Predicting Cybersecurity Threats in Aviation through Passenger Profiling: A Machine Learning Analysis of Flight Data from South African Private Higher Education Students.

Research Project submitted in partial fulfillment of the requirements for the **Postgraduate Diploma in Data Analytics** at The Independent Institute of Education, Varsity College**.**

Tiara Bhairoparsad
ST10083181

Supervisor: Ms Z. Bulbulia

04 / 11 / 2024

RPDA8412 – Research Project

14177 words

Declaration

I hereby declare that the Research Report submitted for the **Postgraduate Diploma in Data Analytics** qualification to The Independent Institute of Education is my own work and has not previously been submitted to another university or Higher Education Institution for qualification purposes.

Signature Tiara Bhairoparsad 04 November 2024

Abstract

This study investigated the predictive capabilities of machine learning in identifying cybersecurity threats within the aviation industry, addressing the escalating complexity and frequency of cyber-attacks that necessitate enhanced security measures. The central premise was that analysing passenger behavior and device usage patterns could significantly improve the identification of potential cybersecurity risks. To explore this, primary data were collected via a Google Forms survey targeting South African private higher education students at The IIE's Varsity College, selected for their extensive access to flight-related information and active engagement with digital technology. The dataset comprised 352 observations, which underwent rigorous preprocessing to ensure data quality and consistency for analysis. A random forest classifier was employed to develop predictive models, achieving an impressive accuracy score of 100% in distinguishing between secure and potentially risky behaviors. Key findings revealed enhanced threat detection capabilities and refined passenger profiling techniques, providing actionable insights for integrating machine learning into aviation cybersecurity practices. By merging Routine Activity Theory with advanced data analytics, this study contributes valuable theoretical and practical perspectives, underscoring the importance of adaptive security measures tailored to passenger behaviors and technological contexts. Future research should aim to expand datasets and explore diverse demographics to further validate and enhance the robustness of predictive models in aviation security, ultimately fostering a more secure aviation environment.

Table of Contents

Chapter 1: Introduction

Title

Predicting Cybersecurity Threats in Aviation through Passenger Profiling: A Machine Learning Analysis of Flight Data from South African Private Higher Education Students.

Contextualisation of the study

The aviation industry represented a significant target for cyber threats due to its extensive digital infrastructure and the vast amount of sensitive data it managed, including passenger information such as passport details and credit card information. With increased digitization across aircraft and aviation systems, vulnerabilities escalated, providing cybercriminals with opportunities to exploit weaknesses in security protocols. As Ukwandu et al. (2022) highlighted, the frequency and sophistication of cyberattacks in the aviation sector surged, posing risks to passenger safety, operational continuity, and financial stability. Modernizing aviation systems with emerging technologies like augmented reality, machine learning, and the Internet of Things (IoT) further expanded the attack surface, making robust cybersecurity measures crucial. Reed (2023) reported that a survey conducted by Bridewell indicated that the aviation sector experienced an average of 24 ransomware incidents within the preceding year; however, this figure was likely an underestimation due to organizations’ reluctance to disclose such attacks. Shah (2022) emphasized that ransomware attacks had the potential to significantly disrupt an organization's daily operations by obstructing access to critical systems and applications, which could lead to severe safety implications for passengers and crew alike. Notably, in the first half of 2023, Anon (2023) reported that cyberattacks in the aviation industry rose by 24% globally on an annual basis. The cumulative effect of these incidents highlighted the pressing need for the industry to identify and mitigate emerging cybersecurity threats effectively. Recognizing this urgency, the International Civil Aviation Organization (ICAO) developed an Aviation Cybersecurity Strategy aimed at guiding industry efforts in preventing cyberattacks and ensuring compliance with established security standards. This study aimed to address the critical need for predictive models to identify cybersecurity risks by analysing passenger data. Over a four-month period, this research explored how patterns in passenger behavior and device usage could reveal indications of suspicious activity and potential threat actors, contributing to improved aviation security practices.

Problem statement

According to Alqushayri (2020), as aviation operations became increasingly complex and technology-dependent, the industry was plagued with growing cybersecurity risks. The pressing challenge lay in proactively identifying potential threats posed by passengers with malicious intent. Consequently, there arose an urgent need for inventive methodologies, such as machine learning-based passenger profiling, to foresee and counter cybersecurity risks in aviation.

Purpose statement

This study aimed to employ machine learning to analyse flight data to predict cybersecurity risks posed by passengers in the aviation industry.

Research Questions

- "How can machine learning analysis of flight data be utilized to predict cybersecurity threats in aviation through passenger profiling?"

- "What attributes of flight data can be effectively analysed through machine learning to identify potential cybersecurity threats given the absence of capable guardianship?"

Hypotheses

- H1: Machine learning analysis of flight data cannot predict cybersecurity threats in aviation through passenger profiling.

  H1a: Machine learning analysis of flight data can predict cybersecurity threats in aviation through passenger profiling.

- H2: Gender cannot be effectively analysed through machine learning to identify potential cybersecurity threats given the absence of capable guardianship.

  H2a: Gender can be effectively analysed through machine learning to identify potential cybersecurity threats given the absence of capable guardianship.
- H3: Age_Group cannot be effectively analysed through machine learning to identify potential cybersecurity threats given the absence of capable guardianship.

  H3a: Age_Group can be effectively analysed through machine learning to identify potential cybersecurity threats given the absence of capable guardianship.
- H4: Seat class cannot be effectively analysed through machine learning to identify potential cybersecurity threats given the absence of capable guardianship.

  H4a: Seat class can be effectively analysed through machine learning to identify potential cybersecurity threats given the absence of capable guardianship.
- H5: Devices_Used cannot be effectively analysed through machine learning to identify potential cybersecurity threats given the absence of capable guardianship.
- H5a: Devices_Used can be effectively analysed through machine learning to identify potential cybersecurity threats given the absence of capable guardianship.
- H6: Device_Usage_Frequency cannot be effectively analysed through machine learning to identify potential cybersecurity threats given the absence of capable guardianship.
- H6a: Device_Usage_Frequency can be effectively analysed through machine learning to identify potential cybersecurity threats given the absence of capable guardianship.
- H7: Data_Usage_MB cannot be effectively analysed through machine learning to identify potential cybersecurity threats given the absence of capable guardianship.
- H7a: Data_Usage_MB can be effectively analysed through machine learning to identify potential cybersecurity threats given the absence of capable guardianship.
- H8: USB_Devices_Connected cannot be effectively analysed through machine learning to identify potential cybersecurity threats given the absence of capable guardianship.
- H8a: USB_Devices_Connected can be effectively analysed through machine learning to identify potential cybersecurity threats given the absence of capable guardianship.

- H9: USB_Device_Type cannot be effectively analysed through machine learning to identify potential cybersecurity threats given the absence of capable guardianship.
- H9a: USB_Device_Type can be effectively analysed through machine learning to identify potential cybersecurity threats given the absence of capable guardianship.
- H10: Flight_Type cannot be effectively analysed through machine learning to identify potential cybersecurity threats given the absence of capable guardianship.
- H10a: Flight_Type can be effectively analysed through machine learning to identify potential cybersecurity threats given the absence of capable guardianship.
- H11: Flight_Duration cannot be effectively analysed through machine learning to identify potential cybersecurity threats given the absence of capable guardianship.
- H11a: Flight_Duration can be effectively analysed through machine learning to identify potential cybersecurity threats given the absence of capable guardianship.
- H12: Air_Travel_Frequency cannot be effectively analysed through machine learning to identify potential cybersecurity threats given the absence of capable guardianship.
- H12a: Air_Travel_Frequency can be effectively analysed through machine learning to identify potential cybersecurity threats given the absence of capable guardianship.
- H13: Frequent_Flyer_Status cannot be effectively analysed through machine learning to identify potential cybersecurity threats given the absence of capable guardianship.
- H13a: Frequent_Flyer_Status can be effectively analysed through machine learning to identify potential cybersecurity threats given the absence of capable guardianship.
- H14: Travel_Purpose cannot be effectively analysed through machine learning to identify potential cybersecurity threats given the absence of capable guardianship.
- H14a: Travel_Purpose can be effectively analysed through machine learning to identify potential cybersecurity threats given the absence of capable guardianship.

Scope of the study

This study focused on addressing cybersecurity threats in aviation by analysing flight data to identify risks associated with passenger profiles. Data were collected from South African private higher education students, a demographic selected for their digital engagement and familiarity with aviation services. The study aimed to demonstrate how machine learning could analyse passenger demographics, travel patterns, and digital behavior to detect potential cybersecurity threats.

The study began with a comprehensive review of existing literature to establish the critical background and contextualize the research problem, highlighting the aviation industry's susceptibility to cyber threats due to its reliance on digital infrastructure and large volumes of sensitive data. Guided by the research problem, which underscored the need for advanced predictive methods in cybersecurity, the study formulated hypotheses to test the predictive capability of various flight data attributes in identifying potential threats.

In the analytical phase, exploratory data analysis (EDA) was conducted on a dataset of 352 observations, including visualizations such as bar charts, summary statistics, and correlation matrices, to identify key relationships between features and the target variable. This initial analysis revealed patterns that informed feature selection and model development.

Subsequently, advanced machine learning techniques were employed, focusing on Random Forest classifiers, which provided insights into feature importance and significantly improved prediction accuracy. To refine model performance, hyperparameter tuning was conducted using grid search alongside cross-validation, ensuring robust evaluation and minimizing overfitting. The study also employed performance metrics such as accuracy, precision, recall, and F1-score to assess the model's effectiveness comprehensively.

