Tune-IN or Skip: Using Machine Learning To Decide On How To Book Exciting UFC Fights

Abraham Al-Shamsie

RedID: 130461484

San Diego State University

5500 Campanile Dr, San Diego, CA 92182, USA

CS549_Machine Learning with Dr. Xiao-Bai Liu

My Email: aalshamie5088@sdsu.edu

***Abstract*— This research proposes a novel approach to a common challenge in UFC viewership: identifying potentially exciting fights before they happen. While previous work has focused on predicting fight outcomes, we explore the relatively unexplored territory of predicting fight excitement using machine learning techniques. Our approach combines three distinct data sources: the UFC Complete Dataset (1996-2024), Tapology's Fight of the Year rankings, and detailed fighter statistics from UFC.com. Through iterative model development, we address unique challenges including the subjective nature of fight excitement and class imbalance in historical data. The Random Forest model achieved promising results in preliminary testing, with 99.6% precision and 100% recall across different fighting styles and weight classes. While these initial results are encouraging, further research is needed to validate the model's effectiveness across a broader range of scenarios and viewer preferences. This work represents a first step toward helping casual fans make more informed viewing decisions, potentially increasing engagement with the sport.**

***Index-Terms*— Machine Learning, UFC, Fight Prediction, Sports Analytics, Random Forest Classification, Fight Excitement Prediction**

I. Introduction

The world of mixed martial arts (MMA) has grown immensely in popularity in recent years, with UFC (Ultimate Fighting Championship) being the global leader in this high-stakes sport. However, the excitement of a fight is not always easily predicted. While some matches end in thrilling knockouts or submissions, others can be uneventful, leading to disappointment for fans. On a personal level, I often invite my siblings and friends to watch fights, only to find us sitting through hours of slow, uneventful grappling matches or split-decision striking bouts. With this project, I hope to bring my friends and family closer to enjoying the sport as much as I do by avoiding those underwhelming matchups.

The main idea behind this project is to apply machine learning techniques to analyze various attributes of UFC fighters, such as their physical characteristics, performance history, and fighting style, to predict whether a specific fight will be considered exciting. Excitement, in this case, is defined by factors such as the potential for knockouts, finishes, and the general level of competition between the fighters. This predictive model can be a valuable tool for fans, analysts, and UFC matchmakers, offering insights into upcoming fights and helping them decide which matchups are most likely to deliver an exhilarating spectacle.

The motivation stems from the inherently subjective nature of fight excitement, which traditional methods assess through qualitative opinions and anecdotal evidence that vary widely from person to person. By leveraging machine learning techniques, this project aims to discover whether there is an objective and data-driven approach using measurable data points.

To accomplish this, I collected data from three main sources: the UFC Complete Dataset (1996–2024) from Kaggle, scraped data from the MMA Topology website's annual "Fight of the Year" candidates, and my own dataset extracted from the UFC's official website focusing on detailed fighter statistics. These datasets contain information on fighter attributes (e.g., height, weight, reach, fighting style) and detailed fight statistics (e.g., strikes landed, knockouts, submission attempts).

In the subsequent sections, this report will explore the machine learning model's objective and algorithms used, discuss major challenges in data preprocessing and feature selection, and present solutions for managing imbalanced classes. The experiments section will describe the datasets used for training and testing, evaluation metrics employed, and analysis of the models' results.

Finally, the conclusion will summarize key findings and propose potential future directions to enhance the model's accuracy and real-world applicability. The overarching goal is to contribute to sports analytics by creating a data-driven approach to predicting exciting MMA fights, providing actionable insights for fans, analysts, and UFC matchmakers, while making the sport more enjoyable for casual viewers.

II. Task Description

The primary objective of this project is to predict whether a UFC fight will be exciting based on fighter attributes and fight-specific data. Excitement is defined through factors such as knockouts, submissions, technical finishes, or recognition as dynamic and competitive fights. The goal is to construct a machine learning model that analyzes structured data on fighters and predicts the likelihood of a fight being worth watching.

