Weekly NFL Fantasy Football Point Forecasting Using Gradient Boosting Models

Course: AAI 595 – Applied Machine Learning
Institution: Stevens Institute of Technology
Presenters: Steven DeFalco, Liam Guske, Nick Obiso
Total Time: ~9–10 minutes

Speaker Assignments

- **Speaker 1:** Slides 1–2 (Motivation & Problem Context)
- **Speaker 2:** Slides 3–4 (Data, Pipeline, Methodology)
- **Speaker 3:** Slides 5–7 (Results, Insights, Example)
- **Speaker 1:** Slide 8 (Limitations, Future Work, Conclusion)

Slide 1 – Title & Project Context

Speaker 1
~1 minute

Good afternoon. Today we're presenting our project titled **"Weekly NFL Fantasy Football Point Forecasting Using Gradient Boosting Models."**

Fantasy football is a domain where people constantly make predictions—who to start, who to bench, and who to pick up off waivers. Most fantasy managers rely on very simple heuristics like season averages, last week's score, or gut feeling. Our goal was to ask a more systematic question: *Can machine learning meaningfully improve weekly fantasy point predictions in a realistic setting?*

Importantly, we focus on **weekly forecasting**, not end-of-season totals. Each week is its own prediction problem with high uncertainty. To address this, we built regression models using NFL play-by-play data from 2009 to 2016, engineered time-aware features, and evaluated whether these models can outperform common baselines that fantasy players actually use.

[Advance to Slide 2]

Slide 2 – Why Weekly Fantasy Forecasting Is Hard

Speaker 1
~1–1.25 minutes

Weekly fantasy forecasting is challenging for several fundamental reasons.

First, there is **extreme week-to-week variance**. Football is a low-sample sport—each player only plays one game per week. A single broken tackle, red-zone opportunity, or weather event can drastically change fantasy output.

Second, player performance is **highly context-dependent**. Matchups against strong or weak defenses, changes in game script, injuries to teammates, or even coaching decisions all affect usage and production. Many of these factors are difficult to capture with simple averages.

Third, there is **limited historical data per player**. Even established players only have 16 or 17 observations per season. Rookies, backups, or players switching teams have even less data, making overfitting a real risk.

Because of these challenges, naive models struggle. This motivates the need for **time-aware machine learning models** that emphasize recent trends while still generalizing across seasons.

[Advance to Slide 3]

Slide 3 – Data & Pipeline Overview

Speaker 2
~1.25 minutes

I'll walk through the data and pipeline we used.

We started with an NFL play-by-play dataset from Kaggle spanning the 2009 through 2016 seasons. This dataset includes every offensive play—passes, rushes, touchdowns, turnovers—with detailed metadata like yards gained and player involvement.

We filtered this raw data down to fantasy-relevant plays and then aggregated it to the **player-week level**. Instead of predicting individual plays, each row in our dataset represents one player's total performance in one week.

Fantasy points were computed directly from the play-by-play data using **Half-PPR scoring**, which awards half a point per reception. This avoids relying on precomputed fantasy totals and ensures consistency.

From there, we engineered **51 features** capturing recent performance, usage, and trends. All features are strictly historical—meaning that when predicting Week *t*, we only use data from Week *t-1* and earlier.

In total, this results in roughly **40,000 player-week observations** across eight seasons, giving us a strong foundation for time-series regression.

[Advance to Slide 4]

Slide 4 – Methodology: Features, Splits, and Baselines

Speaker 2
~1.25 minutes

Our modeling approach is designed to closely mirror real-world forecasting.

We grouped features into several categories. **Lag features** capture what happened most recently, such as last week's fantasy points. **Rolling averages** over 3, 5, and 8 weeks smooth out noise and highlight sustained performance. **Season-to-date features** track cumulative production, while **trend indicators** capture momentum—whether a player's role is increasing or declining. Finally, **usage metrics** like targets and carries quantify opportunity.

To avoid data leakage, we use a **strict time-based split**. We train on seasons from 2009 to 2014, validate on 2015, and test on 2016. The test set represents a completely unseen future season.

We also compare against three simple but realistic baselines: last week's score, a 3-week rolling average, and a 5-week rolling average. These baselines reflect how many fantasy managers actually make decisions, so outperforming them is a meaningful benchmark.

[Advance to Slide 5]

Slide 5 – Results: Model Comparison

Speaker 3
~1.25 minutes

Now let's look at how the models performed.

We evaluate performance using **mean absolute error**, or MAE, measured in fantasy points on the 2016 test season. Lower values indicate better predictions.

The baselines perform as expected. Using last week’s score performs worst, with an MAE above 5.5 points. The 3-week and 5-week rolling averages improve on this but still hover around 4.8 to 4.9 points.

All machine learning models outperform these baselines by a wide margin. Ridge Regression achieves an MAE of about 3.2, Random Forest improves further to 2.69, and the best-performing model—**Histogram Gradient Boosting**—achieves an MAE of **2.62 fantasy points**.

That represents roughly a **45% improvement** over the best baseline. The gradient boosting model also achieves an **R-squared of 0.726**, meaning it explains nearly 73% of the variance in weekly fantasy outcomes.

[Advance to Slide 6]

Slide 6 – What the Model Learned

Speaker 3
~1 minute

One advantage of gradient boosting is that it allows us to inspect **feature importance**.

The most important features are **rolling averages of fantasy points**, particularly over 3 and 5 weeks. This tells us that recent sustained performance is far more predictive than single-week spikes.

Interestingly, raw lag features—like just last week’s points—are less important. The model has effectively learned that smoothing performance over multiple weeks reduces noise and improves stability.

This confirms our intuition: a player who has been consistently productive over the last month is more reliable than someone coming off one unusually big game.

[Advance to Slide 7]

Slide 7 – Player-Level Example: Aaron Rodgers

Speaker 3
~1 minute

To make this more concrete, this slide shows a player-level example using Aaron Rodgers’ 2016 season.

The blue line represents actual fantasy points, while the orange line shows the model's predictions. You can see that the model tracks the overall trend quite well, staying close during consistent stretches of performance.

However, the model struggles with **extreme outliers**. In very high-scoring weeks, it underestimates production, and in unusually low weeks, it sometimes overestimates.

This highlights a key tradeoff: the model is optimized for typical performance, not rare, unpredictable events. That's a fundamental limitation of weekly forecasting.

[Advance to Slide 8]

Slide 8 – Limitations, Future Work, and Conclusion

Speaker 1
~1.25 minutes

To conclude, this system is best suited for **weekly decision-making**—helping fantasy managers choose between players based on recent trends rather than intuition alone. In that setting, it consistently outperforms simple heuristics.

That said, there are clear limitations. The model does not include contextual features like opponent strength, injuries, or weather, all of which clearly matter. Missing Week 1 data also limits early-season predictions, and like most statistical models, it struggles with truly unpredictable outliers.

For future work, we'd like to incorporate matchup-level features, integrate external data sources such as injury reports, and explore ensemble approaches that combine multiple models. We could also extend this to multi-week forecasting rather than just one week ahead.

Overall, this project demonstrates that machine learning—when designed with proper time-awareness—can meaningfully improve weekly fantasy football forecasting while remaining interpretable and practical.

Thank you.