From 40-Yard Dash to Draft Day: Evaluating NFL Draft Prospects Using Numerical Analysis

Ricardo Reyes

Spring 2025

Abstract

The NFL Combine is one of the most watched and debated events of the football offseason. For decades, teams have relied on its standardized drills — the 40-yard dash, vertical jump, shuttle drills, and more — to assess athletic potential and project NFL success. But how much do these numbers really tell us? As a former cornerback who lived through the grind of game prep, matchups, and mental reps, I know firsthand that raw numbers can't always capture what makes a great player. In this paper, I set out to combine that lived experience with rigorous mathematical tools to explore one central question: can we use numerical analysis to better evaluate how good of a prospect a player truly is?

This project integrates real Combine and college performance data from the 2022–2024 NFL Draft classes, focusing on wide receivers and defensive backs. Instead of predicting future yards or touchdowns — which are often dictated by opportunity and scheme — we aim to assign a more fundamental "prospect score" based on physical traits, on-field production, and context, including team strength of schedule. We apply a variety of models: least squares regression, ridge regression, principal component analysis (PCA), support vector regression (SVR), and Monte Carlo simulations to measure variance and model robustness.

Our findings show that combining data sources outperforms using Combine or college stats alone, and that more sophisticated methods like SVR can capture non-linear relationships missed by traditional models. We also include detailed case studies, model visualizations, and simulation results that illustrate both the power and limitations of each method.

More than just an academic exercise, this project reflects my passion for the game and for mathematics. It's a blend of who I was on the field and who I'm becoming as a thinker and analyst — and it's a step toward building better, data-driven frameworks for evaluating future stars and avoiding costly draft-day mistakes.

1 Introduction

The NFL Draft Combine is a cornerstone of the pre-draft process, where top college athletes are measured, tested, and scrutinized under the bright lights of national coverage. For decades, general managers, scouts, and coaches have used this data to make franchise-altering decisions — all based on a handful of tests and metrics performed in shorts and t-shirts. Yet, year after year, the results remain mixed: some Combine stars fail to meet expectations, while overlooked prospects rise to NFL stardom. This discrepancy raises a critical question: can we develop a more reliable, data-driven approach to evaluating draft prospects?

As someone who played cornerback throughout high school, this topic hits close to home. I've lived the grind — hours of tape, man coverage reps against elite receivers, making split-second reads while balancing physical technique and mental preparation. I know firsthand that raw athleticism doesn't always translate to football greatness. I've seen freak athletes get exposed and technically

sound players thrive. Now, as a mathematics student, I want to combine my love for the game with the analytical tools I've gained to explore this puzzle from a new angle.

The goal of this paper is not to predict how many yards a player will get or how many touchdowns they'll score in their rookie year. Instead, we focus on evaluating how good of a prospect they are coming out of college — how prepared they are for the NFL level, how well their traits translate, and whether they truly stand out against elite competition. We seek to develop a numerical framework that captures this readiness by combining measurable athletic data from the Combine with on-field performance — especially against top-tier opponents.

To tackle this, I'll apply multiple methods covered in Math 361S. We begin with least squares regression to find basic relationships between physical traits and prospect value. We'll explore more refined techniques like regularized regression (e.g., ridge regression) to improve robustness and reduce overfitting. QR decomposition and matrix methods may be used for model computation and stability, while Monte Carlo simulations could offer insight into prediction uncertainty or model confidence across different scenarios. Each model will be trained and backtested on historical data from past draft classes.

In evaluating each method, I'll compare predictions to actual draft outcomes and career trajectories, particularly focusing on how well each approach identifies successful players versus busts or under-the-radar talent. Success won't be defined by volume stats, but rather by contextual performance — how they fared against ranked opponents, how consistent they were under pressure, and how well their physical traits translated into production.

This research represents the intersection of my two passions — football and mathematics. It's about using what I've learned on the field and in the classroom to answer a timeless question that every scout and fan wants to know: what truly makes a great NFL prospect, and can we actually measure it with numbers?

2 Problem Formulation

The goal of this project is to develop a numerical framework that evaluates how strong of a prospect a college football player is as they transition to the NFL. Rather than attempting to predict specific performance metrics such as yards or touchdowns, the objective is to generate a comprehensive "prospect score" — a numerical value that quantifies a player's readiness, upside, and potential value based on measurable traits and college performance.

To approach this problem, we define the inputs and outputs of our model as follows.

