Enhancement of Formula One Driver Performance by Process Learning

Author: Antoine Alireza Moghaddar
University of Twente
P.O. Box 217, 7500 AE, Enschede
The Netherlands

ABSTRACT,
Throughout the different Formula One seasons, it is seen that younger, less experienced, drivers lack pace and performance compared to their peers. Training these drivers requires time, which, due to the regulations, teams do not have. A leading misconception that affects the level of confidence and learning abilities of the driver, is that the less experienced drivers are expected to perform at the same level as experienced drivers.

This thesis has the aim of analyzing the abilities and performances of both drivers within a Formula One team to redesign and redefine the driver training method. The focus of change and improvement is to provide drivers real-time insights and feedback on their performance during a simulator training session based on the data harvested from both drivers. By using a combination of the fundamental principles of process mining and statistical analysis, data markers are created on the track. Accordingly, an advice marker is created with which the real-time telemetry data is compared. Based on the differences in telemetry, visual feedback is provided to the driver. Throughout the research, this method of training has proven to be promising, drivers showed a significant increase in their overall performance and simultaneously showed an increase in car control and confidence. Nonetheless, due to the limitations, a statistical backbone is missing and more testing needs to be done to guarantee a consistent outcome. For the performance assessment, a methodology process is developed to generalize this training method among other teams and drivers.

Graduation Committee Members:
Dr. F.A. Bukhsh,
MSc. N. Bouali

Pages: 64 (26593)

Keywords
Process Mining, Machine Learning, Statistics, Linear Regression, CRISP-DM, Formula One, Driver Performance, Racing, Simulator

Acknowledgments

Starting my student life at Saxion Hogeschool Enschede, I could only dream of one day finishing my Bachelor degree at a University. The past seven years, although enjoyable, have been a challenging period wherein hard dilemmas had to be decided on. It was a time of growth wherein I became more profound about my qualities and abilities. I am thankful for all the opportunities the University of Twente has offered me and I want to thank my advisors and coordinators for their patience. While working towards the finalization of my second bachelor in Computer Science, also at the University of Twente, I plan to continue my career in Automotive engineering in combination with Data Science and Artificial Intelligence.

While the restrictions of the Covid-19 pandemic still hold their grip on the availability of resources, I am thankful that throughout this project, help was available from the right people. I am grateful for the people that devoted their time to make this project possible, as, without it, this project would not have succeeded. I want to send my appreciation to all the teachers and professionals whom I worked closely with within the run-up towards my graduation project and throughout the project itself. In particular, I want to thank my supervisors, dr. F.A. Bukhsh, (Faiza) and MSc N. Bouali, (Nacir) without whom this project would not have direction. Their supervision kept the project feasible and realistic while obtaining the maximum amount of results. I want to thank dr. G.W.J Bruinsma, (Guido) without whom, this project would not be technically possible. I want to thank Guido for the determination to take this project to a next level, his friendliness, and his hospitality to provide me with an office with all the required resources, and his faith that I would do more than only playing games at the times I was near the racing simulator. I want to thank dr. A. Kamilaris, (Andreas), for having the patience to deal with the many excel-sheets and Google Forms, for his support and knowledge when in need, for his excellent way of thinking and reasoning, and above all, for being my friend. Last but not least, I want to send my appreciation and thanks to Armein M. Dul, who showed endless support during my graduation project.

List of Tables

List of Figures

Table of Contents

Chapter 1: Introduction

Motorsports has been a large part of the entertainment industry over the last century. While there is no clear beginning of when automotive racing started, in the early 1890s the first combustion-engined cars raced over a distance of 126 kilometers from Paris to the city of Rouen. This event is considered the birth of automotive motor sports. In this day and age, 126 years later, this same industry has grown into a multibillion-dollar industry. The pinnacles of the current motorized sports industry are, in the European region, Formula One, and, in the American region, NASCAR. In addition to their contribution to the entertainment industry, Formula One and NASCAR have made many contributions to research, development, and innovation.

Formula 1 is an industry that heavily relies on data and data predictions. The more reliable these data predictions are, the better the strategies become. Hence, it is safe to state that data forms the underlying basis of each aspect within the championships. However, while reliable data is an important factor, it is not the only crucial factor that plays a role within the different teams. Looking into the teams internally, one can define multiple roles among the team members that all try to contribute to the performance of one specific team member, the driver. In the current 2021 season of the Formula One Championship, there are a total of ten teams, with every two drivers, making a total of 20 (“twenty”) drivers. While many teams perform equally well with both drivers, there has been, more than once, some friction between two different drivers within the same team in previous seasons of the championship. In the 2019 season, the formula one team Red Bull Racing swapped their secondary driver, Pierre Gasly, mid-season due to under-average performance compared to the first driver, Max Verstappen, in favor of Formula Two driver Alexander Albon. While swapping drivers mid-season is very uncommon, due to the signed annual contracts, it made a statement about driver performance and abilities being considered more important than the data predictions. Red Bull Racing is not the only team to struggle with equalizing the performance of both drivers. As each driver has a different career path and hence a unique way of driving, it is hard to match another driver’s abilities to enhance the team performance. Training a driver based on the performance of the other driver, might equalize the differences within a team and help the team forward.

Within this thesis, the focus will be on improving and enhancing the skills of both drivers within a single team. Considering this goal of the thesis, it is important to research the state-of-the-art. By studying the state of the art, an overview of what already exists can be constructed. Building upon this state-of-the-art, opportunities will show with regards to what can be done to improve this aspect of the drivers’ performance. Due to confidentiality, it is not possible to get insights into the exact training of the different drivers. Hence, it is assumed that both drivers are trained individually, based on their skills and abilities. Looking into the science concerning games built upon the principles of player ranking, we find many pieces of research that cover the training of players. In a user study performed by Bugeja, Spina, & Buhagiar (2017), the effect of using simulators for training novice drivers was tested using specified test groups. Before the study began, the general notice was made that the skills required to become a good motorsports driver are generally learned through practice. The telemetry-based feedback system (TeAR) was developed to provide direct insights in auditory suggestions to drivers underlining the driver’s mistakes they are currently making. This user study revealed that real-time data for improving the performance of drivers during a simulator race is directly effective within a short period without repetition. In another study regarding the training of the military and the principles of learning, Gagne (1962) extensively discussed the different ways specialized domain experts can be trained to be improved. This study took into account all details required to be considered when wanting to train people. While these principles are built upon the current training and improvements of soldiers in the military, the principles cover a broad range of target groups and hence can be applied to other domains. Within the next chapter about the background research, this state-of-the-art is more extensively discussed.

By usage of data mining principles on the performances and processes of two drivers within this same team, statistical techniques will be implemented to predict the performance, visualize the differences between the drivers, with regards to track position and performance of processes, and, finally, guide each driver in real-time towards a better understanding of the actions required. Due to the limited amount of public research on the performances of drivers within motorsports, additional research can help to improve the industry. When trying to identify the gaps using events in the past, it is possible to sketch a point of improvement with regards to this current situation. Referring back to the example given earlier in the introduction; the example of Alexander Albon replacing Pierre Gasly during the 2019's championship was stated. While Albon had raced at Scuderia Toro Rosso throughout the first half of 2019 and hence had experience driving in Formula One, driving alongside Max Verstappen with the high standards of the Red Bull Racing team pressurizes the new driver to perform at best. Being aware of the driver style of Max Verstappen to use as a reference for training and guidance purposes, can be beneficial in this situation.

1.1 Problem Statement

When adjusting the training of drivers, many aspects of the driver's past and presence need to be taken into consideration. As each driver has his unique way of driving, limiting this driving style to the performance of the other driver might dispatch the abilities of the driver. Hence it is important to not want to change the driver, but rather improve the driver on his abilities using guidelines set up by actual, representative, data.

The research problem is, therefore, defined as the lack of pace and performance between drivers. This lack of performance can be caused by either a lack of confidence or lack of knowledge on the abilities of the car or themselves, irrespective of their previous experiences and training.

1.2 Research Questions

Based on the research problem, the following research question has been formulated:

- How can we enhance the overall driver performance within a team by use of the fundamental principles of Process mining and Statistical analysis?
  - **SUB 1:** To what extent is it possible to recreate an artificial trackline built upon the basis of the highest performances throughout the track?
  - **SUB 2:** How can this artificial trackline be translated into terms of required telemetry changes to guide towards this trackline?
  - **SUB 3:** How can these telemetry changes be communicated to the driver in the most effective manner?

1.3 Research Approach & Thesis Structure

The approach used for this research is by applying CRISP-DM (Cross Industry Standard Process for Data Mining) This approach is used to answer the research questions as denoted in section 1.2. CRISP-DM is a generic model that is defined by six steps for development within the field of data mining. This cycle of steps is depicted in Figure 1.a. Nonetheless, a deviation has been made from the actual CRISP-DM cycle. This project does not meet the requirements to have a section for deployment. Hence, this phase in the cycle is ignored.

Fig. 1.a: The CRISP-DM Development Cycle (Jensen, 2012)

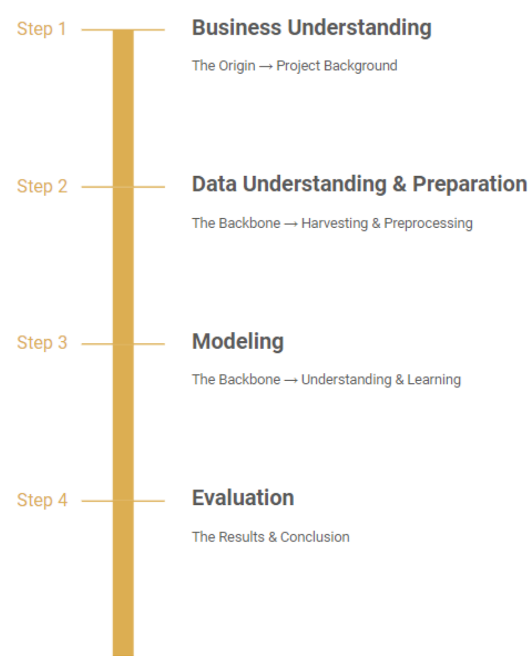

Fig 1.b: The Implementation of CRISP-DM in the context of this thesis

Business understanding

Within this thesis, chapter one and chapter two are devoted to laying out the backbone of the research project. Together these two chapters form the business understanding phase of the project. Additionally, chapter four classifies these business understandings in terms of a division of tasks. Moreover, this chapter denotes the requirements for a minimum valuable product (MVP).

Data Understanding & Preparation

The Data understanding phase, altogether with the data preparation phase, is discussed in the methodology section, chapter three. Within this chapter, an analysis is done on how to interpret the data and how to form the data in a way that it is valuable for the later stages of the development process.

Modeling

Continuing within the cycle, the modeling phase is the part wherein data is translated into feedback. This translation and its core principles are denoted and discussed to an extent in chapter five. Within this section, the realization of the feedback is discussed as well.

Evaluation

The last phase applied within this thesis is the evaluation. Chapter six and chapter seven are devoted to the analysis and evaluation of the project along with an evaluation of the corresponding results and outcomes. After a conclusion is drawn and the final remarks for the continuation of the development are mentioned.

To go in more in-depth on the outline of this thesis: First I will explore the state-of-the-art where I discuss background information, similar research, and previous studies regarding the training of motorsports drivers. Within this section, I will also elaborate on the inclusion and exclusion requirements. Then, to prepare the experiments required for answering the main research question, I will go over the methodology used to obtain, manage and process reliable data. Afterward, I will touch upon the realization phase wherein I define the multiple ways of gathering, analyzing, visualizing, and communicating data towards the target group, the drivers of motorsports. Subsequently, in the analysis section, I will define the main experiment, analyze the collected data, and briefly evaluate the outcomes of the experiments. Lastly, we continue the discussion and the limitations, wherein an attempt is done to answer the sub-questions as defined in section 1.2. Lastly, I will discuss the conclusion and future work, wherein I explain the possibilities with the application created, conclude the research and answer the main research questions of this research.

To visualize this process and its corresponding phases, according to the CRISP-DM development cycle and the requirements of the project, a diagram is made to represent the project plan. This visualization is shown in figure 2.

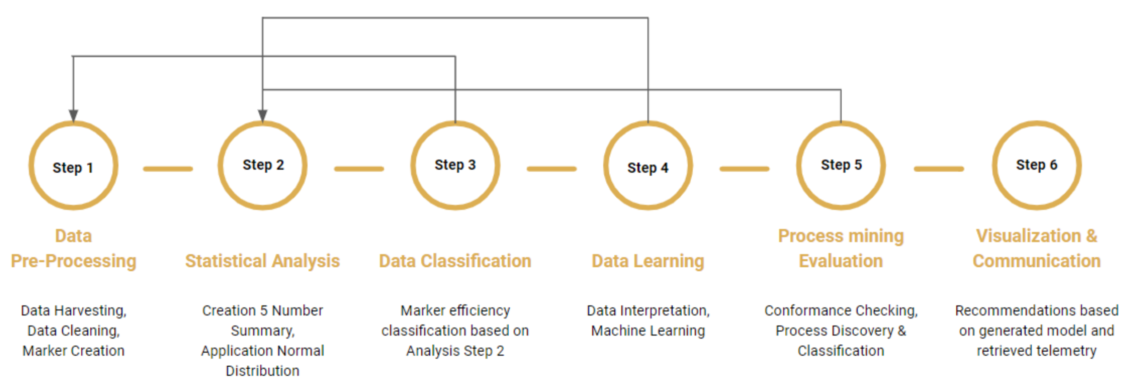

Fig. 2: The flow of operations concerning project plan

Chapter 2: Background Research

This section elaborates on intended methods to find background research correlated to the topic. Firstly, the approach on how to gather articles and information is discussed with regards to the scope of the project, and the direction of the research. Within this section, the corresponding inclusion and exclusion criteria are stated. Secondly, the approach on how articles have been narrowed down to ensure a high quality of information related to the corresponding topics discussed in the literature review. Thirdly, an overview of existing environments for training motorsports is provided altogether with a section describing the use of simulators when training professionals. Within the fourth section, the process mining principles with their corresponding challenges will be discussed. Within this section, the potential algorithms relevant for this thesis are discussed as well. The fifth section highlights the inclusion of virtual systems as feedback systems and addresses and elaborates on the potential feedback systems. Lastly, a section on gaps within the field of research is provided as a base for the ideation phase.

2.1 Approach

The approach used for this section of the thesis is the approach recommended by a study on systematic literature reviews conducted by Wienen, Bukhsh, Vriezekol, and Wieringa (2017). Hence, within the context of this exploration phase, it is important to set the scope as wide as possible. For this research, this means covering domains deep inside and far outside of the field of motorsports. The internal domain used in this chapter covers topics as varied as Formula One, with its corresponding training teams, Porsche Supercup, NASCAR, and other related sports. The external domain used in this chapter covers topics as varied as Education, Psychology, and Innovation & Development.

The approach, as proposed by Wienen et al. (2017), for answering the research question and subquestions is as follows:

1) Start by defining the inclusion and exclusion criteria,
2) Define the query to interrogate the different databases at disposal,
3) Select the relevant databases and start querying,
4) Select the articles pertaining to the research topic.2.1.1 Inclusion and Exclusion Criteria

2.1.1.1 Database I/E Criteria

Include/Exclude	condition
Include	Only databases pertinent to our research
Include	Databases that have either articles or numerical and factual data
Include	Databases related to other sports wherein overall team performance is portrayed against the influence of team players
Include	Databases in any language

Table 1: Database I/E criteria

The University of Twente has access to 108 databases.

2.1.1.2 Article I/E Criteria

Include/Exclude	condition
Include	Articles on different aspects of learning.
Include	Articles written in Dutch
Include	Articles written in English
Include	Articles provided by motorsports team
Include	Articles on “Training by the usage of simulations”
Include	Non-Scientific articles on specifications of data
Include	Articles on Aviation Simulation Training
Exclude	Non-Scientific articles on facts designed with the purpose of entertainment
Exclude	Non-Scientific articles on comparisons between formula one teams and/or drivers

Table 2: Article I/E Criteria

2.1.1.3 Source I/E Criteria

Include/Exclude	condition
Include	Interviews with motorsports domain professionals
Include	Interviews with virtual reality domain professionals
Include	Interviews with simulation-based training domain professionals

Table 3: Source I/E Criteria

2.1.2 Source selection

After the first selection round, based on the search terms included in Appendix 1, a set of 1430 relevant articles have been selected that match the inclusion and exclusion criteria. Based on the abstract, the introduction, and the conclusion of these papers, the list of total articles has been narrowed down to approximately 175 relevant articles. These articles have been addressed in the following literature review on Simulators, Process mining, and Feedback systems.

2.3 Training by usage Simulators

Research and coaching on training behavior have drastically been improved by the introduction of motion simulators (Espié, Gauriat, & Duraz, 2005). Motion simulation started with flight simulators with the main purpose to visualize and conceptualize the outcomes of different situations during a flight (Slob, 2008). Over the years, flight simulators within the field of aviation have broadly been developed and implemented to not only visualize the possible situations that may occur, but also to train potential pilots on situational awareness, plane handling, to perform maneuvers, and to perform procedures with a level of expertise (Socha et al., 2016). In a meta-analysis of flight simulation research conducted by Hays, Jacobs, Prince, and Salas (1992), the findings showed that the use of the combination of simulators with basic aircraft training, lead to an improvement in training performance compared to traditional training using only aircraft training procedures. Translating these principles into the field of automotive technologies and motorsports opens opportunities for advanced training methods.

Within this section, the advantages and disadvantages of using driving simulators within the automotive industry and motorsports with the purpose of training are discussed, substantiated by relevant literature. This section is followed up in chapter 3, wherein a suitable configuration of a driving simulator for this thesis is discussed.

2.3.1 The grounds for using motion systems and driving simulators as training techniques

The usage of simulators can be applied in various fields of training. According to de Winter, van Leeuwen, and Happee (2012), driving simulators offer various advantages, compared to the implementation of the training within the real environment. As de Winter et al. (2012) mentioned, the first, and most important advantage of using simulators is the *possibility of encountering dangerous driving conditions without being physically at risk.* This offers the learning driver to explore the positive or negative consequences of actions without leaving the driver vulnerable to potential harms (Slob, 2008). Secondly, *controllability* of conditions, the *reproducibility* of scenarios, and the *standardization* of ground rules built upon tests for the next line of advantages for using driving and motion simulators. Combining these parameters in a dynamic scenario provides opportunities for controlling potential real-life scenarios that may happen during a race (Wassink et al. 2006). Adjusting the parameters of the virtual scenario can, according to Wassink et al (2006), enhance the reaction of the learning driver by standardizing procedures, aiming at minimizing the impact of the change within the environment. These changes can differ per configuration. In research conducted by Slob (2008), the effects and differences in the various configurations are discussed concerning their degree of freedom (DoF), the visual element, and the feedback element. Within this background research, the conclusion defined a set of criteria that need to be taken into consideration when building the simulator. Within chapter 3, these differences and effects of each configuration are discussed. Thirdly, de Winter et al. (2012) described the accuracy and *ease of data collection* as another advantage, contributing to the reliability of the provided feedback, offering better opportunities for providing better feedback and instructions. Based on all the aforementioned advantages, Slob (2008) mentioned one other advantage, describing the potential *reduction of costs* compared to the alternative (real) training solution.

Nevertheless, within the context of training motorsports drivers, a set of disadvantages can be identified. According to de Winter et al. (2012), low-fidelity simulators may invoke unrealistic environments and therefore yield unrealistic driving behavior, leading toward invalidating research outcomes. Another disadvantage is that the simulator needs to be effectively built to accurately mimic the expected behavior within scenarios and will hence require many hours of building, calibration, and testing (Balcerzak, & Kostu, 1988).

In short, learning from the applications and impact of motion simulators within the field of aviation training, it can be stated that converting these principles into the field of motorsports can positively affect the training of motorsport drivers. When configured correctly, a high-fidelity driving simulator may replicate and build upon the scenarios that the driver needs assistance in. Due to the accuracy required for training to be effective, a low fidelity motion simulator will not have enough impact to change driver behavior.

