Using supervised learning techniques to predict kicking outcomes in the NFL

MSc Research Project
MSc in Data Analytics

Rory Gibney
Student ID: 20167482

School of Computing
National College of Ireland

Supervisor: Mohammed Hasanuzzaman

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Rory Gibney

Student ID: 20167482

Programme: MSc in Data Analytics **Year:** 2022

Module: Research Project

Supervisor: Mohammed Hasanuzzaman
Submission Due Date: 15/12/2022

Project Title: Using supervised learning techniques to predict kicking outcomes in the NFL

Word Count: **5523** **Page Count**: 17

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ……………………………………………………………………………………………………

Date: ……………………………………………………………………………………………………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Using supervised learning techniques to predict kicking outcomes in the NFL

Rory Gibney
20167482

Abstract

This purpose of this research project is to use supervised machine learning techniques in order to predict the outcome of kicks in the National Football League (NFL). There are 2 types of kicks that will be analysed in parallel; field goals and point after touchdowns (PATs). The motivation for this project came from a personal interest in the sport, and the knowledge that the sport is becoming every increasingly influenced by data driven decision making. Kickers are amongst the most important players on any NFL team. Understanding the conditions in which kickers perform better or worse can greatly help in the decision-making process for NFL team coaches. Models that predict kicking outcomes can also be retrospectively used to analyse whether a kicker, in the past, has made the kicks the model has predicted they should male. This paper aims to investigate various classification techniques to establish an optimal model in predicting whether a kicker will successfully make a kick or not, and potentially use this model to make real-time, in-game decisions. Feature selection will play a pivotal role in this process and will distinguish this research from other similar research undertaken to date. This analysis will help NFL team's make decisions around what kickers are considered better than others, and ultimately help with difficult roster decisions. Various models were implemented, including naïve bayes, logistic regression and random forest. The overarching results summissed that predicting unsuccessful kick attempts is a much more difficult proposition than successful kick attempts; with a ranging accuracy between 25 and 90% respectively.

1 Introduction

The National Football League is an annual competition held between 32 competing franchises in the sport of American football. This league is considered the highest level of American football in terms of skill in the world, with the overarching goal of any team competing to win the Superbowl. This affirms their goal to be considered the best team competing in the competition that year. The method by which teams win the Superbowl is by winning enough games during the regular season to qualify for the play-offs, a knockout competition eventually concluding with the Superbowl held every February.

American football is a game played where the primary objective is to score more points than your opponent. There are 2 ways to achieve this; possessing the ball in the opponents end zone (commonly referred to as a touchdown) or kicking the ball between the uprights. Due to the scoring system, most spectator and indeed academic focus is on scoring touchdown. However, the contribution kickers make to their team's success cannot by misconstrued. (Stites, 2017) comments that kickers contribute around 30% of an NFL team's total points scored in each NFL season. In a sport where fine margins regularly dictate the outcome of a game, any way in which teams can optimize performance in this area is decisive.

Coming from a background in predominantly kick-based sports, this researcher always found the kicking elements of the NFL game very intriguing. Questions such as what makes a good kicker, how valuable are kickers and how significant are kickers contribution to team's success all came to mind. There has been some interest in academic circles around predicting the performance of NFL kickers; this dates back as early as 1985 when (Berry & Berry, 1985) conducted an analysis on the probability of an NFL kicker successfully making a kick or not. Although this and many other related works will be explored further in the literature review, the gap this analysis aims to bridge is creating an up-to-date, comparative series of model across recent season in the NFL over a period of several recent seasons to assist team's decision making around whether to or not to attempt a field goal. The main beneficiaries of the analysis will be head coaches of NFL teams whoa re charged with selecting their kicker and when they should kick. Kicker's themselves can also use the analyse to analyse their own individual perfroamnce.

The major contribution f this research will be to create a model that predicts whether a kicker' attempt will be successful or not, which in machine learning terms is a classic binary classification task. The results of this modelling process will be used to see can the model predict when a kick attempt will be successful or not. In particular, the ability to predict unsuccessful field goals will be of much interest, as optimizing these events can be the difference in winning and losing. A series of modelling techniques will be used; amongst these will include the likes of logistic regression, k-nearest neighbours and decision tress for more simplistic modelling techniques. This will be supplemented with the likes of random forest and support vector machine models for more complex approaches. Once models are fitted and trained, they will be used to predict the outcomes on unseen data.

The structure of this report will assume the following shape: the related work section will delve into the modern literature on the topic, highlighting the strengths and weaknesses of current work. The research methodology section will explain how the data that is being used for the analysis is captured, analysed and transformed into suitable data for modelling. This section will also capture the statistical methods and modelling performed on this refined data. The design specification will justify the technologies and techniques used to develop and underpin the ICT solution used to answer the research question. The Implementation section will include only the finalized model used and the steps undertaken as part of this process. The Evaluation will include a detailed account about the findings of the study and analysis of the results achieved through various performance metrics. Finally, the Evaluation section will assess how well the research did in answering the research question, discuss limitations of the work and comment on how future work in this area might be undertaken.