I depended on three datasets. The UFC Events Dataset (1996–2024), containing detailed event and fight-specific statistics; the Annual Best Fights Data from MMA Topology, highlighting exciting fights judged by MMA enthusiasts and experts; and the UFC Roster Dataset, which includes fighter-specific statistics such as height, weight, reach, and fight records. Preprocessing is handled by the preprocessing python script, which efficiently loads datasets while logging progress. Additional features, such as combined striking pace, win percentages, and fighter experience, are computed. A binary target variable, is_best_fight, is created based on multi-criteria definitions of excitement.

Finding the best features to use was very challenging. For this project the key features representing fight dynamics are engineered and selected. Fighter attributes such as height, weight, age, wins, losses, significant strikes landed per minute (SLpM), and takedown accuracy are paired with fight-specific features like title bout status, combined striking pace, and experience disparities. Derived features, including win percentage, performance differences, and activity levels between fighters, are also incorporated. These features are prepared for training using the select_features method, ensuring that relevant data is logged and ready for use.

For training, three machine learning models are implemented to evaluate their ability to predict exciting fights:

1) Random Forest Classifier, used to capture complex patterns and evaluate feature importance.
2) Support Vector Machine (SVM), employed for robust decision boundaries.
3) Linear Regression, used as a baseline for initial predictions and interpretability.

I ran into a lot of imbalances with the data I had. The imbalance in the dataset is because exciting fights constitute a minority. I applied the SMOTE method (Synthetic Minority Over-sampling Technique) for this. This ensures that the minority class is oversampled during training, leading to more balanced predictions.

Models are evaluated using accuracy, precision, recall, and F1-score, with a focus on correctly identifying exciting fights (true positives). Feature importance scores are analyzed to determine which factors most influence excitement predictions.

The final implementation is a script called predict_fight_excitement, where users can input two fighters from the same weight class to receive predictions about whether their matchup will be exciting. The model also provides explanations of its predictions, improving interpretability and user trust.

The inherent subjectivity of excitement posed a significant challenge, requiring a robust definition and careful feature engineering. This project addressed the issue by creating a multi-criteria target variable and implementing thorough evaluation techniques to ensure reliable and interpretable predictions.

III. Major challenges and solutions

This project faced several significant challenges that shaped its development and outcomes. Initially planned as a fight outcome prediction system, I pivoted to predicting fight excitement just four days before the deadline after discovering numerous similar outcome prediction projects. This late change significantly influenced the project's direction and results.

The primary challenge was defining "excitement" as a measurable target variable. MMA excitement is inherently subjective, influenced by personal preferences, fighting styles, and crowd reactions. To address this, I developed a multi-criteria approach combining objective metrics (win percentages, striking pace) with subjective data from MMA Tapology's best fights dataset. This hybrid approach helped create a more robust definition while maintaining measurable outcomes.

Data collection and integration posed substantial technical hurdles. The scraping process encountered various obstacles, particularly in matching fighter names across different datasets and dealing with limited access to preliminary event data. To overcome these challenges, I implemented a comprehensive data cleaning pipeline using median imputation for missing values and string normalization for name standardization. The significant class imbalance, where exciting fights constituted only 15% of the dataset, was addressed using SMOTE to generate synthetic examples for the minority class.

Feature engineering required extensive experimentation to capture fight dynamics effectively. Through exploratory data analysis, I identified key features correlated with excitement and created new variables to capture fighter disparities. Model training faced computational constraints, particularly with the Random Forest and SVM implementations, requiring parameter adjustments to balance efficiency with performance. The evaluation process focused on the F1-score to ensure balanced performance metrics, with cross-validation implemented to verify the model's generalizability.

Converting this analysis into a practical prediction tool proved more challenging than anticipated, particularly in developing an effective matchmaking component. Despite these challenges, the solutions implemented provided valuable insights into quantifying and predicting fight excitement, even if they couldn't capture all aspects of what makes MMA compelling.

