LSU Softball: Called Strike Probability Model

Addeson Zappa, Ashlynn Simmons, Dina Taing, Madden Gleason, and Madison Ziegler

Fall 2025 Final Presentation

Outline

Introduction

Strike Zone Model

- Softball umpires do not always call the NCAA rulebook strike zone

Strike Zone Model

- Softball umpires do not always call the NCAA rulebook strike zone
  - Umpires tend to apply their own individualized strike zones, which shift based on pitch location, count, player handedness, and game situation

Introduction

Strike Zone Model

- Softball umpires do not always call the NCAA rulebook strike zone
    - Umpires tend to apply their own individualized strike zones, which shift based on pitch location, count, player handedness, and game situation
- We built a **called strike probability model** for LSU Softball that predicts how likely a taken pitch (no swing) is to be called a strike

Introduction

Strike Zone Model

- Softball umpires do not always call the NCAA rulebook strike zone
  - Umpires tend to apply their own individualized strike zones, which shift based on pitch location, count, player handedness, and game situation
- We built a **called strike probability model** for LSU Softball that predicts how likely a taken pitch (no swing) is to be called a strike
- The model was originally trained on MLB Statcast data (2020–2024) and later adapted to LSU TrackMan data (2025)

Strike Zone Model

- Applies machine learning to analyze pitch type, count, and player handedness to help us understand how these factors influence the zone, pitch decisions, and swing decisions

Strike Zone Model

- Applies machine learning to analyze pitch type, count, and player handedness to help us understand how these factors influence the zone, pitch decisions, and swing decisions
- Will help improve trainings, player evaluations, scouting, and in-game decision-making for LSU Softball

Objectives

- Transfer a model trained on MLB Statcast data to LSU Softball TrackMan data

Goals

Objectives

- Transfer a model trained on MLB Statcast data to LSU Softball TrackMan data
- Predict the probability a taken pitch is called a strike using pitch location and handedness

Goals

Objectives

- Transfer a model trained on MLB Statcast data to LSU Softball TrackMan data
- Predict the probability a taken pitch is called a strike using pitch location and handedness
- Visualize LSU-specific strike zone patterns using probability heatmaps

Goals

Objectives

- Transfer a model trained on MLB Statcast data to LSU Softball TrackMan data
- Predict the probability a taken pitch is called a strike using pitch location and handedness
- Visualize LSU-specific strike zone patterns using probability heatmaps
- Use results to help players understand the zone better and make better decisions in games

Data & Preprocessing

Data & Preprocessing

MLB Statcast Data (2020–2024)

- Used initially because LSU TrackMan data was not yet available

Data & Preprocessing

MLB Statcast Data (2020–2024)

- Used initially because LSU TrackMan data was not yet available
- Includes pitch location at the plate, pitch type, velocity, count, handedness, and pitch result

Data & Preprocessing

MLB Statcast Data (2020–2024)

- Used initially because LSU TrackMan data was not yet available
- Includes pitch location at the plate, pitch type, velocity, count, handedness, and pitch result
- Filtered to taken pitches and defined $Y = 1$ for called strikes and $Y = 0$ otherwise

Data & Preprocessing

MLB Statcast Data (2020–2024)

- Used initially because LSU TrackMan data was not yet available
- Includes pitch location at the plate, pitch type, velocity, count, handedness, and pitch result
- Filtered to taken pitches and defined $Y = 1$ for called strikes and $Y = 0$ otherwise

LSU TrackMan Data (2025)

- Arrived late in the semester and is significantly smaller than the MLB dataset

Data & Preprocessing

MLB Statcast Data (2020–2024)

- Used initially because LSU TrackMan data was not yet available
- Includes pitch location at the plate, pitch type, velocity, count, handedness, and pitch result
- Filtered to taken pitches and defined $Y = 1$ for called strikes and $Y = 0$ otherwise

LSU TrackMan Data (2025)