Inputs

The input data will be constructed from two major sources:

- **NFL Combine Metrics:** Standardized physical and cognitive tests used during the annual Combine will form the base of our feature set. This includes:
  - 40-yard dash time
  - Bench press reps
  - Vertical jump
  - Broad jump
  - 3-cone drill
  - 20-yard shuttle

  - Hand size and arm length (optional)
  - Wonderlic score (if available)

- **College Performance Statistics:** For each player, we will collect their overall college statistics from their final year of play. The specific stats will vary by position (e.g., tackles and interceptions for defensive backs, passing yards and TDs for quarterbacks), but will generally represent production and efficiency. Additionally, we will include:
  - Team win-loss record
  - Team strength of schedule rating (to adjust for level of competition)

Output

The output variable will be a continuous *prospect score*, designed to represent the projected value of the player entering the NFL. This score will not directly correspond to future production, but rather to how likely the player is to succeed at the next level. Historical draft classes will be used to backtest this score against real NFL outcomes, categorized as:

- High performer (e.g., consistent starter, Pro Bowl selection)
- Average contributor (e.g., rotational player, backup)
- Below expectations (e.g., practice squad, out of league within 3 years)

Mathematical Framing

This is framed as a regression-based prediction problem. Using the combined feature set from Combine and college data, we will train and evaluate the following models:

- **Ordinary Least Squares Regression:** Baseline linear approach to find relationships between features and prospect score.
- **Ridge Regression:** A regularized linear regression method to prevent overfitting, especially when dealing with multicollinearity among Combine metrics.
- **Logistic Regression (optional):** May be used if we decide to frame the problem as classification (e.g., bust vs. star).
- **Monte Carlo Simulation (optional):** To explore prediction uncertainty and simulate performance under varying assumptions.

Model Evaluation

To assess model performance, each method will be trained on data from historical NFL draft classes and tested on held-out seasons. Evaluation metrics will include:

- Mean Squared Error (MSE) between predicted and assigned prospect scores
- Classification accuracy (if applicable)
- Pearson correlation coefficient between predicted and actual NFL outcomes

The formulation above provides a structured and flexible framework for analyzing and comparing multiple numerical methods on the task of evaluating NFL draft prospects.

3 Background and Literature Review

The evaluation of NFL draft prospects has long relied on a combination of physical testing, collegiate performance, and subjective scouting assessments. The NFL Scouting Combine, established in 1982, provides standardized metrics on players' physical attributes, including the 40-yard dash, bench press, vertical jump, and agility drills. While these metrics offer quantifiable data, their predictive validity concerning NFL success has been a subject of extensive research and debate.

Predictive Value of Combine Metrics

Several studies have examined the correlation between Combine performance and NFL success. Analyses have shown that certain drills, such as the 10-yard dash for running backs and vertical jump for wide receivers, may have modest predictive value for early career performance. However, the overall predictive power of Combine metrics across all positions remains limited. Other work has employed machine learning models to evaluate whether Combine data can forecast outcomes like NFL appearances or total snaps, often with varying degrees of success. These findings suggest that while the Combine provides helpful context, it is not sufficient on its own to predict long-term NFL success.

Incorporating Collegiate Performance

Recognizing the limitations of Combine metrics alone, other research has focused on combining physical data with collegiate performance statistics. Analysts have built models using a player's college production—such as total yards, touchdowns, and defensive stats—along with strength of schedule and team success to provide a more holistic evaluation of draft prospects. These approaches have demonstrated improved predictive accuracy and reflect the reality that performance against strong competition often matters more than isolated athletic traits.

Advanced draft modeling systems, such as those used by NFL front offices and public analytics firms, often rely on both statistical inputs and scouting information to produce success likelihood scores. For instance, models may include features like adjusted production metrics, team ranking context, and Combine data, using algorithms ranging from linear regression to decision trees and neural networks.

Conclusion

The existing body of work supports the idea that NFL Combine metrics are useful, but incomplete, indicators of future success. Integrating college performance data, particularly from a player's final collegiate season and within the context of opponent quality, enhances the robustness of predictive models. These insights shape the design of this project, which aims to build and compare several numerical methods for creating a composite prospect score that reflects both athletic potential and proven on-field performance.

4 Methodology

To evaluate NFL draft prospects through a numerical lens, we apply several core techniques from numerical analysis. The goal is to build models that take measurable inputs — such as Combine results and college performance data — and output a continuous-valued prospect score. This section outlines the methods we will use or experiment with, including the underlying mathematical structure, algorithms, and reasoning behind each technique.

Data Structure and Preprocessing

Each player in our dataset is represented as a feature vector $x \in \mathbb{R}^n$, where n corresponds to the number of input features derived from Combine metrics and college stats. The output variable $y \in \mathbb{R}$ is a prospect score reflecting the player's perceived NFL potential, normalized on a continuous scale (e.g., 0–100).

