Season-Wise NBA Draft Ranking Using Collegiate Performance

Tenzin Tsundue (tt3124)
MSCS Courant
New York University

Aryaman Sharma (as20476)
MSCS Courant
New York University

Chuyang Ye (cy3064)
MSIS Courant
New York University

December 2025

1 Abstract

Predicting NBA draft order from pre-draft data is challenging due to uncertainty, contextual effects, and heterogeneous evaluation criteria. In this work, we formulate NBA draft prediction as a season-wise learning-to-rank problem and develop an end-to-end pipeline using collegiate performance statistics. Models are trained on draft years 2000–2024 and evaluated on the held-out 2025 draft class under temporally consistent splits. We compare pointwise, pairwise, and listwise ranking approaches using Spearman's rank correlation and pairwise accuracy as evaluation metrics. Listwise methods outperform pairwise and pointwise baselines, highlighting the importance of modeling within-season ranking context. We also study a complementary binary classification task to predict whether a prospect is drafted or undrafted, finding moderate discriminative ability but persistent limitations in identifying undrafted players. Overall, our results show that collegiate box-score statistics capture a substantial portion of draft ordering signal when appropriate ranking objectives are applied, while also revealing the influence of non-statistical factors. The proposed framework provides a transparent baseline for future work on context-aware and outcome-driven draft analytics.

2 Introduction

The NBA Draft is a high-stakes decision process in which teams allocate limited draft capital to select prospects expected to generate long-term on-court value. Despite the availability of extensive pre-draft information—including collegiate performance, physical measurements, age, and scouting evaluations—draft outcomes remain highly uncertain. Player development trajectories vary widely, competition levels differ across leagues, and teams weigh positional fit, upside, and organizational context differently. From an analytics perspective, this setting naturally favors a ranking-based formulation: teams are fundamentally deciding who should be selected before whom, rather than estimating an absolute measure of player quality.

In this work, we develop an end-to-end predictive pipeline for NBA draft analysis that explicitly models draft prediction as a season-wise learning-to-rank problem. Given a cohort of draft-eligible prospects from a particular year, the objective is to produce an ordering that aligns with historical selection patterns using only pre-draft collegiate performance statistics. This formulation emphasizes relative comparisons within a draft class and reflects the way draft decisions are made in practice, where players are evaluated in relation to their contemporaries rather than in isolation.

A key design choice in our approach is to preserve draft-year grouping throughout both training and evaluation. Each draft class is treated as a distinct ranking group, and models are trained using temporally consistent splits that mirror real-world deployment: historical draft years are used for training and validation, while an unseen future draft class is reserved for testing. This time-aware evaluation strategy prevents information leakage across seasons and provides a more realistic estimate of generalization performance than random splits that mix players from different eras.

Data quality and feature reliability pose additional challenges in sports analytics. Many commonly cited pre-draft features—most notably NBA Draft Combine measurements—suffer from substantial missingness and protocol changes over time. To ensure robustness and temporal consistency, we deliberately restrict our feature set to collegiate box-score statistics from players' final college seasons. These features capture scoring, efficiency, rebounding, defensive activity, and playing time, and are available with relatively stable coverage across the study period.

Within this framework, we systematically compare three classes of learning-to-rank methods: pointwise, pairwise, and listwise approaches. Pointwise models score players independently and induce rankings through sorting, providing interpretable and computationally efficient baselines. Pairwise models learn relative preferences between pairs of players within the same draft year, directly optimizing local ordering consistency. Listwise models operate at the level of entire draft-year lists, optimizing objectives that reflect global ranking structure. Performance is evaluated using rank-based metrics—Spearman's rank correlation and pairwise accuracy—computed within each draft class.

In addition to draft-order prediction, we study a complementary binary classification task that predicts whether a prospect is drafted or undrafted based on collegiate performance alone. This task provides insight into the limits of box-score features for capturing draftability and highlights the role of non-statistical factors in selection decisions.

Overall, this work contributes a transparent and reproducible framework for NBA draft ranking that respects temporal structure, emphasizes relative ordering, and enables principled comparison across ranking paradigms. The results illustrate both the explanatory power and the limitations of collegiate statistics for draft prediction, providing a foundation for future extensions that incorporate richer context, downstream career outcomes, and uncertainty-aware decision support.

3 Literature Review

Research on NBA draft analytics addresses multiple related but distinct objectives, including predicting the revealed draft order, estimating whether a player will be drafted, and forecasting post-draft professional performance. These tasks are not interchangeable: draft order reflects teams' decision-making under uncertainty and contextual constraints, whereas outcome prediction aims to measure underlying player value. Consequently, methods optimized for performance forecasting or draftability classification do not directly solve the problem of season-wise draft ranking.

