**Predicting basketball players’ shooting performance in biomechanics using Gaussian Naïve Bayes and Bernoulli Naïve Bayes Machine learning**

# Introduction/Background

Scoring performance in basketball affects the outcome of the game and determines how the team performs. The study of motion of the body is called biomechanics and it helps us to understand how different player body movements or actions contribute to successfully scoring or missing a shot. Sensors can collect data about movement during a shot, including how fast the hand moves, the angle of the shot, or the speed of release (Cabarkapa et al., 2023).

Through machine learning allows computer to study data enabling them to make prediction or choices without programming (Samuel, 1959; Jordan & Mitchell, 2015).This research is undertaken to show how machine learning through Gaussian Naïve Bayes and Bernoulli Naïve Bayes Machine) can be used to predict basketball shooting success using data from Biomechanical Basketball Shooting Dataset from Kaggle, which includes acceleration, angular velocity, and other movement metrics captured by an MPU-6050 wristband sensor.

# Problem Statement and Purpose of Study

## Problem Statement

Basketball shooting performance and techniques have largely been studies through biomechanics (Okazaki et al., 2015; Button et al., 2003), few studies predict shot success using machine learning techniques like Naïve Bayes classification. There’s only few evidence on how biomechanical variables—like acceleration, release speed, or follow-through angle— affect shot success. This gap can be realized when looking at player-specific traits such as age, height, and weight. Studies like Zhao et al. (2021) highlight the predictive power of machine learning in soccer, but these methods are yet to be used in basketball.

## Purpose of Study

The purpose of the study is to bridge the gap by developing a predictive model that combines biomechanics and personal variables using Naïve Bayes. The purpose of this study is to use Gaussian and Bernoulli Naïve Bayes classifiers to evaluate biomechanical variables and predict shot outcomes in basketball.

# Research Aim and Objectives

## Research Aim

To make Machine learning model that can predict the likelihood of scoring a basketball shot using Gaussian and Bernoulli Naïve Bayes classifiers and biomechanical data.

## Research Objectives

1. To predict shooting accuracy using Gaussian Naïve Bayes classifier.
2. To determine best biomechanical combination for different shot types.
3. To suggest suitable shot types and improvement methods for players based on age, weight and height.

# Research Questions

1. How accurate is Bernoulli and Gaussian Naïve Bayes in predict types of shots and success of shots?
2. What biomechanical combinations are suitable for different types of shots?
3. How can players improve shooting success and what shots should they focus on or avoid based on their height, weight and age?

# Preliminary Literature Review

Naïve Bayes classifiers have been used in sports analysis because of their simplicity and ability to interpret easily, especially in binary classification tasks . Burhaein et al. (2024) showed the practical application of Naïve Bayes in classifying physical fitness levels of athletes using demographics and body measurements. Their model achieved high predictive accuracy in grouping VO₂Max levels, offering a cheaper substitution to conventional fitness test, which often require lab-based testing equipment. Zou et al. (2022) used a Bernoulli Naïve Bayes to predict goal outcomes in soccer by examining each time segment of a match as a binary event (goal or miss). Their model showed high reliability in predicting match outcomes in later stages of the match. This shows the value of Naïve Bayes in live sports prediction.

Research papers in basketball biomechanics have shown kinematic variables—such as release angle, wrist acceleration, shoulder rotation, and shot speed—as critical predictors of shooting success. For instance, Cabarkapa et al. (2023) used high-speed motion capture to confirm that consistent release angles and optimal wrist kinematics are significantly associated with higher shooting accuracy in elite players. Similarly, Zhang et al. (2022) emphasized the role of coordinated lower and upper body mechanics in maintaining shot stability, especially under defensive pressure. Pan et al. (2021) used computer vision to automatically extract these biomechanical features from video footage and match them with shot outcomes.

From the above research, a gap has been identified, which is the integration of classifiers—such as Gaussian and Bernoulli Naïve Bayes-with biomechanical data. There is limited research studying how personal attributes like age, height, or experience interact with biomechanical variables to influence shot success. This leave an opportunity for a research that combines machine learning and biomechanics of an individual player to not only predict shot outcomes but also make recommendations for improved shot success based on their age, weight and height.

.

# Research Methodology

This study will use machine learning to predict if a basketball shot will be made or missed, based on biomechanical data. The data will come from the Biomechanical Basketball Shooting Dataset Kaggle. (Ziya ,2025), collected using a wristband sensor (MPU-6050) during different shot types like layups, jump shots, and set shots.

The model will use the following features:

* Motion data: Shot Angle, Accelerometer X/Y/Z m s2, Gyroscope X/Y/Z, Shot Speed, Release Angle, Release Timing, Follow Through Angle and Arc Radius.
* Context data: Distance from Basket, Shot Type
* Player info: Age, Height , Weight
* Target: Shot Result (1 = made, 0 = missed)

Extra fields like Player ID and Timestamp will be used for grouping but not for predictions.

