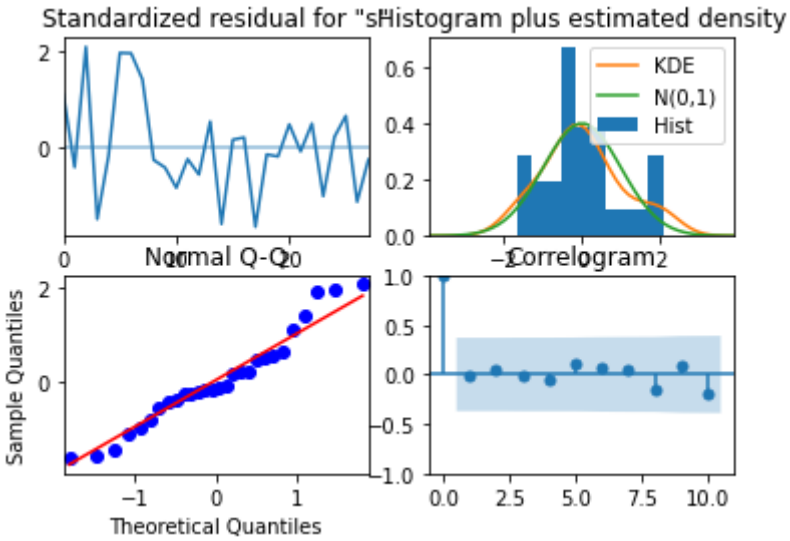
3.3.4 Decision Tree Classifier

I have added a new column called risk where I split the data into two classes, class 1 stands for high risk and class 0 for low risk. Using decision tree model, I have made classification.

```
******************Decision Tree classifier**************
Accuracy = 0.9962962962962963
Train Accuracy= 1.0
CM
 [[144   1]
 [  0 125]]
classification report for decision tree
              precision    recall  f1-score   support

           0       1.00      0.99      1.00       145
           1       0.99      1.00      1.00       125

    accuracy                           1.00       270
   macro avg       1.00      1.00      1.00       270
weighted avg       1.00      1.00      1.00       270

# of leaves 4
 Depth 3
```

3.4 Evaluation of models

Evaluation of model is as important as making the model. I have created 3 models in ARMA , Auto regression and Vector Auto regression. Using mean squared error and R-squared error I took the error rates of different models. Accuracy also calculated to understand how efficient and precise my model is. Initial modeling was done on time series of “Republic of Russia”. Final modeling will be done on live website using python Dash or Streamlit.



Finalizing my models, I would be able to add more evaluation techniques. Above are the four diagnostic plots I have created after running ARIMA Model. Looking at the plots I can say there is no pattern in the standardized model. Looking at the Histogram there is no gaussian distribution (green and red line should be almost the same for gaussian). The QQ Plot seems to be not normally distributed, if it is normally distributed all the blue dots will be aligned over the line (except some values in the either end)

3.5 Data Storage

Thirdly, we need a database server for data to be stored on the server. I will be using PSQL or MySQL servers for data storage and management. I want the data in my DB to be updated time to time and my model has to be updated based on the new data injected in each time. The reason

for choosing these DB's is the flexibility of usage and its syntax matching with Structured Query Language (SQL) minute differences.

Initially data was stored in the csv format in different files. Later I have uploaded them into mochahost Psql server.

3.6 Web Server and Hosting

We know there are thousands of hosting companies providing hosting services, I have chosen Mochahost

one of the best service providers for small businesses website. My goal is to make highly dynamic web application on the server. I have purchased a VPS service which allows to run PIP Packages on their server making the server IDE more suitable for the Dash App. Mochahost cPanel will be connected to the github repository where my application wll be updated time to time. Using git technology for the hosting making the process more sophisticated and professional in terms of version control.

3.7 Data Security

Data Security has become an Important concern in this era. Even though the suicide dataset is publicly available on the internet, I have followed the best practices in data security to ensure there is no data leakage. I have used an encrypted windows drive to store the data. Whole project codes are updated time to time to GitHub private repository. Any information related to this study has been considered for data security and ethical practices before actually using them. No personal information is used in this study. For web applications files kept in a private repository and used that repo to pull changes to the live mochahost server.

3.8 Applications and software's

Microsoft Visual Studio is the main IDE used for coding. For version control I have used Github. I also have used other tools like Jupyter Notebook, Spyder, Atom for coding purposes. All the testing are done with the local Anaconda environment. Python version 3.8.8 is used for the whole research. PIP package is used for configuring the IDE. Microsoft excel is also used for minor csv file inspections. Visual Studio's inbuilt git version control features are used time to time for managing the branches in the repository.

