# Suicide Prediction Model

Edvardsen

*IE, IIR, NTNU*

Ålesund, Norway

[joakied@ntnu.no](mailto:joakied@ntnu.no)

Eide

*IE, IIR, NTNU*

Ålesund, Norway

[torstei@ntnu.no](mailto:torstei@ntnu.no)

Andrei Cristea

*IE, IIR, NTNU*

Ålesund, Norway

[eduarac@ntnu.no](mailto:eduarch@ntnu.no)

**Suicide across the globe is a big concern and people are continuously investigating the matter. In this article we develop a suicide prediction model based on data collected across the world from 1985 to 2016 with the end goal of finding causes to why suicides takes place so prevention method can be developed.**

## Introduction

The following report is meant to provide an insight into global suicide rates. One person commits suicide every 40 seconds. Suicide is a worldwide issue that happens at all stages of life. To prevent suicide and suicide attempts, effective and evidence-based interventions may be adopted at the community, sub-population, and individual levels. There is evidence that for every adult who died by suicide, more than 20 others attempted suicide.

Data collection can help us combat complex issues such as suicide, such that prevention methods can be crafted. For these methods to be crafted; coordination and collaboration across fields must materialize.

## Methods

### Reviewing data

When a new dataset is presented, it is important to review the data within the dataset. When we started viewing the dataset, we reviewed python code displaying statistics from the dataset [1]. It’s important to visualize the data to get a better understanding of what you’re working with. It’s also interesting to see all the factors that plays a role of the outcome.

If we look at figure 1, we can see a significantly higher suicide rate for males. In fact, if we look a little closer, the suicide rate is almost three times higher for males. This number has been pretty much been consistent from 1990 to 2016.

Chart, line chart

Description automatically generated

Figure 1, number of suicides from year 1985 to 2015, categorized in male and female

Chart, line chart

Description automatically generated

Figure 2, suicide rate from 1985 to 2915 grouped by age

Chart, line chart

Description automatically generated

Figure 3, suicide rate from 1985 to 2015 group be generation

We looked at features such as age (figure 2) and found that the number of older people taking their own lives are higher than younger, showing a clear pattern that the older people are the more likely to take their own lives.

When checking the correlation between generations and suicide numbers (figure 3) we also saw that the suicide rates differed from each. Showing that for example the G.I generations spiked in suicide rates about the same time as the worldwide depression right before WW2. From this we can see a relation between generations and suicide rates. When designing our model, it’s important to take notes of such spikes and abnormal tendencies to make sure our model is not overfitted.

We also looked at rates from each country. Here we discovered that the suicide rates increased in countries with greater alcohol consumption, depression and so on. There is also a clear connection between the suicide rates in the 1990s and the disintegration of the soviet union.Chart, line chart

Description automatically generated

Figure , Suicide rates from top 10 countries from 1985 to 2016

### The model

Our model consists of nine input features and one target feature. Since we know the outcome, this is known as a supervised model. Furthermore, we split our data into test, train and validation sets using sklearn’s “train\_test\_split” function which randomly shuffles the data for us and splits it into the sets we needed. The data was split in a 4:1 ration, meaning 80% of the data was used for training the model, and the 20% leftovers were used to validate it.

After the data was successfully split into train, test, and validation sets, we created our model. The model we used was XGB Regression model. We opted for a regression model over classification since it’s hard and doesn’t make sense to classify our problem.

## Results

Our model ended up with a 99.4% accuracy during training and 98.8% during testing. Meaning it performed well both during and after training.

## Discussion

Suicide is a well-known issue and needs to be taken seriously. From the dataset we can see that the problem is a global problem, however some countries are more affected than others. Non the less measures should be made to help lower the suicide rate and stop the current growth.

As mentioned earlier in this report, location is a great factor and is also used as a feature in our model. This can be caused by the development of infrastructure in the area. Well developed countries might have more resources to handle patient that suffer from depression. If there is a big correlation between suicide rate and the number of resources in the area, this can be a pointer to some actions that can be made to lower the suicide rate.

## Link to code

<https://colab.research.google.com/drive/1YJhe36YC_gcEbm3FT6oHqSEJFFmNE9Fl>

# References

|  |  |
| --- | --- |
| [1] | C.-C. Yeh, "Suicide Rates Overview 1985 to 2016," 1 1 2019. [Online]. Available: https://www.kaggle.com/code/chingchunyeh/suicide-rates-overview-1985-to-2016. [Accessed 26 9 2022]. |
| [2] | K. K. S. S. K. Vishnu Kumar, "Mental health research," 01 06 2022. [Online]. Available: https://learn-eu-central-1-prod-fleet01-xythos.content.blackboardcdn.com/5def77a38a2f7/18731250?X-Blackboard-Expiration=1664193600000&X-Blackboard-Signature=BBI3YvUvjiUCJR3AGE1p1H5tWNLeSsBHxZnPfsiV4Lo%3D&X-Blackboard-Client-Id=303508&response-cache-contro. [Accessed 26 09 2022]. |
| [3] | shreyagopai, "Google Colab," Google, 10 5 2020. [Online]. Available: https://colab.research.google.com/github/shreyagopal/Suicide-Rate-Prediction-with-Machine-Learning/blob/master/2.%20Suicide%20Rate%20Prediction%20with%20Machine%20Learning.ipynb#scrollTo=OwtQg-D0PoSa. [Accessed 30 08 2022]. |