Formula 1(F1) Race Winner Prediction using Artificial Neural Network

Atharva Vichare
*CIS*6020: Artificial Intelligence*
University of Guelph
Under the Guidance of: Prof. Bruce Neil

***Abstract*—Artificial Neural Networks (ANNs) have demonstrated significant success in predictive modelling across various fields, including sports analytics. This project leverages ANNs to predict the winners of Formula 1 (F1) races in 2022, utilizing historical data from 1980 to 2021. The dataset comprises of variables such as qualifying, status, driver and constructor statistics, and performance metrics. The primary objective is to develop a feedforward neural network model capable of predicting race winners in upcoming races. By employing advanced feature engineering and optimizing model parameters, the project seeks to maximize the accuracy of race outcome predictions. Through this work, we aim to deepen our understanding of the factors influencing F1 race results and enhance our ability to fine-tune neural networks for accurate predictive modelling.**

***Index Terms*—neural networks, model architecture, F1 race prediction**

I. Introduction

Artificial Neural Networks (ANNs), inspired by biological neural structures, have demonstrated remarkable effectiveness in a wide range of applications, particularly in predictive modelling. Their capacity to handle complex and nonlinear relationships makes them a powerful tool in domains like classification, pattern recognition, optimization, and time-series forecasting. Unlike traditional models such as regression, which often struggle to capture intricate patterns due to their linearity, ANNs offer a flexible and scalable approach to address these challenges. The success of ANNs has been evident in diverse prediction tasks, including weather forecasting, financial trends, and even sports results.

This project applies ANNs to predict Formula 1 (F1) race results, specifically focusing on predicting race winners during the 2022 season. The model is trained using a rich dataset that includes historical data from 1980 to 2022, featuring variables like weather conditions, driver and constructor statistics, and race performance metrics. The primary goal is to develop a neural network capable of predicting the finish position of drivers in upcoming races. In this project, hyperparameter tuning is conducted through a grid search to identify the optimal neural network architecture, with the evaluation metric being accuracy. The chosen architecture is designed to maximize predictive accuracy, allowing reliable predictions of race outcomes.

By applying advanced feature engineering and model optimization, the project not only aims to improve the prediction of the F1 race results, but also improves the understanding of the factors influencing these results. This research contributes to the advancement of the application of ANN in sports analytics and predictive modelling.

This report is organized in the following Sections:

- Understanding F1 Racing
- Methodology
- Related Work
- Data Structure
- Feature Engineering
- Imputing Missing Value
- Feature Scaling and One Hot Encoding
- Architecture
- Hyperparameter Tuning
- Model Predictions
- Model Evaluation
- Conclusion & Future Work

II. UNDERSTANDING F1 RACING

A. F1 Racing Overview

Formula 1 (F1) racing is the highest class of single-seater auto racing, governed by the FIA. Cars race on circuits and street tracks, reaching extreme speeds. Races, known as Grands Prix, are held worldwide, with drivers competing for the World Championship title. Strategic elements like tire management and pit stops add to the complexity and excitement of the sport.

B. Characteristics of F1 Racing

In Formula 1 racing, cars and drivers must adhere to strict regulations. Exceeding track limits or causing collisions can lead to penalties or disqualification. Drivers typically begin their careers in junior formulas before stepping up to F1, with no specific age restrictions, although most start in their early twenties.

Fig 1: Formula 1 Racing

C. Track Surfaces

Formula 1 races are held on a variety of surfaces, including permanent circuits, street tracks, and hybrid layouts. Track characteristics like asphalt composition, elevation changes, and corner types can significantly affect car performance and driver strategy.

D. The Start

The start of an F1 race is critical for the outcome. The race begins with a standing start from a grid, where drivers line up based on their qualifying times. A series of red lights signals the start: all lights illuminate and then go out simultaneously, indicating the drivers should start.

Fig 2: Starting grid

Standing Start: Drivers line up according to their qualifying positions. Then a series of red lights signals the start. Lights out means go.

Qualifying Determines Grid Positions: A. Held over three sessions (Q1, Q2, Q3). B. Top 10 positions are decided in Q3. C. Penalties can alter starting positions.