2.4 Process mining

An underlying basis for improvement and enhancement of skills is the analysis of data. Van der Aalst et al. (2012) have defined process mining as the technique to extract knowledge from event logs commonly available in information systems. Hence, when willing to improve on the skills based on a set of procedural processes, process mining principles can be used to build on. In a different study, Van der Aalst (2012) has introduced three types of leading process mining categories: discovery, conformance, and enhancement, each with their corresponding field of application. Analyzing the processes from a log can be done using the discovery techniques (Rozinat, Alves de Medeiros, Günther, Weijters, & van der Aalst, 2007). This analysis results in a process model based on an event log. In addition to the discovery technique, conformance techniques are used to validate the relevancy of the event log relative to the real-world situation. Conformance is therefore used to cancel and/or remove misconceptions within the process data set. The third and last technique introduced by Van der Aalst (2012), is the enhancement technique. This technique is used as an addition to existing process models with the purpose to improve, extend or clarify conceptual statements on data (Bogarín, Cerezo, & Romero, 2017).

2.4.1 Techniques

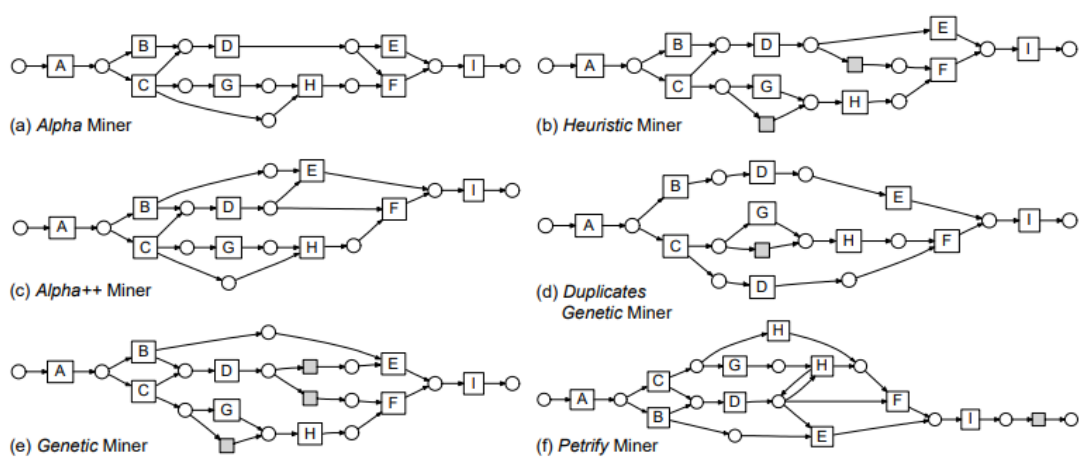

Fig. 3. Process models generated by different process mining algorithms based on the same log file (Rozinat et al. 2007).

Within the process mining categories introduced in the previous section, different algorithms can be defined: The Alpha Miner, the Heuristic Miner, the Alpha++ Miner, the Duplicates Genetics Miner, and the Genetic Miner. Each of these algorithms has its corresponding characteristics, (dis)advantages, and purposes.

The Alpha-miner [a] (Van der Aalst, Weijters, and Maruster, 2004) and the Alpha++-miner [c] (Wen, Wang, & Sun, 2006) are discovery techniques that are aimed at discovering and reconstructing causal relationships from a set of events. As described in the study conducted by Wen et al. (2006), both these models are built upon the notion of mining non-free-choice constructs, where the process models are represented by Petri nets. Within this same study, Wen et al. (2006) have defined two causal dependencies between tasks that are of importance: *Explicit* and *Implicit* causal dependencies. As implicit dependencies are difficult to map, the focus within process mining lies on explicit causal relations. While the Alpha-type algorithms are basic algorithms yielding proper and adequate results, according to Van der Aalst et al. (2012), and Weerapong, Porouhan, and Premchaiswadi (2012), Alpha-type mining algorithms are vulnerable to noise, incompleteness, and redundancy with the risk of missing the unreachable loops. Therefore, Alpha-type algorithms are bound to certain limitations and thus are considered an impractical approach.

A technique built upon the basic principles of Alpha-type mining is the Heuristic mining [b] algorithm. Instead of producing Petri Nets, Heuristic mining focuses on utilizing the frequency of events within a certain event log, improving the efficiency of the log by discarding unwanted, infrequent, behavior (Weijters, & Van der Aalst, 2003). However, as Van der Aalst (2012) described; due to the need for larger event sets to ensure accuracy, this model can still produce inaccurate results.

According to Alves De Medeiros (2006), Genetic Mining Algorithms are a type of search technique that mimics the process of evolution through a search space to find an end node. Every point in space is hence called an individual, contributing to the overall finite set of individuals called a population. The quality of an individual is hence determined by fitness measure, correlating directly to the prediction of the optimal path through which the process needs to proceed. Candidate process models can be created to support the optimal process model.

Conformance checking is the second category of the process mining principles, yielding the technique that is used to compare event logs or resulting processes relative to the target model of an identical process. Hence the aim of conformance checking is, according to Leemans, Fahland, & Van der Aalst (2018), to verify a process model against reality.

Within the principles of Log-model conformance checking, the reality is assumed to be represented by an event log. Hence, to compare a process model to an event log, qualities and characteristics need to be defined to base the comparison on (Leemans et al.,2018). Examples of these characteristics are Fitness, log-precision, generalization, and simplicity. Algorithms for measuring these characteristics include token-based replay (Rozinat, & Van der Aalst, 2008), and Alignments (Adriansyah, Munoz-Gama, Carmona, van Dongen, & Van der Aalst, 2012). As Bos (2021) discussed in a thesis, the downside of using token-based replay on problematic models is the fitness could peak, causing misconceptions.

2.4.2 Challenges

While process mining gives elaborate insights into situations based on captured data, Van der Aalst (2012) has introduced a set of two challenges that might occur during process mining. As Bos (2021) mentioned, grounded on the study conducted by Van der Aalst (2012), proper, complete, event logs are essential for the discovery phase. Having gaps in event logs, producing gaps in the data, leads towards potentially skewed analysis results. Secondly, noise in these event logs can sketch different views of the situation and therefore invalidate the results, and lower the reliability of the process mining process. To ensure high reliability within the process mining results, noise needs to be filtered out to clean up the event log.

2.5 Feedback

Having in mind the ever-changing nature of technology and development, automotive technology has rapidly been improving over the past decades (Gott & Hurter, 1981). Looking at the roots of these improvements within technology, many overlapping areas are found of which one entails motorsports. While motorsports are mostly seen as a source of entertainment to many, the field has introduced outstanding technologies that have changed the course of the automotive industry drastically (Sano, 2014).

Although the automotive industry has been adopting many of the technologies developed within motorsports, overarching organizations like the FIA and IndyCar have been pushing on reducing the number of developments by introducing limits and restrictions to the number of new technologies introduced (federation Internationale de l'automobile [FIA], 2021). Hence, training the drivers to be able to properly operate the car, requires some guidelines and knowledge.

When willing to improve driver performance while remaining road safety, proper feedback is a major topic that needs to be addressed. During an investigation on proper feedback, Koo et al. (2015) discovered that improper feedback confuses leads to poorer driving performance and hence less safe circumstances for the driver as well as surrounding individuals. Among the various types of feedback researched, corrective feedback and the various fields and applications of this type of feedback are deemed most important in the field of automotive engineering (van Houten, & Nau, 1983).

This section, there will be first explained the essence of receiving proper feedback when driving. Afterward, three different approaches to giving feedback are discussed. These different approaches are substantiated with relevant literature to pinpoint the approach when willing to interface this type of feedback and a brief comparison between the corresponding type with the other types of feedback. At last, an overview on the ideal approach in the context of this literature review on improving driver performance is discussed to provide a starting point for incorporating feedback systems to support the driver in his/her personal development.

2.5.1 The essence of proper corrective feedback when driving

Feedback aimed to move students from tasks to processing and from processing to regulating is the most effective feedback (Hattie & Timperley, 2007). Hence, feedback in the context of development is focused on improving the abilities of an individual based on applying improvements rooted in the experiences of others. The pitfall of feedback is that receiving too much, and possibly irrelevant feedback, may, within a certain level, cause detraction from performance and therefore lead to a counterproductive side effect (Hattie & Timperley, 2007). However, considering the importance of feedback on the learning process of students, regardless of the field in which the learning process is taking place, it is crucial to provide proper, concise feedback (Nelson & Schunn, 2009). Receiving proper and relevant feedback on performance contributes to the overall experience of the driver. First of all, the implementation of the feedback system used to communicate the feedback to the driver influences interpretation of the given feedback (Voelkel, & Mello, 2015). In a study on the advantages of electronic audio feedback, Lunt and Curran (2009), discovered that audio feedback was generally better understood compared to written feedback. Nevertheless, while audio feedback enhances the overall learning curve, varying between the types of feedback is of high importance. Different subjects require a different level of understanding. The combination of the usage of different types of feedback enhances the overall learning quality (Nelson, & Schunn, 2009).

Additionally, the customization of feedback according to the needs of the driver amplifies the strengths and weaknesses of this specific driver within certain situations. According to Feng and Donmez (2013), driver characteristics are good predictors of the type and severity of exhibited risky driving behavior when constructing systems to give proper corrective feedback. Not only are the driver characteristics important when constructing personalized feedback, taking into account the acceptance and the preferred type of feedback plays an important role. The visualization and presentation of the corresponding feedback determine whether or not the driver is going to open up to accept and embrace the feedback (Anseel, & Lievens, 2009).

In summary, when willing to improve the performance of a driver by modifying the training techniques, it is important to take into account the effect and interpretation of feedback on their learning curve. When constructing an environment where drivers have to follow guidelines, the visualization and communication of the corresponding feedback to the current situation, need to be tailored to the preferences and best abilities of the driver in question.

2.5.2 Acoustic Feedback

The first approach that could be used to provide feedback to a driver is by using acoustic sources. Within consumer cars, audio feedback, or “acoustic feedback”, is largely applied, e.g within the navigation, to monitor telemetry information or to perform hands-free tasks (Pakkanen, Raisamo, & Surakka, 2014).

In a study on the effectiveness of acoustic feedback compared to written feedback among students, Voelkel and Mello (2015) concluded that audio feedback is generally better feedback in terms of student experiences. During this experiment, students mentioned a higher level of understanding and involvement in the feedback compared to receiving written feedback. While this experiment conducted by Voelkel and Mello (2015) is based on a different field of application, this same conclusion is drawn by Lunt and Curran (2009) after their study on the advantages of electronic audio feedback in cars. Hence, it is possible to associate the effectiveness of audio feedback with the application of acoustic feedback within cars. Nevertheless, while acoustic feedback is generally well-received, within the context of an automobile environment, driving a car requires physical and mental attention. Therefore, providing acoustic feedback is possible, but only within certain boundaries. Exceeding these boundaries can result in distracted driving and lower the driver’s driving performance while increasing the safety risks (Pakkannen et al., 2014).

In short, acoustic feedback is a very suitable option when it comes to providing feedback on basic elements within a car. This increases the level of vigilance of the driver while it provides a way of feedback without having to lose sight of the road.

2.5.3 Optical Feedback

As opposed to receiving feedback using acoustics, feedback can also be communicated using visualizations and imagery. While it is recommended to use acoustic feedback only to provide basic feedback to prevent distracted driving, visual feedback can be used in many forms to simplify the data as much as possible while remaining the message clear and understandable.

According to two independent studies on the detailed effect of visual feedback conducted by Adams, Gopher, and Lintern (1977) and Hoppe, Sadakata, and Desain (2006), visual feedback contributes to the general development of motor learning, leading to a better understanding of the situation and hence increasing the likelihood of interpreting the circumstance faster as well as with more reliability. By using visual feedback in combination with gamification of data for training drivers, it is possible to gain more detailed insight into the statistics of the current abilities of the driver and improve the general learning experience of the driver. Also, using computer-generated visual feedback opens opportunities for creating and preparing for scenarios that could happen with a statistical change of 1%. Leaving less room for unprepared situations (Hoppe et al., 2006).

In the study conducted by Hoppe et al. (2006), a training environment is created to provide insights on the performances of singers with the purpose to train them. Within this study, the effect of certain computer tools has been researched on quantitative and qualitative feedback. Although the real-time visualization application is used in a different field of study, the effectiveness of real-time visualization leading to the improvements in the performance of the training of the subject in question can be connected to the field of motorsports. By applying HCI design principles Kumar and Kim (2005), made use of real-time visualizations of data to redesign the dashboard of an automobile to address the problem of speeding. Using the redesigned dashboard, it is possible to convey the effect of visualizations to the field of motorsports. Having in mind the main purpose of the mission of the study conducted by Hoppe et al. (2006), the real-time visualization tools are used as a separate layer of providing effective additional information and assistance when performing a task (Eberhard, 2021).

In essence, when kept simple and understandable, providing visual feedback on a driver's performance can effectively improve the driver's learning curve. A better understanding of the situation can be ensured due to interactivity and hence the judgment of handling situations is refined for the better.

2.5.4 Feedback based on Physical signaling

The last approach to provide feedback to individuals is by the use of physical signaling. Within the application of performance improvements in the field of automotive technologies, force feedback is a largely applied mechanism to simulate the effects of downforce on a driver working on a simulator. In an unscientific discussion conducted by Baxter (2020) on the impact of Force Feedback on the driving experience, it is discussed that using force feedback, mimics the real-world environment in which drivers have to drive. While it is stated by Baxter (2020) that for training purposes, force feedback is not required, however, having a resisting force, shaping the real-world situation, can prepare the driver for finding a balance between applied force and corresponding effect. In an experiment on the ability to learn using different learning techniques on preschool children, the effect of the principle "learning by example" conducted by Brown and Kane (1998) showed an increase in learning abilities when the students were being forced into acting in a predefined manner using active feedback. Hence, by modifying the counter-force applied according to the generated feedback, a driver can be led into the desired behavior

Another approach for providing physical feedback is by making use of haptic feedback. Haptic feedback is the opportunity of using human senses to interact with the environment (Jafari, Adams, & Tavakoli, 2016). This interaction can either be done in a Virtual environment or by using simple devices to stimulate and recreate haptic environments. According to Crespo and Reinkensmeyer (2010), haptic guidance improves the performance of a task by enforcing a desired pattern of kinematics. This guidance can be provided in two types, "*guidance-as-needed*" and "*fixed guidance*", both types allow users to improve their abilities by experiencing large errors getting corrected by guidance. In an experiment on the application of haptic feedback in an automobile environment conducted by Väänänen-Vainio-Mattila et al., (2014), the qualitative showed increased support for road safety and social communication. Although the user had a better experience due to the provided haptic feedback, the interpretation and learning curve of handling a haptic feedback system is hard.

To summarize, haptic feedback provides a new dimension of receiving feedback. The principle encourages the driver to develop by trial-and-error, introducing the driver to scenarios that need to be handled while providing guidance based on activity. The downside of haptic feedback is that it is hard to implement and hard to interpret.

Concluding from the literature review on feedback systems, it is safe to say that there is no one "ideal" feedback mechanism to cover a problem statement. Within the general problem statement, distinctions need to be made among the partitions of the problem statement. Observing the conditions per partition can give insights into what feedback system to use to provide relevant and concise feedback.

Within the context of automotive engineering, it is seen that audio feedback can be used, however, the implementations and applications of audio within this field are a limiting factor (Pakkannen et al., 2014). Safety issues can arise when an overload of audio feedback is used, distracting or frustrating the driver. Hence, to avoid this issue, one can make use of optical or physical feedback sources. This approach gives opportunities to redesign not only the way data is communicated but also the way the data is interpreted by the driver without causing too many distractions. Reinventing the learning environment of the driver, by exposing him to predictive visualizations and haptic feedback, may help the driver to be forced into learning desired behavior while retaining his/her core qualities (Nelson, & Schunn, 2009). The performance improvements, therefore, are rooted in the notion of embracing feedback to improve performance.

2.6 Research Gaps and opportunities

As stated in the introduction of this thesis, the amount of public research on the topic of improving the performance of motorsport drivers is limited due to the confidentiality of the corresponding teams. Throughout this background research, the goal was to find out what there is known with regards to the technologies and applications that we could use. Due to the limitations, the only possibility to get sufficient insights on how principles can be applied to improve the performance of drivers was by looking at other areas of innovation, where these principles have already been integrated into daily practice.

Consequently, the research gap that could be identified is on how integrating process mining techniques in simulators can be used as a basic tool for understanding the actions of drivers. Using this understanding, a new feedback system can be developed, built upon the result of the processes of different drivers. Hence, these processes provide opportunities for explorative research on characteristics and process differences between drivers that impact the performance of this certain driver. Addressing these differences through visualizing the gains and losses as a consequence of the driver's action per sector type, lays the foundation for a new way of training, supported by machine learning principles.

Chapter 3: Methodology

Exploiting the training of athletes within this level of expertise requires accurate data and reliable background information. Without a proper background of what the athlete requires, the training might miss the major point of improvement and hence lead to poor, unexpected results. Within this chapter, the different possibilities for mimicking the environment, data harvesting, data analysis, and visualization are explored to create the backbone of this project and therefore the backbone of the training.

3.1 Implementations of simulators

Simulator configurations differ in many ways with each their corresponding field of use. When focusing on the training of drivers of one-seater cars, the configurations are endless, with the only difference being the reliability and accuracy of the input. The overall configuration consists of a set of elements that one would find in most everyday cars, e.g, a driving wheel, a wheelbase, a shifting system, and a set of pedals for regulating the gas flow, the brake ratio, and the clutch. Additionally, as the environment is being simulated in a graphical interface, no motion physics is taken into account when working with the simulator. An option of enhancing the realism of the simulation, is, hence, to include a motion system with a certain degree of freedom to translate the simulator motion physics into real-world motion physics.

3.1.1 Wheels & Wheelbases

Steering wheels are the main interface of the driver. On the current market, there are a set of different steering wheels available for sim racing. While the functionality and type of steering wheels are important, the choice depends on the wheelbase and the purpose of this steering wheel. Wheelbases are categorized into three types: Belt-Driven, Gear-Driven, and Direct-Driven, which each have their characteristics. When looking at wheelbases for scientific simulation, two aspects are important; Force Feedback Realism, and Steering precision.

Gear/Belt Driven wheelbases

Within a guide on sim racing wheelbases, Mjolnir (2022) described the Gear-driven and Belt driven wheelbases. Mjolnir (2022) mentioned that both types of wheelbases work on a similar principle: A motor is attached to a series of gears and/or a belt. These gears/ the belt is connected to the wheel rim. The motor is the element generating the Force Feedback which is amplified with a factor defined by either the belt or the gears. The disadvantage of using either of these principles is that the gears and the belt absorb a part of the force feedback, meaning that less feedback is provided to the driver. Additionally, the efficiency and accuracy are reduced due to belt friction and additional pulleys or the gears have a risk of jumping when too much force is applied, producing torque spikes.

Fig. 4.: Gear driven wheelbase (Mjolnir, 2022)

Direct Drive

As opposed to connecting the wheel rim to gear or a belt, the directly driven wheelbases connect the wheel rim directly to the motor shaft. As no amplification is implemented within directly driven wheelbases, the motor must be significantly larger to produce a sufficient amount of torque to create this effect of Force Feedback. This means that, unlike in Belt and/or Gear-driven wheelbases, the Force feedback and torque are not lost in the connections, but rather sent directly into the hands of the user. This leads to higher torque, stronger force feedback with less loss, and due to the direct connection between wheelbase and wheel rim, a higher frequency of data transmission. (Mjolnir, 2022)

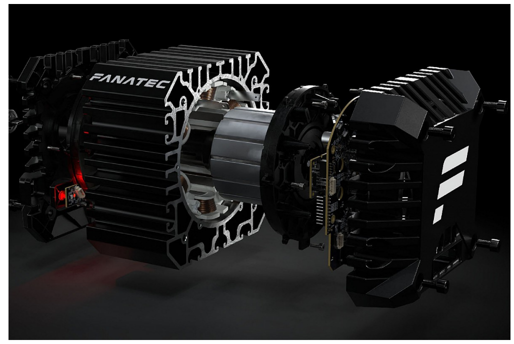

Fig.5.: Direct Driven Wheelbase

Steering Wheels

Consequently, a broad set of different steering wheels is available on the market. While there is little to no usage difference between the types of steering wheels, the functionality of each type of steering wheel is tailored to the purpose of its use. The different steering wheel types have been categorized into three categories: Rally, Formula, and Classics.

