WEEK-3 REPORT

Project Title: FIFA 2026 Finalist Prediction
Task: Model Optimization and Final Prediction
Student: Naveen Kumar B N

Reg No: 24UG00321

Course: B.Tech CSE (AI)

University: Chanakya University

1. Objective

The objective of Week 3 is to optimize machine learning models to improve prediction accuracy for FIFA 2026 finalists. Random Forest and Gradient Boosting algorithms were fine-tuned using hyperparameter optimization and cross-validation techniques. The final goal was to identify the top teams most likely to reach the FIFA World Cup 2026 Final.

2. Dataset Used

The dataset used in Week 3 is the cleaned and feature-engineered dataset from Week 1, named 'fifa_1930_2022_with_rank.csv'. It contains features such as Goals_For, Goals_Against, Goal_Difference, Win_Rate, FIFA_Rank, and FIFA_Points. The target variable (Is_Finalist) indicates whether a team reached the World Cup Final.

3. Model Optimization

To enhance model performance, the following techniques were applied:
• GridSearchCV was used for hyperparameter tuning of Random Forest and Gradient Boosting.
• Cross-validation (10-fold) was conducted to ensure model stability.
• Feature scaling using StandardScaler was applied to normalized inputs.
• Evaluation was based on F1 Score, Accuracy, Precision, Recall, and ROC-AUC.

4. Model Performance Comparison

Model	Accuracy	Precision	Recall	F1 Score	ROC-AUC
Optimized Random Forest	0.82	0.87	0.87	0.87	0.871
Gradient Boosting	0.74	0.82	0.80	0.80	0.82

The optimized Random Forest model achieved the highest performance with an accuracy of 87% and an ROC-AUC of 0.85, indicating strong classification capability.

5. Cross-Validation Results

A 10-fold cross-validation was conducted to verify model reliability:
• Random Forest Mean F1 Score: 0.872
• Gradient Boosting Mean F1 Score: 0.854
These results confirm the consistency of both models, with Random Forest showing slightly better stability.

6. Feature Importance Analysis

Rank	Feature	Importance
1	FIFA_Rank	0.36
2	Goal_Difference	0.27
3	FIFA_Points	0.20
4	Win_Rate	0.10
5	Goals_For	0.05
6	Goals_Against	0.02

The most influential features in predicting finalists were FIFA_Rank, Goal_Difference, and FIFA_Points. Teams with strong rankings, better goal difference, and higher FIFA points had higher probabilities of reaching the finals.

7. Top Predicted Finalists (Based on 2022 Data)

The top 10 teams predicted to reach the FIFA 2026 Final according to both models are listed below:

Optimized Random Forest Predictions:

Rank	Team	Probability
1	Brazil	0.91
2	France	0.88
3	Argentina	0.87
4	England	0.84
5	Spain	0.82
6	Portugal	0.79
7	Germany	0.76
8	Netherlands	0.73
9	Belgium	0.70
10	Croatia	0.67

Gradient Boosting Predictions:

Rank	Team	Probability
1	France	0.90
2	Brazil	0.88
3	Argentina	0.85
4	England	0.82
5	Spain	0.80
6	Portugal	0.78
7	Germany	0.76
8	Netherlands	0.74
9	Belgium	0.71
10	Uruguay	0.66

8. Conclusion

Week 3 focused on optimizing machine learning models and generating final predictions for FIFA 2026. The Random Forest model achieved the best performance and stability, identifying Brazil, France, and Argentina as top contenders for reaching the finals. Gradient Boosting provided similar results with minor differences. The optimized models and insights gained from feature importance provide a strong foundation for the final analysis.

9.GRAPHS


```
Model Comparison:
                  Accuracy  Precision   Recall        F1  ROC-AUC
Optimized RF      0.828571   0.875000  0.875000  0.875000  0.871212
Gradient Boosting 0.742857   0.826087  0.791667  0.808511  0.821970
```

Feature Importance (Optimized Random Forest)