Szombat *et al.* explicitly model draft-order prediction by framing it as a pointwise learning-to-rank problem using collegiate performance data [1]. Their approach treats draft position as a continuous target and reports improved MAE and R^2 relative to baseline regression models, showing that ranking-aware formulations can outperform naive regression. However, because predictions are optimized independently for each player, the method does not explicitly encode within-draft ordering constraints or emphasize errors at the top of the draft.

Other studies focus on identifying factors associated with draft outcomes rather than reconstructing complete rankings. El-Hajj *et al.* analyze NCAA player statistics to determine which metrics influence the likelihood of being drafted, highlighting offensive indicators as more influential than certain defensive metrics [2]. While this work improves understanding of draft-related signals, its primary target is draftability rather than producing a season-wise ordering of prospects.

A separate body of work centers on predicting post-draft performance. Czasonis *et al.* propose a relevance-based framework to forecast NBA outcomes using similarities between prospects and previously drafted players [3]. This outcome-oriented perspective clarifies which pre-draft statistics translate to early NBA success, but it models underlying talent rather than the revealed draft order chosen by teams.

Empirical evaluations further show that draft position is an imperfect proxy for long-term value. Koz *et al.* find that while earlier draft rounds correlate with greater career participation, draft order explains only a limited fraction of variance in professional outcomes [4]. This supports treating draft order as a noisy label shaped by uncertainty and contextual influences. Consistent with this view, Barden and Kozlak demonstrate that organizational factors and decision stakes affect draft selections, particularly for foreign players, independent of purely on-court performance [5].

Related work has also examined how teams should evaluate prospects across heterogeneous pathways. Lewis proposes a statistics-based framework that standardizes evaluation across NCAA, international, junior college, and G League entrants, emphasizing age-, efficiency-, and context-adjusted metrics over raw box-score volume [6]. Although aimed at improving decision quality rather than predicting draft order, this work motivates careful feature design when modeling draft outcomes from collegiate data. Similarly, Cleary's analysis of NBA Draft Combine data shows that anthropometric and movement measures have selective predictive value, particularly for defensive impact, but explain only a modest share of overall NBA performance, suggesting that combine metrics are best used as complementary signals [7].

Finally, ranking-oriented evaluation has been explored in adjacent domains such as mock drafts. Fisher and Montague treat draft predictions as ranked lists and introduce rank-based distance measures that handle incomplete lists and emphasize top-of-draft accuracy [8]. Although their work does not map collegiate features to draft order, it highlights the importance of ranking-appropriate evaluation.

Overall, the literature demonstrates substantial progress in draft-related modeling but leaves a clear gap for season-wise learning-to-rank based on collegiate performance. Existing studies focus on pointwise draft-position prediction, draftability classification, or post-draft outcomes, and often rely on regression-style evaluation. This motivates modeling the NBA draft explicitly as a season-conditioned ranking problem and evaluating predictions in a manner consistent with ranked outputs.

4 DataSets

4.1 Data Sources

We integrate multiple publicly available datasets, each serving a distinct role in the draft-ranking pipeline:

- **NBA Draft and Combine Data (nba.com):** Historical NBA Draft records, including drafted players, draft order, team assignments, and related metadata, are collected from the official NBA website. NBA Draft Combine measurements are also obtained from the same source using the `nba_api` Python package (`https://github.com/swar/nba_api`), which provides programmatic access to official NBA statistics endpoints.
- **Collegiate Statistics for Drafted Players (Basketball-Reference):** Season-by-season NCAA performance statistics for drafted players are collected from Basketball-Reference (`https://www.basketball-reference.com/`). These data include scoring, efficiency, rebounding, defensive contributions, and playing time metrics.
- **Collegiate Statistics for Undrafted Combine Participants (Sports-Reference):** To obtain collegiate statistics for players who participated in the NBA Draft Combine but were not drafted, additional data are scraped from Sports-Reference (`https://www.sports-reference.com/`). Including these players allows us to analyze coverage differences and missingness patterns across data sources during preprocessing.

.

4.2 Data Coverage and Scope

The datasets collected for this study span the years 2000 through 2025. Each NBA Draft year typically consists of 60 selections, with one pick per team across two rounds. However, in certain years the total number of selections is reduced when teams forfeit draft picks due to league penalties or rule violations. For example, in the 2000 NBA Draft, only 58 players were selected because two teams lost their draft rights.

In addition to draft selections, many prospects participate in the annual NBA Draft Combine, an event designed to evaluate players' physical measurements, athletic performance, and on-court skill tests. The combine provides standardized metrics such as height, wingspan, agility times, vertical jump results, and shooting drill performance. However, not all drafted players attend the combine, and the set of tests administered has evolved over time, resulting in uneven coverage across seasons.

Furthermore, in each draft year there are players who are selected directly from high school or who enter the NBA from international professional leagues, such as the EuroLeague or Spain's top domestic league. For these players, collegiate statistics from U.S.-based NCAA competitions are unavailable, leading to missing college-performance data in the merged dataset.

