Suicide Rate Prediction

Mini Project Report

Submitted by

Lakshmi C

Submitted in partial fulfillment of the requirements for the award of the degree of

Master of Computer Applications
Of
A P J Abdul Kalam Technological University

FEDERAL INSTITUTE OF SCIENCE AND TECHNOLOGY (FISAT)®
ANGAMALY-683577, ERNAKULAM(DIST)
MARCH 2022

DECLARATION

I, **Lakshmi C**, hereby declare that the report of this project work, submitted to the Department of Computer Applications, Federal Institute of Science and Technology (**FISAT**), Angamaly in partial fulfillment of the award of the degree of Master of Computer Application is an authentic record of our original work.

The report has not been submitted for the award of any degree of this university or any other university.

Date : 04-03-2022

Place: Angamaly

FEDERAL INSTITUTE OF SCIENCE AND TECHNOLOGY (FISAT)®

ANGAMALY, ERNAKULAM-683577

DEPARTMENT OF COMPUTER APPLICATIONS

CERTIFICATE

This is to certify that the project report titled **"Suicide Rate Prediction"** submitted by **Lakshmi C** towards partial fulfillment of the requirements for the award of the degree of Master of Computer Applications is a record of bonafide work carried out by them during the year 2022.

Project Guide **Head of the Department**

Submitted for the viva-voice held on ..... at .....

Examiner1 :

Examiner2 :

ACKNOWLEDGEMENT

Gratitude is a feeling which is more eloquent than words, more silent than silence. To complete this project work I needed the direction, assistance and cooperation of various individuals, which is received in abundance with the grace of God.

I hereby express our deep sense of gratitude to **Dr. Manoge George**, Principal of FISAT and **Dr. C Sheela**, Vice principal of FISAT, for allowing us to utilize all the facilities of the college.

Our sincere thanks to **Dr. Deepa Mary Mathew**, Head of the department of Computer Applications FISAT and scrum master and our Internal guide for this project **Dr.Rakhi Venugopal and Ms.Sona Mary Louis** for giving valuable guidance, constructive suggestions and comment during my project work. I also express my boundless gratitude to all the lab faculty members for their guidance.

Finally I wish to express a whole heart-ed thanks to my parents, friends and well-wishers who extended their help in one way or other in preparation of my project. Besides all, I thank GOD for everything.

ABSTRACT

In this world everyone has set of their own problems. It is really depend on how they are emotionally strong. Support from their family and friends also play an important role to prevent them to make such a bad decision. Youth are more in world and youth are back bone to the nation. They can change the future of the society with their well being and courageous behavior . But most of the youngsters are being affected by suicide. Day by day increasing the suicide rates worldwide. The suicide rate prediction is vital factor to the government. The citizens decide the strength of the country. So The life of each person is important. The project concentrated on predicting the suicide Rate at a particular Country. It can be helpful to take the preventive action to reduce suicide risk rate based on some criteria. It can predict the suicide risk rate more accurate than the existing system. The suicide rate Prediction is very important factor for the government because if we already predict potential suicidal conditions through surveys we can try to stop them. In this model, I have used regression techniques like Decision Tree , Random Forest and XGBoost and to analyzing significant patterns features that result in increase of suicide rates globally. The project is implemented with three regression algorithms and compare then in order to finalize which regression algorithm will give better accuracy and predict the suicide rate.

Contents

Chapter 1

INTRODUCTION

Worldwide, suicide rate is considered one of the most significant issue. Suicide is a serious public health problem. The World Health Organization (WHO) estimates that every year close to 800 000 people take their own life, which is one person every 40 seconds and there are many more people who attempt suicide. With each passing year, the number of suicide is getting increased. Because of that reason, this project is carried out to predict the suicide rate to reduce the suicide risk rate by using machine learning algorithms. The suicide rate prediction is very important factor for the government because if we already predict potential suicidal conditions through surveys we can try to stop them. The passing years the suicide rate is increased day by day. India recorded 153,052 suicides an average of 418 daily in 2020, according to the latest central government data. Every country recorded suicide rate ,the increasing of suicide is really affecting the country badly. It is important that take actions to reduce the suicide cases. If we can predict it early it is helpful to the government to take actions and reduce the suicide risks.