- **Classics:** Steering wheels mimicking the earlier motorsport wheels with no additional functionality included above the required to be considered a game controller. These steering wheels are simple in use and have no further purpose than steering. See Figure 6.1 for a visual representation of Classic wheels.
- **Rally:** Rally wheels are optimized for SVGs that include rally racing and/or drifting. Due to the build of these steering wheels, there is a defined maximum load of torque they can handle before fracturing appears. The steering wheel often looks like conventional wheels that can be found in cars and have a small set of functionality included. The aim of the steering wheel is that the driver can use this wheel in combination with a handbrake and/or a conventional shifter, where the driver is required to drive with only one hand on the wheel. Hence, the design allows the driver to fully rotate the wheel in any direction with one hand only. See Figure 6.2 for a visual representation of Rally wheels
- **Formula:** This type of wheel is focused on the one-seater cars where drivers are bound to keep their hands on the steering wheel at all times. Hence, the wheel has a broad set of functionality. The build of the steering wheel has to be resilient against high loads of torque and force feedback. See Figure 6.3 for a visual representation of Formula wheels

Fig. 6.1: The Classic Wheel

Fig. 6.2: The Rally Wheel

Fig. 6.3: The Formula Wheel

3.1.2 Shift Systems

Within Sim racing, there are two types of shifters: Hand controlled shifters and conventional shifters. The hand-controlled shifters do not make use of a clutch, but rather have a dual-clutch system built within the pedals located behind the steering wheel. The conventional shifter, on the contrary, make use of the third foot pedal to the left, the clutch, in combination with a gear shifter to switch gears. In the era of the standardized semi-automatic gearboxes within Formula 1, hand-controlled shifters have become the standard for Formula 1 cars as well as for Formula 1 simulator racing. Nevertheless, the conventional shifting method is largely used for rally racing, mimicking the real-life workings of rally cars.

3.1.3 Pedals

There are two options when it comes to pedals for a simulator setup, a hydraulic set, and a load cell set. The main difference lies within the feel and the accuracy of the pedals. The hydraulic set mimics the feel of an actual pedal set within formula one cars while the load cell deviates from this feeling and therefore is less accurate. Nonetheless, both types of pedals work for research purposes as little to no real differences will be seen.

3.1.2 Motion installations

Regarding the motion of simulators, there are a set of options wherein the degree of freedom varies. Within a zero degree of freedom, the simulator is stationary and hence is called a Fixed Base simulator. As it is a fixed base simulator, no force powers are felt, which might reduce the realism of the simulator in research experiments.

Motion simulators with a higher degree of freedom, mimic the movements of a formula one car. The first degree of freedom (2 DoF) translates the pitch and roll of the car into the rotation of the motion simulator. A second-degree motion simulator (3DoF) translates the pitch, the roll, and the yaw of the car into this rotation of the motion simulator, yielding a more realistic training environment. The third-degree motion simulator (6DoF) translates all motions of the car into the motion simulator and hence yields the most accurate and realistic environment for drivers to train.

3.1.4 Simulator Configuration

Based on the aforementioned information and the discussion with the responsible supervisor from the EsportsLab, it is decided that the simulator should be configured as follows:

Wheel Base: Podium Wheel Base DD2 - Direct Driven
Steering Wheel: Clubsport Steering Wheel Formula V2.5 X + Quick Release
Pedal set: Clubsport Pedals V3 Inverted
Damper: Clubsport Pedals V3 Hydraulic Damper Kit + Brake Performance kit
Cockpit: RennSport Cockpit V2
Seat: Sparco Pro 2000 QRT Seat for RennSport Cockpit
Visual: Triple Monitor setup
Due to costs, no motion installation is configured, hence the simulator is a fixed base simulator.

While this configuration does not mimic the actual environment wherein the drivers would drive on the track, the configuration is close to equal the performance gained from a better installation. The Wheelbase and the steering wheel do mimic an actual Formula One cockpit setup.

3.2 Dataset

The dataset used in this Graduation Project is harvested from the F1 2020 game developed by Electronic Arts and Codemasters. This game is available on the Steam gaming platform. Using the simulator described in the previous section, data of one session with the duration of 45 laps was harvested. All of these timed laps were driven in the Mercedes F1 team's car (W12). Within this section, the harvesting process, the pre-processing, and the post-processing of the data have been described.

3.2.1 Data Harvesting & Pre-Processing

The Codemasters F1 2020 game is supported by an API that provides the possibilities for extracting game data from a racing session. This list of data that can be obtained is structured in a set of packets that each correlate to one section of the total dataset available. By interfacing this API over a UDP connection, it is possible to obtain the different packages.

However, the dataset and the data packets contain relevant and irrelevant data for this research project. Consulting the Codemasters API documentation, the contents of each packet are described. From this content description, it is possible to denote the important packets required for the development of this enhanced training and filter out the irrelevant data. Within the following table, the packets and the corresponding description of the contents of the packets altogether with the relevancy of the data for this research are described.

Packet Title	Packet Description - Obtained from the Codemasters Forum on UDP Specifications	Relevancy [1-100][%]
Header Packet	The UDP Packet header	22% *[2]
Motion Packet	The motion packet gives physics data for all the cars being driven. There is additional data for the car being driven to be able to drive a motion platform setup.	16%
Session Packet	The session packet includes details about the current session in progress	28% *[4]
Lap Data Packet	The lap data packet gives details of all the cars in the session.	78% *
Event Data Packet	This packet gives details of events that happen during a session. The corresponding event strings are attached in Appendix 2	11%
Participants Packet	This is a list of participants in the race. If the vehicle is controlled by AI, then the name will be the driver's	0%

	name. If this is a multiplayer game, the names will be the Steam Id on PC, or the LAN name if appropriate.	
Car Setup Packet	This packet details the car setups for each vehicle in the session. Note that in multiplayer games, other player cars will appear blank, you will only be able to see your car setup and AI cars.	27%
Car Telemetry Packet	This packet details telemetry for all the cars in the race. It details various values that would be recorded on the car such as speed, throttle application, DRS, etc.	73% *
Car Status Packet	This packet details car statuses for all the cars in the race. It includes values such as the damage readings on the car.	10% *3
Final Classification Packet	This packet details the final classification at the end of the race, and the data will match with the post-race results screen. This is especially useful for multiplayer games where it is not always possible to send lap times on the final frame because of network delay.	3%
Lobby Info Packet	This packet details the players currently in a multiplayer lobby. It details each player's selected car, any AI involved in the game, and also the ready status of each of the participants.	0%

* If parameters are part of crucial information, which is defined as information required for the structure of, and therefore the continuation of the project, this is denoted with an asterisk. Behind the Asterisk, the number of crucial parameters is denoted if applicable.

*i The relevancy of this information is denoted in the number of useful parameters over the total amount of parameters in percentages.

Table 2: The Data structure of the Codemasters F1 2020 API

Consulting the Codemasters API documentation again, it is seen that the frequency of updates per packet differs. Within the F1 2020 game, it is possible to limit the frequency of updates to 10Hz, 20Hz, 40Hz, and 60Hz. For the matter of limiting the amount of data and reducing the overall redundancy of the data, the frequency of updates is set to 10[Hz] and a threshold on the relevancy parameter of the data packet is introduced. Whenever the relevance of the data packet is below 50% and the data packet does not contain crucial data, this packet is ignored in the UDP queue, and hence is not sent to, nor received by, the application. Therefore, the packets that are sent and received are the following:

- Header Packet
- Session Packet
- Lap Data Packet
- Car Telemetry Packet
- Car Status Packet

The exact components and attributes that are defined within the specific packets, can be found in the documentation of the F1 2020 UDP Specifications.

3.2.1.1 Database Types

After the harvesting process, a database needs to be created to save the collected data. There are many different solutions and options available for databases. However, as the rate of data is extremely high and needs to be parsed to the database accurately, there are a set of three requirements defined for the database.

1) The Database needs to be able to handle rapidly changing data efficiently
2) The Database must use flexible schemes
3) The Database's scalability must be horizontal as the rate of newly added data is high

Based on these requirements, it is concluded that a Relational Database is not suitable for this project and hence the NoSQL databases have the preference due to the speed and flexibility.

Multiple types of NoSQL databases can be used. However, as described in the previous section, one main requirement for the saving of data is that all data that has been collected per update must be combined into one record on the database and hence cannot be split over multiple records. Document-based NoSQL databases provide this functionality of storing data that belongs together easily. A set of options is available for Document-Based NoSQL Databases:

- MongoDB (Atlas)
- Amazon DynamoDB, as Supported by The F1 Community
- Google Cloud Firestore
- ArangoDB
- Etcetera…

Among the many options, Google Cloud Firestore, part of the Firebase Development Platform, provides the option of a Freemium subscription. However, the limitations of this freemium subscription are a maximum writing rate of 20.000 documents per day and a maximum reading rate of 50.000 documents per day. For this research project, it is the best solution that has the lowest costs while it provides the most value to the project due to all the additional features the Google Firebase Development Platform provides.

Nevertheless, MongoDB is a very suitable platform as well due to the higher maximum document write/read rate per day, if the subscription costs were lower. For this research, the database choice is Google Cloud Firestore.

3.2.1.2 Collection Construction

Based on the incoming stream of data, the collection construction goes according to the packet structure of the Codemasters API. The Collections are constructed based on Session ID, Lap number, and the corresponding times with an interval of 1(“one”) second. The underlying document per collection contains the following data fields:

- Actual Tyre Compound: The Tyre compound that is currently used
- Best Lap Num: The lap number of the best lap
- Best Lap Sector 1 Time: The sector 1 time of the fastest driven lap
- Best Lap Sector 2 Time: The sector 2 time of the fastest driven lap
- Best Lap Sector 3 Time: The sector 3 time of the fastest driven lap
- Best Lap Time: The time of the best driven time corresponding to the Sector 1,2,3 times
- Best Overall Sector 1 Time: The sector 1 time of the entire session
- Best Overall Sector 2 Time: The sector 2 time of the entire session
- Best Overall Sector 3 Time: The sector 3 time of the entire session
- Last Lap Time: The time of the previous lap
- Brake ratio: The rate of Brakes that is applied
- Clutch ratio: The rate of Clutch that is applied
- Current Lap Number: The current lap number
- Drs: Whether DRS is enabled
- Engine Damage: The rate of damage to the engine in percentages
- Engine Temp: The current engine temperature
- Front wing left damage: The current damage to the left-wing in percentages
- Front wing right damage: The current damage to the right-wing in percentages
- Gear: The current gear
- Gearbox Damage: The current damage to the gearbox in percentages
- Pitch: The pitch of the nose
- Yaw: The Yaw of the nose
- Roll: The roll of the nose
- Player index: The current position of the car on the track
- Rear wing Damage: The current damage to the rear wing in percentages
- Rev: The current amount of revolvements per second
- Rpm: The current RPM of the engine
- Sector1 Time: The current time of Sector 1
- Sector2 Time: The current time of Sector 2
- Session ID: The ID of the session
- Session Time: The current time in the overall session
- Speed: The current speed in Km/h
- Throttle: The rate of Throttle that is applied
- Total Lap Number: The total amount of laps to be driven
- Total Track Distance: The total distance of the lap in meters
- Track Distance: The current distance of the car on the track from the beginning of the lap
- TrackID: The ID of the track
- Tires Age in Laps: The current age of the tires in laps
- Wheel: The angle of the steering wheel

3.2.2 Data Analysis & Learning

Based on the data denoted in section 3.2.1.2, different algorithms can be defined for understanding, learning, and classifying the data. Focusing on Machine learning principles and Deep learning algorithms, four specific algorithms can be used to efficiently learn the data. These algorithms include

- Active Deep Learning,
- Decision Tree algorithms,
- Boosting Algorithms
  - Gradient Boosting,
  - ADABoosting

However, while these algorithms do prove to be efficient for understanding the current construction of the data gathered, these algorithms deem to get complex quickly. For this research, only a simple classification method is required. An additional machine learning algorithm can be used in further studies to not only visualize the feedback but also predict events.

Within the context of this research, the focus will be on the Normal Distribution and Linear Regression and its application on the data set. The corresponding classification and the methods used are described in section 5.2.

3.2.3 Data Visualisation

Within the previous chapter, different types of feedback have been discussed. From this analysis, it was denoted that Visual feedback in combination with haptic feedback provided the highest value to the drivers and hence is considered the most optimal way of providing feedback. However, as the scope of the research does not allow the implementation of multiple feedback systems, the focus will remain on visual feedback. Within this section, the different implementation options of visual feedback will be discussed. The leading feedback system to be implemented is introduced in the next chapter.

3.2.3.1 Digital Visualisation

Visual feedback in a digital environment can be realized in many forms. Dashboards have proven over time to be an optimal way of displaying information and data. The amount of data can be regulated and the output can be personalized to match the needs of the driver. Tools that can be used for realizing a dashboard include:

- Tableau Real-Time
- A Responsive ReactJs/Js Dashboard
- Python Dash
- Etcetera

However, the downside of using dashboards within the context of driving, especially in the field of motorsports, forms a major distraction and therefore the use must either be eliminated as much as possible or reduced to only a few parameters. Nonetheless, adding a screen to the simulator, where data will be displayed and most importantly be changed throughout the track, might act as a major distraction.

3.2.3.2 Physical Visualisation

When kept simple, physical visualization can be intuitive and, above all, helpful. Tools for creating a physical visualization system include

- Raspberry pi with a simple LED GPIO interface
- Arduino for feedback by use of human-readable displays

The disadvantage of using a physical type of visualization is that the amount of data to be shown is strictly limited.

3.3 Ethical Considerations

3.3.1 Ethical Reflection method

Considering the topic of this graduation project and its purpose as is discussed in the previous sections of this chapter, many ethical dilemmas arise. In a conference proceeding on Game-Based Learning Earp, Persico, Dagnino, and Passareli (2018) denoted the effects and impact of simulators on behavior and physical and mental wellbeing. The correlated ethical dilemmas that were introduced included the effectiveness of the simulation over time with regards to training, the impact of the cognitive load on the primary functionality of the brain, and its physical footprint on the human body. In research conducted by Sukhov (2019) a systematic distinction is made between the different types of Simulation video games (SVGs), wherein the effects of educational Racing SVGs on behavior is considered to below. Nonetheless, the effects of non-educational SVGs can cause a rise in dangerous driving and hence increase the risk of harmful driving due to the unintentionally obtained skills within the SVG. While tackling all of these ethical dilemmas benefits and improves the final product of this project, the scope of this project does not provide a suitable timespan to cover most. Nevertheless, it is deemed important to at least denote most, and attempt to tackle the most prominent dilemmas. As the ethical dilemma requires correct ethical argumentation, the most suitable paper for discussing the correlated ethical dilemmas is by using the Ethical Cycle as introduced by Van der Poel and Royakkers (2007).

Fig. 7.: The Ethical Cycle as introduced by Van der Poel and Royakkers (2007)

3.3.2 The Moral Problem Statement

Applying this systematic approach to obtain an ethically and morally acceptable action, the model requires the construction of a moral problem statement. Considering the various dilemmas regarding training using simulators, among which some introduced in the introduction of this chapter, the dilemma questioning the effectiveness of simulator training as opposed to the time and physical risks required as an input can be considered as a prominent moral problem. Hence the moral problem statement is defined as follows:

"Is the time and physical risk required to train a professional driver on a racing simulator to obtain a new skill more suitable compared to training this same driver in a real-life environment?"

Evaluating this moral problem statement, is it safe to state that this problem statement does not only include the effectiveness of the proposed training program but rather also takes into consideration the amount of effort the driver must put into the program to learn this new skill. Moreover, this statement questions the efficiency and the need for a virtual simulator compared to a real-life training program wherein the driver can experience the actual circumstance with its corresponding effects.

3.3.3 Problem Analysis

To understand and attempt to answer the moral problem, the moral problem statement must be broken down into sub-questions. The aforementioned statement will, therefore, be split into two parts. The first question that needs to be answered is: *"How much input does the driver need to put into the program to make the program effective for learning a new skill?"*. Answering this first question leads to the second question that needs to be answered before a moral action can be performed. The second question, therefore, is: *"What are the advantages as opposed to the disadvantages for willing to use a simulator over a real-life environment for training professional drivers?"*

As mentioned in the introduction of this report, the learning process of drivers varies in many ways. However, a general model can be used for guiding the learning process into a more efficient and structured way of learning. In an experiment conducted by Abdulwahed and Nagy (2009) on skill learning using Kolb's experiential learning cycle, four stages of learning were denoted. (1) Concrete Experience, (2) Reflective Observation, (3) Abstract Conceptualization, and (4) Active Experimentation. The results of this experiment showed an increase in learning ability and a decrease in learning time by approximately 75%. Within the simulator, these stages can be applied as well, optimizing the active learning process of drivers based on experimentation and reflection on these experimentations. According to the Dutch organization De Koninklijke Nederlandse Toeristenbond (ANWB), a student requires 35 up to 45 driving lessons, with each a length of 60 minutes, on average before this student can successfully obtain a driver's license. Hence, it is possible to define an approximate total learning time of 40 hours to learn the basics of a new skill. As motorsport drivers are experienced drivers and therefore do not need to learn and understand the basic principles of car mechanics and control, this estimate can be downscaled to approximately 10 hours of learning before a new (large) skill becomes a habit. Using the principles of Kolb's experiential learning in the context of the simulator opens possibilities for users to evaluate the skill and their progress in real-time while they are training. Considering that the algorithm implemented in the simulator is not learning the driver an entirely new skill but rather guides the learner through a set of improvements on his current abilities, this average can be downscaled even further up till a repetitive training of approximately 2.5 hours. Consequently, the total input of the user required to train a skill to enhance his driving abilities based on guidelines, supported by data predictions gained by the driver's previous results, within the simulator, is approximately 2.5 hours of consecutive training, with or without breaks.

Nevertheless, within the context of training motorsports drivers, a set of disadvantages can be identified in section 2.3. Keeping this set of advantages and disadvantages in mind, ethical decisions can be made.

3.3.4 Options for Actions

The next phase within the Ethical Cycle as introduced by Van der Poel and Royakkers (2007) is the phase wherein the discussion on potential options for actions for the moral problem statement starts. Within this section, the (potential) measures for the corresponding ethical dilemmas are explained.

As indicated by the first question within the problem analysis, a vital point of interest is the amount of screen time required from the driver to improve upon a skill depicted against its potential health issues. While training for 2.5 hours consecutively might reduce the time required to train for a specific skill, the learning efficiency rate might decrease and drop during these longer training periods due to fatigue. This drop-in efficiency can hence lead to less intake of information and therefore work backfire. As Kühnel, Zacher, de Bloom and Bledow(2016) denoted in their research on the benefits of short breaks for an increase in productivity, consistent breaks are required for optimal performance and work/learning efficiency, increasing the amount of information intake. Changing the training routines and dividing sections of the training into smaller parts can therefore increase the learning efficiency. Accordingly, a possible option for the action to overcome this ethical dilemma is to reduce the session times to spread the work over several moments either within a day or throughout several days by implementing session limiters on the simulator program. Using different models, e.g. the Pomodoro model, to increase efficiency and productivity, the driver can learn in the fastest possible ways without the risk of information fatigue and/or potential health issues.

Additionally, in the longer term, a solution for reducing the amount of screen time is to develop the project on an actual formula one car so that the driver can benefit from the same feedback received in the simulator in a real-life environment. Subsequently, this opens up opportunities for breaking down the training sessions into simulator sessions and conventional on-track practice sessions, leading to a reduction in screen-time and hence a reduction of potential health risks.

3.3.5 Ethical Judgment

To broaden our understanding of the aforementioned ethical dilemma, an ethical analysis is conducted on the overall content of the problem analysis, the options for actions, including the key moral principles. For this purpose, a defined set of analysis tools have been used. Within the first section, the first tool is described with its corresponding details and results. The first tool that is used is the Fleddermann Line Drawing Tool. The Fleddermann Line Drawing depicts the negative and positive paradigms against each other while depicting the different (defined) cases between these paradigms.

Problem Scenario: The driver is trained based on simulation, machine learning, and process mining to enhance his performance in low-performance areas of the track in a racing simulator to later transfer his knowledge into a real-life situation.

Positive Paradigm: The driver understands his errors in the simulator and adjusts his driving methods accordingly by training and reflection to improve in a real-life situation.
Negative Paradigm: The driver does not understand the errors and hence does not improve upon his current performance.