2 Related Work

2.1 In-Domain Literature

There have been several papers released in this domain; both directly related to the research questions, but also in very similar domains. It is crucial to analyse both as part of this research to see what has been achieved to date in this problem to domain, but also to see what other related problem domains have used to solve similar problems to potentially leverage the solutions.

Central to any modelling process is the features chosen; (Hsu et al., 2019) used factors such as distance from filed goal, environmental variables and situational context to analyse an NFL phrase known as ‘choking’; when pressure of the situation causes kickers to miss attempts, they would normally make. Various modelling techniques such as logistic regression and beta-binomial regression was used to ascertain what factors most influence kickers performance. This analysis was useful for identifying potential factors that should be taken into consideration for this researcher’s project, yet it fails to comprehensively cover different modelling techniques and exists in the slightly different domain of ‘choking’ rather than accurately forecasting successful and unsuccessful attempts. A more similar domain was explored by (Pasteur & Cunningham-Rhoads, 2014), who use multivariate logistic regression to compare kicking attempts across 3 seasons. They also, based on their model, correlate kicker salaries to how well their model deemed their performance, an idea that produces very interesting results in that there was a weak correlation between good kickers and higher salaries. There are some notable limitations with this analysis, however; the data used is a small, outdated subset of the population data. This would most likely lead to a failure to capture change in kicker performance over time. The research was also conducted in 2014, which calls into question the validity of the research in the modern game. Kicking performance has improved over time, meaning the usability of this analysis to make decision in the modern game has waned.

More work was carried out by (LeDoux, 2016), who again carried out research as to what factors impacted kickers performance, but their work was motivated by a rule change in the placement of PAT kicks (a specific type of kick a kicker is tasked with). Factors such as distance, climate and stadium were amongst the most important. These factors appear in most modeling features, however the limitation with these types of analysis is it doesn’t take into consideration other external factors a kicker must deal with. These factors include how much pressure the opposition places on the kicker through their rush, and what position the ball is in when the kicker makes their attempt. This author investigated other types of models such as random forests and neural networks. What was of particular intrigue here was the lack of quantifiable performance in deep learning methods versus less complex approaches.

From analysing in-domain previous work, the above were selected as the most relevant material in which to analyse. Patterns amongst them include their use of features, and the techniques used to model the data. Models also seemed to perform well across metrics such as mean squared error and root-mean squared error.

However, there are a couple of clear gaps in this literature; the most glaring being the outdated nature of the data being used. Secondly, there are certainly more factors that can be explored when assessing kicker performance not seen in these. Finally, the extent to which different modelling techniques can be explored should be accounted for in this research; multiple models should be trained and tuned.

2.2 Binary Classification in related domains

The first section of this review was useful in ascertaining what has been done in the past in terms of features used, models explored, and results evaluated within the domain. As this is binary classification task, it is important to evaluate closely related domains to see what potential findings can be brought to this analysis. A more macro level implementation of binary classification in the NFL was undertaken by (Bosch, 2018), who used a plethora of modelling techniques to predict outcomes of NFL games across several years. What was interesting about this was the artificial neural network and convolutional neural network models did not significantly outperform the more simplistic random forest and support vector machine models. This was also noted in the closer related precious work explored in the previous section. In the case where there is no discernible difference in results, simplistic models should be preferred, something which the author noted. This is an important finding to bring forward into the valuation of any models built as part of this research.

The value of analytics in sport as whole was discussed by (Gifford & Bayrak, 2022). The author's noted that ever increasingly sports teams reply and depend upon analytics to strategize pre-game and evaluate post game. This sentiment helps validate the necessity for up-to-date, relevant work in this domain. The author's used decision trees and binary logistic regression to help predict the overall outcome in an NFL game. Whilst their model's achieved serviceable results in correct predicting approximately eighty percent of game outcomes, the authors did note limitations with their work. They offer the idea that, particularly in their decision tree model, the strong correlation between their outcome and predictor variables might be due to causation rather than correlation. Evert modelling process will have a degree of limitation, it is important to convey to consumers of the analysis what these limitations are and how they might impact interpretability of the results.

Valuable insights into some generic problems in the binary classification domain were gained from the research conducted by (Roumani, 2022). This analysis, however, was even more macro than the previous research in this section; the author looked to predict the winner of the Superbowl using data across 15+ seasons. However, it was not the model selection or hyperparameter tuning that was of major interest; it was the use of oversampling of data to account for a deficiency in the prevalence in one of the outcome classes. They used difference rates of oversampling to increase the amount of data passed through to the model and compared model predictions using these different rates. They also quantified the effect increasing the oversampling rate had on model accuracy metrics such as precision and recall. This technique is not something that appeared in any of the related domain literature, which may be interpreted as a potential oversight and indeed opportunity in tis research to implement. The classes should be analyzed in the test data to evaluate whether any other data augmentation technique should be used.

