**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.
   1. 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.

* 1. Trained the model using Tree-Based Regression models such as Random Forest, CatBoost, XGBoost, and Light GBM.
  2. Tuned the hyper-parameters using cross-validation.
  3. Created an ensemble of the above-mentioned models using StackingRegressor to improve the model performance.
  4. Predicted the target **FitmentPercent** on the test dataset.

1. Predict the **BiasInfluentialFactor** on test dataset using Multi-Class Classification.
   1. Trained the model using Tree-Based Classification models such as Random Forest, CatBoost, XGBoost, and LightGBM.
   2. Tuned the hyper-parameters using cross-validation.
   3. Created an ensemble of the above-mentioned models using StackingClassifier to improve the model performance.
   4. Predicted the target **BiasInfluentialFactor** on the test dataset.
2. Feature Re-Engineering of data.
   1. The bias influential factors include ‘Gender’, ‘Marital Status’, and ‘Ethnicity’.
   2. Set the values of ‘Gender-Female’ to ‘Gender-Male’ and ‘Gender-Other to ‘Gender-Male’ to avoid bias on Ethnicity.
   3. Set the values of ‘Marital Status - Married’ to ‘Marital Status - Single’ to avoid bias on Ethnicity.
   4. Set the values of ‘LanguageOfCommunication\_Hindi’ and LanguageOfCommunication\_Native to ‘LanguageOfCommunication\_English’ to avoid bias on Ethnicity.
   5. Trained the model using Tree-Based Regression models such as Random Forest, CatBoost, XGBoost, and LightGBM.
   6. Tuned the hyper-parameters using cross-validation.
   7. Created an ensemble of the above-mentioned models using StackingRegressor to improve the model performance.
   8. Predicted the target **FitmentPercent** on the test dataset.