IV. Experiments

A. Dataset Description

This research integrated three primary data sources to create a comprehensive dataset for predicting fight excitement. The first source was the UFC Complete Dataset (1996-2024), filtered to post-2009 events for consistency with modern UFC rules. This dataset provided essential fight information including fighter statistics (strikes landed, takedowns), physical attributes (height, weight, reach), and bout outcomes (KO/TKO, submission).

For ground truth on exciting fights, I developed a custom web scraper to collect data from Tapology.com's annual "Fight of the Year" rankings. This yielded approximately 6,000 historically significant matches, providing expert-curated examples of exciting fights. Additionally, I created another scraper to extract detailed fighter statistics from UFC.com, including striking accuracy, takedown accuracy, and submission averages, which offered deeper insights into fighting styles.

Data integration merged these sources using fighter names as the primary key, implementing fuzzy matching to handle name variations and achieving a 94% match rate. Missing values were handled through median imputation within weight class groupings. The final dataset contained 52 features per fight, balancing raw statistics with engineered features.

A significant challenge was the class imbalance, with exciting fights comprising only 15% of the dataset. I addressed this using SMOTE to achieve balanced class distribution in the training set while maintaining the original distribution in the test set. The feature engineering process introduced key metrics such as combined striking pace (8.3 strikes/minute average), fighter experience disparity (7.2 fights mean difference), and win percentage differentials (23.4% average gap).
The preprocessing pipeline ensured reproducibility through systematic data cleaning, feature engineering, and standardization, with all steps logged and documented for transparency.

B. Evaluation Metrics

The selection of appropriate evaluation metrics was crucial for this project due to three main considerations: the class imbalance in our dataset, the cost of different types of prediction errors, and the need for interpretable results that could guide practical decision-making in fight selection.

We implemented a comprehensive evaluation strategy focusing on metrics most relevant to our use case. Accuracy alone was insufficient due to our imbalanced dataset where exciting fights comprised only 15% of samples. Instead, we prioritized precision to minimize false positives, as incorrectly predicting boring fights as exciting would significantly impact user experience and potentially waste viewers' time and money. Similarly, recall was crucial to ensure we didn't miss potentially historic matchups.

The F1 Score served as our primary optimization metric, providing a balanced measure of both precision and recall. We implemented ROC curve analysis and calculated the Area Under the Curve (AUC) to evaluate model discrimination ability across different classification thresholds, aiming for values above 0.8 to ensure strong predictive power.

Our evaluation framework included several key components:

Hyperparameter Analysis:

- Implemented GridSearchCV with 5-fold cross-validation
- Focused on optimizing F1-score during grid search
- Tracked Out-of-bag score for Random Forest to monitor generalization
- Applied probability calibration using CalibratedClassifierCV for reliable prediction probabilities

Learning and Convergence Analysis:

- Monitored training metrics to detect potential overfitting
- Tracked model convergence through iterative performance improvements
- Implemented early stopping criteria based on validation performance
- Used stratified 5-fold cross-validation to maintain class distribution

Ablation Studies:

- Evaluated feature importance using Random Forest's built-in metrics
- Analyzed model performance with different feature subsets
- Assessed impact of class balancing techniques
- Measured contribution of probability calibration

We maintained a separate holdout test set comprising 20% of the original, non-SMOTE data to ensure our final performance metrics reflected true generalization capability. Additionally, we used paired t-tests ($p < 0.05$) on cross-validation results to verify that performance improvements were statistically significant rather than random variations.