This section has predominantly focused on macro level implementations of binary classification in the NFL; the research being undertaken is indeed micro in nature, so it is important to evaluate other related work. (Wu et al., 2021) analyzed data across all elements of an NFL game (offense, defense and special teams) to predict outcomes of plays. This differed from previously analyzed work as this was a multi-class classification rather than binary. SMOTE was again used to address class imbalance, and the accuracy of the random forest model reached almost ninety percent, which is admirable. However, the authors did note the limitation that data quality coupled with limited features in their dataset proved problematic; this limitation conveys just how important the data exploration and data cleaning phase is in any implementation of machine learning. Identifying and remediating such issues can greatly increase the impact and usability of any research project.

The purpose of analyzing other domains within the NFL was to understand better the techniques used in their binary classification tasks, so that they could be leveraged in this research. It is clear also from the abundance of related work in the NFL machine earning space that there is just cause and validation in the research being undertaken as part of this project.

2.3 Classification in other sport domains

The NFL, as previously mentioned, is becoming ever richer with quality academia due to significant investments from teams as well as widespread data accessibility. Other sports, however, have seen similar growth and should be analysed under this literature survey. One paper of particular interest was written by (Yang et al., 2021), who used binary classification to analyze basketball trajectories to critique player technique. This could then be used to help athletes train and correct potential flaws in their technique. Through analyzing the results, the optimal shooting angle was determined. What was particularly interest about this research was that images from videos were transformed using the Hough transform algorithm[1] into data passed through to support vector machines for classification. (Ji, 2020) cited previous poor image detection technology as a reason why this area had not been further explored, but with recent advancements this type of analysis is becoming more abundant.

Image detection and feature extraction is something used across the NFL, particularly in analyzing offenses and defenses. Examples include the hybrid convolutional neural network and LSTM model employed by (Cheong et al., 2021) in predicting defensive player movements and (Skoki et al., 2021) innovative work with GPS tracking data in the automatic detection of defensive pass interference. It could therefore be prudent to potentially use a level of video or GPS driven data as part of this analysis to determine what impact other players' positions have on the probability of a kicker making a field goal.

Baseball is another American sport which has been heavily influenced by machine learning in the past; (Sidle & Tran, 2018) cited the ever-increasing availability of pitching

[1] Hough transform algorithm: https://homepages.inf.ed.ac.uk/rbf/HIPR2/hough.htm

data as a fundamental behind their multi-class pitch predictor. What was particularly of interest here was the idea that this model could be used to make real-time game decisions; all the features identified were readily available before the outcome occurred, meaning it could be used during a game. It is vital to consider the business context of how your modelling process should be used, and whether that ultimately helps solve the problem from the research question. In the case of the research problem being investigated here, the problem being solved is how relatively good kickers are to one another, and what decisions should NFL teams make about rostering these kickers. There is no necessity for this to be a real-time modelling process like the work undertaken by Sidle and Tran, and NFL teams cannot drop kickers during a game. Keeping the business context firmly in mind when creating any sort of modelling process is often an overlooked but vital consideration.

As the nature of the analysis being undertaken in this research has to do with kicking, analyzing other sports with kicking as part of the fundamental may uncover potential learnings. Football, like the NFL, has many critical analysis pieces on macro level implementation of binary classification. What was of interest about the work undertaken by (Fahey-Gilmour et al., 2019) was their use of automatic feature selection and non-linear relationships between variables to successfully outperform bookmaker performance in Australian football matches. The extensive limitations of their model were explored, and this is a learning to take into this work. Despite achieving relatively solid results, a list of potential improvements and future work was provided. This should be central to any analysis piece so that future work may lean on and indeed improve upon the work carried out.

2.4 Literature Review Synopsys

The key takeaways from the literature review are centered around the idea that there is a gap for relevant literature in the kick outcome prediction space. Any literature undertaken to date would not be applicable in making key decisions in the modern game due to suspect model performance or the outdated nature of when the analysis took place. Looking at literature in other domains highlighted the need for model diversity, careful pre-processing of data and holistic interpretation of what the model is and is not capable of doing. All these variables will be factored into this analysis

3 Research Methodology

3.1 Data Collection

The primary source of data for this project was extracted from Kaggle[2]; a crowd sourced platform in which data scientists can share both data and code in a collaborative fashion.

[2] https://www.kaggle.com/getting-started/44916

There are various datasets of interest regarding NFL and indeed NFL kickers, a summarized list of which can be seen below:

Game data: The `games.csv` contains the teams playing in each game. The *key* variable is **`gameId`**.
Play data: The `plays.csv` file contains play-level information from each game. The *key* variables are **`gameId`** and **`playId`**.
Player data: The `players.csv` file contains player-level information from players that participated in any of the tracking data files. The *key* variable is **`nflId`**.
Tracking data: Files `tracking[season].csv` contain player tracking data from season `[season]`. The *key* variables are **`gameId`**, **`playId`**, and **`nflId`**.
PFF Scouting data: The `PFFScoutingData.csv` file contains play-level scouting information for each game. The *key* variables are **`gameId`** and **`playId`**.