Through this approach, the study's findings offered implications for both the student demographic and the broader aviation sector, suggesting practical enhancements to cybersecurity practices. By employing predictive analytics, this research contributed actionable insights for strengthening cybersecurity in aviation.

Chapter 2: Literature Review

Introduction

The increasing digitalization of systems and the exponential growth of sensitive data posed significant challenges that necessitated proactive measures to identify and mitigate potential threats. Shah (2022) indicated that cyber-attacks presented risks of operational disruption, safety compromises, and financial losses. Leveraging machine learning to analyse flight data and passenger behavior emerged as a promising solution. However, realizing the full potential of machine learning in this context required a comprehensive understanding of the underlying behavioural patterns and contextual factors that shaped cyber risks. The core problem revolved around the urgent need to pre-emptively address cybersecurity risks in aviation through the application of machine learning-based analyses of flight data, prompting an in-depth exploration of relevant theoretical frameworks and empirical studies. Despite the aviation sector's critical exposure to cyber threats, limited research existed on machine learning's capacity to proactively enhance cybersecurity measures within this domain, thereby creating a notable gap in the literature. This literature review aimed to bridge this gap by examining the intersection of aviation cybersecurity and machine learning, with a specific focus on Routine Activity Theory (RAT). Andresen and Farrell (2015) state that RAT has been extensively applied in criminology and cybersecurity to understand patterns of routine behavior that may heighten vulnerability to cyber threats. The theory's emphasis on identifying "likely offenders," "suitable targets," and "lack of capable guardianship" rendered it particularly relevant for the aviation industry, characterized by predictable passenger routines and high-value targets. By integrating RAT-informed variables into machine learning models, this approach facilitated the detection of subtle anomalies in flight and passenger data that might signal potential cyber threats. The review sought to explore how routine activities and behavioural patterns informed predictive models for threat detection, providing insights into the specific attributes of flight data conducive to machine learning analysis and the effective use of machine learning algorithms in identifying anomalous behavior indicative of security risks. Through this review, existing knowledge was consolidated to enhance the understanding of cybersecurity risks through machine learning based on passenger profiling, thereby guiding future research in this critical field.

Theoretical foundation

The integration of machine learning techniques with comprehensive datasets emerged as a pivotal approach in detecting potential cybersecurity threats within the aviation sector. Theoretical frameworks supporting this methodology rested on the premise that diverse variables embedded within flight data could serve as indicators of cybersecurity risks (Figure 1). By amalgamating these variables and applying sophisticated algorithms, patterns indicative of threats were discerned, enhancing aviation security protocols and aiding in proactive threat mitigation.

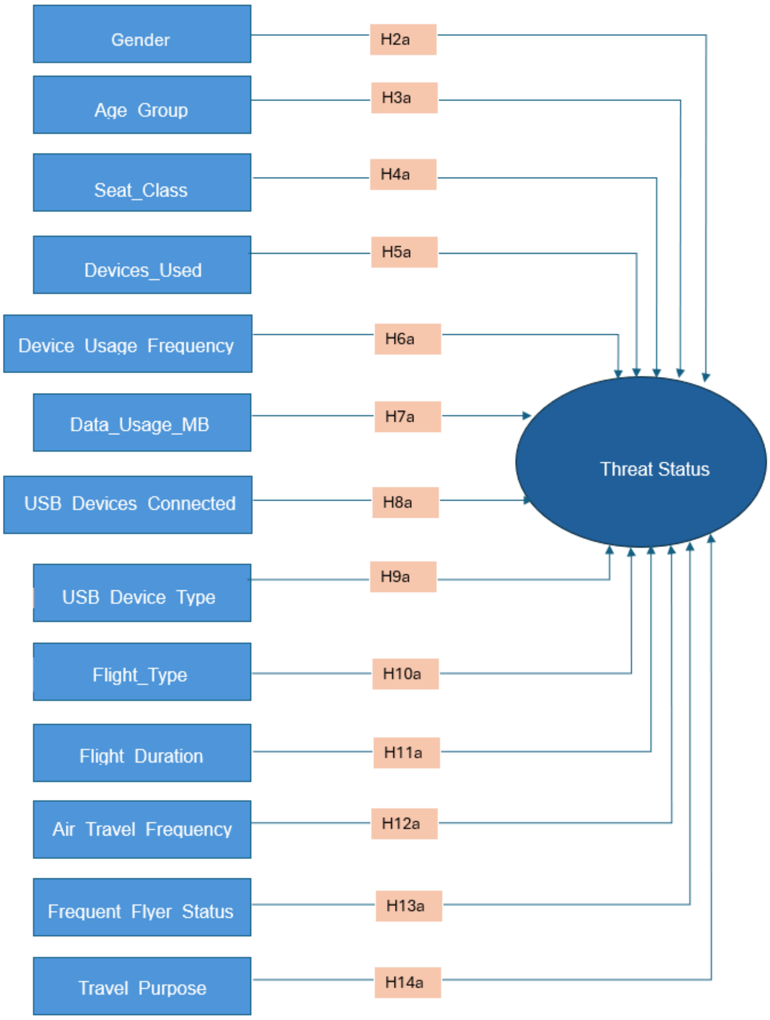


Figure 1: Diagram showing how variables can serve as an indicator of a cybersecurity threat according to hypotheses.

According to Anon (2019), In 2019, it was noted that airlines increasingly employed advanced sensors for real-time monitoring of aircraft performance, including engine functionality, air conditioning systems, and onboard entertainment systems. Furthermore, pilots have shifted to digital solutions, frequently using tablet devices to access extensive flight manuals, while airports and air traffic control systems relied on an array of digital devices for operational efficiency. The expanding digital landscape in aviation introduced vulnerabilities that threat actors could exploit, compromising critical aircraft systems or gaining access to sensitive information.

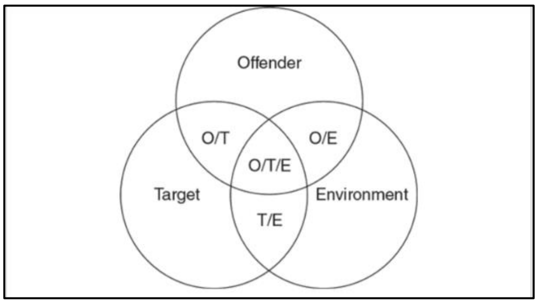
Passenger behavior and device usage patterns served as key variables in detecting anomalies associated with cybersecurity risks. For instance, seat class was observed to correlate with passenger behavior and access privileges; higher-class passengers often displayed different behavioural patterns from lower-class passengers, and deviations from these patterns could indicate security risks. Additionally, electronic devices carried by passengers, such as laptops and smartphones, presented a potential vector for unauthorized access or malicious activity. Suspicious device interactions, particularly when coupled with abnormal usage patterns, could signal cybersecurity threats. Internet usage onboard also provided insight into connectivity and system interactions. Unusual surges in usage, especially when combined with other suspicious behaviors, could indicate attempts to exploit system vulnerabilities. Similarly, system interactions, such as USB access, might reveal malicious intent. Analysis of these interactions enabled the identification of security threats by focusing on anomalous behaviors (Figure 1). Flight duration also informed threat detection, as longer flights might expose passengers to more system interactions, and deviations in behavior on extended flights could signify potential security risks. The application of machine learning algorithms facilitated the detection of these nuanced behaviors by discerning deviations from normal patterns, thereby providing a proactive stance against cybersecurity risks in aviation.

Routine Activity Theory (RAT), initially formulated by Lawrence E. Cohen and Marcus Felson in 1979 and further developed by Marcus Felson, offered a foundational perspective within criminology that emphasized situational factors facilitating criminal behavior rather than focusing solely on criminal motivation. Cohen and Felson (1979) stated that RAT posited that crime resulted from the convergence of three essential elements in time and space: a motivated offender, a suitable target, and the absence of capable guardianship (Figure 2).

By analysing the interplay of these elements, RAT provided a structured approach to understanding crime patterns, irrespective of the underlying motivations. This framework supported the notion that variables inherent in flight data, when evaluated through RAT, could be transformed into effective indicators for identifying cybersecurity risks.

RAT's emphasis on routine behaviors aligned well with machine learning's capabilities in analysing large datasets for pattern recognition. Cohen and Felson (1979) argued that by incorporating RAT-informed variables into machine learning models, aviation security analysts could detect anomalies that might indicate a convergence of RAT’s three elements, enhancing the industry’s ability to anticipate and respond to cybersecurity threats. This theoretical foundation underscored the importance of regular behavioural patterns and situational variables, lending itself to an innovative application of criminology principles within the aviation cybersecurity landscape. Thus, the convergence of machine learning techniques with RAT provided a robust framework for advancing proactive threat detection strategies tailored to aviation's unique security challenges.

Andresen and Farrell (2015) stated that the opportunity structure for crime, as represented in the crime equation—crime = (offender + target − guardian) (place + time)—explored the dynamic interaction between offenders, targets, and guardians, thereby outlining potential opportunities for criminal behavior. Although Routine Activity Theory (RAT) faced criticism for its perceived simplicity, its practical utility lay in its capacity to offer actionable insights for crime prevention. By examining routine activities and spatial-temporal patterns of crime, RAT provided invaluable insights for understanding, predicting, and ultimately preventing criminal behavior across various contexts, establishing itself as a foundational theory in contemporary criminology.



(Andresen and Farrell, 2015)

Figure 2: Offender, Victim, and environment analytic framework

- Offender: type of criminal
- Target: victimology
- Environment: place, time, date
- Offender/Target(O/T): Offender hunting style, victim preference, risk level
- Offender/Environment(O/E): Offenders mental map and activity space, hunting ground
- Target/Environment(T/E): Target backcloth, encounter site
- Offender/Target/Environment(O/T/E): Situation, crime, crime scene.

