EAI Endorsed Transactions
on Pervasive Health and Technology
Research Article EAI.EU

Suicidal Behavior Prediction and Socioeconomic Suicide Indicators

Muhammad Nouman[1*], Kareem Ullah[2] and Muhammad Azam[3]

[1]University of Agriculture, Faisalabad, Pakistan
[2]University of Agriculture, Faisalabad, Pakistan
[3]University of Agriculture, Faisalabad, Pakistan

Abstract

INTRODUCTION: According to the WHO (World Health Organization), nearly 0.8 million people commit suicide each year, with more than 20 suicide attempts for every self-immolation. Suicidal behaviors have a profound effect on communities, societies, families, friends, and colleagues. After the recognition of public health as a priority by the WHO, various studies are being conducted to prevent it.

OBJECTIVES: The investigation's goals were to improve understanding of suicide by identifying socioeconomic indicators correlated with rising suicide rates among divergent legions globally and to develop a prediction model for those who are at a higher risk of suicide by using different predictors of suicide such as tension, depression, anxiety, and so on.

METHODS: We used a variety of data mining techniques to create a prediction model for suicide, including Logistic Regression, Multilayer Perceptron, Polynomial/Gaussian/Sigmoid SVM, Decision Tree, and K-Nearest Neighbors. For identifying socioeconomic suicide indicators, we used various descriptive and exploratory analysis techniques such as mean, regression, and correlation.

RESULTS: Classification through the Gaussian Kernel - SVM has been shown to have the best results relative to others. Results also stated that many countries saw a decrease in suicide rates between 2006 and 2015, compared to 1996 to 2005. The highest concentrations have been reported in Europe, while the lower has been observed in South America.

CONCLUSION: Things are improving, at least according to the statistics. The performance of Gaussian Kernel-SVM has been demonstrated to be superior to the other algorithms for suicide prediction. Data on suicide and suicide attempts are imprecise and difficult to gather. Suicide and suicide attempt monitoring, and surveillance must be improved for suicide prevention initiatives to be effective.

Keywords: Suicidal Behavior, Risk Factors, Socioeconomic Indicators

Received on 02 August 2021, accepted on 15 November 2022, published on 02 December 2022

doi: 10.4108/eetpht.v8i5.3175

1. Introduction

Attempted suicide is defined as an end-of-life impulse and cognition, ranging from the conviction that life is not worth living to intense delirious preoccupation with self-destruction, to deliberate plans to kill oneself. As a result, the immense scientific and practical benefit lies in the ability to understand the power, ubiquity, and characteristics of the concept, which can predict the likelihood of suicide attempts later **[1].** One of several main known risk factors is a mental illness, with 80 percent of those who attempted or died of suicide citing some sort of mental disorder. Almost all people with mental illness, however, do not have suicidal thoughts.

The social, economic, and demographic variables of individuals have been studied mainly through current attempts to identify and recognize suicidal elements. This research also explores the question of the economic status of mental health. It is a big economic problem with millions of people and billions of euros a year. Over the last decade, the major economic influence of mental health issues has become increasingly evident and acknowledged **[2]**.
Suicide is a major public health problem all over the world. This is among the top twenty leading causes of death. Suicide kills more people than malaria, chest cancer, conflict, and murder combined. The WHO has designated suicide fatality as a worldwide priority and has included it as an index in the United Nations Sustainable Development Goals (SDGs).

*Corresponding author. Email: m.nouman909@gmail.com

Committing suicide is the second-largest reason for demise between 15 to 29 years of age [3]. This is the 10th major cause of death in the USA. There are socioeconomic risk factors as well as other risk factors such as depression, helplessness, anxiety, and so on. With a suicide mortality rate of 2.9 in 2016, over 5,500 people died by suicide in 2016. Experts estimate that the death toll is somewhere between the two figures, but the truth remains unknown [4].

Thoughts of loneliness and futility are also thought patterns associated with suicide activities. This could lead to severe academic disability [5]. Symptoms that can be easily understood in daily life include the failure to concentrate, to reason rationally, or even to make small decisions; other apparent symptoms include difficulties in delivering the requisite tasks, self-harm, withdrawal from regular relationships or depression, and elevated absenteeism in education [6].

The Global Burden of Disease is a major global report published in The Lancet on the causes of death and illness [7]. The GBD report shows the number of deaths per year for each cause. Suicide kills roughly twice as many people as assassination. Suicide is more common than murder in most countries, often 10 to 20 times more common [8].

It is critical to take a rigorous and organized approach to prevent suicide attempts and ensuring that the tragedy of suicide does not continue to cost lives. Suicide is a particularly complicated matter that causes millions of persons around the globe to suffer every year [8]. The goal of this research is to spark an informed and open discussion about suicide prevention. If you are having suicidal thoughts, you can get emergency help by visiting the website Suicide.org [9]. As per the WHO and the Global Burden of Disease report, nearly 0.8 million people go dead of suicide each year. It is one human in 40 seconds. Suicides may be avoided by prompt, evidence-based approaches. Due to the stigma related to suicide the fact that it is banned in certain countries, this number is often likely to be exaggerated, with some suicides classified as accidental accidents.

1.1 Problem Statement

The problem addressed in this study was titled "Suicidal Behavior Prediction and Socioeconomic Suicide Indicators." Every year, there are far more suicide attempts than suicide deaths. Previous suicide attempts are the single most significant risk factor for suicide in the general population. Suicide remains a serious public health problem; however, suicide can be avoided through timely, evidence-based, and sometimes cost-effective approaches. Suicidal behavior can occur for a variety of reasons and predicting suicidal behavior is difficult. The current study examines socioeconomic indicators of the increase in suicide rates across different regions and develops a prediction model for those who are at a higher risk of suicide by analyzing different predictors of suicide such as tension, depression, anxiety, and so on using various data mining techniques.

1.2 Objectives

- To identify indicators related to the escalating suicide rates amongst divergent legions across the socio-economic spectrum.
- To develop a predictive model for those at higher risk of suicide by examining various suicide predictors such as stress, helplessness, anxiety, and so on.

The remaining of this research article is organized as follows. Methods are covered in Section 2. Section 3 describes the findings of our research as well as a discussion of them. Section 4 summarizes our findings and future research directions.

2. Overview of the Framework

The procedure for predicting suicidal behavior and socioeconomic aspects of suicidal behavior is depicted in Figure 1, beginning with pre-processing method-level datasets and ending with predicting suicidal behavior and socioeconomic aspects of suicidal behavior. Furthermore, the pre-processed datasets are utilized to evaluate the classifier's overall performance [10]. To begin the pre-processing of the dataset, related features from the dataset are chosen. The purpose of this type of feature extraction is to choose solely valuable features that are strongly associated with the target class. Features that are no longer needed or relevant are removed. After that, outlier extraction is performed. Outliers are metrics that are located outside of the main data cluster (s). Outliers in datasets can happen because of evaluation variations or because of experimental error [11].

Outliers were erased during the study's previous step because they could impair model performance. Datasets that have been cleaned and pre-processed are used for inference, correlation, regression, and explanatory analysis [11]. Dataset was divided into a proportion of 80 percent for training and 20% for testing of classifiers. The motive for model training with an 80-20 ratio is to guarantee that the models learn effectively and separately through cross-validation subsampling, which helps to eliminate biases. Following that, models are evaluated to predict suicidal behavior using the remaining 20% of the data, and the mean classifier performance is calculated using the testing data [12].

