Somali Football Match Prediction Dataset Explanation

Why the Dataset Was Updated

The old dataset had only four features: `Date`, `Home_Team`, `Score`, and `Away_Team`. This data was very limited because it only showed match results without providing deeper insights.

To make better predictions, we needed more details on team performance, goal differences, and head-to-head records. The new dataset includes 16 features, allowing us to analyze past performances and trends for more accurate predictions.

New Dataset Features and Their Importance

1. **Season** – Helps identify patterns over different seasons.
2. **Date** – Used to track match timing and trends over time.
3. **Home_Team** – Identifies the home team.
4. **Away_Team** – Identifies the away team.
5. **Home_Score** – Number of goals scored by the home team.
6. **Away_Score** – Number of goals scored by the away team.
7. **Home_Goal_Diff** – Goal difference for the home team. (Higher values indicate a stronger team performance.)
8. **Away_Goal_Diff** – Goal difference for the away team. (Lower values may indicate weak away performance.)
9. **Result** – Indicates if the home team won, lost, or if the match was a draw.
10. **Home_Last_5_Wins** – Number of matches won by the home team in their last 5 games. (Shows recent form.)
11. **Away_Last_5_Wins** – Number of matches won by the away team in their last 5 games. (Shows recent form.)
12. **H2H_Home_Wins** – Number of times the home team has beaten the away team in past meetings. (Shows historical dominance.)
13. **H2H_Home_Losses** – Number of times the home team lost to the away team in past meetings. (Indicates weakness against the opponent.)
14. **H2H_Away_Wins** – Number of times the away team has beaten the home team. (Shows if the away team performs well in this matchup.)
15. **H2H_Away_Losses** – Number of times the away team lost to the home team. (Indicates the away team struggles in this matchup.)
16. **H2H_Draws** – Number of times the teams played a draw in past encounters. (Important for estimating draw probabilities.)

Approaches for Prediction

To make accurate predictions, we analyze key features using different approaches:

- **Recent Form Analysis** → `Home_Last_5_Wins` & `Away_Last_5_Wins` help understand team momentum.
- **Goal Difference Impact** → `Home_Goal_Diff` & `Away_Goal_Diff` indicate scoring strength and defense ability.
- **Head-to-Head Performance** → `H2H_Home_Wins`, `H2H_Away_Wins`, and `H2H_Draws` reveal historical match trends.
- **Season-Based Trends** → Season helps analyze trends over multiple years.
- **Score Analysis** → `Home_Score` & `Away_Score` determine attacking strength.
- **Result Evaluation** → `Result` is used as the outcome variable for predictions.

By combining these factors, we can improve match predictions and identify trends that affect match outcomes.

Conclusion

Updating the dataset with these new features allows us to analyze teams more effectively and make more secure predictions. Instead of only relying on raw match scores, we now consider goal differences, recent form, and head-to-head records to understand team performance better. This structured approach makes the predictions more reliable and data-driven.