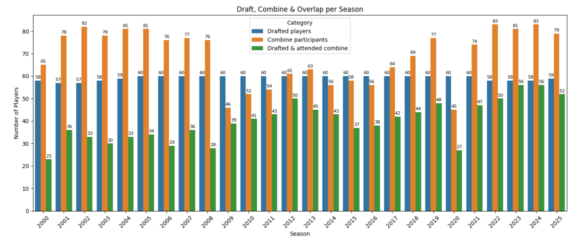

Figure 1: **Summary of NBA Draft Combine Statistics.** Year-by-year distribution (2000–2025) of drafted players, NBA Draft Combine participants, and their overlap (players who both attended the combine and were drafted). Across 2000–2025, the aggregated season-level totals are: drafted players (1,542), combine participants (1,795), drafted and attended combine (1,040), and attended combine but not drafted (755). The figure highlights substantial variation in combine attendance across seasons and a consistent gap between total combine participation and the number of players ultimately selected in the draft.

4.3 NBA Draft and Combine Data

The NBA Draft dataset contains one record per drafted player per season and provides the ground-truth ranking used in this study. Each entry includes a unique player identifier (`PERSON_ID`), player name, draft year (`SEASON`), round and pick information (`ROUND_NUMBER`, `ROUND_PICK`, `OVERALL_PICK`), and draft type. Team-related metadata such as team identifier, city, team name, and abbreviation are also included. In addition, the dataset records each player's pre-draft organization and organization type, distinguishing between collegiate programs, high school entrants, and other pathways.

Across the 2000–2025 period, the aggregated dataset contains the following season-level totals:

- Drafted players: 1,542
- NBA Draft Combine participants: 1,795
- Drafted & attended combine: 1,040
- Attended combine but not drafted: 755

Figure 1 illustrates the year-by-year distribution of drafted players, combine participants, and their overlap. The figure highlights substantial variation in combine attendance across seasons, as well as a consistent gap between total combine participation and the number of players ultimately selected in the draft.

As a result, the merged dataset exhibits both missing values and year-to-year variability arising from differences in combine participation, draft eligibility pathways, and the evolving structure of pre-draft evaluation. These characteristics motivate the preprocessing and feature-selection strategies described in the following section.

4.4 College Statistics Data

To obtain collegiate performance statistics for both drafted players and undrafted NBA Draft Combine participants, we rely on different data sources and access routes. College statistics for drafted players are collected from Basketball-Reference, while collegiate data for undrafted combine participants are obtained from Sports-Reference. Although sourced from different links, both repositories provide collegiate performance records with largely consistent statistical fields, enabling joint analysis after alignment.

After merging and filtering, the resulting dataset contains collegiate records for both drafted and undrafted players

- **Drafted players:** 1,210
- **Undrafted players:** 559

For each player, collegiate statistics may span multiple seasons depending on the length of their NCAA career. For the purposes of this study, we retain only each player's final college season, corresponding to the year in which they declared for the NBA Draft. Focusing on the final season provides the most up-to-date snapshot of a player's pre-draft performance and ensures fair comparison across players with different developmental timelines.

The resulting college statistics dataset provides season-level records for individual players and captures a broad range of performance indicators. These include demographic information such as age and collegiate affiliation, participation measures such as games played and minutes logged, cumulative production totals covering scoring, rebounding, playmaking, and defensive actions, and efficiency metrics such as shooting percentages. In addition, per-game averages for minutes, points, rebounds, and assists are included to normalize performance across different usage levels. Together, these features provide a comprehensive representation of each player's on-court contributions during their final collegiate season prior to draft entry.

4.5 Missing Data and Limitations

The NBA Draft Combine dataset contains a substantial amount of missing data for several reasons. First, not all drafted players participate in the NBA Draft Combine, resulting in missing combine measurements for many players who were ultimately selected. Second, the set of physical tests and on-court drills conducted at the combine has changed over time, leading to inconsistent feature availability across seasons. As a result, many combine-related variables exhibit high missingness and limited temporal consistency.

In addition, the dataset exhibits an inherent imbalance between drafted and undrafted players. In real-world draft scenarios, team decision-makers primarily evaluate and rank players who formally declare for the draft and are expected to be selected. However, in the collected data, the number of drafted players significantly exceeds the number of undrafted combine participants, creating an imbalance in representation. This reflects the structure of the draft process itself rather than a data collection artifact, but it nonetheless constrains the ability to model undrafted outcomes symmetrically.

Further limitations arise from changes in player development pathways over time. In the early 2000s, a nontrivial number of players entered the NBA directly from high school, while in more recent years prospects increasingly arrive from international professional leagues or alternative development programs such as the G League. For these players, U.S.-based collegiate statistics are unavailable, leading to additional missing data in the college performance dataset.