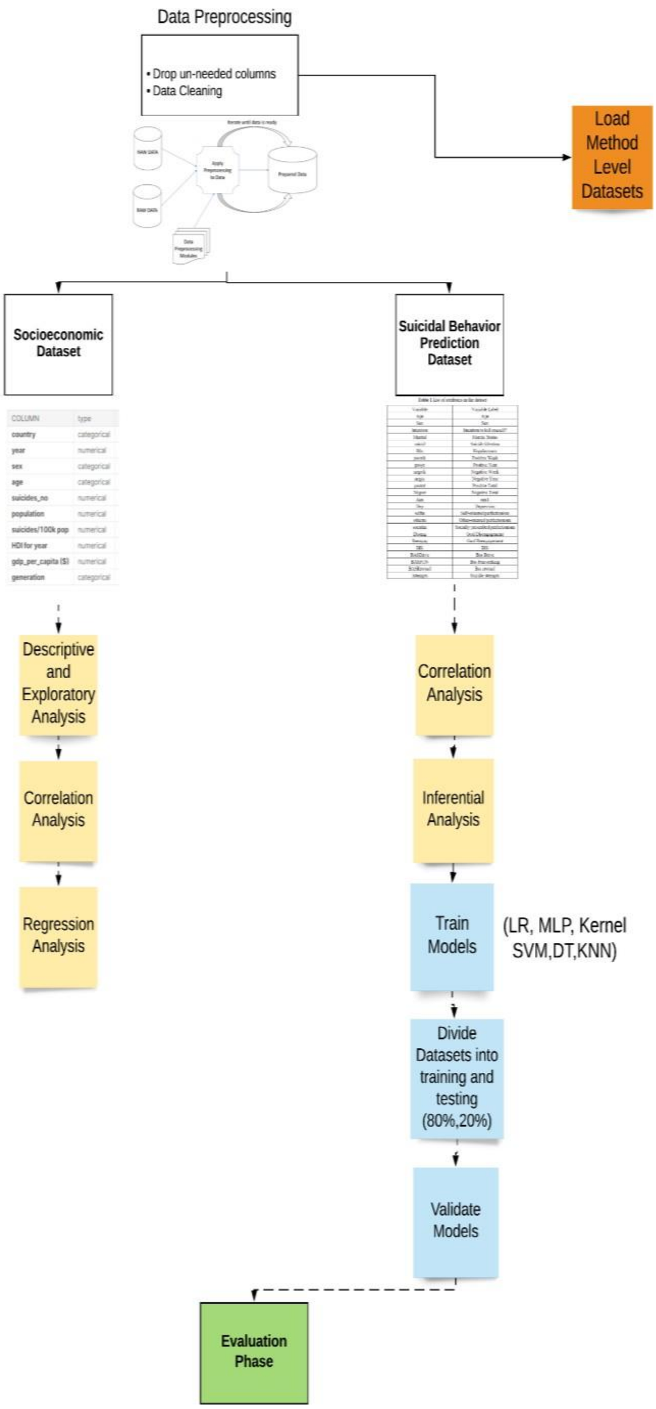

Figure 1: Overview of the Research Framework

3. Methods

During the paper's preparation, we aimed to select the approach that is best suited to the issue at hand. The nature of research is dependent on the accuracy of configuration as well as the productivity of the review technique while keeping the nature and goals of the review in mind [13]. Despite this, observational verification of the proposed hypotheses is required, first and foremost, on the reliable estimation of the factors of extreme interest, and secondly, on the methods and procedure for concluding such measurements.
This required:

- Selection of a suitable sample.
- Selection of proper tools that could be gainfully utilized for solid measures; and
- Selection of appropriate statistical techniques for analyzing the data.

Different datasets were used for our purposed research.

3.1 Socioeconomic Dataset

This dataset was compiled from the WHO mortality database [13] and the World Databank database (World Bank, 2018) to find correlations to rising suicide rates among divergent legions worldwide across the socioeconomic spectrum.
The database contains a list of suicides for each nation each year, sorted by different age groups, as well as the overall population of each community, allowing us to only quantify the rough statistics per 100,000 people. This dataset contains suicide data from 101 countries and spans 32 years by comparing socioeconomic data with annual and regional suicide rates. The dataset includes roughly half of the world's countries, so it is far from complete, but it is an excellent resource for mapping and visualizing the data [14]. We can categorize the categorical and numerical data based on the dataset as follows:

Table 1: Attributives of Socioeconomic Dataset

Features	*Data type*	*Description*
Country	Categorial	Country name
Year	Numerical	Year to date
Gender	Categorial	Gender of suicide victims
Age	Categorial	The age range of suicide victim
Suicide no	Numerical	No of suicides
Population	Numerical	Population of country
Suicides/100k	Numerical	Suicides per 100k population
HDI for year	Numerical	Human Development for year
Gdp_per_capita ($)	Numerical	Gross Domestic Product
Generation	Categorial	Population born between certain years

3.2 Suicidal Behavior Prediction Dataset

The dataset contains information on parasuicidal patients, which can be used to predict suicide behavior (aged above 15). The study used data mining to analyze the data and created a model for predicting suicidal attitudes in an individual by analyzing different predictors of suicidal behaviors such as stress, helplessness, and so on.

The data is derived from the UK data archive [15]. The data set includes information on 267 patients from Edinburgh Liaison Psychiatry who attempted suicide between the ages of 16 and older [16]. The participants completed the various psychological assessments, and the feedback they provided was recorded. Table 2 displays the list of attributes. Suicide behavior was assessed using a subscale of suicide probability comprised of "Beck's Despair Scale," which has been assessing sadness for the next few years, e.g. 'I try to be positive with hope and excitement.'

Table 2: Attributives of Suicidal Behavior Dataset

Variable	*Variable Label*
Age	Age
Gender	Gender
Intention	The intention to kill own self?
Martial	Marital Status
Suicide	Suicide Ideation
BHS	Hopelessness
Pock	Positive week
Poseur	Positive year
New	Negative week
Negar	Negative year
Poston	Post total
Negotiate	Negative total
An	Anx1
Dep	Depression
Self	Self-cantered Perfectionism
Other	Others-adapted Perfectionism
Socialism	Socially approved perfectionism
Design	Goal disengagement
Re-engage	Goal Reengagement
BIS	BIS
Bedrive	BAS Drive
BaFin	BAS Fun-seeking
Barware	BAS reward
Attempts	Suicide attempts

3.3 Machine Learning Models

After modeling each classifier, we compared the results of various learning algorithms applied to datasets. Several classifiers, including Logistic Regression (LR), Multilayer Perceptron (MLP), Polynomial/Gaussian/Sigmoid SVM, K- nearest neighbors (KNN), and Decision Tree (DT), were used in the current study.

i. Logistic regression (LR)

Logistic Regression is a sort of Generalized Linear Model that models a binary variable using a logistic function and any number of independent variables. Logistic regression indicates the probability of action with two alternative outcomes and is therefore ideal for statistical research in which the independent variables can have values of zero or one. The LR model is used to analyze data and explain the link between one or more independent variables and one or more binary variables that are dependent on each other [17].

Logistic regression, like linear regression, represents the data with an equation. To forecast an output value, input data (x) are linearly averaged with weights or coefficient values (referred to as the Greek capital letter Beta) (y). It varies from linear regression in that the output result is a binary value rather than a numeric number [18].

Here's an example of a logistic regression equation:

$$y = e^{\wedge}(b0 + b1 * x) / (1 + e^{\wedge}(b0 + b1 * x))$$

Where y is the expected output, b0 represents the bias or intercept term, and b1 represents the coefficient for a single input value (x). Each column in your input data represents a b coefficient that must be learned from your training data [19].

The stronger the coefficients, the better the formula, which predicts a value near to one for the default class and a value close to zero for the other class. The notion behind maximum-likelihood logistic regression is that a search strategy seeks coefficient values that minimize the difference between the probabilities expected by the model and those in the findings, such as a probability of 1 if the data is of the primary class [20].

ii. Multilayer Perceptron (MLP)

Humans have an astonishingly high level of precision when it comes to identifying trends within available knowledge. You know what you're looking at when you see a car or a bicycle. This is because we have observed how a car and a bicycle look and what distinguishing features they have over time. Artificial neural networks are computing structures that aim to mimic human learning capabilities by using a dynamic architecture like the human nervous system [21].