Figure 1- Description of datasets

This can be further complemented with additional datasets from Kaggle; these involve data extracted from MeteoStat and Weather Underground. They contain metadata about weather conditions during NFL games from 2000 to 2020. From the literature review it is clear from previous work that weather has helped explain kicker performance in the past, which justifies its inclusion in this analysis as a potential dataset. Stadium metadata can also be found as a separate dataset on Kaggle; this includes features such as longitude, latitude and stadium design which should also be investigated. Kaggle offers out-of-the-box native functionality to extract data from its website. In this case, all data was presented in the form of comma separated value (CSV) files. Each of these files was extracted to a relative path of a series of Jupyter notebook files.

3.2 Raw Data Analysis

Each dataset went through a series of data quality checks; predominantly at this stage the most important being identification of missing or incomplete data. As was seen in most of the related work, the data was of good quality and there was no need for imputation of any missing values. Other useful metrics such as shape of the data i.e., number of columns, rows, visualization of relevant information and high-level summary statistics were calculated to help shape a better understanding of what the data represented.

What was crucial, however, was the identification of join keys between each of the datasets. As there are several features from each of the datasets that will be used as predictors in the model building process, a method of joining each of the data together was identified for each of the datasets. Predominantly, this was a combination of a game identifier and play identifier. AN important aspect that had to be considered here was the grain of data for each of the datasets; our outcome variable is at the play level of NFL data. Some of the datasets are at different levels and had to be transformed appropriately. An example of the hierarchal levels of data that existed in the tracking_data.csv files can be seen below in Figure 2:

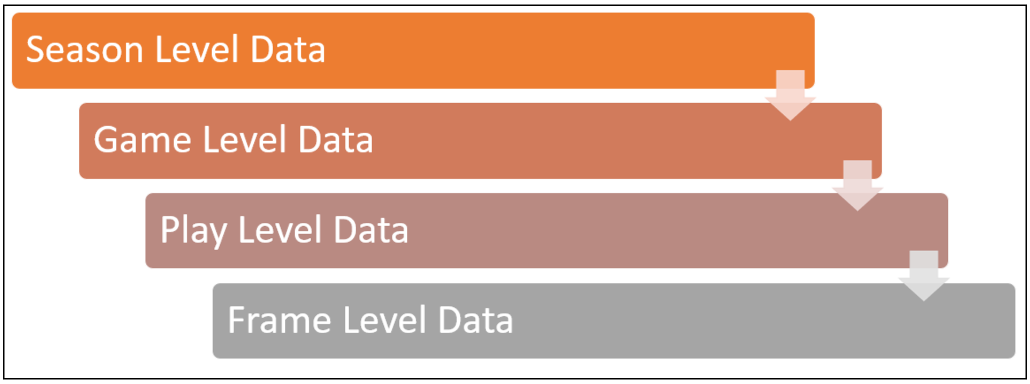

Figure 2- Hierarchal structure of trackingdata

The above hierarchy highlights the level of detail available today in the analysis of NFL games; frame by frame data of NFL games was aggregated over 10 seasons; these frames captured at a fraction of a second rate. Again, this was highlighted as a contributing factor to the rapid ascent of data science usage in the NFL.

3.3 Feature Extraction and Engineering

As part of the data pre processing stage, a set of features needed to be identified and processed accordingly. Leaning on the literature review heavily, the set of features were identified as follows in Figure 3:

Field Name	Description
specialTeamsResult	The outcome variable used to describe whether a kick was successful or not
kickDistance	The distance away from the goal in which the kick was taken from
Temperature	The temperature recorded during the game (Fahrenheit)
DewPoint	Dew point during game (Fahrenheit)
Humidity	Humidity in percentage
Precipitation	precipitation in inches
WindSpeed	speed of wind in miles per hour
Pressure	pressure in inches of mercury
kicker_speed	Speed of kicker kicking ball
kicker_acceleration	Acceleration of kicker kicking the ball
defense_player_speed	Average speed of defending team attempting to block ball
absolute_eucl_distance	Average distance defense is away from kicker when kicking ball
Longitude	Geo location of stadium
Latitude	Geo location of stadium
time_left_seconds	How much time is left in the game when kick occurs
points_differential	The difference in score between team executing kick and opposition
RoofType_Indoor	Roof Type of Stadium (One Hot encoded)
RoofType_Outdoor	Roof Type of Stadium (One Hot encoded)
RoofType_Retractable	Roof Type of Stadium (One Hot encoded)

Figure 3- Field description for refined dataset

Some of these feature’s came directly from the datasets provided, including the dependent variable. This was pre-processed via vectorization for the models to interpret the data. Other predictive features such as Temperature, DewPoint and Precipition came directly from datasets; they did have to be aggregated to the appropriate level, as previously mentioned, but no further feature engineering was needed on them.

