National College of Ireland

Synthetic Attack Evaluation of Anomaly Based Intrusion Detection System for Formula 1 Telemetry

MSc Research Project
MSCCYBE_JANO24_O

Niamh Daly
Student ID: x19370553

School of Computing
National College of Ireland

Supervisor: Ross Spelman

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: ……Niamh Daly……

Student ID: ……x19370553……..……

Programme: ……MSCCYBE_JANO24_O……**Year:** ……2……..

Module: ……MSc (Research) Practicum/Internship Part 2……..……

Supervisor: ……Ross Spelman……..……

Submission Due Date: ……11th December 2025 2pm……..……

Project Title: Synthetic Attack Evaluation of Anomaly Based Intrusion Detection System for Formula 1 Telemetry

Word Count: ……9200…… **Page Count**……23……..……..

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

Signature: …Niamh Daly……

Date: …8th December 2025……

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

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

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

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

Synthetic Attack Evaluation of Anomaly Based Intrusion Detection System for Formula 1 Telemetry

Niamh Daly
x19370553

Abstract

Telemetry in Formula 1 provides the foundation for both race strategy and driver safety, transmitting thousands of data points per second from cars to engineers in real time. While essential, these data streams also represent a potential attack surface for adversaries capable of intercepting, manipulating, or delaying telemetry. Malicious tampering could mask critical safety issues such as brake overheating or tyre degradation, or mislead teams into making flawed strategic decisions. Despite the importance of telemetry to modern motorsport, there has been limited academic attention to its cybersecurity. This project investigates how anomaly based intrusion detection systems (IDS) can be applied to Formula 1 telemetry to detect unauthorised tampering in real time. Using open source telemetry data obtained via the FastF1 library, synthetic attack scenarios were simulated, including false value injection, delayed transmissions, and packet loss. A range of detection methods were implemented, combining physics informed consistency checks with machine learning approaches such as Principal Component Analysis and Isolation Forest. This work contributes both academically, by extending cybersecurity research into the underexplored context of motorsport telemetry, and practically, by outlining a framework for integrating IDS into Formula 1 operations.

Keywords: Anomaly Detection, Machine Learning, Formula 1, Cyber Physical Systems (CPS), IDS

1 Introduction

Formula 1 is widely recognised as the pinnacle of motorsport, defined by its technological sophistication and the demanding intensity of its competitive environment. At speeds exceeding 350km/h, drivers rely not only on their skill but also on a continuous stream of telemetry data transmitted in real time between car and pit wall. This data driven decision making underpins critical safety and performance strategies, providing teams with a decisive competitive edge. Split second assessments by analysts often involve detecting brake overheating, monitoring tyre degradation or optimising pit stop planning. If such telemetry were to be tampered with, even subtly, the consequences could be catastrophic: a masked brake failure[1], a miscalculated fuel strategy[2], or a delayed tyre change that could endanger not only the driver but also those around them[3].

Formula 1 is a sport in which both safety and performance are critically dependent on telemetry[4]. Each car transmits thousands of data points per second from on board sensors, to engineers present either at the track, or remotely at team headquarters. Over the course of a

[1] https://www.formula1.com/en/latest/article/brake-disc-failure-to-blame-for-leclerc-hartley-collision.3Kl1jE4UV2CMaiwA8yEw8i
[2] https://www.racefans.net/2019/12/01/leclerc-keeps-abu-dhabi-gp-podium-after-fuel-penalty-investigation/
[3] https://www.bbc.com/sport/formula1/22294880
[4] https://www.watsonpost.com/telemetry-in-formula-one/

race weekend, a single car can generate more than 1 terabyte of data[5]. This information is crucial for making decisions relating to both race strategy and driver safety. The confidentiality, integrity, and availability of telemetry are therefore essential[6].

In recent years, motorsport has increasingly faced the same cybersecurity challenges that have been observed in other cyber physical systems (CPS) such as autonomous vehicles, aviation, and industrial IoT[7]. During the 2012 Formula 1 season, the Bahrain Grand Prix unfolded under heavy political tension, with protestors challenging the events legitimacy while calling attention to human rights concerns[8]. During this same period, the official Formula 1 website was briefly hacked using a denial of service attack by activists seeking to disrupt the race and draw global attention to the unrest[9]. Although the site compromise was short lived, it exposed the sports susceptibility to digital interference. Whether for activism, sabotage, or competitive disruption, cyberattacks on motorsport platforms could impact team strategy, competitive integrity, and even on track safety, making cybersecurity as critical to modern racing as aerodynamic efficiency or tyre management.

While academic research has addressed cybersecurity in connected vehicles and industrial control systems, very little work has examined the protection of telemetry in motorsport, particularly Formula 1. This gap in the literature presents an opportunity to study Formula 1 as a real world case study for anomaly based intrusion detection in safety critical environments. The motivation for this project stems from both academic and industrial needs. Academically, it extends cybersecurity research into a new and underexplored application domain, contributing to the literature on anomaly detection in high frequency, multidimensional telemetry streams. Industrially, it offers a possible basis that motorsport organisations might draw on to enhance their cyber resilience against various potential threats. Beyond Formula 1, the findings have the potential to generalise to other cyber physical systems where secure telemetry is vital for human safety. My personal motivation for this research is rooted in a long standing personal interest in Formula 1, having followed the sport closely for many years and attended several races. This passion has provided not only an appreciation of the technological and strategic complexity of the sport, but also an awareness of the critical role telemetry plays in both performance and safety.

1.1 Research Question and Objectives

The central research question guiding this work is: How can unsupervised, anomaly based intrusion detection be applied to Formula 1 telemetry streams, and to what extent can hybrid statistical and physics models detect synthetic tampering attacks in real time. To address this, the project will pursue the following objectives:

---

[5] https://medium.com/formula-one-forever/data-the-unseen-driver-in-formula-1-cars-63f31c16f2fe

[6] https://www.techtarget.com/whatis/definition/Confidentiality-integrity-and-availability-CIA

[7] https://www.cloudsek.com/blog/racing-into-danger-advanced-cyber-threats-targeting-formula-1-fans-and-teams-ahead-of-the-dutch-grand-prix

[8] https://www.motortrend.com/news/hackers-take-down-f1-site-due-to-protests-in-bahrain-195785

[9] https://phys.org/news/2012-04-hackers-f1-websites-bahrain-gp.html

1. Analyse the cyber threat landscape of Formula 1 telemetry, focusing on manipulation attacks such as false data injection, signal freezing, delay, and packet loss.
2. Construct a synthetic attack injection framework to introduce controlled tampering events into real telemetry data using the FastF1 dataset.
3. Develop an unsupervised anomaly based intrusion detection system combining statistical learning (PCA + Isolation Forest) with lightweight physics based consistency rules.
4. Evaluate detection performance in terms of true positive rate, false positive control (via calibrated thresholding), and anomaly type coverage across multiple tampering scenarios.
5. Propose a hybrid IDS architecture for motorsport telemetry security, including layered detection (statistical and physics) and real time operational considerations.

The study aims to demonstrate that unsupervised anomaly based intrusion detection can be applied to Formula 1 telemetry streams to detect tampering in real time. Open source FastF1 telemetry data will be used as a baseline for the analysis, and synthetic adversarial perturbations, including sensor falsification, packet freeze events, and timing jitter, will be injected to simulate real world cyber physical attacks. The system is to be trained exclusively on clean race telemetry, reflecting operational constraints where labelled attack samples are unavailable.

A hybrid approach will be adopted, combining statistical learning (Principal Component Analysis and Isolation Forest) with lightweight physics based consistency checks (e.g., speed RPM gear agreement, and freeze detection). These choices will allow the IDS to capture both statistical deviations and physically implausible behaviour. Model effectiveness is to be assessed in terms of true positive rate at controlled false positive rates, anomaly type coverage, and real time feasibility.

Results aim to demonstrate that the hybrid approach can detect a range of tampering behaviours, particularly abrupt or correlated anomalies, while physics checks should significantly improve detection of subtle or sustained manipulation such as brake loss or signal freezing. These findings will support the viability of unsupervised hybrid IDS for safety critical motorsport telemetry.

1.2 Report Structure

The remainder of this research paper is organised as follows:

- **Literature Review**: The paper begins by reviewing existing research on motorsport telemetry security, cybersecurity in cyber physical systems (CPS), telemetry in motorsport, and anomaly based intrusion detection techniques.
- **Research Methodology**: Will then describe the research approach, including the use of open source Formula 1 telemetry data and synthetic tampering techniques. The rationale for employing unsupervised hybrid anomaly detection will also be outlined.
- **Design Specification:** The architecture of the intrusion detection pipeline is outlined

- **Implementation:** Technical details of the implementation, including anomaly injection framework, model training workflow, and run time consideration for real time suitability
- **Evaluation**: Experimental setup is presented and the model performance is evaluated across multiple anomaly types, reporting true positive rates under fixed false positive thresholds, and fusing physics based rules with statistical detection.
- **Conclusion and Future Work**: Summarises the findings of the project, the practical viability for motorsport telemetry monitoring, and proposes future research directions, and the limitations of the project.

2 Literature Review

The literature review provides the foundation for this research by situating Formula 1 telemetry security within the broader field of cybersecurity in cyber physical systems (CPS). It surveys existing research across five main domains: cybersecurity in CPS, human and sociotechnical context in F1, telemetry in motorsport, cyber threats in motorsport, and intrusion detection.

2.1 Cybersecurity in Cyber Physical Systems (CPS)

Cyber physical systems (CPS) are systems that combine computational components with physical processes, enabling seamless interaction between the digital and physical worlds[10]. Systems such as the Internet of Things (IoT), industrial control systems (ICS), and vehicular networks exemplify CPS and have become key subjects of cybersecurity research. This research highlights their inherent vulnerabilities, which stem from their dependence on real time communication, distributed architectures, and the integration of physical and digital elements.