- Arrived late in the semester and is significantly smaller than the MLB dataset
- Contains PlateLocSide, PlateLocHeight, pitch type, count, pitcher/batter handedness, and swing indicator

Data & Preprocessing

MLB Statcast Data (2020–2024)

- Used initially because LSU TrackMan data was not yet available
- Includes pitch location at the plate, pitch type, velocity, count, handedness, and pitch result
- Filtered to taken pitches and defined $Y = 1$ for called strikes and $Y = 0$ otherwise

LSU TrackMan Data (2025)

- Arrived late in the semester and is significantly smaller than the MLB dataset
- Contains PlateLocSide, PlateLocHeight, pitch type, count, pitcher/batter handedness, and swing indicator
- Cleaned by removing rows with missing locations and keeping variables relevant to the model

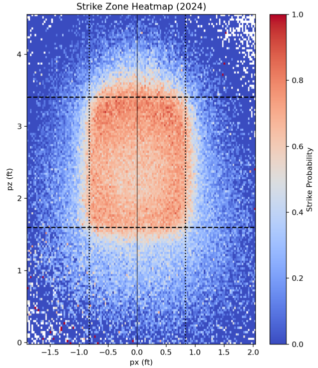
Exploratory Visualizations

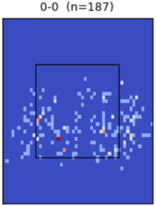
MLB Statcast Data (2020–2024)

- We first studied MLB pitch locations to understand typical strike-zone patterns before adapting anything to softball

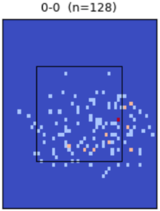
MLB Statcast Data (2020–2024)

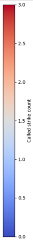
- We first studied MLB pitch locations to understand typical strike-zone patterns before adapting anything to softball
- Heatmaps show the distribution of taken pitches and where umpires most often call strikes



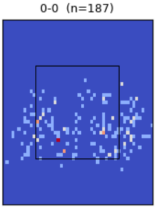


0–0 Count, LHP Curveball

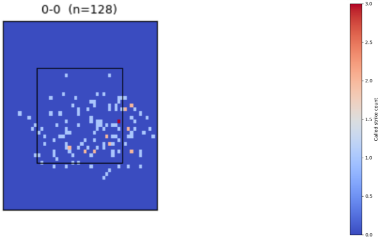




0–0 Count, RHP Curveball



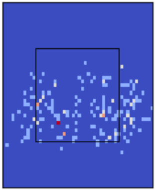
0–0 Count, LHP Curveball



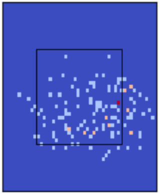
0–0 Count, RHP Curveball

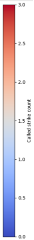
LSU TrackMan Data (2025)

- LSU heatmaps help reveal raw pitch tendencies before modeling. For example, LHP curveballs show wider horizontal variability, and both groups tend to finish low in the zone



0–0 Count, LHP Curveball





0–0 Count, RHP Curveball

LSU TrackMan Data (2025)

- LSU heatmaps help reveal raw pitch tendencies before modeling. For example, LHP curveballs show wider horizontal variability, and both groups tend to finish low in the zone
- Patterns differ from MLB due to softball-specific pitch types, release points, and strike-zone geometry, as well as the much smaller size of the LSU dataset

Model Architecture & Methods

Model Architecture

Neural Network Structure

- Multilayer Perceptron (MLP) that outputs the probability a taken pitch is called a strike

Model Architecture

Neural Network Structure

- Multilayer Perceptron (MLP) that outputs the probability a taken pitch is called a strike
- **Inputs:** horizontal and vertical location, swing indicator, pitcher handedness, batter handedness

Model Architecture

Neural Network Structure

- Multilayer Perceptron (MLP) that outputs the probability a taken pitch is called a strike
- **Inputs:** horizontal and vertical location, swing indicator, pitcher handedness, batter handedness
- **Hidden layers:** 32 ReLU neurons, then 64 Sigmoid neurons