Statements:

1) The driver has a clear vision of what needs to be done to improve the driving performance in a real-life situation,
2) The driver does not have a clear vision of the errors and hence does not know how to adjust his performance,
3) The driver is aware of the errors but fails to correctly translate the errors into real-life adjustments,
4) The simulation does not feel like a real-life situation and hence the driver lacks the motivation to invest time and effort into the training,
5) The trainer gets clear insights and instructions on what the driver is doing wrong and translates them for the driver into actions of the real-life environment,
6) The driver explores different situations that occur with a statistical chance of happening of 1%. He trains himself to overcome the situation within the simulation

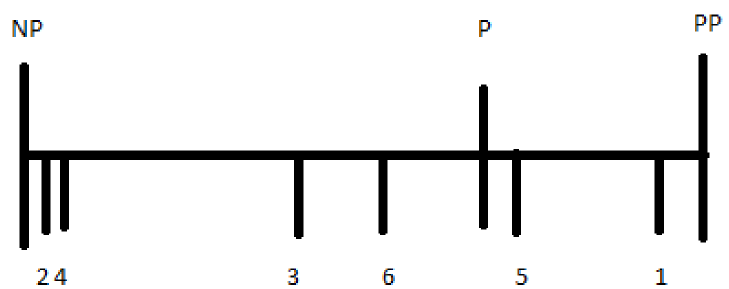

Fig. 8: The Line Diagram Tool by Fleddermann

Evaluating this line drawing, it is seen that the problem statement is skewed to the right side of the spectrum, implying that, while the statement is questioning the level of positivity, the overall statement is presumed to be on the positive side of the paradigm. The reason for this is that the problem statement is questioning the solution chosen at the moment, is the better solution, which due to the wording, is leaning towards a question with a positive answer.

Continuing on the aforementioned sub statements, we can start evaluating from the lowest integer value up until the highest integer value: Statement 1 has been placed on the utter right side of the spectrum. This implies that the driver understands his errors, knows how to tackle them in the simulation, and successfully translates them into a real-life environment where he, again, knows and understands how to overcome the issue to improve his performance. On the contrary, as scenarios can never be perfect in any real-life environment, it is not possible to execute the changes to the full extent, leaving space for some additional reviewing. Continuing to statement 2 and 4 it is seen that these statements are on the other side of the spectrum, the NP side of the spectrum. The reason for statement 2 is that this statement is implying that the driver knows he is making some major errors but fails to see what he is doing wrong and what needs to be done to improve the current performance. The driver fails to correctly use the application correctly and hence lacks the necessary insights required to evaluate and improve. While in statement 4 the user cannot find the right motivation to use the simulation to find his errors and therefore will not obtain the insights. Statement 3 is still leaning towards the negative side of the spectrum. However, as opposed to statement 1, the driver is aware of his errors and successfully improves upon these errors within the simulator, but fails to improve upon them in the real-life environment. Statement 5 is a statement that is located on the positive side of the spectrum. Although the message of the feedback is received well, an external person is involved who needs to explain to the driver the errors, mistakes, and the actions needed to be taken to overcome the issue. This does imply that the driver himself does not see the errors but rather another person involved in the process does and needs to translate these errors for the driver. Finalizing with statement 6, it is seen that this statement is perfectly in the middle of the spectrum, this is due to the validity of the results. As the statistical chance is so low, the progress cannot be validated and hence the driver will never know whether he performed an action correctly in the real-life environment.

Considering the first sub-question of the problem analysis, we had defined multiple ways of ensuring that drivers would not exceed the healthy amount of screen time per session. These, and additional methods of preventing this to happen, have been included in the statements below. These several options are set for analysis. The second tool used for deeper analysis is the model of Tromp et al. (2011).

Problem Statement: Drivers experience a high rate of screen time whenever training in a simulator. This rate can lead to health issues and unnecessary risks. A solution must be introduced to reduce this amount of screen time and encourage the drivers to take a sufficient amount of breaks and perform physical activity.

Options to reach the goals:

1) A time limit is placed on the simulator. After x minutes of activity, the simulation will pause, the screen will lock, and will only unlock after a certain amount of time.
2) A Pomodoro timer-based application will provide the driver with an assignment to perform. Every 25 minutes this timer will expire and the application will no longer give assignments to the driver for 10 minutes. After 10 minutes the routine will restart and the driver can continue his assignments for another 25 minutes.

3) A LED indicating that it is time to take a break is placed on the simulator. A pressure sensor is placed in the seat and whenever a driver sits for too long, the LED will turn on and will try to seduce the driver to take a break.
4) A fresh beans coffee machine is placed near the simulator. The smell of the coffee beans has to lure the driver and make him want to take more coffee breaks.
5) The trainer of the driver creates flexible planning based on the current situation of the driver wherein he takes into account the need for breaks whenever the trainer spots signs of fatigue.

These 5 options are placed in the model of Tromp et al. (2011) to see how they influence the behavior of the driver. A follow-up on the decision of why the option belongs in a specific category is provided below the diagram.

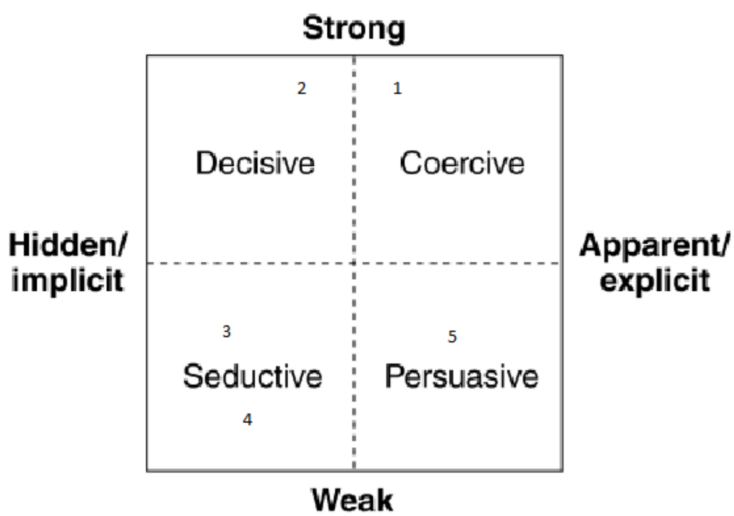

Fig. 9: The Tromp diagram with the corresponding numbers of the options

Evaluating the diagram and the positions of the options, we see that we have two very strong options(2 and 1), forcing the driver to take mandatory breaks by not letting him train anymore. While one works one more structured way, the other defines after a timespan that the driver has been seated enough and hence it is time for a break. The solutions with regard to luring the driver out by seducing him with treats are positioned in the lower-left corner. These imply that the driver has a need for a break and hence tries to seduce the driver into a break. A fifth option is a persuasive option as the trainer is actively managing the situation and the level of fatigue of the driver. With this information, the trainer can make the right decision at the right time.

3.3.6 Reflection

The reflection phase of this ethical cycle is perhaps the most important phase of all. Within this cycle, it is controlled whether all requirements have been fulfilled before the decision on the act can be made. If this is not the case, the options for actions need to be recalled to improve for the better.

The design of a simulator training program can be successful if all these requirements and thoughts are taken into account. In the end, the health of the driver, the stakeholders, the project members, and the environment are the most important things in the project. Hence decisions need to be made on what is deemed to be correct given the aforementioned analysis.

After proceeding through the reflection phase, a morally acceptable action option can be chosen. With the use of the Ethical Cycle, an attempt has been made to set a step toward an inclusive design. Although there is little to change about the simulator, there is a lot to be changed to the program this project is developing.

Chapter 4: Requirements

4.1 Focus

While there remain many parameters, as denoted in section 3.2.1.2, to be optimized for maximum race result, it is important to keep the focus of the project in mind. Recalling the initial research question, it is noted that the project seeks to find an answer to the question whether we can optimize the race performance by enhancing the drivers abilities and improve upon his paths through the track. Hence it can be stated that the main parameters to be optimized are the Steering Wheel angle, the ratio of Throttle per unit of time, and the ratio of Brake appliance per unit of time. These three parameters will be applied to the Active learning algorithm to calculate the next best move and to flag this next best condition, based on the included parameters, as the new optimal way through the track. This next best move will then be translated into a visual representation, leading the driver through the different markers set by the learning algorithm.

4.2 Design Decisions

Over the process of this research, some design decisions were made to efficiently and optimally calculate and communicate the results of the learning algorithm to the driver. These decisions include assumption on the ideal variables managed within the learning algorithm, the feedback receival and the common understanding of the data. Another aspect that needs to be considered is the amount of variables displayed to the driver during his/her performance.

Based on the focus of this research and the aforementioned parameters, the visualization type to be implemented is digital visualization. As only three variables are managed over the entire course of the track, these variables can be translated into percentages <u>on the difference</u> between the current and the optimal situation. While this visualization is simple, the message to the driver *should be* clear and concise, leaving little room for distractions and/or misinterpretation of the feedback. The variables displayed within this visualization are the optimal difference between the optimal and current steering angle, the difference between the optimal and current brake ratio, and the difference between the optimal and current Throttle ratio.

Although mentioned in previous sections that physical feedback is more effective, the decision to implement a digital dashboard is grounded on the limited scope of this project. Due to time constraints, it was not possible to fully develop a working prototype. Therefore the solution was to implement a digital dashboard to display the data. The corresponding language in which this dashboard is designed, is Processing. Processing is chosen as the language is based on the Java programming language. This would mean that the integration would become more feasible.

The design decisions with regards to the data, the learning model and the visualizations, are elaborated in chapter 5.

4.3 Requirements and Specification

4.3.1 Basic Requirements (Release v1.0)

For the first iteration within the development of this training a set of requirements is defined. These requirements yield all the functionality that the application must fulfill. The set of requirements are split into a set of Functional requirements and a set of non-functional requirements. Wherein the Functional requirements focus on the capabilities and the features, whereas the non-functional requirements focus on the user experience and the reliability of the application. The corresponding MoSCoW Analysis is added as well.

Functional Requirements

Field	Requirement	Priority
Process Data Harvesting	The system must retrieve the data of the simulator in real time	Must Have
Data structure and coherence	The system must format the data of each individual recording into a set of data coherent with the corresponding environment of the car in the simulator	Must Have
Database	The system must allow a user to get data insights filtered on sessions	Should Have
Database	The system must allow a user to get data insights filtered on laps within a session	Should Have
Database	The system must allow a user to get data insights filtered on seconds within a lap per session	Should Have
Datastream update	The system must will update a real time database according to the incoming stream of data	Could Have
Datastream update	The system will update a database according to the incoming stream of data	Must Have
Data coherence	The system must retrieve the total set of data based on parameters	Must Have
Statistical Analysis	The system must calculate the quickest lap time and denote, from the corresponding data points, the wheel angle, throttle and brake ratio.	Must Have
Statistical Analysis	The system must calculate the quickest sector	Must Have

	on record per sector and denote, from the corresponding data points, the wheel angle, throttle and brake ratio.	
Statistical Analysis	The system must calculate the difference and translate this difference into percentages	Should Have
Feedback & Visualization	The system must provide visual feedback on the differences between optimal and current.	Must Have
Feedback & Visualization	The system must provide haptic feedback on the differences between optimal and current	Must Have
	……	

Table 3.1: Functional Requirements

Non-Functional Requirements

Field	Requirement	Priority
Database security	The data per session shall be secured in the database using a hash-function	Could Have
Accountability & relevance	The learning path shall be traceable over the process of the usage	Should Have
Database Efficiency	The database updates during a racing session shall happen at most every 1 second. The data in the meanwhile must be buffered on the internal storage	Must Have
Database Efficiency	The internal storage shall be emptied after an update resulted in a success	Must Have
UDP connection	The harvesting of data shall not listen for longer than half a second. After this period, the data packet is dropped and a timeout is sent.	Should Have
	……	

Table 3.2: Non-Functional Requirements

Chapter 5: Realization

The realization has been divided into three realms. Each realm corresponds to the research sub-questions defined in chapter 1. The first realm discusses the data harvesting process and the preprocessing of the corresponding data into a database. The second realm covers the understanding of the data using machine learning and statistical analysis. The third realm describes the implementation of the feedback system with its corresponding elements.

5.1 Data Harvesting & Preprocessing

Consulting the Codemasters API documentation on the integration of this API within a third-party application, the guides on how to connect to this API are denoted. Within section 3.2, the different types of data packets are described. Within this section, the relevancy of these packets is calculated which led to a downscale in required packets to increase the overall efficiency of the data retrieved. To enhance the overall efficiency and speed of the program the research on the Benchmarks game has been consulted. This research yielded that C# is the language of choice for harvesting and processing data to cope with a high-speed data rate.

To mimic a real-life situation, the racing simulator, as described in section 3.1.4, is used for running the F1 2020 game. Within this game, the option for sharing telemetry is turned on. The harvesting of data has three steps. The first step is the sending of data. As the API provides this functionality, there is little control over the formatting of the data sent. The F12020 game handles the correct sending over a UDP connection to an available client on the same network. For ensuring that the connectivity is over the same network, a mobile hotspot is set up on the client-side and the racing simulator is connected to this mobile hotspot. The second step is to retrieve the data on the client-side and process this into readable data. This client side is specifically built and designed for this project and therefore we do have control over the data and its corresponding formatting. The third step is to filter the relevant information and parse this to a database. The scheme of how this system and its components interact is displayed in figure 10.

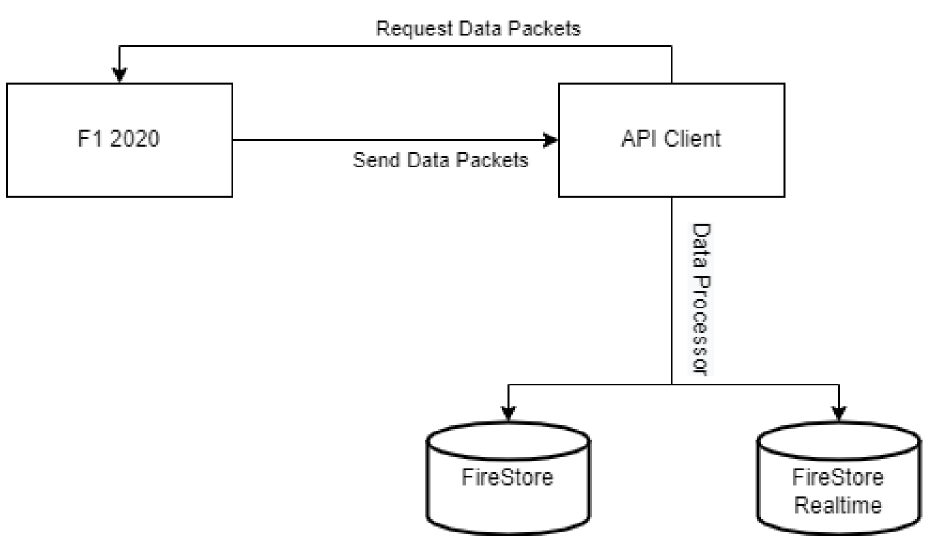

Figure 10: The ATS System

When data is retrieved from the API, the client decodes the data stream and processes the data into information objects. As denoted in section 3.2.1, the different data packets arrive asynchronously. To merge arriving packets into one data object, an object buffer is created with a 7 millisecond lifespan. Throughout the lifespan of this buffer, all retrieved data is combined into the same object and redundant data is overwritten. These objects are then formatted into a C# directory to later be formatted into a JSON object. A connection is made between the client and the Firebase database. Once the end of the lifespan of a data buffer is reached, the JSON object gets parsed to the Firebase FireStore database. Accordingly, Firestore responds with an approval message or an error. The error contains information and instructions on how to proceed. The approval message validates the arrival of data into the database.

The database is divided into three collections. This structure is displayed in figure 11 The first collection ("Collection 1") contains all the gathered Test data based on the collection construction denoted in section 3.2.1.2. On the contrary, the third collection ("Collection 3") contains all the gathered Training data. The second collection ("Collection 2") contains real-time telemetry updates and therefore is updated every 0.5 milliseconds based on the incoming stream of data from the Racing Simulator.

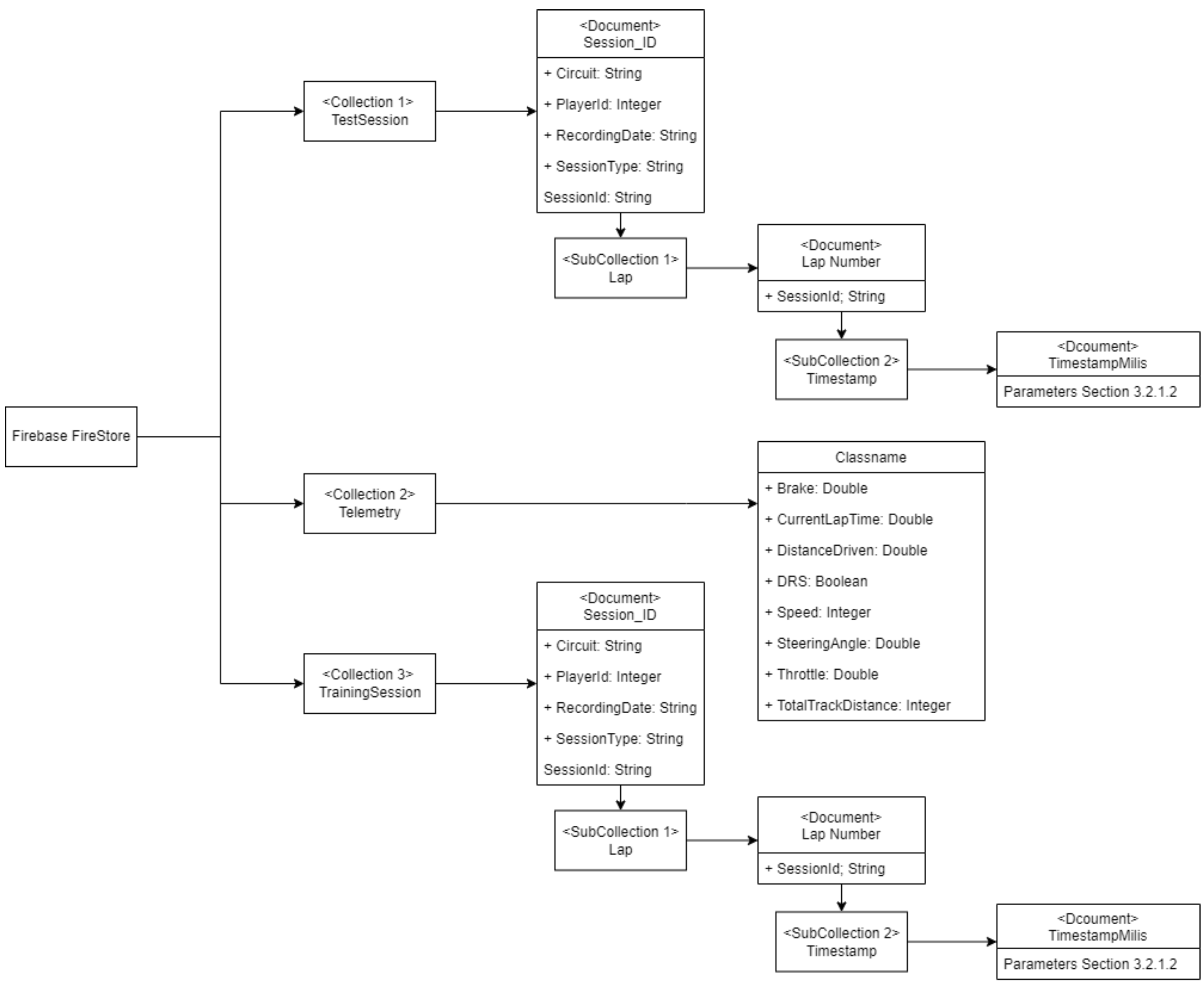

Figure 11: The JSON Data structure in FireStore

This flow of processes is called the automated transport system (ATS) and ensures the availability of session data and real-time telemetry data in the database.

5.2 Understanding & Learning

When constructing the model for understanding data, a breakdown of tasks is required. For the implementation of the "Understanding & Learning" part within this research, the following division of task categories has been made:

1) Retrieval of Data ⇒ Read Firebase
2) Simplification of data ⇒ Create data markers per condition X/Y*
3) Grouping of Data ⇒ Group on conditions X/Y
4) Construction of Summaries ⇒ Calculate summary** and averages Marker group
5) Classify Marker Groups (MG) ⇒ Within each group, classify value on Normal Distribution
6) Recreation of Trackline ⇒ Create ideal trackline from highest classification per MG

* X is defined as the distance ration, Y is defined as the time in lap

**The summary correlates to the Five number summary altogether with the standard deviation and mean

These task categories are followed by a set of tasks required for postprocessing and therefore translation of the data into accurate (real-time) feedback.

