PLAY THE BARÇA WAY

Think you can play the Barça way? Let's find out.

By Yoganshu Vaidya

DATASET & FILTERING LOGIC

- **SOURCE:** STATS OF PLAYERS FROM TOP LEAGUES (2024–25)

- **INITIAL SIZE:** 2854 ROWS × 165 COLUMNS

- **FILTERED:**
  - ONLY MIDFIELDERS (MF) AND FORWARDS (FW)
  - PLAYERS FROM 24 TOP CLUBS
  - MINIMUM 1100 MINUTES PLAYED

- **RESULT:** QUALITY-FOCUSED PLAYER SUBSET FOR PREDICTION

- **HIGHLIGHT COLUMNS:**
  - "PLAYER", "NATION", "POS", "AGE", "PRGC", "PRGP", "PRGR", "G+A", "XG+XAG", "MIN", "SQUAD"

FEATURE ENGINEERING

CREATED NEW PER 90-MINUTE FEATURES

ADDED A BINARY COLUMN: ISBARCA → 1 IF PLAYER IS FROM BARÇA, 0 OTHERWISE

TRAINED THE MODEL ON BARÇA VS NON-BARÇA STYLE INDICATORS

SIDE-BY-SIDE BAR CHART: AVG PER-90 STATS (BARÇA VS OTHERS)

MODELING & TOP PLAYERS

STEP 1: CREATED BINARY LABEL:

- ISBARCA → 1 (BARÇA PLAYER), 0 (NON-BARÇA)

STEP 2: ENGINEERED PER 90-MINUTE FEATURES:

- PROGRESSIVE_PER_MIN, CREATIVE_PER_MIN, GOALS_PER_MIN, AND RUNS_PER_MIN.

STEP 3: SMOTE MODEL FOR IMBALANCE DATA

- APPLIED SMOTE TO BALANCE THE CLASSES

(BARÇA PLAYERS WERE VERY FEW COMPARED TO OTHERS)

STEP 4: USED A CLASSIFICATION MODEL:

- (LIKE RANDOMFOREST) TO PREDICT HOW WELL A PLAYER FITS THE BARÇA STYLE
- MODEL RETURNED BARÇA FIT PROBABILITY

SCOUTING THE FUTURE

LIST OF FUTURE ENHANCEMENTS:

- LET USERS SELECT PLAYERS TO COMPARE, VISUALIZE STATS, AND SEE MODEL PREDICTIONS INSTANTLY.

- TEST ADVANCED MODELS LIKE XGBOOST, ENSEMBLE VOTING, OR EVEN NEURAL NETWORKS.

- INTEGRATE FAN OPINIONS WITH ML MODEL TO SIMULATE PUBLIC + DATA HYBRID SCOUTING.

- FETCH REAL-TIME PLAYER STATS VIA APIS FOR CONTINUOUS MODEL UPDATES.