

Analytics in the Octagon: Modeling UFC Fight Prediction

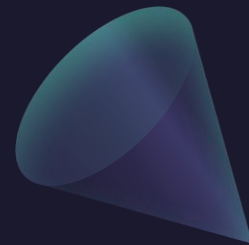
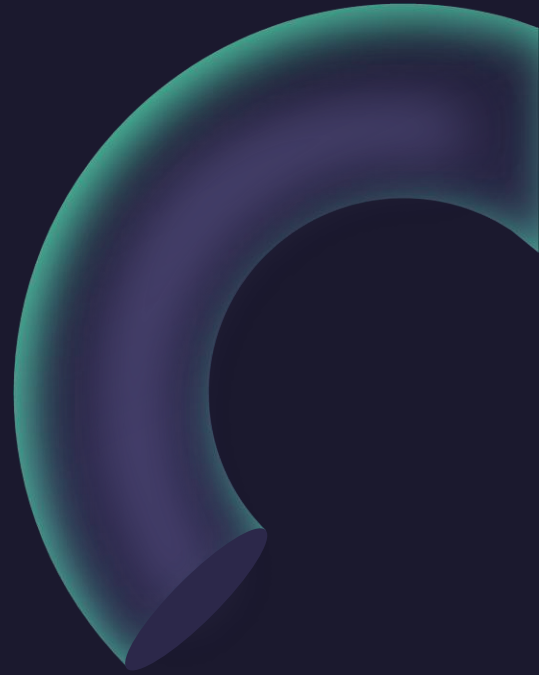
Group 4 -
Evan Embry

Description

My aim is to build a predictive model that determines which fighter (Red or Blue corner) will win an upcoming UFC match based on historical fight data. My project will investigate the impact of fighter attributes (e.g., height, reach, stance, etc.), fight context (title bout vs. non-title), and performance metrics (strikes, takedowns, etc.) on the probability of victory. By exploring data patterns and testing various machine learning models, I hope to gain insights into the key factors that drive success in the Octagon.

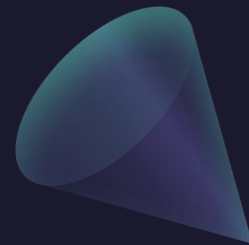
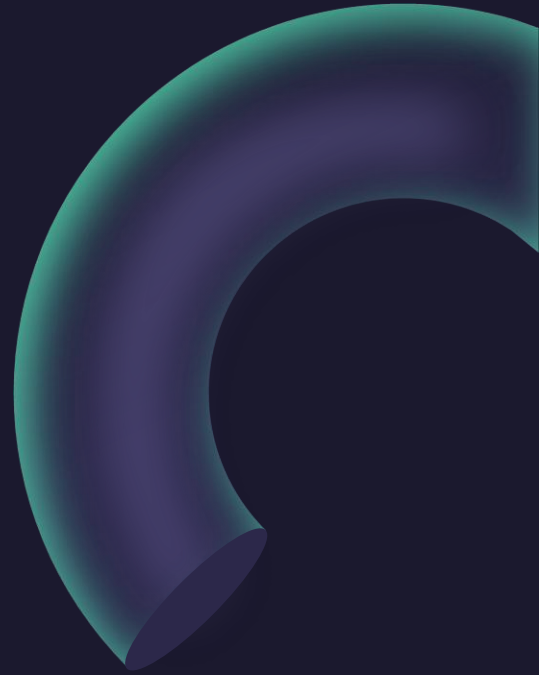
4. Prior Work

- **Existing UFC or MMA Predictions:** Sports analysts and betting markets already use fighter metrics and expert opinion to set odds.
- **Related Sports Analytics:** Approaches from boxing and other team sports have also demonstrated how advanced statistics can improve outcome predictions.



5. Datasets

- **List of Datasets:**
 - I am primarily using the “Ultimate UFC Dataset” from Kaggle.
 - **Where Found:**
 - <https://www.kaggle.com/datasets/mdabbert/ultimate-ufc-dataset>.
- **Download:**
 - The CSV file is downloaded and stored locally on my machine.



