Environmental Effect on Historical Suicide Rate[*]

Extended Abstract

Ayberk Ersoy[†]

Affiliation
Ankara, Turkey
aersoy@etu.edu.tr

ABSTRACT

Committing suicide is horrible thing and it requires to determine which factors are cause of this behaviour. The factors can use for pre-determine suicide attempts and save people's life. This work considers environmental factors as reasons of suicide attempts. [1]

KEYWORDS

Correlation Analysis, Suicide

1 INTRODUCTION

People can be in depressed cause of personal problems, unemployment, job's requirements etc. Sometimes this reasons becomes very raid in their life and this cause of suicide attempt. if it is success or not, this situation is really horrible and it requires to determine which factors are cause of this behaviour.

Generally, works in literature, focus on psychologic factors of the attempter. But, especially after the milenium, world has become a more depressed place with developed technology. And also environmental conditions like weather has changed with global warning. So that, factors must not reduced in

---

[*]Produces the permission block, and copyright information
[†]Note
[1]This is an abstract footnote

---

psychology. This work considered environmental factors as reasons of suicide attempts.

As a reference of developed technologies, CO_2 emissions will taken. Also, weather conditions will been referenced of global warming. The aim of this work is determining relationship between suicide rate and combined of these datasets according to the biggest 50 countries.

This paper is organised as follows. A detailed dataset description is given in section 2. Methodology is explained in section 3. Results are presented in section 4 and conclusions are shared in section 5.

2 DATASET DESCRIPTION

This analysis is required weather conditions, suicide rate, CO_2 emissions of the biggest 50 countries between 1991-2014. Weather conditions of these countries is collected from average weather conditions of their capital cities for each year. For achieving the conditions, I have used Visual Crossing Weather API. Then I have found suicide rate and CO_2 emissions datasets from Kaggle. These datasets is also for the biggest 50 countries. Features of the datasets are shown in Table 1. The biggest 50 countries are shown in Table 2.

Input Datasets		
Weather Conditions	CO_2 Emissions	Suicide Rate
Country	Country	Country
Year	Year	Year
Temperature	Annual CO_2 Emissions	Gdp Per Capita
Heat Index		Suicide count
Wind Chill		Population
Precipitation		
Snow Depth		
Wind Speed		
Wind Gust		
Visibility		
Cloud Cover		
Relative Humidity		

Table 1: Attributes of input datesets

Top 50 Countries		
United States	Netherlands	Ireland
China	Saudi Arabia	Malaysia
Japan	Turkey	Singapore
Germany	Switzerland	South Africa
India	Taiwan	Philippines
United Kingdom	Poland	Denmark
France	Thailand	Colombia
Italy	Sweden	Bangladesh
Brazil	Belgium	Egypt
Canada	Iran	Chile
Russia	Austria	Pakistan
South Korea	Nigeria	Finland
Spain	Argentina	Vietnam
Australia	Norway	Czech Republic
Mexico	United Arab Emirates	
Indonesia	Israel	
Portugal	Portugal	

Table 2: The biggest 50 countriess

3 METHODOLOGY

Following section is two part. First part is about preprocessing of the data. Next part is about techniques used for determining relationship of features.

3.1 Data preprocessing

Several sources was used for collecting data. So that, data preprocessing had become the key point of the work.

First step of the process was combining all data that comes from different sources. Country name and year was used as a key for this purpose. With the key all datasets were combined.

Second step of that was handling missing values in the combined dataset. This step was also the most critical part of the preprocessing. Two different strategies are proposed for it:

  i) Removing data which includes missing values.
 ii) Filling data:
    • with mean of group by country, if at least one value exist in the feature of specific group.
    • with mean of feature, if there is not any value in the feature of specific group.

All numerical features of datesets after implemented these two strategies were normalized. Then they were used for experiments and the achieved results with the datasets were presented in Section 4.

3.2 Correlation Analysis

Correlation Analysis is a technique for evaluating the relationship between two features by calculating correlation coefficient (also known as Pearson's product moment coefficient). This value are calculated as below.

$$r_{X,Y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

In above formula, n is the number of tuples, x_i and y_i are the respective values of X and Y in tuple i. and are the respective mean values of A and B, and denominators are respective standard deviations of A and B. $r_{X,Y}$ can be:

  i) $r_{X,Y} > 0$: X, Y are positively correlated, meaning that the values of A increase as the values of B increase.
 ii) $r_{X,Y} = 0$: X and Y are independent and there is no correlation between them.
iii) $r_{X,Y} < 0$: then X and Y are negatively correlated, where the values of one attribute increase as the values of the other attribute decrease.

Although my purpose was finding feature that related suicide rate, I have calculated correlation between all numerical features for different the two datasets and results were presented in Section 4.

3.3 Results

Results are calculated from two different datasets by calculating correlation coefficient. After filling with mean value, the correlations are as shown in Figure ??. I was focused on suicide no row but the correlation values was not notable since they was not enough large.

In the other method, removing data that contains missing value, the results of correlations was appeared large enough little bit more than the other method in Figure ??. I focused suicide no row again, realized that suicide no was negatively related with wind chill.

3.4 Conclusions

This work purpose was investigating especially weather effect on suicide no. For this aim, correlation analysis was made with two different datasets made with different two strategies. According to results, although wind chill had enough correlation value, it was not sense. This work can be reference of next works.

REFERENCES