Race Procedures: A. False starts or jumping the start incur penalties. B. Drivers must maintain their positions until they cross the starting line. C. Incidents at the start may trigger safety cars or race restarts.

Grid Position:

Position	Determination
1st row	Fastest qualifiers
2nd row	Next fastest
3rd row	Continuing down

Strategic elements and split-second decisions at the start play a vital role in the overall race strategy and outcome.

E. The Race Dynamics

In Formula 1 racing, the main objective is to complete the designated laps in the shortest possible time. Drivers must maintain control and optimize their speed while adhering to strict racing rules. Securing victory requires not only crossing the finish line first, but also skilful management of tire wear, fuel load, and strategic pit stops. Deviations from rules or errors can lead to penalties, significantly affecting the race outcome.

F. Strategies

Driver strategies in an F1 race vary significantly. Some drivers aim to take an early lead and maintain it throughout the race, leveraging their position to control the pace. Others employ a more tactical approach, conserving tires and fuel, waiting for the optimal moment to overtake competitors, often during pit stops or towards the final laps. Decisive maneuvers and strategic decisions play a crucial role in securing victory.

G. Disqualification Rules

In Formula 1 racing, disqualification can result from various infractions to ensure fair competition and maintain the integrity of the sport. Key disqualification reasons include:

- Exceeding track limits: Persistently going off track can lead to penalties or disqualification. Drivers must keep at least one wheel within track boundaries.
- Unsafe driving: Causing avoidable collisions. Ignoring flag signals or race marshals' instructions.
- Technical regulations: Non-compliance with car specifications. Illegal modifications or using unapproved parts.
- Race conduct: Jumping the start or false starts. Failing to complete the race in the allotted time.

Fault	Definition
Exceeding Track Limits	Going beyond track boundaries repeatedly
Unsafe Driving	Causing collisions or ignoring safety instructions

Technical Violations	Non-compliance with car specs or illegal modifications
Jumping Start	Moving before the starting lights go out

Understanding these rules and maintaining compliance is crucial for success in Formula 1 racing.

H. Points Distribution

The Formula 1 (F1) points distribution system rewards drivers based on their finishing positions in each race. Points are allocated as follows:

Position	Points
1st	25
2nd	18
3rd	15
4th	12
5th	10
6th	8
7th	6
8th	4
9th	2
10th	1

Points Distribution

In addition to race position points, drivers can earn an extra point for setting the fastest lap if they finish in the top 10. Moreover, F1's sprint qualifying races award additional points to the top three finishers, with 3 points for the winner, 2 points for second place, and 1 point for third place. These points contribute to the Drivers' and Constructors' Championship standings, determining the season's champions. Understanding the points distribution system is crucial for grasping the competitive dynamics of Formula 1 racing.

III. METHODOLOGY

The approach is as follows:

- Data Collection: Extract and prepare the dataset collected. The dataset consists of all information on the Formula 1 races, drivers, constructors, qualifying, circuits, lap times, pit stops, championships from 1950 till the latest 2024 season.
- Data Pre-processing: Perform data cleaning, data preprocessing and join the datasets to form one final dataset.
- Feature Engineering: Create new features based on the existing data to help improve model accuracy
- Model Tuning and Training: Perform Grid Search to get the best hyperparameters and training the model using these hyperparameters on the training set.
- IsRaceWinner Prediction: Deploy model to predict race winners for 2022 season

This methodology aims to uncover patterns for accurate predictions in F1 Racing.

IV. RELATED WORK

Predicting outcomes in motorsports, including Formula 1 (F1) racing, has been a focus of research in both sports analytics and artificial intelligence. Various approaches have been explored to predict race results, employing techniques ranging from traditional statistical methods to more advanced machine learning algorithms such as artificial neural networks (ANNs).

Machine Learning in Motorsports:

The application of machine learning to motorsports has gained traction in recent years, particularly in predicting race outcomes and optimizing strategies. For example, Liu and Fotouhi (2020) explored the use of artificial neural networks (ANNs) combined with Monte Carlo tree search for developing race strategies in Formula E, focusing on optimizing pit stop decisions and overall race performance. Their work demonstrates the potential of AI in enhancing decision-making in racing contexts by predicting key events such as pit stops and race outcomes [1].

