College Football Model Pack: Notebook Fact Sheet

Train better predictive models using 7 curated, ready-to-run Jupyter notebooks built on cleaned, opponent-adjusted college football data (2016–2024, Week 5 onward).

Each notebook is self-contained and demonstrates a core modeling technique with real-world targets like win probability, score margin, and more.

Notebook Guide

`01_linear_regression_margin.ipynb`

Predict final score margin using a simple, interpretable linear regression model. Ideal as a fast baseline or starting point for feature selection.

- Model: `LinearRegression`
- Target: `margin`
- Includes: Evaluation (MAE, RMSE, R^2), scatterplot of actual vs predicted

---

`02_random_forest_team_points.ipynb`

Predict final team scores (home and away) using two Random Forest models. Also computes implied margin from predicted scores.

- Model: `RandomForestRegressor`
- Targets: `home_points`, `away_points`
- Includes: MAE/RMSE evaluation, margin visualization

`03_xg_boost_win_probability.ipynb`

Predict home win probability with XGBoost. Outputs predicted probabilities and derives win/loss predictions.

- Model: `XGBClassifier`
- Target: `home_win`
- Includes: Accuracy, AUC, log loss, calibration curve

`04_fast_ai_win_probability.ipynb`

Train a tabular neural net using FastAI to predict whether the home team wins. Useful for deep learning experimentation on structured data.

- Model: `fastai.tabular_learner`
- Target: `home_win`
- Includes: AUC, accuracy, F1, FastAI-native pipeline

`05_logistic_regression_win_probability.ipynb`

Build a fast, interpretable classifier to predict home wins. Ideal for benchmarking or understanding key drivers of victory.

- Model: `LogisticRegression` (sklearn)
- Target: `home_win`
- Includes: Feature importance plot of coefficients

`06_shap_interpretability.ipynb`

Explain model predictions using SHAP (Shapley values) on an XGBoost model trained to predict final score margin.

- Tools: `shap.Explainer`, `shap.plots.beeswarm`, `force_plot`
- Includes: Global feature importance and individual prediction breakdowns

---

`07_stacked_ensemble.ipynb`

Combine predictions from Logistic Regression, Random Forest, and XGBoost using a stacking meta-model.

- Models: Logistic + RF + XGBoost → `LogisticRegression` stacker
- Target: `home_win`
- Includes: Performance comparison of all models vs ensemble

Notes

- All models use cleaned, opponent-adjusted features derived from 2016–2024 games (Week 5 onward only)
- Feature set includes advanced stats: EPA, success rate, explosiveness, havoc, field position, and more
- Ideal for modelers, analysts, bettors, students, and Pick'em participants

---

For support and updates: collegefootballdata.com | Discord | Twitter | Patreon