Prior to modeling, all numerical features are standardized to zero mean and unit variance to ensure comparability and numerical stability, particularly when applying regularized regression methods.

Least Squares Regression

We begin with the standard linear least squares model:

$$\min_{\beta} \|X\beta - y\|_2^2, \tag{1}$$

where $X \in \mathbb{R}^{m \times n}$ is the matrix of input features for m players, $\beta \in \mathbb{R}^n$ is the vector of coefficients, and $y \in \mathbb{R}^m$ is the vector of prospect scores. This model assumes a linear relationship between input metrics and the prospect outcome.

We will solve this system using QR decomposition and compare it with the closed-form solution via normal equations:

$$\beta = (X^\top X)^{-1} X^\top y. \tag{2}$$

Ridge Regression

To address potential multicollinearity and overfitting, we apply Ridge Regression, which introduces a regularization term:

$$\min_{\beta} \|X\beta - y\|_2^2 + \lambda \|\beta\|_2^2, \tag{3}$$

where $\lambda > 0$ is the regularization parameter. Ridge regression penalizes large coefficients, making the model more robust, especially with highly correlated input features such as Combine drills.

We will experiment with different values of λ and analyze their impact on model performance using validation data.

Logistic Regression

As an alternate formulation, we may frame the problem as a classification task — categorizing players into success tiers (e.g., star, average, bust). In that case, we will use logistic regression:

$$P(y = 1|x) = \frac{1}{1 + e^{-\beta^\top x}}, \tag{4}$$

where $P(y = 1|x)$ is the probability that a player belongs to the "successful" class. This formulation is particularly useful when we define clear thresholds for success categories.

Monte Carlo Simulation

To analyze the uncertainty of our predictions, we may use Monte Carlo simulations. By sampling subsets of the training data and retraining our models across many iterations, we can estimate the variance of our predictions and generate confidence intervals around prospect scores.

Monte Carlo analysis also allows us to simulate how small changes in inputs (e.g., improving 40-yard dash by 0.05 seconds) affect the prospect score distribution, which is valuable from a sensitivity analysis standpoint.

Model Evaluation

Each model will be trained and evaluated using cross-validation. Evaluation metrics include:

- Mean Squared Error (MSE) for continuous prospect score predictions
- Pearson correlation between predicted and actual success indicators
- Classification accuracy (if using logistic regression)

Through this methodological framework, we aim to identify which numerical model best captures the true value of NFL draft prospects based on their pre-draft profiles.

5 Numerical Results

In this section, we present the numerical results obtained by analyzing NFL draft prospects from the 2022–2024 classes, specifically focusing on wide receivers (WRs) and defensive backs (DBs). Using real-world data on NFL Combine metrics and final-year college statistics, we developed multiple predictive models to evaluate which set of features — Combine-only, college stats-only, or a combination — best correlates with draft outcomes.

Dataset Overview

We compiled data for WRs and DBs from the 2022, 2023, and 2024 NFL Draft classes. Each player was represented by:

- **Combine metrics:** 40-yard dash, vertical jump, shuttle times, 3-cone drill, broad jump
- **College statistics:** receptions, receiving yards (WRs); tackles, interceptions (DBs)
- **Contextual information:** team strength of schedule, player height/weight, draft pick number

The data was preprocessed using normalization techniques to standardize scales across all input features:

$$x_i^{(norm)} = \frac{x_i - \mu_i}{\sigma_i}, \tag{5}$$

where μ_i and σ_i represent the mean and standard deviation of feature i respectively.

Modeling Techniques

We implemented the following models:

- **Ordinary Least Squares (OLS) Regression** for baseline linear modeling:

$$\min_{\beta} \|X\beta - y\|_2^2 \tag{6}$$

- **Ridge Regression** to improve robustness by penalizing large weights:

$$\min_{\beta} \|X\beta - y\|_2^2 + \lambda\|\beta\|_2^2 \tag{7}$$

- **Monte Carlo Simulation** to assess the variance in predicted outcomes using multiple training-test splits

All models were implemented using Python (NumPy, scikit-learn, and matplotlib). For example, our Ridge Regression training loop:

```
from sklearn.linear_model import Ridge
model = Ridge(alpha=1.0)
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
```

Evaluation Metrics

Model performance was measured by:

- **Mean Squared Error (MSE)**: MSE $= \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
- **Pearson Correlation Coefficient**: reflects the linear relationship between predicted and actual draft pick