Similarly, Kumar and Preethi (2023) applied machine learning techniques to analyse Formula 1 races, developing models to predict race results using historical performance data, weather conditions, and driver statistics. Their study highlights the importance of incorporating various dynamic factors, such as driver skill, team strategies, and environmental conditions, to improve the predictive accuracy of models [4].

Horse Racing Prediction Using ANNs:

While not directly related to F1, several studies have applied ANNs to predict outcomes in horse racing, which shares similar challenges such as class imbalance and the dynamic nature of races. Davoodi and Khanteymoori (2010) used ANNs to predict horse racing outcomes, incorporating factors such as past performance, weather, and track conditions. This approach showed the capability of ANNs to learn from complex, non-linear data, an essential feature when dealing with unpredictable outcomes in competitive sports [2]. This work provides insight into feature selection and model tuning, which are similarly relevant in the context of F1 race prediction.

Virtual Strategy Engineer for Motorsport:

Heilmeier et al. (2020) proposed the "Virtual Strategy Engineer," a model that uses ANNs to make race strategy decisions in motorsports, such as determining the optimal time for pit stops based on real-time data and predictions about driver performance. Their research emphasizes the importance of real-time data integration and strategic decision-making, like how race-winning predictions in F1 could benefit from incorporating dynamic race data like weather changes, tire wear, and pit stops [3].

Deep Learning Models for Race Prediction:

The use of deep learning models, including deep neural networks (DNNs), for predicting race results has shown promising results in various domains. The study by Williams and Li (2008) applied neural networks to predict horse racing outcomes, showcasing how deep learning can be leveraged to identify complex patterns and dependencies within historical data. Their findings underscored the ability of neural networks to predict non-linear outcomes and the potential benefits of using deep architectures for better prediction performance in sports [11]. This resonates with the approach adopted in our project, where deep learning is used to predict F1 race winners by processing large amounts of historical race data.

The literature reviewed reveals a growing interest in using machine learning and artificial neural networks to predict outcomes in motorsports, particularly in Formula 1. Previous studies have explored various aspects of race prediction, from optimizing race strategies to forecasting results based on historical data. Our project builds on these advancements by utilizing a feed-forward neural network (ANN) model to predict F1 race winners, using data from multiple seasons to understand the relationships between key factors such as driver statistics, constructor performance, and race conditions.

By reviewing related work in the field, we see that the combination of feature engineering, model optimization, and deep learning techniques holds great promise for improving the accuracy of sports predictions, and we aim to continue building on these foundations to create more robust prediction models for F1 racing.

V. Data Structure

We fetched data from Ergast API but since Kaggle had the same dataset we originally decided to pick the data from Kaggle. The dataset consists of all information on the Formula 1 races, drivers, constructors, qualifying, circuits, lap times, pit stops, championships from 1950 till the latest 2024 season. The data was then segmented to include only the years 1996, 2000, and 2006-2022, as the qualifying pattern was similar for these years and since the build of the cars in 1950's was way to different we decided to go from the year 1996.

The final data frame after combining various datasets has total 22 columns before Feature Engineering.

Data 1: Results

race_Id: A unique identifier for each race event. This column helps in distinguishing between different races across seasons.

driver_Id: A unique identifier for each driver participating in the race. It corresponds to the driver's specific details, such as nationality, age, and career statistics.

constructor_Id: A unique identifier for the constructor (team) that the driver is representing in the race. This refers to the car manufacturer and the team behind the vehicle, such as Ferrari, Mercedes, etc.

grid: The position that the driver achieved during the qualifying session. This is crucial in determining the starting grid for the race and can often correlate with a driver's race performance.

position: The final position of the driver in the race. It indicates the rank or finishing position that the driver achieved at the end of the race (e.g., 1st place, 2nd place, etc.).

status Id: A numerical identifier representing the status of the driver during the race. This could indicate whether the driver finished the race, retired early, or did not finish (DNF). Common statuses include finished, retired, and disqualified.

Data 2: Races

race_Id: A unique identifier for each race event. This column helps to distinguish between different races in the dataset and is used to link the race results with other race-related data.

year: The year in which the race took place. This helps categorize races by season, which is essential for tracking driver and constructor performance over time.