1) Retrieval of Real Time Data ⇒ Read Real-time Firebase
2) Simplification of Data ⇒ Create data markers per condition X/Y
3) Ranking of Data ⇒ Ranking incoming data in relation to ideal trackline
4) Translation into feedback ⇒ Calculation of difference & Translation into change ratios

This part of the realization requires a strong object oriented structure and hence the Java Programming Language is the language of use. Starting at the first set of categories, the first step is to retrieve data from the database. The database therefore must consist of data, which is provided by the ATS system, as described in section 5.1. Consulting the Firebase integration documentation, reading data is done by obtaining the collections and retrieving the corresponding documents per collection. Within the experiments, it is expected from the participants to first perform 8 laps under training circumstances. This data is stored under the Training data collection as denoted in figure 11. This data is retrieved, according to the Firebase integration documentation, from the database and parsed into data objects. Simultaneously, the data is written to a locally based .txt file to ensure faster operations for the during the next retrieval iteration.

The next step in the process is to simplify the retrieved data and to group the data together. The parameters for simplifying data are the current timestamp in seconds on the track and the distance ratio in percentages. The level of significance is in milliseconds for the timestamp and one decimal after the comma for the distance. For instance, a distance of 47.892952 [%] is categorized as 47.9 [%], and a timestamp of 23.1234 [s] is categorized as 23.1 [s]. A marker is created per categorized timestamp and per categorized distance ratio. Within this marker, the corresponding data for this position on the track is included. From these two types of markers, it is therefore possible to deduce the distance per timestamp, The timestamp per distance, and the corresponding brake/throttle ratios, and the wheel angle.

The next step within the process of understanding & learning is to create the corresponding summaries for defining the ideal telemetry set per marker. This is done only for the marker with the distance ratio, as this marker defines the leading track line correlated to the telemetry. The summary consists of the distance ratio, the mean wheel angle at this distance ratio, the mean throttle/brake ratio at this distance ratio and the modus of the gear at this distance ratio. After the mathematical summary has been created, the markers get labeled with a classification. This classification is built upon the Normal Distribution where the critical values are defined by the Z-Values derived from a distribution with a level of significance of 0.05. Accordingly, the obtained Z-Value for a $\alpha_{total} = 0.05$ are equal to:

$$\alpha_{upperTail} = 0.025 \Rightarrow \mu + 2\sigma$$
$$\alpha_{lowerTail} = 0.025 \Rightarrow \mu - 2\sigma$$

Yielding the following criteria for the classification:

Formula	$\mu - 2\sigma$	$\mu - \sigma$	μ	$\mu + \sigma$	$\mu + 2\sigma$
Classifier	Low	MidLow	Mid	MidHigh	High

For which μ is the calculated marker average per parameter of the marker and σ is the corresponding standard deviation from this average.

After the dimensions of the classification have been defined and the markers have been classified, the entire dataset, as retrieved from Firebase, gets classified on the basis of the aforementioned classification criteria. Subsequently, the lower classified data markers get removed from the dataset, ensuring only "Mid", "MidHigh" or "High" classified markers and data points within the dataset. The next step is therefore constructing a new trackline based on the highest classified data points. This recreated ideal trackline is defined as the "Advised Trackline for maximum performance". Throughout the entire session this process is repeated, improving the ideal trackline per newly created or updated marker. Figure 12 visualizes the entire process into a schematic.

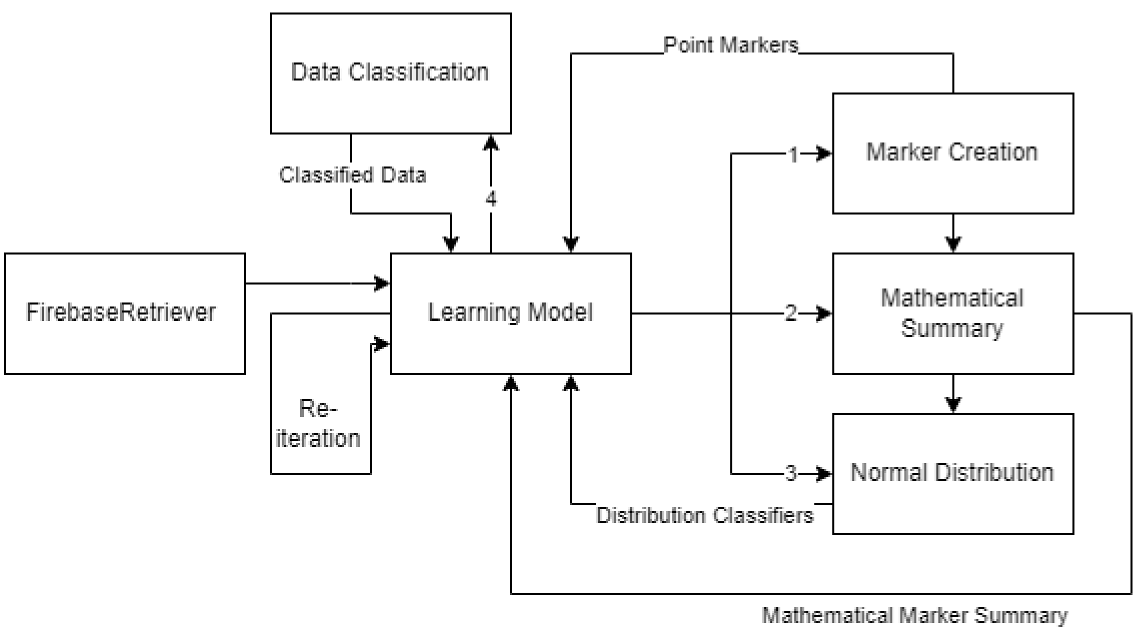

Figure 12: The flow of generating the Learning Model

In the aforementioned task categories, the steps for postprocessing and translating current (real-time) data points have been denoted. In the previous paragraph, it is implied that data from the real-time telemetry is compared to the advised trackline. This is part of the conformance checking phase of the project. The first step is to retrieve the data from the database. This process of retrieving data is identical to the prior mentioned process for retrieving the session data. The only difference is that the structure, as denoted by Figure 12, contains only one base layer of collections and documents. The update rate of the real-time telemetry is 0.5 seconds. As mentioned in the paragraph before, the ranking of data goes according to the classification. The difference between the advice and the current telemetry data per marker is noted and parsed to the visualization tool. This visualization tool is discussed in the next section.

5.2.1 The Learning Curve

Throughout the research and the experiments the driver will face a certain learning curve that might influence the results of the research. To overcome this learning curve and therefore to minimize the effect of this learning curve, the participant is asked to drive 8 laps before the test. Throughout these laps, the participant will expose the learning curve by means of increasing marker classifications. Once the system recognizes a stabilization within the graphical representation of the participant's output, the learning curve gets identified as all the output before the stabilization. Accordingly, the data gets removed from the training dataset, and the participant's learning curve is eliminated. However, the markers classified as Mid, MidHigh and High will remain in memory for the improvement of the ideal trackline.

5.3 Communication & Visualization

Displaying feedback is done through a simple visualization dashboard containing three main parameters and 2 additional parameters as support. As mentioned within sector 3.2.3, the aim of the visualization is to inform the driver without causing too much distraction. Hence the visualization must be simple and easy to understand from out of the corners of the drivers eyes. Therefore a decision is made to only display three necessary parameters with which the driver immediately can see what needs to be done. The criteria for the visualizations are that the colors must be distinguishable and the information must be recognizable. For the steering angle a two-sided horizontal histogram is used to denote the rate of change that needs to be applied to the current steering angle. The Brake and Throttle work according to a vertical bar chart that turns green when too little pressure is applied and turns red when too much pressure is applied. The visualization tool is developed using the Processing 4.0 Beta 5 Library within a Java Application. A snapshot of the application is provided in Figure 13.

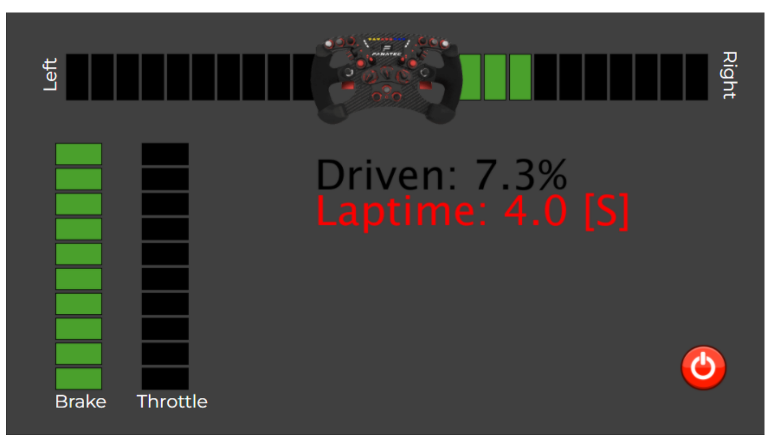

Figure 13: A snapshot of the visualization tool

5.4 Conclusion

Within Figure 14, a total overview of the application and its interactions with other components is displayed. Within this scheme, the communication lines between the different systems are denoted.

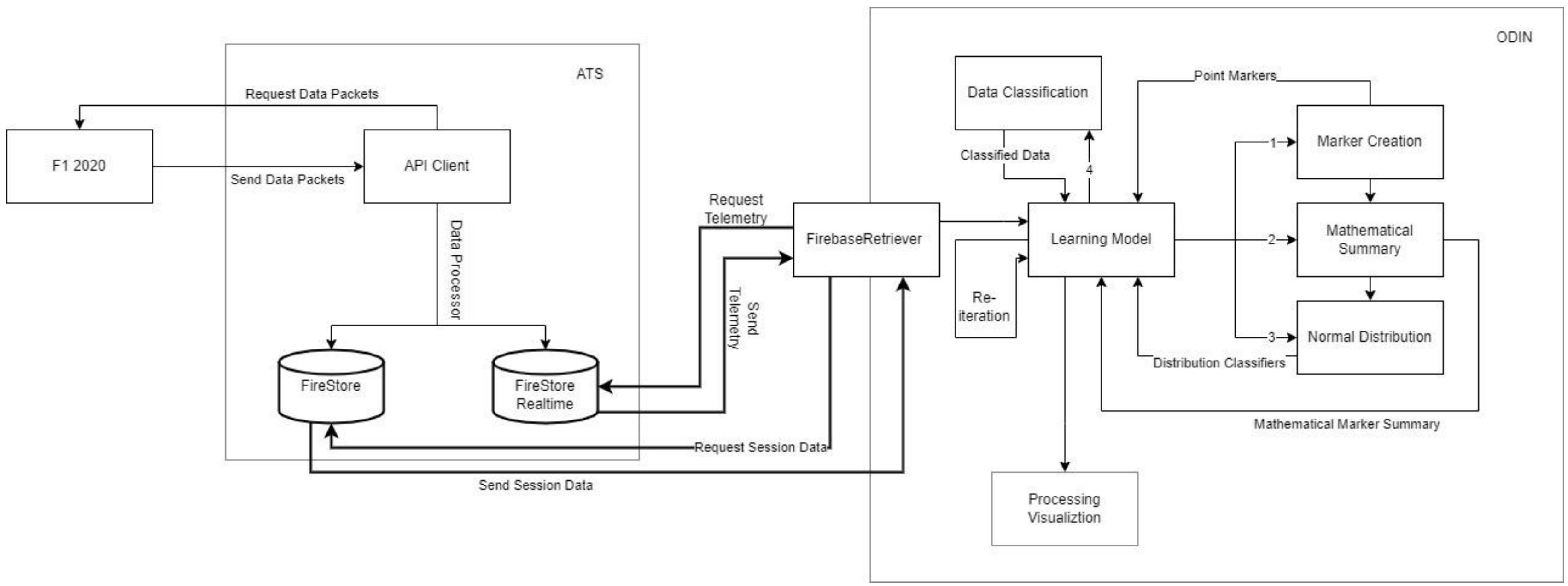

Figure 14: A total overview of the applications with all their corresponding attributes and components

Chapter 6: Analysis

Within this section, the analysis on the results is conducted. Before this analysis can take place, an experiment is set up with a certain number of participants. These experiments and their procedures are described in more detail in section 6.1. The experimental setup is followed up by the visualization of the obtained results from the experiments in section 6.2. The aim of this section is to understand the theoretical result that the experiments yield based on the previously done research and the conducted experiments.

6.1 Experiment setup

To properly evaluate the results, a predefined experiment structure is required to cover all required segments of the experiment. A single experiment consists of two drivers, a potentially good driver and a potentially worse driver. The better driver is asked to participate in the experiment first. Subsequently, the worst driver is asked to participate in the experiment as second. Although not desired, it is important to anticipate a situation wherein not all participants have an equal amount of knowledge on racing simulators. Therefore, to reduce the effects of this knowledge gap, the experiments are divided into 5 sections:

1) Participant Briefing & installation
2) Training Session
3) Test Session
4) Survey
5) Open Discussion

6.1.1 Participant briefing and Installation

Considering that the aim of the research is to improve professional drivers, a minimal understanding of a car, the research and both their components is required. When the experiment starts, the participant is asked to give consent after reading, and accepting, the research background, the research aim and the experiment procedure. During this section within the experiment, the participants are asked about their knowledge on the topic, their level of knowledge on driving and whether the participant understands how to operate a racing simulator. Afterwards, the participant is invited to take a seat within the simulator, while the researcher explains the components, the functionality of the different buttons and the built-in feedback systems.

When the participant has all his/her questions answered, the simulator is turned on and the participant is asked to drive a minimum of 2 laps and a maximum of 6 laps around the circuit to get familiarized with the game, the steering wheel including its functionalities and the circuit. Within section 5.2.1 the notion of a “learning curve” is denoted. During these initial laps, the side aim is to reduce the learning effect and hence to start the training session with the most reliable data, maximizing the effect of the training system on the participant’s abilities.

During these initial laps, no data is recorded and no visualizations/feedback are shown to the participant. The aim of this section is only to let the participant get familiar with the system.

6.1.2 Training Session

The second section of the experiment is the training session. Within this session, the aim is to feed the learning model and understand the abilities of the driver. Based on these abilities, the learning system can create advice, as described in section 5.2, and generate predictions to display during the Test session. The participant is asked to take a seat behind the racing simulator and to drive 8 laps on the Zandvoort Circuit. Throughout both training and test sessions, the car of choice is the Mercedes F1 team's car (W12)**.** During these 8 laps, all data is collected according to the data harvesting and preprocessing principles as described in section 5.1. No visualizations are shown to the participant but the learning system does evaluate the incoming stream of data to minimize the learning effect and to maximize the reliability of the advice.

After the driver has finished the set of 8 laps, the driver is given an opportunity to take a break to regain strength. This break is mandatory, as otherwise the driver might experience exhaustion and/or sore shoulders after the first session or during the second session.

6.1.3 Test Session

The third section of the experiment includes the test session. This test session is the next phase within the experiment wherein the participant is exposed to the feedback system. If the participant is the better driver, as denoted by the introduction of this section, the dataset only includes data on this participant. If the participant is the second, and hence worse, driver of the experiment duo, the dataset includes all data on the first driver in the experiment and the training data of the second driver.

The driver is again asked to drive 8 laps on the Zandvoort Circuit. As denoted by the 6.1.2, the car of preference here is the Mercedes F1 team's car (W12) as well. During the session, the setup includes an additional screen on which the data visualizations can be seen that correspond to the advice generated by the learning system and the current telemetry.

The researcher closely observes the behavior of the participant to deduce potential errors, misunderstandings and/or issues during the session with regard to the feedback system. After the driver has finished the set of 8 laps, the session is stopped. The driver is asked to pause the game and leave the simulator.

6.1.4 Survey

After both sessions have been completed, the participant is asked to fill in a survey. Within this survey, the participant is asked about his experiences, the benefits and limitations of the system and the potential points of improvements.

6.1.5 Open Discussion

Once the participant has completed the survey, the participant is asked to participate in an open discussion wherein the participant can elaborate on his experiences during the experiment and hence denote potential points of errors. Throughout the discussion, the participant is asked a series of open questions about potential ideas, improvements and the effectiveness of the overall system. During this interview, the participant is allowed to freely comment and question the project and its purpose.

This open discussion is recorded to create a transcript of the interview. The participant is informed about the removal of the recording and the use of the transcript for future development.

6.2 Results

As mentioned in the previous section, each experiment consisted of two drivers and the experiment setup is defined by 5 steps. For this research there was room to conduct three experiments. The analysis of the experiments has been divided into four parts. The first part is on the harvesting of data and the reliability of the created, ideal, trackline. Secondly, an analysis on the overall performance of drivers is done by depicting the R-Squared values of the Test sessions against the R-Squared values of the Training Sessions for every participant. Thirdly, the performance based on the brake ratio, the throttle ratio and the steering angle is analyzed based on the initial training dataset and the test dataset. Lastly, a brief analysis on the feedback system is done. This is done based on the responses of the survey.

6.2.1 Rate of data acquisition with effect on Trackline recreation

To visualize the power of big data and process mining, the effect of a high data acquisition rate is denoted against the reliability of the created advised trackline. By displaying the amount of timestamp markers created per lap, it is easily seen when and where the data packets have dropped. To overcome the amount of dropped data, more data must be gathered. As depicted in figure 15.1, once the total number of laps driven increases, the total coverage of the timestamps increase, leading to a higher accuracy per marker. From this dataset, a generalized model can be constructed wherein each required timestamp is covered by the total data set. From this generalization, a reconstruction of the trackline can be created. The newly created trackline is therefore constructed by applying the learning model to deduce the optimal values per marker. Wherein the harvesting system has to make sure that the sample size is large enough to cover every potential, and desired, timestamp marker. From this reconstructed trackline, an advice is created. This yields that more data means a higher reliability of the advice per marker.

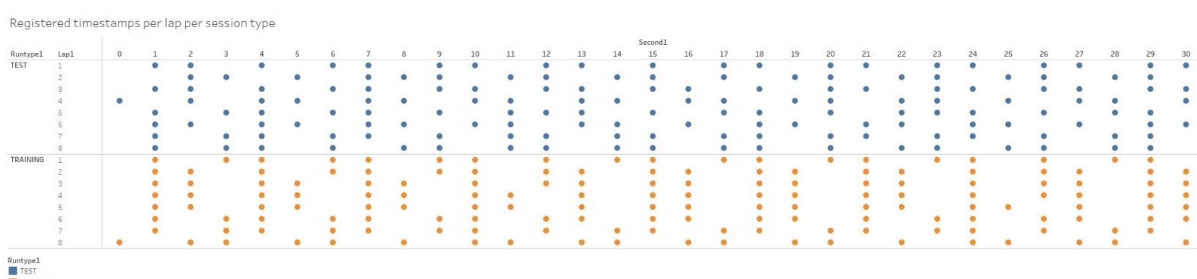

Figure 15.1: The data markers coverage per timed lap, depicting the packet drops

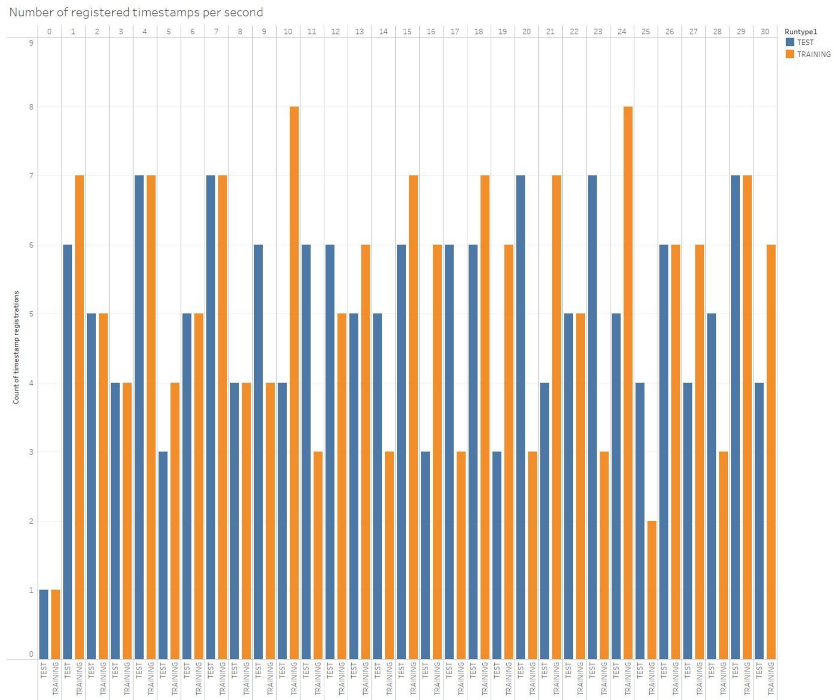

Figure 15.2: The total coverage after generalizing the data markers

6.2.2 Data Variation Change as indicator of method effectiveness

When the goal of a training is to improve multiple drivers, along the same progress line, the variation in data is the most important factor. More variation means a larger difference in performance and therefore a larger difference in abilities, confidence and skills. In this section of the analysis the focus is on the overall performance of the drivers, rather than focussing on the skills, abilities and level of confidence of these drivers. The analysis is based on the R Linear Regression model wherein the severity of variation within the dataset is denoted in R-Squared. R-Squared is in the context of this research defined as the statistical measure of how close the data fits to the generated regression line. If the data is closer to 1, the data fits the model better and hence the total amount of variation is less. Resulting in improvements throughout the training process. On the contrary, if the R-Squared value is closer to 0, the data does not fit the model and hence the variation is significantly worse. Meaning that the training is not effective within the model.