An MLP is a type of neural network that works in tandem with a feed-forward neural network. It is composed of three layers: a reference layer, a medium layer, and a hidden layer. The input layer receives the interpreted input signal [22]. The output layer oversees tasks such as prediction and categorization. The true computational

the engine of the MLP is made up of an endless number of hidden layers located between the input and output layers. In an MLP, which is analogous to a feed-forward network, data flows forward from the input to the output layer [23].
The neurons of the MLP are trained using the backpropagation learning technique. MLPs are designed to replicate any continuous function and can tackle issues that cannot be solved linearly. MLP's most typical uses include pattern detection, identification, estimate, and approximation. Each neuron in the output and hidden layers performs the following computations [24].

$$o(x) = G\big(b(2) + W(2)h(x)\big)$$
$$h(x) = \varphi(x) = s\big(b(1) + W(1)x\big)$$

iii. Kernel Support Vector Machine (SVM)

Support Vector Machines (SVMs) are a more versatile classification algorithm that can do both linear and non-linear classification [25]. An SVM classifier calculates the optimum match line to optimize the margin between two linearly separable groups. The SVM classifier treats each instance of a class as a vector for the relevant input variables [26]. So,

$$v = (x_1, x_2, \ldots\ldots, x_1)$$

As a result, the data set came in the formation of vector sets and classes:

$$D = \{(x_1, y_1), (x_2, y_2), \ldots\ldots, (x_n, y_n)\}$$

When x is a vector, y is a vector-related group (positive or negative), and n is the number of occurrences.

SVM is distinguished from other classification algorithms in that it selects the decision boundary that minimizes the distance between the nearest data points in all groups. An SVM considers not just a decision boundary, but the most optimal decision boundary. The optimal judgment border is the one with the greatest margin between the two grades nearest points [27].

The basic SVM technique, on the other hand, can be used to determine the decision boundary for linearly separable outcomes. A straight line, on the other hand, cannot be employed as a decision boundary in the event of non-linearities separate results. Kernel SVM, a customized variant of SVM, is utilized instead of plain SVM. The kernel SVM effectively projects non-linearly separable data in lower dimensions to linearly separable data in higher dimensions, assigning data points from different groups to different dimensions. Kernel SVM employs a set of mathematical functions known as the kernel. The kernel's strategy is to take in data and turn it into the format required [28]. Different types of kernel functions are used by various SVM algorithms. These functions might be linear, nonlinear, polynomial, radial basis function (RBF), or sigmoid.

Kernel function can be defined as

$$K|\bar{x}| = \begin{cases} 1, & if\,\|\bar{x}\| \le 1 \\ 0, & otherwise \end{cases}$$

Within the closed ball of radius, one centered at the origin, this equation has a value of one and a value of zero elsewhere [29].

iv. Decision Tree (DT)

A decision tree is a supervised machine learning technique that can perform regression and classification problems. The decision tree algorithm's intuition is simple, but it is extremely powerful. Decision trees define instances by dragging them down from the root tree to a leaf node, with the leaf node displaying the categorization example. Every node in the tree is a case study for any feature, and each edge that descends from the node refers to the correct selections to the testing ground. The decision tree technique generates a node for each attribute in the dataset, with the most important attribute serving as the root node. We initiate at the root node and work our way down the tree, looking for the node that corresponds to our condition or "choice." This technique is repeated until a leaf node containing the decision tree's forecast or result is reached [29].

The choice to make strategic splits has a major impact on the accuracy of the tree. The judgment criterion for grouping and regression trees is distinct. Decision tree algorithms employ a variety of approaches to divide a node into two or more sub-nodes. The establishment of a site increases the homogeneity of the sub-nodes that result. In other words, we can infer that the node's integrity rises as the response variable increases. The decision tree separates the nodes based on all available variables and then selects the division that results in the most uniform sub-nodes [30].

v. K-Nearest Neighbors (KNN)

The K-nearest neighbors (KNN) algorithm is a machine learning algorithm that is supervised. In its most basic form, KNN is incredibly simple to use when performing extremely complex classification tasks. Because it lacks a specialized training process, it is a lazy learning algorithm. Rather it utilizes all the data for training when classifying a new data point or case. KNN is a non-parametric learning algorithm, which means it makes no assumptions about the data. Because most real-world evidence does not adhere to statistical assumptions such as linear separability, uniform distribution, and so on, this is a very useful function [31].
It computes the difference between a new data point and all prior training data points. The gap can be of any size, including Euclidean and Manhattan gaps. It then chooses the data points that are K-nearest to each other, where K

can be any integer. Finally, the data point is assigned to the class that contains most of the K data points **[32]**. Other methods of calculating distance exist, and one approach may be preferred based on the situation. A well-known and popular alternative is the straight-line distance (also known as the Euclidean distance **[33]**.
The main downside of KNN is that it gets much slower as the number of data points rises, making it an inefficient solution in instances when predictions must be produced rapidly. Faster algorithms can also give more consistent classification and regression results **[34]**. KNN, on the other hand, can be beneficial in tackling issues whose solutions rely **[35]** on discovering related artifacts if you have the adequate computational power to manage the data, you're utilizing to make predictions rapidly.

3.4 Metrics of Assessment

In this section, we give an overview of the assessment criteria used to measure the success of the classification models used in this report **[36]**. Many metrics have been applied for the use of many measurement metrics to increase the probability of receiving valuable knowledge on model precision, along with mean efficiency, because of which estimates are more accurate in this section, the notations listed below are utilized. [37]:

Classification accuracy (CA): The ratio **[38]** of the number of accurate forecasts to the overall number of forecasts. Classification accuracy should now be computed as seen:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.1}$$

Precision: The fraction **[39]** of the entire positive predictions which are accurately projected

$$Precision = \frac{TP}{TP + FP} \tag{3.2}$$

Recall (sensitivity): The portion of all currently positive occurrences that were accurately estimated **[40].**

F-measure: The harmonic means of accuracy and recall.**[41]** It shall be determined utilizing the method given:

$$F - score = \frac{2 \times (Precision \times Recall)}{(Precision + Recall)} \tag{3.3}$$

4. Results and Discussion

The research and evaluation were carried out on a personal computer running Windows 10 and equipped with an AMD Radeon graphics processing unit (GPU). The evaluation was created in Anaconda Guide using the Python programming language in a Jupiter notebook.

4.1 Suicide Rates per Country

The use of averages appears to be intuitive because the suicide numbers are proportional to population size: the larger the country, the more suicides occur. What is less evident is that these thresholds, which are often referred to as crude rates, do not make a strong distinction between nations. Suicide rates differ greatly between ethnic groups, so we must first understand a country's current demographic structure.

According to the findings, the global suicide rate is 15.17 per 100,000 people, with the Lithuanian rate being significantly higher: >41 suicides per 100,000 people (annually). There is a significant overrepresentation of EU countries with high rates, with only a few having lower rates. Suicide rates are significantly higher in European countries such as Latvia, Hungary, and Belarus. These findings build on existing evidence from multiple types of research (WHO, 2016) that show that these countries are leading to suicidal deaths.

4.2 Suicide Rates per Continents

We plot population and average suicide rates for each continent, allowing us to learn suicide rates in continents after inferring suicidal deaths by country. The findings show that Europe has the most prominent countries in the data, which skews the global suicide average. The average suicide rate per 100,000 people per continent is calculated as follows.