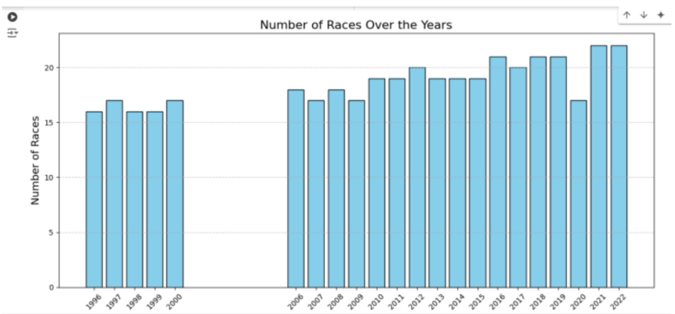
round: The round number within the season (e.g., Round 1, Round 2, etc.). This indicates the specific race within the championship series for that year, as the Formula 1 season typically consists of multiple races.

circuit_Id: A unique identifier for the circuit (track) where the race took place. This ID links to the details of the circuit, such as its name, location, and characteristics.

Circuit: The name of the circuit or track where the race occurred. This can be a famous circuit, such as Monaco, Silverstone, or Suzuka.

date: The exact date when the race took place. This is important for analysing race performance and correlating with external factors like weather conditions on the race day.

Fig 3: No. of races over the years



Data 3: Qualifying

race_Id: A unique identifier for each race event. This column links the qualifying data to the specific race event.

driver_Id: A unique identifier for each driver participating in the qualifying session. It allows you to associate the qualifying results with specific drivers.

q1: The time achieved by the driver in the first qualifying session (Q1). This is typically the first part of qualifying where all drivers compete, and the slowest drivers are eliminated at the end of the session. The lower the time, the better the performance.

q2: The time achieved by the driver in the second qualifying session (Q2). After Q1, the fastest drivers progress to Q2, where the slowest are again eliminated. The times in Q2 determine the starting positions for the race grid, with the fastest securing spots near the front.

q3: The time achieved by the driver in the third qualifying session (Q3). This is the final round of qualifying, and the top drivers compete for the best starting positions on the grid. The fastest time in Q3 determines the pole position (starting in 1st place).

Data 4: Drivers

driver_Id: A unique identifier for each driver. This ID is used to distinguish between drivers and link their data across various race-related datasets.

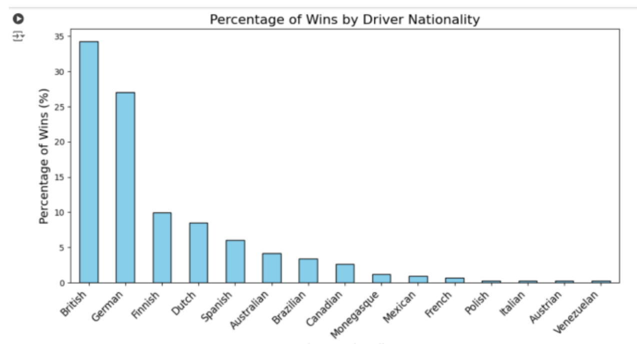
forename: The first name (given name) of the driver. This is part of the driver's personal identification, alongside their surname.

surname: The last name (family name) of the driver. It is used together with the forename to identify the driver in official records.

dob: The date of birth of the driver. This helps in identifying the driver's age and can be important for demographic analysis of the drivers in the F1 dataset.

nationality: The nationality of the driver. This indicates the country the driver represents in Formula 1, often tied to the flag they race under and their home country's motorsport culture.

Fig 4: Wins by nationality



Data 5: Constructors

constructor_Id: A unique identifier for each constructor (team) participating in Formula 1. This ID is used to link the constructor's data with the corresponding team and their performance throughout the season.

name: The name of the constructor (team) in Formula 1. This could be the name of the car manufacturer or the team, such as Ferrari, Mercedes, Red Bull Racing, etc. The constructor's name is key in associating the team with its respective drivers and performance data in the races.