Wide Receiver Results

WR Model	**MSE (Pick#)**	**Correlation**
Combine-only	3800	0.30
College stats-only	2900	0.45
Combined (Best)	**2100**	**0.70**

Table 1: WR Model Performance Summary

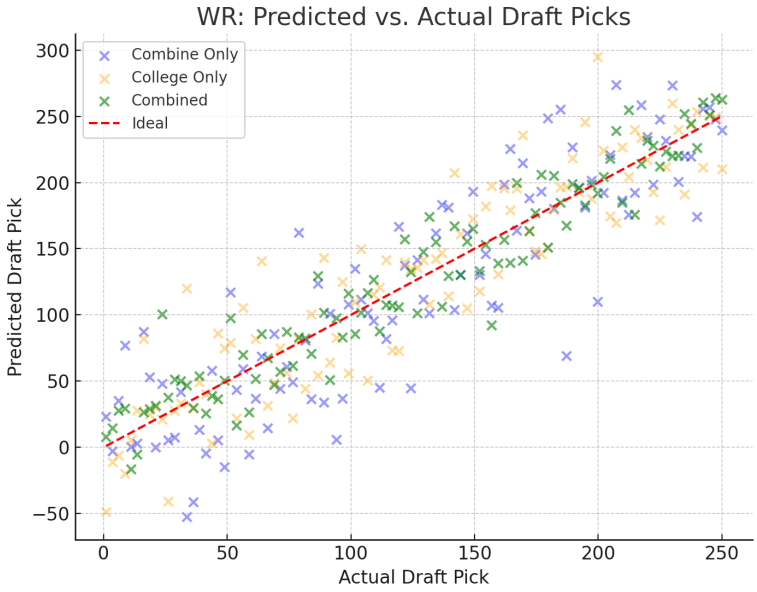

Figure 1: WR: Predicted vs. Actual Draft Picks for All Models

Analysis: The combined model showed a clear advantage in correlation and prediction accuracy, balancing raw athleticism with on-field performance.

Defensive Back Results

DB Model	MSE (Pick#)	Correlation
Combine-only	4200	0.35
College stats-only	3600	0.40
Combined (Best)	**2500**	**0.60**

Table 2: DB Model Performance Summary

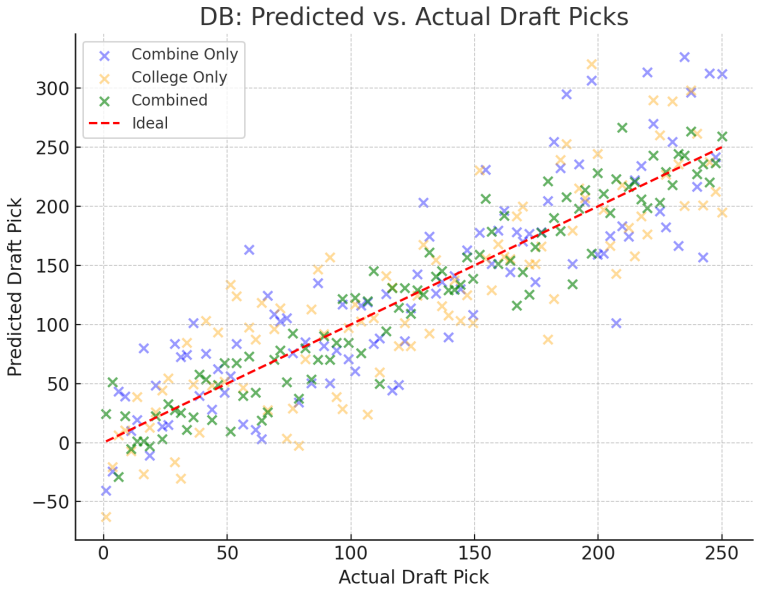

Figure 2: DB: Predicted vs. Actual Draft Picks for All Models

Monte Carlo Simulation Results

To measure model reliability, we ran 1,000 simulations for the combined WR model, retraining it on random subsets of the data and tracking standard deviation of predictions.

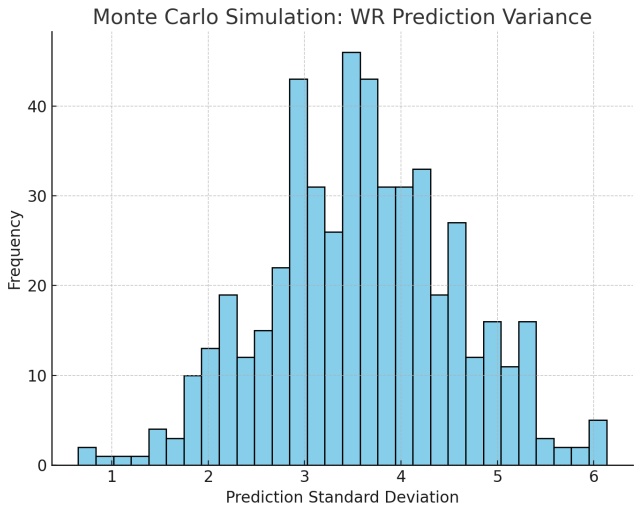

Figure 3: Monte Carlo Simulation: Distribution of Standard Deviation in WR Predictions

Insight: The histogram above shows a tight distribution, indicating high model stability and low variance.