(Longo, 2025) investigates cybersecurity challenges in computerised transportation CPS, focusing on maritime and aviation domains. By integrating theoretical analysis with testbed experimentation, the study identifies several previously unknown vulnerabilities in automation, radar, and collision avoidance systems. The study highlights the importance of systematic testing, attack simulation, and mitigation strategies, providing methodological guidance for developing anomaly based intrusion detection tailored to safety critical, real time telemetry environments such as motorsport, although its domain specificity limits direct applicability. Similarly, (Westbrook, 2025) examines cybersecurity challenges in autonomous vehicle (AV) systems. The study highlights the risks to AV systems, such as systems vulnerabilities, data breaches, and threats to public safety that necessitate robust cybersecurity frameworks and regulatory measures to secure AV based CPS. Although largely conceptual and lacking empirical validation, this study contextualises motorsport telemetry security within broader CPS challenges.

(Pervez, et al., 2020) examined cybersecurity vulnerabilities in modern day automotive CPS, focusing on internal vehicle networks such as the Controller Area Network (CAN). Their study identified the absence of message authentication as a critical weakness, leaving vehicles susceptible to malicious attacks and unauthorised data manipulation. Research into CPS and

[10] https://www.fortinet.com/resources/cyberglossary/cyber-physical-systems

connected vehicles demonstrates parallels with Formula 1 telemetry, particularly in terms of high frequency data exchange, safety critical decision making, and exposure to adversarial manipulation. These insights provide transferable methodologies for identifying threats and developing defensive measures in motorsport.

2.2 Human and Sociotechnical Context in Formula 1 Cyber Physical Systems

Beyond technical mechanisms, the resilience of Formula 1 cyber physical systems (CPS) is influenced by human, organisational, and sociotechnical factors. (Mishra, 2018) highlights the role of tacit knowledge transfer within Formula 1 teams, demonstrating how expertise and collaboration enhance adaptability in these high performance environments. The study shows how organisational and human factors play a crucial role in ensuring system resilience, revealing that individuals with high technical and managerial expertise significantly influence team performance. These insights underscore that effective cybersecurity and anomaly detection depend not only on advanced technical systems but also on the efficient flow of knowledge and decision making across Formula 1 teams

Complementing this, (Evans, 2013) examines how Formula 1 is constructed as a live mediatised spectacle through broadcast narratives and visual representations. This external perspective reveals how telemetry and data systems operate under intense public scrutiny, where reliability and integrity are essential not just for safety and performance but also for maintaining the sport's credibility. (Kyösti, 2023) demonstrated how Formula 1 race outcomes have measurable economic effects, with race victories influencing short term stock market performance. Together these studies show that Formula 1 CPS exist within a broader ecosystem where technical performance, human expertise, media exposure, and market perception all interact. A telemetry breach or anomaly isn't just a technical failure it can ripple into public image and financial consequences.

2.3 Telemetry in Motorsport

Telemetry plays a central role in modern motorsport, enabling the transfer of thousands of data points per second from a car's onboard sensors to trackside engineers and remote team headquarters[11]. While dedicated research on telemetry security in motorsport is limited, existing studies outline the operational and analytical significance and structure of these systems in Formula 1.

(Komov, 2024) provides an overview of the software ecosystem underpinning Formula 1 performance, describing how telemetry data is gathered from onboard sensors, transmitted via the ECU and CAN bus, and processed using tools such as ATLAS for real time and post race analysis. The study highlights telemetry's central role in decision making as well as driver training, and strategy development, highlighting the teams reliance on continuous, high fidelity data streams. (Msakamali, 2024) further investigates telemetry in Formula 1 and its use with performance data to inform race strategies. The study utilises the Fast F1 and Anaconda frameworks to analyse performance and driver telemetry across case studies. The findings outlined that precise race strategy predictions were a challenge, however it demonstrated that

[11] https://www.jalopnik.com/formula-one-telemetry-explained-1845819678/

telemetry insights are crucial for tactical decision making race outcome differentiation. (Akin, 2024) analyses driver input and vehicle telemetry to reveal meaningful behavioural relationships between throttle, brake, gear, and RPM signals that inform race strategy and performance. The work combines domain engineering insight with semi unsupervised modelling, and while not cyber security focused, the study highlights the need for domain informed multivariate analysis to detect behavioural or system deviations. These studies establish the baseline operational and analytical context of Formula 1 telemetry, outlining the environment in which intrusion detection systems (IDS) must operate. Their collective emphasis on the dependence on accurate, uninterrupted data highlights a significant research gap. Despite telemetry's essential role in performance and safety, there is limited academic investigation into methods for protecting these data streams against tampering or cyberattacks in real time.

2.4 Cyber Threats in Motorsport and Sporting Events

The motorsport industry faces increasing exposure to cyber threats. With documented cases ranging from espionage, such as the theft of confidential performance data[12], to acts of sabotage[13], including ransomware attacks and targeted system disruptions[14]. High profile incidents include the Marussia Formula 1 teams infection with a Trojan virus[15] and phishing campaigns targeting the FIA and personal driver data[16], illustrate the sectors exposure to both targeted and opportunistic threats. These cases highlight cybersecurity as a critical and emerging concern within the motorsport ecosystem.

(Lambert, 2022) provides a retrospective analysis of cybersecurity breaches in Formula 1, identifying human error, phishing, insider threats, and insecure telemetry systems as the recurring sources of vulnerability. The study highlights the importance of anomaly detection as a defence against cyber incidents. With the analysis being solely on publicly reported cases, there is limited quantitative depth, however the findings offer practical insights into real world vulnerabilities within the domain. (Bongiovanni, et al., 2024) broadens the focus beyond motorsport, providing an integrative review of cybersecurity across major sporting events. The study identifies key challenges faced by sporting organisations, including infrastructure vulnerabilities, data breaches, and the inherent lack of tailored cybersecurity measures suited to the complex, real time operational demands of high profile sporting events. These insights position Formula 1 within the broader landscape of sports cybersecurity, where real time data exchange and global visibility amplify both operational vulnerabilities and reputational risks.

2.5 Intrusion and Anomaly Detection in Automotive Telemetry Systems

In data intensive environments like Formula 1, traditional security methods are insufficient to protect high speed, real time telemetry streams. Investigations into intrusion and anomaly detection within telemetry systems reveals various different approaches for identifying

[12] https://www.bbc.com/sport/formula1/63575321
[13] https://techinformed.com/accelerating-cybersecurity-in-the-world-of-motorsport-formula-one/
[14] https://www.racingpost.com/news/britain/bha-believed-to-be-latest-organisation-to-be-hit-by-a-cyber-attack-arubK5n9AzDw/
[15] https://www.techcentral.ie/marussia-f1-testing-disrupted-by-trojan-malware/
[16] https://www.grandprix.com/news/fia-confirms-f1-driver-data-hacking-breach.html

abnormal patterns, thereby maintaining data integrity and ensuring reliable system performance under potential cyber threats.

(von Schleinitz, et al., 2021) proposed the VASP framework, a hybrid anomaly detection approach that combines a variational autoencoder (VAE) based detector with LSTM forecasting, to identify anomalies in multivariate time series telemetry data from motorsport. The system detects and selectively repairs corrupted telemetry signals, leaving intact signals untouched, to maintain forecasting accuracy on clean data. In some areas the accuracy was improved up to 70%. The framework was designed for high frequency and time dependent telemetry streams and demonstrates how anomaly based intrusion detection can enhance data integrity and forecasting reliability in motorsport.

In another study (Carratù, et al., 2025) proposed a generative adversarial network (GAN) for anomaly detection in automotive time series sensor data. This study is specifically applied to motorcycle suspension stroke sensors. The architecture consists of two competing neural networks, a generator for reproducing realistic sensor signals, and a discriminator for determining real measures from generated ones. This work has demonstrated the effectiveness of deep generative models for time series sensor data through reconstruction based scoring and improved false positive reduction. (Widanage, et al., 2019) developed a real time anomaly detection framework for IndyCar telemetry using Hierarchical temporal memory (HTM) algorithm and deployed on Apache Storm. The system analysed multichannel metrics such as speed, RPM, and throttle, for detecting race events and irregularities. The model was focused on operational anomalies rather than cybersecurity, however, the study was able to demonstrate a scalable and low latency architecture with high relevancy to intrusion detection in motorsport telemetry.

2.6 Research Gap

The existing literature aligns around key insights that directly inform this project. Studies such as (Lambert, 2022) and (Bongiovanni, et al., 2024) highlight that cybersecurity in sporting events and more specifically motorsport are shaped by both human and technical vulnerabilities, emphasising the need for proactive data driven protection strategies. Within cyber physical systems (CPS), the research conducted by (von Schleinitz, et al., 2021), (Widanage, et al., 2019), and (Carratù, et al., 2025) demonstrate that telemetry is inherently multivariate and time dependent, thus meaning that anomaly detection models must consider correlations between features over time. As labelled intrusions are rare in these domains, unsupervised learning approaches are the best suited for effective anomaly detection and analysis, as recommended in (Pervez, et al., 2020) and as (Akin, 2024) and (von Schleinitz, et al., 2021).

Given the characteristics of the data in this domain, hybrid intrusion detection architectures are identified as the most suitable approach, demonstrating greater robustness and adaptability than the single method systems (von Schleinitz, et al., 2021) (Widanage, et al., 2019). The literature review additionally supports the use of controlled attack simulation for evaluation of models. (Longo, 2025) provides a conceptual taxonomy of CPS attack types, while (Widanage, et al., 2019) and (von Schleinitz, et al., 2021) validate synthetic anomaly injection as an experimental methodology. Despite these studies, there remains a gap in research applying such techniques to Formula 1 telemetry data under controlled cyber attack conditions. This

project addresses the gap by implementing and testing a hybrid IDS architecture on Formula 1 telemetry data, using simulated attack conditions to assess its effectiveness in protecting data integrity and system reliability within high speed motorsport environments.

3 Research Methodology

The following section outlines the chosen research procedure and evaluation methodology adopted to design, implement, and evaluate whether unsupervised anomaly based intrusion detection techniques can reliably identify synthetic tampering in Formula 1 telemetry streams.