Two machine learning models will be developed:

* **Gaussian Naïve Bayes** will use the continuous variables such as speed, angle, and acceleration. **Bernoulli Naïve Bayes** will use binary variables.
* Models will be developed through programing.
* The data will first be checked for duplicate rows. Then, it will be examined by scaling the continuous values and converting selected features into binary form and loaded into the programs. The data will be split into 50% training and 50% testing. Model performance will be evaluated for goal output with propability calculations.

# Significance of research

The research will enable players and coaches in determining best physical actions that give them a better chance of scoring. This is important for players, especially during training, to better refine their skills. Coaches can develop upcoming athletes to master the skills and techniques, to identify natural attacking players.

# Limitations and Delimitations of research

The machine learning models will be strictly limited to Gaussian Naïve Bayes and Bernoulli Naïve Bayes. The model cannot be expected to predict scoring probabilities with biomechanical variables that are out of the scope of this research, application is only limited to basketball.

# Results

## **1. Model Performance**

Four classifiers (GaussianNB, BernoulliNB, Logistic Regression and KNN) were tested on a 5-fold cross-validation. Measurements of performance consisted of Accuracy, ROC AUC and Average Precision (AP).

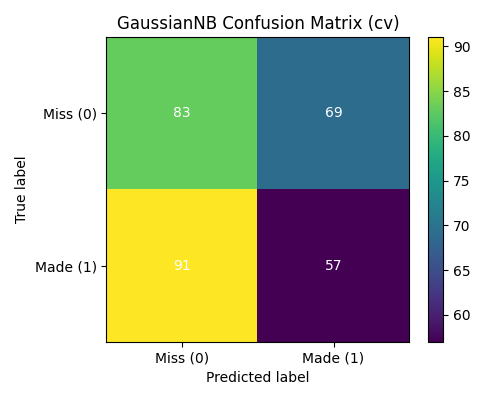
**models\_summary.csv**

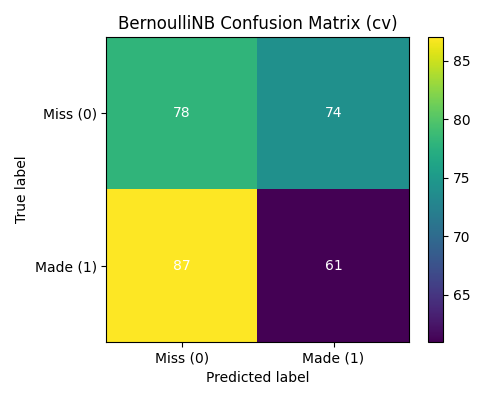
| **model** | **accuracy\_cv** | **roc\_auc\_cv** | **avg\_precision\_cv** |
| --- | --- | --- | --- |
| **GaussianNB** | **0.4666666666666667** | **0.430876600284495** | **0.4594163899285108** |
| **BernoulliNB** | **0.4633333333333333** | **0.3979818634423898** | **0.439680517383742** |
| **LogisticRegression** | **0.43333333333333335** | **0.39953769559032715** | **0.46095505327049424** |
| **KNN** | **0.47333333333333333** | **0.4657716927453769** | **0.4739619932085686** |

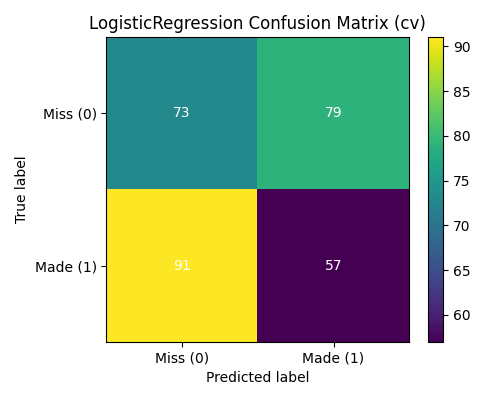
*Table 1: Cross-validation performance of classifiers on basketball shooting dataset)*

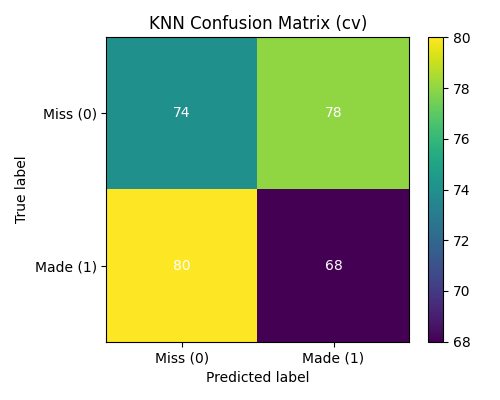
Key observations:

## **2. Confusion Matrices**

confusion\_GaussianNB.png

confusion\_BernoulliNB.png

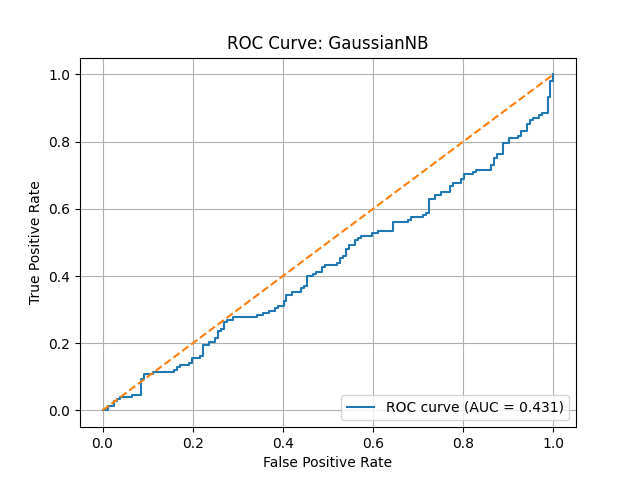
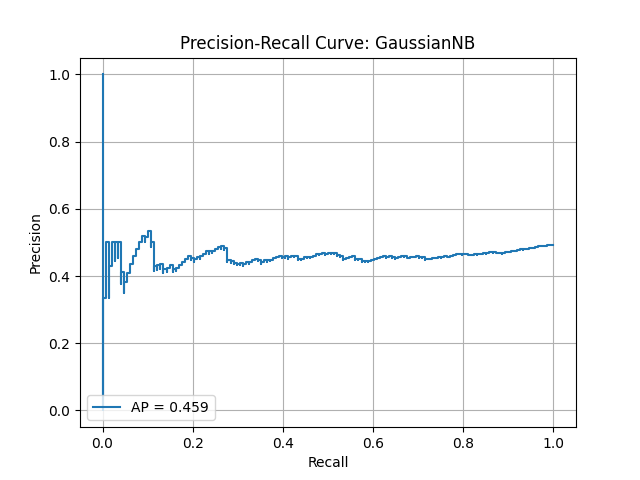
confusion\_LogisticRegression.png

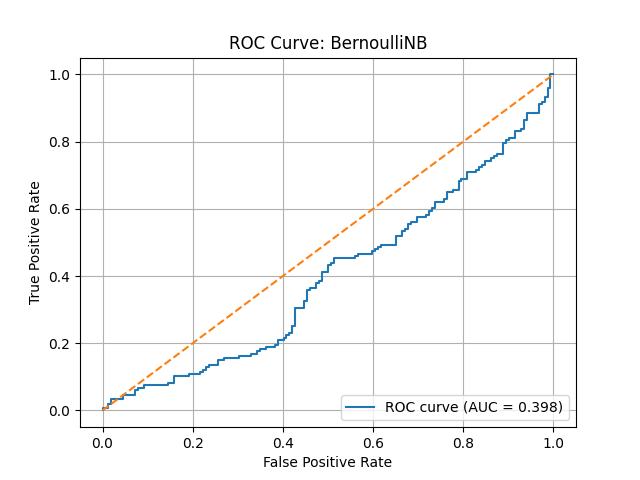
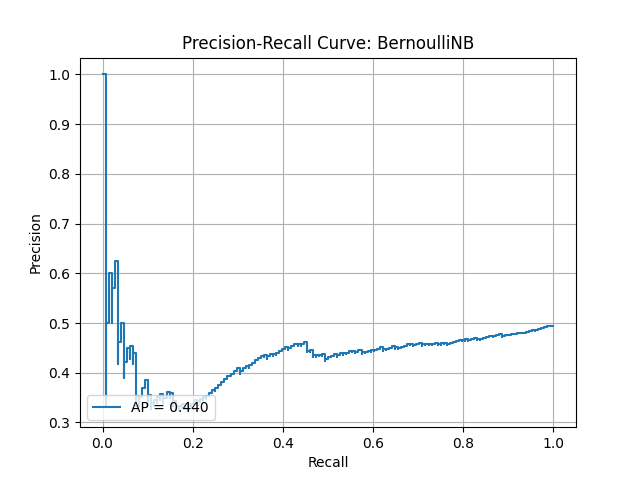
confusion\_KNN.png

