Young People Survey Solution Report

Task Description

Given the Young People Survey Dataset, task is to predict a person's "Empathy" on the scale of 1 to 5 (Option 1)

Young People survey captures various aspects of a person such as hobbies, health habits, personal traits and so on. Most of this aspect are given rating by a person on the scale of 1-5.

Data Preprocessing

Missing Values:

There are two types of Input Data – Numeric and Categorical. We follow different approach on both the types of Data.

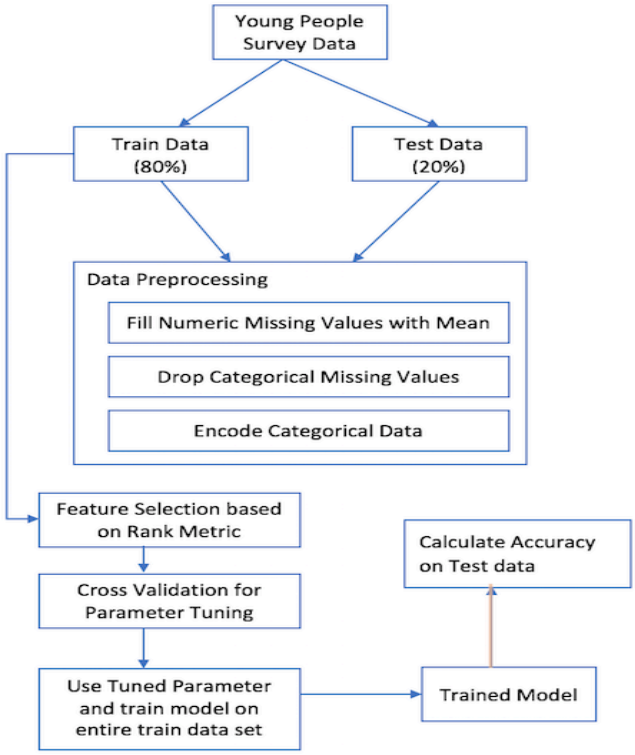
- Categorical: As there are very few rows which are empty for categorical data, so the best solution in such a scenario is to drop the rows.
- Numerical: Here we perform the imputation using the column means.

Categorical Data Conversions:

- In order to convert categorical data to numeric we perform one hot encoding technique.

Solution

First, we test the Baseline accuracy by creating a baseline model which returns the most common class label.

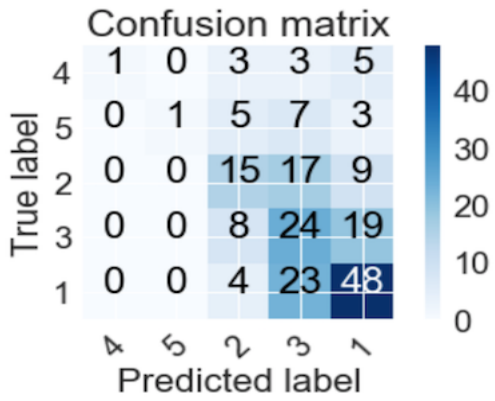


Steps followed in solution:

1. Split Data into Train (80%) and Test Data (20%)
2. Apply Data preprocessing techniques applied above.
3. Perform feature selection based on feature importance using different Algorithms like Random forest and RFE (Recursive Feature Elimination Techniques)
4. Apply 10-Fold Cross Validation on train data and tune hyper parameters for SVM and XGBoost Classifier.
5. Select hyper parameter and train model using entire train data
6. Evaluate model on test data.

Results

Confusion matrix of SVM



Our baseline model gives accuracy of around 38%. Below we present the accuracy of our SVM and XGBoost.

To test the model, we use the sklearn's accuracy_score metric to compute the accuracy of the model.

SVM Model with RBF Kernel: Accuracy ranges in between 44% - 48%

XGBoost Classifer: Accuracy ranges in between 45% - 50%

To further evaluate the model, we calculate the confusion metric which is plotted in the notebook uploaded on bitbucket.

References

Libraries used: Pandas, Numpy, sklearn, seaborn, and matplotlib

Extra Credit Bitbucket Link: **https://bitbucket.org/mmulla4/hw5_mohammed-noman-mulla/src/master/**