Chapter 2

PROOF OF CONCEPT

2.1 Existing System

Existing suicide rate prediction models appeared sufficiently accurate to be cost effective . Suicide rates between nations can vary by more than an order of magnitude.The existing systems are in a statistical approaches . And some Paper works are mentioning about the suicide rates .There is no systems better to predict the suicide rates . The data are collecting by the year endings . It is not a good approach . The existing systems are predicting the suicide rate using machine learning algorithms with giving low accuracy rate . The research paper give a table that shows 4 out of 6 algorithms are giving the accuracy in below 5

2.2 Proposed System

The existing systems are Predict suicide rate without enough features that relationship with suicide count. And not analysis features importance for predict suicide.A small number of machine learning models still apply for predict the suicide rate.Used algorithms are with very low accuracy . The proposed system that predict the suicide rate used by the Random Forest algorithm. The XGBoost algorithm is used to increase the performance and speed. Compare the other ma-

chine learning algorithms, Random forest regression and Decision tree regression algorithm with XGBoost.The performance and speed is increased well. It can be helpful to take the preventing action to reduce suicide risk rate and also it can predict the suicide risk rate more accurate than the existing system by applying Algorithms in machine learning .

2.3 Objectives

The objective of this project is to predict the suicide rates at a particular country using Machine Learning algorithms and to analyzing significant patterns features that result in increase of suicide rates globally. The project is implemented with three regression algorithms in order to finalize which regression algorithm will give better accuracy.

Chapter 3

SCRUM MEETINGS

On 17-11-2021

On this day started my project discussion that was accepted on 15th December 2021. My project topic is ”Suicide Rate Prediction”.

On 20-12-2021

On this day we discussed about the basic requirements needed for the project. Then I decided to do the project in python using Spyder.

On 22-09-2021

This day we started learning Machine Learning algorithms. Also searched some videos and research papers related to the project for getting more informations.

On 10-01-2022

First Review

On 17-01-2022

Started ML model development

On 19-01-2022

Model Evaluation

On 31-01-2022

Started developing of UI

On 23-02-2022

Final Presentation

On 25-02-2022

Started documentation

Chapter 4

IMPLEMENTATION

This project aims to develop a website for Suicide Rate Prediction. The Suicide rate is predicted using the data like Country,year,gender,age group,generation,gdp for year ,gdp per capita,population and suicide count.A Machine learning algorithms is used here for predicting the Suicide rate. The algorithm used here is Random Forest Regression algorithm. This is an algorithm which ensembles the less predictive model to produce better predictive models. It aggregates the base model to create a large model. The features are sampled and passed to trees without replacement to obtain the highly uncorrelated decision trees. To select the best split it is required to have less correlation between the trees. The main concept that makes random forest different from the decision tree is aggregated uncorrelated trees. Suicide Rate Predictionis an user friendly website.The project uses spyder to develop the data science code and Flask server as framework to connect the model code and the UI .The back-end is python and server is python flask.

ALGORITHM

Random Forest Regression

This is an algorithm which ensembles the less predictive model to produce better predictive models. It aggregates the base model to create a large model. The features are sampled and passed to trees without replacement to obtain the highly uncorrelated decision trees. To select the best split it is required to have less correlation between the trees. The main concept that makes random forest different from the decision tree is aggregated uncorrelated trees.

XGBoost Regression

XGBoost is one of the most popular machine learning algorithms these days. XGBoost stands for eXtreme Gradient Boosting. Regardless of the type of prediction task at hand; regression or classification. XGBoost is an implementation of gradient boosted decision trees designed for speed and performance.

Decision Tree Regression

Decision tree builds regression or classification models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes.

	ML_Model	Train Accuracy	Test Accuracy	Train RMSE	Test RMSE
2	XGBoost Regression	0.994	0.988	0.096	0.133
1	Random Forest	0.988	0.981	0.133	0.170
0	Decision Tree	0.967	0.951	0.220	0.274

4.1 System Architecture

The use case diagram that describes the operation of the system .

4.2 Dataset

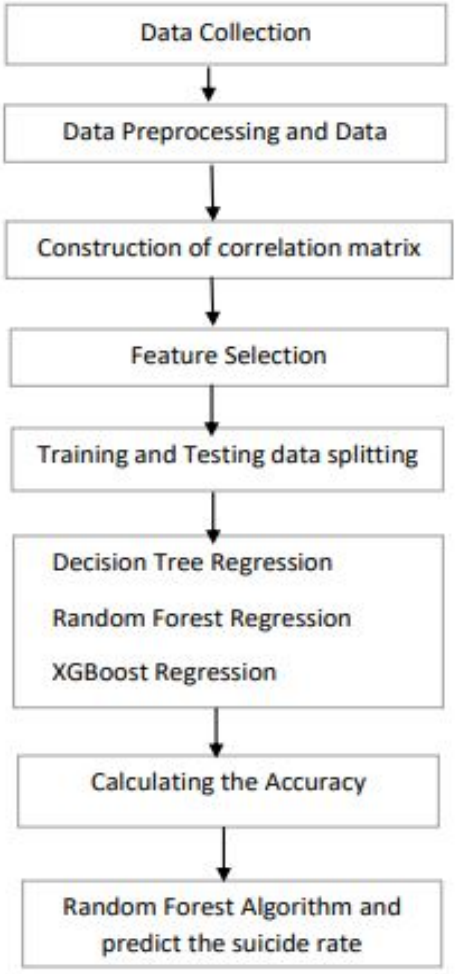
The data requirements is very high for the project.We get a dataset that contains the component like:-