Model Architecture

Neural Network Structure

- Multilayer Perceptron (MLP) that outputs the probability a taken pitch is called a strike
- **Inputs:** horizontal and vertical location, swing indicator, pitcher handedness, batter handedness
- **Hidden layers:** 32 ReLU neurons, then 64 Sigmoid neurons
- **Output layer:** 1 neuron returning the called–strike probability

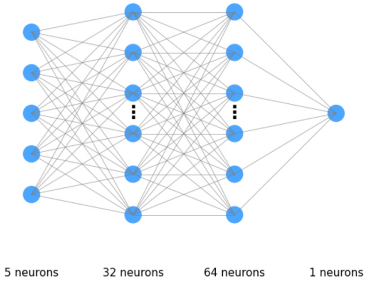
Neural Network Structure

- Multilayer Perceptron (MLP) that outputs the probability a taken pitch is called a strike
- **Inputs:** horizontal and vertical location, swing indicator, pitcher handedness, batter handedness
- **Hidden layers:** 32 ReLU neurons, then 64 Sigmoid neurons
- **Output layer:** 1 neuron returning the called–strike probability

$$W_1 = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,5} \\ \vdots & \vdots & \ddots & \vdots \\ w_{32,1} & w_{32,2} & \cdots & w_{32,5} \end{bmatrix}$$

$$W_2 = \begin{bmatrix} w_{1,1} & \cdots & w_{1,32} \\ \vdots & \ddots & \vdots \\ w_{64,1} & \cdots & w_{64,32} \end{bmatrix}$$

$$W_3 = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,64} \end{bmatrix}$$



Activation Functions

- ReLU introduces nonlinearity and allows flexible boundary shapes

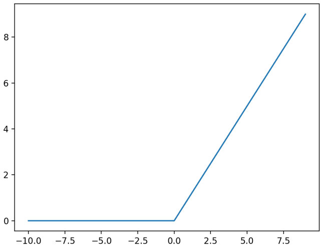
Activation Functions

- ReLU introduces nonlinearity and allows flexible boundary shapes
- Sigmoid compresses outputs into probabilities between 0 and 1

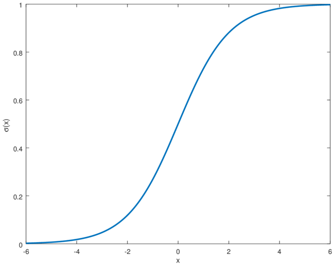
Activation Functions

- ReLU introduces nonlinearity and allows flexible boundary shapes
- Sigmoid compresses outputs into probabilities between 0 and 1

ReLU



Sigmoid



LSU Softball: Dataset 1 to Dataset 2

- We first fitted the model on a smaller LSU softball set to learn a stable, location-based strike zone

Training & Transfer Learning

LSU Softball: Dataset 1 to Dataset 2

- We first fitted the model on a smaller LSU softball set to learn a stable, location-based strike zone
- Once we received the second, larger dataset, we wanted to preserve the model's progress from the first dataset while adding new features from the larger one.