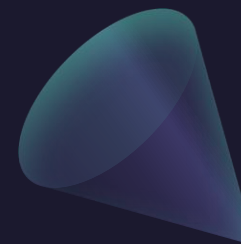
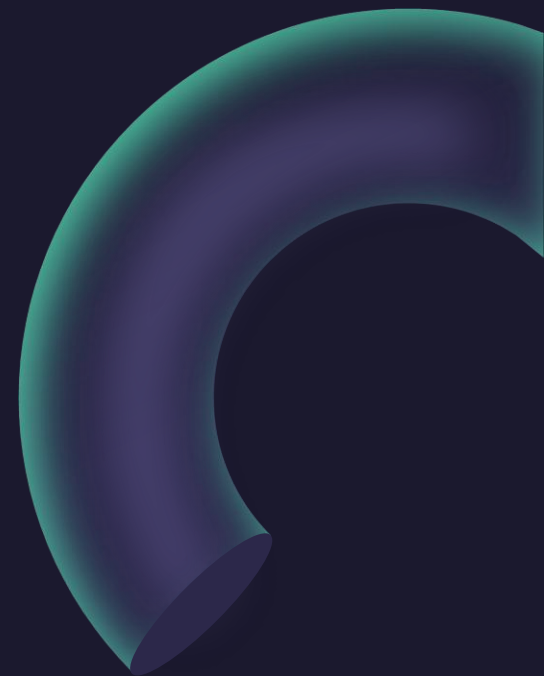
6. Proposed Work

1. Data Cleaning

- Remove duplicate entries, handle missing or invalid fields (e.g., missing reach or stance).
- Convert columns to appropriate data types (e.g., numeric, categorical).

2. Data Preprocessing

- Derive new features such as reach difference, height difference, etc.
- Normalize or standardize numeric columns if needed.



Proposed work cont.

1. Data Integration

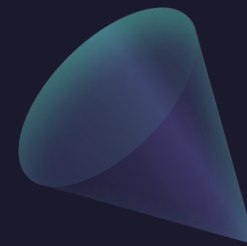
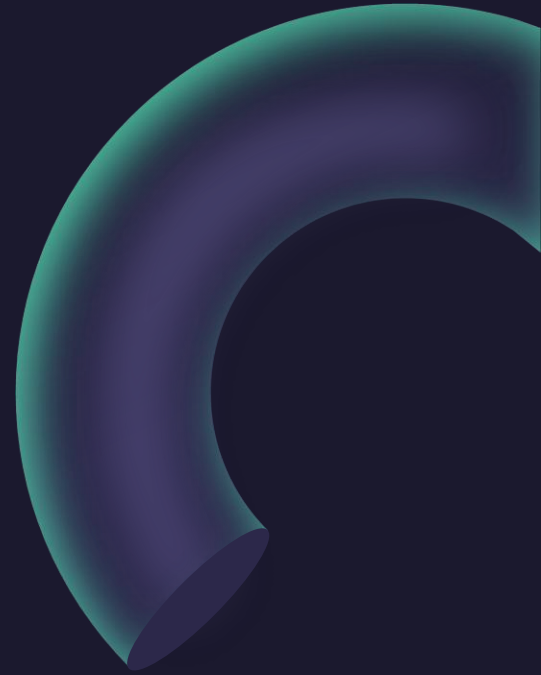
- Combine fighter stats with contextual info (e.g., date, location, odds).
- Possibly incorporate external data (e.g., historical fighter records) if available.

2. Modeling

- Train at least one classification model (e.g., logistic regression, random forest, or XGBoost).
- Tune hyperparameters using validation methods.

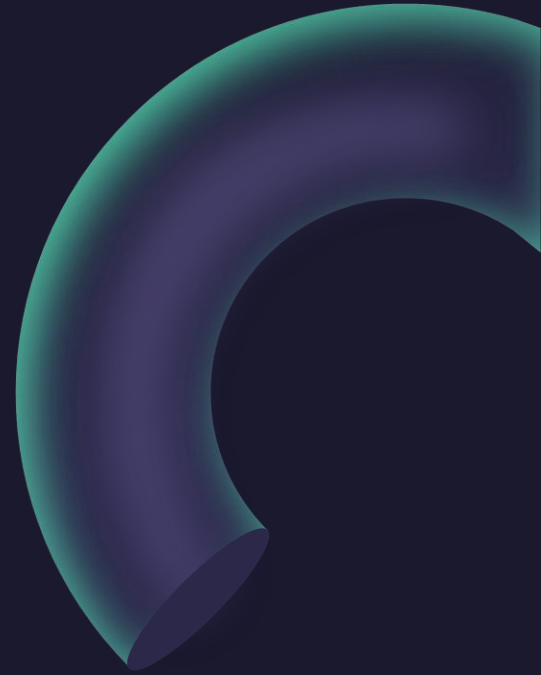
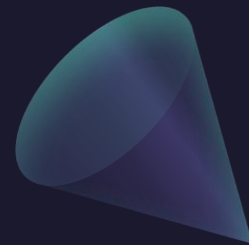
3. Analysis

- Inspect feature importance.
- Compare baseline (simple odds-based prediction) vs. advanced models.