For reports and notes, MS Word and notepad are used. PowerPoint is used to create slides for presentation. Adobe Acrobat DC is used for managing PDF Files. Visual Studio in-built terminal, Anaconda Prompt and windows terminal are used for running PIP and git commands. Windows Operating System is used for the whole work. Google Chrome and Mozilla Firefox are the two browsers used in this project.

Ethical Considerations

A harm is inflicted when an action of a researcher affects the participants or society. There are several reasons behind the actions of researchers that create huge trouble for society or individuals Furthermore (Human Radiation Experiments | Atomic Heritage Foundation n.d.) was one of the biggest examples of this. In 1994 US President Clinton created an advisory team to research human radiation that has been conducted over the years. In this study, doctors injected Plutonium into the body of many patients and many of them did not consent to be part of this study. Also, there was a company called Quaker Oats which is also part of this study included radioactive components in oatmeal and were unknowingly fed to the children.

In my study, no such experiment is done on humans in the process of data collection or analysis. An aggregated suicide dataset only provides information about the country's general population is used throughout the research. No prior experiment is conducted to gather data for this research. No harm is made to any subject in this regard. There are several benefits related to the data. Data provides an overview of how many suicides are happening from time to time. Talking about the societal impact of this research is enormous. For example, (Study: Benefits of Electric Cars Add Up—in the Billions! | NRDC n.d.) has created a significant impact on how this research has benefited society to help understand the carbon footprint reduction

and cost-saving. In suicide analysis I am trying to make use of data to leverage suicide attempts by helping the govt to take measures or policies from the outcome of my study it's going to help create plans to tackle such acts in coming years.

My research about 'suicide analyses was based on the dataset which is open source in Kaggle which anyone can download. Thinking about the data storage and security I would not say it's a very much sensitive dataset because first of all this dataset is not private on the internet, so the author has kept the access to the public. Secondly, this suicide dataset is not using any confidential information about any individual rather it's more of a summary dataset providing general information about the country's deaths rates. Also, information like the age group, internet users, human development index etc who is more likely to commit suicide.

SWOT Analysis: (SWOT Analysis based on the Guidelines of ACL as per ethics)

As I previously stated, the suicide dataset did not contain any personal information. I would highly recommend for future studies we need to incorporate more humans in the experimentation to collect data from people in real-time. The most important thing we need to follow taking consent from each person who is willing to participate in the study. There are some major concerns in this regard, let's imagine if we have not taken any consent from these human beings who are participating in the study/experiment. They might later go to court and file a complaint against us for doing illegal use of personal information. Also, we need to clearly state what are the acts or dangers involved in the experiments. So that they are aware before they participate in these activities. (Bogod 2004) is one of the real-life examples where in 1942 prisoners were asked to undergo dangerous experiments to understand the survival chance of soldiers sometimes even leading to deaths. Understanding personal, social, and business impacts of data practice.

In addition, even sharing information of individual sharing with any other colleagues or third party would be through proper procedure and getting signs on consent forms.

Strength: - In my study, I am trying to see suicide rates in different countries from time to time. My research strength is its dynamic nature. Similar weather forecast of google or Microsoft, my model will be run from time to time based on the latest data. This research aims at tackling suicide tendencies in every country's population. My research is going to predict

how many people are going to commit suicide in the next 5 years in different countries or continents. When working with a socially responsible research project, it is going to stand out in the world of the internet. Similar to the websites showcasing covid trends live, my website is also going to show the same impact of suicide numbers and create respective visualizations for any general audience to easily understand what the trend in data would be.

Weakness: - The data is aggregated and no specific information of individuals available for forecasts. So, I think the data must be having more specific features which could make accurate predictions about the suicides. But the model would have been more accurate if more specific features could have been added to the dataset. Things like diagnostic information of subjects, population happiness index, education index, happiness index each country. So, my prediction would be more of general understanding about the trends in data.

Opportunity: - It's unexplainable how much we can make use of the suicide data analysis. Govt. is trying to find out the reasons behind suicides or how to reduce the number of suicides every year. We can develop new strategies that can mitigate the effect of suicides through analysis and understanding of existing data. We can make use of ML models act as smart applications which can guide mobile users based on user activity data. Suicide analysis creates a new era of AI where we can keep an eye on who is more vulnerable to death.