However, some of the other features were derived from other data points present in the data. For example, the straight-line distance between the opposing players and the kicker during the action of kicking the ball utilized the frame-by-frame data of each play; this type of innovation is only made possible by advancements in data capture quality.

3.4 Final Data Processing

After further analysis and pre-processing techniques, data from 8 separate csv files was used to create the finalized output for modelling. Once the data was aggregated into this coherent dataset, the data then had to be appropriately prepared for the modelling process.

Data was separated into 2 distinct types: numeric and categorical. Numeric data was analysed and, in some cases, transformed using the appropriate scalers, and categorical data was vectorized for interpretation by the model. The most important of these variables being the dependent variable, which was divide into the 2 classes of successful kick attempt and non-successful kick attempt. As was to be expected from the literature review, there was a clear class imbalance between successful and non-successful kicks. This class balance from the finalized data set can be seen below:

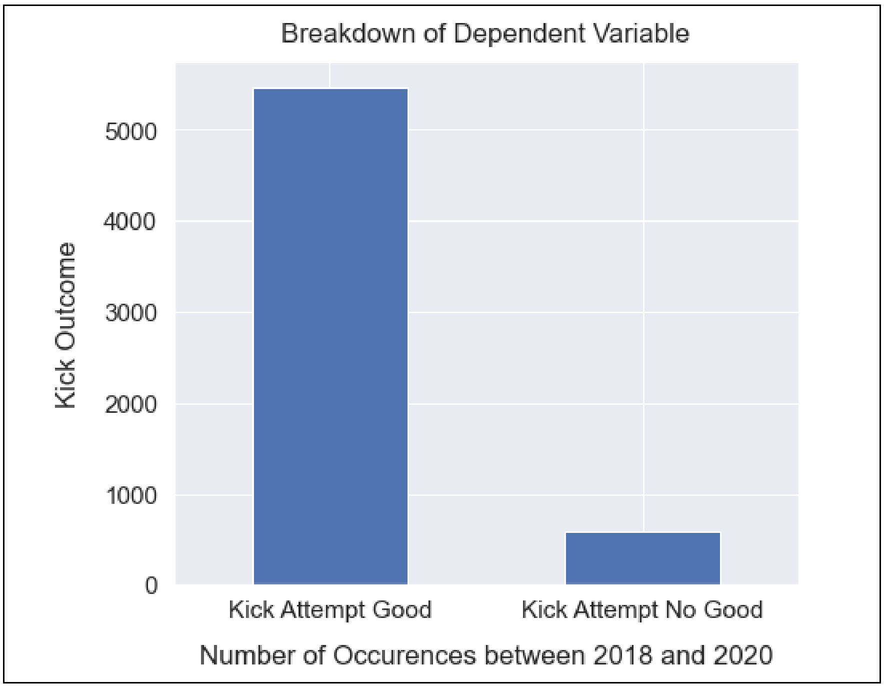

Figure 4- Breakdown of class imbalance present in the dependent variable

When modelling, this type of information is very useful in interpreting the results; a method of dealing with this type of class imbalance may be needed to achieve the desired results.

Most data was numeric, with the exception being the variable donated to describing the roof of the stadium. This was transformed to be represented across multiple variables, once again so that the model could be interpreted correctly. The data was split into train and test with a split of 75:25.

After preliminary models were built, tuned and subsequently analysed, it was discovered that there was a clear bias caused by the class imbalance. As there was no authentic data that could be further gathered, an artificial way of creating more samples of data was explored.

3.5 SMOTE Data Transformation

(Fernandez et al., 2018) inferred that Synthetic Minority Oversampling Technique was the 'de-facto' method for addressing class imbalance within a dataset. The duplication of existing samples within the dataset for the minority class was considered; however, this would not add any new variance or information from which our models to learn from. Using this data augmentation algorithm, the test data was transformed to include equal amounts of both classes. This new sample of test data was used to train the models to predict upon the class of the test dataset.

4 Design Specification

The process model that was adapted for the purpose of this research was CRISP-DM- cross industry standard process for data mining. When deciding which model to follow, the work of (Saltz, 2021) played a pivotal role as the benefits highlighted in their article suited the work being undertaken as part of this research. Succinctly, the adaptable and cyclical nature of CRISP-DM allowed several instances of project to take place. Additionally, the flexibility afforded in areas such as stakeholder involvement, which this research will not have, makes it the outstanding candidate. Having the six stages, as depicted in Figure 5, helped shape this project into measurable milestones:

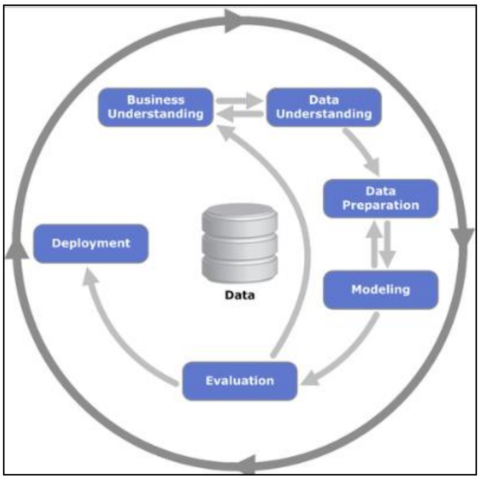

Figure 5- CRISP-DM Methodology

The research was conducted on a Dell XPS 13 7390 machine using the Windows 11 operating system. The various data analysis and modelling steps were conducted in the Python programming language. The main reason supporting this decision was the vast array of libraries relating to data processing provided by Python, in particular the Pandas data frame object suited the tabular nature of the data. Pandas' data frames are an essential part of the toolkit of data scientists, according to (Bantilan, 2020), due to their robust nature. Some of the datasets used in this analysis are quite large in terms of volume, so the ability to process, clean, visualize and ultimately model data in pandas' data frames justifies the use of Python as the programming language of choice for this analysis.

5 Implementation

The main output of this research is trained models that help classify the research question. They are presented in a Jupyter notebook, entitled 03_modelling_and_evaluation.ipynb. The models are based on another output; this is a .csv file generated by aggregating, pre-processing and cleaning datasets mentioned previously in this report.

5.1 Modelling Process

Several models were created to evaluate which of these could potentially best describe the dependent variable. A list of these can be seen below

- Logistic Regression
- Naïve Bayes
- k-nearest neighbors
- Decision Tress
- Random Forest
- Support Vector Machines

For the evauklation3 of the most performant models will be evaluated and the main findings of these models explored with regards to the research question.

6 Evaluation

The whole purpose of this project was to predict the outcome of kicks made by NFL kickers; as this results in a binary classification task, many of the models were evaluated using the following metrics:

- Precision
- Recall
- F1 Harmonic Mean

This allowed comparison intra models to evaluate which performed better at solving the research question.

6.1 Naïve Bayes Model

The Naïve Bayes was selected to be analysed due to its simplicity; understanding and implementing this model is relatively straightforward compared to other modelling techniques. There is no method to hyperparameter tune this type of model, so once implemented the results below were interpreted:

Classification Report for fitted Naive Bayes model on Test Dataset

	precision	recall	f1-score	support
0	0.94	0.60	0.74	1296
1	0.13	0.60	0.21	126

Figure 6- Naive Bayes summary statistics

From the results above, it can be inferred that the model is in fact good at predicting successfully field goal attempts. It predicts 94% of successful field goals attempts correctly. However, the accuracy of the model drops substantially when predicting failed field goal attempts.

6.2 K-nearest neighbor Model

Another model that was implemented was a k-nearest neighbor classifier. This was the first model was discernible hyperparameter tuning; several iterations of the model were run for different values of k; an optimal value of 15 was found, which was used to help generate the below confusion matrix:

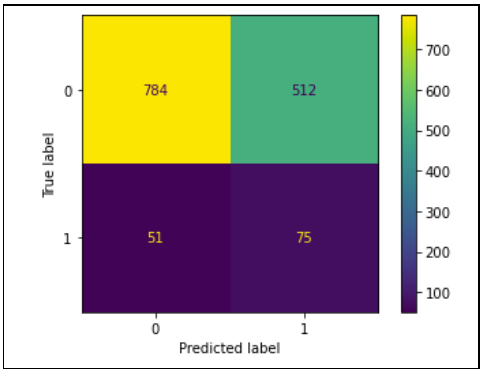

Figure 7- k-nearest neighbor confusion matrix

This model again did well in predicting the successful attempts; but again, there is a dramatic drop in the accuracy of unsuccessful kick attempts. This was a commonality across many of the train models; including the logistic, support vector machine a decision tree models.

6.3 Random Forest Model

The most performant model across both desired classes was the random forest model; it achieved over ninety percentage efficiency in identifying true accurate kicks, and twenty five percent accuracy in identifying inaccurate kicks. Different numbers of trees were used in this ensemble model; random forest build a series of weaker decision trees and uses voting principles to estimate the class in classification tasks. The confusion matrix below in Figure 8 depicts visually the results of this model:

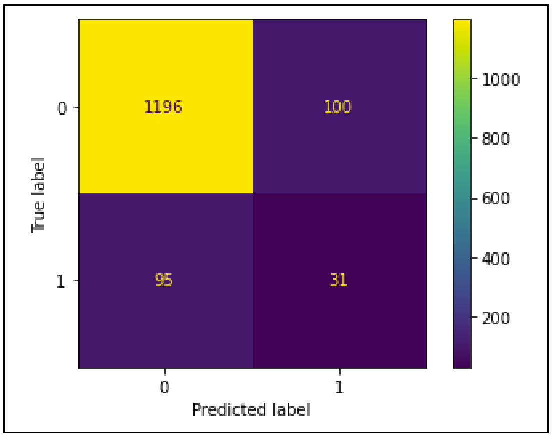

Figure 8- Random Forest Confusion Matrix

7 Discussion

A subset of the most important models evaluated for this process was described under the Evaluation section. What is clear from these models, and indeed all the models, was that accurately predicting successful field goals was commonplace. However, the accuracy of the models fell substantially when predicting unsuccessful field goals. Thinking about the domain application of these models, it would be much more useful to understand the factors which cause kicks to be inaccurate rather than accurate.