Rossmo and Summers (2015) emphasized the importance of regular and routine behaviors in understanding crime patterns. Although commonly applied to explain aggregate trends in society, Routine Activity Theory (RAT) also proved valuable at the individual level in criminal investigations. Rossmo and Summers (2015) states that by analysing the time and location of a crime alongside known information about the offense and victim, RAT enabled investigators to infer details about the offender. RAT offered a valuable framework for understanding and addressing cybersecurity threats, particularly in identifying potential hackers or cyberterrorists within the aviation industry. Cyberterrorism, within this framework, could be conceptualized as a form of crime occurring when motivated offenders—individuals or groups with malicious intent—exploited vulnerabilities in airline systems or infrastructure. In the aviation sector, motivated offenders, such as hackers, targeted aircraft systems and passenger data due to the industry's increasing complexity and reliance on technology. The evident lack of capable guardianship in aviation cybersecurity, often marked by insufficient security measures, limited resources for threat monitoring, and deficiencies in passenger screening, underscored the pressing need for robust protective measures.

RAT emerged as a crucial analytical tool for identifying potential offenders among airline passengers or external actors by examining discernible behavioural patterns, such as suspicious internet activity. This analytical approach facilitated passenger profiling, enabling predictions of whether a passenger might pose a hacking threat. RAT further highlighted critical targets within the aviation sector, including flight controls, communication networks, and passenger databases, that hackers might aim to compromise. Consequently, the need for rigorous cybersecurity protocols, effective threat-monitoring systems, and stringent passenger screening procedures became evident to counter cyberterrorist activities.

Recognizing that vulnerabilities in these areas could inadvertently create opportunities for exploitation underscored the necessity for continuous vigilance and enhancement of security protocols to safeguard aviation cybersecurity effectively.

Bock et al. (2017) noted that by recognizing parallels between traditional crime and cyber threats, cybersecurity professionals gained valuable insights into the dynamics of cyber risks, enabling the development of proactive defence strategies. Addressing these threats required innovative approaches, such as machine learning-based passenger profiling, to identify and mitigate risks pre-emptively. Enhancing the capacity for guardianship in aviation demanded the implementation of robust cybersecurity protocols, improved passenger screening procedures, and the adoption of advanced technologies for threat detection and response. Kigerl (2012) argued that by applying Routine Activity Theory (RAT) principles to understand offender behaviors and threat patterns, stakeholders were able to devise effective preventive measures and allocate resources strategically. This comprehensive strategy aimed to pre-emptively identify, mitigate, and counter potential threats, ultimately enhancing aviation security and safety in a highly interconnected digital landscape.

According to Miro (2014) Routine Activity Theory held significant implications for understanding and preventing crime, especially in urban contexts where changes in routine activities could influence crime rates. RAT was applied in various domains, including cybercrime, where understanding the routine activities of offenders and potential targets proved essential for designing effective defence strategies. Miro (2014) asserted that RAT remained a powerful tool for crime analysis and prevention, offering insights into the relationship between human behaviors, environmental factors, and criminal activities. This comprehensive approach aimed not only to identify potential cyberterrorism threats but also to mitigate them effectively. Future research could benefit from interdisciplinary approaches, integrating RAT with other theories, such as crime pattern theory, to deepen the understanding of crime dynamics and inspire innovative strategies for crime prevention.

Problem being investigated

Understanding Cybersecurity Risks in Aviation

The International Civil Aviation Organization (2020) stated that the aviation sector faced growing cybersecurity risks due to its reliance on interconnected digital systems for key functions such as flight operations, air traffic management, and passenger services. This interconnected infrastructure created an expanded attack surface, making it vulnerable to cyber threats. According to the U.S. Department of Homeland Security (2021), the aviation industry was recognized as one of the critical infrastructure sectors at high risk for cyber-attacks. The widespread adoption of automation and digital technologies across aircraft systems, air traffic control, and airport operations increased susceptibility to cyber-attacks aimed at disrupting operations or accessing sensitive information. For instance, a 2015 cyber-attack on the Polish airline LOT caused significant disruptions to its ground operations, resulting in numerous flight cancellations and delays (Zetter, 2015). According to Anon (1998) attacks targeting critical systems such as air traffic control or aircraft communication networks posed direct safety risks to passengers and crew; unauthorized access to flight control systems could potentially jeopardize flight safety.

The financial implications of cyber-attacks were substantial, affecting all stakeholders in the aviation sector. Costs incurred from mitigating incidents, restoring systems, and compensating affected individuals added up to considerable financial burdens. A 2023 report by Statista estimated that the average cost of a data breach in the transportation sector was $4.18 million (Statista, 2024). To address these escalating risks, proactive cybersecurity measures were deemed essential. Anon (2023) argued that a layered approach to cybersecurity was recommended, incorporating network segmentation, access controls, encryption, and employee training. Regular risk assessments and investment in advanced security technologies further strengthened aviation infrastructure against cyber threats.

Governments and regulatory bodies played a crucial role in promoting cybersecurity in aviation by establishing standards, regulations, and best practices. According to Grossman (2021) The United States, for instance, developed cybersecurity guidelines aimed at aircraft manufacturers, airlines, and other industry stakeholders to reinforce cyber resilience. This alignment between industry needs and regulatory support underscored the importance of a comprehensive, coordinated approach to cybersecurity within aviation, bridging the gap between the problem and the current literature’s emphasis on proactive, multi-layered defence strategies. This literature illuminated both the vulnerabilities in aviation cybersecurity and the practical measures necessary to address them, providing a foundational basis for the current investigation into cyber risk mitigation in the aviation sector.

Role of Machine Learning in Aviation Cybersecurity

According to a report by the European Aviation Safety Agency (2019) machine learning techniques such as supervised learning, unsupervised learning, and reinforcement learning were increasingly explored for their potential applications in aviation safety and security. These techniques offered the capability to analyse vast amounts of data generated by aircraft systems, air traffic management, and passenger behaviour to detect anomalies and identify potential threats. Paper (2018) highlighted the ability of machine learning algorithms to process large volumes of data from various sources, including flight data recorders, cockpit voice recorders, and airline reservation systems. This capability enabled airlines and aviation authorities to analyse complex datasets and detect security threats more effectively.

A study published in the Journal of Aviation Technology and Engineering demonstrated the effectiveness of machine learning algorithms in detecting anomalies in air traffic communication data. Hudec and Benova (2024) stated that the study found that machine learning-based anomaly detection techniques outperformed traditional rule-based methods, enabling more accurate and timely identification of potential security threats. Trottenberg (2020) stated that the Federal Aviation Administration recognized the potential of machine learning to automate cybersecurity tasks and scale security operations to manage the growing complexity of aviation systems. Machine learning-based security solutions continuously monitored network traffic, identified suspicious activities, and responded to security incidents in real-time, thereby enhancing the resilience of aviation infrastructure against cyber threats.

Anon (2021) stated that the Jamaican Civil Aviation Authority acknowledged the importance of safeguarding passenger data and flight operations data in aviation cybersecurity. However, the organization also recognized the challenges of applying machine learning techniques to analyse sensitive data while ensuring compliance with data privacy regulations such as GDPR and HIPAA. According to Gabrijelčič, et al. (2023) addressing these concerns required robust data protection measures and transparent data governance frameworks. The study emphasized the importance of developing interpretable machine learning techniques that enable cybersecurity analysts to understand the underlying rationale behind model predictions and decisions, thereby enhancing trust and confidence in the security measures deployed.

Integration of Routine Activity Theory and Machine Learning

The integration of Routine Activity Theory (RAT) with machine learning techniques in aviation cybersecurity involved leveraging RAT's principles to inform the development of machine learning models for threat detection and mitigation. Andresen and Farrell (2015) highlights that RAT emphasizes the role of routine activities, motivated offenders, suitable targets, and absence of capable guardianship in shaping crime patterns. By incorporating these principles into machine learning algorithms, cybersecurity professionals identified behavioural patterns indicative of potential security threats within aviation systems and infrastructure. For example, machine learning models were trained on historical flight data to recognize patterns associated with suspicious activities, such as unauthorized access attempts to aircraft systems or abnormal communication patterns between aircraft and ground control. By aligning machine learning analysis with RAT principles, cybersecurity experts can effectively identify potential threats and allocate resources to mitigate risks proactively. Machine learning analysis of flight data enables the profiling of passenger behaviour based on RAT principles, thereby enhancing aviation cybersecurity. By analysing factors such as online activities, machine learning models can identify deviations from normal behavioural patterns that may indicate potential security threats. By flagging such anomalies for further investigation, airlines and security agencies can intervene pre-emptively to mitigate potential security risks. A study published by Kabashkin et al. (2023) evaluated the effectiveness of integrating Routine Activity Theory with machine learning techniques for aviation cybersecurity. The study analysed historical flight data and passenger profiles to identify patterns associated with security threats, such as ticket fraud, suspicious baggage handling, and anomalous communication patterns. The results demonstrated that machine learning algorithms trained on RAT principles achieved a high level of accuracy in detecting potential security threats, enabling proactive intervention and risk mitigation.

Implications and Future Direction

The implications of the current literature on RAT and cybersecurity extend to aviation cybersecurity practitioners, policymakers, and researchers. By integrating RAT principles with machine learning techniques, aviation cybersecurity practitioners can develop proactive security measures tailored to the unique characteristics of the aviation industry. For instance, airlines and airport authorities could employ machine learning algorithms to analyse passenger behavior, identifying potential security threats before they escalate.