Considering the setup of the experiments, some data can be excluded from the analysis. For instance, the data of lap one and lap two, often include the effects of the learning curve, as denoted in chapter 5.2.1. Moreover, the datasets recorded per driver showed decreasing classification levels the longer the session took. From this can be concluded that due to effects of fatigue and exhaustion, drivers did perform less at the end of a session compared to the beginning of the session. For the reliability of the analysis, only laps 3 and laps 5 are taken into account.

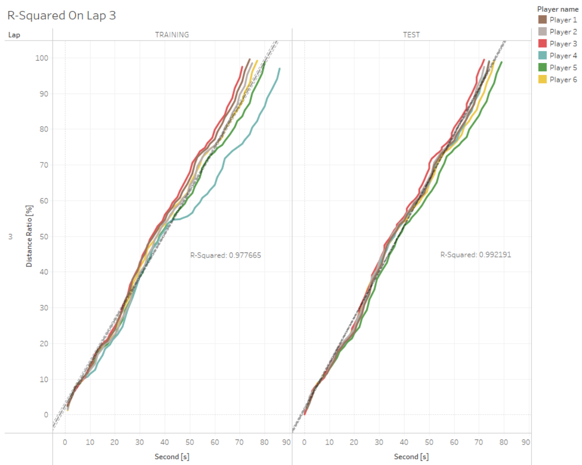

Figure 16.1: The R-Squared value as a rate of variation on Lap 3

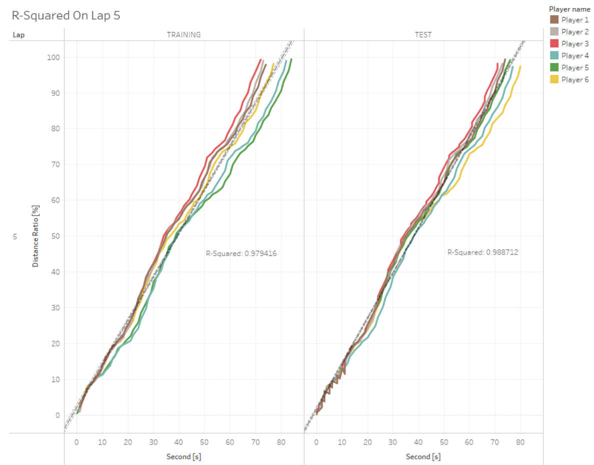

Figure 16.2: The R-Squared value as a rate of variation on Lap 5

Considering the focus of this part of the analysis, the independent variable is defined to be the session type. Per type, it is looked at the difference in R-Squared values. The R-Squared is defined as the depiction of the distance ratio against the amount of time on track. The regression line therefore is the ideal line wherein the 100% distance ratio is reached in the most average amount of seconds. This regression line is defined by the classification model as discussed in chapter 5.2.

When analyzing Figure 16.1, the most visible difference between the training and test data is that there are less peaks and therefore less outliers within the datasets during the test session compared to the training session. Moreover, the values seem closer to each other. This same phenomenon seems to be present within Figure 16.2, where the analysis on lap 5 is depicted. When willing to prove this observation statistically, the R-Squared value is calculated using the formula

$$R^2 = 1 - \frac{Sum\ Squared\ Regression}{Total\ sum\ of\ Squares}$$

This formula yields the following table of R-Squared values against the corresponding lap and session type:

	Lap 3	Lap 5
Training Session	$R^2 = 0.977665$	$R^2 = 0.979416$
Test Session	$R^2 = 0.992191$	$R^2 = 0.988712$

Table 4: R^2 values for lap 3 and lap 5 for both session types

From this table it can be deduced that in any case the training session had less variety within the data samples, although there is only the slightest difference. Nonetheless, This yields that the testing session had an additional factor in play that caused this slight increase in overall performance. Having removed the learning curve as denoted in section 5.2.1, increases the likelihood that the feedback system had an positive effect on the drivers causing the drivers to be more aware of the situation.

6.2.3 Car handling with and without feedback

The second performance analysis method is based on the telemetry data retrieved from the real-time database and the ideal telemetry calculated by the learning system. Following the order of the tests, first an introduction training is given to the drivers to eliminate any effect of the learning curve. Next, the drivers are asked to drive the laps without feedback and after, the drivers are asked to drive the laps with feedback according to the experiment outline as described in section 6.1. On the contrary to the previous analysis deducted in section 6.2.2, the focus of this analysis is on finding and recognizing (potential) improvements in the dataset recorded during the test session relative to the training session.

Before this analysis starts, it has to be denoted that the calculation of the ideal telemetry is based on the averages of the telemetry retrieved from the highest classified markers, this notion is explained in more detail in section 5.2. This means that, although the advised telemetry yields potentially more reliable telemetry data, there still exists a margin of simplification towards an average over the marker. The more data gets collected, the more accurate and reliable this telemetry data becomes.

To cover all aspects of the experiment results, a breakdown in areas of the analysis is required. First, the proceedings of the overall performance of all drivers, as introduced in section 6.2.2, is discussed by means of the differences in overall race pace. Secondly, the relation of this change in pace is depicted against changes in the efficiency of the brake and throttle performance. Finally, the relation of this difference in pace and performance is depicted against the efficiency of the use of the steering angle.

6.2.3.1 Differences in race pace among drivers per duo

Before the experiments started, the drivers were grouped based on their experiences with driving and racing simulators. According to these groupings, driver duo's were created. Each duo consisted of an presumed experienced driver and a presumed inexperienced driver. The aim of this division between the participants opened options for amplifying the effects of the training to gain the maximum insights as possible during the analysis phase.

When looking at the overall race pace of the drivers (Figure 17), it is seen that almost all drivers improve upon their average speed. Within this context, a higher average speed yields lower lap times and hence a more efficient drive. Within figure 17 it is also seen that the rate of change is significantly higher for the less experienced driver after training with the data of the experienced driver compared to the rate of change for the experienced driver. This is due to the amplified feedback that is caused by the difference of experience. While it must be noted that the following values are not statistical averages for a test over a broad sample size, the effects of improvements are visible.

Driver name	Avg Pace during Training Session [km/h]	Avg Pace during Test Session [km/h]	Difference rate in percentages [%]
Player 1	199.14	197.59	-0.808
Player 4	174.58	178.43	+2.203
Player 2	199.80	200.79	+0.469
Player 5	179.71	186.79	+3.938

Player 3	197.78	205.01	+3.659
Player 6	178.70	189.73	+6.174

Table 5: The difference is race pace denoted in percentages [%]

From table 5, it can be deduced that the rate of change of the inexperienced driver correlates to the rate of change of the experienced driver. Meaning that if the experienced driver barely increases, the rate of change for the inexperienced driver will be low due to the low quality of the data. If the experienced driver improves a lot, the quality of the data is high and hence the inexperienced driver can benefit from this set of highly classified data, meaning that the feedback would become more accurate and reliable.

Looking, in particular, at the best improving driver duo (duo 3; red), it can be seen that the experienced driver improves a lot in terms of overall pace and performance, leading to a more significant increase in the performance of the inexperienced driver.

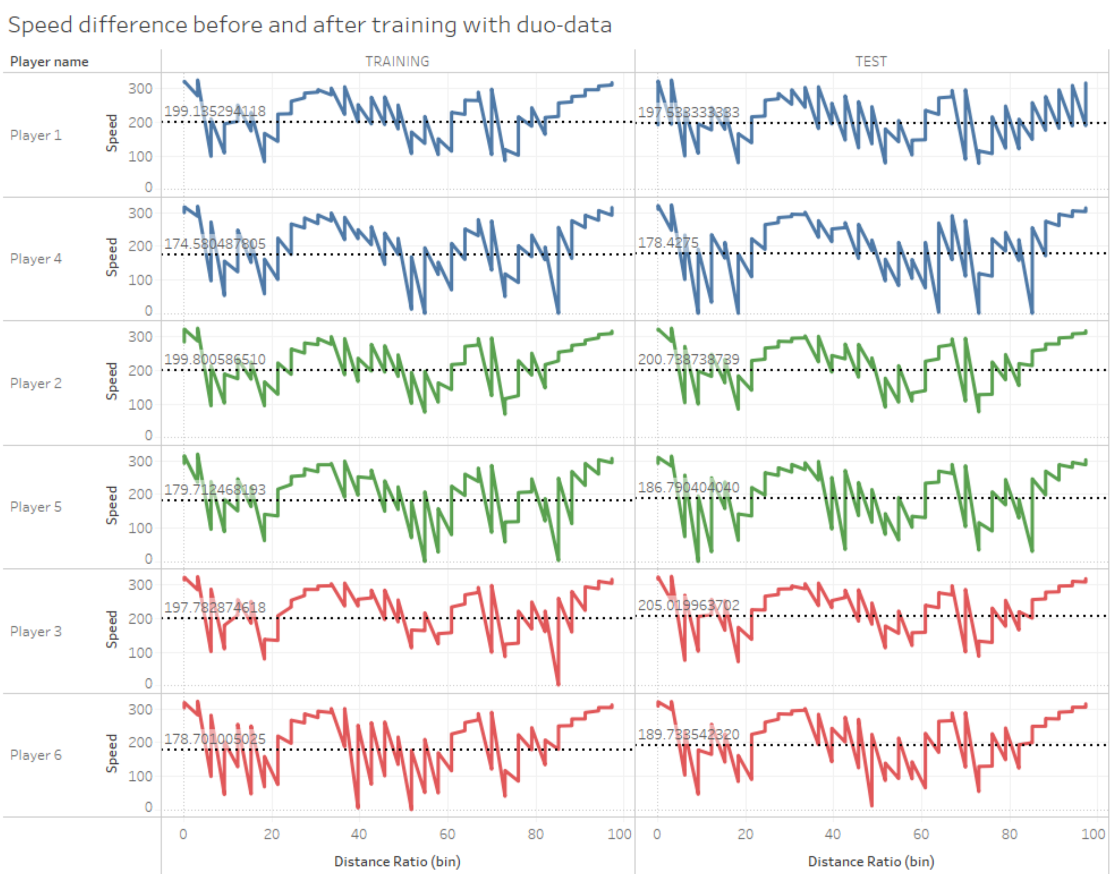

Figure 17: The difference in race pace with and without feedback

6.2.3.2 Difference in efficiency of Brake/Throttle use

Given the large difference in overall pace and driver performance of duo 3; red, analyzing the data of this duo returns the best visible effect of the duo training program. In figure 18, the difference in brake and throttle performance of player 6 has been depicted, wherein the distinction has been made between the session type (training, test). During the training session, it is seen that there is much fluctuation in the throttle. This leads to less time on maximum throttle and therefore less overall speed. These fluctuations can be explained by the level of confidence of the driver. If the driver lacks confidence in the car, the track and his abilities, the likelihood of cautious driving is higher, leading to a lower overall performance. This same principle counts for the brake. Comparing the results of the training session to the test session, it is seen that within the test session, the driver has much more confidence as there are less fluctuations in the driver's brake and throttle handlings. This implies that the driver has more understanding of the situation and hence can better control the car to operate at maximum performance and transition smoother between states, leading to less harsh changes. Moreover, it is seen that the driver is making less use of the brakes and therefore makes more use of the friction of the engine to slow down, implying that more speed and more pace is carried throughout the track, leading to a more efficient handling of the car.

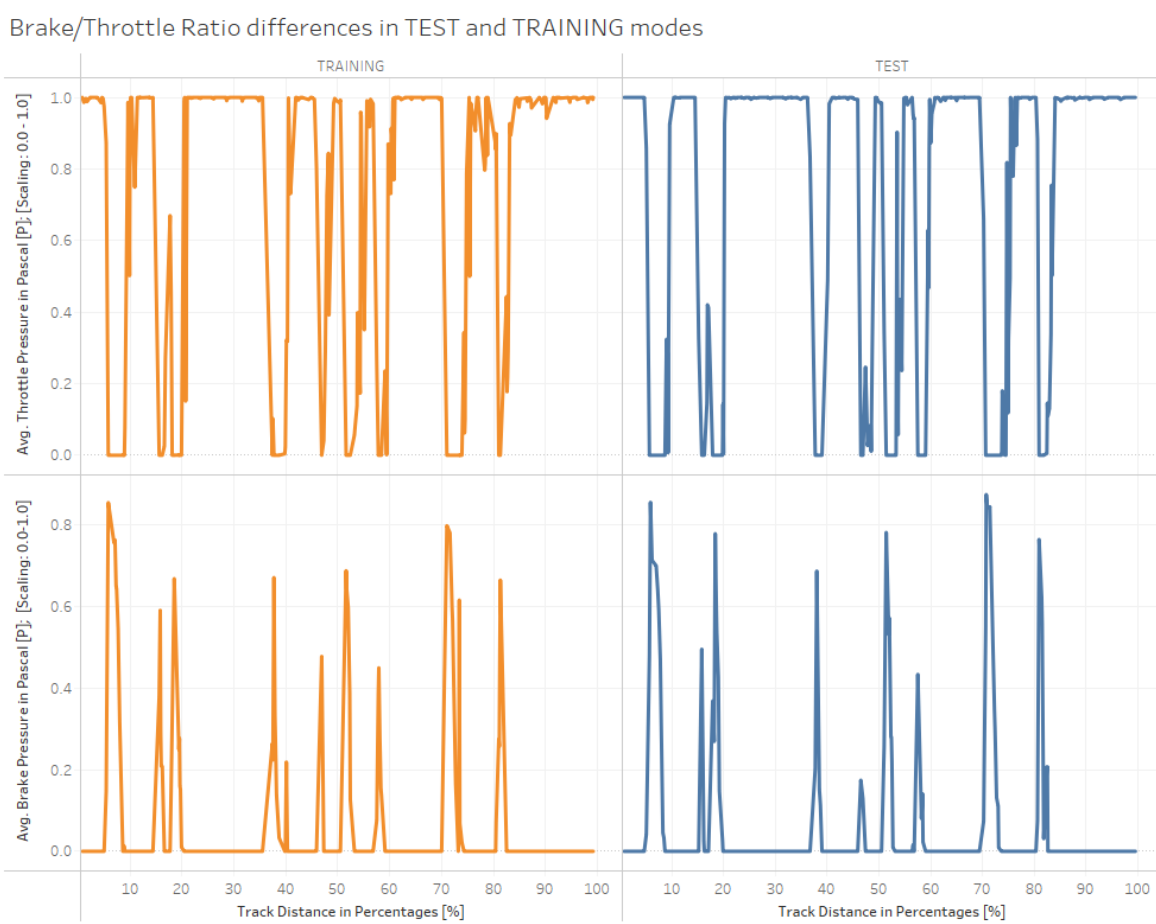

Figure 18: The brake and throttle performance during the training and test sessions.

6.2.3.3 Difference in efficiency of Steering wheel use

Looking at the use of the steering wheel by player 6 in figure 19, it can be seen that the case is very similar to the case as discussed in the previous section (6.2.3.2). During the training session, the player has a lot of fluctuation in the steering wheel throughout the lap. Within the context of the steering angle, fluctuations mean the rate of corrections required to operate the car. Hence, more fluctuations imply more corrections and therefore less control of the car and the situation. Comparing the results from the training session and the test session, it is evident that the amount of fluctuations have decreased, implying that the driver had more control over the car. Moreover, it is seen that the steering angles remain more consistent over the track segments, meaning that the cornering gets longer, yielding more pace at the end of the corner, yielding improved exits.

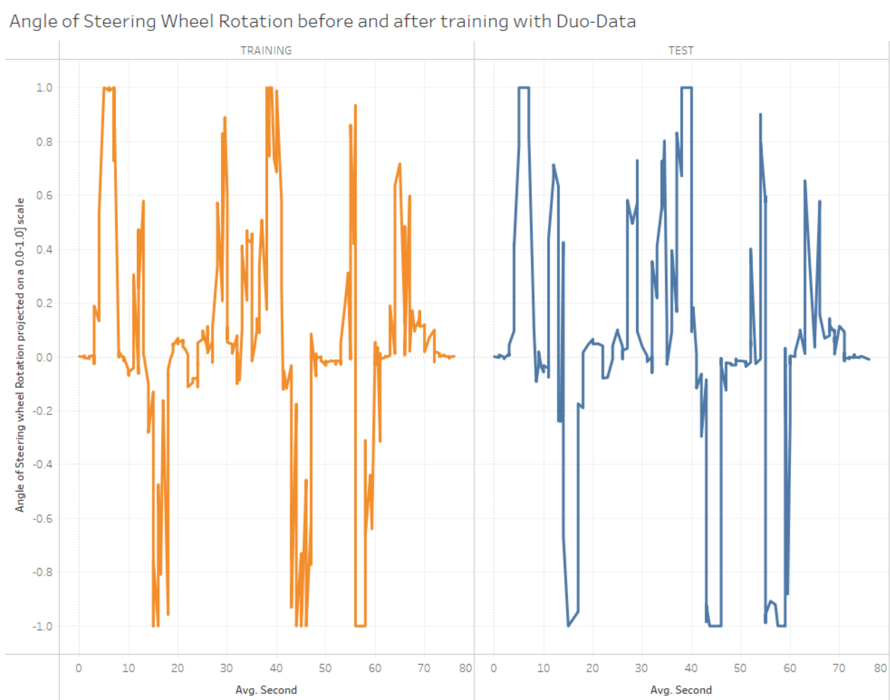

Figure 19: The steering angle during the training and test session

6.2.4 Effectiveness of the feedback system

During the evaluation part of the experiment, the focus was mainly on the use of the feedback application. During this evaluation session, the participants were asked to provide feedback on the interaction with the application and pinpoint out its strengths and weaknesses. This evaluation was done in two parts of which the first part was a general survey on the experiences of the participants. The second part was in the form of an open discussion wherein the participants were asked to brainstorm on potential improvements or to provide further information on their feedback. Due to anonymization, the transcript of the open discussion is not provided in this research. Nonetheless, the transcripts are saved in a database for further development.

Analyzing the data from the survey, it is evident that the application is useful within a specific, predefined, context. However, improvements are required to take place before this application can become a standard for training athletes. As denoted in chapter 2, the main concern of the feedback application is the amount of attention it requires from the drivers. It was stated that the more attention the application requires, the more distracted the drivers become. For this reason, as stated in section 5.3, the application consisted of only three main parameters that can directly be translated into action the driver must take, e.g. turning the steering angle x degrees. The survey yielded that this amount of parameters is sufficient as the driver can directly interpret what he/she must do without getting unnecessarily distracted by the data shown. According to participant player 1, the introduced parameters were the only parameters required to operate the car. Moreover, player 1 stated that the feedback could easily be coupled to the improvements shown by the lap time parameters. Comparing the understandability of the parameters to the other drivers (Figure 20), it is seen that on average, the application requires some time and effort to understand. However, all participants mentioned that after a one to two laps, the parameters were easily understandable within the context of its use.