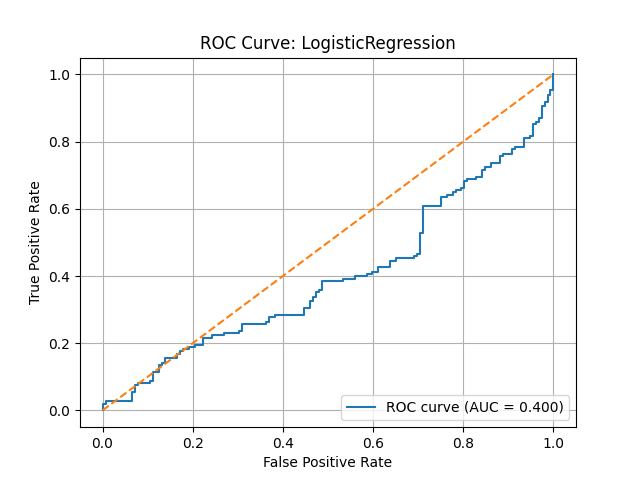
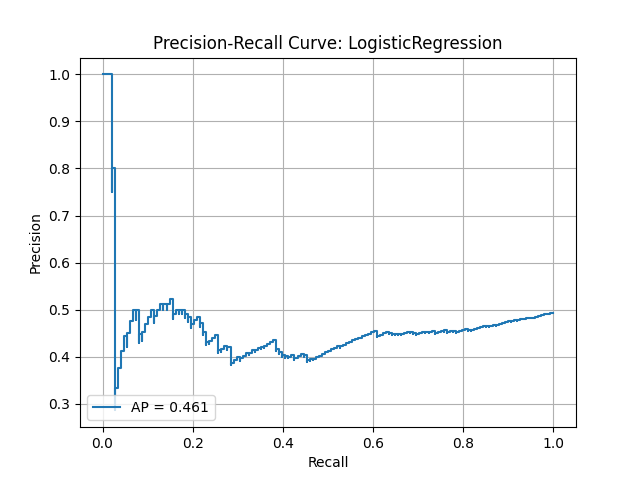
Observation: All models struggled to cleanly separate the classes, with many false positives and false negatives.

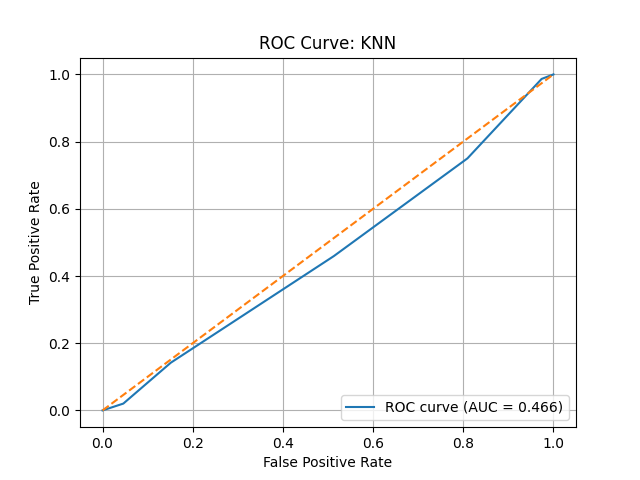
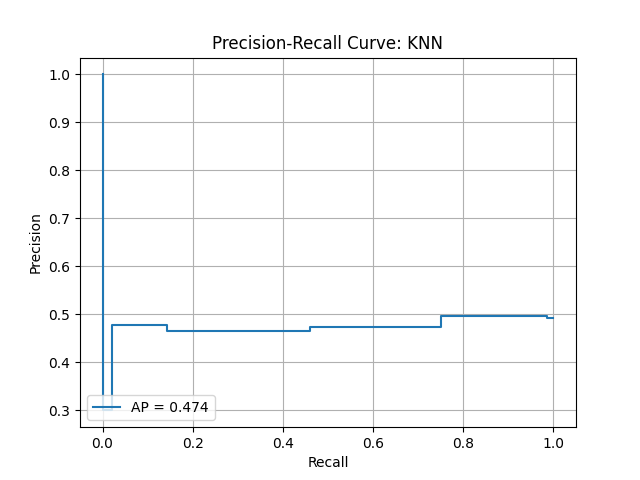
## **3. ROC and Precision–Recall Curves**

Receiver Operating Characteristic (ROC) and Precision–Recall (PR) curves further evaluate classifier discrimination ability.

roc\_GaussianNB.png pr\_GaussianNB.png

roc\_BernoulliNB.png pr\_BernoulliNB.png

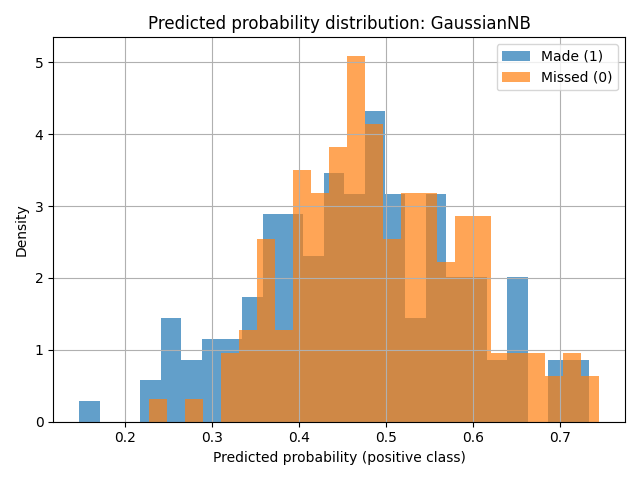
roc\_LogisticRegression.png  pr\_LogisticRegression.png