The research adopts a design science methodology to investigate whether unsupervised anomaly based intrusion detection can detect synthetic tampering in Formula 1 telemetry streams, combined with an experimental evaluation using controlled attack injection on real telemetry. The design science methodology focuses on creating and evaluating innovative artefacts to solve real world problems via a structures cycle of problem identification, artefact design, implementation and evaluation, thus is a suitable methodology for the following research[17]. Guided by cyber physical system (CPS) security literature, (Longo, 2025) and (Pervez, et al., 2020) this dual approach reflects common practice by developing a lightweight intrusion detection system (IDS) and evaluating it using real telemetry enriched with controlled adversarial perturbations. The study uses real publicly available Formula 1 telemetry collected via FastF1 Python API[18], and introduces targeted cyber physical anomalies to simulate adversarial manipulation. The IDS pipeline combines Principal Component Analysis (PCA) for multivariate feature compression[19], Isolation Forest (IF) for unsupervised anomaly scoring[20], and physics based consistency checks to detect mechanically implausible states[21]. Models are trained exclusively on clean telemetry and evaluated on tampered data to replicate real world operational constraints where labelled attack data is unavailable.

Detection performance is assessed using true positive rate at fixed false positive rate thresholds and event level detection success**,** consistent with unsupervised anomaly detection evaluation standards[22] [23]. Per the design science methodology, the aim is to ensure that the results are reproducible. A lightweight, real time capable detection pipeline is prioritised to reflect the latency requirements of motorsport telemetry systems. The below sections additionally outline the stages of the design science methodology, including the datasets, attack simulation framework, IDS architecture, and evaluation design.

3.1 Data Collection and Compilation

The FastF1 Python library provides access to publicly broadcast Formula 1 telemetry and timing data via the API. This data accurately reflects the structure and behaviour of real racing

[17] https://medium.com/@yassin.lazar/design-science-research-methodology-4577f732a1fa

[18] https://docs.fastf1.dev/

[19] https://www.geeksforgeeks.org/data-analysis/principal-component-analysis-pca/

[20] https://www.geeksforgeeks.org/machine-learning/what-is-isolation-forest/

[21] https://aithor.com/essay-examples/taking-it-to-the-limit-physics-in-formula-one-car-design#2-the-role-of-physics-in-formula-one-car-design

[22] https://www.iguazio.com/glossary/true-positive-rate/

[23] https://www.harness.io/harness-devops-academy/false-positive-rate

data and is obtained from official broadcast timing feeds. The data is made available for non commercial, educational, and research purposes under the terms of the FastF1 project[24]. A summary of the dataset size is outlined in Table 1, with the total number of attributes listed.

The data points selected for analysis are outlined in Table 2. The dataset was generated by selecting specific session parameters within FastF1: session type (e.g., Race, Practice, Qualifying), event location (e.g., Monaco, Abu Dhabi), and year. Given that the Bahrain Grand Prix was previously targeted in a cyber attack in 2012, as discussed above, the 2024 Bahrain Grand Prix Race session was chosen for this analysis. The 2024 Bahrain race represents the most recent Bahrain race from a fully completed Formula 1 season and therefore provides a relevant, contemporary reference point for examining telemetry in a security focused context. The telemetry data was taken for all 20 drivers during that session and then preprocessed into a clean, synchronised, multivariate time series dataset in preparation for the analysis. These data points were chosen because they exhibit strong physical correlations and are commonly referenced in vehicular anomaly detection literature (Msakamali, 2024), (Akin, 2024). The relationships between the telemetry features are outlined in Figure 1 via the correlation heatmap.

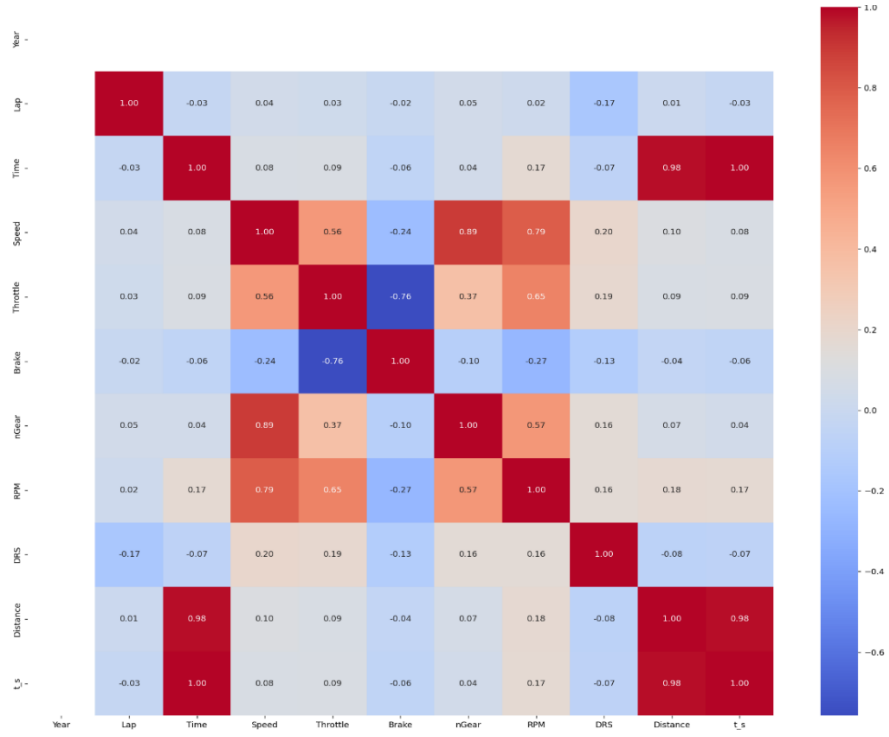

Figure 1: Correlation Heatmap on Chosen Telemetry

Table 1: Summary of Dataset

Race	Year	No. of Attributes	No. of Instances	Categorical Attributes	Numerical Attributes	Missing Values
Bahrain Grand Prix	2024	13	413,014	5	8	0

[24] https://pypi.org/project/fastf1/

Table 2: Summary of Data Selected from FastF1[25]

Data Field	Description (units)	Typical Range	Role in Vehicle Behaviour	Relevance to IDS
Speed	Ground speed of the car, representing how fast the vehicle is travelling at each timestamp (km/h)	~0-350 km/h	Reflects the combined effects of throttle, braking, gear selection, aerodynamic drag, and DRS	Sudden or unrealistic changes may indicate tampering. Speed is strongly correlated with RPM and Gear, suitable for physics based consistency checks
RPM	Engine revolutions per minute, indicating crankshaft rotational speed (rev/min)	~1,500-14,000+ RPM (varies by team/era)[26]	Directly influenced by gear selection and throttle application. A higher RPM typically means more power, translating to faster lap times[27]	False data injection attacks on RPM can mislead power unit analysis; RPM to Speed correlation is a key anomaly detection feature
Throttle	Driver's throttle pedal position as a percentage of full acceleration (%)	0-100%	Controls power output from the engine and directly affects RPM and Speed	Inconsistent values, e.g., 100% throttle and brake reveal possible tampering or implausible states
Brake	Boolean value for the brake state (True/False)	0/1 (binary)	Decelerates the car and shifts the vehicle balance under braking	Helps detect unrealistic behaviour. e.g., simultaneous brake and throttle, useful context for anomaly scoring
Gear (nGear)	Current gear selected by the car's transmission control system	Integer 0–8	Gearbox state; relates RPM to Speed via ratios[28]	Tampered values break mechanical plausibility, making Gear useful for physics invariant detection
DRS	The state of the DRS, whether the Drag Reduction System is active, partially active, or closed (commonly represented as 0 or 14)	0 = Closed, 14 = Open	Reduces aerodynamic drag, allowing higher top speed on straights[29]	DRS activation outside permitted zones or at low speed indicates suspicious behaviour or faulty telemetry

3.2 Scenario and Case Study Setup

To support controlled experimentation, and to overcome the absence of labelled attack data, a synthetic tampering framework was built on the testing data frame by injecting attacks aligned with CPS literature (Longo, 2025), (Pervez, et al., 2020). Table 3 outlines the three attack categories that were utilised for the research. Table 4 outlines the types of motorsport attacks that were simulated within the three chosen attack categories, and an example of the observed data anomaly as a result of the manipulation.

Table 3: Summary of Chosen Attack Categories

Attack Category	Description	Purpose
False Data Injection (FDI)	Sudden spikes, drift, or bias in RPM or Speed	Compromise integrity[30]
Delay/Jitter Attack	Timestamp modification of a subset of samples	Desynchronise telemetry[31]
Packet Loss Attack	Random and burst dropping of samples	Degrade availability[32]

[25] https://docs.fastf1.dev/core.html#fastf1.core.Telemetry
[26] https://f1chronicle.com/what-rpm-do-f1-drivers-shift-at/
[27] https://driver61.com/motorsport-engineering/why-f1-engines-cant-rev-past-20000-rpm/
[28] https://f1chronicle.com/what-rpm-do-f1-drivers-shift-at/#Technical-Aspects-of-Gear-Changes-in-Formula-1
[29] https://www.espn.com/espn/feature/story/_/id/36180474/f1-terms-explained
[30] https://sudip-says-hi.medium.com/what-is-false-data-injection-ffccfca76719
[31] https://obkio.com/blog/latency-vs-jitter/
[32] https://www.checkpoint.com/cyber-hub/network-security/what-is-packet-loss-and-how-to-prevent-it/

Table 4: Summary of Simulated Motorsport Attack Mechanisms

Attack Category	Manipulation Method	Observed Data Anomaly	System Level Outcome
False Data Injection	***Spike injection***[33]	Sudden 80 km/h addition	Data Spoofing
	Speed zeroing[34]	Forced stall	Engine stall/ denial of control
	Throttle stuck[35]	Constant 100% throttle	Throttle actuator manipulation
	Brake loss[36]	Brake suppressed during deceleration	Brake actuator fault
	Gear flicker[37]	Incorrect gear jitter	Gear sensor corruption
Packet Loss/ Stale Data	***Data Freeze***[38]	Packet loss or stale value	Network packet loss
Delay/ Jitter	***Timing jitter***	Reordering jitter simulations	Control desynchronisation

Each attack was parameterised (timing, severity, duration) to produce multiple experimental runs and ensure robustness testing. These three realistic cyber physical attack categories, false data injection, delay/jitter, and packet loss, are introduced into the telemetry stream. Thus emulating adversarial conditions observed in CPS literature and providing an experimental basis for evaluating IDS behaviour under different threat scenarios.