Together, these factors introduce missing values, distributional imbalance, and temporal heterogeneity into the merged dataset. These limitations motivate the preprocessing, feature selection, and modeling choices described in subsequent sections and should be considered when interpreting the ranking results.

5 Methodology

5.1 Task A: Draft Order Prediction

5.1.1 Problem Formulation

We formulate NBA draft prediction as a season-wise learning-to-rank problem. Each draft year defines a ranking group consisting of all players selected in that season, with the observed draft order serving as the ground-truth ranking. The goal is to learn a scoring function that ranks players within the same draft year using only pre-draft collegiate performance features.

Models are trained using data from draft years 2000 through 2024. The final evaluation is performed on the 2025 draft class, which is held out entirely from training. This setup reflects the real-world use case in which historical draft data is used to rank prospects in an unseen future draft year while preserving temporal consistency.

5.1.2 Data Preprocessing

Data preprocessing consists of feature selection and feature standardization steps designed to improve robustness, interpretability, and training stability.

Feature Selection We excluded NBA Draft Combine data from our feature set because many players do not participate in the combine, resulting in substantial missing values. In addition, the combine tests themselves have changed over the years, leading to inconsistent measurements across seasons. To avoid noise and bias from these irregularities, we relied solely on college statistics moving forward.

To reduce redundancy and avoid multicollinearity, we computed the Pearson correlation between each feature and the target variable (*Overall Pick*). Features within the same statistical category were often highly correlated; therefore, from each group, only one representative feature was retained. This ensures that the final feature set remains compact, interpretable, and statistically robust.

The final feature selection is summarized below:

1. **Scoring**
   Original Features: `Totals_FG`, `Totals_PTS`, `Totals_FGA`, `Totals_FT`, `Totals_FTA`, `Per Game_PTS`
   Selected Feature: `Totals_FG`

2. **Free Throws**
   Original Features: `Totals_FT`, `Totals_FTA`
   Selected Feature: `Totals_FT`

3. **Rebounding**
   Original Features: `Totals_TRB`, `Totals_ORB`, `Per Game_TRB`
   Selected Feature: `Totals_TRB`

4. **Defense**
   Original Features: `Totals_BLK`, `Totals_STL`
   Selected Features: `Totals_BLK`, `Totals_STL`

5. **Efficiency**
   Original Features: `Shooting_FG%`
   Selected Feature: `Shooting_FG%`

6. **Turnovers**
   Original Features: `Totals_TOV`
   Selected Feature: `Totals_TOV`

7. **Playing Time**
   Original Features: `MP`, `G`
   Selected Feature: `MP`

8. **Personal Fouls**
   Original Features: `Totals_PF`
   Selected Feature: `Totals_PF`

9. **Age / Experience**
   Original Features: `Age`
   Selected Feature: `Age`

Abbreviation	**Full Form**
FG	Field Goals
FGA	Field Goal Attempts
FT	Free Throws
FTA	Free Throw Attempts
PTS	Points
TRB	Total Rebounds
ORB	Offensive Rebounds
BLK	Blocks
STL	Steals
TOV	Turnovers
PF	Personal Fouls
MP	Minutes Played
G	Games Played
FG%	Field Goal Percentage

Table 1: Full Forms of Statistical Abbreviations

Feature Standardization

To place all features on a comparable scale, we applied z-score normalization to the dataset. For each fold in cross-validation, the mean and standard deviation were computed exclusively from the training seasons, and these statistics were then used to standardize both the training and test sets. This prevents information leakage and ensures that no information from future seasons influences the scaling of earlier data.

The z-transformation used in this process is defined as:

$$z = \frac{x - \mu}{\sigma},$$

where x is the raw feature value, μ is the mean of the feature in the training data, and σ is the corresponding standard deviation. Standardizing the inputs in this manner improves model stability, accelerates convergence, and ensures that the ranking algorithm focuses on meaningful differences between players rather than differences in feature magnitude.

5.1.3 Ranking Algorithms

For ranking tasks, we explore pointwise, pairwise, and listwise approaches.

5.1.3.1 Pointwise Ranking As a baseline approach, we employ *pointwise* learning-to-rank methods. In the pointwise formulation, each player is scored independently by a regression model, and rankings are obtained by sorting these scores within each draft year. Unlike pairwise or listwise methods, pointwise ranking does not explicitly model relative comparisons during training; instead, it relies on the model's ability to learn a monotonic mapping from player features to draft value.

Formulation: Let $x_i \in R^d$ denote the feature vector of player i in a given draft year. A pointwise ranker learns a scoring function

$$s_i = f(x_i),$$

where higher scores indicate earlier draft selection. We define the regression target as the negative overall pick,

$$y_i = -OverallPick_i,$$

so that earlier selections correspond to larger target values. At inference time, players within the same season are ranked by sorting s_i in descending order.