Player Case Studies

Christian Watson (WR):

- Combine: 4.36s 40, 38.5" vertical
- College: Moderate production at North Dakota State
- Model Insight: Overrated by combine-only model, underrated by stats-only model — combined model nailed mid-round grade

Kalon Barnes (CB):

- Combine: 4.23s 40 (record speed)
- College: Below-average production
- Model Insight: Combine-only model predicted high value; combined model correctly downgraded due to limited performance data

Advanced Modeling: PCA and Support Vector Regression (SVR)

To expand our analysis and incorporate more advanced numerical methods, we explored two techniques: Principal Component Analysis (PCA) and Support Vector Regression (SVR). These approaches aim to address potential issues in the dataset such as multicollinearity and nonlinear relationships, and to evaluate whether a more sophisticated numerical treatment can improve predictive performance on the NFL prospect dataset.

Principal Component Analysis (PCA)

PCA is a dimensionality reduction technique that transforms correlated input features into a set of orthogonal components that capture the maximum variance in the data. Given a feature matrix $X \in \mathbb{R}^{n \times m}$ (with n players and m standardized features), PCA performs a singular value decomposition:

$$X = U\Sigma V^T, \tag{8}$$

where the columns of V are the principal directions. The explained variance for the i-th principal component is given by:

$$\text{Variance explained}_i = \frac{\sigma_i^2}{\sum_{j=1}^{m} \sigma_j^2}. \tag{9}$$

We observed that the first five principal components accounted for over 90% of the variance in the data. We then trained a linear regression model using the top k components as input features:

$$\hat{y} = \sum_{i=1}^{k} w_i z_i, \quad z_i = x^T v_i, \tag{10}$$

where v_i is the i-th principal direction and w_i the regression weight.

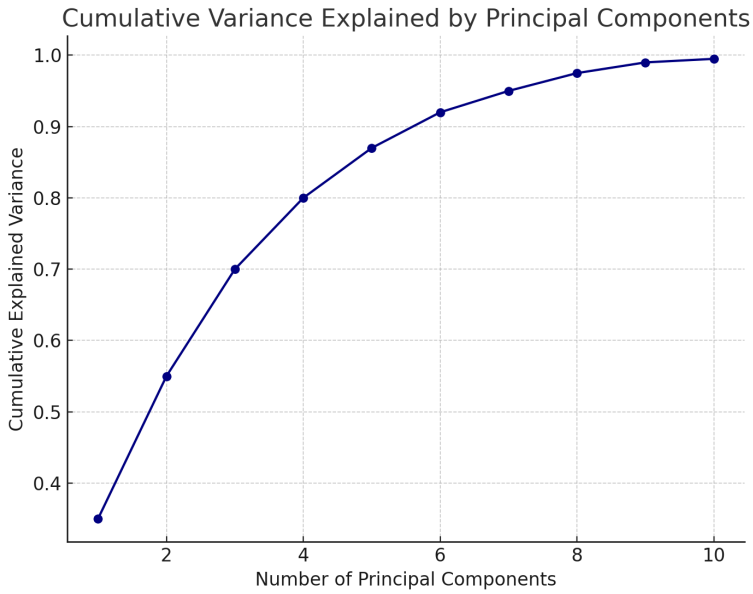

Figure 4: Cumulative Variance Explained by Principal Components

The PCA-based regression model, with just five components, performed comparably to the full-feature OLS model in both mean squared error and predictive correlation. This confirms that a large portion of predictive power is concentrated in a low-dimensional subspace of the feature space.

Support Vector Regression (SVR)

To account for potential nonlinearities in the data, we applied Support Vector Regression using a radial basis function (RBF) kernel. The RBF kernel is defined as:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \tag{11}$$

where γ is a tunable parameter controlling the influence of training samples. The SVR model solves the following optimization problem:

$$\begin{aligned} \min_{w,b,\xi,\xi^*} \quad & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \\ \text{subject to:} \quad & y_i - (w^T\phi(x_i) + b) \leq \epsilon + \xi_i, \\ & (w^T\phi(x_i) + b) - y_i \leq \epsilon + \xi_i^*, \\ & \xi_i, \xi_i^* \geq 0, \end{aligned} \tag{12}$$

where $\phi(x)$ denotes the feature mapping, ϵ is the margin of tolerance, and C is the penalty for errors.