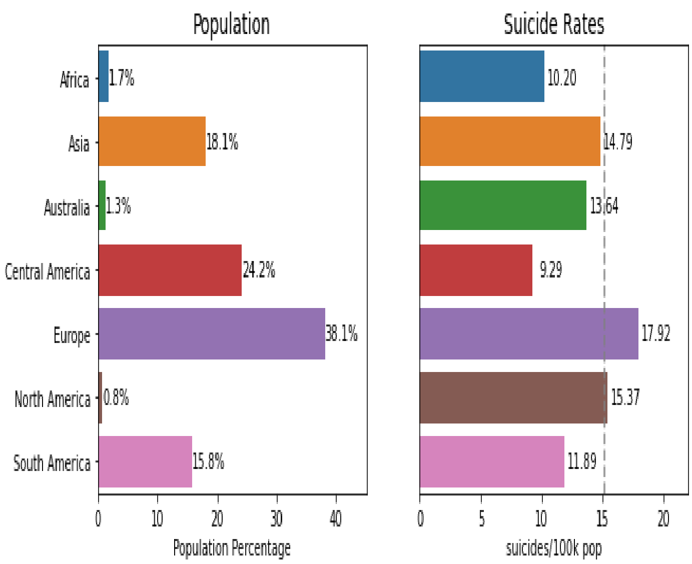

Figure 2: Suicide Rates per Continent

Figure 2 shows that Europe has the highest percentage (38.1%) of suicides among all continents, which is concerning given that Europe has the most first-world countries with the most advanced technology and facilities, as well as the top economic trends (Black et al., 2013). North America outperforms the previous percentage and continues to improve in terms of suicidal trends, with only 0.8 percent. Suicidal crude rates are also rising in Europe (17.92) Over time, the population and mean suicide rates for males and females per continent show that Central American countries have a disproportionately higher percentage of male suicides than females. Males are far more likely than females to attempt suicide. Asian countries have a significantly higher rate of female suicide than other countries.

4.3 Suicide Rates Across the Globe

Suicide rates increased in the early years, peaked in 1995 (figure 3), and have since been decreasing globally. Suicide rates are stable in North and South America, with highs and lows in Asia and Oceania. The highest concentrations have been found in Europe, while the lowest has been found in South America, which is quite surprising.

The graph clearly shows that the global estimated suicide rate from 1985 to 2015 was 13.1 fatalities (per 100,000, annually). In 1995, the maximum suicide rate was 15.3 deaths per 100,000 people. Reduced gradually (by 2 eentrcp5) to 11.5 per 100,000 in 2015. Tariffs are only now returning to pre-90 levels. Because there were few statistics available in the 1980s, it is impossible to say whether the rate was truly representative of the global population at the time.

Things are getting better, at least according to the numbers. Suicide rates have generally decreased. Suicide rates fell in most countries between 2006 and 2015, compared to 1996 to 2005. This pattern can be seen in nearly all age groups for both genders. While previous research has concentrated on clustering countries due to the scope and domain of researchers, these findings highlight specific areas. Using a decade's median suicide rate (for a given nation, age, and gender) reduces the possibility that the observed result is spurious due to randomness. I hope this movement continues into the twenty-first century.

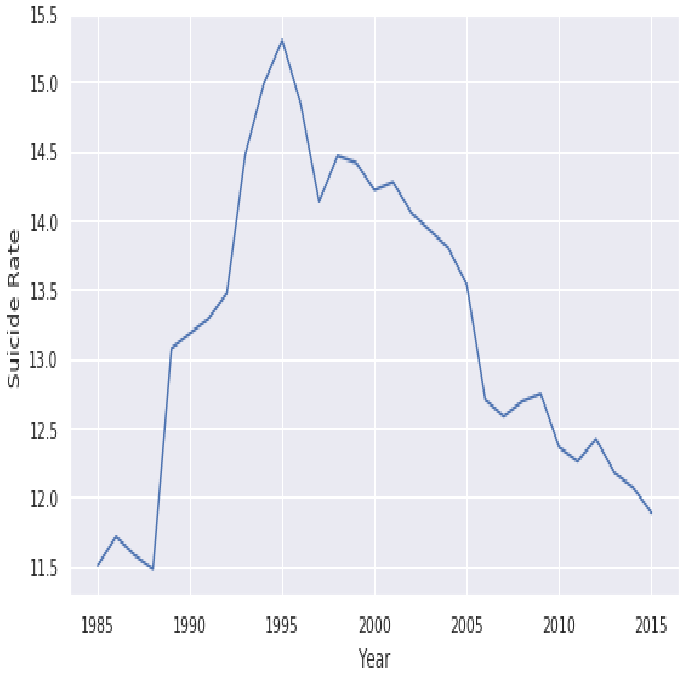

Figure 3: Global Suicide (per 100k)

4.4 Suicide Trends and Age Group

Suicide risk rises with age around the world. Since 1995, the suicide rate for those over the age of 15 has decreased linearly. Since 1990, the suicide rate among people over the age of 75 has dropped by more than half. The suicide rate among children aged 5 to 14 remains relatively stable and low (one per 100,000 people per year).

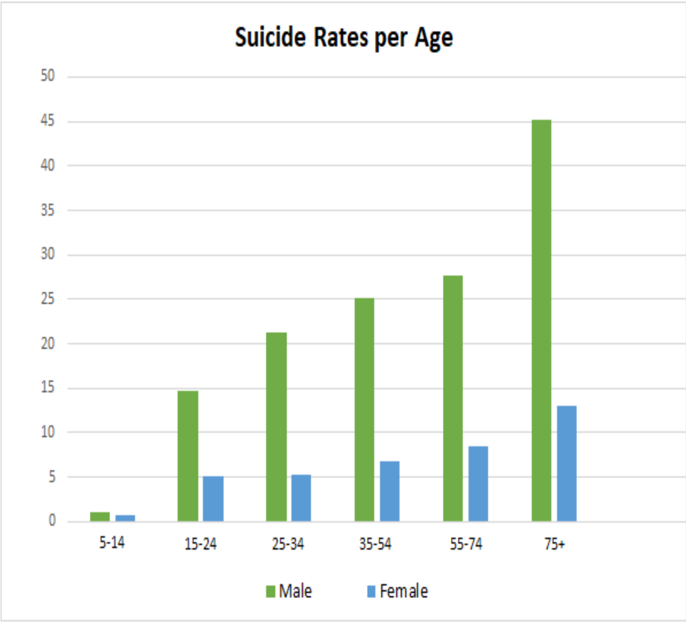

Figure 4: Suicide Rates per Age (Male & Female)

The rate of suicides has been declining mutually in men and women; however, they have been a significant uptick in the ratios of females in age groups 55+ in recent years. Males have a greater risk of suicide compared with women. After a spike in 1995, suicide rates globally have been on a steady decline for all age groups. There is some indication of an uptick in suicides from 2013 in older people. Females in the 5-14 years age group have had a very shallow uptick in suicides for the past few decades. People within the 50 years of 25-75 years have a steady incidence of suicides. People that are 75+ years have a significantly higher incidence of suicides, in men and women. Africa saw a significant increase in man's suicide in 2005-2006, however, there has been a steady decline in suicides for both genders.

However, the trends are very volatile, and these results build on existing evidence of research in WHO (World Health Organization, 2017)

- ✓ Overall, male suicide rates were ~3.5x higher than females.
- ✓ Men's and women's suicide rates reached their peak in 1995, dropping since.
- ✓ This proportion of 3.5:1 (men: women) has stayed essentially stable since the mid-1990s. Though, during the 1980s, this ratio was as low as 2.7:1. (men: women)

4.5 Suicide Trends and Human Development Index (HDI)

The values in figure 5 show that the overall HDI has risen while the suicide rate has steadily decreased. There is an inverse relationship between HDI and suicides, but there is also a lag between when HDI rises, and suicides fall. Because the HDI for the year segment contains approximately 70% of the missing values, the data for recent years, namely 2013, 2014, has been purified, and the states where data is missing for those years have been deleted. It was discovered that there is a 0.43 (weak positive association) relationship between suicide rates and the Human Development Index, implying that suicide rates are very high in countries with high human development indexes.