Models. We evaluate five representative pointwise rankers. These include four classical regression models—Ridge regression, Random Forest, Extremely Randomized Trees (ExtraTrees), and Histogram-based Gradient Boosting (HistGB)—as well as a neural pointwise ranker implemented as a multilayer perceptron (MLP). Ridge regression serves as an interpretable linear baseline, while the tree-based ensemble methods capture nonlinear interactions among collegiate performance features. The MLP provides a neural baseline that learns a nonlinear scoring function through stacked fully connected layers and gradient-based optimization.

Neural pointwise ranker. The MLP consists of multiple fully connected layers with ReLU activations and a final linear output layer that produces a single scalar score per player. The network is trained using a regression loss (Smooth L1 loss) to predict y_i, and rankings are induced by sorting the predicted scores within each season. This architecture represents the most common neural approach to pointwise ranking and serves as a bridge between classical regression baselines and more complex neural ranking models.

5.1.3.2 Pairwise Ranking To model draft ordering with a comparison-based objective, we employ *pairwise* learning-to-rank methods. In the pairwise formulation, the training signal is derived from relative preferences between two prospects from the same draft year. This matches the structure of draft decisions, where teams implicitly compare players against alternatives available in the same class.

Formulation: Let $x_i \in R^d$ denote the feature vector for player i in a given season, and let $pick_i$ be the observed overall pick (smaller is better). For any two players i and j from the same season, we define a binary preference label

$$y_{ij} = \begin{cases} 1 & \text{if } pick_i < pick_j \quad (i \succ j), \\ 0 & \text{otherwise.} \end{cases}$$

We train models on difference vectors $\Delta x_{ij} = x_i - x_j$, so that the classifier learns to predict whether i should be ranked ahead of j. Because the number of within-season pairs grows quadratically, we cap the number of training pairs per season to ensure tractable training and stable run time.

Models. We evaluate three pairwise rankers:

1. **RankSVM:** a linear support vector classifier trained on Δx_{ij} with hinge loss, producing a linear scoring function $s(x) = w^\top x$.
2. **Pairwise Logistic:** a linear logistic regression classifier trained on Δx_{ij} with log-loss, also yielding $s(x) = w^\top x$.
3. **RankNet:** a neural pairwise model that learns a nonlinear scoring function $s(x) = f_\theta(x)$ and minimizes a pairwise logistic loss on score differences.

Season-wise Pair Data Preparation All pairwise training data are constructed by grouping players within the same draft season from 2000 to 2024. The resulting models are then evaluated on the held-out 2025 draft class. This season-wise grouping preserves temporal structure and reflects a more realistic real-world draft prediction setting.

5.1.3.3 Listwise Ranking To model the ordering of players within each NBA draft class, we employ *listwise* learning-to-rank methods. Unlike pointwise regression (which scores players independently) or pairwise methods (which learn from sampled comparisons), listwise ranking optimizes the quality of an *entire* within-season ordering. This is a natural fit for the NBA Draft, where teams evaluate a cohort of prospects in the same year and decisions depend on relative comparisons among all available players.

Season-wise grouping. Each season s is treated as a ranking group containing all prospects in that draft year, with the observed *Overall Pick* defining the ground-truth order. The model learns a scoring function $f(x_i)$ such that for players i and j within the same season, the predicted ordering matches the observed draft ordering as closely as possible.

Listwise models. We evaluate three representative listwise algorithms: ListNet, ListMLE, and LambdaMART. ListNet and ListMLE are probabilistic listwise methods that optimize ranking quality by learning distributions over permutations (or top-weighted approximations), while LambdaMART is a boosted decision tree ranker that directly optimizes a listwise surrogate of ranking quality through gradient-boosting updates. All models operate on the same standardized college feature set used throughout our pipeline.

5.2 Task B: Draftability Prediction

In addition to predicting within-draft ordering via ranking, we study a complementary binary classification task: given a prospect's final-season college statistics, can we predict whether the player will be drafted (selected in picks 1–60) or undrafted? This classifier serves as a screening module that estimates draft likelihood before constructing a full draft board.

5.2.1 Problem Setup and Label Definition

Let $x_i \in R^d$ denote the feature vector corresponding to player i's final-season college performance. We define the binary target variable as

$$y_i = \begin{cases} 1 & \text{if } \mathit{OverallPick}_i \leq 60 \quad \text{(drafted)}, \\ 0 & \text{otherwise} \quad \text{(undrafted)}. \end{cases}$$

After merging collegiate statistics for both drafted players and undrafted NBA Draft Combine participants, the resulting dataset contains 1,769 players, consisting of 1,210 drafted and 559 undrafted prospects, resulting in a moderately imbalanced class distribution.