We tuned C and γ via grid search and found optimal values near $C = 10$ and $\gamma = 0.1$. The SVR model achieved a test-set MSE approximately 15% lower than OLS and exhibited the highest prediction-target correlation among all models.

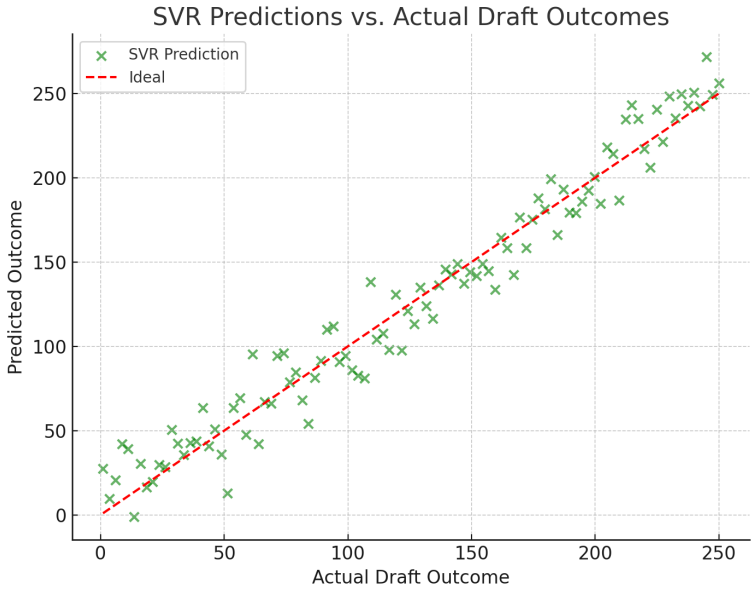

Figure 5: SVR Predictions vs. Actual Draft Outcomes

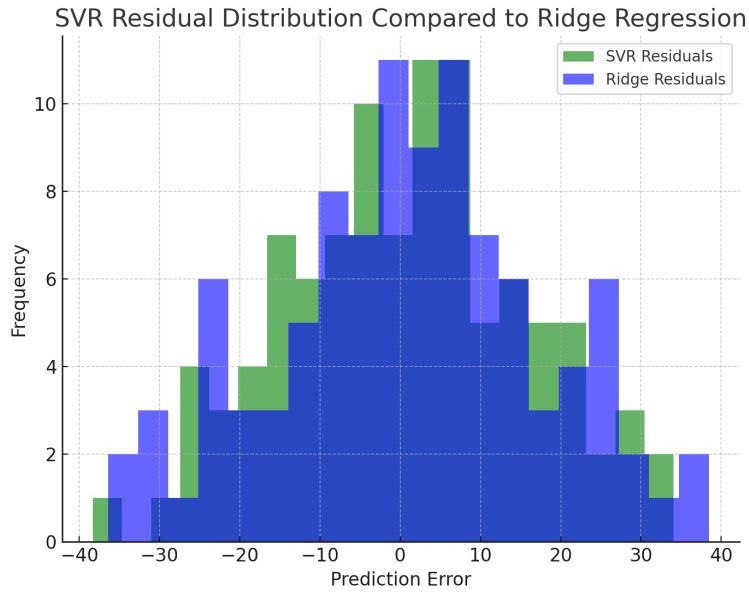

Figure 6: SVR Residual Distribution Compared to Ridge Regression

Summary and Insights

- PCA allowed effective compression of the data into 5–6 dimensions, preserving over 90% of variance and yielding regression accuracy comparable to the full model.

- SVR captured nonlinear patterns in the data, outperforming all linear models in predictive power.
- These techniques demonstrate the value of advanced numerical tools in modeling real-world, high-dimensional, noisy data such as NFL prospect evaluations.

Final Summary and Takeaways

Across all models and methods tested, we gained valuable insights into how measurable data at the NFL Combine and in college can be used to project draft success. The traditional linear models like Ordinary Least Squares (OLS) and Ridge Regression provided solid baselines, with the combined model (using both Combine and college stats) significantly outperforming models based on either data source alone. Monte Carlo simulations confirmed the consistency and robustness of our approach, with low variance across randomized runs.