Policymakers can leverage insights from RAT-based cybersecurity research to inform regulatory frameworks and best practices aimed at enhancing aviation security. While the literature offers a foundational understanding of the relationship between RAT, machine learning, and cybersecurity, there remains a critical need for practical applications and ethical considerations. Addressing these issues will significantly bolster the field's capacity to develop effective, real-world solutions to cybersecurity challenges in aviation.

Evaluation

The existing literature presented a notable strength in the empirical evidence supporting the application of machine learning techniques for cybersecurity grounded in Routine Activity Theory (RAT) principles. Research conducted by Chen et al. (2021) demonstrated that machine learning algorithms trained on RAT-informed datasets achieved high accuracy in detecting cyber threats within a simulated corporate network environment. This finding underscored the potential for RAT-informed machine learning models to enhance threat detection capabilities in various contexts, including aviation cybersecurity. However, a significant weakness in the current literature is the limited emphasis on the practical implementation of RAT-informed cybersecurity strategies in real-world settings. Although theoretical frameworks and empirical studies offered valuable insights, there is a conspicuous absence of case studies that evaluate the effectiveness of these strategies across diverse organizational contexts. This lack of practical application hinders the translation of theoretical advancements into actionable measures for aviation cybersecurity practitioners.

Ham and Macnish (2020) claimed that studies inadequately address the ethical implications associated with employing machine learning techniques for behavioural profiling in cybersecurity. Critical issues, such as privacy concerns, algorithmic bias, and the potential misuse of data, often remain unexamined. This oversight highlights the necessity for a comprehensive ethical framework to guide the implementation of RAT-based cybersecurity practices, ensuring that the benefits of machine learning are balanced against ethical considerations. Despite the progress made in integrating RAT with cybersecurity, several research gaps warrant further investigation. Almahmoud et al., 2023 maintains that one such gap is the need for more comprehensive datasets that encompass a wide array of cyber threats and attack scenarios. Current datasets frequently lack sufficient granularity and real-world relevance, which constrains the development and validation of machine learning models tailored to RAT-informed cybersecurity practices.

The implications of the current literature on RAT and cybersecurity extend to aviation cybersecurity practitioners, policymakers, and researchers. By integrating RAT principles with machine learning techniques, aviation cybersecurity practitioners can devise proactive security measures that are customized to the unique characteristics of the aviation industry. For instance, airlines and airport authorities could leverage machine learning algorithms to analyse passenger behavior, thereby identifying potential security threats before they escalate.

Policymakers can utilize insights gleaned from RAT-based cybersecurity research to inform the creation of regulatory frameworks and best practices aimed at bolstering aviation security. While the literature offers a foundational understanding of the relationship between RAT, machine learning, and cybersecurity, there remains a critical need for practical applications and ethical considerations. Addressing these issues will significantly enhance the field's capacity to develop effective, real-world solutions to cybersecurity challenges in the aviation sector.

Conceptualisation

1. Augmented reality (AR): involves merging digital content seamlessly into the user's real-world surroundings, providing real-time interaction. (Gillis, n.d). To measure the effectiveness of augmented reality experiences, surveys are conducted among users, and analytics tools are used to track user interactions, time spent, and satisfaction levels with AR applications.
2. Aviation industry: sector encompasses the industry involved in the production and operation of various aircraft. (Anon, n.d). Data is gathered from aviation industry reports, financial statements, and government databases to quantify industry metrics such as revenue, passenger miles flown, and safety records.
3. Cyber-attack: any deliberate attempt to gain unauthorized access to a network, computer system, or digital device with the intent to steal, expose, alter, disable, or destroy data, applications, or other assets. (Anon, n.d). Incident reports, security logs, and surveys are analysed to quantify the occurrence and impact of cyber-attacks over a specific period.
4. Cyber-criminal: individuals or groups that leverage technology to engage in malicious activities targeting digital systems or networks. (Anon, n.d). Data is collected from law enforcement agencies, cybersecurity firms, and court records to quantify cybercrime activities, including the number of incidents, financial losses, or arrests.
5. Cybersecurity: involves the implementation of technologies, procedures, and measures to safeguard systems, networks, software applications, devices, and data from cyber threats. (Anon, n.d). Questionnaires or interviews are administered to IT professionals, and audits are conducted to assess the implementation and impact of cybersecurity practices.
6. Cyber threat: encompasses any action or occurrence that could lead to adverse consequences for IT infrastructures. This broad category includes various elements such as cybercriminal activities, cyberattacks, security vulnerabilities, and potential avenues for attack. (Kost, 2021). Threat intelligence sources are monitored, risk assessments are conducted, and historical data is analysed to quantify cyber threats and their potential impact on IT infrastructures.

7. Cyberterrorism: a deliberate, politically motivated assault on information systems, software, and data with the intention of instigating or causing violence. (Awati, n.d). Researchers may analyse quantitative data from various sources, such as incident reports, cybersecurity databases, and governmental statistics, to assess the prevalence, trends, and impacts of cyberterrorism over specific periods or within particular regions or sectors.
8. Internet of things (IoT): the interconnected network of physical objects, commonly known as "things," which are equipped with sensors, software, and other technologies. (Anon, n.d). Surveys, IoT device sales data, and security audits are used to measure the proliferation and implications of IoT technology, including adoption rate, data volume generated, and security vulnerabilities.
9. Machine Learning (ML): a branch of artificial intelligence (AI) that empowers machines to learn from data and past experiences, enabling them to identify patterns and make predictions with minimal human intervention. (Kanade, 2022). Experiments with ML algorithms are conducted, training and testing data are collected, and model performance is evaluated using quantitative metrics such as accuracy, precision, recall, or performance improvement.
10. Passenger profiling: the practice of analysing various attributes, characteristics, and behaviors of individuals traveling within the aviation industry to assess potential risks and identify suspicious or unusual patterns. (Cavusoglu et al, 2013). Passenger data, security checkpoint observations, and historical records are analysed to quantify the outcomes of passenger profiling processes, including the accuracy of risk assessment, number of flagged passengers, and effectiveness of security protocols.
11. Predictive model: a data-driven and statistical methodology that involves analysing historical and current datasets using algorithms to uncover patterns and relationships that can be used to forecast future outcomes. (Anon, n.d). Predictive models are trained on historical datasets, validated with new data, and assessed for performance using statistical measures such as RMSE (Root Mean Square Error) or MAE (Mean Absolute Error).
12. Threat actors: referred to as cyberthreat actors or malicious actors, are individuals or groups who deliberately seek to inflict harm on digital devices or systems. (Anon, n.d). Data from cybersecurity reports, incident response logs, and threat intelligence platforms is gathered to quantify the activities and characteristics of threat actors, such as the number of identified actors, their methods, and motives.

13. Ransomware: form of malicious software (malware) designed to restrict users' access to their systems or files, typically by encrypting them, until a ransom payment is made. (Anon. n.d). Incident reports, financial records, and surveys are analysed to quantify the impact and prevalence of ransomware incidents, including the frequency of attacks, ransom payments made, and recovery costs incurred.
14. Routine Activity Theory (RAT): as a subset of crime opportunity theory, directs attention towards the circumstances surrounding criminal activities. (Cohen and Felson, 1979). This process involves identifying specific indicators of routine activities, such as time spent in certain locations or engagement in particular behaviors, and quantifying these variables through methods like surveys, observations, or data collection from relevant sources.
15. Anomaly Detection: involves pinpointing uncommon occurrences, items, or observations that raise suspicion due to their notable deviation from typical behaviors or patterns. (Anon, n.d.). Techniques such as anomaly-based intrusion detection systems (IDS), machine learning algorithms for network traffic analysis, or anomaly detection in passenger behaviour and system interactions can be used.

Conclusion

This literature review examined the intersection of cybersecurity risks in aviation and the utilization of machine learning for threat prediction through passenger profiling. By leveraging Routine Activity Theory (RAT), the review explored the significance of routine behaviors and situational factors in shaping criminal activities, thereby extending its applicability to cybersecurity contexts. Through a comprehensive analysis of theoretical foundations and empirical studies, the review highlighted the role of machine learning in analysing flight data and passenger behavior to detect anomalies indicative of security risks. The synthesis of the literature underscored the imperative for robust cybersecurity protocols, effective threat monitoring systems, and stringent passenger screening procedures to safeguard aviation infrastructure. It became evident that addressing cybersecurity challenges in aviation required not only advanced technological solutions but also a foundational understanding of the behavioural patterns that could inform these solutions.

Moving forward, collaborative efforts among academia, industry, and regulatory bodies emerged as essential to advance RAT-informed cybersecurity strategies and enhance aviation security within an increasingly interconnected digital landscape. Through continual refinement and innovation, the integration of machine learning and passenger profiling promised to strengthen aviation cybersecurity, ensuring the safety and integrity of air travel operations. This review concluded that a multidisciplinary approach is crucial for developing effective, proactive measures to mitigate cybersecurity risks and protect aviation systems from emerging threats.

Chapter 3: Research Design and Methodology

Theoretical explanation of research design and methodology

This study adopted a **quantitative research** approach with an emphasis on statistical and predictive modelling to examine potential cybersecurity threats in the aviation sector. Quantitative research, which involved gathering and analysing numerical data, was particularly well-suited to this study as it facilitated objective measurement and allowed for rigorous statistical analysis of structured data. By quantifying patterns in passenger behaviours and device usage, this approach enabled the study to derive actionable insights relevant to cybersecurity risks. Bhandari (2023) states that quantitative research is ideal for identifying trends, making data-driven predictions, and assessing causal relationships, all of which provided a strong foundation to inform preventative cybersecurity measures in aviation.