Data 6: Status

status: This column describes the status of the driver during the race. It indicates whether the driver finished the race, retired early, or faced any other specific race conditions like disqualification or an accident. Common statuses include:

- " Finished" – Driver completed the race.
- " Retired" – Driver did not finish the race due to mechanical failure, crash, etc.
- " Disqualified" – Driver was removed from the race for rule violations.

status id: A numerical identifier for the status of the driver. It is a coded representation of the different possible statuses that a driver can have during the race.

Data 7: Lap times

race_Id: A unique identifier for each race event. This column helps link the lap data to a specific race.

driver_Id: A unique identifier for each driver. This column links the lap data to a specific driver and their performance during the race.

lap: The lap number during the race. Formula 1 races typically consist of multiple laps, and this column indicates the specific lap for which the data is recorded (e.g., lap 1, lap 2, etc.).

position: The position of the driver at the end of the lap. This indicates the driver's rank in the race at the specific lap, showing whether they are in 1st, 2nd, 3rd, etc.

time: The time taken by the driver to complete the lap. This is usually recorded in a standard time format (e.g., m: ss.sss) and indicates the driver's performance on that specific lap.

milliseconds: The lap time recorded in milliseconds. This provides a precise measurement of the time taken to complete the lap, which can be crucial for fine-tuned analysis of driver performance.

These datasets were then merged to form a final dataset for our prediction.

VI. Feature Engineering

DidNotFinish: This column is a binary indicator that denotes whether a driver finished the race. A value of 1 indicates the driver did not finish due to reasons such as accidents, mechanical failures, or disqualification, while a value of 0 indicates the driver completed the race. This column simplifies race outcome analysis

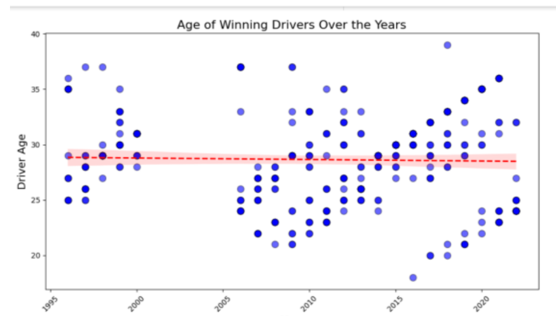
IsRaceWinner: This is the target variable we aim to predict, indicating whether a driver won the race. A value of 1 indicates the driver finished in 1st place (race winner), while a value of 0 indicates the driver did not win the race. This column helps identify race winners and can be used for performance analysis or as a target variable in predictive models.

podium_last_5 races: This column counts how many times a driver has finished in the top 3 (podium positions) across their last 5 races. It is calculated by checking if the rank is 1, 2, or 3 in each of the last 5 races. A higher count indicates consistent strong performances.

wins_last_5_races: This column counts the number of wins (1st place finishes) a driver has achieved in the last 5 races. It checks whether the driver's rank is 1 (indicating a win) in each of the last 5 races

driver_age: This column captures the age of a driver at the time of each race. This feature is calculated by subtracting the driver's date of birth (dob) from the race date and converting the difference into years. This feature provides insight into the driver's age during the race, which can be relevant for analysing performance trends across different age groups.

Fig 5: This figure shows the distribution of ages for race winners, illustrating the relationship between driver age and success in races.



driver_experience: This column represents a driver's experience, calculated as the product of the number of unique years a driver has participated in races (unique on the year column) and their age at the time of each race. This feature gives a measure of both the driver's longevity in the sport and their age, reflecting the potential influence of experience and maturity on performance.

avg_pos_last_5_race: This feature provides an indication of a driver's recent performance trends, with a lower average position suggesting better recent form.

career_wins: This feature provides a running total of the driver's wins over time, reflecting their overall success in the sport.

constructor_wins_last_5_race: This feature highlights the recent success of a constructor, reflecting their performance trend in the most recent races.

Quali_position_impact: This column represents the historical win rate associated with each grid position. It is calculated as the ratio of races won from a specific grid position to the total races started from that position. This feature reflects the advantage or disadvantage of starting from a particular grid position, providing insight into how qualifying performance impacts race outcomes.

driver_circuit_win _rate: This feature provides insight into a driver's performance and familiarity with particular circuits, which can influence their likelihood of success in future races on the same tracks.