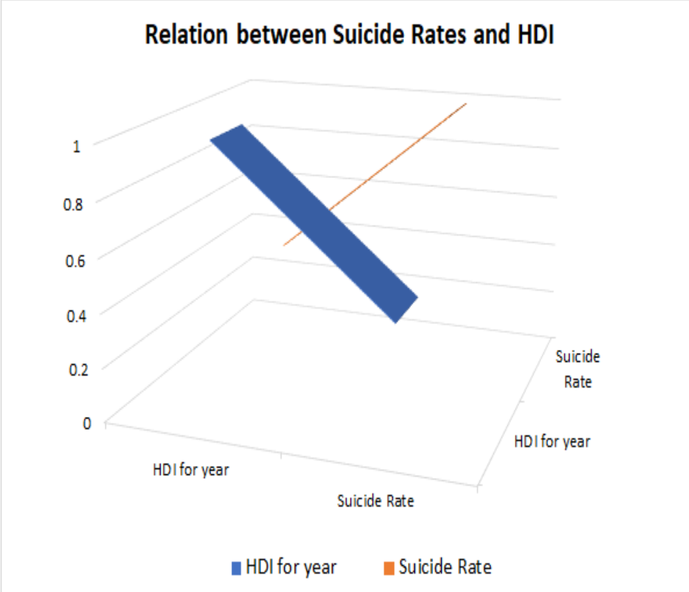

Figure 5: Relation between Suicide Rates and HDI

4.6 Correlation between Socioeconomic Factors

Correlation is a mathematical method for determining how dependent various variables are on one another. I can look at the average association as well as the male-female similarities in the study. To establish a connection, I must first complete two critical tasks:

- Integer value qualitative properties are encoded.
- Statistical standardization.

Age, generation, and gender all play a role in determining suicide rates. Figure 6 depicts all the socioeconomic features as well as all the correlation values between them.

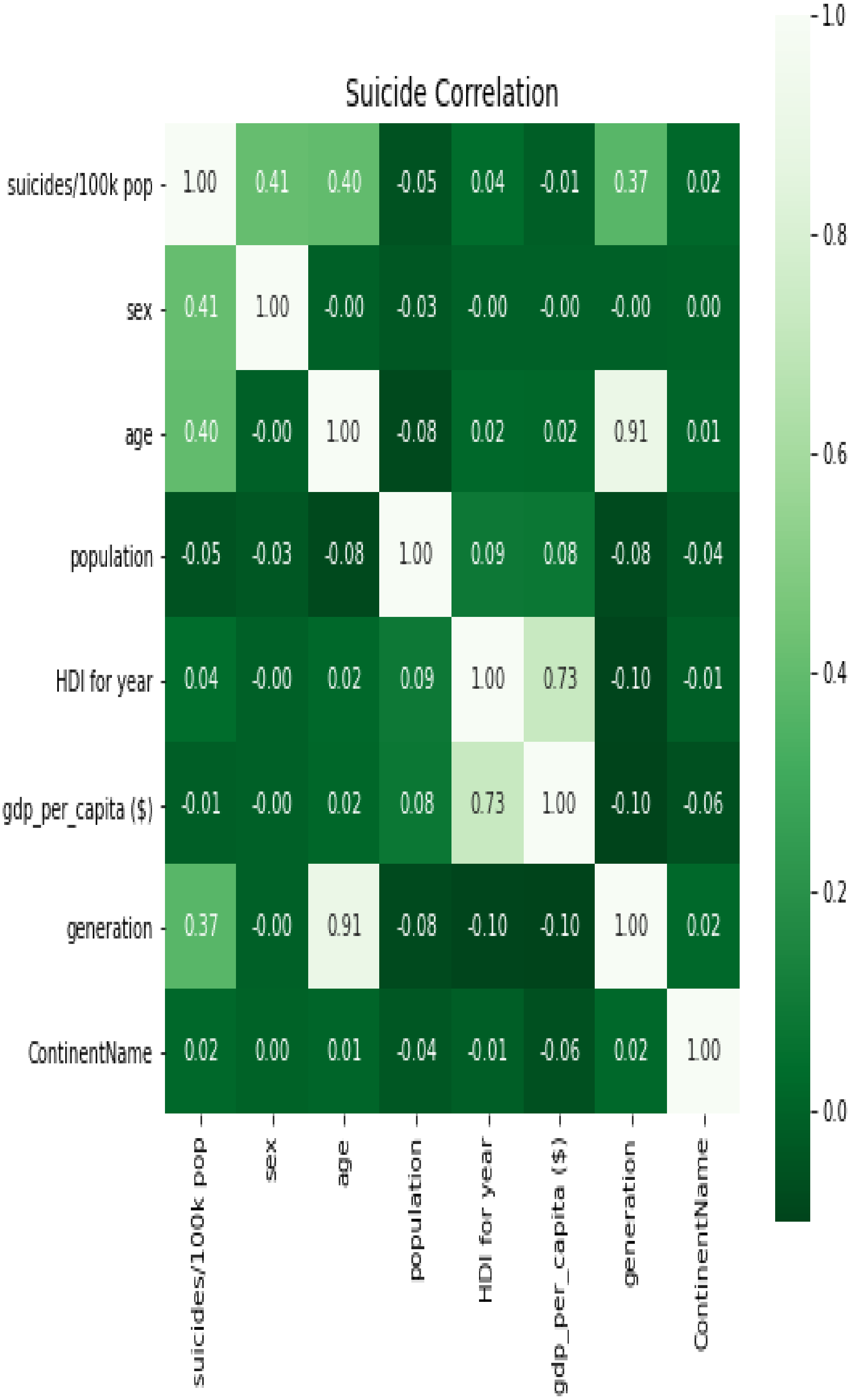

Figure 6: Suicide Overall Correlation

As shown in figure 6, age, generation, and gender all play a role in determining suicide rates for both men and women. Gender is a significant factor, and males are more likely than females to commit suicide in recent years. Suicide rates have been steadily declining across all age groups since an increase in 1995. Suicides among the elderly appear to have increased since 2013. Suicides among females aged 5 to 14 have increased only slightly over the last few decades. Suicides occur at a steady rate in people aged 25 to 75.

4.7 Suicides relation with Country's GDP

According to figure 7, there is a very poor positive relationship between the rate of suicide and the country's GDP per capita, where it has frequently been discovered that countries with a higher GDP have suicide rates of more than ten. This simply indicates that within a country, the question "does a rise in income (per citizen) have an impact on suicide incidence?" is equivalent to asking, "does a country's suicide rate increase over time?"

As a result, they pose the following question: Does the suicide rate decrease as a country's wealth increases? It is dependent on the region—for almost any nation, there is a strong relationship between the year and GDP per capita, implying that GDP per capita rises linearly over time. I calculated the Pearson correlations between 'year' and 'GDP per capita' in each region and then averaged the results. According to Antonio et al. (2011), the GDP and suicide rates are inextricably linked.

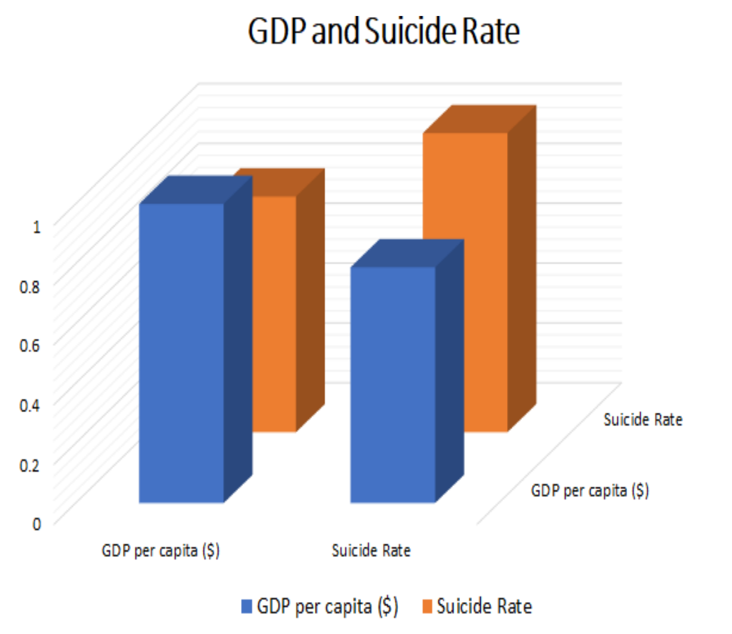

Figure 7: GDP and Suicide Rate

The average correlation coefficient was 0.7878. This entails asking within a country, "Does an increase in wealth (per citizen) influence the suicide rate?" is equivalent to asking, "Does a country's suicide rate rise over time?" It is up to the rest of the world! Any country can rise over time, but most of them will fall. Instead, we have included a reasonable question below.