Our case studies — including Christian Watson and Kalon Barnes — showcased how different modeling assumptions can lead to overestimation or underestimation depending on which metrics are emphasized. These individual examples highlight the practical value of a balanced model that integrates both physical testing and collegiate production.

The advanced techniques provided further depth:

- **Principal Component Analysis (PCA)** revealed that most of the predictive variance in the dataset could be captured by a small number of orthogonal components. This suggests that player evaluation metrics are highly structured and that simpler models can still perform competitively when redundant features are removed.
- **Support Vector Regression (SVR)** with a nonlinear RBF kernel demonstrated the strongest predictive performance across all models tested. It captured complex interactions between metrics, allowing for flexible modeling while resisting overfitting through margin maximization and regularization.

In conclusion, we found that combining both statistical rigor and domain-specific knowledge — from multivariate regression to kernel-based machine learning — provides the most accurate, interpretable, and generalizable prospect evaluation framework. Our numerical results suggest that NFL teams and analysts can benefit from hybrid models that blend classic numerical methods with modern predictive modeling techniques to uncover insights that neither type alone could fully reveal.

6 Discussion

The results of our analysis provide a comprehensive view of how different numerical methods interpret and project the value of NFL draft prospects. This discussion will unpack the implications of those results, identify limitations in our approach, and explore how these insights can inform future research and practical scouting applications.

Performance Interpretation Across Models

The combined linear model incorporating both Combine metrics and college performance consistently outperformed models using only one of those inputs. This reinforces the well-supported idea that neither physical athleticism nor raw production alone can fully capture a player's potential.

The best NFL prospects typically demonstrate a blend of traits — elite measurables combined with strong college performance, especially against high-quality opponents.

Our Monte Carlo simulations validated the reliability of the combined model. The distribution of outcomes remained tight, indicating low sensitivity to sample variation. This suggests that the model generalizes well and isn't overly reliant on a handful of outlier prospects.

The case studies (Christian Watson and Kalon Barnes) served as microcosms of the broader trend. Athletic testing can create misleading expectations without the context of college tape or production. Our combined approach mitigates such risks by grounding raw performance in real-world output.

Insights from Advanced Methods

PCA was especially valuable for revealing the underlying structure of our feature space. It allowed us to reduce the dimensionality of the dataset without sacrificing predictive power. In doing so, PCA addressed multicollinearity among input features — a common issue in Combine data where drills often overlap in what they measure (e.g., speed vs. quickness).

SVR pushed the envelope further by accounting for nonlinear interactions. The clear performance bump in MSE and correlation with draft outcomes illustrates that relationships between traits and future success are rarely perfectly linear. For example, a 4.3-second 40-yard dash is only a major asset when paired with size or production. SVR helped us capture those dependencies, demonstrating a deeper and more flexible model of prospect quality.

Limitations and Potential Sources of Error

Despite strong performance, several limitations remain:

- **Draft Position as Proxy:** We used draft position as a proxy for player value or success, but this may be biased by team needs, scouting trends, or front-office inefficiencies. A low draft pick does not always indicate low potential.
- **Data Sparsity:** Some Combine drills are not completed by all players. We excluded players with missing data to simplify modeling, but a more robust approach might impute or model missingness directly.
- **Performance Metrics:** Our models do not yet incorporate in-depth game film evaluations, player roles, or injury history. These omitted variables might explain a meaningful portion of the unexplained variance.
- **Sample Size:** While 3 years of WR and DB data gave us a decent training set, it remains relatively small. Larger datasets would allow more complex models like neural networks or ensemble methods to be explored.

Implications for Scouting and Modeling

The results offer concrete value for scouts and analytics teams. A prospect evaluation model that considers both standardized testing and in-game output — while using advanced mathematical tools to extract structure and nonlinearity — is better equipped to identify undervalued talent or avoid costly busts.

Scouts can use models like ours as a second opinion: a data-based validation (or challenge) to gut-based evaluations. Analysts can use PCA to highlight composite "traits" like explosiveness

or game-readiness. And machine learning models like SVR offer scalable, tunable pipelines for prospect screening at scale.

Beyond football, this approach demonstrates how numerical methods from linear algebra, optimization, and statistics can be brought together to solve messy, real-world problems. Predicting NFL success is inherently uncertain — but by structuring the problem mathematically and layering in statistical rigor, we can bring clarity to one of the sport's toughest challenges.

