Predicting Life Satisfaction

This project uses data retrieved from Jong et al. 2020 (https://www.nature.com/articles/s41597-020-0482-y) to see if life satisfaction scores can be predicted using a variety of psychological and religious questionnaires. The psychological questionnaires assess several factors, such as depression, stress, anxiety, and traumatic experiences, while the religiosity questionnaires examine things like religious behaviors, affiliations, and religious views. Multiple regression and classification models were performed to predict life satisfaction scores. Each model underwent feature selection and grid searches to find the best parameters for each. Out of all of the models used, only the Ordinary Least Squares (OLS) Regression model performed well on the data.

Project Features

This project uses several different regression and classification models to see which is the most efficient at predicting life satisfaction scores for the participants. Recursive feature elimination with cross-validation was performed for each model to determine the best features to keep in the analysis, and each model underwent a grid search to find the best parameters to use. The performance of each model was evaluated to determine the best-performing model.

Installations and Requirements

This project will require the following Python libraries to be imported in order to manipulate and analyze data, transform the data, create visualizations, and run and evaluate models:

- pandas
- numpy
- matplotlib.pyplot
- statsmodels.api
- sklearn.feature_selection: RFECV, RFE
- sklearn.tree: DecisionTreeRegressor, DecisionTreeClassifier
- sklearn.linear_model: LinearRegression
- sklearn.svm: SVC, SVR
- sklearn.model_selection: train_test_split, GridSearchCV
- sklearn.metrics: accuracy_score, classification_report, mean_squared_error, mean_absolute_error, confusion_matrix, r2_score

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, manipulate and transform the data for modeling, and run multiple classification and regression models on the data to predict life satisfaction. Each model's metrics can be compared to determine the best model for predicting life satisfaction scores.

Contact

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