Review and Insight Paper on S.H.I.E.L.D (Suicide Health Insights for Early Learning & Detection)

# Abstract

This paper provides a comprehensive review and analysis of the S.H.I.E.L.D (Suicide Health Insights for Early Learning & Detection) project,   
which is a Streamlit-based machine learning and data visualization platform designed to analyze global suicide data. The system leverages   
interactive dashboards and a Decision Tree regression model to offer valuable insights into demographic suicide trends and predict future   
suicide rates based on key parameters such as age, gender, year, and country.

# 1. Project Overview

The S.H.I.E.L.D project integrates data science and machine learning to assist in understanding and predicting suicide patterns worldwide.   
The platform is designed to be both analytical and predictive, enabling real-time exploration through an interactive Streamlit dashboard.   
It utilizes datasets containing demographic and temporal information on suicides and provides visualizations such as maps, bar charts,   
and heatmaps for deeper insight.

# 2. Data Analysis and Visualization Insights

The dashboard offers multiple layers of analysis:  
• Global Suicide Distribution: Interactive choropleth maps display country-wise suicide statistics.  
• Gender and Age-Based Trends: Bar charts highlight demographic variations, revealing patterns such as higher suicide rates among   
specific gender and age groups.  
• Yearly Trends: Line charts track changes in global suicide counts across years.  
• Generational Insights: Pie charts classify suicide trends according to generational cohorts, such as Gen Z, Millennials, Gen X,   
Boomers and the Silent Generation.  
These visual tools help in understanding socio-demographic patterns, thereby supporting mental health awareness and policy decisions.

# 3. Machine Learning Model

The predictive component of the project employs a Decision Tree Regressor model. The model uses features such as year, age, gender, and   
country (encoded via one-hot encoding) to estimate suicide rates per 100k population. Data preprocessing steps include outlier removal,   
data cleaning, and transformation using log scaling for better stability.  
The model achieves an R² score\*\* indicating reasonable predictive capability, with evaluation metrics such as Mean Squared Error (MSE),   
Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) were calculated to assess its performance.

# 4. Key Insights and Applications

• The integration of machine learning with visualization enables early detection and proactive measures for suicide prevention.  
• By filtering parameters such as gender, country, and age, the tool highlights vulnerable groups, assisting policymakers and researchers   
in targeted mental health interventions.  
• The project emphasizes the value of explainable AI, as Decision Trees allow transparency in feature importance.  
• Predictive analytics further extends the project’s utility by providing potential forecasts for suicide rates under varying demographic   
scenarios.

# 5. Conclusion

The S.H.I.E.L.D project demonstrates how data-driven methodologies can support mental health awareness and policy formulation.   
Its combination of interactive visualization and machine learning prediction offers a holistic perspective on global suicide data.   
Future improvements could include integrating deep learning models, adding real-time data updates, and expanding the dashboard to cover psychological or socioeconomic factors