3.3 Preprocessing and Feature Engineering Techniques

The raw telemetry collected from the FastF1 API consists of high frequency, multivariate time series measurements representing vehicle dynamics and control states. Therefore, before developing the intrusion detection pipeline, the data required preprocessing to ensure numerical suitability for unsupervised modelling. As telemetry streams are time dependent and multivariate, careful preparation is recommended to avoid misleading model behaviour or biased statistical results.

3.3.1 Data Cleaning and Structuring

Telemetry sessions were loaded lap by lap for each of the twenty drivers using the FastF1 API. The raw telemetry stream contained multiple features sampled at high frequency, and it was determined that minimal preprocessing was required to make it suitable for anomaly based detection, all while preserving the real time characteristics. The raw telemetry was structured into session aligned driver datasets in order to create driver specific baselines, allowing for individual driving styles and behaviour to be individualised.

Only completed telemetry rows were retained, excluding sessions or laps that contained incomplete data. This method was conducted to avoid introducing unnecessary noise into the model. Time series preprocessing typically involves utilising temporal smoothing. However, for this data it was decided that no smoothing, interpolation, or resampling were performed, (von Schleinitz, et al., 2021). This approach preserved the natural structure and timing of the telemetry data, keeping the feature space closely aligned with live race feeds and supporting the study's real time anomaly detection objective. To support downstream numerical processing, time data was converted from the FastF1 time delta format into floating point

[33] https://grassrootsmotorsports.com/articles/how-decipher-all-data-you-logged/

[34] https://www.motorsportweek.com/2025/10/28/how-start-contact-caused-carlos-sainz-to-incur-two-speeding-penalties-f1-mexico-gp/

[35] https://www.motorsport.com/f1/news/scared-leclercs-throttle-was-stuck-open-up-to-30-in-austrian-gp/10336286/

[36] https://www.mercedesamgf1.com/news/formula-one-brake-systems-explained

[37] https://www.quora.com/How-do-F1-drivers-know-what-gear-they-are-in-when-changing-gears-while-driving

[38] https://statusneo.com/f1-telemetry-the-hidden-battle-for-data-and-performance/

seconds, Figure 2 displays a selection of two drivers laps during the race, plotting their speed, throttle and brake.

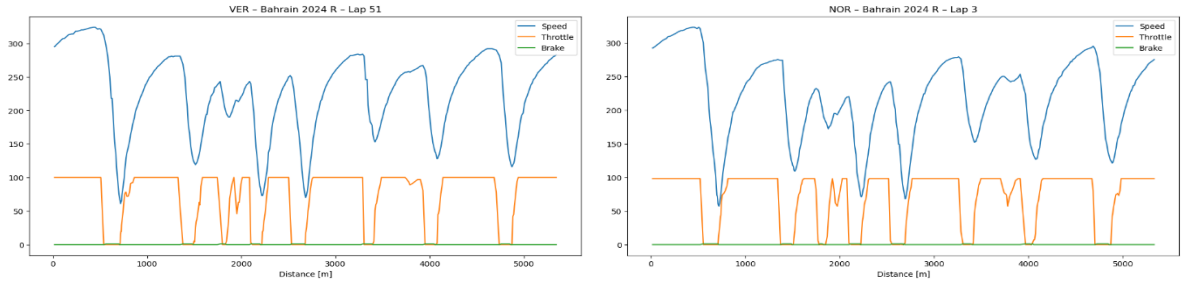

Figure 2: Fastest Lap for Drivers VER and NOR

3.3.2 Per Driver Train Test Partitioning

Formula 1 drivers are described as having distinct driving styles and approaches to racing[39]. As such it was expected that the driving styles and car setup variability would lead to naturally distinct telemetry distributions. To avoid cross driver data leakage, and ensure realistic deployment, all preprocessing and modelling was performed on a per driver basis. To do this, telemetry data was grouped by driver. Each drivers dataset was split into 80% training and 20% testing segments. The training set was assumed to represent clean driving behavior, and all transformations (standardising, PCA, IF) were fitted only on each driver's training data. This approach reflected a practical motorsport scenario in which each vehicle and driver pair generates an evolving statistical profile. The correlation heatmaps outlined in Figure 3 represent the relationships between six of the twenty drivers from the 2024 Bahrain race. General relationships are seen to be consistent across all drivers, with speed throttle, gear and RPM all being strongly positively correlation with each other. The correlation strength however varies by driver, confirming the differences in driving style or car setup being visible across the data, thus justifying the need for per driver partitioning.

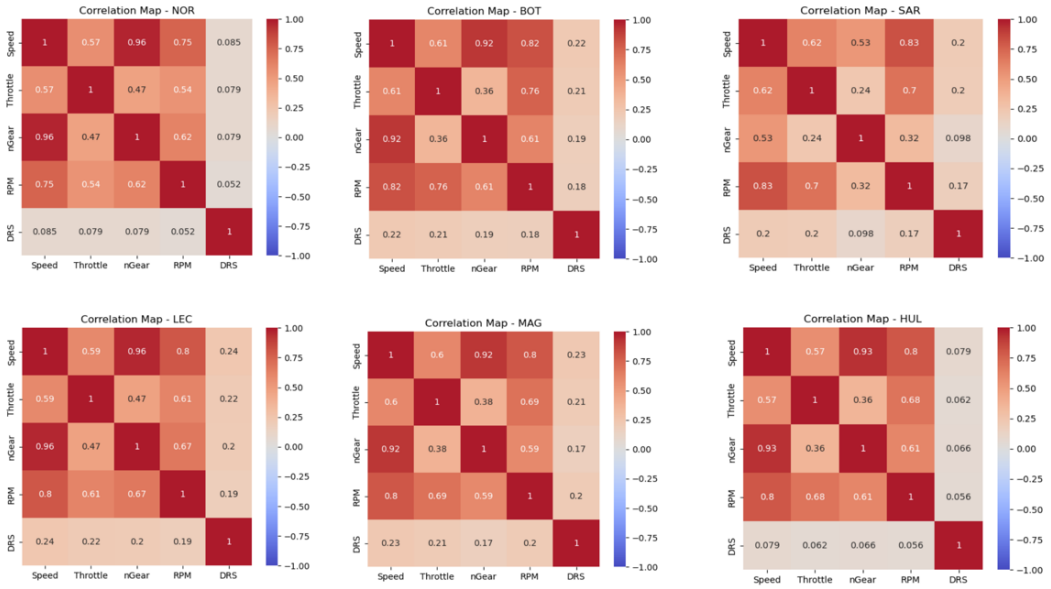

Figure 3: Correlation Heatmaps of 6 Drivers from Bahrain Race

[39] https://www.formula1.com/en/latest/article/watch-verstappens-distinct-style-and-alonsos-very-different-approach-to.2zKrFJDRsRjVXZZfgevdjp

3.3.3 Feature Standardisation

Unsupervised methods such as PCA and Isolation Forest, are sensitive to feature scales. To prevent bias within the data toward the high magnitude features such as RPM, and to ensure each input contributed proportionally, standardisation was applied to the data. As outlined in Table 2, all chosen features display differing scales, thus all continuous variables were standardised using z-score normalisation[40]:

$$x' = \frac{x - \mu_{train}}{\sigma_{train}}$$

The standardisation was fitted only on the training portion of the data to avoid information leakage, and transformed onto each drivers test set. This method ensures real time compatibility by learning statistics only from clean data. As a results the physical interpretability was maintained while preventing leakage from future observations.

3.3.4 Feature Engineering Strategy

The feature engineering approach prioritised real time robustness and interpretability. Raw telemetry was utilised where possible to support low computational overhead, thus reducing the delay and avoiding distortion of sharp transient events that may represent faults or attacks. This project intentionally preserved raw telemetry structure to maintain signal fidelity and ensure applicability in live streaming environments. Only minimal necessary preprocessing was introduced to avoid distorting physical behaviour, as mentioned above for effective model usage. This structured approach produced a robust input feature set suitable for unsupervised multivariate anomaly detection, while respecting real world telemetry constraints in Formula 1.

3.4 Machine Learning Models

The project applies an unsupervised machine learning strategy to detect abnormal telemetry behaviour in a Formula 1 race environment. Because labelled tampering or fault data is not available in real world motorsport systems, the model is trained exclusively on clean historical telemetry. This anomaly detection pipeline combines Principal Component Analysis (PCA) with Isolation Forest (IF). Hybrid methods have been identified as a well established approach for multivariate anomaly identification in high dimensionality sensor streams (von Schleinitz, et al., 2021), (Akin, 2024). As mentioned above, to reflect real racing conditions and avoid cross driver leakage, PCA and Isolation Forest models are trained and applied independently per driver.

3.4.1 Principal Component Analysis

PCA is used in the project as dimensionality reduction step to transform correlated telemetry inputs into a lower dimensional set of components. PCA serves two main purposes. Firstly, capturing dominant behavioural patterns across multiple vehicle signals. And secondly, reducing the noise and redundancy present within the data, to improve anomaly detection stability.

3.4.2 Isolation Forest

The reduced PCA features are passed into an Isolation Forest (IF) model, which isolates the anomalous points by recursively splitting the data space. Points that can be isolated with fewer

[40] https://www.geeksforgeeks.org/data-analysis/z-score-normalization-definition-and-examples/

splits are more likely to be anomalous[41]. IF is well suited for this domain due to its, unsupervised nature, meaning that no attack labels are required within the data. IF is efficient on streaming or real time data due to its robustness to high dimensional telemetry features.