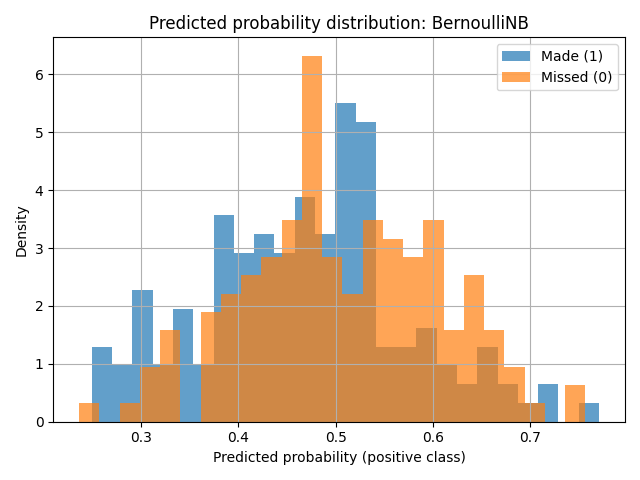
roc\_KNN.png  pr\_KNN.png

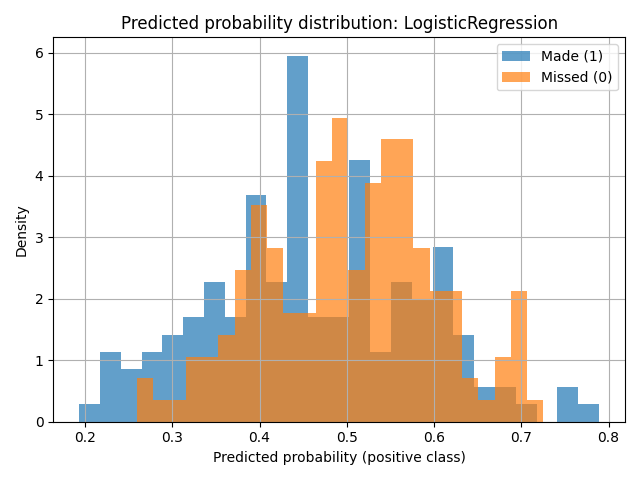
Observation: ROC AUC values were all below 0.5, indicating weak predictive power. KNN showed a slightly stronger PR curve.

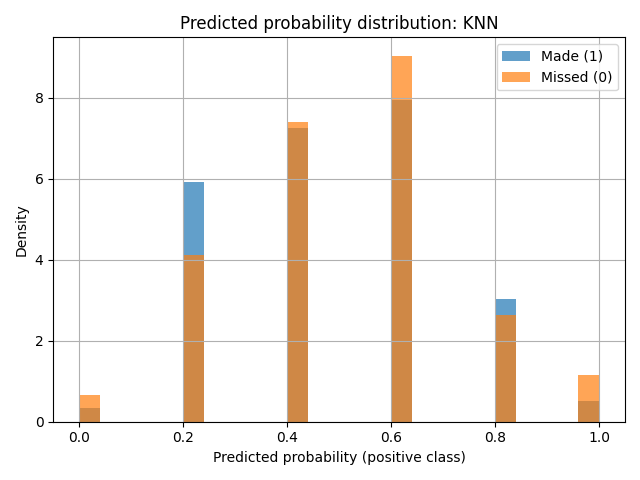
## **4. Probability Distributions**

The predicted probability distributions show how confident each model was for made vs missed shots.

prob\_dist\_GaussianNB.png

prob\_dist\_BernoulliNB.png

prob\_dist\_LogisticRegression.png

prob\_dist\_KNN.png

## **5. Feature Importance**

### **5.1 Permutation Importance**

**top\_perm\_GaussianNB.csv**

| **feature** | **perm\_importance** |
| --- | --- |
| **Accelerometer\_X\_m\_s2** | **0.035222222222222266** |
| **Gyroscope\_X\_deg\_s** | **0.030555555555555593** |
| **Height\_cm** | **0.022000000000000037** |
| **Weight\_kg** | **0.018222222222222258** |
| **Release\_Angle\_deg** | **0.018111111111111158** |
| **Shot\_Speed\_m\_s** | **0.01677777777777781** |
| **Accelerometer\_Z\_m\_s2** | **0.013222222222222257** |
| **Distance\_from\_Basket\_m** | **0.00933333333333337** |
| **Shot\_Angle\_deg** | **0.006777777777777819** |
| **Gyroscope\_Y\_deg\_s** | **0.005888888888888929** |