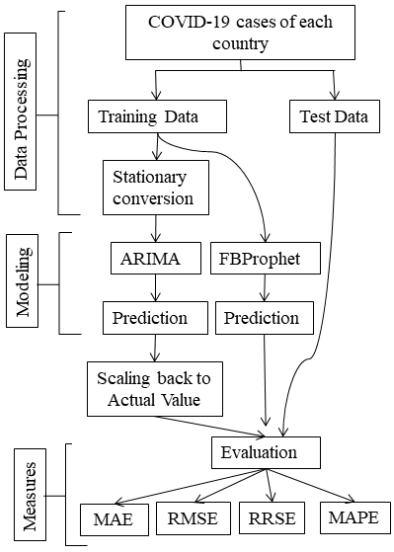
Now let's look at my data and its opportunities. Have you ever thought of having a suicide prediction model for each country? The wide range opportunities using AI and Time Series model on big data is possible using current technologies. Internet of things, cloud computing and ML are the best example of state-of-the-art technologies. Suicide prediction model and live dashboard visualization is a great analysis model which any growing business can take inspiration from. Just imagine a burger selling vendor creating a live predicting model of specific kind of burger that are sold at particular season of a year? or may be checking bestselling milk shakes in each month? Wouldn't these analyses make them grow? or even predict how much products are going to be sold in coming months so they can prepare their store for the coming period to avoid lack of materials. Thus, this model is ultimately showing what kind of predictions or analysis our business and health industry need today to go smarter and do smarter businesses.

Threat: - Data can be used in many different ways. Some people have used it for good reasons others did differently as well. Suicide dataset could be misused in some way. But in my point of view as long as we are not providing specific information about individuals, they are less likely to occur. In my analysis what I would say is incase more features are added to the model in the future, I will have to alter model statically and make them dynamic using cron jobs. Also, when it comes to storing individual information in in near future, more storage space might be required as well as my model could perform poor because of the server requirements. Even though we have other options to buy cloud storage space, it will still be costing more money on the other hand I will have to figure out ways to improve the requirements.

Talking about analysis of suicides in previous years, there could be political impact because of the difference in counts during different political administration periods.

Literature Review

ARIMA Model and FBProphet models used for predicting suicide deaths around the world.(Kumar and Susan 2020). Covid-19 was a very sensitive topic in recent year. Many people have affected by it and lost their life. The dataset is very similar to what I have chosen for my suicide prediction as well. In my view This is an excellent model to take inspiration from. The paper talks about countries including India and to understand the patten in deaths happened around the world. Time series is definitely helped to make predictions about coming years. ARIMA and FBProphet models are used and for analysis data has been split into training and testing.



Another study was done on predicting birth (Włodarczyk et al. 2021), study was trying to figure out the preterm births. This study used machine learning algorithms like support vector machine(svm), random forest, K-Nearest Neighbor, and Convolutional Neural Network (CNNs).

Another study I can point out was done on predicting mortality of predicting attributable to cancer in Qingdao, China: (Qi et al. 2021). They have also used ARIMA Model for prediction of deaths. ARIMA model is combination of autoregressive model and moving average model.Another study was done on prediction o exchange rate (Airiti . 2012). Artificial Nueral Network and ARIMA are used to predict the model.

SVM has been used Time series analysis, like (Huang et al. 2017) has clearly studied classification problem on Breast cancer dataset. This study also checked for different kernel function that used in the SVM Classifier. The outcome of their shows that for large scale datasets RBF kernel based SVM ensembles based onboosting perform better than the other classifiers. SVM is first introduced by (Cortes and Vapnik 1995) shown that it's better for two-group classification problems.

PCA-KNN model is used in (2018) for financial time series prediction, we could use output from sliding window as input for the KNN Model. Principal Component Analysis (PCA) is used in the transformation of the data as well. Suicide dataset will have to undergo above methods to achieve

efficiency and accuracy in modeling or achieve optimum results. Empirically, my assumptions on the previous studies may vary along my research but still this literature review on previous studies has help me improve my preparations for the suicide research in achieving my project goals.

Working with Multivariate time series data, I was looking for models which can making predictions on more than one variable. For example, the (Vector Autoregressive Models for Multivariate Time Series 2006) showed me how relevant is VAR (Vector Autoregressive Models) model for the suicide analysis. It also gave me.