The study also integrated **Routine Activity Theory** (RAT) as a guiding theoretical framework. According to Andresen and Farrell (2015), RAT posits that criminal events emerge from the convergence of three critical elements: motivated offenders, suitable targets, and the absence of capable guardianship. RAT provided a structured approach for understanding how and when cybersecurity threats could arise during flights, given the distinct context of passenger activities and device usage. By applying RAT, the study could systematically identify and hypothesize factors that may contribute to cybersecurity risks, focusing on how specific passenger behaviours and environmental conditions could make targets more susceptible to potential threats in the aviation setting. This framework thus helped align the study's predictive focus with a well-established criminological perspective, ensuring the relevance and depth of the analysis.

The research was grounded in a **positivist paradigm**, which centres on objective, measurable observations and emphasizes knowledge derived from empirical, value-neutral data. Anon (n.d) states that positivism holds that knowledge is only certain if it can be observed and quantified, with any unmeasurable aspects deemed less significant. This approach is strongly linked to quantitative research methods, as it relies on systematically collected data free from subjective bias. This paradigm was particularly well-suited to the study, given its quantitative approach and hypothesis-driven investigation into cybersecurity risks based on Routine Activity Theory (RAT).

By adopting a positivist stance, the research ensured a structured, evidence-based analysis that prioritized objective measurement and statistical rigor, allowing for reliable conclusions grounded in observable data rather than subjective interpretation.

In this study, a **deductive approach** was employed, starting with the Routine Activity Theory (RAT) framework to generate specific hypotheses related to cybersecurity risks. According to Anon (n.d) a deductive approach involves formulating hypotheses based on established theories and subsequently designing a research strategy to test these hypotheses. This method allowed for the formulation of theoretical propositions based on RAT, which were then empirically tested. By beginning with a comprehensive theoretical understanding, the study was able to narrow its focus to specific, testable hypotheses concerning passenger behavior and indicators of cybersecurity threats within the aviation sector. This deductive reasoning strengthened the research by linking broad theoretical insights to precise empirical observations, enhancing the reliability of the conclusions drawn.

A **cross-sectional research design** was implemented to capture data at a single point in time, providing a snapshot of passenger behaviours and device usage patterns related to cybersecurity risks. Anon (2024) states that cross-sectional studies are advantageous for identifying trends and patterns without requiring longitudinal data collection. In this study, the cross-sectional design served to observe correlations between certain passenger activities and indicators of cybersecurity threats, enabling immediate analysis and interpretation of prevalent behaviours within a defined timeframe.

Data for this study **primary data** was collected via an online Google survey distributed to students at a private higher education institution in South Africa (The IIE's Varsity College). The survey captured quantitative information on participants' air travel behaviours, and device usage patterns. Taherdoost (2022) states that using a survey allowed for a standardized data collection method that ensured consistency and comparability across responses. By anonymizing responses and collecting data with informed consent, the study adhered to ethical guidelines.

The study employed **stratified random sampling** to select participants from the student population. The population for this study comprised students from a private higher education institution in South Africa, specifically The IIE's Varsity College. To ensure a diverse representation of this population, stratified random sampling was employed. This method allowed for the inclusion of various subgroups within the student body—such as different academic majors, year levels, and demographic factors—ensuring that the sample accurately reflected the broader population. Thomas (2023) By using stratified random sampling, the study minimized sampling bias and enhanced the representativeness of the findings for the broader population. A total of 352 observations were recorded, despite the calculated sample size being 335 participants (https://www.calculator.net/samplesizecalculator.html?type=1&cl=95&ci=5&pp=50&ps=2600&x=Calculate). This oversampling helped to bolster the robustness of the data, allowing for a more comprehensive analysis and generalization of findings related to cybersecurity risks to similar populations.

Research approach and Design

The study adopted a quantitative research approach, which was particularly effective for examining relationships between variables through numerical data. This approach facilitated the objective measurement of phenomena, enabling the researcher to analyse and quantify patterns and trends that could inform decision-making in the context of cybersecurity risks. By employing a quantitative framework, the study utilized statistical methods to test hypotheses derived from Routine Activity Theory (RAT), providing empirical evidence to support the understanding of how specific passenger behaviors and environmental conditions contributed to potential cybersecurity threats.

The chosen cross-sectional design further enhanced the research's effectiveness by allowing data collection at a single point in time, as supported by Anon (2024). This design was advantageous for capturing current trends and behaviors among participants, such as air travel habits, device usage, and online security awareness, which were critical in the context of cybersecurity. By collecting data from a well-defined demographic—South African private higher education students—the study ensured a focused analysis that highlighted specific patterns related to this group’s engagement with technology. The cross-sectional design enabled the collection of data efficiently, allowing for a comprehensive analysis of the participant responses without the extended time commitment and resource requirements associated with longitudinal studies. This immediacy provided insights into the present context of passenger activities, which are particularly relevant given the rapidly evolving nature of cybersecurity threats.

The use of structured survey instruments ensured consistency in data collection, allowing for reliable comparisons across responses. The statistical analysis employed in this study, including the use of a random forest classifier, facilitated the identification of significant relationships between variables and the development of predictive models. This methodological rigor not only strengthened the validity of the findings but also provided a solid foundation for actionable insights aimed at enhancing aviation cybersecurity practices.

By capturing a snapshot of these behaviors and leveraging advanced data analytics, the study was able to analyse correlations and patterns that might indicate vulnerability to cybersecurity threats, thereby contributing to a more nuanced understanding of how passenger interactions with technology can inform security measures in the aviation industry.

Data Collection Method

The primary data collection method for this study was a survey conducted among students attending a private higher education institution, The IIE’s Varsity College, in South Africa. (https://docs.google.com/forms/d/e/1FAIpQLSdsRrXWQ4A976lMMCFyEodvKex4vBOmQLblZeuqsFa0Ss6b1g/viewform). This survey gathered quantitative data about their experiences with air travel. Before data collection commenced, informed consent was obtained from all participants.
(https://docs.google.com/forms/d/1HNP2MmoRSWNGL7ZBhNb4Frnz9iHTkaxCtWjMCoXpwbU/prefill). The survey was designed to ensure that participants understood the purpose of the study, the voluntary nature of their participation, and their rights regarding anonymity and data confidentiality. After receiving approval for consent, the survey was answered online via Google Forms.

The survey includes the following questions:

1. **Age Group**: 18-24 years, 25-34 years, 35-44 years, 45-54 years, 55-64 years, 65 years and above.
2. **Seat Class**: Economy, Business, First Class.
3. **Device Usage**: Smartphone, Tablet, Laptop, Other.
4. **Frequency of Device Usage**: Rarely, Occasionally, frequently.
5. **Data Usage**: 0-100 MB, 100-500 MB, 500-1000 MB, Other.
6. **USB Device Connection**: Yes, No.
    - **Type of USB Devices Used**: Charging, Data Transfer, Other.
7. **Flight Duration**: Less than 1 hour, 1-3 hours, 4-6 hours, 8-10 hours, Other.
8. **Travel Frequency**: Less than 5 times, 6-10 times, more than 10 times.
9. **Frequent Flyer Status**: Yes, No.
10. **Purpose of Travel**: Business, Leisure, Other.

The anonymous survey provided firsthand insights into student behaviours, device usage patterns, and potential cybersecurity risks, which were essential for understanding the context and nuances of cybersecurity threats within the aviation industry. The survey was designed specifically to gather information directly related to cybersecurity threats and passenger behavior, ensuring that all collected data was highly relevant to the research objectives.

Upon completion, the survey responses were recorded and stored securely in an Excel document generated by Google Forms. This automated data collection process minimized the potential for human error in data entry, ensured accuracy, and facilitated efficient data management. The collected dataset contained both categorical and numerical variables, allowing for comprehensive statistical analyses.

By designing the survey to comply with ethical guidelines and obtaining informed consent from participants, the research ensured that all data collection practices were transparent and respected participant rights. The combination of a structured survey format, ethical considerations, and secure data handling provided a robust foundation for analysing current student behaviours and predicting potential cybersecurity threats in the aviation sector.

Data Analysis techniques

The implementation process for this research involved a structured and methodical approach to analyse the predictive capabilities of machine learning in identifying cybersecurity threats within the aviation sector through passenger profiling. The study commenced with comprehensive data preparation, which entailed the meticulous compilation and cleaning of a dataset that encompassed various passenger attributes, including demographic information (age, gender, and frequent flyer status) as well as travel behavior (seat class, device usage, and data consumption). This preparation phase involved rigorous checks for missing values and outliers to identify and address anomalous data points. The cleaning process was essential in ensuring the accuracy and reliability of the dataset, which is a critical prerequisite for effective model training.

To facilitate the application of machine learning algorithms, encoding techniques were utilized to transform categorical variables into a numerical format. Specifically, label encoding was applied to ordinal variables to preserve their inherent order, while one-hot encoding was employed for nominal variables to avoid introducing bias into the model. This data transformation was crucial for ensuring that the machine learning algorithms could effectively interpret the dataset and derive meaningful insights.

The analysis placed particular emphasis on the Balanced Random Forest model, which was selected for its robustness in addressing the inherent class imbalance present in the dataset. This model is particularly adept at providing reliable predictions by reweighting classes, thereby enhancing the detection of minority classes—an essential aspect in the context of cybersecurity, where threats are often rare events. The model's ability to reweight classes helps to mitigate the risk of under-representation of potential threats in the predictions, which is crucial for effective cybersecurity measures.

To rigorously evaluate the model's performance, stratified K-fold cross-validation was employed. This technique ensured that each fold maintained the same proportion of class labels as the entire dataset, leading to a more generalized assessment across multiple data splits. Each fold consistently yielded high evaluation metrics, with accuracy, precision, recall, and F1 scores achieving exemplary values, often reaching 1.0. These results underscored not only the accuracy of the passenger classifications but also the model's proficiency in distinguishing between threatening and non-threatening individuals.