3.5 Evaluation Methodology

The project adopts an experimental evaluation design to assess the ability of the intrusion detection system to recognise abnormal telemetry behaviour in a Formula 1 context[42]. Since labelled real world attack data is unavailable, controlled synthetic anomalies are injected into clean telemetry to create a test environment with known ground truth events. The evaluation methodology focuses on measuring how accurately the system identifies these synthetic deviations while maintaining low false positive rates on normal driving data. Table 5, outlines the evaluation methods that were used throughout the project.

Table 5: Summary of Selected Evaluation Methods

Evaluation Method	Description	Purpose / Insight	Output Examples
Baseline Statistical Evaluation	Trained PCA Isolation Forest on clean driver data and test on untouched telemetry	Established normal behaviour scoring distribution	Baseline anomaly scores, false alarm rates
Synthetic Anomaly Injection	Injected controlled cyber physical anomaly patterns into test telemetry	Created ground truth for benchmarking without real attack data	Fault tagged test sets for each driver
Performance Metrics	Computed key anomaly detection metrics (TPR, FPR)	Quantified detection effectiveness and reliability	Detection rate tables, ROC like comparisons
Per Driver Evaluation	Evaluated each driver's model independently	Prevented cross driver leakage and capture behavioural variability	Driver specific performance summaries
Physics Based Consistency Checks	Applied rule based physical plausibility checks (e.g., throttle brake conflict, frozen speed detection)	Assessed value of domain knowledge as model enhancement	Comparison of baseline vs physics augmented performance
Ablation Testing	Compared PCA IF (baseline) vs PCA IF + physics constraints	Measured contribution of each component	Improvement in TPR/FPR across anomaly types
Visualisation Analysis	Visually inspect anomaly scores, PCA space, and event overlays	Qualitatively validate detection behaviour and interpretability	Bar charts, line charts, and correlation matrices

4 Design Specification

For the project, a hybrid Intrusion Detection System (IDS) has been proposed which employs an unsupervised anomaly detection pipeline designed specifically for Formula 1 telemetry streams. This approach models each driver's normal driving behaviour independently, learning statistical baselines exclusively from clean historical data from the 2024 Bahrain Formula 1 Race. During operation, incoming telemetry is continuously scored against these per driver profiles to identify deviations indicative of faults or malicious manipulation. The IDS integrates two complementary detection layers via a hybrid approach. The first layer is the data driven anomaly detection which is carried out using Principal Component Analysis (PCA) followed by Isolation Forest (IF). This layer captures multivariate deviations from normal telemetry behaviour. The second layer is the physics informed rule checks, that enforce domain

[41] https://www.analyticsvidhya.com/blog/2021/07/anomaly-detection-using-isolation-forest-a-complete-guide/

[42] https://drogaieva.medium.com/designing-machine-learning-experiment-and-interpreting-experimental-results-cbf870caf68d

constraints which are based on known vehicle dynamics, for example, throttle brake conflicts. Unlike supervised intrusion detection pipelines, this approach does not require labelled attack data, making it suitable for zero day fault or tampering scenarios. The pipeline currently evaluates telemetry frame by frame, instead of sliding window embedding, supporting low latency inference without temporal buffering.

4.1 IDS Purpose and Requirements

The purpose of this design is to define a hybrid anomaly based IDS from Formula 1 telemetry that is capable of detecting cyber tampering and abnormal telemetry behaviour in real time. The design integrates physics informed rules with unsupervised machine learning (ML) techniques to detect deviations in car telemetry caused by false data injection, delay, or packet loss attacks. The IDS has three main system requirements categories that are needed to meet the project specifications. These requirements can be described as functional, non functional, and security requirements, and are outlined in Table 6.

Table 6: IDS System Requirements

Requirement	Description
Functional Requirements (Core Capabilities)	Detection of anomalous telemetry values Alert generation in real time Support multiple attack types Per driver adaptive modelling
Non Functional Requirements (Performance Expectations)	Fast response, low latency Easily adaptable, modular and scalable pipeline Robust to noise and telemetry variation Explainable decision making, interpretability
Security Requirements	Support integrity validation Detect previously unseen zero day tampering patterns Minimise false positives

4.2 IDS Architecture Description and Model Structure

As mentioned above the deployed IDS was a hybrid unsupervised anomaly detection system, as outline in Figure 4. The different stages of the IDS architecture are detailed in Table 7 below:

Table 7: Stages of the IDS Architecture

Stages	Method	Purpose
Baseline modelling	Per driver clean training split (80:20)	Avoids cross driver leakage, learns individual driving behaviour
Normalisation	StandardScaler (fitting on training data only)	Balances feature contributions and prevent magnitude bias
Dimensionality reduction	PCA (per driver)	Captures correlated telemetry behaviour and reducing noise
Anomaly scoring	Isolation Forest	Detect statistical outliers in PCA space
Physics constraints	Bounds, throttle brake conflict, frozen speed detection	Identifies physically implausible or mechanical inconsistent values
Decision logic	IF score and physics flags	Combines statistical and physical signals for robust alerts

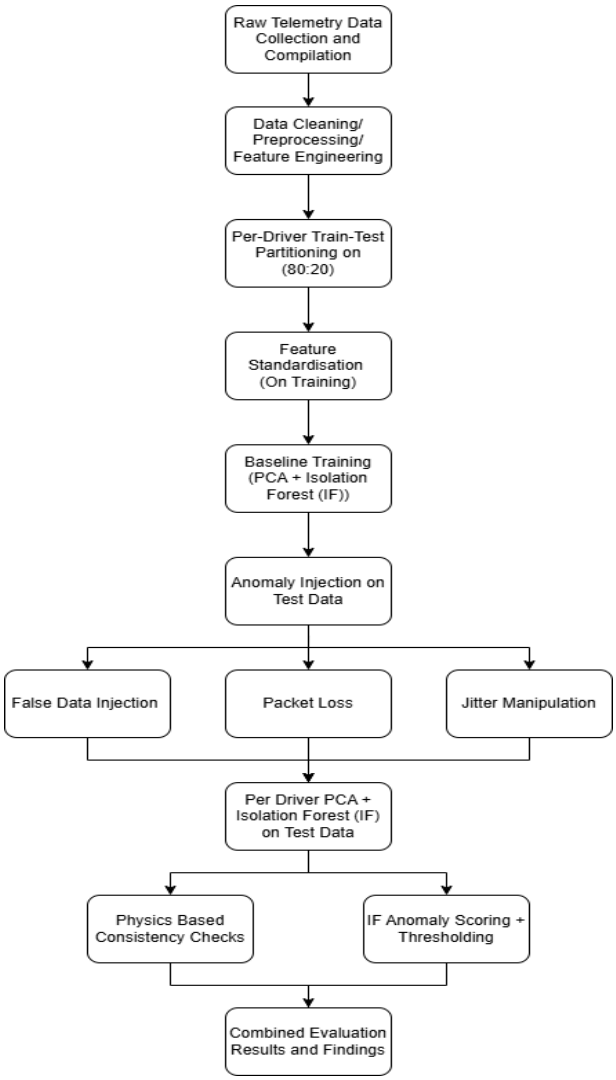

Figure 4: IDS Process Flow

4.3 IDS Detection Logic

The IDS uses a dual layer anomaly decision mechanism, from the PCA and Isolation Forest detector along with the physics based consistency checks, to increase robustness. Anomalies are detected if either the Isolation Forest scores exceeds the learned thresholds, or the relevant physics constraint are violated. There is a high confidence alert when both the Isolation Forest and the physics layers successfully detect anomalous behaviour. This layered design balances sensitivity of subtle anomalies, and precision while avoiding false positives from noise, which are both crucial considerations for real time F1 telemetry.

4.4 Design Rationale

A hybrid IDS is chosen because no single model reliably detects all CPS attack types. The system is designed using modular components to support reusability and future development. The architecture directly supports the threat model goals by protecting data integrity and availability. As outlined in the literature review, few studies evaluate real time IDS performance against synthetic attack scenarios involving data tampering, delay, or packet loss. This project addresses that gap by implementing and testing a hybrid PCA and IF IDS on Formula 1 telemetry data, using simulated attack conditions.

5 Implementation

The final implementation phase focused on developing and executing the hybrid intrusion detection pipeline for the Formula 1 telemetry data. This stage integrated the pre processing framework, anomaly injection mechanisms, the unsupervised machine learning models, and the physics based rule checks into a combined unified system which was capable of detecting abnormal telemetry behaviour in a near real time setting.

5.1 Tools and Technologies

All experiments were conducted on a workstation running Windows 11 (64-bit) with Python 3.9.13, using the following software ecosystem outlined in Table 8. The environment was held constant for all runs to avoid variation in results and ensure reproducibility.

Table 8: Summary of Tools and Technologies

Tool/ Library	Version	Purpose
Python	3.9.13	Programming language used to run data analysis, ML, and telemetry scripts
FastF1	3.6.1	Python Library for acquisition of real Formula 1 telemetry data
NumPy	2.0.2	Time series data handling and preprocessing
Pandas	2.3.3	Data handling and transformation[43]
Scikitlearn	1.6.1	ML library for modelling (PCA, IF) and data preprocessing (scaling)[44]
Matplotlib	3.9.2	Visualisation of telemetry and anomaly markers (line plots, etc.)
Seaborn	0.13.2	Built on Matplotlib, used for plotting (heatmaps, etc.)[45]
Jupyter Notebook	7.4.5	Interactive coding environment for Python used for experiment execution, reproducibility, and iterative development

5.2 Synthetic anomaly injection

A function was created to take the real time telemetry data, and add fake faults into it, as described in Table 4. These faults simulated real world sensor and vehicle problems so that anomaly detection can be tested in the models. Each of the anomalies were applied to a short instance of time within the data. Data was never deleted from the rows present within the data, in order to keep the timestamps aligned. Each injected segment is marked so that there is record of where they faults are.