My test includes checking seasonality, tends, stationarity and testing statistical models for finding the best model for prediction. AIC and BIC

References

(3) 1/4: What is Streamlit - YouTube. Available from: https://www.youtube.com/watch?v=R2nr1uZ8ffc [accessed 4 June 2022].

Airiti Library_Comparative+Study+of+Artificial+Neural+Network+and+ARIMA+Models+in+Predicting+Exchange+Rate. Available from: https://www.airitilibrary.com/Publication/alDetailedMesh?docid=20407467-201211-201512080011-201512080011-4397-4403 [accessed 8 April 2022].

Bellman, V. and Namdev, V. (2022). Suicidality Among Men in Russia: A Review of Recent Epidemiological Data. *Cureus* [online], 14(3). Available from: https://www.cureus.com/articles/88128-suicidality-among-men-in-russia-a-review-of-recent-epidemiological-data [accessed 8 June 2022].

Bogod, D. (2004). The Nazi Hypothermia Experiments: Forbidden Data?. *Anaesthesia* [online], 59(12), pp.1155–1156.

Brunello, A., Marzano, E., Montanari, A. and Sciavicco, G. (2019). J48SS: A Novel Decision Tree Approach for the Handling of Sequential and Time Series Data. *Computers 2019, Vol. 8, Page 21* [online], 8(1), p.21. Available from: https://www.mdpi.com/2073-431X/8/1/21/htm [accessed 7 June 2022].

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning* [online], 20(3), pp.273–297.

How suicide became the hidden toll of the war in Ukraine - BBC News. Available from: https://www.bbc.com/news/world-europe-60318298 [accessed 8 June 2022].

Huang, M.W., Chen, C.W., Lin, W.C., Ke, S.W. and Tsai, C.F. (2017). SVM and SVM ensembles in breast cancer prediction. *PLoS ONE*, 12(1).

Human Radiation Experiments | Atomic Heritage Foundation. Available from: https://www.atomicheritage.org/history/human-radiation-experiments [accessed 15 April 2022].

John, A., Okolie, C., Eyles, E., Webb, R.T., Schmidt, L., McGuiness, L.A., Olorisade, B.K., Arensman, E., Hawton, K., Kapur, N., Moran, P., O'Connor, R.C., O'Neill, S., Higgins, J.P.T. and Gunnell, D. (2020). The impact of the COVID-19 pandemic on self-harm and suicidal behaviour: a living systematic review. *F1000Research 2020 9:1097* [online], 9, p.1097. Available from: https://f1000research.com/articles/9-1097 [accessed 7 June 2022].

Kumar, N. and Susan, S. (2020). COVID-19 Pandemic Prediction using Time Series Forecasting Models. *2020 11th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2020* [online], 1 July 2020.

Mochahost Review 2022: Mocha Host Details, Pricing & Features | Sitechecker. Available from: https://sitechecker.pro/web-hosting/mochahost.com/ [accessed 7 June 2022].

Qi, F., Xu, Z., Zhang, H., Wang, R., Wang, Y., Jia, X., Lin, P., Geng, M., Huang, Y., Li, S. and Yang, J. (2021). Predicting the mortality of smoking attributable to cancer in Qingdao, China: A time-series analysis. *PLOS ONE* [online], 16(1), p.e0245769. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0245769 [accessed 8 April 2022].

Study: Benefits of Electric Cars Add Up—in the Billions! | NRDC. Available from: https://www.nrdc.org/experts/luke-tonachel/study-benefits-electric-cars-add-billions [accessed 15 April 2022].

Suicide Statistics 2011 - CSO - Central Statistics Office. Available from: https://www.cso.ie/en/releasesandpublications/er/ss/suicidestatistics2011/ [accessed 7 June 2022].

Tang, L., Pan, H. and Yao, Y. (2018). K-nearest neighbor regression with principal component analysis for financial time series prediction. *ACM International Conference Proceeding Series* [online], 12 March 2018, pp.127–131.

Vector Autoregressive Models for Multivariate Time Series. (2006). *Modeling Financial Time Series with S-PLUS®* [online], 9 October 2006, pp.385–429. Available from: https://link.springer.com/chapter/10.1007/978-0-387-32348-0_11 [accessed 4 June 2022].

Włodarczyk, T., Płotka, S., Szczepański, T., Rokita, P., Sochacki-Wójcicka, N., Wójcicki, J., Lipa, M. and Trzciński, T. (2021). Machine learning methods for preterm birth prediction: A review. *Electronics (Switzerland)*, 10(5).