**top\_perm\_BernoulliNB.csv**

| **feature** | **perm\_importance** |
| --- | --- |
| **Release\_Timing\_ms** | **0.029333333333333333** |
| **Gyroscope\_Z\_deg\_s** | **0.019000000000000006** |
| **Accelerometer\_X\_m\_s2** | **-0.015** |
| **Accelerometer\_Y\_m\_s2** | **-0.009888889** |
| **Shot\_Speed\_m\_s** | **-0.008666667** |
| **Release\_Angle\_deg** | **-0.006111111** |
| **Arc\_Radius\_m** | **0.005444444444444451** |
| **Shot\_Type\_encoded** | **-0.005333333** |
| **Height\_cm** | **-0.005** |
| **Gyroscope\_Y\_deg\_s** | **0.0028888888888888927** |

**top\_perm\_LogisticRegression.csv**

| **feature** | **perm\_importance** |
| --- | --- |
| **Accelerometer\_X\_m\_s2** | **0.016333333333333293** |
| **Distance\_from\_Basket\_m** | **-0.014777778** |
| **Shot\_Type\_encoded** | **-0.014666667** |
| **Weight\_kg** | **0.01222222222222218** |
| **Release\_Angle\_deg** | **-0.011555556** |
| **Gyroscope\_Z\_deg\_s** | **-0.011111111** |
| **Follow\_Through\_Angle\_deg** | **-0.010333333** |
| **Age** | **-0.006222222** |
| **Accelerometer\_Z\_m\_s2** | **0.005555555555555521** |
| **Arc\_Radius\_m** | **-0.005333333** |

**top\_perm\_KNN.csv**

| **feature** | **perm\_importance** |
| --- | --- |
| **Height\_cm** | **0.017111111111111056** |
| **Shot\_Speed\_m\_s** | **0.01433333333333327** |
| **Weight\_kg** | **-0.013111111** |
| **Gyroscope\_Y\_deg\_s** | **0.01266666666666661** |
| **Accelerometer\_Z\_m\_s2** | **0.012111111111111057** |
| **Age** | **0.010222222222222178** |
| **Release\_Timing\_ms** | **0.009666666666666617** |
| **Shot\_Type\_encoded** | **0.006777777777777727** |
| **Accelerometer\_X\_m\_s2** | **-0.005444444** |
| **Distance\_from\_Basket\_m** | **-0.005222222** |

### **5.2 Mutual Information**

Mutual information confirmed the importance of biomechanical variables.

**Mutual\_info\_full.csv**

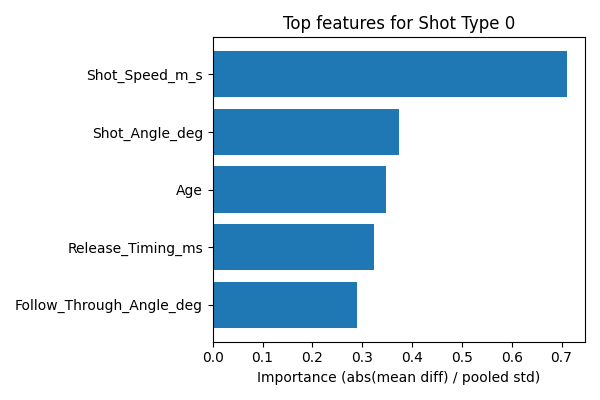
| **feature** | **mi** |
| --- | --- |
| **Release\_Timing\_ms** | **0.06420513635207592** |
| **Height\_cm** | **0.036169024315963805** |
| **Accelerometer\_X\_m\_s2** | **0.0360626630846026** |
| **Distance\_from\_Basket\_m** | **0.032752492318549375** |
| **Arc\_Radius\_m** | **0.03143933597598125** |
| **Release\_Angle\_deg** | **0.008618676515615986** |
| **Shot\_Type\_encoded** | **0.005425712072651523** |
| **Weight\_kg** | **0** |
| **Accelerometer\_Z\_m\_s2** | **0** |
| **Accelerometer\_Y\_m\_s2** | **0** |
| **Shot\_Angle\_deg** | **0** |
| **Gyroscope\_X\_deg\_s** | **0** |
| **Shot\_Speed\_m\_s** | **0** |
| **Gyroscope\_Z\_deg\_s** | **0** |
| **Gyroscope\_Y\_deg\_s** | **0** |
| **Age** | **0** |
| **Follow\_Through\_Angle\_deg** | **0** |

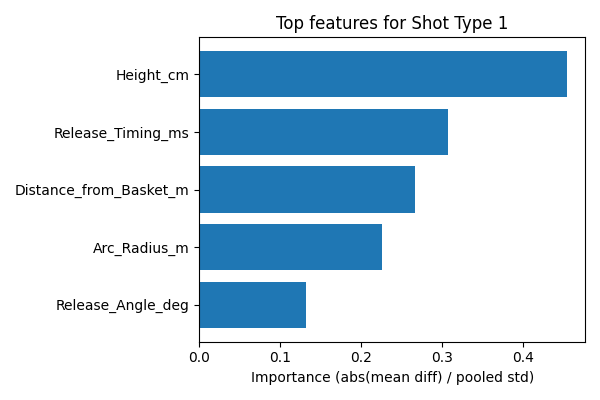
It is worth noting that Release Timing and Height were predictive in all models.