5.2.1 Anomaly 1: Speed Spike

The first anomaly mimicked a speed sensor spike, where the measured vehicle speed momentarily jumped to an unrealistically high value. For each driver, a contiguous 15 sample window was selected and 40, 80, or 120 km/h was added to the recorded ‘Speed’ values. The modified values were then capped at 350 km/h to keep them within plausible operating range

[43] https://pandas.pydata.org/

[44] https://www.ibm.com/think/topics/scikit-learn

[45] https://seaborn.pydata.org/

of an F1 car. This anomaly created a short, sharp spike that mimicked transient sensor corruption or telemetry noise, allowing the detection pipeline to be tested on sudden non physical speed increases.

5.2.2 Anomaly 2: Speed Drop Out or Severe Sensor Failure

This anomaly reproduced a speed sensor failure where the speed suddenly reads near zero, despite the car continuing to move. To simulate this, contiguous windows of 10, 20, and 40 samples were selected, and the 'Speed' values within each window were replaced with uniformly sampled values between 0 and 5 km/h. This emulated a short term sensor failure or communication outage and allowed evaluation of the models ability to detect implausible low speed readings during active motion.

5.2.3 Anomaly 3: Throttle Stuck During Deceleration

The third anomaly simulated a throttle position sensor fault in which the throttle reading remained high despite vehicle deceleration. Candidate windows were selected during periods of negative speed gradient, and the 'Throttle' signal was forced to a high constant value for approximately 40 samples. This created a physically inconsistent input pattern characteristic of a stuck pedal or frozen throttle sensor.

5.2.4 Anomaly 4: Brake Loss During Deceleration

The brake loss anomaly represents a brake signal dropout, where braking effort is not recorded even though the vehicle is decelerating. Deceleration periods were identified via negative speed gradients, and within windows of 20, 30, or 40 samples, the 'Brake' signal is set to zero. This emulated brake pressure sensor failure or braking line signal loss, generating a physically contradictory state for anomaly detection validation.

5.2.5 Anomaly 5: Gear Flicker

The fifth anomaly captured gear selection sensor instability, producing rapid oscillations in reported gear. For each affected window of approximately 20 samples, integer jitter of plus or minus one to three units was added to the 'nGear' values and clipped to the valid range (1-8). This introduced high frequency gear noise typical of encoder bounce or corrupted transmission state signals.

5.2.6 Anomaly 6: Packet Loss/ Frozen Values

Packet loss anomaly injection, simulated signals that remained constant temporarily due to data not updating. For selected windows of 10 and 20 samples, each applicable sensor value (e.g., 'Speed', 'Throttle', 'Brake', 'nGear', 'RPM') was overwritten with its preceding value, producing a hold last sample effect. This mimicked communication latency or telemetry buffering faults.

5.2.7 Anomaly 7: Delayed Data/ Sensor Lag

Lastly the final injected anomaly emulated temporal misalignment in telemetry signals, where sensor updates arrived late. For each selected 30 sample window, the relevant channels were replaced with values shifted by plus two or five samples, which was effectively applying a

positive time delay. This recreated timestamp jitter or delayed CAN telemetry updates without altering index alignment.

5.3 Physics Based Consistency Checks

To ensure the telemetry remained physically plausible and internally consistent, a series of physics based validation checks were applied prior to anomaly analysis. These checks were implemented to detect violations of fundamental vehicle dynamics and sensor behaviour, including impossible speed values, excessively large acceleration or jerk magnitudes, contradictory throttle brake inputs, and prolonged periods of frozen speed readings indicative of telemetry packet loss. Acceleration and jerk thresholds were not hard coded; instead, robust, session adaptive limits were used by flagging values above the 99.5th percentile of observed magnitudes, ensuring sensitivity to abnormal behaviour without over reacting to legitimate extremes in race conditions. Together, these heuristics provided a lightweight but effective filter that highlights inconsistent or physically impossible data, reinforced confidence in downstream anomaly detection, and mirrored real world motorsport engineering quality assurance practices.

Table 9: Summary of Physics Based Consistency Checks

Flag Column	What it detects	How its Computed	Parameters/ Thresholds
phys_out_of_range	Speed that is outside the physically plausible bounds	Flagging if 'Speed' < 0 or 'Speed' > 'max_speed'	'max_speed' Default is set to 360 km/h
phys_high_accel	Implausibly high acceleration magnitude	Converting 'Speed' (km/h) to m/s, and checking if finite difference with the 'dt' column	Threshold is taken from values greater than the 99.5th percentile in the session
phys_high_jerk	Implausibly high jerk (rate of change of acceleration)	Calculating the finite difference of acceleration, using the 'dt', and flagging if true	Threshold is taken from values greater than the 99.5th percentile in the session
phys_tb_conflict	Throttle brake inconsistency (hard throttle while braking)	Flag when 'Throttle' > 'throttle_hi'* and Brake = 1	*'throttle_hi' = 80, which is a throttle threshold on a 0–100 scale
phys_speed_freeze	Frozen speed (no updates across many samples)	Groups consecutive True runs and counts their lengths, producing a running counter of how long the speed has remained constant	flags runs with length ≥ 'freeze_window'
phys_any	Row wise aggregate of all above	Logical 'OR' of the five flags	Any physics issue present convenience mask

5.3.1 Fusion of Physics based checks with IF

A fixed false positive rate is set for each driver, and the Isolation Forest is calibrated on clean data to meet it. The physics based consistency checks were then added to flag impossible telemetry behaviour. An anomaly is counted if either the model or physics triggers. Performance is measured at the event level, judging whether each injected anomaly window is caught. These results were visualized in the Evaluation section below.

6 Evaluation

The following section presents the results obtained by applying the IDS to controlled attack scenarios. The implemented PCA and Isolation Forest Intrusion Detection System (IDS) was

evaluated using the injected anomalies that were representative of critical cyber physical faults within Formula 1 telemetry data. The goal of the study is to analyse how unsupervised, anomaly based intrusion detection can be applied to Formula 1, and investigate to what extent can hybrid statistical and physics models detect synthetic tampering attacks in real time. The anomaly detection system is evaluated using a set of complementary methods, as summarised in Table 5. Baseline statistical measures were computed alongside key true positive rate (TPR) and false positive rate (FPR) performance metrics. These results were integrated with physics based consistency checks and visualisations.

In the first set of evaluations, each of the anomaly types from A1 to A7, were independently injected into the clean driving data. For each experiment, a series of score thresholds were computed based on the fixed score quantiles. For this, 10%, 5%, 2%, 0.2%, 0.1%, 0.05% were used, to represent varying levels of model strictness, as seen in Table 10. A detection is counted once a model score falls below the corresponding quartile listed above. .

The lower the quantile, the more strict the threshold is. These thresholds, e.g. 0.1% and 0.05%, yield lower TPR but corresponding near zero FPR. Because these thresholds were computed separately for each experiment on the full dataset (i.e., both clean and anomalous samples), the values reflect the score distribution in each of the specific experiments. Table 10, assists in highlighting the detectability of each of the anomaly types across a range of threshold aggressiveness. By assessing the values obtained under the 10% and 5% thresholds, it can be seen that anomaly 2, speed zeroing, was the anomaly that obtained the highest values. With A3, throttle stuck being close behind. A3 however does not perform well during the stricter thresholds. A6, the packet loss hold shows as the weakest overall from the below.

Table 10: Detection Sensitivity of Each Anomaly Type Under Increasingly Strict Thresholds

Anomaly Type	10%	5%	2%	0.2%	0.1%	0.05%
A1: Speed Spike	0.866667	0.733333	0.533333	0.1	0.016667	0
A2: Speed Zero	0.983333	0.916667	0.583333	0.033333	0.016667	0.016667
A3: Throttle Stuck	0.983333	0.833333	0.316667	0	0	0
A4: Brake Loss	0.85	0.466667	0.216667	0.033333	0	0
A5: Gear Flicker	0.816667	0.633333	0.3	0.016667	0	0
A6: Packet Loss Hold	0.625	0.35	0.15	0	0	0
A7: Delay Jitter	0.95	0.65	0.25	0.05	0.05	0.05

Similarly, Figure 6 and Figure 7, show that the anomaly detection model exhibits strong performance for large high signal anomalies, such as speed zero, throttle stuck and speed spike, but performance degrades sharply for subtle or intermittent faults, such as gear flicker, brake loss, and packet loss. A2 speed zero is the best detected anomaly overall, with A6, packet loss hold, remaining the most challenging to detect, with consistently lower detection across all operating points. Across all the anomaly types, detection performance continually decreases as the thresholds tighten, reflecting the expected relationship between TPR and model strictness. As mentioned, the model demonstrates strong performance on the high signal anomalies, while the more subtle and intermittent behaviours show reduced sensitivity. These results highlight that different anomaly types occupy different positions in the relationship between detection sensitivity and false positive tolerance. Thus emphasizing the value of evaluating across multiple operating points rather than relying on a single threshold.