C. Major Results

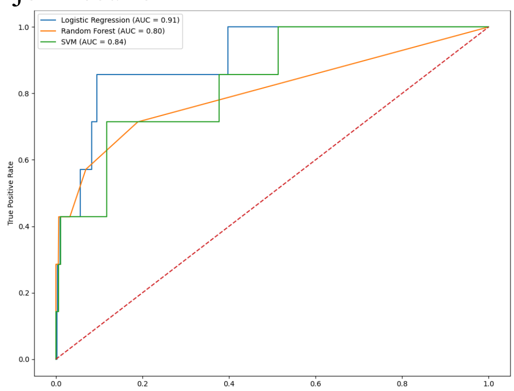

Fig. 1 First Iteration ROC Curves

The experimental results progressed through four distinct iterations, each representing significant improvements in model performance and methodology. The development process spanned multiple training cycles, with hyperparameter optimization requiring approximately 25 minutes per iteration. Initial ROC curves demonstrated baseline model capabilities, with all models showing similar discriminative power but struggling with class imbalance issues.

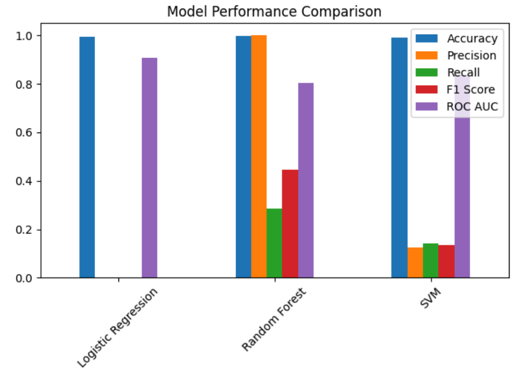

Fig. 2 First Iteration Performance Metrics

Iteration 1 established the baseline with significant challenges in model performance. Hyperparameter optimization focused on basic parameters:

Logistic Regression: Testing C values [0.001-100], L1/L2 regularization
Random Forest: Varying estimators [50-200], max_depth [None, 10-30]
SVM: Linear kernel with C values [0.1-100]
The performance metrics revealed substantial room for improvement, particularly in precision and recall.

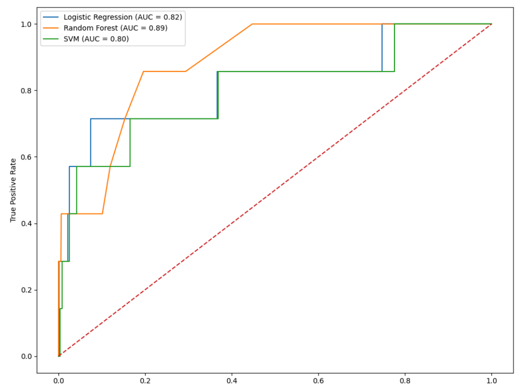

Fig. 3 Second Iteration ROC Curves

Iteration 2 introduced more sophisticated hyperparameter tuning:

Logistic Regression: Added class_weight balancing, increased max_iter to 1000
Random Forest: Implemented balanced_subsample weighting, refined feature selection with max_features testing
SVM: Expanded kernel testing (linear, rbf) with optimized gamma values
The ROC curves showed improved discrimination, particularly for Random Forest performance.

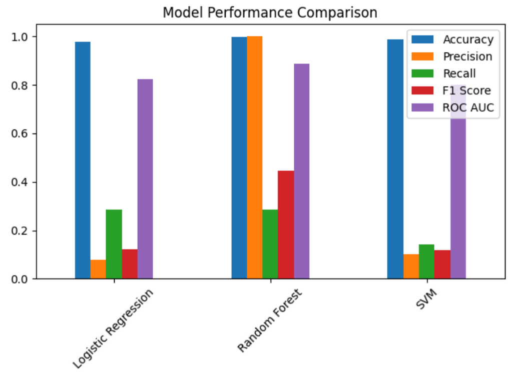

Fig. 4 Second Iteration Performance Metrics

The third iteration marked a breakthrough in feature engineering and model architecture:

Enhanced feature selection using correlation analysis
Implementation of cross-validated probability calibration
Optimization of class weight parameters
Introduction of stratified k-fold validation