Do wealthier nations have a higher suicide rate?

Rather than focusing on patterns within countries, I take each country and calculate its average GDP (per capita) overall years for which data is available. I can then calculate how this relates to the suicide rate in various countries over those years. As a result, one data point per country is intended to provide a broad picture of the country's prosperity and suicide rate.
There are also some countries with very high leverage and residuals that may have a significant impact on the fit of my regression equation (e.g., Lithuania). Except for countries with a Cooks Distance value greater than 4/n, I defined and omitted this using Cooks Distance.
The model's p-value is 0.0288 0.05. This assumes that we should reject the idea that a country's GDP (per capita) has nothing to do with its suicide rate. Although the r-square is 0.05, GDP explains just a small portion of the entire variation in suicide rates. The study used the data mining method to analyze the data and generated a model for predicting suicidal tendencies in the person based on the findings. Prediction is made based on risk factors such as depression, fear, hopelessness, stress, and so on, and is measured using a variety of psychiatric scales. For prediction purposes, numerous data mining algorithms for classification are matched.

4.8 Machine Learning Classifiers

Logistic Regression (LR), Multilayer Perceptron (MLP), Polynomial/Gaussian/Sigmoid SVM, K-nearest neighbors (KNN), and Decision Tree (DT), were used in the current analysis for prediction purposes.

I. Logistic Regression

Table 3 summarizes the effective evaluation of the Multinomial Logistic Regression model by selecting function subsets from each feature selection process. According to the table, the accuracy, F-measure, precision, support, and recall values of models produced by automated feature selection and combined feature selection are generally higher than the baseline. The best classifier output is obtained using preliminary measures. The accuracy of the results is shown in the table.

Comprehensive Precision by Class:

Table 3: Logistic Regression Classifier

	0	1	ACC	macro Avg	weighted Avg
Precision	0.58	0.90	0.83	0.74	0.84
Recall	0.64	0.88	0.83	0.76	0.83

F-Measure	0.61	0.89	0.83	0.75	0.83
Support	11.00	43.00	0.83	54.00	54.00

Logistic regressions have an accuracy of 83%, which is quite good. Expert-based feature discovery is thought to be the most efficient in terms of running time, possibly because features are pre-selected, and calculations are involved. A classifier's confusion metrics are as follows:

$$\text{Confusin Metrics} = \begin{vmatrix} 7 & 4 \\ 5 & 38 \end{vmatrix}$$

II. Multilayer Perceptron (MLP)

We also discovered that the average recall and accuracy score for both strategies are extremely low, whereas the specificity score is extremely high. Precision and recall are calculated using the number of positive cases (TP or FP) and a high specificity score (which indicates high TN, low FP counts).

Table 4: MLP Classifier

	0	**1**	**ACC**	**macro Avg**	**weighted Avg**
Precision	0.59	0.92	0.82	0.75	0.84
Recall	0.77	0.83	0.82	0.80	0.81
F-Measure	0.67	0.90	0.82	0.77	0.82
Support	13.00	41.00	0.82	54.00	54.00

Table 4 summarizes the performance evaluation of the MLP Neural Network model with feature subsets chosen from each feature selection process. In general, the selected features performed well on the MLP Neural Network, but for the baseline, the logistic regression model outperformed the MLP.

III. Kernel Support Vector Machine (SVM)

An SVM considers not just a data point, but the most optimal decision boundary. The optimal judgment boundary is the one with the weight difference between the two nearest spots.

In contrast, the basic SVM technique can be used to establish the decision boundary for linearly separable outcomes. A straight line, on the other hand, cannot be employed as a decision boundary in the event of non-linearities separate results. Instead of simple SVM, Kernel SVM, a tweaked variant of SVM, is used. To determine which kernel SVM worked best for our problem, we implemented polynomial, Gaussian, and signed kernels.

Table 5: Polynomial Kernel-SVM Classifier

	0	**1**	**ACC**	**macro Avg**	**weighted Avg**
Precision	0.83	0.75	0.76	0.80	0.78
Recall	0.30	0.97	0.76	0.63	0.76
F-Measure	0.43	0.85	0.76	0.64	0.72
Support	17.00	37.00	0.76	54.00	54.00

In the case of the polynomial kernel, you must also specify a value for the SVC class's degree parameter. Table 5 shows an accuracy of 76%, which is insufficient for our model due to the sensitivity of the research problem. Now I'll take it a step further and add the Gaussian kernel to our model. To use a Gaussian kernel, enter 'ruff' as the value for the SVC class's Kernel parameter.

Table 6: Gaussian Kernel-SVM

	0	**1**	**ACC**	**macro Avg**	**weighted Avg**
Precision	0.0	0.83	0.84	0.41	0.69
Recall	0.0	1.00	0.84	0.50	0.84
F-Measure	0.0	0.90	0.84	0.46	0.76
Support	9.0	45.00	0.84	54.00	54.00

The Gaussian Kernel-SVM (table 6) achieves an excellent accuracy of 84%. Expert-based feature discovery is thought to be the most efficient in terms of running time, possibly since features are pre-selected, and calculations are involved. This kernel's performance is very good when compared to all classifiers implemented for this problem.

Finally, to improve accuracy, recall, F-measure, and support values, we implemented the Kernel SVM with a signed kernel (table 7). I specified signed as the value for the SVC class's kernel parameter to use the signed kernel.

Table 7: Sigmoid Kernel-SVM

	0	1	ACC	macro Avg	weighted Avg
Precision	0.0	0.78	0.78	0.39	0.60
Recall	0.0	1.00	0.78	0.50	0.78
F-Measure	0.0	0.87	0.78	0.44	0.68
Support	12.0	42.00	0.78	54.00	54.00

When the output of the various kernel types is compared, the polynomial kernel performs the worst. Because the sigmoid function returns two numbers, zero and one, it is better suited for binary classification problems. However, in our instance, we had three manufacturing groups. In terms of prediction rate, we can see that the Gaussian kernel outperformed the signed kernel, while the signed kernel misclassified some cases. As a result, the Gaussian kernel performed marginally better than the other kernels. There is, however, no hard and fast rule as to which kernel works best in each scenario. It ultimately boils down to testing each kernel and choosing the one that provides the best results on your test dataset.

IV. Decision Tree (DT)

Classification through the Decision Tree has been shown to produce the best results when compared to others. It cannot demonstrate higher accuracy, precision, recall, or F-measure values and is, therefore, best suited for forecasting suicidal attempts. It has a normal accuracy of 78 percent (Joseph et al., 2018). Expert-based feature discovery is thought to be inefficient in terms of running time, possibly since features are pre-selected, and calculations are required.

Table 8: Decision Tree Classifier

	0	1	ACC	macro Avg	weighted Avg
Precision	0.85	0.76	0.78	0.80	0.79
Recall	0.53	0.94	0.78	0.73	0.78
F-Measure	0.65	0.84	0.78	0.74	0.76
Support	21.00	33.00	0.78	54.00	54.00

V. K-Nearest Neighbors (KNN)

The results show that our KNN algorithm was unable to identify any of the test collection records with high precision (74 percent), which is insufficient for suicidal prediction. While the algorithm performed admirably on this dataset, it does not predict the same results in other applications. As previously stated, KNN does not always perform well when dealing with high-dimensional or categorical elements.

Table 9: KNN Classifier

	0	1	ACC	macro Avg	weighted Avg
Precision	0.50	0.78	0.74	0.64	0.70
Recall	0.29	0.90	0.74	0.59	0.74
F-Measure	0.36	0.83	0.74	0.60	0.71
Support	14.00	40.00	0.74	54.00	54.00

Following the implementation of all the classifiers for the suicidal prediction model. In this method of evaluation, we base our assessment of function subset consistency on how well a specific classifier performs when certain features are used to build the classifier. Because the class distribution in the suicidal dataset is balanced, accuracy is not an appropriate evaluation criterion in all situations. In some cases, a single recall value can provide a more precise picture of how well a classifier distinguishes between positive and negative instances, but it is usually prone to data imbalance in others.