We employ the same compact, college-only feature set used in the ranking pipeline, consisting of box-score production, efficiency, and usage indicators: `Totals_FG`, `Totals_FT`, `Totals_TRB`, `Totals_STL`, `Totals_BLK`, `Totals_TOV`, `Totals_PF`, `Shooting_FG%`, and `MP`. All features are standardized using statistics computed exclusively on the training data to prevent information leakage.

5.2.2 Temporal Split and Evaluation Protocol

To reflect real-world deployment, we adopt a season-wise temporal split. Earlier draft seasons are used for training, while the most recent five seasons are held out as an unseen test set. This setup ensures that no future-season information influences training and mirrors how a model would be applied to an upcoming draft class.

We evaluate classification performance using multiple complementary metrics. ROC-AUC is used to measure the model's ability to discriminate between drafted and undrafted players across classification thresholds. In addition, F1-score, precision, recall, and confusion matrices are reported to assess class-specific performance under imbalance.

5.2.3 Model

We use a GradientBoostingClassifier, a tree-based ensemble method that captures nonlinear interactions between collegiate performance features. Model hyperparameters are selected using randomized search with stratified cross-validation on the training split. Gradient boosting provides a strong baseline for tabular sports data while remaining robust in moderate-data regimes.

5.2.4 Handling Class Imbalance

Because drafted players are more prevalent than undrafted players, naive accuracy-based evaluation can be misleading. To diagnose the effect of class imbalance, we evaluate the classifier under two complementary conditions: (i) using the natural class distribution of the test set, and (ii) using a balanced test set created via random under-sampling to equalize drafted and undrafted examples. This dual evaluation highlights how performance shifts when equal importance is placed on identifying both classes.

6 Experimental Results

6.1 Task A: Draft Order Prediction

6.1.1 Experimental Setup

All models are evaluated under a temporally consistent, season-wise holdout protocol. Training is performed using data from NBA draft years 2000 through 2024, while the 2025 draft class is held out exclusively for testing. This evaluation strategy mirrors the real-world deployment scenario in which a model trained on historical drafts is applied to rank prospects in an unseen future draft year.

Within each draft year, ranking performance is assessed by comparing the model-predicted ordering against the observed draft order. Evaluation metrics are computed separately for each season and then reported for the held-out 2025 test year. By enforcing a strict temporal split, this setup prevents information leakage from future seasons and provides a realistic estimate of generalization performance.

We evaluate overall ranking quality using two complementary metrics: Spearman's rank correlation coefficient (ρ) and pairwise accuracy. Spearman's ρ measures the degree of monotonic agreement between the predicted ranking and the ground-truth draft order within a season, capturing how well the global ordering of prospects is recovered. Pairwise accuracy evaluates the fraction of correctly ordered player pairs, providing a more local measure of ranking consistency that reflects how often the model correctly prefers one prospect over another.

Both metrics are computed within the held-out 2025 draft class by comparing the model-predicted scores to the observed draft order. Using these two measures together allows us to assess both global ranking fidelity and local ordering correctness across the evaluated models.

6.1.2 Pointwise Ranking Results

Table 2 summarizes the performance of the pointwise ranking models on the held-out 2025 draft class. All models demonstrate substantial ability to recover the observed within-class ordering using collegiate box-score features alone, including both classical regression baselines and the neural pointwise ranker.

Table 2: Pointwise ranking performance on the 2025 draft class.

Model	Pairwise Acc.	Spearman's ρ
HistGB	0.736	0.663
Random Forest	0.734	0.641
Ridge	0.728	0.640
ExtraTrees	0.736	0.636
MLP	0.726	0.646

Among the classical models, Histogram-based Gradient Boosting achieves the strongest overall performance, indicating that nonlinear feature interactions play an important role in recovering draft order. Random Forest and ExtraTrees exhibit comparable pairwise accuracy, suggesting that ensemble averaging over decision trees provides stable local ordering even when global rank correlation differs slightly. Ridge regression performs competitively despite its linear structure, highlighting that a meaningful portion of draft ordering signal is linearly separable in collegiate statistics.

The neural pointwise ranker (MLP) achieves competitive performance relative to the classical baselines, demonstrating that a simple feedforward neural network can learn a useful scoring function for draft ranking. However, its performance does not consistently exceed that of the best tree-based models, which aligns with prior findings that gradient-boosted trees often outperform neural networks on small-to-medium tabular datasets.

Overall, the pointwise results establish a strong baseline across both classical and neural models: independent scoring approaches can recover a large fraction of the observed draft ordering when evaluated season-wise. Nevertheless, because pointwise methods do not explicitly optimize relative comparisons, they are normally outperformed by pairwise and listwise approaches in subsequent experiments, particularly in resolving ordering among closely ranked prospects.

6.1.3 Pairwise Ranking Results

Table 3: Pairwise learning-to-rank results on the held-out 2025 draft class (trained/validated on 2000–2024).