• Country
• year
• sex
• age group
• suicide count
• Population
• country - year
• HDI for year
• gdp for year
• gdp per capita
• generation
• suicide rate

4.3 Modules

4.3.1 Data Preprocessing

Explore and analysis the data set. Firstly read the data set. Then analyse the data set using the describe method and check if any null values are present our data set using “isnull()” function.


Data Collection
Data Preprocessing and Data
Construction of correlation matrix
Feature Selection
Training and Testing data splitting
Decision Tree Regression
Random Forest Regression
XGBoost Regression
Calculating the Accuracy
Random Forest Algorithm and predict the suicide rate

4.3.2 Data Transformation

Before run the models, we have to convert all categorical values (text values) to numerical values.
In our data set, we can see that we have fields country, year, gender , age group and generation which is categorical. So we have to convert this field into numerical. It do using the label encoder function .

4.3.3 Splitting Data

We will use 80in our system two splitting are done. X and y splitting to predict the weight and w, z values are splitting to predict the BMI. The splitting is used to do the train test split() function.

4.3.4 Model Training

After the splitting the data set we are apply to the data model. Here I use to predict the suicide the data model is RandomForestRegressor() model .

4.3.5 Model Evaluation and Testing

Score and Root Mean Square Error(RMSE) are used to see the accuracy of the model.

Chapter 5

RESULT ANALYSIS

The Suicide Rate Prediction helps the persons to know their countries suicide rate when giving the inputs .

Chapter 6

CONCLUSION AND FUTURE SCOPE

6.1 Conclusion

The project predict the suicide rate in the world by country wise using machine Learning algorithms.The suicide rate is depending the GDP of a country . the GDP increases the suicide rate is decreased. The suicide rate prediction is vital factor to the government. By applying the Random Forest Regression algorithm it gives more accuracy and Xgboost regression algorithm it increases the speed and performace of the model.

6.2 Future Scope

Train the model with the data regarding countries which were not included in the dataset. Hence predict the suicide rate in those countries for all unseen data. To develop a system for the rectification of varied accuracy of the model in different platforms. .

Chapter 7

SOURCE CODE

```
    # Decision tree regression model
from sklearn.tree import DecisionTreeRegressor
tree = DecisionTreeRegressor(max_depth=9)
tree.fit(x_train, y_train)
y_test_tree = tree.predict(x_test)
y_train_tree = tree.predict(x_train)
print(y_test_tree)
print(y_train_tree)
acc_train_tree = tree.score(x_train, y_train)
acc_test_tree = tree.score(x_test, y_test)
rmse_train_tree = np.sqrt(mean_squared_error(y_train, y_train_tree))
rmse_test_tree = np.sqrt(mean_squared_error(y_test, y_test_tree))
print("Decision Tree Regression Algorithm")
print("")
print("Decision Tree: Accuracy on training Data:{:.3f}".format(acc_train_tree))
print("Decision Tree: Accuracy on testing Data:{:.3f}".format(acc_test_tree))
print("\nDecision Tree: The RMSE of training Data:{:.3f}".format(rmse_train_tree))
print("Decision Tree: The RMSE of testing Data:{:.3f}".format(rmse_train_tree))
```

Figure 7.1: Deision Tree Regression code

```
# Random forest regression model
from sklearn.ensemble import RandomForestRegressor
forest = RandomForestRegressor(max_depth=9)
forest.fit(x_train, y_train)
y_test_forest = forest.predict(x_test)
y_train_forest = forest.predict(x_train)
print(y_test_forest)
print(y_train_forest)
# computing the accuracy of the model performance
acc_train_forest = forest.score(x_train, y_train)
acc_test_forest = forest.score(x_test, y_test)
# computing the root mean square error(RMSE)
rmse_train_forest = np.sqrt(mean_squared_error(y_train, y_train_forest))
rmse_test_forest = np.sqrt(mean_squared_error(y_test, y_test_forest))
print("Random Forest Regression Algorithm")
print("")
print("Random Forest: Accuracy on training Data:{:.3f}".format(acc_train_forest))
print("Random Forest: Accuracy on testing Data:{:.3f}".format(acc_test_forest))
print("\nRandom Forest: The RMSE of training Data:{:.3f}".format(rmse_train_forest))
print("Random Forest: The RMSE of testing Data:{:.3f}".format(rmse_train_forest))
```