What would help this kind of analysis is some of the findings from the literature review; some of the relevant features mentioned in related work such as individual kicker history, wind direction and more time agnostic weather conditions. The design of this research leant firmly on the assumption, gained from the literature review, the neural networks and other forms of deep learning did not significantly outperform those of other machine learning techniques. These more complex models may have been able to distinguish the patterns in the data more and ultimately help with the classification task.

The explanatory variables that were used in the analysis could have gone through more layers of transformations; one of the key assumptions of this research was the importance of avoiding a black box model process, as consumers would want to understand each of the features. There is a trade-off here, however, as the use of some dimensionality reduction techniques may have helped correlated the explanatory variable with the target variable. In particular, the use of an algorithm such as principal component analysis might have been useful and help achieve better, more accurate modelling results.

Overall, the results are mixed in terms of meeting the research question. The models built, the Random Forest model, had an overall prediction success rate that was serviceable. However, with the clear bias in predicting successful kicks much more accurately than unsuccessful, there is a clear caveat with this analysis, using some of the different features and techniques mentioned in this discussion, more serviceable results could be attained by future work.

8 Conclusion and Future Work

The research question that was proposed at the beginning of this project was how could supervised learning techniques be applied to help predict the outcome of kicks in the NFL. To a degree, this question was answered through the work undertaken; throughout all the various models created, there was a high accuracy overall in total outcome predictions. However, when analysed more closely the recall and precision of our binary classification models did not meet expectations. This was in part due to the vast class imbalance in the data; almost eight five percent of kicks in the NFL are successful in the data between the 2018 and 2020 seasons.

This was a limiting factor in the analysis, as the amount of data used to both train and test the models was not vast. The reason behind this design decision was to include the video frame by frame data that was not present in datasets prior to 2018. One potential area for further work would to be use more data from the past to build a larger pool for the models to work with. This would have to be pre-processed accordingly to merge with the data used in this analysis, but more of the variation in kicking tendencies could potentially be identified using this.

From the literature reviewed, it is clear there is a huge future for analytics in the whole domain of the NFL. In particular, the use of data derived for videos could be used to help determine the likelihood of a kick to be successfully made by a kicker or not. A CNN or LTSM neural network model could be used to deliver a second-by-second analysis on whether the kick should be made; this would have great commercial for teams trying to strategize how to give their kicker the best opportunity at making a kick.

Commercialization for any outcome prediction process in the NFL will never be an issue; in the recently completed 2021 season, the 32 competing teams in the NFL generated revenue northwards of seventeen billion US dollars. Coupled that with is the ever-growing partnership between the NFL and Amazon Web Services to provide 'Next Gen Stats'- a proprietary service provided by Amazon Web services which leverages data science in order to provide real time analytics during NFL games. Figure 9 below shows a process flow:

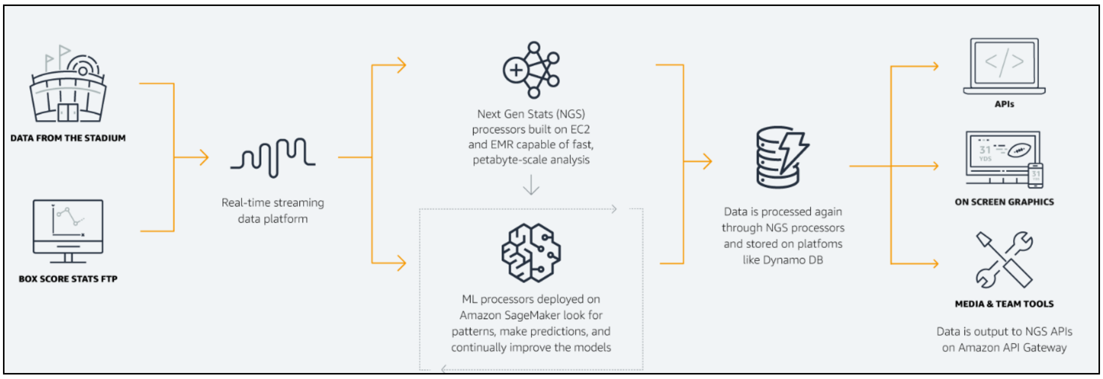

Figure 9- Next Gen Stats Process Flow

Amazon cite formations, routes and events as areas they can leverage machine learning models to provide real time analytics during games. The scope for kicking outcome predictions is extremely large due to both the demand and infrastructural set-up already in place to analyse the data, so future work in this domain should be explored.