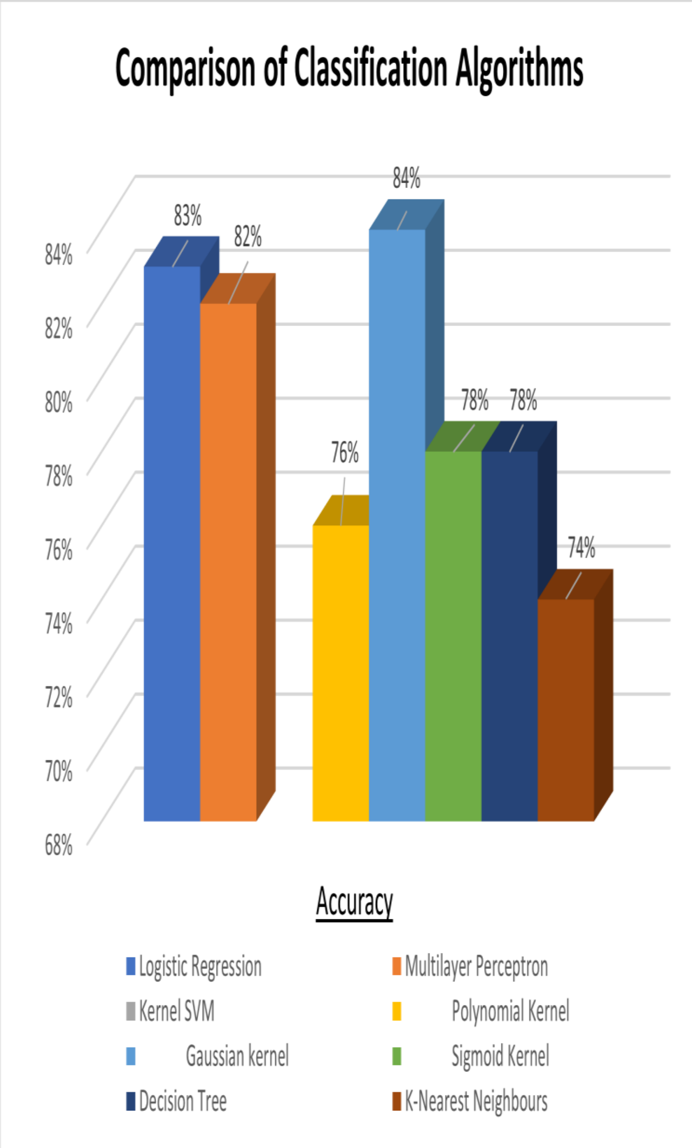

Figure 8: Comparison of Classification Algorithms

As we've seen, all classifiers and interruptions result in the search for a better model to predict suicidal behavior using machine learning models. Classification using the Gaussian Kernel - SVM has been shown to produce the best results when compared to others. It outperformed the competition in terms of accuracy, precision, recall, and F-measure values, and it is best suited for forecasting suicidal attempts.

5. Conclusion

My question was whether the situation would improve or deteriorate. And the conclusion is that things are improving, at least according to the statistics. Suicide rates have declined in general. Many countries saw a decrease in suicide rates between 2006 and 2015, compared to 1996 to 2005. This trend can be seen in almost all age groups of both sexes. The use of the 10-year median suicide rate (for a given country, age, and gender) reduces the possibility that the observed outcome is spurious due to randomness. I hope that this trend continues for all of us in the twenty-first century. In terms of the reasons for the observed transition, there are several scenarios that I can define in broad strokes as follows:

- ✓ Mechanisms, whether intentional or not, that make it difficult to commit suicide.
- ✓ Actual sociological, economic, and other factors keep people from thinking about suicide in the first place.
- ✓ Faults, stereotypes, methodology improvements, or deception, beginning with local data collection.

Various computational classification algorithms are available for predicting suicidal behavior in a person for behavioral forecasting. The performance of Gaussian Kernel-SVM has been demonstrated to be superior to the other algorithms used. Suicide prevention is a key goal on the worldwide public health agenda, yet data on suicide and suicide attempts is imprecise and difficult to gather. Suicide and suicide attempt monitoring, and surveillance must be improved for suicide prevention initiatives to be effective. This includes suicide attempt registration, clinical suicide attempt registries, and representative surveys of national self-reported suicide attempts.

Proper data availability, prediction accuracy, and more and more effective suicide prevention strategies can then be implemented. Suicidal ideation, hopelessness, isolation, addiction, psychoticism, neuroticism, extraversion, and the family setting are all significant areas of higher education, and the current research has yielded very promising results in this regard. There is no such thing as a perfect or final analysis, and each study is presented as a starting point for future research.

References

[1] D. J. Koncak *et al.*, “A suicide prevention strategy for youth presenting to the emergency department with suicide related behaviour: protocol for a randomized controlled trial,” *BMC Psychiatry*, vol. 20, no. 1, p. 20, Dec. 2020, Doi: 10.1186/s12888-019-2422-y.

[2] A. H. de Gannet, J. M. Rodríguez, and J. M. Benda, “Adrián S. Gemmate-Welsh H.,” p. 43.

[3] L. Hong *et al.*, “Bullying, Depression, and Suicidal Ideation Among Adolescents in the Fujian Province of China: A Cross-sectional Study,” *Medicine*, vol. 95, no. 5, p. e2530, Feb. 2016, Doi: 10.1097/MD.0000000000002530.

[4] C. L. Perry, S. H. Kielder, D. M. Murray, and K. I. Kleppe, “Communitywide smoking prevention: long-term outcomes of the Minnesota Heart Health Program and the Class of 1989 Study.,” *Am J Public Health*, vol. 82, no. 9, pp. 1210–1216, Sep. 1992, Doi: 10.2105/AJPH.82.9.1210.

[5] B. Cigar and D. Anal, “Comparison of Data Mining Classification Algorithms Determining the Default Risk,” *Scientific Programming*, vol. 2019, pp. 1–8, Feb. 2019, Doi: 10.1155/2019/8706505.

[6] M. H. Allen *et al.*, “Contributors to Suicidal Ideation Among Bipolar Patients with and Without a History of Suicide Attempts,” *Suicide and Life-Threatening Behaviour*, vol. 35, no. 6, pp. 671–680, Dec. 2005, Doi: 10.1521/suli.2005.35.6.671.

[7] G. Liotta, M. Mozella, and F. DiSanto, “Does democracy improve human development? Evidence from former socialist countries,” vol. 9, p. 19, 2018.

[8] G. Karnowski, “Demographic changes in post-socialist countries in the period 1992-2017,” *Price Nau owe Universität Economizing we Wroclaw*, vol. 64, no. 6, pp. 87–102, 2020, Doi: 10.15611/pn.2020.6.07.

[9] N. Panji and A. Bayrakdarian, “Depression and suicide ideation among secondary school adolescents involved in school bullying,” *Primary Health Care*, vol. 11, no. 04, pp. 349–362, Oct. 2010, Doi: 10.1017/S1463423610000307.

[10] K. Hawtin *et al.*, “Epidemiology and nature of self-harm in children and adolescents: findings from the multicentre study of self-harm in England,” *Eur Child Adolescent Psychiatry*, vol. 21, no. 7, pp. 369–377, Jul. 2012, Doi: 10.1007/s00787-012-0269-6.

[11] E. Turban, T. Nandi, M. Aslan, and C. Zeren, “Epidemiology of attempted suicide in Hate, Turkey,” p. 6.

[12] J. R. Cutcliffe, “Evaluating the success of clinical supervision,” p. 1.

[13] E. Hamilton and B. Climes-Dougan, “Gender Differences in Suicide Prevention Responses: Implications for Adolescents Based on an Illustrative Review of the Literature,” *IJERPH*, vol. 12, no. 3, pp. 2359–2372, Feb. 2015, Doi: 10.3390/ijerph120302359.