Model	Pairwise Acc.	Spearman's ρ
RankSVM	0.726	0.642
PairwiseLogistic	0.725	0.642
RankNet	0.718	0.630

The close agreement between RankSVM and Pairwise Logistic suggests that, for this feature set, the dominant ordering signal is well captured by a linear separator in the pair-difference space. In other words, a single weight vector w that increases scores for stronger statistical profiles is sufficient to reproduce a large portion of within-class preferences. RankNet provides additional nonlinear capacity, but does not outperform the best linear pairwise models in this setting, which is consistent with the small-to-medium tabular regime where linear and tree-based methods often remain highly competitive.

Compared to pointwise scoring, pairwise training directly optimizes relative comparisons and therefore reduces local inversion errors within crowded tiers of prospects. However, pairwise objectives still do not optimize the full list structure end-to-end, which can limit global coherence when many correlated comparisons interact. This motivates listwise approaches as the next step when the goal is to directly optimize permutation-level ordering quality for an entire draft class.

6.1.4 Listwise Ranking Results

Table 4: Listwise ranking performance on the 2025 draft class.

Model	Pairwise Acc.	Spearman's ρ
LambdaMART	0.743	0.660
ListMLE	0.735	0.656
ListNet	0.715	0.631

ListMLE consistently improves upon ListNet, suggesting that directly optimizing the likelihood of the observed permutation yields better recovery of within-year structure than the softer distributional objective used by ListNet. LambdaMART further improves upon both neural listwise methods, achieving the highest pairwise accuracy and the highest Spearman correlation.

A likely reason for LambdaMART's advantage is its ability to capture nonlinear interactions among collegiate statistics via boosted decision trees while still optimizing a listwise ranking objective. This combination helps resolve difficult ordering decisions within crowded tiers (where many players have similar box-score profiles), improving both local correctness (pairwise accuracy) and global agreement (Spearman's ρ). Overall, these results support LambdaMART as the most reliable listwise approach for reproducing historical NBA draft orderings in this study.

6.1.5 Comparative Analysis of Ranking Algorithms Results

A comparison of the strongest models from each ranking paradigm—Histogram-based Gradient Boosting (pointwise), RankSVM (pairwise), and LambdaMART (listwise)—shows that LambdaMART delivers the best overall performance on the held-out 2025 draft class. The pointwise HistGB model achieves the highest Spearman rank correlation ($\rho = 0.663$), but LambdaMART attains the highest pairwise accuracy (0.743) while maintaining a comparable Spearman correlation ($\rho =$

0.660). RankSVM performs moderately, with pairwise accuracy of 0.726 and $\rho = 0.642$.

This distinction is important in the draft context, where correct local ordering of closely matched prospects often outweighs marginal gains in global rank correlation. While pairwise methods improve upon independent scoring, they do not optimize the full list structure. By directly optimizing draft-year permutations and capturing nonlinear feature interactions, LambdaMART reconciles correlated ordering decisions across the entire draft class, resulting in the most reliable overall reconstruction of historical NBA draft orderings in this study.

6.2 Task B: Draftability Prediction

6.2.1 Unbalanced Test Set (Natural Distribution)

On the held-out test set with the natural drafted-to-undrafted class distribution, the classifier demonstrates moderate discriminative ability. Performance is substantially stronger for drafted players than for undrafted players, indicating that collegiate box-score features more reliably capture signals associated with being selected than signals associated with not being selected.

Table 5: Binary drafted-vs-undrafted classification performance on the unbalanced test set (last five seasons).

Metric	Value
ROC-AUC	0.565
F1 (drafted, class 1)	0.701
Macro F1	0.527

The corresponding confusion matrix is

$$\begin{pmatrix} 39 & 39 \\ 104 & 168 \end{pmatrix},$$

where rows correspond to true labels (undrafted, drafted) and columns to predicted labels. The classifier exhibits high recall for drafted players but struggles to correctly identify undrafted prospects, frequently misclassifying them as drafted.

6.2.2 Balanced Test Set (Under-sampled 50/50)

To assess performance when both classes are weighted equally, we evaluate the classifier on a balanced test set obtained via random under-sampling. Under this more challenging setting, overall accuracy and drafted-class F1 decrease, while performance on the undrafted class improves relative to the unbalanced evaluation.

Table 6: Binary drafted-vs-undrafted classification performance on a balanced (50/50) under-sampled test set.

Metric	Value
ROC-AUC	0.577
F1 (drafted, class 1)	0.528
F1 (undrafted, class 0)	0.595
Macro F1	0.562
Accuracy (balanced)	0.564

The confusion matrix for the balanced test set is

$$\begin{pmatrix} 50 & 28 \\ 40 & 38 \end{pmatrix}.$$

These results show that the classifier identifies undrafted players more frequently when class imbalance is removed, but misclassifications remain common in both directions.