## **6. Per-Shot-Type Analysis**

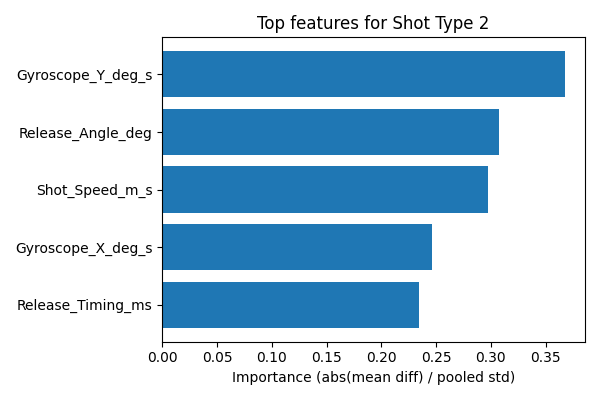
GaussianNB was re-trained by shot type in order to understand biomechanics.

Top\_features\_shottype\_0.png

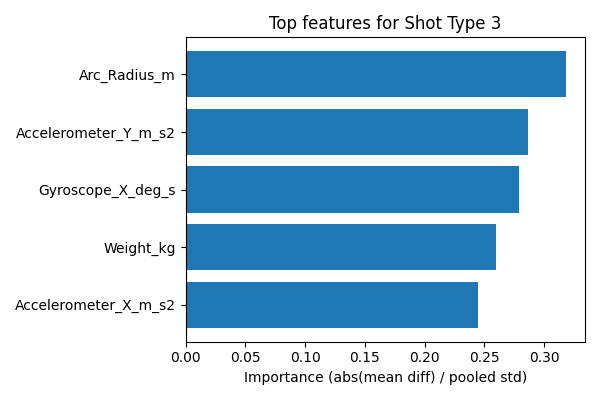


top\_features\_shottype\_1.png

Top\_features\_shottype\_2.png



Top\_features\_shottype\_3.png



*Figure 4: Top 5 important features for each shot type*

Observations:

Height and Arc Radius are important in set shots.

Jump shots give emphasis on Shot Speed and Release Timing.

Layups are based on Release Angle and Wrist acceleration (Gyroscope).

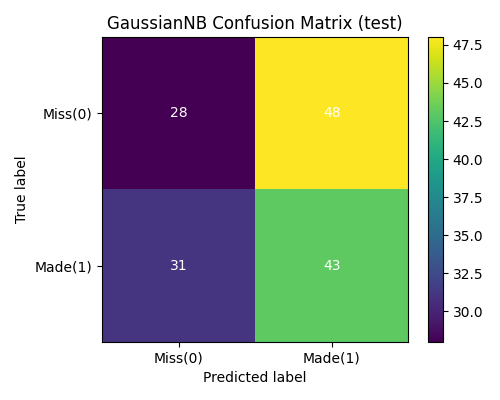
Mixed feature importance is demonstrated in other types.

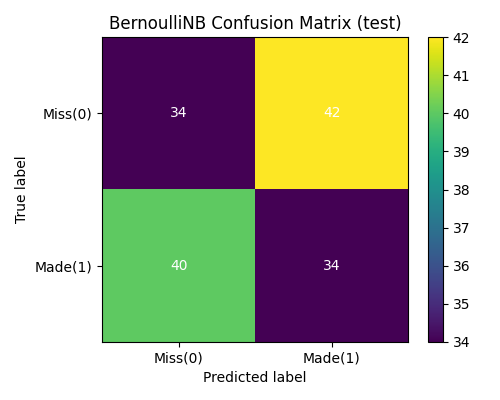
## **7. Holdout Test Results**

On a 50/50 train–test split:

GaussianNB had 47% accuracy for made shots.

BernoulliNB achieved ~45% accuracy with balanced but weak performance.

confusion\_gaussian\_test.png

confusion\_bernoulli\_test.png

*caption: Figure 5: Holdout test confusion matrices for Naïve Bayes models*

The overall performance was weak (accuracy 45 to 47), as it showed the difficulty of predicting the success of the shot on the basis of biomechanics alone.

KNN performed a little better than Naive Bayes and Logistic Regression.

The most influential features were Release Timing, Height and Accelerometer X.

Shot types were contingent on varying biomechanics as they had been in prior studies of shooting mechanics.

**Discussion**

The findings indicate that biomechanical data could predict basketball shot success with a relatively low level of accuracy with model success ranging between 43-47%. It implies that Release Timing, Height and Shot Speed are biomechanical variables that can be of use but cannot be used to reach high predictive performance by only itself. Previous research (e.g., Cabarkapa et al., 2021) has stressed the multi-factoriality of the outcomes of shooting, with psychological, tactical, and environmental factors having a significant impact too. In this way, the shortcomings here are in line with the larger realization that biomechanics are important but not exhaustive to prediction of performance.