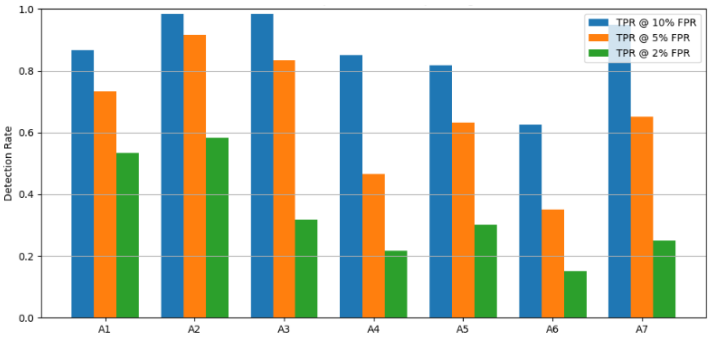

Figure 6: Detection Comparison Across Operating Points

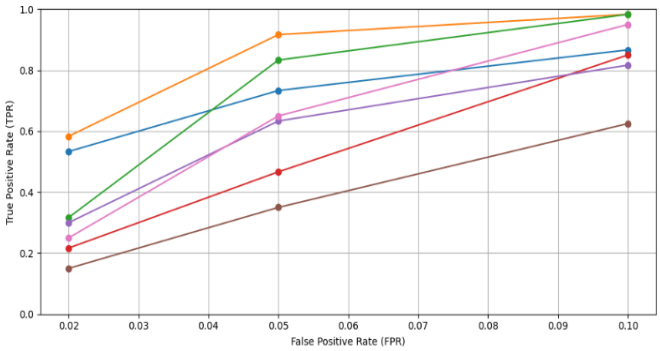

Figure 5: ROC Style Visualisation per Anomaly Type

The second analysis used a more realistic and operationally meaningful evaluation procedure. The thresholds are calibrated exclusively on clean driving data, separately for each driver. The threshold is chosen such that 5% of the drivers clean samples would be flagged. These calibrated thresholds are then applied to the full set of injected anomaly events. Detection in Table 11, is assessed at an event level, whereby the anomalies are considered detected once any sample within the event crosses the threshold. Thus creating an event level TPR, which is a more realistic measure of whether anomalies are successfully recognized during their occurrence. Table 11, evaluates both PCA and IF only detection at 5%, and the fused detection of the model flags with the physics based rules. Anomaly 3, the throttle stuck injection, showed the highest results, with 55% of events being detected when solely looking at PCA IF performance. Once the physics based consistency check were introduced, the fused system was able to detect 100% of those same anomaly events. A6, the packed loss hold, results in the lowest TPR of the injected anomalies, with the fusion of the physics based consistency checks, not greatly improving the TPR.

Overall, the visualisation in Figure 7 highlights which anomaly types are naturally well captured by the Isolation Forest and where physics checks provide meaningful gains. In most cases, the fused detector achieves equal or higher detection rates, demonstrating the complementary nature of domain driven physical constraints and the data driven model. Differences across anomaly types also help illustrate how certain faults, particularly those that violate physical constraints or produce frozen sensor behaviour, benefit significantly from the physics layer, while others are already well captured by the machine learning model alone. The physics layer that is introduced never harmed the detection cases, instead it improved the performance and detection rate, showing it is complementary to the models accuracy.

Table 11: Summary of IF and Physics Fused Results

Anomaly Type	TPR at 5% FPR	TPR Fused
A3: Throttle Stuck	0.550000	1.000000
A2: Speed Zero	0.650000	0.833333
A7: Delay Jitter	0.700000	0.800000
A5: Gear Flicker	0.633333	0.750000
A1: Speed Spike	0.542373	0.559322
A4: Brake Loss	0.433333	0.450000
A6: Packet Loss Hold	0.282051	0.384615

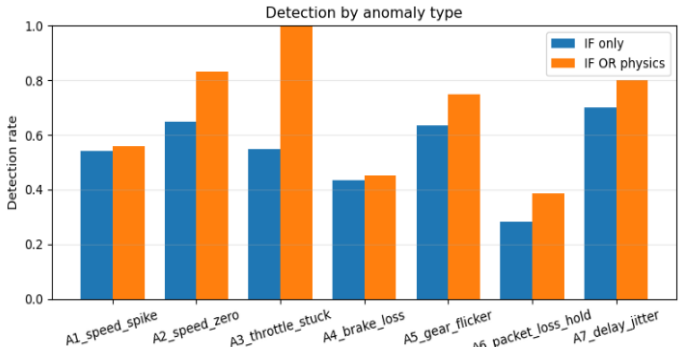

Figure 7: Detection by Anomaly Type

A 5% per driver FPR was chosen to account for model sensitivity, robustness, driver variability and comparability. 5% is a low enough FPR to avoid excessive noise and alerts within the data, but high enough that subtle anomalies can still be detected. The thresholds being calibrated per driver allowed for the natural behavioural difference between the drivers, that had the potential to inflate the false positives, if taken at an overall level. By evaluating this operating point the project is provided with a realistic assessment of system performance under conditions that mirror the expected operational constraints.

7 Conclusion and Future Work

Returning to the central research question: How can unsupervised, anomaly based intrusion detection be applied to Formula 1 telemetry streams, and to what extent can hybrid statistical and physics models detect synthetic tampering attacks in real time. The study set out to achieve several objectives. These included analysing the cyber threat landscape surrounding Formula 1 telemetry, designing a synthetic attack injection framework for controlled tampering experiments, developing an unsupervised anomaly based intrusion detection system, and evaluating its effectiveness in terms of true positive and false positive performance.

Building on these objectives the research demonstrated how telemetry specific statistical modelling, when fused with physics informed consistency checks, can characterise normal vehicle driver behaviour with sufficient accuracy to expose subtle deviations introduced via the synthetic tampering. The introduction of physics based consistency checks into the models, overall improved the anomaly detection performance. These physics based enhancements were achieved without needing to retrain or modify the underlying model, demonstrating that simple physics layer checks are a practical and effective extension to unsupervised PCA Isolation Forest systems in real time telemetry. The attack injection framework enabled systematic evaluation of a wide range of anomaly types, from the high signal failures observed in speed zeroing, and throttle sticking, to the more subtle low magnitude disruptions such as sensor delay, and packet loss.

While the PCA and Isolation Forest IDS achieved moderate detection rates, this behaviour is desirable in this research context as it demonstrates that the model generalises conservatively, avoiding overfitting to synthetic noise while still identifying a majority of true anomalies at a low false positive rate. The results highlight the IDSs suitability for unsupervised monitoring of CPS telemetry, with strong detection rates observed for abrupt tampering (e.g., speed zero, delay jitter), and improved performance for subtle control attacks (e.g., throttle stuck, packet loss) when combining the statistical and physics aware rules. These reduced TPR values observed from Table 10 to Table 11, after introducing scale consistent and context aware anomaly injections indicated a more realistic evaluation of how the model was able to effectively distinguish between physically plausible variations and genuine faults within the data, rather than trivially large signal perturbations.

The results obtained from the model validate both the methodological choices that were made for the project, such as the controlled anomaly injection framework along with the per driver training approach, and the theoretical assumptions identified within the literature review. These theoretical assumptions emphasises the importance of multivariate modelling for capturing the complexity of the correlations present within the telemetry data. The experiments also confirmed that the unsupervised hybrid models are well suited to unlabelled CPS telemetry. Thus offering a practical means of detecting the tampering within the data without prior attack signatures. The use of synthetic attacks enables a controlled yet realistic evaluation of IDS robustness, with the hybrid approach consistently outperforming the single method

alternative. The IDS was able to effectively meet the three core requirements that were set out for it in Table 6 above. Collectively the findings in this research demonstrates that a hybrid anomaly based intrusion detection approach, combining PCA driven feature compression, Isolation Forest anomaly scoring, and lightweight physics consistency checks, can detect a range of synthetic telemetry manipulation attacks in Formula 1 car data streams.

7.1.1 Limitations

The results of this study highlighted both the promise and the practical limitations of unsupervised intrusion detection in high velocity motorsport environments and offer a foundation for future work incorporating multivariate modelling, adaptive thresholds, and integration with team telemetry engineering workflows. The research makes several assumptions to scope and manage system complexity, it assumes that the attacker is limited to tampering with in transit telemetry data, rather than gaining access to onboard ECUs or vehicle actuators[46]. Although FastF1 provided convenient access to authentic Formula 1 telemetry, it has limitations that must be acknowledged. The data is not raw CAN bus or ECU level telemetry, but broadcast timing data intended for public viewing, therefore, it lacks many low level vehicle parameters that would be available to engineers in a real security environment. The data used is therefore assumed to be attack free, thus enabling unsupervised models to learn normal behaviour. These assumptions lead to certain limitations. The system does not attempt to prevent attacks, but only to detect and flag abnormal behaviour. As a result, while FastF1 is suitable for experimentation and proof of concept IDS development, it cannot perfectly represent all operational, security, or engineering aspects of a real Formula 1 telemetry environment.

7.1.2 Future Work

With additional time and resources, the research could focus on extending the IDS towards richer temporal modelling and improved sensitivity to subtle control plane anomalies. While PCA successfully captured static correlations across telemetry channels, it does not model temporal dynamics. Incorporating sequential methods such as LSTM Autoencoders or temporal convolution networks (TCNs) may enable better detection of gradually developing faults such as brake degradation or throttle drift, which exhibited lower recall in this study. Additionally, future evaluations could look at incorporating different injections of realistic cyber attacks. Testing against more advanced attack profiles would more rigorously evaluate the IDS's resilience and further bridge the gap between academic experimentation and operational deployment in safety critical environments.

References

Ajmera, V., 2025. *Racing Into Danger: Advanced Cyber Threats Targeting Formula 1 Fans and Teams Ahead of the Dutch Grand Prix.* [Online]
Available at: https://www.cloudsek.com/blog/racing-into-danger-advanced-cyber-threats-targeting-formula-1-fans-and-teams-ahead-of-the-dutch-grand-prix

Akin, L., 2024. Data-Driven Vehicle Performance Optimization for Formula Student Racing. October.
akshara_416, 2025. *Anomaly detection using Isolation Forest – A Complete Guide.* [Online]
Available at: https://www.analyticsvidhya.com/blog/2021/07/anomaly-detection-using-isolation-forest-a-complete-guide/

Anderson, G., 2023. *Gary Anderson explains how DRS actuators work on F1 cars.* [Online]
Available at: https://www.the-race.com/formula-1/gary-anderson-explains-how-drs-actuators-work-on-f1-cars/

[46] https://www.the-race.com/formula-1/gary-anderson-explains-how-drs-actuators-work-on-f1-cars/

Barber, B., 2025. *BHA believed to be latest organisation to be hit by a cyber attack.* [Online]
Available at: https://www.racingpost.com/news/britain/bha-believed-to-be-latest-organisation-to-be-hit-by-a-cyber-attack-arubK5n9AzDw/

Benson, A., 2013. *British GP: Why did six F1 tyres blow up at Silverstone?.* [Online]
Available at: https://www.bbc.com/sport/formula1/22294880

Benson, A., 2022. *F1 'spygate': Fifteen years on from the sporting scandal that had everything.* [Online]
Available at: https://www.bbc.com/sport/formula1/63575321

Bhambwani, R. N., 2024. *Data: The Unseen Driver in Formula 1 Cars.* [Online]
Available at: https://medium.com/formula-one-forever/data-the-unseen-driver-in-formula-1-cars-63f31c16f2fe

Blackstock, E., 2020. *Formula One Telemetry: Explained.* [Online]
Available at: https://www.jalopnik.com/formula-one-telemetry-explained-1845819678/

Bongiovanni, I., Herold, D. M. & Wilde, S. J., 2024. Protecting the Play: An Integrative Review of Cybersecurity in and for Sports Events. *Computers & Security,* August.Volume 146.