VII. Imputing Missing Value

There is some missing data in our dataset especially in the features we created. We will fill those columns with 0 and drop some of the columns which are not required for prediction.

VIII. Feature Scaling and One Hot Encoding

To prepare the training data for the model, we apply standard scaler to the dataset to ensure that all features have a consistent scale. This process helps to evenly distribute the weights across the network, improving the model's learning efficiency and enabling it to converge more quickly to optimal neuron weights. We also perform one hot encoding to convert categorical columns (identified by their data type) into numerical columns, where each unique category is represented as a binary column. This transformation ensures that categorical data can be effectively utilized by the model, which requires numerical inputs.

IX. Architecture

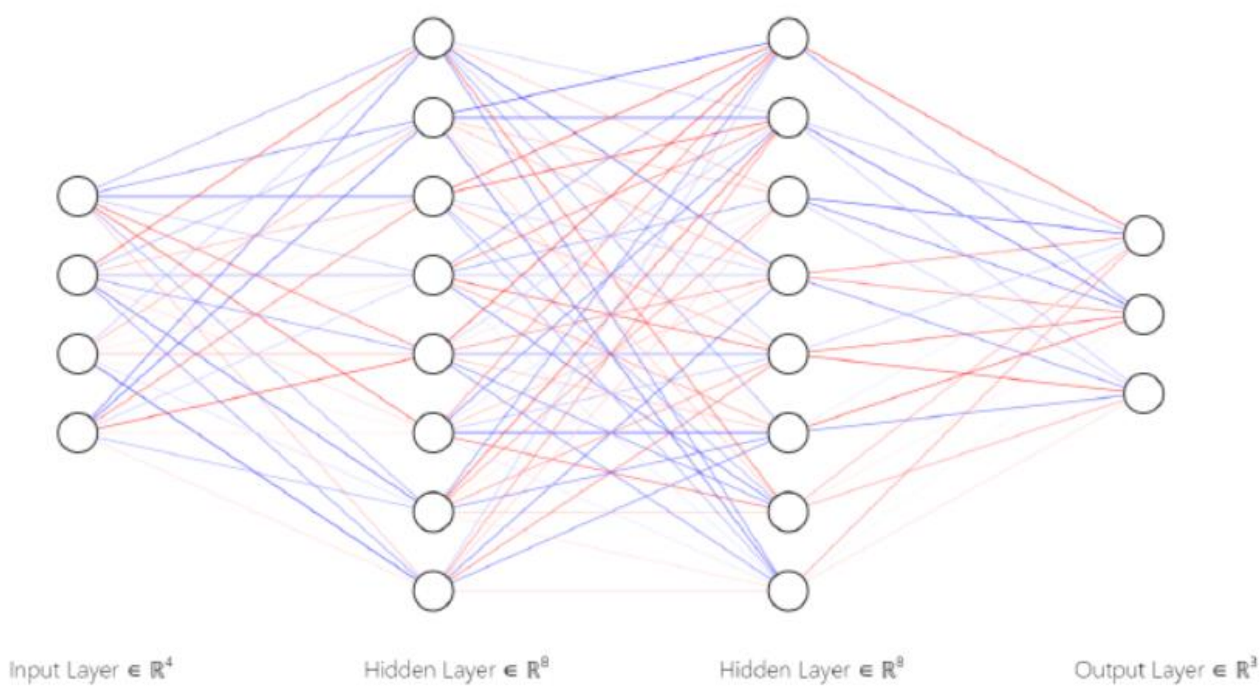
For our F1 race outcome prediction model, we opted for a feedforward neural network, which is well-regarded for its ability to detect and learn complex patterns from diverse data inputs. The structure of the network was fine-tuned using various optimization techniques to ensure the best possible performance.

The model consists of several essential layers:

Input Layer: The input layer consists of neurons corresponding to the features, such as driver statistics, weather conditions, and constructor details. Each feature is represented by a unique neuron, forming the foundation of the model's input.