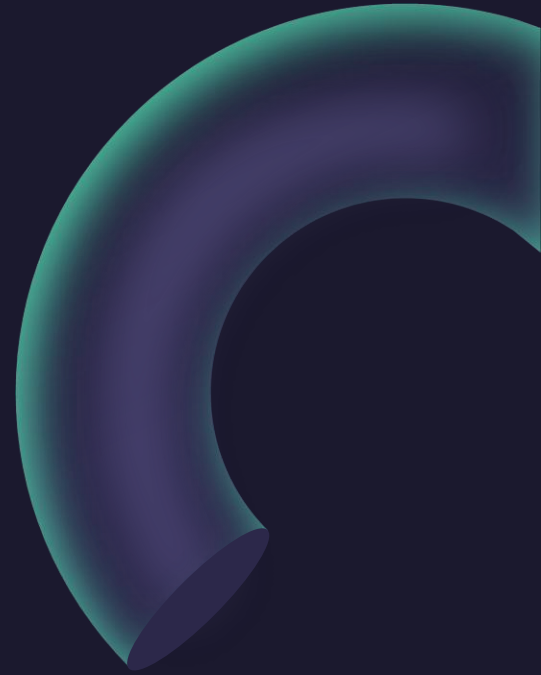
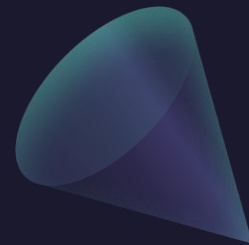
7. List of Tools

- **7. List of Tools**
- **Python** (for data wrangling and analysis).
- **Pandas** (data cleaning, EDA).
- **scikit-learn / XGBoost** (machine learning).
- **Jupyter Notebook** (exploratory analysis, documentation).
- **GitHub** (version control and collaboration).



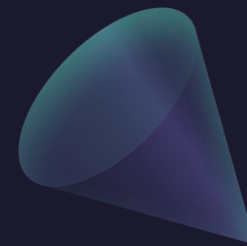
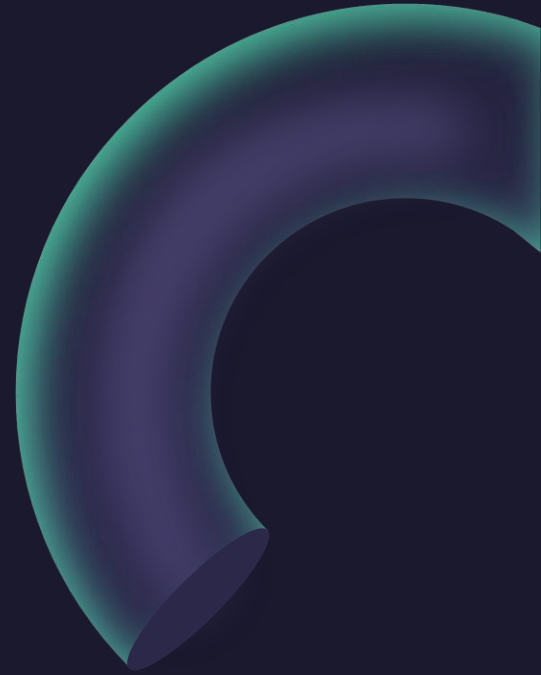
7. List of Tools

- **7. List of Tools**
- **Python** (for data wrangling and analysis).
- **Pandas** (data cleaning, EDA).
- **scikit-learn / XGBoost** (machine learning).
- **Jupyter Notebook** (exploratory analysis, documentation).
- **GitHub** (version control and collaboration).



8. Evaluation

- **Model Accuracy:** Percentage of correct predictions on a held-out test set.
- **Precision/Recall / F1 Score:** If class imbalance is notable (e.g., one corner wins more often).
- **ROC Curve / AUC:** To measure model performance in a threshold-agnostic way.
- **Comparison to Betting Odds:** See if the model can match or beat the implied probabilities from bookmaker odds.



Thank you