The goal of this research was to see if supervised learning could be used to predict kicking outcomes in the HFL; this question was partially answered through the results attained but what this research also does is highlight and pave the way for future work to be done.

References

Bantilan, N. (2020) "Pandera: Statistical data validation of Pandas Dataframes," *Proceedings of the Python in Science Conference* [Preprint]. Available at: https://doi.org/10.25080/majora-342d178e-021.

Berry, D.A. and Berry, T.D. (1985) "The probability of a field goal: Rating kickers," *The American Statistician*, 39(2), p. 152. Available at: https://doi.org/10.2307/2682830.

Bosch, P., (2018). Predicting the winner of NFL-games using Machine and Deep Learning. *Vrije universiteit, Amsterdam*.

Cheong, L. L., Zeng, X., & Tyagi, A. (2021). Prediction of defensive player trajectories in NFL games with defender CNN-LSTM model. MIT Sloan School Sports Analytics Conference 2021. https://www.amazon.science/publications/prediction-of-defensive-player-trajectories-in-nfl-games-with-defender-cnn-lstm-model

Dangeti, P. (2017) "Logistic Regression Versus Random Forest," in Statistics for Machine Learning: Techniques for exploring supervised, unsupervised, and reinforcement learning models with python and R. Birmingham, UK: Packt Publishing.

Fahey-Gilmour, J. *et al.* (2019) "Multifactorial analysis of factors influencing elite Australian football match outcomes: A machine learning approach," *International Journal of Computer Science in Sport*, 18(3), pp. 100–124. Available at: https://doi.org/10.2478/ijcss-2019-0020.

Fernandez, A. *et al.* (2018) "Smote for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary," *Journal of Artificial Intelligence Research*, 61, pp. 863–905. Available at: https://doi.org/10.1613/jair.1.11192.

Gifford, M. and Bayrak, T. (2022) "Application of decision tree and logistic regression models in predicting outcomes in the NFL," *SSRN Electronic Journal* [Preprint]. Available at: https://doi.org/10.2139/ssrn.4263776.

Hsu, N.-W., Liu, K.-S. and Chang, S.-C. (2019) "Choking under the pressure of competition: A Complete Statistical Investigation of pressure kicks in the NFL, 2000–2017," *PLOS ONE*, 14(4). Available at: https://doi.org/10.1371/journal.pone.0214096.

Ji, R. (2020) "Research on basketball shooting action based on image feature extraction and machine learning," IEEE Access, 8, pp. 138743–138751. Available at: https://doi.org/10.1109/access.2020.3012456.

LeDoux, J. (2016) "Extra point under review: Machine Learning and the NFL field goal," *Elements*, 12(2). Available at: https://doi.org/10.6017/eurj.v12i2.9448.

Pasteur, R.D. and Cunningham-Rhoads, K. (2014) "An expectation-based metric for NFL field goal kickers," *Journal of Quantitative Analysis in Sports*, 10(1). Available at: https://doi.org/10.1515/jqas-2013-0039.

Roumani, Y.F. (2022) "Sports analytics in the NFL: Classifying the winner of the Superbowl," *Annals of Operations Research* [Preprint]. Available at: https://doi.org/10.1007/s10479-022-05063-x.

Saltz, J.S. (2021) "CRISP-DM for data science: Strengths, weaknesses and potential next steps," *2021 IEEE International Conference on Big Data (Big Data)* [Preprint]. Available at: https://doi.org/10.1109/bigdata52589.2021.9671634.

Sidle, G. and Tran, H. (2018) "Using multi-class classification methods to predict baseball pitch types," *Journal of Sports Analytics*, 4(1), pp. 85–93. Available at: https://doi.org/10.3233/jsa-170171.

Skoki, A., Lerga, J. and Stajduhar, I. (2021) "ML-based approach for NFL defensive pass interference prediction using GPS tracking data," *2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO)* [Preprint]. Available at: https://doi.org/10.23919/mipro52101.2021.9596877.

Stites, A. (2017) *Why do NFL teams ignore kickers?, SBNation.com*. SBNation.com. Available at: https://www.sbnation.com/2017/11/15/16596068/nfl-kickers-coaching-special-teams (Accessed: December 8, 2022).

Wu, ason, Gunnell, E. and Sun, Y. (2021) "PlayGuessr: Commercial Application of machine learning in football play prediction," *Computer Science and Information Technology Trends* [Preprint]. Available at: https://doi.org/10.5121/csit.2021.111714.

Yang, Q., Shao, J. and Zuo, H. (2021) "Automatic analysis of basketball shooting based on machine learning," *2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)* [Preprint]. Available at: https://doi.org/10.1109/icaica52286.2021.9498159.