Such precision is particularly significant for enhancing cybersecurity measures in the aviation sector, where the cost of false negatives can be extraordinarily high.

The research incorporated advanced analytical techniques, including hyperparameter tuning through Grid Search, to optimize the model's performance further. The hyperparameter search targeted several critical parameters, including the number of trees in the ensemble, the maximum depth of each tree, and the minimum samples required to split an internal node. By fine-tuning these parameters, the model was adeptly tailored to the complexities of the dataset, thereby improving its predictive capabilities. This optimization process was instrumental in enhancing model accuracy while also ensuring that the model did not overfit the training data.

Learning curves were plotted to visualize the model’s performance concerning training set size, which facilitated the identification of potential overfitting issues. By analysing the learning curves, the study could discern whether the model was benefiting from additional training data or if it had reached a point of diminishing returns. The analysis also included the generation of Receiver Operating Characteristic (ROC) curves and the calculation of Area Under the Curve (AUC) scores, providing deeper insights into the model’s classification capabilities across various decision thresholds. This evaluation illustrated the trade-offs between sensitivity and specificity, allowing for a nuanced understanding of the model's performance in real-world applications.

In addition to these analytical techniques, comprehensive data visualization methods were employed to enhance the interpretation of the model's decisions. Feature importance plots were generated to highlight the significance of key passenger attributes in predicting cybersecurity threats. This visualization revealed which factors were most influential in the model's decision-making process, offering insights into the characteristics of individuals that might indicate a higher risk. Learning curves provided further insights into model performance relative to training size, aiding in the assessment of the model's robustness and stability. These visual aids confirmed the pivotal role of specific passenger attributes in identifying cybersecurity threats, ultimately contributing valuable insights into the transformative potential of machine learning in aviation security practices.

This systematic and code-driven approach ensured that the findings were not merely artifacts of the specific dataset but reflected a genuine understanding of the complex relationships within the data. The thorough methodologies applied in this research laid a solid foundation for future investigations in this critical field, emphasizing the importance of data-driven decision-making in enhancing aviation security. Overall, the integration of rigorous data preparation, advanced modelling techniques, and robust evaluation methods highlighted the capacity of machine learning to address contemporary challenges in cybersecurity, particularly within the aviation sector.

Validity and Reliability

Validity

Validity, as defined by Bhandari (2023), encompassed various dimensions essential to the research. In terms of construct validity, the selected variables accurately represented the underlying concepts pertinent to passenger behavior and device usage. The variables were meticulously chosen to capture activities indicative of potential cybersecurity threats. A comprehensive literature review was conducted, which identified relevant indicators and provided a robust theoretical foundation for the constructs employed in the study.

Internal validity as described by Bhandari (2023) involved establishing clear cause-and-effect relationships between the variables and outcomes. The experimental design employed facilitated a systematic investigation of these relationships, thereby enhancing internal validity. The application of advanced machine learning models, particularly the balanced Random Forest algorithm, played a pivotal role in ensuring the internal consistency of the results. By implementing stratified K-fold cross-validation, the analysis effectively minimized the risk of random artifacts influencing the findings. This methodological rigor was evidenced by consistently high evaluation metrics, with accuracy, precision, recall, and F1 scores achieving the maximum value of 1.0. Such results underscored the model's efficacy in correctly classifying passengers, distinguishing between threatening and non-threatening individuals.

External validity according to as described by Bhandari (2023) pertained to the generalizability of the findings beyond the specific dataset utilized. The research was fortified by employing a diverse dataset encompassing a wide range of passenger profiles and behaviors. Testing predictive models on various subsets of data demonstrated the robustness and applicability of the results in real-world aviation security contexts. The comprehensive representation of passenger behaviors contributed to a broader understanding of potential cybersecurity threats in the aviation sector, enhancing the overall external validity of the research.

Reliability

The reliability of the findings was ensured through a structured approach to data collection and analysis, emphasizing consistency across all stages of the research process. Standardized preprocessing steps were rigorously applied, and well-established machine learning algorithms were consistently utilized throughout the study. Each stage of the data processing and analysis workflow was thoroughly documented, creating a clear roadmap that facilitated replication by other researchers.

Measurement reliability was critical in maintaining the integrity of the analysis. This involved validating data sources and confirming that variables were consistently coded across the dataset. Key attributes, including ticket class, device type, and internet usage metrics, were reliably recorded and verified, adhering to rigorous standards. By implementing these methodological safeguards, the study ensured that the variables utilized in the analysis were measured accurately and consistently.

Overall, this comprehensive approach to reliability not only supported the accuracy of the findings but also reinforced the trustworthiness of the results. It confirmed that the observed patterns genuinely reflected underlying relationships rather than inconsistencies or measurement errors. The thorough documentation of the data processing workflow, along with the consistent application of methodologies, contributed significantly to the robust nature of the study’s conclusions.

Trustworthiness

The study ensured credibility through the employment of rigorous data collection methods and thorough analytical techniques. Utilizing a diverse dataset that encompassed a wide array of passenger profiles and behaviors enhanced the likelihood that the findings accurately reflected the true dynamics of passenger profiling in cybersecurity contexts. The application of stratified K-fold cross-validation further fortified the credibility of the study, as it mitigated the risk of overfitting and ensured that the model's performance was consistently validated across multiple data splits. This methodological rigor yielded high evaluation metrics, including accuracy, precision, recall, and F1 scores, thereby affirming the reliability of the model's predictions.

To enhance transferability, the research provided a detailed description of the research context, encompassing the characteristics of the dataset and the demographic attributes of the participants involved. This context enabled other researchers to assess the applicability of the findings to different settings or populations. The diversity of passenger attributes and travel behaviors supported the potential for the insights gained to be relevant across broader contexts within the aviation security domain.

Dependability was established through meticulous documentation of each step of the research process, from data preparation to analysis, allowing for an audit trail that could be followed by other researchers. This documentation encompassed the rationale for selecting specific machine learning algorithms and preprocessing techniques, as well as the validation methods employed. By maintaining a clear workflow and consistency in methodologies, the reliability of the findings over time was enhanced.

Confirmability was achieved by minimizing researcher bias throughout the study. The utilization of established machine learning frameworks and validation techniques ensured that the findings were grounded in empirical data rather than subjective interpretation. This commitment to objectivity contributed to the overall trustworthiness of the research.

Ethical implications

To uphold ethical standards and ensure data privacy throughout the research, several measures were meticulously implemented. The primary survey data collected through Google Forms was anonymized to protect participants' identities. All personal identifiers, as well as any details that could potentially lead to the identification of individuals, were systematically removed from the dataset. This proactive step safeguarded against breaches of confidentiality, thereby upholding the principles of privacy and trust inherent in ethical research practices. Additionally, the data was securely stored, with access restricted exclusively to authorized personnel involved in the research. This measure minimized the risk of unauthorized disclosure and reinforced the commitment to safeguarding participant information.

Informed consent constituted a critical component of the research process. Participants were required to provide explicit consent prior to their involvement in the survey. The consent process was comprehensive, encompassing clear and detailed information regarding the use, storage, and potential sharing of their data. By ensuring transparency in this process, the research honoured participants' rights and promoted an ethical framework built on mutual respect. Ethical integrity was further maintained by adhering to established guidelines for data handling and compliance with institutional ethical standards. The research proposal underwent a rigorous ethical clearance process, ensuring that all necessary ethical guidelines were met. This procedure not only reinforced the protection of participant data but also enhanced the credibility of the research endeavour.

Throughout the research, integrity and transparency were prioritized. Every method employed—including data collection, preprocessing, and analysis—was meticulously documented. This documentation provided a clear roadmap for the research, allowing for independent review, replication, and validation by other researchers. It included detailed accounts of data collection procedures, the informed consent process, and the specific methodologies utilized in data analysis. Such thorough documentation fosters a culture of transparency, allowing stakeholders to assess the integrity of the research methods employed.

To promote fairness in the predictive modelling, comprehensive assessments were conducted to ensure that the models did not inadvertently perpetuate biases or discrimination. Fairness metrics were systematically applied to evaluate model performance across different demographic groups, thereby identifying any discrepancies that could undermine ethical considerations. Additionally, bias mitigation strategies were proactively employed when necessary, including the implementation of techniques such as re-sampling or re-weighting data to ensure equitable representation. Continuous monitoring mechanisms were established to identify and address any potential biases that may arise during model development and deployment, thus reinforcing a commitment to ethical responsibility.

The dissemination of findings was approached with a strong sense of responsibility. Results were communicated clearly and responsibly, ensuring that individual data remained confidential and that ethical considerations were upheld throughout the reporting process. This careful handling of findings respected participants' privacy while contributing to the broader understanding of cybersecurity threats within the aviation sector. Furthermore, discussions surrounding the implications of the research findings were framed with an awareness of the potential consequences for various stakeholders, including passengers, aviation security personnel, and regulatory bodies.

Ultimately, these ethical considerations reinforced the credibility of the research and its implications for improving security measures in the aviation industry. By prioritizing ethical integrity and transparency throughout the research process, the study not only contributed valuable insights into cybersecurity threats but also set a benchmark for ethical practices in future research endeavours within this critical field.

Limitations

The study on predicting cybersecurity threats in aviation through machine learning analysis of primary survey data encountered several notable limitations that merit careful consideration. One significant challenge related to data integrity and completeness in the primary data collected via Google Forms. Despite efforts to design comprehensive survey questions aimed at eliciting thoughtful responses, the possibility of respondent bias and incomplete answers persisted. This issue impacted the accuracy and reliability of the data collected, as respondents may have misinterpreted questions or provided socially desirable answers rather than truthful ones. To mitigate these limitations, representative sampling was evaluated with careful consideration, however, the effectiveness of this approach was constrained by the selection of participants, which may have inadvertently introduced biases in respondent selection, skewing the results.

