Does The Weather Affect The Amount Of people Infected / Dead From Covid-19?

Authors: Or Shemesh & Snir Shaharabani

About the project:

The "covid_19_deaths" database includes the number of deaths in each country from 22/1/20 to 22/6/20. Each country has a latitude and longitude that indicate its geographical location.

The "covid_19_confirmed" database includes the parameters of the previous database only this time it is about the number of infected instead of the number of deaths.

The "daily_weather_2020" database includes the maximum and minimum temperatures in each country from 31/12/19 to 20/4/20.
Each country has latitude and longitude as in the previous databases.

By using the latitude and longitude method, we connected the databases and passed our data to three parameters: the number of deaths, the number of those infected and whether it was hot or cold on a particular date (1 or -1).
For example, in Israel on 20/4/20 there were 177 death cases, 13,713 infected cases and the average temperature for that day was 67 degrees Fahrenheit (19 degrees Celsius), which means it was a cold day.

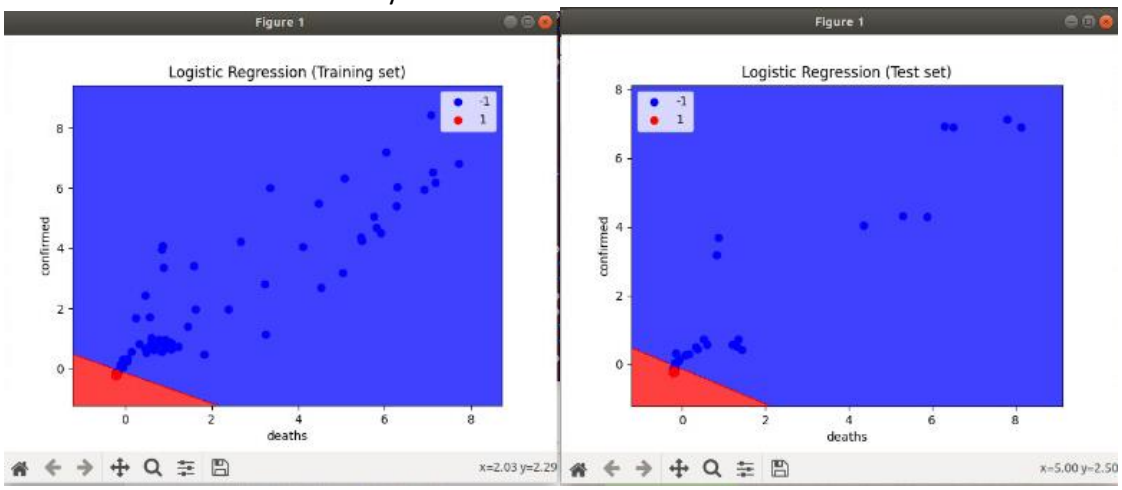
For the sake of getting the best results from the research we used three types of machine learning algorithms which are:
Testing: 25%
Training: 75%
Red means hot day and blue means cold day

- Logistic Regression

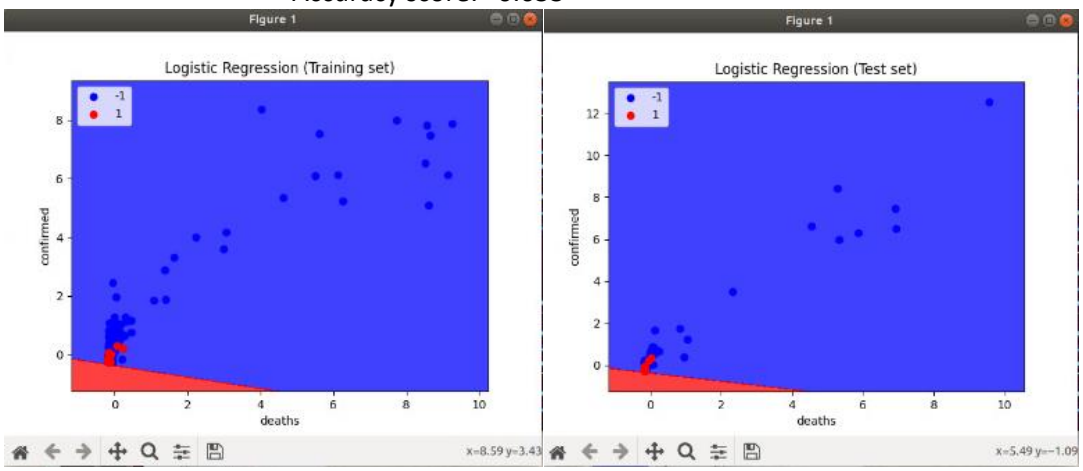
  Definition: Logistic Regression is the appropriate regression analysis to conduct when the dependent variable is dichotomous (binary). Like all regression analyses, the logistic regression is a predictive analysis. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.