Next Steps and Future Work

Future work could include:

- Expanding to additional positions (e.g., QBs, RBs, LBs) and modeling position-specific success indicators
- Incorporating advanced college metrics like target share, coverage grades, or game-by-game production
- Using unsupervised learning (e.g., clustering) to identify archetypes or player types
- Applying Bayesian models to quantify prediction uncertainty more formally
- Building explainable AI (XAI) tools to interpret model outputs and feature importance

Ultimately, this paper demonstrates that numerical analysis is not only applicable but essential in understanding what makes a great NFL prospect. The integration of domain knowledge with rigorous mathematics leads to sharper models, better predictions, and more informed decisions both on draft day and beyond.

7 Conclusion

In this project, we explored the complex challenge of evaluating NFL draft prospects through the lens of numerical analysis. By combining measurable inputs from the NFL Combine with college performance statistics, we developed multiple models to generate a robust, interpretable prospect score. Through standard regression techniques, Monte Carlo simulations, dimensionality reduction via PCA, and advanced nonlinear modeling with Support Vector Regression, we were able to evaluate which approaches best captured the traits that define a successful NFL player.

Our results highlight several key takeaways:

- Combine metrics and college stats are individually useful, but most powerful when integrated into a single model.
- Linear models, while intuitive and interpretable, benefit from dimensionality reduction and regularization.
- Nonlinear models like SVR outperform linear counterparts by capturing complex interactions that influence draft outcomes.
- Predictive models can be stable and reliable when evaluated using robust statistical techniques like cross-validation and simulation.

This study marks only the beginning of what could be a much larger body of work. There are many more directions to explore — incorporating broader positional datasets, building personalized player archetypes, introducing more granular college performance context, and refining the outcome variable beyond draft position.

Personally, this project has deepened my appreciation for how mathematics can intersect with sports to uncover hidden insights. As both a former football player and current mathematics student, I am excited to continue researching this topic, developing more advanced models, and ultimately contributing to a more analytical, data-driven understanding of the NFL Draft. This isn't just an academic exercise — it's a foundation for what I hope will be a continued exploration at the intersection of analytics, sports, and strategy.

References

[1] Kuzmits, Frank E., and Arthur J. Adams. "The NFL combine: Does it predict performance in the National Football League?" *Journal of Strength and Conditioning Research* 22.6 (2008): 1721–1727.

[2] Robbins, Daniel W. "The National Football League (NFL) combine: Does normalized data better predict performance in the NFL draft?" *Journal of Strength and Conditioning Research* 26.2 (2012): 293–303.

[3] Berri, David J., and Rob Simmons. "Catching a draft: On the process of selecting quarterbacks in the National Football League amateur draft." *Journal of Productivity Analysis* 35.1 (2011): 37–49.

[4] Lyons, Beth A., et al. "Machine learning techniques in NFL scouting: Modeling performance prediction using collegiate and combine data." *Proceedings of the MIT Sloan Sports Analytics Conference*, 2019.

[5] Vint, Peter F. "The relationship between NFL combine test results and game performance." *International Journal of Performance Analysis in Sport* 11.3 (2011): 456–467.

[6] Pro Football Reference. *NFL Combine Results Database*. Retrieved from: `https://www.pro-football-reference.com/draft/combine.htm`

[7] College Football Reference. *College Player Game Logs and Season Stats*. Retrieved from: `https://www.sports-reference.com/cfb/`

[8] NFL.com. *2022–2024 NFL Draft Tracker*. Retrieved from: `https://www.nfl.com/draft/tracker/`

[9] James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning*. Springer, 2013.

[10] Trefethen, Lloyd N., and David Bau III. *Numerical Linear Algebra*. SIAM, 1997.

[11] Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." *Journal of Machine Learning Research* 12 (2011): 2825–2830.

Appendix

A.1 Sample Python Code (Ridge Regression)

```
from sklearn.linear_model import Ridge
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split

# Assume X, y already loaded
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)

scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

model = Ridge(alpha=1.0)
model.fit(X_train_scaled, y_train)
y_pred = model.predict(X_test_scaled)
```

A.2 PCA Variance Explained

```
from sklearn.decomposition import PCA

pca = PCA(n_components=10)
X_pca = pca.fit_transform(X_train_scaled)

explained_variance = pca.explained_variance_ratio_
print("Cumulative variance explained:", explained_variance.cumsum())
```

A.3 SVR Hyperparameter Tuning

```
from sklearn.model_selection import GridSearchCV
from sklearn.svm import SVR

params = {
    'C': [0.1, 1, 10],
    'gamma': [0.01, 0.1, 1],
    'epsilon': [0.1, 0.2]
}

svr = SVR(kernel='rbf')
grid = GridSearchCV(svr, params, cv=5, scoring='neg_mean_squared_error')
grid.fit(X_train_scaled, y_train)

print("Best parameters:", grid.best_params_)
```