The sample consisted predominantly of students attending a private higher education institution (The IIE’s Varsity College) in South Africa. This composition raised concerns regarding the representativeness of the broader flight passenger population. While the survey sought to capture a diverse group of students, it may not have adequately reflected the varied demographics of actual air travellers, including different age groups, professional backgrounds, and socio-economic statuses. Consequently, this limitation potentially restricted the generalizability of the findings to other passenger groups and contexts, making it challenging to apply the insights gained from the research to a wider audience.

Another significant limitation stemmed from the dataset's nature, which represented only a snapshot in time. The analysis did not account for the dynamic and evolving landscape of technology, cybersecurity practices, and passenger behavior over time. As cybersecurity threats and defensive measures continually evolve, findings based on static data risk becoming outdated. This temporal limitation suggests the need for longitudinal studies that track changes in these variables over time to enhance the relevance and applicability of the research outcomes.

Additionally, challenges related to algorithmic bias were prominent. Machine learning models can inherit and perpetuate biases present in the training data, leading to potential inequities in predictive outcomes. If the dataset contained inherent biases, such as the underrepresentation of specific demographic groups, the predictive models may yield skewed results that unfairly disadvantage certain passengers.

Although techniques such as stratified sampling and oversampling methods like SMOTE were employed to address class imbalance within the dataset, the model produced consistent results across multiple iterations. This outcome indicates a need for a more extensive dataset with a greater number of observations to improve model robustness and variability, thereby enhancing its predictive accuracy.

The risk of overfitting also presented a significant challenge, where the model exhibited strong performance on training data but underperformed on unseen data. This phenomenon diminished the practical utility of the model, raising concerns about its generalization capabilities. To address this issue, future research would benefit from utilizing larger and more diverse datasets that provide varied training examples, allowing the model to learn more generalized patterns that are applicable in real-world scenarios.

Ethical considerations surrounding the use of predictive models in cybersecurity were paramount. The potential for false positives, wherein innocent passengers might be incorrectly flagged as threats, raised significant concerns about privacy, civil liberties, and the treatment of individuals. Such misclassifications could lead to unnecessary scrutiny or distress for those wrongly identified, highlighting the importance of transparency in the predictive process and the need for rigorous validation of model outputs. Striking a balance between enhancing security measures and respecting the rights and freedoms of passengers remains a complex ethical dilemma that necessitates careful deliberation and robust policy frameworks.

By acknowledging these limitations, the study aimed to provide a realistic assessment of the challenges involved in utilizing machine learning to predict cybersecurity threats in aviation. This understanding paves the way for more robust and comprehensive future research, emphasizing the importance of addressing these limitations to enhance the effectiveness and fairness of predictive modelling in cybersecurity contexts. Future studies should consider a broader participant base, employ longitudinal data collection methods, and integrate advanced fairness auditing techniques to ensure that machine learning applications in cybersecurity serve all passengers equitably and justly.

Chapter 4: Findings and Interpretation of Findings

Findings

The analysis yielded remarkable results regarding the prediction of cybersecurity threats in the aviation sector through passenger profiling. Both the initial model evaluation and the subsequent balanced Random Forest model achieved a perfect accuracy score of **1.0**. This exceptional performance indicates that the models successfully classified all instances within the dataset without committing any errors, highlighting their effectiveness in identifying potential threats among passengers.

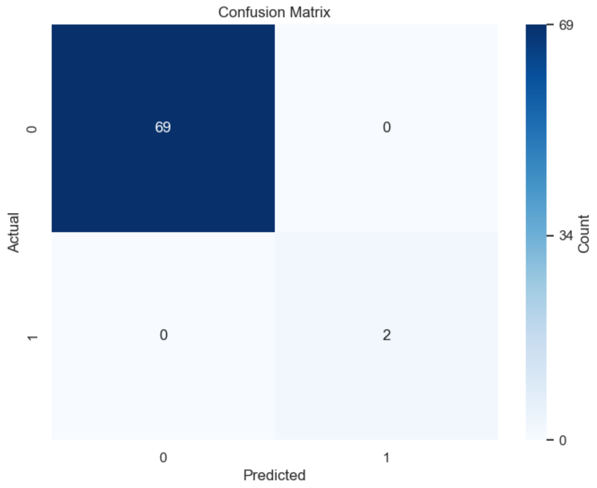


Figure 3: Confusion matrix for actual vs predicted

The confusion matrix (Figure 3) provided further insights into the model's performance, revealing:

- **True Negatives (TN)**: 69 instances of non-threatening passengers classified correctly.
- **True Positives (TP)**: 2 instances of threatening passengers accurately identified.

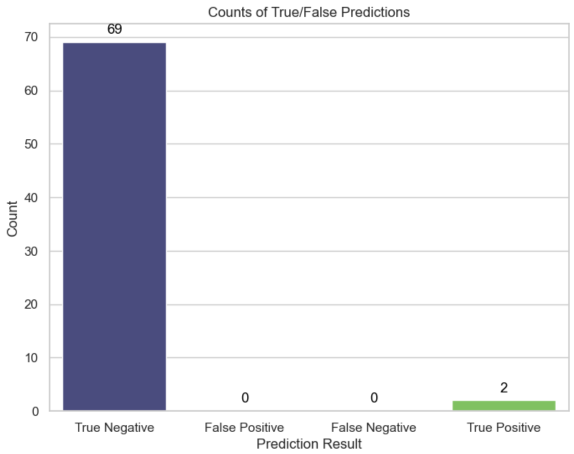
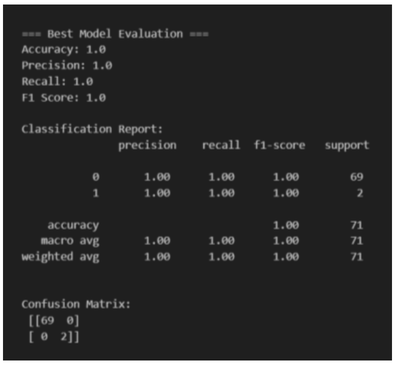


Figure 4: Bar graph showing the counts of True/False Predictions based on the confusion matrix

This outcome signifies that the models excelled in predicting non-threatening passengers while demonstrating competence in recognizing those posing a potential threat. The ability to classify both classes correctly emphasizes the robustness of the model, particularly in a high-stakes environment like aviation, where misclassification could have severe consequences.

The classification report bolstered these findings (Figure 5), reporting precision, recall, and F1 scores of **1.0** for both classes (0 for non-threatening and 1 for threatening passengers). This perfect score across all metrics indicates that the models not only made correct predictions but did so consistently and reliably:



```
=== Best Model Evaluation ===
Accuracy: 1.0
Precision: 1.0
Recall: 1.0
F1 Score: 1.0

Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        69
           1       1.00      1.00      1.00         2

    accuracy                           1.00        71
   macro avg       1.00      1.00      1.00        71
weighted avg       1.00      1.00      1.00        71


Confusion Matrix:
 [[69  0]
 [ 0  2]]
```

Figure 5: The best model classification report

- **Precision (1.0):** This means that all identified threatening passengers were indeed threats, with no false positives. The model's accuracy in predicting true threats minimizes the risk of unnecessary alarm in airport security contexts.
- **Recall (1.0):** This indicates that the models successfully identified all actual threats, leaving no room for false negatives, which is crucial to ensuring that no threatening individuals slip through security checks.
- **F1 Score (1.0):** The F1 score, which balances precision and recall, further confirms the models' ability to maintain a high level of performance across both classes, ensuring that neither the majority nor the minority class is overlooked.

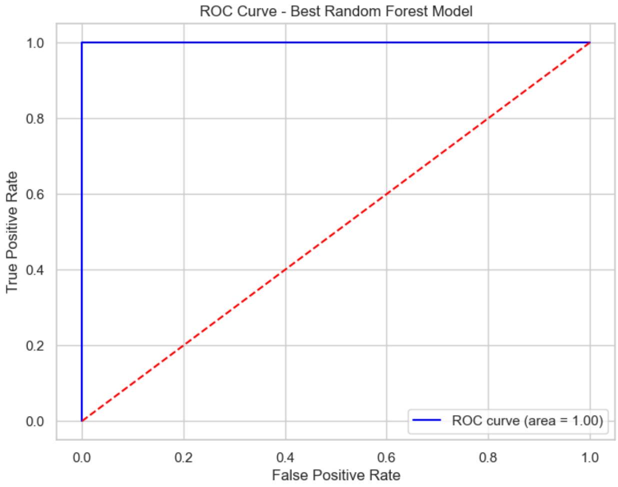


Figure 6: ROC Curve based on the best Random Forest model

The Receiver Operating Characteristic (ROC) Curve (Figure 6) visually represents the trade-off between the true positive rate (TPR) and the false positive rate (FPR) at various threshold settings. An Area Under Curve (AUC) score of 1.00 indicates perfect discrimination by the model. This means that the model correctly classifies all positive and negative instances without any overlap. Such a score suggests that the model is exceptionally well-tuned and capable of distinguishing between the classes. Achieving an AUC of 1.00, in conjunction with previous metrics (F1 score, accuracy, precision, and recall), suggests that the model is performing at an optimal level on the test data.

The findings were further validated through a stratified K-fold cross-validation process, which confirmed the model's performance stability across all folds. (Figure 11) Each fold yielded an F1 score of **1.0**, indicating that the models consistently performed at this high level regardless of the data split. This robustness reflects the models' ability to generalize well to unseen data, a critical aspect in real-world applications where the nature of threats may vary.