Figure 7.2: Random Forest Regression code

```
# XGBoost regression model
from xgboost import XGBRegressor
xgb = XGBRegressor(learnging_rate=0.2, max_depth=4)
xgb.fit(x_train, y_train)
y_test_xgb = xgb.predict(x_test)
y_train_xgb = xgb.predict(x_train)
print(y_test_xgb)
print(y_train_xgb)
acc_train_xgb = xgb.score(x_train, y_train)
acc_test_xgb = xgb.score(x_test, y_test)
rmse_train_xgb = np.sqrt(mean_squared_error(y_train, y_train_xgb))
rmse_test_xgb = np.sqrt(mean_squared_error(y_test, y_test_xgb))
print("XGBoost Regression: Accuracy on training Data:{:.3f}".format(acc_train_xgb))
print("XGBoost Regression: Accuracy on testing Data:{:.3f}".format(acc_test_xgb))
print("\nXGBoost Regression: The RMSE of training Data:{:.3f}".format(rmse_train_xgb))
print("XGBoost Regression: The RMSE of testing Data:{:.3f}".format(rmse_train_xgb))
```

Figure 7.3: XGBoost Regression code

```
app = Flask(__name__)

@app.route("/")
def index():
    return render_template("index.html")

@app.route("/predict")
def predict():
    return render_template("prediction.html")

@app.route("/predict", methods=["POST"])
def predicts():

        country =int(request.form["country"])
        #country1 = str(country)
        year = int(request.form["year"])
        gender = int(request.form["gender"])
        age_group = int(request.form["age_group"])


        population = int(request.form["population"])
        gdp_for_year = int(request.form["gdp_for_year"])
        gdp_per_capita = int(request.form["gdp_per_capita"])
        generation = int(request.form["generation"])

       # country = label_country.transform([int(country)])[0]
        #year = label_year.transform([int(year)])[0]
        suicide_count = int(request.form["suicide_count"])

        year = int(request.form["year"])
        arr = np.array([[country,year,gender,age_group,suicide_count,population,gdp_for_year,gdp_per_capita,generation]])
       # output = round(prediction[0],
        prediction = model.predict(arr)
        return render_template('Result.html',prediction_text="Suicide Rate is {} ".format(prediction))
```

Figure 7.4: Suicide Rate Prediction code

```
46
47      <title>PREDICTION!INDEX</title>
48  </head>
49
50  <body>
51
52      <div class="container d-flex main-content">
53          <div class="row card p-4">
54              <div class="col-12 text-center h1 mt-4">
55                  SUICIDE RATE PREDICTION
56              </div>
57
58              <div class="col-12 d-flex buttons-group">
59                  <button type="button" onclick="location.href = '/predict'"
60                      class="main-btn btn btn-outline-primary">PREDICTION</button>
61                  <button type="button" onclick="fireAbout()"
62                      class="main-btn btn btn-outline-success">ABOUT</button>
63              </div>
64
65          </div>
66      </div>
67
```

Figure 7.5: Home page html code

```
    <title>PREDICTION!RESULT</title>
</head>

<body>

    <div class="container d-flex main-content">
        <div class="row card p-4">
            <div class="col-12 text-center h1 mt-4">
                {{ prediction_text }}
            </div>

            <div class="col-12 d-flex buttons-group">
                <button type="button" onclick="location.href = '/predict'"
                    class="main-btn btn btn-outline-primary">GO BACk TO PREDICTION</button>


            </div>

        </div>
    </div>
```