Brodowicz, M., 2024. *Taking It to the Limit: Physics in Formula One Car Design.* [Online]
Available at: https://aithor.com/essay-examples/taking-it-to-the-limit-physics-in-formula-one-car-design#2-the-role-of-physics-in-formula-one-car-design

Carratù, M., Gallo, V., Pietrosanto, A. & Sommella, P., 2025. Deployment of an Anomaly Detection Methodology Based on Generative Adversarial Network. *2025 IEEE International Workshop on Metrology for Automotive (MetroAutomotive),* June.

Check Point, n.d. *What Is Packet Loss and How To Prevent It?.* [Online]
Available at: https://www.checkpoint.com/cyber-hub/network-security/what-is-packet-loss-and-how-to-prevent-it/

Clark, B., n.d. *What is Scikit-Learn (Sklearn)?.* [Online]
Available at: https://www.ibm.com/think/topics/scikit-learn

Cleeren, F., 2022. *"Scared" Leclerc's throttle was stuck open up to 30% in Austrian GP.* [Online]
Available at: https://www.motorsport.com/f1/news/scared-leclercs-throttle-was-stuck-open-up-to-30-in-austrian-gp/10336286/

Collantine, K., 2019. *Leclerc keeps Abu Dhabi GP podium as Ferrari are fined for fuel infringement.* [Online]
Available at: https://www.racefans.net/2019/12/01/leclerc-keeps-abu-dhabi-gp-podium-after-fuel-penalty-investigation/

Deslandes, N., 2024. *Accelerating cybersecurity in motorsports.* [Online]
Available at: https://techinformed.com/accelerating-cybersecurity-in-the-world-of-motorsport-formula-one/

Dunn, J. E., 2014. *Marussia F1 testing disrupted by Trojan malware.* [Online]
Available at: https://www.techcentral.ie/marussia-f1-testing-disrupted-by-trojan-malware/

ESPN, n.d. *A BEGINNER'S GUIDE TO F1 TERMS AND TECH.* [Online]
Available at: https://www.espn.com/espn/feature/story/_/id/36180474/f1-terms-explained

Evans, C., 2013. The media representation of Formula One as 'spectacle' : constructing sport as a live mediatised event.

FastF1, n.d. *FastF1.* [Online]
Available at: https://docs.fastf1.dev/

FastF1, n.d. *Timing and Telemetry Data - fastf1.core.* [Online]
Available at: https://docs.fastf1.dev/core.html#fastf1.core.Telemetry

Formula1, 2018. *Brake disc failure to blame for Leclerc-Hartley collision.* [Online]
Available at: https://www.formula1.com/en/latest/article/brake-disc-failure-to-blame-for-leclerc-hartley-collision.3Kl1jE4UV2CMaiwA8yEw8i

Formula1, 2024. *WATCH: Verstappen's 'distinct' style and Alonso's 'very different' approach to cornering.* [Online]
Available at: https://www.formula1.com/en/latest/article/watch-verstappens-distinct-style-and-alonsos-very-different-approach-to.2zKrFJDRsRjVXZZfgevdjp

Fortinet, 2025. *Cyber-Physical Systems (CPS).* [Online]
Available at: https://www.fortinet.com/resources/cyberglossary/cyber-physical-systems

Geeksforgeeks, 2025. *Principal Component Analysis (PCA).* [Online]
Available at: https://www.geeksforgeeks.org/data-analysis/principal-component-analysis-pca/

Geeksforgeeks, 2025. *What is Isolation Forest.* [Online]
Available at: https://www.geeksforgeeks.org/machine-learning/what-is-isolation-forest/

grandprix.com, 2025. *FIA confirms F1 driver data hacking breach.* [Online]
Available at: https://www.grandprix.com/news/fia-confirms-f1-driver-data-hacking-breach.html
[Accessed 2025].

Harness Team, 2025. *What is False Positive Rate?.* [Online]
Available at: https://www.harness.io/harness-devops-academy/false-positive-rate

Hashemi-Pour, C., 2023. *What is the CIA triad (confidentiality, integrity and availability)?.* [Online]
Available at: https://www.techtarget.com/whatis/definition/Confidentiality-integrity-and-availability-CIA

Howson, G., 2025. *What RPM Do F1 Drivers Shift At?.* [Online]
Available at: https://f1chronicle.com/what-rpm-do-f1-drivers-shift-at/#Technical-Aspects-of-Gear-Changes-in-Formula-1

iguazio, n.d. *What is the True Positive Rate in Machine Learning?.* [Online]
Available at: https://www.iguazio.com/glossary/true-positive-rate/

Ishan, N., 2025. *F1 Telemetry: The Hidden Battle for Data and Performance.* [Online]
Available at: https://statusneo.com/f1-telemetry-the-hidden-battle-for-data-and-performance/

Kateryna, 2021. *Designing Machine Learning Experiment and Interpreting Experimental Results.* [Online]
Available at: https://drogaieva.medium.com/designing-machine-learning-experiment-and-interpreting-experimental-results-cbf870caf68d

Kernitskyi, A., 2024. *Latency vs. Jitter: Understanding Network Metrics.* [Online]
Available at: https://obkio.com/blog/latency-vs-jitter/

Komov, D., 2024. *The role of software in Formula 1 racing.* [Online]
Available at: https://lutpub.lut.fi/bitstream/handle/10024/167439/bachelorsthesis_komov_daniil.pdf?sequence=1

Kyösti, A., 2023. Win on Sunday, Sell on Monday - The Winning Formula?

Lambert, O., 2022. A retrospective of breaches in the cyber security of Formula 1, the impact these situations had on the sport, and can these situations be used as lessons to predict possible future breaches?. January.

Lazaar, Y., 2024. *Design Science Research Methodology.* [Online]
Available at: https://medium.com/@yassin.lazar/design-science-research-methodology-4577f732a1fa

Longo, G., 2025. *On the cyber security analysis of computerized transportation systems.* [Online]
Available at: https://iris.uniroma1.it/handle/11573/1734600

Mandal, A. A., 2025. *How start contact caused Carlos Sainz to incur two speeding penalties in F1 Mexico GP.* [Online]
Available at: https://www.motorsportweek.com/2025/10/28/how-start-contact-caused-carlos-sainz-to-incur-two-speeding-penalties-f1-mexico-gp/

Mansell, S., 2024. *Why Engines Can't Go Beyond 20,000 RPM in Formula 1.* [Online]
Available at: https://driver61.com/motorsport-engineering/why-f1-engines-cant-rev-past-20000-rpm/

Mercedes AMG F1, n.d. *Formula One Brake Systems, Explained!.* [Online]
Available at: https://www.mercedesamgf1.com/news/formula-one-brake-systems-explained

Mishra, D., 2018. Tacit Knowledge Transfer in Inter-Organisational Networks: A social network analysis of Formula 1. February.

MotorTrend Staff, 2012. *Hackers Protest Bahraini Government by Taking Down F1 Site.* [Online]
Available at: https://www.motortrend.com/news/hackers-take-down-f1-site-due-to-protests-in-bahrain-195785

Msakamali, B., 2024. F1 Data Analysis and Tactical Insights: Exploring Formula 1 Race Performance Strategies. May.

Pandas, n.d. *Pandas.* [Online]
Available at: https://pandas.pydata.org/

Pasterjak, J., 2025. *How to decipher on-track data for faster lap times.* [Online]
Available at: https://grassrootsmotorsports.com/articles/how-decipher-all-data-you-logged/

Pervez, L., Khan, A. & Ain, N. u., 2020. Cyber Security Challenge in an Automobile. *International Journal of Scientific and Research Publications,* December, 10(12), pp. 162-166.

Phys Org, 2012. *Hackers target F1 websites over Bahrain GP.* [Online]
Available at: https://phys.org/news/2012-04-hackers-f1-websites-bahrain-gp.html

pmishra01, 2025. *Z-Score Normalization: Definition and Examples.* [Online]
Available at: https://www.geeksforgeeks.org/data-analysis/z-score-normalization-definition-and-examples/

pypi, 2025. *fastf1 3.6.1.* [Online]
Available at: https://pypi.org/project/fastf1/

Quora, 2024. *How do F1 drivers know what gear they are in when changing gears while driving?.* [Online]
Available at: https://www.quora.com/How-do-F1-drivers-know-what-gear-they-are-in-when-changing-gears-while-driving

Seaborn, n.d. *seaborn: statistical data visualization.* [Online]
Available at: https://seaborn.pydata.org/

Sengupta, S., 2022. *What Is False Data Injection?.* [Online]
Available at: https://sudip-says-hi.medium.com/what-is-false-data-injection-ffccfca76719

von Schleinitz, J., Graf, M., Trutschnig, W. & Schröder, A., 2021. VASP: An autoencoder-based approach for multivariate anomaly detection and robust time series prediction with application in motorsport. *Engineering Applications of Artificial Intelligence,* September.Volume 104.

Watson, J., 2023. *Telemetry in F1: Unveiling Peak Performance Secrets.* [Online]
Available at: https://www.watsonpost.com/telemetry-in-formula-one/

Westbrook, T., 2025. Cyber Security Of Autonomous Vehicles: The Implications For City Planning. *Journal of Strategic Security,* Volume 18, pp. 77-88.
Widanage, C. et al., 2019. Anomaly Detection over Streaming Data: Indy500 Case Study. *2019 IEEE 12th International Conference on Cloud Computing (CLOUD),* July.