DES Pune University, Pune
F.Y. M.Sc. Data Science

Project Topic: Machine Learning Approaches to Understand Suicide Cases and Patterns in India

Project by: Alaissa Shaikh

Problem Statement: The goal of this project is to determine how many people commit suicide in each of India's states. Furthermore, by pinpointing the main causes of high suicide rates, we can influence the formulation of intervention plans and legislation to lower the number of suicide deaths.

Data Description: The dataset offers a comprehensive analysis of suicide instances, classifying incidences in depth according to the reasons or circumstances that lead people to commit suicide. The data is arranged by particular states or Union Territories (UTs). This means carefully categorising suicides based on variables including relationship stress, mental health issues, financial difficulties, substance misuse, demands from the workplace or school, and other recognisable causes.

Data Source Ministry/ Department/ Organisation - Ministry of Home Affairs Collection

Sector - Home Affairs & National Security

Year Range - CY 2004 - CY 2022

Link to data - https://ndap.niti.gov.in/dataset/9386

Summary of Insights:

1. Gender Analysis:
    - Higher Suicide Rates Among Males: Compared to females and transgender individuals.
    - Support Needed for Transgender Individuals: The number of suicides among transgender individuals is low which indicates a need for targeted mental health support.

2. State-Wise Analysis:
    - High Suicide States: Maharashtra, Tamil Nadu, and West Bengal have the highest total number of suicides.
    - Targeted Interventions: These states require tailored interventions and policies to address the high suicide rates.

3. Temporal Trends:
    - Spike Around 2020: There was a noticeable increase in suicides around the year 2020, potentially due to external factors such as the COVID-19 pandemic.
    - Yearly Influence: The year of the data (YearCode) is an important factor, reflecting the impact of temporal changes on suicide rates.

4. Feature Importance:
    - State lgd code: This indicates geographical location significantly impacts suicide rates.
    - Gender_Transgender: This suggests the need for special attention to the mental health of transgender individuals.
    - Gender_Male: Highlighting the gender disparity in suicide rates.

5. Model Performance:
    - Accuracy: The Decision Tree model has an accuracy of 87%.
    - Class Performance: The model performs well in predicting suicide rates.
        1. Class 0 - Low suicides: precision - 0.89 and recall - 0.98
        2. Class 1 - High suicides: precision - 0.53 and recall - 0.15.

Recommendations:

1. Targeted State Interventions: Create policies and programmes for mental health that are unique to each state, particularly for those like West Bengal, Tamil Nadu, and Maharashtra.

2. Support for Transgender Individuals: Provide specialised social and mental health assistance to transgender people to address their particular needs.

3. Temporal Analysis: To adjust policy to evolving circumstances and outside variables influencing suicide rates, annual trends should be tracked and analysed.

4. Model Improvements: Dataset Balance: To balance the dataset and increase prediction accuracy for the minority class (high suicides), apply techniques such as SMOTE.

5. Advanced Models: For improved performance, investigate more sophisticated models like XGBoost, Gradient Boosting, and Random Forest.

6. Hyperparameter Tuning: To improve prediction performance, optimise the model's parameters.