**Accident analysis using machine learning.**

**Introduction:** Here I analyzed some data from the previous 5 years. In Bangladesh, the accident rate is increasing day by day. So I take data from the internet and different types of paper publication. Collecting the data and using the rules of statistical analysis I analyzed those data.

After I analyzed the data by the use of basic statistical theory , I learned some prediction formulas with the help of probability. As machine learning is basically based on probability and statistics. I try to analyze those analyses with more accuracy. So I tried to regenerate a machine learning model to make a prediction of a 5 year Bangladesh accident rate. For that I need to learn python programming. As for machine learning, there are python libraries. Individual tutorial is quite tough for a beginner. So I collected that data using my machine learning model. I tested the data based on different parameters. And for different parameters the accuracy rate of predication the model analyzes the statistical results are getting different. From that I make a decision and make a machine learning model with relative parameters based.

**Procedure:**

1. **Data collection**
2. **Statistical math**
3. **Regression model**
4. **Use the data in Histogram**
5. **Import python libraries**
6. **Hard coding.**

**Data Collection:**

First I collect the data from google. I viewed news- paper and sources about the past 5 years of accidents in Bangladesh.  There are different categories of accidents. People killed, some are severe and some are minor injuries. So for calculating errortic issues we need to consider the minor accident rate. So in my calculation I just work with the death and severe accident rate of the last 5 years.

I collect data from different types of resources then I make a spreadsheet or csv file that will be directly imported in my python code so my interpreter can easily collect those data and make a model and the regression model.

I am adding the spreadsheet link which was created by me with the web sources .

#The spreadsheet link created by me from different sources.

[Accident-Rate Bangladesh](https://docs.google.com/spreadsheets/u/1/d/1TQNySzsuBDTngBb1UUInmW3DSs4HGI6wBO9xECcJBY0/edit)

In this link I analyzed a spreadsheet and counted the number of accident occurrences from 2018 to 2022. In this spreadsheet I consider the killing of numbers and major injuries. This is how I collect the data. Also using statistical theory I code the regression model histogram of the accident rates. Using machine learning python libraries help to predict the next 5 years approximate accident occurrence from 2023 to 2027.

**Challenges:**

To execute this code and to make a machine learning model I faced different types of challenges.

One of the major challenges I faced was collecting the raw data with validity. As if we don’t train the model with accurate data then the prediction of the machine will not be accurate. Which will be a very bad threat for the prediction. So one of the challenges was to search the webs and make sure the news was accurate.

**Coding Issue:** The next challenge was to make my processor environment setup. Yes, it's true that there are different IDEs to code python and machine learning code.But in my case it was different. I coded in vscode which is Microsoft IDEs.  For python programming it has its own interpreter. So that it consumes a good amount of time to make an environment for python programming. Machine learning actually works with python different setup of libraries. **Numpy,Pandas,Tensorflow etc.** To analyze the prediction of this thing one the major problem was to setup  the libraries of python in vs code . First of all i tried with the PIP packages of python but the interpreter could not match with the installed packages. So to fix this issue I changed the pip package system to anaconda. As far anaconda packages are setup with many libraries whereas pip needs to install individual packages.

**Code-Analysis:**

In coding here I use python different libraries to complete the programming. A little bit of hard code was done by me. To solve this I use three libraries of python. **Numpy , Pandas, Matplotlib, Scipy.optimize.** Here I also import python curve fit to draw scatter diagrams and regression analysis.

**i)    Data -set**

**Ii ) Data-frame**

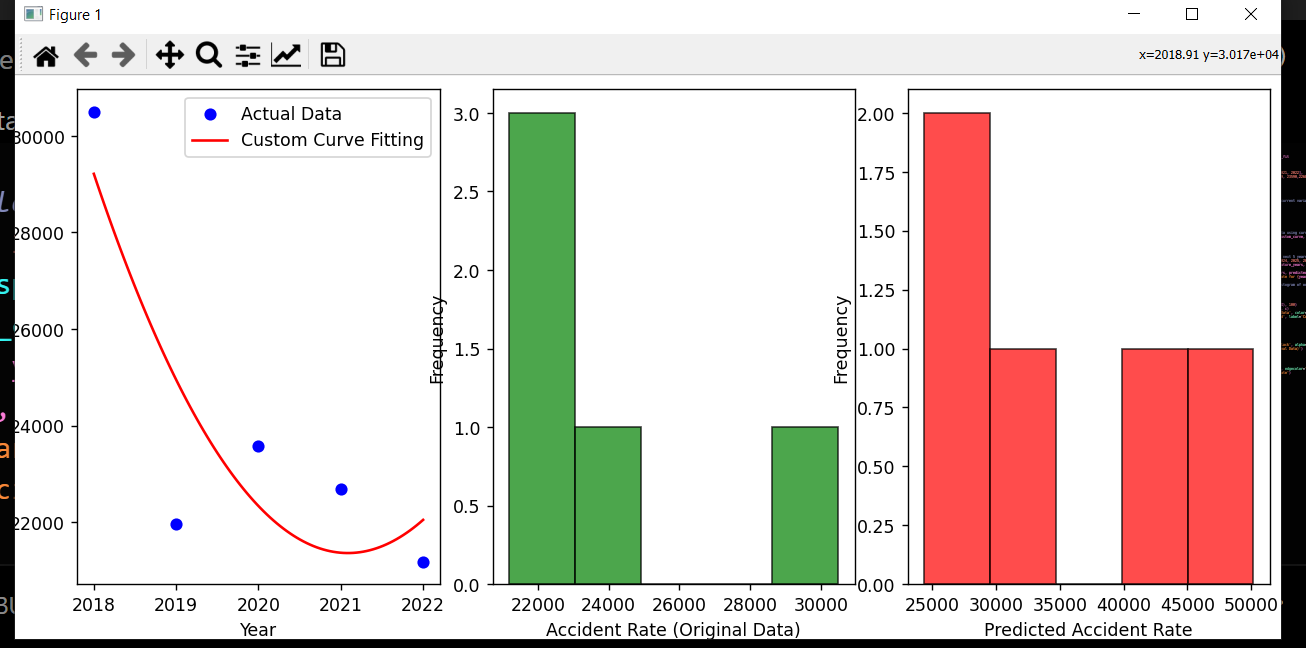
**iii)  Non-linear custom data set.**

**iv)  passing the fitted parameters.**

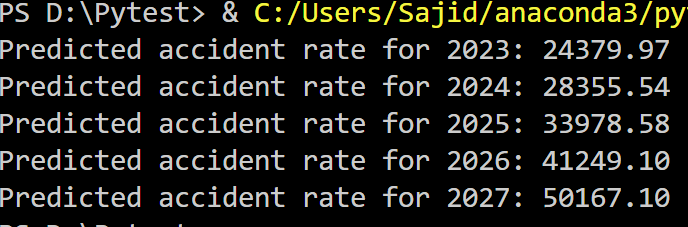
**v)   Regression plot.**

**vi)  Histogram plot.**

Attaching the graph and model after the compilation.



Also giving the predicted rate of next 5  years that machine learning analysis give it to me



This is how I analyzed the next 5 years rate of prediction using machine learning models and statistical theory.  The fractional issue can be omitted because the statistical formula actually analyzes the fact of median.