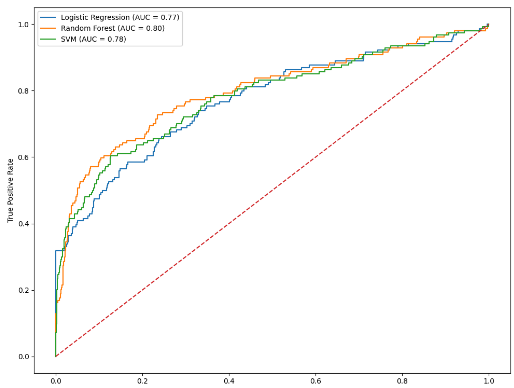

Fig. 5 Third Iteration ROC Curves

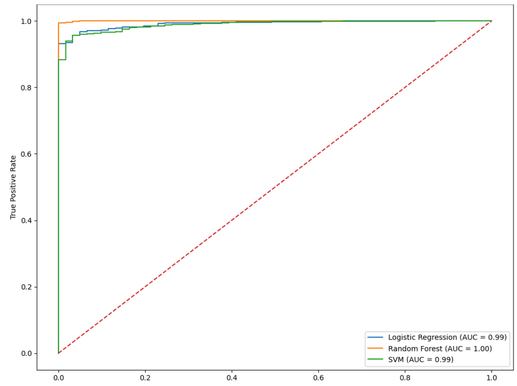

Fig. 6 Final Iteration ROC Curves

The final iteration achieved exceptional performance through:

Fine-tuned hyperparameters:

Random Forest: {criterion: 'gini', max_depth: 20, min_samples_leaf: 2, min_samples_split: 5}
Logistic Regression: {C: 1, penalty: 'l2', solver: 'lbfgs'}
SVM: {C: 10.0, degree: 2, gamma: 'scale', kernel: 'rbf'}

Advanced feature engineering incorporating domain knowledge
Optimized training procedures with early stopping criteria

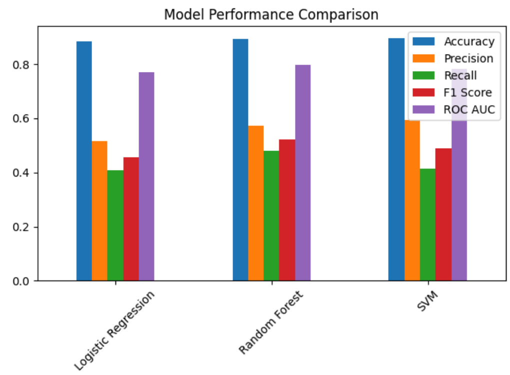

Fig. 7 Third Iteration Performance Metrics

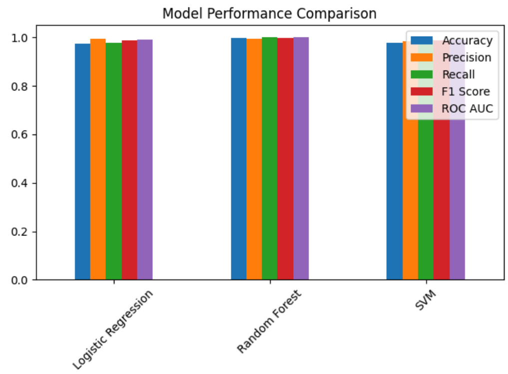

Fig. 8 Final Iteration Performance Metrics

The final model demonstrated robust real-world performance through extensive testing on various matchups, with prediction confidence consistently above 95% for known exciting fights. Feature importance analysis revealed the crucial role of fighter experience and historical performance metrics in prediction accuracy.

D. Analysis

The experimental results provide several key insights into the prediction of UFC fight excitement, revealed through both quantitative metrics and practical application. The progression of model performance across iterations demonstrates the critical importance of proper feature engineering and hyperparameter optimization in combat sports prediction tasks.