Training & Transfer Learning

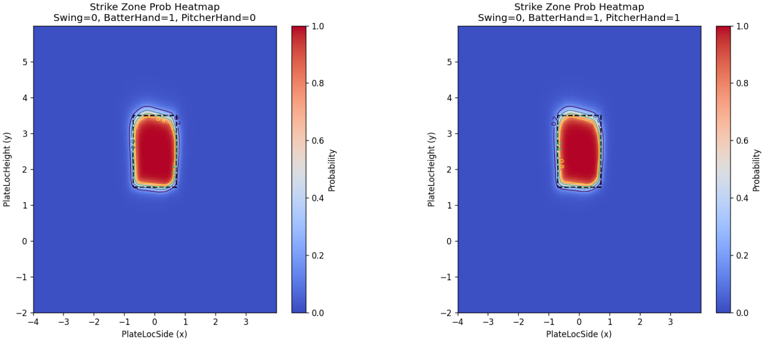
LSU Softball: Dataset 1 to Dataset 2

- We first fitted the model on a smaller LSU softball set to learn a stable, location-based strike zone
- Once we received the second, larger dataset, we wanted to preserve the model's progress from the first dataset while adding new features from the larger one.
- We used neural network weight surgery to expand the input layer:
    - Copied the original weight matrix into a larger matrix.
    - Added new columns for swing indicator, pitcher hand, and batter hand
    - Initialized the new entries so the models original behavior was preserved

Training & Transfer Learning

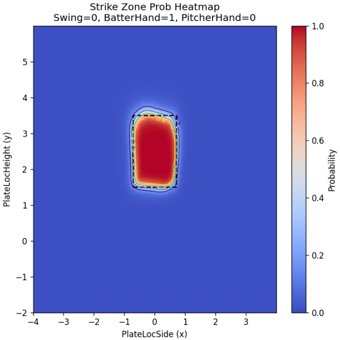
LSU Softball: Dataset 1 to Dataset 2

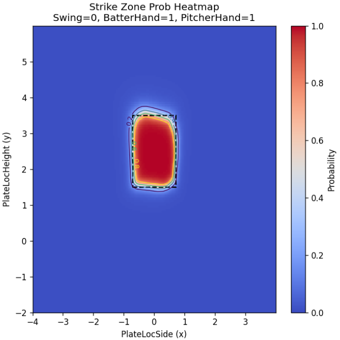
- We first fitted the model on a smaller LSU softball set to learn a stable, location-based strike zone
- Once we received the second, larger dataset, we wanted to preserve the model's progress from the first dataset while adding new features from the larger one.
- We used neural network weight surgery to expand the input layer:
  - Copied the original weight matrix into a larger matrix.
  - Added new columns for swing indicator, pitcher hand, and batter hand
  - Initialized the new entries so the models original behavior was preserved
- After surgery, we fine-tuned the expanded model on LSU Softball data to adapt the strike zone to NCAA softball



- The model captures the "true" zone LSU pitchers are actually getting in games, not just the rulebook version.

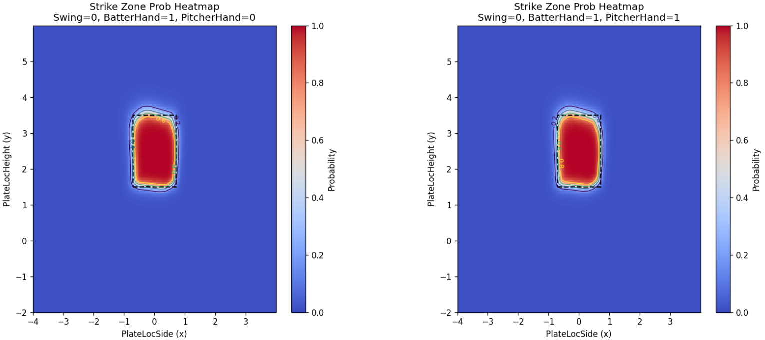
Model Results & Observations





- The model captures the “true” zone LSU pitchers are actually getting in games, not just the rulebook version.
- Called strike probability is highest in the zone center

Model Results & Observations



- The model captures the “true” zone LSU pitchers are actually getting in games, not just the rulebook version.
- Called strike probability is highest in the zone center
- We received 97.5% accuracy

Limitations & Challenges

Challenges Encountered

- LSU Softball data was received late in the semester

Limitations & Challenges

Challenges Encountered

- LSU Softball data was received late in the semester
- The LSU Softball dataset was significantly smaller than the MLB dataset

Limitations & Challenges

Challenges Encountered

- LSU Softball data was received late in the semester
- The LSU Softball dataset was significantly smaller than the MLB dataset
  - This resulted in challenges adapting and fitting the model with the same level of reliability as the MLB-trained version