Hidden Layers: The network includes two hidden layers: the first layer has 25 neurons, and the second layer has 10 neurons. These layers are responsible for processing the relationships within the data, enabling the model to capture complex patterns and improve prediction accuracy

Fig 6: Basic Architecture View of our model with 2 Hidden layers



Output Layer: The final layer contains a single neuron, providing the model's output as a probability indicating whether a specific driver is likely to win the race. This is calculated using a sigmoid activation function, which converts the model's output into a value between 0 and 1, representing the chance of a win

A key feature of this architecture is its full connectivity, meaning that each neuron in each layer is connected to all neurons in the subsequent layer. This design ensures that the information flows seamlessly through the network, allowing the model to make well-informed predictions by understanding the intricate relationships between different features.

This flexible feedforward architecture enables the model to accurately capture the complex dynamics of F1 races, ultimately providing reliable predictions based on various race-related factors.

X. Hyperparameter Tuning

To identify the optimal architecture, we employed a method known as network growing. This technique involves incrementally adding layers to the network and testing different configurations of neurons in each layer until the best-performing structure is found, based on model accuracy. Our initial exploration began by testing a multilayer perceptron with a single hidden layer, experimenting with different numbers of neurons (10, 25, 50, 100). This allowed us to observe how varying the number of neurons in the first layer impacted the model's accuracy.

Next, we systematically introduced additional hidden layers, and for each network, we tested different combinations of neurons across the layers, such as [10,10] or [10,25]. We evaluated the model's performance based on its accuracy. Below table is attached for grid search findings.

Through this process, the best-performing configuration was found, achieving an accuracy of 95.33 %.

Epochs	Neurons_Layer1	Neurons_Layer2	Neurons_Layer3	Accuracy
20	25	10	0	0.953265428
20	100	0	10	0.953265428
50	10	0	50	0.953265428
100	10	0	100	0.953265428
100	10	10	25	0.953265428
100	50	10	10	0.953265428
100	50	10	25	0.953265428
100	100	50	10	0.953265428
20	25	25	0	0.951092796
20	25	0	25	0.95000648
50	10	0	100	0.947290689
20	25	10	10	0.944574899
20	25	50	100	0.944574899
20	50	100	10	0.944574899
20	100	25	100	0.944574899
50	25	50	100	0.944574899
50	50	50	25	0.943488583
100	50	100	0	0.943488583
20	25	10	50	0.942945425
50	10	50	50	0.942945425
20	50	25	100	0.942402267
20	100	25	10	0.942402267
20	10	10	100	0.941859108
20	25	50	50	0.941859108
20	25	100	25	0.941859108
20	50	25	10	0.941859108

The optimal architecture consisted of:

- 20 epochs,
- 25 neurons in the first hidden layer,
- 10 neurons in the second hidden layer, and
- No third hidden layer.
- This configuration delivered the highest accuracy, making it the best model for predicting outcomes.

XI. Model Prediction

To predict the winner of an F1 race using the model, we first select the race using the test dataset (df test), which includes all the drivers participating in the race. The model automatically prepares the feature data for these drivers and splits it into training and validation sets. It uses this training data to build a neural network model based on the best hyperparameters found earlier (such as the number of neurons and epochs).

Once the model is trained, it makes predictions on the validation and test data. The predictions are probabilities representing the likelihood of each driver winning the race. To identify the race winner, the model compares the probabilities for each driver and selects the driver with the highest predicted probability as the winner.

Thus, the model uses the trained neural network to predict the race winner based on the features of the participating drivers, and the driver with the highest probability of winning is predicted as the race winner.

XII. Model Evaluation

After performing a grid search over the training parameters, we found that the lowest accuracy achieved on the validation precision for class 1 being 0.50. The recall for class 0 was 0.99, while for class 1, it was 0.33, showing some imbalance in performance across the two classes.

The classification report for the test set shows an accuracy of 0.98. Class 0 achieved a precision of 0.99 and a recall of 0.99, indicating that the model was very good at identifying class 0. For class 1, precision was 0.76 and recall was 0.73, indicating the model's ability to correctly identify winners was decent, though still imperfect.

The confusion matrix for the validation set showed that the model correctly identified most of the non-winners, with 84 true negatives and 1 false positive. However, there were 2 false negatives and 1 true positive. For the test set, the matrix showed that the model was still highly accurate, with 413 true negatives and 5 false positives, as well as 6 false negatives and 16 true positives.