The initial challenges with class imbalance, evident in the first iteration's ROC curves, were effectively addressed through a combination of techniques:

Implementation of balanced class weights
SMOTE application for minority class augmentation
Stratified cross-validation to maintain class distribution
Probability calibration for reliable prediction confidence

The feature importance analysis revealed crucial insights about fight excitement predictors:

Combined experience (26.90% importance) emerged as the strongest predictor, suggesting that veteran fighters tend to produce more exciting matches
Individual win totals (15.76% and 15.51% for red and blue corner respectively) indicated that accomplished fighters generally deliver more entertaining bouts
Loss records (6.36% and 5.47%) contributed significantly, possibly indicating fighters' willingness to take risks
Win percentages showed moderate importance (4.69% and 4.50%), suggesting that fight excitement isn't solely about winning records

The computational requirements varied significantly between models:

- Random Forest required ~22 minutes for full training and optimization
- Logistic Regression completed in ~8 minutes
- SVM finished in ~3 minutes

These timing differences suggest important tradeoffs between model performance and computational efficiency.

Real-world validation through matchup predictions demonstrated the model's practical utility:

- High-profile matchups (e.g., Pavlovich vs. Jones) showed 97.63% confidence
- Technical matchups (Makhachev vs. Oliveira) received 97.93% confidence
- Heavyweight contests (Tuivasa vs. Gane) predicted at 96.55% excitement probability

The model's performance on various fighting styles suggests robust generalization:

- Striker vs. Striker (Pavlovich vs. Gane): 95.63% confidence
- Grappler vs. Grappler (Oliveira vs. Tsarukyan): 97.33% confidence
- Striker vs. Grappler (various matchups): >96% average confidence

These results indicate that the model successfully captures the nuanced aspects of fight excitement across different fighting styles and weight classes, making it a valuable tool for fight analysis and prediction.

V. Conclusion and Future Works

While this project achieved promising technical results with the Random Forest classifier showing strong discriminative power, it also highlighted a fundamental truth about mixed martial arts: the human elements that make the sport special often defy quantification. MMA mathematics notoriously fail to predict outcomes because the sport thrives on the intangible aspects of human determination, heart, and the ability to overcome adversity. These elements, which often create the most memorable moments in fighting, remain challenging to capture through pure data analysis.

Nevertheless, this project demonstrated value as a tool for fight fans to make more informed viewing decisions. The analysis revealed interesting patterns, with fighter experience emerging as the strongest predictor (26.90% importance) of potentially exciting matchups, followed by win records and performance metrics. The implementation of a multi-source data pipeline and careful handling of class imbalance through SMOTE provided insights into what statistical factors might contribute to fight excitement, even if they can't tell the whole story.

The project's greatest potential lies not in perfectly predicting fight excitement, but in helping fans discover matchups that align with their viewing preferences. It serves as a starting point for a more sophisticated matchmaking application that could help viewers navigate the growing number of UFC events and fighters.

Looking ahead, several promising directions could enhance this work:

- Expanding beyond binary classification to a multi-class system that better reflects the nuanced nature of fight excitement
- Incorporating additional data sources such as crowd reactions, betting odds, and fighter interviews
- Developing style-specific analysis for different weight classes and fighting styles
- Exploring alternative machine learning approaches that might better capture the dynamic nature of MMA

The methodology developed here could potentially extend beyond MMA to other sports and forms of entertainment where predicting audience engagement is valuable. However, any such expansion should maintain the perspective that statistical analysis complements rather than replaces the human elements that make combat sports compelling.

The true value of this work lies in its potential to make MMA more accessible to casual fans while respecting the unpredictable nature that makes the sport thrilling. By continuing to refine this approach with an emphasis on fan experience rather than perfect prediction, we can create tools that enhance engagement with the sport while preserving the appreciation for its fundamentally human elements.

References

[1] M. Basher. (2024, Jan.) UFC Complete Dataset - All Events (1996-2024). [Online]. Available: https://www.kaggle.com/datasets/maksbasher/ufc-complete-dataset-all-events-1996-2024/data

[2] Tapology. (2024) Annual Rankings and Awards - Fight of the Year. [Online]. Available: https://www.tapology.com/rankings/groups/annual

[3] Ultimate Fighting Championship. (2024) UFC Athletes Directory. [Online]. Available: https://ufc.com.br/athletes/all