Figure 7.6: Result page html code

Chapter 8

SCREEN SHOTS

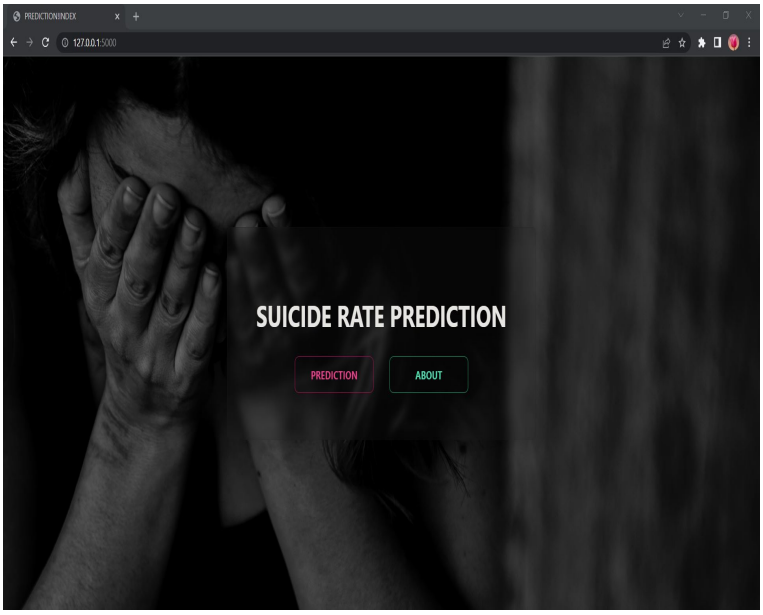


Figure 8.1: Home Page

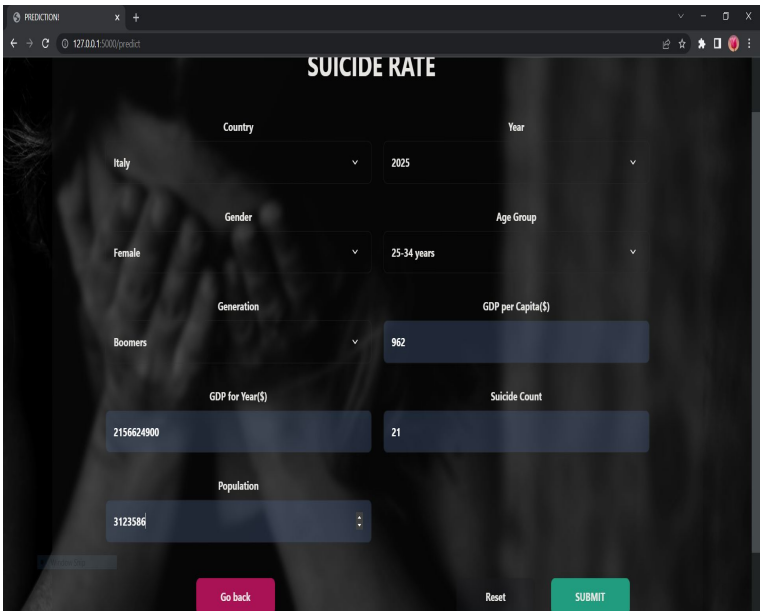


Figure 8.2: Suicide Rate Prediction page

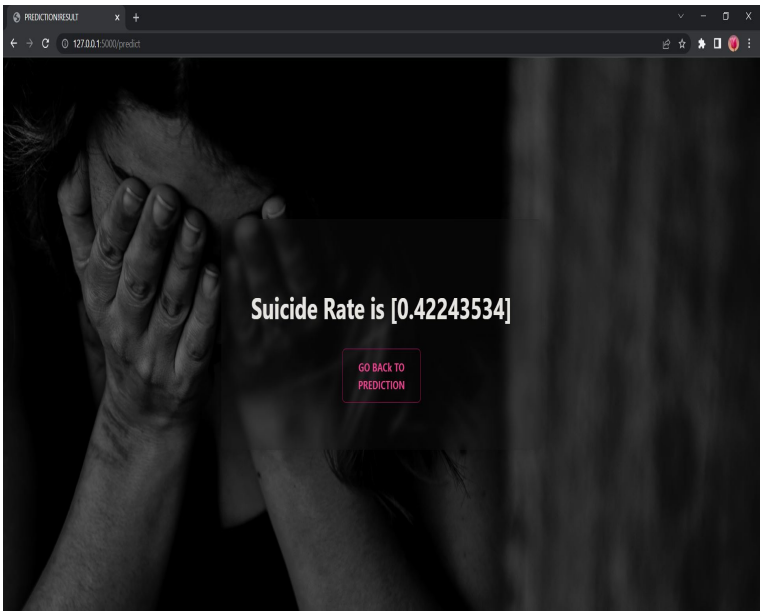


Figure 8.3: Result Page

Chapter 9

REFERENCES

[1] www.youtube.com
[2]www.google.com
[3] http://www.dynamicpublisher.org/gallery/62-ijsrr-d-1769.f.pdf
[4] International Journal of Scientific Research and Review Volume 8, Issue 4, 2019 , Design of Machine Learning Based Suicide Rate Prediction System
[5] https://www.kaggle.com/chingchunyeh/suicide-rates-overview-1985-to-2016
[6]International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249-8958, Volume-10 Issue-1, October 2020,Machine Learning Based Prediction of Suicide Probability