Overall, the model achieved strong accuracy on both the validation and test sets, demonstrating its ability to predict race outcomes, though with some room for improvement on predicting winners (class 1).

Also, if we have a look at actual race winners from the F1 website and compare them, our model gave correct predictions for 16 races out of 22 races which are held in one season which isn't that bad.

XIII. Conclusion & Future Work

In this project, we presented a model for predicting Formula 1 race winners using a feed-forward artificial neural network (ANN) trained on historical race data. The model achieved strong performance with an accuracy of 97% on the validation set and 97% on the test set.

The results demonstrate that neural networks can successfully identify patterns in historical F1 data, offering a promising tool for predicting race outcomes. However, the model showed slight imbalance in its performance, particularly in predicting class 1, which indicates that there is still room for improvement, especially in the prediction of race winners.

Future work can focus on improving the model's ability to predict class 1 (race winners) with higher precision and recall. This can be achieved by exploring techniques such as class balancing or resampling to minimize effect of the class imbalance observed in the data. More experimentation with more complex architectures, such as deeper networks or recurrent models that capture temporal dependencies in race data, could enhance prediction accuracy.

In conclusion, this project demonstrates the potential of Neural Networks for predicting Formula 1 race outcomes, emphasizing the importance of model architecture, feature selection, and learning algorithms. While the model shows promising results, there is significant room for improvement, such as extending the training period, incorporating advanced feature engineering, and utilizing more computational resources for complex architectures. However, Formula 1 race results are influenced by a variety of dynamic factors, including driver performance, weather, strategy, and unforeseen race events, which makes predicting outcomes based solely on historical data challenging. For more accurate predictions, integrating real-time data and accounting for these evolving factors would be essential to improving the model's reliability and performance.

References

[1] Liu, X., Fotouhi, A. Formula-E race strategy development using artificial neural networks and Monte Carlo tree search. Neural Comput & Applic 32, 15191–15207 (2020).

[2] Davoodi, Elnaz & Khanteymoori, Alireza. (2010). Horse racing prediction using artificial neural networks. link.

[3] Heilmeier, Alexander, Andre Thomaser, Michael Graf and Johannes´ Betz. "Virtual Strategy Engineer: Using Artificial Neural Networks for Making Race Strategy Decisions in Circuit Motorsport." Applied Sciences (2020): n. pag.

[4] Keertish Kumar, M., Preethi, N. (2023). Formula One Race Analysis Using Machine Learning. In: Gunjan, V.K., Zurada, J.M. (eds) Proceedings of 3rd International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications. Lecture Notes in Networks and Systems, vol 540. Springer, Singapore.

[5] Shelke, Priya M., Riddhi Mirajkar, Anurag Pande, Srujan Kale and Yash Paralikar. "F1 Race Winner Predictor." 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA) (2023): 1-4.

[6] P. Shelke, A. Pande, S. Kale, Y. Paralikar and A. Kulkarni," F1 Race Winner Predictor," 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Pune, India, 2023, pp. 1-4, doi: 10.1109/ICCUBEA58933.2023.10392224.

[7] J. Kahn; Neural network prediction of NFL football games; World Wide Web Electronic Publication, 2003 (2003), pp. 9-15

[8] Stoppels, Eloy. "Predicting race results using artificial neural networks." (2017).

[9] FRANSSEN, K. (2021). COMPARISON OF NEURAL NETWORK ARCHITECTURES IN RACE PREDICTION (Doctoral dissertation, tilburg university).

[10] Rana, R., Pandey, D., Mishra, S., Nehra, N., Deshwal, D. and Sangwan, P., 2021, December. Predicting Standings in F1 Sports Driver's Championship using Lasso Penalised Regression. In 2021 International Conference on Industrial Electronics Research and Applications (ICIERA) (pp. 1-5). IEEE.

[11] Williams, J. and Li, Y., 2008, January. A case study using neural networks algorithms: horse racing predictions in Jamaica. In Proceedings of the International Conference on Artificial Intelligence (ICAI 2008).