Images:

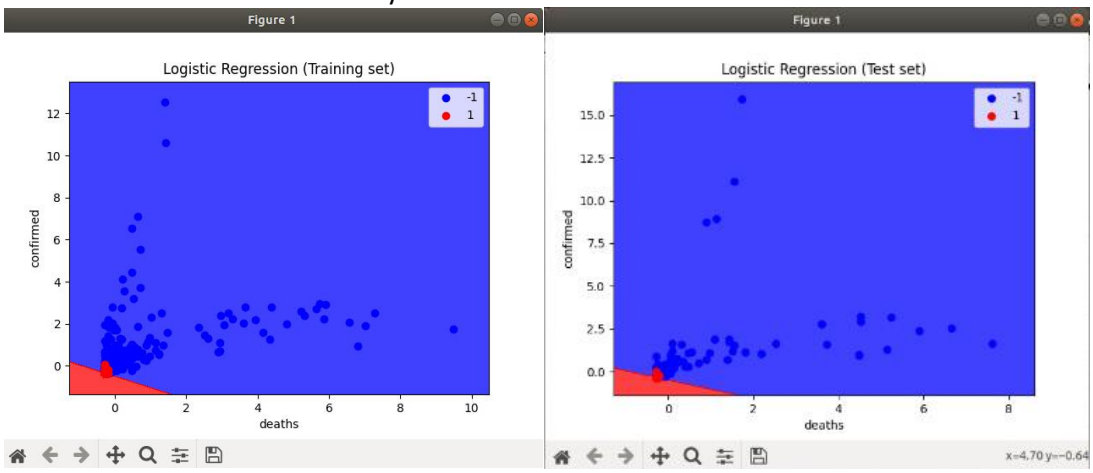
Accuracy score: 0.652



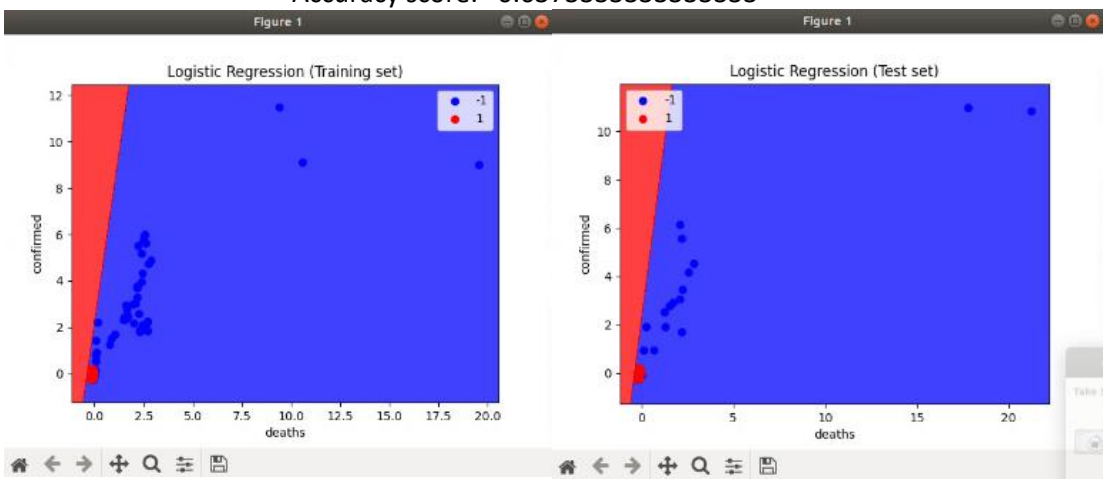
Accuracy score: 0.633



Accuracy score: 0.696



Accuracy score: 0.6373333333333333

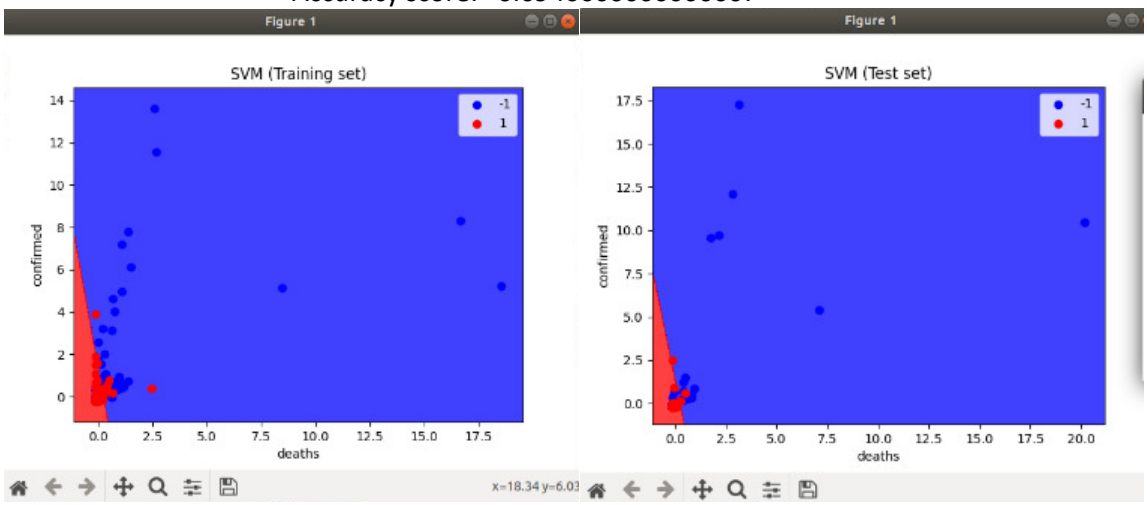


- SVM