[14] S. M. Stewart, B. D. Kennard, P. W. H. Lee, T. Mayes, C. Hughes, and G. Emslie, “Hopelessness and suicidal ideation among adolescents in two cultures,” *J Child Psyche & Psychist*, vol. 46, no. 4, pp. 364–372, Apr. 2005, Doi: 10.1111/j.1469-7610.2004.00364. x.

[15] S. Y. C. Lai Kwok and D. T. L. Shak, “Hopelessness, Family Functioning and Suicidal Ideation Among Chinese Adolescents in Hong Kong§,” *TOFAMSJ*, vol. 1, no. 1, pp. 49–55, Oct. 2008, Doi: 10.2174/1874922400801010049.

[16] R. Carroll, C. Metcalfe, and D. Gunnell, “Hospital Presenting Self-Harm and Risk of Fatal and Non-Fatal Repetition: Systematic Review and Meta-Analysis,” *Plops ONE*, vol. 9, no. 2, p. e89944, Feb. 2014, Doi: 10.1371/journal.pone.0089944.

[17] E. C. Chang, L. J. Sanan, J. K. Hirsch, and E. L. Jelick, “Loneliness and negative life events as predictors of hopelessness and suicidal behaviours in Hispanics: evidence for a diathesis-stress model,” *J. Clin. Psychol.*, vol. 66, no. 12, pp. 1242– 1253, Dec. 2010, Doi: 10.1002/jclp.20721.

[18] A. Khan, M. S. Mustafa, A. R. Hamdan, and R. Ahmad, “Influence of Psychological Factors on Suicide Ideation among Malaysian and Indian Adolescent,” *Procedia - Social and Behavioural Sciences*, vol. 143, pp. 347–351, Aug. 2014, Doi: 10.1016/j.sbspro.2014.07.418.

[19] P. Mendez-Bustos, J. Lopez-Cast roman, E. Baca-García, and A. Ceferino, “Life Cycle and Suicidal Behaviour among Women,” *The Scientific World Journal*, vol. 2013, pp. 1–9, 2013, Doi: 10.1155/2013/485851.

[20] D. A. Lamas, E. D. Ballard, and A. B. Patel, “Loneliness and Suicidal Ideation in Drug-Using College Students,” *Suicide Life Threat Behave*, vol. 44, no. 6, pp. 629–640, Dec. 2014, Doi: 10.1111/sltb.12095.

[21] M. Laggard, L. Gosens, and A. Elk lit, “Loneliness, Depressive Symptomatology, and Suicide Ideation in Adolescence: Cross-Sectional and Longitudinal Analyses,” *J Abnormal Child Psychology*, vol. 39, no. 1, pp. 137–150, Jan. 2011, Doi: 10.1007/s10802-010-9442-x.

[22] B. Sharma, T. Lee, and E. Nam, “Loneliness, Insomnia and Suicidal Behaviour among School-Going Adolescents in Western Pacific Island Countries: Role of Violence and Injury,” *IJERPH*, vol. 14, no. 7, p. 791, Jul. 2017, Doi: 10.3390/ijerph14070791.

[23] T. Kanchan, A. Menon, and R. G. Menezes, “Methods of Choice in Completed Suicides: Gender Differences and Review of Literature,” *Journal of Forensic Sciences*, vol. 54, no. 4, pp. 938–942, Jul. 2009, Doi: 10.1111/j.1556-4029.2009.01054. x.

[24] G. Carter, A. Milner, K. McGill, J. Parkas, N. Kapoor, and M. J. Spittle, “Predicting suicidal

behaviours using clinical instruments: Systematic review and meta-analysis of positive predictive values for risk scales," *Br J Psychiatry*, vol. 210, no. 6, pp. 387–395, Jun. 2017, Doi: 10.1192/bjp.bp.116.182717.

[25] J. B. Palmier, "Prevalence and Correlates of Suicidal Ideation Among Students in sub-Saharan Africa," p. 66.

[26] A.-R. A. Asil and M. M. Abdel-Fattah, "Prevalence, symptomatology, and risk factors for depression among high school students in Saudi Arabia," p. 9.

[27] N. Ibrahim, N. Amit, and M. W. Y. Suen, "Psychological Factors as Predictors of Suicidal Ideation among Adolescents in Malaysia," *Plops ONE*, vol. 9, no. 10, p. e110670, Oct. 2014, Doi: 10.1371/journal.pone.0110670.

[28] H. Bolete, T. Ali, G. Legal, and L. Pavan, "Relapse and Clinical Characteristics of Patients with Bipolar Disorders in Central Ethiopia: A Cross-Sectional Study," *Psychiatry Journal*, vol. 2020, pp. 1–6, Sep. 2020, Doi: 10.1155/2020/8986014.

[29] M. J. Heysel, G. L. Flatt, and P. L. Hewitt, "Social Hopelessness and College Student Suicide Ideation," *Archives of Suicide Research*, vol. 7, no. 3, pp. 221–235, Jul. 2003, Doi: 10.1080/13811110301557.

[30] S. You and K. R. Conner, "Stressful Life Events and Depressive Symptoms: Influences of Gender, Event Severity, and Depression History," *Journal of Nervous & Mental Disease*, vol. 197, no. 11, pp. 829–833, Nov. 2009, Doi: 10.1097/NMD.0b013e3181be7841.

[31] S. Chere, R. Lane, L. Polanco-Roman, E. Gaol, and R. Miranda, "Suicidal ideation among racial/ethnic minorities: Moderating effects of rumination and depressive symptoms.," *Cultural Diversity and Ethnic Minority Psychology*, vol. 21, no. 1, pp. 31–40, 2015, Doi: 10.1037/a0037139.

[32] D. B. Kandel, V. H. Raves, and M. Davies, "Suicidal ideation in adolescence: Depression, substance use, and other risk factors," *J Youth Adolescence*, vol. 20, no. 2, pp. 289–309, Apr. 1991, Doi: 10.1007/BF01537613.

[33] S. Khokhar and M. M. Khan, "Suicidal Ideation in Pakistani College Students," *Crisis*, vol. 26, no. 3, pp. 125–127, May 2005, Doi: 10.1027/0227-5910.26.3.125.

[34] M. M. Previn and N. Ferdowsi, "Suicidal ideation in relation to depression, loneliness and hopelessness among university students," *Dhaka Univ. J. Biol. Sci*, vol. 25, no. 1, pp. 57–64, Jun. 2016, Doi: 10.3329/dubs. v25i1.28495.

[35] J. Cooper, N. Kapoor, M. R. C. Psych, R. Webb, E. Guthrie, and K. MacKay-Jones, "Suicide After Deliberate Self-Harm: A 4-Year Cohort Study," *Am J Psychiatry*, p. 7, 2005.

[36] A. Aria, K. O'Grady, K. Caldera, K. Vincent, H. Wilcox, and E. Wish, "Suicide Ideation Among College Students: A Multivariate Analysis," *Arch. of Suicide Res.*, vol. 13, no. 3, pp. 230–246, Jul. 2009, Doi: 10.1080/13811110903044351.

[37] R. McKeon, "Suicide and Cognition: What We Know and How to Use It," *Securities*, vol. 52, no. 3, Jan. 2007, Doi: 10.1037/a0006339.

[38] R. Brar and R. R. Kaur, "Suicide Ideation in Abused Women as Related to Their Depression," p. 3.

[39] M. F. Steger, P. Frazier, S. Oishi, and M. Kale, "The meaning in life questionnaire: Assessing the presence of and search for meaning in life.," *Journal of Counselling Psychology*, vol. 53, no. 1, pp. 80–93, Jan. 2006, Doi: 10.1037/0022-0167.53.1.80.

[40] T. A. Cato, "The Relationship Among Suicide Ideation, Depression, and Optimism," vol. 10, p. 7, 2012.

[41] A. W. Ashur, "Variable influence of the degree of smoking dependence on adult attention deficit/hyperactivity disorder in Iraqi medical students," *adult ADHD*, p. 7.