PROJECT PROPOSAL
MACHINE LEARNING INSIGHTS FOR SOMERVILLE HAPPINESS LEVEL PREDICTION

Prepared by : Group G

- 15747 - Amila Poorna
- 15665 - Tashini Ramindi
- 15642 - Hasini Nimesha
- 15669 - Ganeshi Umayangana

BACKGROUND OF THE PROJECT

Understanding and predicting happiness levels among residents of Somerville, New Jersey, is essential for enhancing community well-being. Happiness is a fundamental indicator of life satisfaction and reflects the quality of life in a city. This study aims to develop and implement predictive models that can accurately predict happiness levels. By leveraging machine learning techniques, this study seeks to identify key predictors of happiness and create predictive models. Additionally, developing a web-based dashboard or mobile solution will facilitate the presentation and accessibility of these predictive insights, empowering stakeholders to monitor and act upon happiness trends in real time.

OBJECTIVES OF THE PROJECT

1. Develop and implement predictive models that accurately predict happiness levels among residents of Somerville, New Jersey.
2. Develop a web based dashboard/mobile solution to present the final product.

DATA

The dataset contains Somerville Happiness Survey results from 2011 - 2023. It consists of 11191 observations. The survey asks residents of the city to assess their personal happiness, wellbeing, and satisfaction with city services. The dependent variable includes the levels of happiness on a 5 - point scale.

Dataset:
Somerville Happiness Survey Responses

Data dictionary:
Somerville_Happiness_Survey_Data_Dictionary_2011-2023

SIGNIFICANCE OF THE PROJECT

Predicting happiness accurately holds significant value for both researchers and policymakers because by pinpointing specific areas that require attention and improvement, they can effectively address the factors contributing to residents' happiness. Ultimately, such efforts lead to more prosperous and thriving communities.

SUGGESTED METHODOLOGY

Conduct a comprehensive descriptive analysis, identifying patterns and potentially significant variables by using suitable figures and tables.
Implement machine learning models suitable for multiclass classification, such as Multinomial Logistic Regression, KNN, Naive Bayes, Decision Trees, Random Forest, SVM and XGboost classifiers. Use cross validation to fine tune parameters.