CI Project

GROUP MEMBERS –

ADITYA AMBURE – 202301070154

SNEHA KABRA – 202301070162

KRISHNA KEDAR – 202301070177

DEVANG DESHMUKH – 202301070170

MODEL TRAINING AND EVALUATION REPORT

Models Trained:

1. Logistic Regression

2. Decision Tree Classifier

3. Random Forest Classifier

4. Gradient Boosting Classifier

Training Pipeline:

- Dataset split using train_test_split with stratified sampling.

- Applied SMOTE on the training data for class balancing.

- Standardized numerical features using StandardScaler.

- Trained models on the resampled and scaled dataset.

Hyperparameter Tuning:

RandomizedSearchCV was applied on Random Forest with:

- n_estimators: [100, 200, 400]

- max_depth: [None, 6, 10, 20]

- min_samples_split: [2, 5, 10]

- min_samples_leaf: [1, 2, 4]

- max_features: ['sqrt', 'log2']

Best Model:

Random Forest Classifier (after hyperparameter tuning)

Evaluation Results:

The following evaluation metrics were used:

- Accuracy Score

- Precision Score

- Recall Score

- F1 Score

- ROC-AUC Score

- Confusion Matrix

- ROC Curve

Key Observations:

- Random Forest achieved the best performance overall.

- SMOTE improved class balance, increasing recall.

- Confusion matrix showed a significant reduction in false negatives.

- Feature importance revealed key predictors like MonthlyIncome, OverTime, Age, and JobLevel.

- SHAP explainability indicated strong influence of OverTime and Age on model predictions.