Violence Against Women

This project uses data retrieved from Kaggle titled *Violence Against Women and Girls* (https://www.kaggle.com/datasets/andrewmvd/violence-against-women-and-girls/data). It includes data of percentages of agreeance with sentiments of violence against women, or in other words, what percentage of people agreed with the statements of violence against women. The project performs different regression and classification models on the data to see what features are important for the number of people who approve of the sentiment and to try to predict the gender of the respondents who agree with the statements.

Project Features

This project uses several different regression and classification models to find the important features for approving violence against women, as well as a principal component analysis (PCA) to see if the PCA could improve the models' performance.

Installations and Requirements

This project will require the following Python libraries to be imported in order to run the models, evaluate the model performance, and plot the model performance:

- pandas
- numpy
- matplotlib.pyplot
- sklearn.model_selection: train_test_split
- sklearn.metrics: accuracy_score, classification_report, confusion_matrix, mean_squared_error
- sklearn.linear_model: LinearRegression, LogisticRegression
- sklearn.preprocessing: StandardScaler
- sklearn.tree: DecisionTreeClassifier
- sklearn.svm: SVC
- statsmodels.api
- sklearn.decomposition: PCA

Using the Project

You can use this project in either Jupyter Notebook or any other Python IDE, such as PyCharm. This system could also be run in a Python terminal. However, it is recommended to be used in an IDE, as that is where the script was created and run before. If you wish to use it in Jupyter Notebook, download the .ipynb file for use in your own Jupyter Notebook or copy each cell into your Jupyter Notebook. You may also copy and paste the code into another Python IDE if you prefer a different IDE besides Jupyter Notebook.

The script will load the data, prepare the models for analysis, and run multiple classification and regression models on the data to predict the gender of the respondents

and to predict the higher education or no education groups of the respondents. Each model's metrics can be compared to determine the best model for each goal.

Contact

For any questions or concerns, please feel free to contact me, Ahria Dominguez, at ahriadominguez@outlook.com.