6.2.3 Comparative Analysis of Classifier Results

Comparing the unbalanced and balanced evaluation settings highlights the limitations of collegiate box-score features for draftability prediction. Under the natural class distribution, the classifier achieves stronger performance for drafted players (F1 = 0.701) but exhibits weak overall discrimination (ROC-AUC = 0.565), frequently misclassifying undrafted prospects. When evaluated on a balanced test set, performance on undrafted players improves (F1 = 0.595), but gains in overall discrimination remain marginal (ROC-AUC = 0.577).

These results suggest that the baseline classifier is not primarily limited by class imbalance, but by insufficient feature information and restricted undrafted data coverage. Draft selection decisions depend on factors such as physical attributes, competition level, positional context, and scouting evaluations that are not captured by box-score statistics alone, limiting the effectiveness of this baseline approach.

7 Conclusion

We studied NBA draft prediction through a season-wise learning-to-rank framework, treating each draft year as a grouped ranking problem and evaluating models under a temporally consistent holdout. Training on draft years 2000–2024 and testing on the held-out 2025 class, we built a reproducible pipeline based on final-season collegiate box-score statistics.

Across pointwise, pairwise, and listwise paradigms, listwise objectives produced the most reliable overall ordering, with LambdaMART achieving the highest pairwise accuracy on the 2025 class while strong pointwise baselines remained competitive in rank correlation. Pairwise methods were comparable but did not surpass the best pointwise model in this feature regime. These results indicate that modeling within-season ranking structure is beneficial, especially for resolving local ordering among similarly profiled prospects.

We also examined a drafted-vs-undrafted classification task and observed only modest discriminative ability (ROC-AUC $\approx$ 0.56–0.58), with persistent difficulty identifying undrafted prospects from box-score features alone. This limitation is consistent with incomplete coverage of undrafted players and with important pre-draft signals (e.g., athletic testing, competition level, and scouting context) not captured by our inputs.

Overall, our findings show that collegiate statistics contain meaningful draft-order signal when optimized with ranking objectives, while remaining variance underscores the role of contextual and non-statistical factors. The proposed framework provides a transparent baseline for future work that incorporates richer features, pathway coverage, and outcome-driven targets.

8 Future Work

Several extensions of this work are promising. Incorporating team context and positional needs would enable the transition from global draft rankings to team-specific draft boards that more closely reflect real draft decision-making. Extending the prediction targets beyond observed draft order to downstream outcomes—such as rookie playing time, early-career value metrics, or All-Star probability—would further help separate organizational bias from underlying player potential.

The binary draftability classifier introduced in this work can also be improved and expanded. Future directions include incorporating competition-strength adjustments, and position-aware representations to better distinguish drafted and undrafted prospects. Joint modeling of classification and ranking objectives may further improve performance by leveraging shared structure between draft likelihood and relative ordering.

Finally, future work may address data limitations by integrating international league statistics and alternative development pathways, including the G League Ignite program. Incorporating uncertainty-aware classification and ranking methods, as well as alternative listwise objectives, could further improve decision support by explicitly quantifying confidence and risk in draft recommendations.

References

[1] B. Szombat, N. Moniz, and G. Kertész, "Predicting nba draft rankings with machine learning: A pointwise learning to rank approach," in *2025 IEEE 29th International Conference on Intelligent Engineering Systems (INES)*. IEEE, 2025, pp. 000 347–000 352.

[2] M. El-Hajj, B. Kwon, C. J. Infante, J. Steed, V. Gore, N. Phan, and M. Elmorsy, "Leveraging machine learning to predict factors that drive successful basketball team formation," in *Fifth Symposium on Pattern Recognition and Applications (SPRA 2024)*, vol. 13540. SPIE, 2025, pp. 73–80.

[3] M. Czasonis, M. Kritzman, C. Kulasekaran, and D. Turkington, "How to predict the performance of nba draft prospects," 2023.

[4] D. Koz, J. Fraser-Thomas, and J. Baker, "Accuracy of professional sports drafts in predicting career potential," *Scandinavian journal of medicine & science in sports*, vol. 22, no. 4, pp. e64–e69, 2012.

[5] J. Q. Barden and S. Kozlak, "Resource deprivation, decision stakes, and the selection of foreign players in the nba draft," *Sport Management Review*, vol. 27, no. 2, pp. 175–196, 2024.

[6] L. Lewis, "A statistics-based framework for evaluating and drafting nba prospects: A quantitative study," Ph.D. dissertation, Capitol Technology University, 2024.

[7] A. Cleary, "Predictive ability of the nba draft combine in elite basketball players," Ph.D. dissertation, Concordia University Chicago, 2024.

[8] J. D. Fisher and C. Montague, "Improving the aggregation and evaluation of nba mock drafts," *Journal of Quantitative Analysis in Sports*, vol. 21, no. 4, pp. 327–343, 2025.