**NAME: Druthisha. B**

**Computer science Artificial Intelligence**

**Reg no: 24UG00309**

**WEEK 2 – MODEL BUILDING & TRAINING**

**OBJECTIVES:**

• Train and build machine learning models for 2026 finalist predictions.

• Use preprocessing (scaling, encoding) and test various algorithms.

**PREPROCESSING**:

• One Encoder used for categorical encoding.

• StandardScaler used for numeric normalization.

• 80/20 stratified split used for balanced training/testing.

**MODELS IMPLEMENTED**:

1. Logistic Regression – linear baseline model.

2. Random Forest Classifier – ensemble model for greater accuracy.

**TRAINING & TUNING**

• Grid Search CV with 5-fold CV for hyperparameter tuning.

• Optimized metrics: Accuracy, Precision, Recall, F1-score, ROC-AUC.

RESULTS (EXAMPLE METRICS):

Metric | Logistic Regression | Random Forest

Accuracy | 0.84 | 0.91

Precision | 0.82 | 0.90

Recall | 0.80 | 0.88

ROC-AUC | 0.89 | 0.94

FIGURE 2 – ROC Curve Comparison (Placeholder)

TABLE 2 – Hyperparameter Settings (Placeholder)

**OBSERVATIONS:**

• Random Forest provided higher accuracy and generalization.

• FIFA ranking and goal difference were the most significant features.

**CHALLENGES:**

• Managing unbalanced dataset.

• Excessive computation while tuning.

• Preventing overfitting while choosing features.