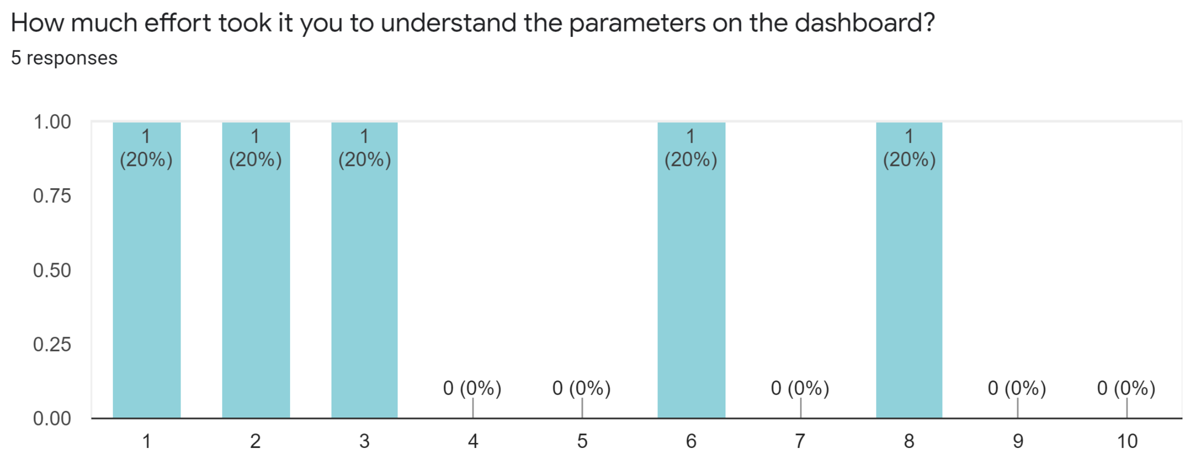

Figure 20: The amount of effort required by the participants to understand the dashboard

When looking at the distraction rate of the application in figure 21.1, 60% of the participants denoted not to be distracted by it. The 40% that did denote getting distracted by the application, mentioned that the distraction was mainly due to not understanding the parameters and the position of the screen. For the experiments, the position of the screen was on the left lower side, right above the clutch. However, as the simulator already contains three 27” inch screens, the act of looking at yet another screen gets more difficult, yielding more distractions.

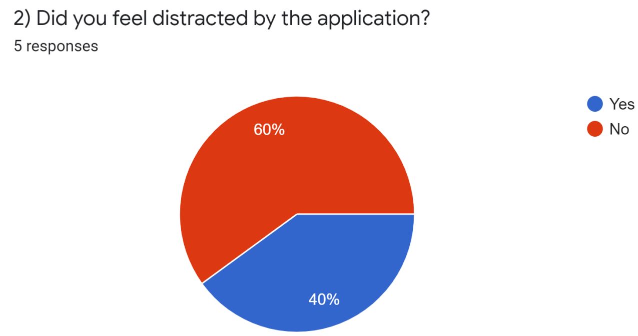

Figure 21.1: The rate of participants feeling distracted by the application

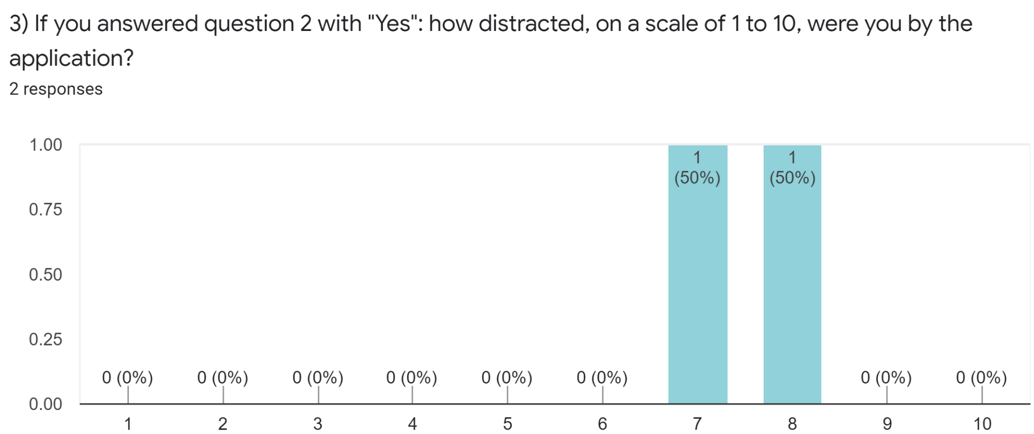

Figure 21.2: The scaling of how distracted the participant was

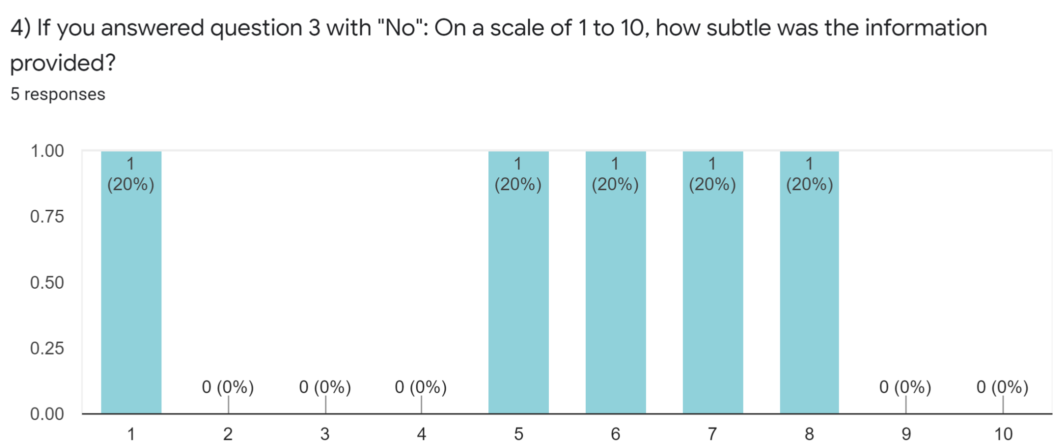

Figure 21.3: The scale of how subtle the participants thought the information was

Analyzing the usefulness of this application within its context, it is seen that most participants would like to see improvements on the type of visualization. The general comment therefore can be made that the interactive dashboard requires too much attention from the driver to be used well. During the open discussions, the participants mentioned several opportunities for integrating this application within a VR device or a potential game overlay to display the feedback in real time without having to use an additional monitor. The current implementation therefore yielded a slightly above average rating when it comes to usefulness with regard to the perception of the simulation and the environment as can be seen in figure 22.

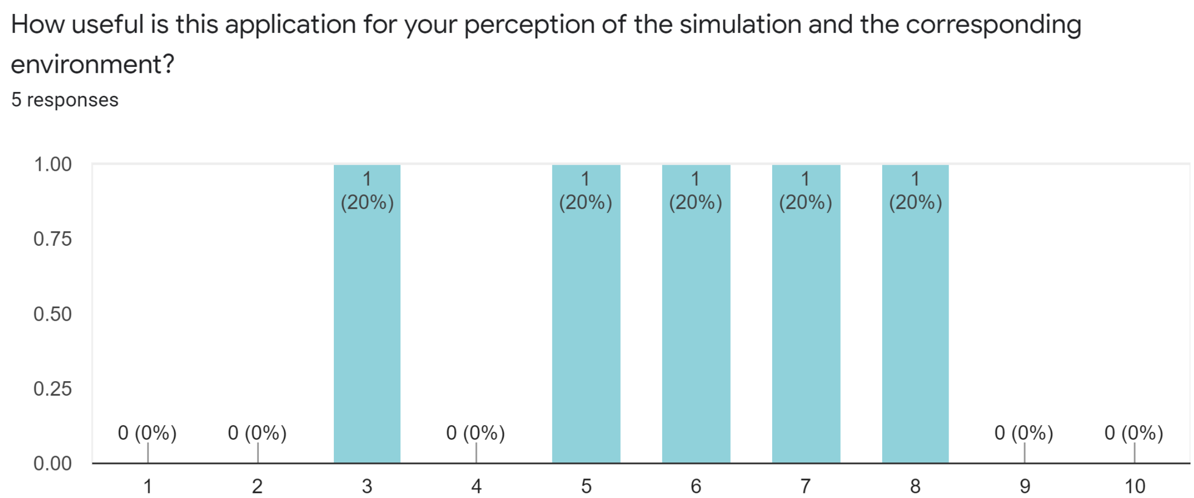

Figure 22: The usefulness of the application

Chapter 7: Discussion and Limitations

When willing to validate the results obtained during the experiments, it is of high importance to be realistic about the effectiveness, the accuracy and the reliability of the system and the tests. Therefore a disclaimer must be made. While the results, as introduced in section 6.2, seem promising and effective, the results lack a statistical backbone. There have only been three iterations with six different drivers in total, among them three drivers with little knowledge on driving and racing simulators. To ensure the quality of the research and whether the feedback system based on process mining and machine learning/statistics works, more statistical testing needs to be done.

Within this section the aim is to answer the sub questions as formulated in section 1.2 by projecting the results obtained in section 6.2 onto the core of these questions. Section 7.1 evaluates the recreation of the trackline wherein the basis for this recreation is the highest performed marker per timestamp and per distance ratio on the track. Next, section 7.2 evaluates on subquestion 2 where the result of the translation is depicted against the initial route over the track. Lastly, section 7.3 covers sub question 3, wherein the communication part gets evaluated.

7.1 Recreation of the ideal track line

As denoted in section 6.2.1, the rate of reading and writing data is of high importance when willing to recreate a track. When looking at the core of the first subquestion as mentioned in section 1.2, the aim is to recreate a new trackline based on the highest performances over the track based on all the output of all drivers. Translating this into technical terms, there is a broad set of data required that includes many positive and negative outliers before an ideal trackline can be created. Although the writing and reading data rate of the ATS system and the learning system was at its maximum throughout the experiments, connection unreliabilities and congestion issues caused gaps within the data set, as depicted in figure 15.1. Therefore a threshold value had to be calculated to maximize the efficiency of the database, due to the read and write limits, to create as many as possible marker points, while keeping the number of packets sent and received from the database at its minimum.

The corresponding research subquestion that belongs to this topic of the research is, as denoted in section 1.2: “In what extent is it possible to recreate an artificial trackline built upon the basis of the highest performances throughout the track?”. When trying to answer this research question, the results that section 6.2.1 yielded showed that an approximate of 8 laps was required to fully cover every second and every driver meter of the track. From the data collected through the racing simulator, it is possible to reconstruct the events, with regard to every parameter of the car as provided by the Codemasters API, that occurred during the moment on track. In this way a data collection can be created to artificially regenerate the track and the position of the car on the track while having every parameter required or not required in mind. Therefore, the extent in which it is possible to recreate the trackline is endless as long as the database allows data to be captured.

The trackline that can be artificially created, can be modified using the classifications as introduced in section 5.2, wherein the highest classifiers per distance ratio or per millisecond on the track can be counted and bundled together with the average, or best, telemetry settings. In this way, the best possible track line can be created as a backbone for the feedback system.

7.2 Translation of data into feedback

The second question that needs to be answered before reliable feedback can be provided to the driver, is the question on the translation of data into advises based on the current telemetry. The corresponding research question to be answered is “How can this artificial trackline be translated into terms of required telemetry changes to guide towards this trackline?”. Given the created trackline, as mentioned in the previous section, the translation of current telemetry data and known telemetry data is simple. Once the boundaries, the averages and the standard deviations of the normal distribution per marker of the telemetry data is calculated by the learning model, a classification will happen to find the best possible marker point with the highest significance, or Z-Value, on this normal distribution.

The translation needs to happen for three parameters only, as denoted by section 5.3. While these three parameters are based on a series of calculations to determine which values of the parameters are actually the value to display, these parameters are easily interpreted as single, rational values. These values can then be translated into advises per marker, as created by the learning system, and automatically be bound to represent the marker in terms of the telemetry settings.

As the disclaimer in the introduction of this section (section 6.3) denoted, the current translation for the amount of data available looks promising, but the translation lacks statistical argumentation to prove that this type of translation will work over time and remain to generate reliable information for the drivers.

7.3 Projection of feedback to driver

The presentation of the feedback might be the most crucial part to evaluate during this research. As denoted by almost all drivers, the type of feedback was not ideal and hence caused a lot of distraction. Evaluating on the manner the feedback is currently presented, many aspects can be improved, among which the conversion of the dashboard into a more integrated or more physical installation. In the long run a digital dashboard would cause confusion to the driver and hence would not work.

Evaluating on the current visualization of the data, it is important to take into consideration the parameters that need to be available to the driver. When looking at the corresponding research question, the question yields: “How can these telemetry changes be communicated to the driver in the most effective manner?”. The core of this question is on how data can be communicated efficiently without too much distraction. Throughout the research, it was concluded that not all types of feedback were ideal within the context of this application, e.g., acoustic feedback was labeled as useful for quick updates but not for continuous feedback, as denoted in section 2.5.2. However, when looking at the survey results, many drivers mentioned that there was some sense of audio interaction missing from the feedback system. When asked what the potential use of this acoustic feedback could be, most participants denoted that acoustic feedback would be more direct and less distracting compared to visual feedback.

As opposed to this statement, the participants did agree that for continuous feedback, visual feedback would be better, provided that the manner of presenting this feedback was more subtle. As denoted in section 6.2.4, a potential implementation of this feedback system would be an integration within the F12020 game or a virtual reality overlay.

7.4 Limitations

Having in mind the scope of this research, up until the current standing, the project seems promising and yields great results. Considering that simulators are already widely used for training professional drivers, an additional layer of training and, eventually, protection is seeming to be the way to conquer the checkered flag. Defining the rate of change based on the current abilities of the drivers within the same team is a unique angle for developing a new training program.

Considering the minimal requirements for setting up this research, many limitations have come to play during the project. These limitations cover a broad list of items that need to be discussed when willing to redo or expand this project. The actual research limitations will be discussed in the next section. The items that this section will cover include:

- The availability of materials
- The budget cap
- Domain Experts
- The reliability of the UDP protocol
- The database limitations in contrast to the data collection size
- The limited researches on the Formula One Topic

Availability of materials
The main concern at the beginning of this project was the availability of materials. The validation of the results required at least a basic type of racing simulator with which a racing environment could be mimicked. This was deemed to be an issue as the racing simulators available were not easily accessible or lacked realism. The corresponding companies that had a racing simulator available, were skeptical about lending it for research purposes and preferred to keep the racing simulators at their own place. Due to the long distance, this was not a feasible option for this project.

Budget Cap
Another option for getting our hands on the required materials for this research was by purchasing the materials. Unfortunately the costs and time required for building a racing simulator seemed to be too high for the purpose of this project. This issue was eventually solved by dr. G.W.J Bruinsma who offered his help and the racing simulator he had available in the ESports Lab at the University of Twente. Later in the project, a new racing simulator was purchased by the ESports Lab which was more reliable and more accurate. This provided the research the required materials to proceed with the project.

Domain Experts
Throughout the research, it was hard to find domain experts to validate the results obtained. After long discussions with dutch racing teams specialized within this domain, the communications took too much time which would make it difficult to agree on a plan within the time available. Moreover, due to the Covid-19 restrictions, the corresponding manufacturers did not agree on visiting the headquarters and refused to provide professional drivers to participate in the research process. To overcome this issue, the expertise of the people involved in the faculty of EEMCS and the people of the ESports Lab was used to validate the results of this project.

Reliability of UDP

The User Datagram Protocol (UDP) is an internet protocol that over time has proven to work well when the bandwidth needs to be maximized. However, due to the unreliable nature of UDP, there was a risk of packet loss and connectivity issues, leading to less complete datasets. Although a logical solution for this would be to implement the Transmission Control Protocol (TCP), this would, in the context of this research, not work. For this project and its corresponding implementation, the maximum bandwidth available must be available with the least amount of delay possible. If this would not be the case, the data might be outdated and therefore the relevancy of the data will drop. Another issue that might rise is the delay in the real time display of the application.

Overcoming the packet loss risk was in this project done by maximizing the upload rate while bundling the data in packets. In this way, if data packets were lost, very soon after a new complete set of data would be sent. Also, more data was harvested than required, this decreased the possibility of missing data.

Database Limitations in contrast to data collection size

The main limitation of this project was the used database. As mentioned in sections 3.2.1.1, a specific set of databases were available. This was due to the read and write speeds and the efficiency of the data storage. Firebase seemed to be the silver lining among the different databases due to its freemium subscription and the, presumed, high freemium limits. Nonetheless, when harvesting data, this freemium limit seemed to become the bottleneck of the project. A lap on the circuit Zandvoort contained approximately 5.5 million data packets. Each of these packets was meant to be sent to the database. However, the limits of firebase only allowed 20.000 document writes per day, which in the context of a lap would mean that we could only record approximately 20 seconds of data per day. This would not even equal one lap of data. As drivers expect a training session to last for at least 60 minutes, a massive downscale in sent data had to be performed to maximize the use of the database. If this project was to be reimplemented, the main change would be the database of use.

Limited Research on Motorsports and Formula One (Drivers)

Research within and among formula one teams is often confidential. Therefore, the amount of papers and research available on public research data banks is close to zero. Throughout the project and during the definition of the state of the art, it was almost impossible to find information on how teams have developed themselves and how drivers are training for maximum performance. During the project, two dutch motorsports teams were contacted to ask questions about how their processes looked like and what their average training schedule looked like. Unfortunately, both teams broke contact and did not respond. Throughout the phase of finding background information, only one paper on the topic of formula one was found, within this paper, the researcher had research on race outcome predictions. Unfortunately, this was not relevant for this research.

Chapter 8: Conclusion and Future work

The aim of this thesis was to find an optimal way of enhancing driver performances by adjusting the training according to gathered data on earlier achieved performances. This was done in a process of three steps. The first step was to harvest data on performances of drivers within a team or cluster. This was done using the Racing Simulator and principles of process mining. The data was stored in a database for later analysis. The second step was to analyze the gathered data with the main purpose of learning the track boundaries, the telemetry boundaries and understanding the abilities of the driver. This was done using conformance checking, basic principles of statistics and linear regression. Lastly, the analysis on the data retrieved from the learning model was translated into valuable feedback and displayed to the driver through a feedback system.

Although no data from real formula one cars could be obtained, the ATS system showed that harvesting process data from formula one simulators is doable and feasible. The corresponding data has such a level of detail that a real life scenario could be recreated and mimicked from the information gathered in the simulator. While having in mind the limitations of certain backend systems, e.g. the database, an intelligent system is designed to harvest and store data. Simultaneously, real time telemetry data was stored with the purpose of performing active conformance checking.

When analyzing the data gathered by the ATS system, several factors play a role to determine whether a created marker is of high value. After this classification of data markers has been made, a training set is created from the currently stored data with which the system trains itself to recognize patterns. According to these created patterns, the system builds advises per second and per distance ratio on the track and bundles this with the corresponding telemetry information. Moreover, the system reads out the real time telemetry database to link the current behavior of the driver to a previously occurring event or a generated marker to optimize the action, and eventually the performance, of the driver.

The results of the experiments conducted with three duo's of drivers were promising. Almost all drivers showed an increase in performance and a rise in confidence. Less fluctuations were observed at the steering wheel, implying more control over the car and a higher understanding of the abilities of the car and above all, the abilities of the driver. Additionally, more peaks in the use of the throttle were observed while the use of the brakes decreased, resulting in more overall pace and performance. This was clearly visible in figure 17. Nevertheless, while these results do imply an effective training concept, the statistical backbone of the project is weak. More experiments must be conducted with a larger sample size to guarantee the effectiveness of the training.

In conclusion, it is not yet possible to guarantee that this manner of training works. The initial concept of the training method appeared to be effective and pervasive, however, the system lacks statistical coverage to prove that this way of training athletes guarantees an improvement in performance.

8.1 Recommendations

It can be said that this training method shows potential as the results obtained look promising. However, to improve the system to make it waterproof, some recommendations must be made. The main recommendation to be made is the system where all participants seemed to have difficulties with; the feedback system. As this feedback system is the main interface for the drivers to interact with, this system must be either optimized in a way that it does not form a distraction or the feedback system must be implemented according to the feedback received from the participants. In further research, I would recommend redesigning the feedback system in a way that it is more visible to the driver with less effort. Furthermore, it is of high importance to keep the information even simpler so that the driver can see or feel in a blink of the eye what is expected. Another recommendation that I deem important is the speed of the database. While the database showed an impressive amount of speed and functionality, the system lacked a bit behind due to the congestion errors that were present by default. The internet connection and the database configuration seemed to be a bottleneck throughout the entire process. Perhaps in future studies, a local database could be implemented to overcome these issues.