Challenges Encountered

- LSU Softball data was received late in the semester
- The LSU Softball dataset was significantly smaller than the MLB dataset
  - This resulted in challenges adapting and fitting the model with the same level of reliability as the MLB-trained version
- Variations in strike-zone definitions, pitch characteristics, and umpire decision-making prevented MLB patterns from transferring perfectly

Conclusion & Future Work

Conclusions

Strike Zone Model

- Successfully adapted a location-based strike probability model from MLB data to NCAA softball data

Conclusions

Strike Zone Model

- Successfully adapted a location-based strike probability model from MLB data to NCAA softball data
- Highlights consistent differences in pitch behavior across pitch types, pitch counts, and pitcher handedness

Conclusions

Strike Zone Model

- Successfully adapted a location-based strike probability model from MLB data to NCAA softball data
- Highlights consistent differences in pitch behavior across pitch types, pitch counts, and pitcher handedness
- Current dataset is relatively small, the patterns are strong and interpretable

Conclusions

Strike Zone Model

- Successfully adapted a location-based strike probability model from MLB data to NCAA softball data
- Highlights consistent differences in pitch behavior across pitch types, pitch counts, and pitcher handedness
- Current dataset is relatively small, the patterns are strong and interpretable
- The model offers a practical, data-driven foundation for future training, player development, and decision-making applications

Future Work

Future Goals

- Extend the model to include catcher framing and swing decisions

Future Work

Future Goals

- Extend the model to include catcher framing and swing decisions
    - This will help quantify how catchers influence umpire decisions and the zone

Future Work

Future Goals

- Extend the model to include catcher framing and swing decisions
  - This will help quantify how catchers influence umpire decisions and the zone
- Further adapt the model to include data from the upcoming 2026 season and continue to update it as more seasons are completed

Future Work

Future Goals

- Extend the model to include catcher framing and swing decisions
  - This will help quantify how catchers influence umpire decisions and the zone
- Further adapt the model to include data from the upcoming 2026 season and continue to update it as more seasons are completed
  - This will improve the accuracy and reliability of the model.

Future Work

Future Goals

- Extend the model to include catcher framing and swing decisions
  - This will help quantify how catchers influence umpire decisions and the zone
- Further adapt the model to include data from the upcoming 2026 season and continue to update it as more seasons are completed
  - This will improve the accuracy and reliability of the model.
  - Produce more precise results for analysis and player-specific insights

Future Work

Future Goals

- Extend the model to include catcher framing and swing decisions
  - This will help quantify how catchers influence umpire decisions and the zone
- Further adapt the model to include data from the upcoming 2026 season and continue to update it as more seasons are completed
  - This will improve the accuracy and reliability of the model.
  - Produce more precise results for analysis and player-specific insights

References

References

References

1. Major League Baseball Advanced Media. *Baseball Savant: Statcast Search and Player Visualizations*. Online, 2025. Retrieved from https://baseballsavant.mlb.com. Accessed: Oct. 21, 2025.
2. Franke, K. *Strike Probability Model and Catcher Framing Using Random Forest*. Medium, 2021. Accessed: Oct. 21, 2025.
3. TrackMan Baseball. *Radar Measurement Glossary of Terms (V3)*. Online, 2025. Retrieved from https://support.trackmanbaseball.com. Accessed: Oct. 21, 2025.
4. Nestico, T. *Classifying MLB Pitch Zones and Predicting MiLB Zones*. Medium, 2021. Retrieved from https://medium.com/@thomasjamesnestico/classifying-mlb-pitch-zones-and-predicting-milb-zones-7e95cf308254. Accessed: Dec. 1, 2025.
5. NCAA. *2024–2025 Softball Rules Book*. National Collegiate Athletic Association, Indianapolis, IN, 2024. Accessed: Nov. 11, 2025.