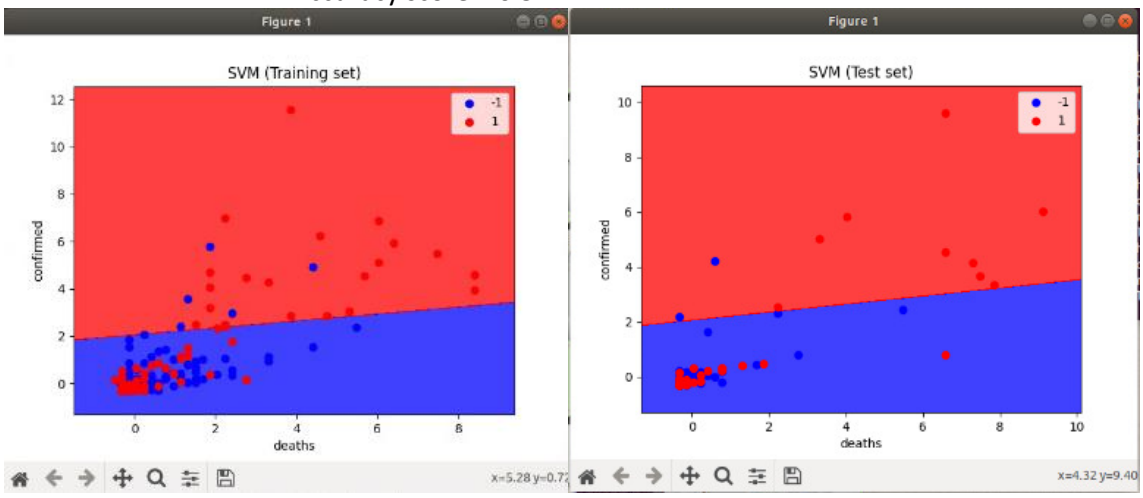
  Definition: A Support Vector Machine (SVM) is a supervised machine learning algorithm that can be employed for both classification and regression purposes. SVMs are more commonly used in classification problems and as such.

  Images:

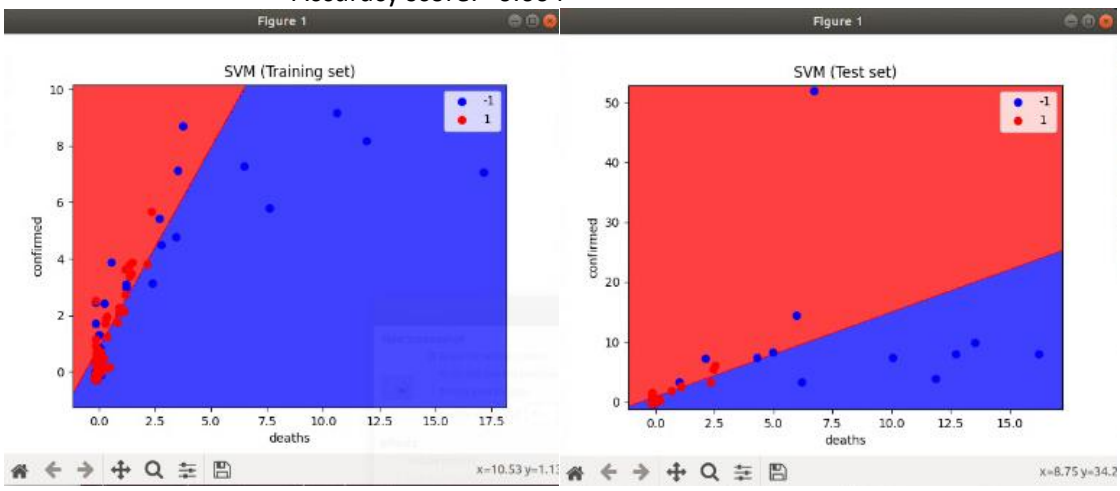
  Accuracy score: 0.6946666666666667



  Accuracy score: 0.52



Accuracy score: 0.664

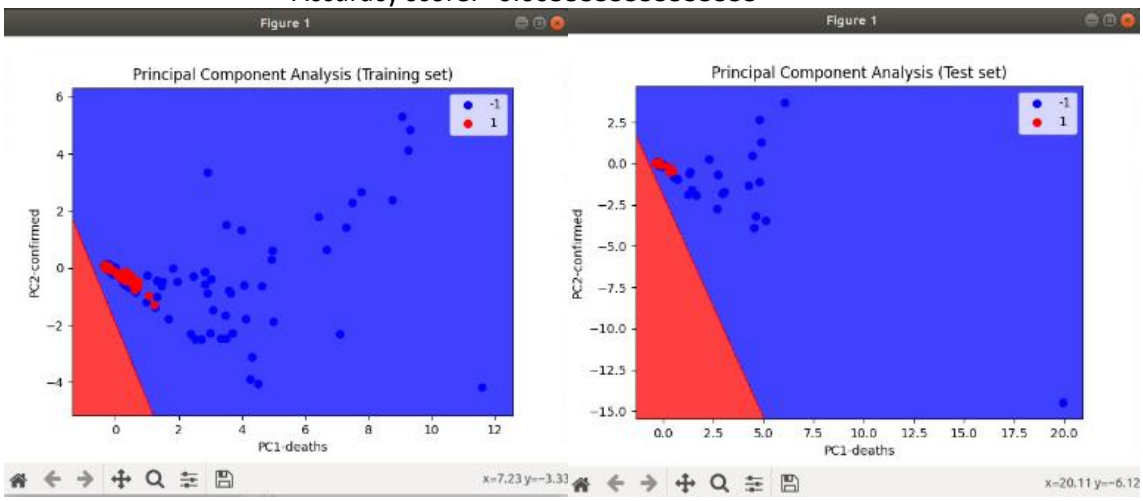


- PCA

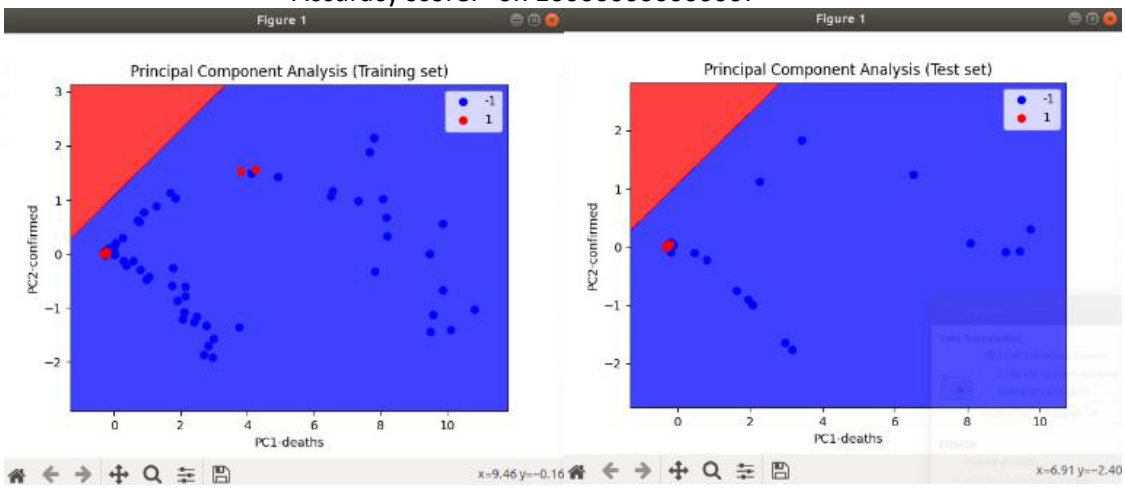
  Definition: Principal Component Analysis, or PCA, is a dimensionality-reduction method that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set.