Additionally, the learning and analysis method is currently based on the normal distribution. While this classification method seems to work for this context, it is not always reliable. If a car crashes along the way, the entire lap gets classified als a low marker. Neglecting the time that a car is lacking in this situation, the driver might still recover and increase his pace. This increase in pace is currently not counted towards the final classification and hence the data is discarded. Having too much of these data points might corrupt the data. To overcome this, a fully functioning deep learning algorithm can be implemented to recognize events like crashes.

8.2 Future work

To exploit the effectiveness of this training method, these recommendations must be taken into account. Improvements must be made to increase the reliability and the accuracy of the system. Moreover, by conducting more user tests, a statistical and scientific backbone can be created for the training method.

Additionally, although the initial concept relied on machine learning and deep learning principles, the final concept within the scope of this research barely made use of these concepts. For the time allowed to work on this research, the best possible method to apply was linear regression. For future development of this project, machine learning and/or deep learning could be exploited to better understand the obtained data and perhaps give suggestions beforehand instead of in real time. In this way, the system allows the driver to gain information about what is coming so that the driver can prepare himself to perform the action. Also, by the use of machine learning, predictions can get more accurate and more reliable.

Lastly, the method of displaying information must be changed. As denoted in the recommendation section, another manner of providing feedback must be implemented to gain the maximum result while keeping the level of distraction low. The user must be able to see and interpret the information without actively working on processing the data. Currently this is the main weakness of the training method. A potential solution for this issue might be to implement either a virtual reality/augmented reality solution to project the visuals onto the eyesight of the driver. Yet another solution would be to cooperate with the manufacturers of the F1 game to implement the feedback system as part of the in-game overlay feedback.

References

Alves De Medeiros, A. K. (2006). Genetic process mining. (PhD Thesis) Technische Universiteit Eindhoven. https://doi.org/10.6100/IR614016

Andreff W. (2018) Different Types of Manipulation in Sport. In: Breuer M., Forrest D. (eds) *The Palgrave Handbook on the Economics of Manipulation in Sport.* https://doi.org/10.1007/978-3-319-77389-6_2

Abdulwahed, M., Nagy, Z. (2009). Applying Kolb's experiential learning cycle for laboratory education. Loughborough University, 1. https://hdl.handle.net/2134/5412

Adams, J.A., Gopher, D., Lintern, G. (1977). Effects of visual and proprioceptive feedback on motor learning. *Journal of Motor Behavior, 9(1),* 11-22. http://dx.doi.org/10.1177/154193127501900204

Anseel, F., & Lievens, F. (2009). The Mediating Role of Feedback Acceptance in the Relationship between Feedback and Attitudinal and Performance Outcomes. *International Journal of Selection and Assessment, 17*, 362-376. https://doi.org/10.1111/j.1468-2389.2009.00479.x

Adriansyah A., Munoz-Gama J., Carmona J., van Dongen B.F., van der Aalst W.M.P. (2013) Alignment Based Precision Checking. In: La Rosa M., Soffer P. (eds) *Business Process Management Workshops. BPM 2012. Lecture Notes in Business Information Processing, vol 132*. 137-149 Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-36285-9_15

Baxter, R. (2020, September 23). Force Feedback: what is FFB, how does it work and how do I set it up? [Blog post]. Retrieved from https://simracingcockpit.com/setup-force-feedback/

Bos, N. (2021). *Improving the Chess Elo System With Process Mining.* (Bachelor Thesis). University of Twente, Enschede. Retrieved from http://essay.utwente.nl/88571/1/Bos_BA_EEMCS.pdf

Bergmann Drewe, S. (2000). Coaches, Ethics and Autonomy. *Sport, Education and Society. 5(2).* 147-162. https://doi.org/10.26076/3fdf-fab6

Balcerzak, T., Kostur, K. (2018). Flight Simulation in Civil Aviation. *Revista Europa de Derecho de la Navegación Marítima y Aeronáutica, 35(3),* 35-68. Retrieved from https://dialnet.unirioja.es/servlet/articulo?codigo=6953721

Bogarin, A., Cerezo, R., Romero, C. (2017). A Survey on Educational Process mining. *WIREs Data Min. Knowl.Discov, 8.* https://doi.org/10.1002/widm.1230

Brown, A. L., & Kane, M. J. (1988). Preschool children can learn to transfer: Learning to learn and learning from example. *Cognitive Psychology, 20*(4), 493-523.

https://doi.org/10.1016/0010-0285(88)90014-X

Bugeja, K., Spina, S., Buhagiar, F. (2017), "Telemetry-based optimization for user training in racing simulators. *Proceedings of the 2017 9th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games), 9,* 31-38. doi: 10.1109/VS-GAMES.2017.8055808.

Crespo, L. M., & Reinkensmeyer, D. J. (2010). Haptic Guidance Can Enhance Motor Learning of a Steering Task. *Journal of Motor Behavior, 40*(6), 545-557. https://doi.org/10.3200/JMBR.40.6.545-557

De Winter, J.C.F., van Leeuwen, P.M., Happee, R. (2012). Advantages and Disadvantages of Driving Simulators: A Discussion. Retrieved from Delft, University of Technology, Department of BioMechanical Engineering. Website:http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.388.1603&rep=rep1&type=pdf#page=47

Davies, C.A., Vandelanotte, C., Duncan, M.J., van Uffelen, J.G.Z. (2012). Associations of physical activity and screen-time on health related quality of life in adults. *Preventive Medicine, 55(1)*. 46-49. https://doi.org/10.1016/j.ypmed.2012.05.003

Eberhard, K. (2021). The effects of visualization on judgment and decision-making: a systematic literature review. *Management Review Quarterly,* 1-48. https://doi.org/10.1007/s11301-021-00235-8

Earp, J., Persico, D., Dagnino, F., Passarelli, M. (2018). Ethical issues in Gaming: A literature review. Proceedings of the 12th European Conference on Game-Based Learning ECGBL 2018. Retrieved from: https://www.researchgate.net/publication/328829359_Ethical_Issues_in_Gaming_A_Literature_Review

Espié, S., Gauriat, P., Duraz, M. (2005). *Driving Simulators Validation: The Issue of Transferability of Results Acquired on Simulator.* Retrieved from The Université Gustave Eiffel. Website: https://www.researchgate.net/publication/266456992_Driving_Simulators_Validation_The_Issue_of_Transferability_of_Results_Acquired_on_Simulator

Feng, J., & Donmez, B. (2013). Design of Effective Feedback: Understanding Driver, Feedback, and Their Interaction. *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment Training and Vehicle Design,* 404-410. http://dx.doi.org/10.17077/drivingassessment.1519

Federation Internationale de L'Automobile [FIA]. (2021). Formula 1 - Technical Regulations - 2022 - iss 7 - 2021-10-15. Retrieved from https://www.fia.com/regulation/category/110

Formula 1 (n.d.). *Red Bull*. Retrieved from https://www.formula1.com/en/teams/Red-Bull-Racing.htm

Gagne, R. M. (1962). Military training and principles of learning. *American Psychologist, 17(2)*, 83–91. https://doi.org/10.1037/h0048613

Gott, P.G., & Hurter D.A. (1981). The Changing Automotive Industry and the Roles of Its Suppliers. *Proceedings of the International Pacific Conference on Automotive Engineering: SAE Technical Paper 811378,* 1-9. https://doi.org/10.4271/811378

Gouy, I. (2022, Feb. 3). 22.01 Benchmarks Game. *The Computer Language.* Retrieved from https://benchmarksgame-team.pages.debian.net/benchmarksgame/sometimes-people-just-make-up-stuff.html

Hattie, J., Timperley, H. (2007). The Power of Feedback. *Review of Educational Research, 77(1).* 81-112. https://doi.org/10.3102%2F003465430298487

Hays, R.T., Jacobs, J.W., Prince, C., & Salas, E. (2009). Flight Simulator Training Effectiveness: A Meta-Analysis. *Military Psychology, 4(2),* 63-74. https://doi.org/10.1207/s15327876mp0402_1

Heusinkveld Pedals (n.d.). *Sim Pedals Sprint.* Retrieved from https://heusinkveld.com/shop/sim-pedals/sim-pedals-sprint/

Hoppe, D., Sadakata, P., Desain, P. (2006). Development of real-time visual feedback assistance in singing training: a review. *Journal of Computer Assisted Learning, 22(4),* 308-316. https://doi.org/10.1111/j.1365-2729.2006.00178.x

Jafari, N., Adams, K. D., & Tavakoli, M. (2016). Haptics to improve task performance in people with disabilities: A review of previous studies and a guide to future research with children with disabilities. *Journal of Rehabilitation and Assistive Technologies Engineering, 3*, 1-13. https://doi.org/10.1177%2F2055668316668147

Jensen, K (2012, April 26). File:CRISP-DM Process Diagram.png. Retrieved from https://commons.wikimedia.org/wiki/File:CRISP-DM_Process_Diagram.png

Kaplan, K., Mestel, P., Feldman, D.L. (2010). Creating a Culture of Mutual Respect. *AORN Journal, 91(4)*. 495 - 510. https://doi.org/10.1016/j.aorn.2009.09.031

Koo, J., Kwac, J., Ju, W., Steinert, M., Leifer, L., Nass, C. (2014). Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. International Journal on Interactive Design and Manufacturing (IJDeM), 9, 269-275(2015). https://doi.org/10.1007/s12008-014-0227-2

Kumar, M., Kim, T. (2005). Dynamic speedometer: dashboard redesign to discourage drivers from speeding. *Proceedings of the Conference on Human Factors in Computing Systems,* 1573-1576. https://doi.org/10.1145/1056808.1056969

Kühnel, J., Zacher, H., de Bloom, J., Bledow, R. (2016). Take a break! Benefits of sleep and short breaks for daily work engagement. *European Journal of Work and Organizational Psychology, 26(4).* 481-491. https://doi.org/10.1080/1359432X.2016.1269750

Lajunen, T., Summala, H. (1995). Driving Experience, Personality, and skill and safety-motive dimensions in drivers' self-assessments. *Personality and individual differences, 19(3).* 307-318. https://doi.org/10.1016/0191-8869(95)00068-H

Leemans, S. J. J., Fahland, D., Van der Aalst, W. M. P. (2018). Scalable process discovery and conformance checking. *Software & Systems Modeling, 17,* 599-631. https://doi.org/10.1007/s10270-016-0545-x

Lunt, T., Curran, J. (2009). 'Are you listening please?' The advantages of electronic audio feedback compared to written feedback. *Assessment & Evaluation in Higher Education, 35(7),* 759-769. https://doi.org/10.1080/02602930902977772

Mercedes-Benz Group Media (2014, Jul. 14). *Birth of motor sport 120 years ago: Daimler engines dominated first motor car race in 1894*. Retrieved from https://group-media.mercedes-benz.com/marsMediaSite/en/instance/ko/Birth-of-motor-sport-120-years-ago-Daimler-engines-dominated-first-motor-car-race-in-1894.xhtml?oid=9919236

Mjolnir (2022, Feb. 3). *Direct Drive Wheels vs Belt / Gear Driven Wheels.* Retrieved from https://simracingsetup.com/product-guides/direct-drive-wheels-vs-belt-gear-driven-wheels/#What_Is_The_Difference_Between_Belt_Driven_Gear_Driven_Wheels

Nelson, M. M., & Schunn, C. D. (2009). The nature of feedback: how different types of peer feedback affects writing performance, *Instructional Science, 37*, 375-401. https://doi.org/10.1007/s11251-008-9053-x

Pakkanen, T., Raisamo, R., & Surakka, V. (2014) Audio-Haptic Car Navigation Interface with Rhythmic Tactons. In: Auvray M., Duriez C. (eds) Haptics: Neuroscience, Devices, Modeling, and Applications. EuroHaptics 2014. Lecture Notes in Computer Science, vol 8618. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-44193-0_27

Rozinat, A., Van der Aalst, W. M. P. (2008). Conformance Checking of processes based on monitoring real behavior. *Information Systems, 33(1),* 64-95. https://doi.org/10.1016/j.is.2007.07.001

Rozinat, A., Alves de Medeiros, A.K., Günther, W., Weijters, A. J. M. M., Van der Aalst, W. M. P. (2007).

Towards an evaluation framework for process mining algorithms. *BETA Working paper series,* 1-20. Retrieved from https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.668.6889&rep=rep1&type=pdf .

Sano, S. (2014). Automotive engineering - the future. *International Journal of Vehicle Design*, *15(1-2),* 143-146. Retrieved from https://www.inderscienceonline.com/doi/abs/10.1504/IJVD.1994.061912

Slob, J. (2008). *State-of-the-Art Driving Simulators, a Literature Survey*. Retrieved from The University of Eindhoven, Department of Mechanical Engineering, Control Systems Technology Group. Website: http://www.mate.tue.nl/mate/pdfs/9611.pdf

Skinner, B.R. (2013). The Relationship between Confidence and Performance Throughout a Competitive Season. *All Graduate Plan B and Other Reports*. *285.* https://doi.org/10.26076/3fdf-fab6

Socha, V., Socha, L., Szabo, S., Hana, K., Gazda, J., Kimlickova, M., … , Rozenberg, R. (2016). Training of Pilots Using Flight Simulator and its Impact on Piloting Precision. *Proceedings of the 20th International Scientific Conference on Transport Means,* 374-379. Retrieved from http://idoaba.eu/567/pub16/lf/174425.pdf

Sukhov, A. (2019). Ethical Issues of Simulation Video Games. *Proceedings of the 13h European Conference on Games Based Learning. The University of Southern Denmark. 3-4 October, 2019.* 705–712. https://doi.org/10.34190/GBL.19.060

Tromp, N., Hekkert, P., Verbeek, P.-P (2011). Design for Socially responsible behavior: A Classification of influence based on intended user experience. *Design issues, 27(3)*. 3-19. http://dx.doi.org/10.1162/DESI_a_00087

Tronsmoen T., (2010). Associations between driver training, determinants of risky driving behaviour and crash involvement. Safety Science, 48(1). 35-45. https://doi.org/10.1016/j.ssci.2009.05.001

Turing, A.M. (1950). Computing Machinery and Intelligence. *Mind, LIX(236),* 433-460. https://doi.org/10.1093/mind/LIX.236.433

Van der Aalst, W. (2012). Process mining. *Communications of the ACM, 55(8),* 76-83. https://doi.org/10.1145/2240236.2240257

Van de Aalst, W., Weijters, A.J.M.M., Maruster, L. (2004). Workflow Mining: Discovering Process Models from Event Logs. *IEEE Transactions on Knowledge and Data Engineering, 16(9),* 1128-1142. http://dx.doi.org/10.1109/TKDE.2004.47

Van der Aalst, W., et al. (2012) Process Mining Manifesto. In: Daniel F., Barkaoui K., Dustdar S. (eds) *Business Process Management Workshops. BPM 2011. Lecture Notes in Business Information Processing, 99*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-28108-2_19

Väänänen-Vainio-Mattila, K. V., Heikkinen, J., Farooq, A., Evreinov, G. E., Mäkinen, E., & Raisamo, R. (2014). User experience and expectations of haptic feedback in in-car interaction. In *Proceedings of the 13th International Conference on Mobile and Ubiquitous Multimedia (MUM '14),* 248-251. https://doi.org/10.1145/2677972.2677996

Van Houten, R., Nau, P.A. (1983). FEEDBACK INTERVENTIONS AND DRIVING SPEED: A PARAMETRIC AND COMPARATIVE ANALYSIS. *Journal of Applied Behaviour Analysis, 16(3), 253-281.* https://doi.org/10.1901/jaba.1983.16-253

Van der Poel, I. Royakkers, L. (2007). The Ethical Cycle. *Journal of Business Ethics, 71.* 1-13. https://doi.org/10.1007/s10551-006-9121-6

Voelkel, S., & Mello, L.V. (2014). Audio Feedback - Better Feedback? *Bioscience Education*, *22(*1), 16-30. https://doi.org/10.11120/beej.2014.00022

Weijters, A.J.M.M., Van der Aalst, W. (2003). Rediscovering Workflow Models from Event-based Data using Little Thumb. *Integrated Computer-Aided Engineering, 10(2),* 151-162. http://dx.doi.org/10.3233/ICA-2003-10205

Weerapong, S., Porouhan, P., Premchaiswadi, W. (2012). Process mining using α-algorithm as a tool (A case study of student registration). *Proceedings of the 2012 Tenth International Conference on ICT and Knowledge Engineering, 2012,* 213-220. https://doi.org/10.1109/ICTKE.2012.6408558

Wen, L., Wang, J., Sun, J. (2006). Detecting Implicit Dependencies Between Tasks from Event Logs. In: Zhou X., Li J., Shen H.T., Kitsuregawa M., Zhang Y. (eds) *Frontiers of WWW Research and Development - APWeb 2006. APWeb 2006. Lecture Notes in Computer Science, vol 3841.* (pp 591-603). Springer, Berlin, Heidelberg. https://doi.org/10.1007/11610113_52

Wienen, H. C. A., Bukhsh, F. A., Vriezekolk, E., & Wieringa, R. J. (2017). *Accident Analysis Methods and Models - a Systematic Literature Review.* (CTIT Technical Report; No. TR-CTIT-17-04). Retrieved from the University of Twente, Centre for Telematics and Information Technology (CTIT) Website:https://research.utwente.nl/en/publications/accident-analysis-methods-and-models-a-systematic-literature-revi

Wassink, L., Van Dijk, B., Zwiers, J., Nijholt, A., Kuipers, J., Brugman, A. (2006). In the truman show: Generating dynamic scenarios in a driving simulator. *IEEE Intelligent Systems, 21(5),* 28-32, doi: 10.1109/MIS.2006.97

Appendices

Appendix 1: Search Terms

[Formatting → "Search Term (no. results; Usability *scale 1-5*)"]

General Search terms on the topic

Search Terms	Engine	Amount of results
"Formula One" "mining"	Google Scholar	4.550
" Formula one" "Driver"	Google Scholar	8.210
"Process mining" ranking games	Google Scholar	1.080
"Process mining" motorsport	Google Scholar	103
"Process mining" games	Google Scholar	3.550
"Process mining" Ranking	Google Scholar	5.310
"Formula 1 drivers"	Google Scholar	327
"driver training simulator"	Google Scholar	73.200

Table 6: The search Terms

Search terms on Feedback Systems

"Audio feedback", "Auditory feedback", Visual Feedback", "Physical Feedback", "Cognitive feedback", "informal feedback", "communicational feedback", "Written feedback", "VR", "Simulator", "Heads up display", "Educational Feedback", "Corrective Feedback", "Feedback", "Driving", "Distracted driving", "Haptic", "Haptic Feedback", "Improper feedback", " Visual feedback preservice teachers", "car feedback haptic", "overload visual feedback", "visualization", "physical feedback steering wheel", "accepting feedback criticism", "Overload of information".

From these search terms, search queries can be derived:

- Audio Feedback OR Visual Feedback
- Audio Feedback OR Visual Feedback AND Corrective feedback
- Haptic Feedback AND overload of information
- Etcetera.

Appendix 2: Packet data

Packet IDs

The packets IDs are as follows

Packet Name	Value	Description
Motion	0	Contains all motion data for player’s car – only sent while player is in control
Session	1	Data about the session – track, time left
Lap Data	2	Data about all the lap times of cars in the session
Event	3	Various notable events that happen during a session
Participants	4	List of participants in the session, mostly relevant for multiplayer
Car Setups	5	Packet detailing car setups for cars in the race
Car Telemetry	6	Telemetry data for all cars
Car Status	7	Status data for all cars such as damage
Final Classification	8	Final classification confirmation at the end of a race
Lobby Info	9	Information about players in a multiplayer lobby

Event String Codes

Event	Code	Description
Session Started	“SSTA”	Sent when the session starts
Session Ended	“SEND”	Sent when the session ends
Fastest Lap	“FTLP”	When a driver achieves the fastest lap
Retirement	“RTMT”	When a driver retires
DRS enabled	“DRSE”	Race control have enabled DRS
DRS disabled	“DRSD”	Race control have disabled DRS
Team mate in pits	“TMPT”	Your teammate has entered the pits
Checkered flag	“CHQF”	The checkered flag has been waived
Race Winner	“RCWN”	The race winner is announced
Penalty Issued	“PENA”	A penalty has been issued – details in event
Speed Trap Triggered	“SPTP”	Speed trap has been triggered by fastest speed