ACC Point Spread Predictions and Prediction Intervals (2026 Season)

Team: YOUR_TEAM_NAME | Team members: YOUR_NAME(S)

1) Problem and target

Goal: predict the game point spread (home score – away score) for 78 ACC men's games, and provide prediction intervals (ci_lb, ci_ub) with at least 70% empirical coverage.

2) Data construction

Historical data were assembled from prior ACC seasons (2023–2025) plus any completed games in the current season (2026) using an ESPN-based scrape (CBBpy). For each played game, we compute the observed point spread and a feature vector built from team-level efficiency ratings and context flags.

3) Model and training procedure (point prediction)

Primary point-spread model: a tuned Gradient Boosting regression model (scikit-learn GradientBoostingRegressor, loss='absolute_error'). Hyperparameters were tuned via RandomizedSearchCV using K-fold cross-validation on the training split only (to avoid leaking test information).

Two evaluation schemes were run:

• Season holdout (train 2023–2025, test 2026): best MAE ≈ 8.36 (Lasso/linear models).
• Random split (mixed 2023–2026): best MAE ≈ 7.89 (GradientBoosting_TUNED).

For the submitted point predictions, we use the random-split tuned Gradient Boosting model because it leverages the largest training set including current-season games and performed best under the mixed-year test.

4) Predictors (features)

Each upcoming game is represented by home-minus-away differences plus context flags:

• net_rating_diff (adj_off_eff – adj_def_eff differential)
• efg_pct_diff
• luck_diff
• adj_off_eff_diff
• home_advantage (team-level home/away margin differential)
• rating_luck_interaction = net_rating_diff × luck_diff
• shooting_strength_interaction = efg_pct_diff × adj_off_eff_diff
• is_conference (1/0)
• is_neutral (1/0)

5) Prediction intervals (method)

Intervals are constructed using an empirical residual calibration approach:

1) On a validation set (held out from training), compute residuals r = y – ■.
2) Let q be the smallest half-width such that the empirical coverage of [■ – q, ■ + q] is at least 70%. Operationally, q is chosen as an empirical quantile of |r| that meets the 70% coverage constraint.
3) For each new game, report: ci_lb = ■ – q and ci_ub = ■ + q.

This method directly targets the competition rule (≥70% coverage) while keeping intervals as tight as possible (small average width).

6) Model assessment and selection

Point predictions were evaluated using MAE and R^2 on out-of-sample splits. Prediction intervals were assessed via empirical coverage (must be ≥70%) and average interval width. The submitted interval half-width q was selected to satisfy the coverage constraint on validation data.