

Gender Neutrality and Equality – Model Presentation

The model is featured engineered and trained on the following aspects to make sure there is no Bias/Factor that influences the Fitment Percent to enable an equal-opportunity and bias-free recruitment process.

1. Predict the **FitmentPercent** on test dataset using Regression.
 - a. Preprocessed the data for missing values and encoding categorical matrix of features.
Note: Dropping the missing values on the **BiasInfluentialFactor** column yielded a better model performance.
 - b. Trained the model using Tree-Based Regression models such as Random Forest, CatBoost, XGBoost, and Light GBM.
 - c. Tuned the hyper-parameters using cross-validation.
 - d. Created an ensemble of the above-mentioned models using StackingRegressor to improve the model performance.
 - e. Predicted the target **FitmentPercent** on the test dataset.
2. Predict the **BiasInfluentialFactor** on test dataset using Multi-Class Classification.
 - a. Trained the model using Tree-Based Classification models such as Random Forest, CatBoost, XGBoost, and LightGBM.
 - b. Tuned the hyper-parameters using cross-validation.
 - c. Created an ensemble of the above-mentioned models using StackingClassifier to improve the model performance.
 - d. Predicted the target **BiasInfluentialFactor** on the test dataset.
3. Feature Re-Engineering of data.
 - a. The bias influential factors include 'Gender', 'Marital Status', and 'Ethnicity'.
 - b. Set the values of 'Gender-Female' to 'Gender-Male' and 'Gender-Other' to 'Gender-Male' to avoid bias on Ethnicity.
 - c. Set the values of 'Marital Status - Married' to 'Marital Status - Single' to avoid bias on Ethnicity.
 - d. Set the values of 'LanguageOfCommunication_Hindi' and 'LanguageOfCommunication_Native' to 'LanguageOfCommunication_English' to avoid bias on Ethnicity.
 - e. Trained the model using Tree-Based Regression models such as Random Forest, CatBoost, XGBoost, and LightGBM.
 - f. Tuned the hyper-parameters using cross-validation.
 - g. Created an ensemble of the above-mentioned models using StackingRegressor to improve the model performance.
 - h. Predicted the target **FitmentPercent** on the test dataset.