  Images:

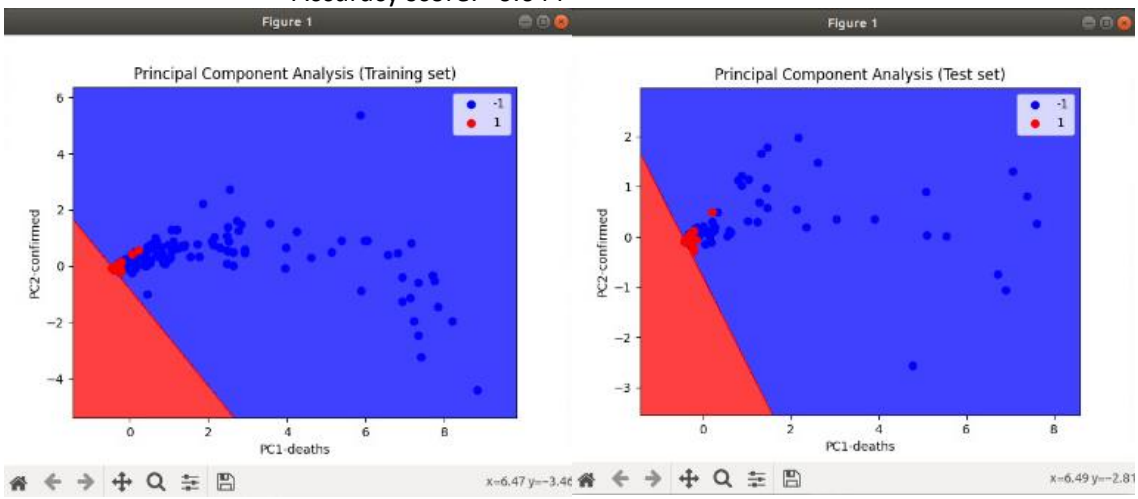
Accuracy score: 0.6053333333333333



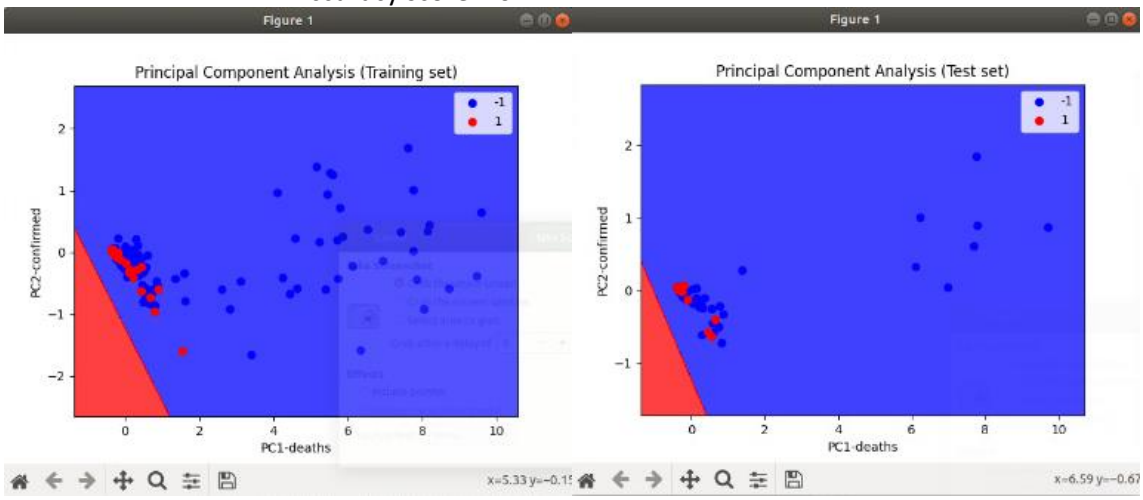
Accuracy score: 0.7106666666666667



Accuracy score: 0.644



Accuracy score: 0.72



Conclusion :

As you can see in the images from each algorithm, the colder the temperature, the higher the number of people infected and dead from the corona virus.

We would like to make an important comment. The data we obtained regarding the weather in each country was updated till April 20, 2020, while the corona virus only gained momentum in late January and early February.
To receive an unequivocal answer to our research question, we will have to wait until the end of 2020 and execute the algorithms on updated weather data.