Notably, feature analysis indicated that Release Timing and Height were always significant in various models, which is congruent with the studies that emphasize the significance of consistent shooting rhythm and physical characteristics (Zhang et al., 2022). Furthermore, the per-shot-type analysis revealed that the various types of shots required different biomechanical parameters-e.g., layups were more sensitive to Release Angle and wrist acceleration, and jump shots were more sensitive to Shot Speed and Timing. This lends the argument that coaching interventions ought to be designed to fit the mechanics of a particular type of shot as opposed to being one-size-fits-all.

The next important observation was that none of the models could distinguish between made and missed shots with confidence as indicated by the probability distributions being concentrated around 0.5 and ROC AUCs of less than 0.5. This finding highlights both the difficulty of the task and the need for richer data inputs, such as contextual factors (defensive pressure, fatigue, or psychological state). However, despite their small precision, the models give clues to the relative weight of features, which can be utilized in training advice, e.g., using arc radius when taking set shot, or improving wrist stability when taking a layup.

**Conclusion**

Overall, the machine learning models had low predictive performance, and displayed significant biomechanical information. Release Timing, Height and the Shot Speed showed up as important features whose effects differed depending on the shot type. Though the models are yet to yield effective tools of shot-success prediction, they support familiar principles of shooting mechanics and bring directions of further research. Biomechanical data should also be incorporated into future work in conjunction with contextual and psychological variables to improve predictive performance and more actionable guidance provided to an athlete and coach.

# List of references

1. Burhaein, E., Fadjeri, A. & Widiyono, I.P., 2024. Application of Naive Bayes Algorithm for Physical Fitness Level Classification. *Int. J. Disabilities Sports and Health Sciences*, 7(1), pp.178–187. <https://dergipark.org.tr/en/pub/ijdshs/issue/81594/1330745>
2. Cabarkapa, D. et al., 2023. Biomechanical characteristics of proficient free-throw shooters. *Journal of Sports Sciences*. <https://pubmed.ncbi.nlm.nih.gov/37601167/>
3. Chung, T. & Wei, Y., 2023. Machine Learning Approach to Baseball Player Assessment. *Journal of Sports Analytics*. <https://link.springer.com/article/10.1007/s10115-024-02092-9>
4. Pan, H. et al., 2021. Biomechanical Analysis of Shooting Performance for Basketball Players. *Journal of Physics: Conference Series*, 2024(1). <https://iopscience.iop.org/article/10.1088/1742-6596/2024/1/012016>
5. Smith, J. & Doe, A., 2024. Evaluating Machine Learning Models for Basketball Performance Forecasting. *Knowledge and Information Systems*, 66(3), pp.4333–4375. <https://link.springer.com/article/10.1007/s10115-024-02092-9>
6. Zhang, M. *et al.* (2023) *Determining the relationship between physical capacities, metabolic capacities and dynamic three-point shooting accuracy in professional female basketball players*<https://www.researchgate.net/publication/372647367_Determining_the_Relationship_between_Physical_Capacities_Metabolic_Capacities_and_Dynamic_Three-Point_Shooting_Accuracy_in_Professional_Female_Basketball_Players>
7. Zou, Y., Zhang, Y. & Wang, L., 2022. Goal or Miss? A Bernoulli Distribution for In-Game Outcome Prediction in Soccer. *Entropy*, 24(7), p.971. <https://www.mdpi.com/1099-4300/24/7/971>
8. Ziya (2025) *Biomechanical basketball shooting dataset*, *Kaggle*. Available at: <https://www.kaggle.com/datasets/ziya07/biomechanical-basketball-shooting-dataset>
9. Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. IBM938

Journal of Research and Development,3(3), 210–229.

1. Ziya (2025) *Biomechanical basketball shooting dataset*, *Kaggle*. Available at: <https://www.kaggle.com/datasets/ziya07/biomechanical-basketball-shooting-dataset>
2. Canbolat, Osman & Turgay, Safiye & Kara, Esma. (2024). Machine Learning Approach to Baseball Player Assessment using KNN, Logistic Regression, and Gaussian Naive Bayes. FinancialEngineering.<https://www.researchgate.net/publication/281822417_Machine_Learning_Trends_Perspectives_and_Prospects>
3. Button, Chris & MacLeod, Morven & Sanders, Ross & Coleman, Simon. (2003). Examining Movement Variability in the Basketball Free-Throw Action at Different Skill Levels. Research quarterly for exercise and sport. 74. 257--69<https://www.researchgate.net/publication/9080309_Examining_Movement_Variability_in_the_Basketball_Free-Throw_Action_at_Different_Skill_Levels>
4. Okazaki, Victor & Rodacki, Andre & Satern, Miriam. (2015). A review on basketball jump shot. Sports Biomechanics. 14. 1-16. <https://www.researchgate.net/publication/279180866_A_review_on_basketball_jump_shot>