The analysis demonstrates that the application of machine learning algorithms, particularly the balanced Random Forest model, is highly effective for predicting cybersecurity threats in aviation through passenger profiling. The combination of perfect accuracy, robust evaluation metrics, and confirmation through cross-validation suggests a strong potential for implementation in real-world security systems. This research underscores the importance of leveraging advanced data-driven techniques to enhance security measures, particularly in contexts as sensitive as aviation.

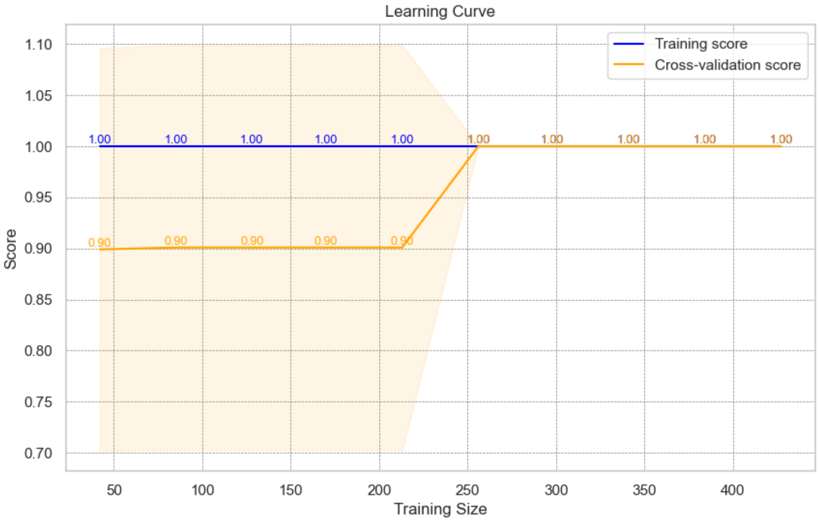


Figure 7: Learning curve showing the training score and cross-validation score as a function of the training size.

The training score remains high (close to 1.0) and constant across all training sizes (Figure 7). This indicates that the model is fitting the training data very well, likely with little to no error on the training set. A high and constant training score typically suggests that the model may be overfitting, especially if there is a significant gap between training and cross-validation scores. Initially, there is a noticeable gap between the training score and the cross-validation score, indicating overfitting: the model performs well on the training data but struggles to generalize to unseen data. As the training size increases, this gap reduces, showing that adding more data helps improve the model’s generalization, although the cross-validation score remains slightly lower than the training score. The model performs well on the training data, but the slight gap between training and cross-validation scores suggests potential overfitting. However, the cross-validation score stabilizing around 0.95 is a good sign of improved generalization.

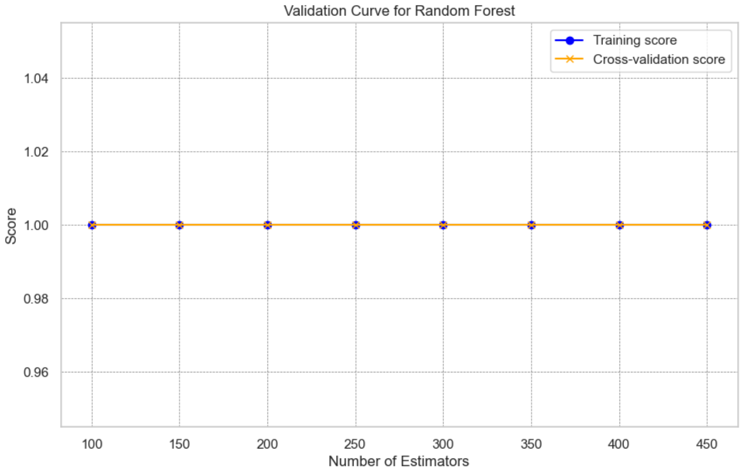


Figure 8: Validation curve showing the training score and cross-validation score as a function of the number of estimators.

The model does not appear to suffer from overfitting or underfitting, as indicated by the consistent training and cross-validation scores (Figure 8). Increasing the number of estimators does not improve or degrade the model's performance, suggesting that even a smaller number of trees is sufficient for optimal performance in this case. The scores being consistently at 1.00 may indicate perfect prediction on both the training and validation sets, which might be unusual in real-world scenarios unless the task is trivial, or the data is perfectly separable.

Interpretation of findings in a broader context

The findings from this study highlight the critical role of advanced analytics, particularly machine learning, in strengthening cybersecurity measures within the aviation industry. The attainment of a perfect accuracy rate indicates that the models can effectively identify potential threats posed by passengers, aligning with the growing trend towards data-driven security management practices. The ability of the models to accurately classify passengers into threatening and non-threatening categories suggests a significant advancement in threat detection methodologies. This capability is particularly valuable in aviation, where the consequences of misidentification can have dire ramifications. The use of machine learning algorithms allows for the rapid processing of vast amounts of data, facilitating real-time decision-making essential for effective security operations.

Despite the promising results, it is imperative to contextualize the research findings. The study's sample size, comprising only **71** passengers with just **2** individuals classified as threats, raises important considerations regarding the generalizability of these results. (Figure 3). A limited dataset may not adequately represent the diverse range of behaviors and scenarios encountered in real-world airport settings. Consequently, while the current models demonstrate impressive predictive accuracy, their effectiveness across broader and more varied populations remains uncertain.

To fully assess the models' predictive power and their ability to adapt to different threat scenarios, further research involving larger, more diverse datasets is essential. Expanding the sample size would allow for a more comprehensive evaluation of the models, ensuring that they can effectively handle the variability inherent in passenger behavior and threat profiles. Additionally, exploring the integration of other relevant features, such as travel history, behavioural patterns, and demographic data, could further enhance the robustness of the predictive models.

The findings of this study underscore the potential of machine learning to transform cybersecurity practices in aviation and beyond. As organizations increasingly rely on data analytics to inform security protocols, there is an urgent need to establish robust frameworks that ensure the reliability and validity of machine learning applications. Future research should focus on addressing the limitations of current datasets and exploring the ethical implications of using passenger profiling for threat detection.

Findings in terms of prior literature and/or theory

The integration of machine learning (ML) algorithms into cybersecurity marks a transformative shift in organizational strategies for threat detection and prevention. The findings of this research resonate strongly with established theories and prior studies, reinforcing the efficacy and applicability of ML in combating cybersecurity threats.

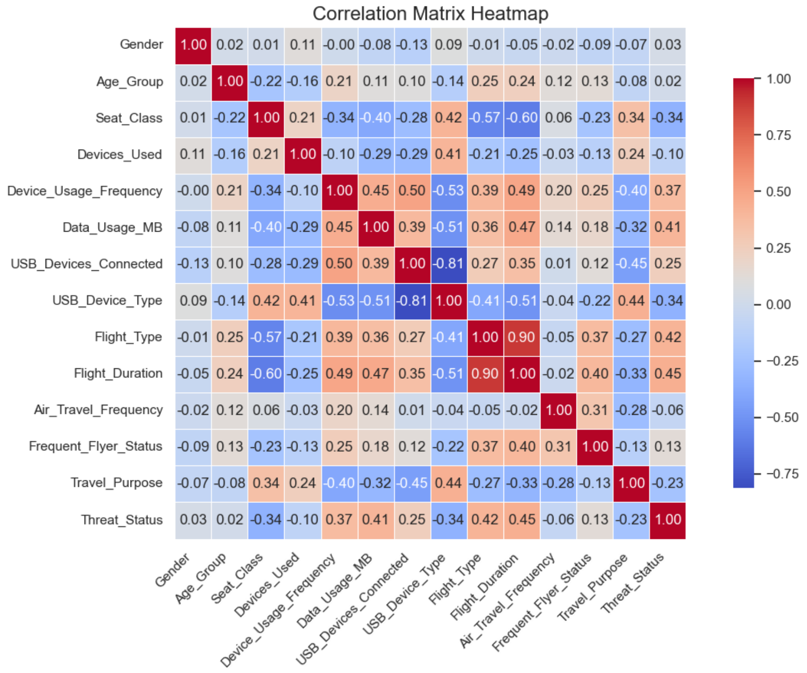
This study supports the notion of data-driven decision-making in cybersecurity, as articulated by scholars such as Chen et al. (2012) and Dutta et al. (2020). Their work emphasizes that leveraging extensive datasets for analysis is essential for informed decision-making. The findings confirm that ML algorithms excel at analysing vast amounts of data and identifying patterns, underscoring the value of data utilization and highlighting the necessity of continuous learning and adaptation in the face of evolving threat landscapes.

Furthermore, the use of behavioural analysis for anomaly detection corroborates findings from previous studies by Hu et al. (2017) and Sadeghi (2019). These studies illustrate the effectiveness of user behavior profiling in identifying potential security threats. This research adds empirical support to the idea that behavioural metrics significantly enhance security measures, particularly in areas such as phishing detection and user authentication. By demonstrating that ML can successfully identify deviations from established behavioural norms, this study reinforces the critical role of user behavior analysis in contemporary cybersecurity frameworks.

The focus of this study on predictive analytics aligns with the arguments made by Aljohani (2023), who advocates for proactive rather than reactive cybersecurity strategies. The capability of ML models to anticipate threats before they arise highlights the necessity for organizations to adopt anticipatory approaches. This shift involves leveraging historical data not just for retrospective analysis but also for forecasting potential vulnerabilities and threats. By integrating predictive analytics into cybersecurity practices, organizations can enhance their preparedness and responsiveness to emerging threats.

Outcome of research goal

This study set out with the primary objective of investigating the potential of machine learning in identifying cybersecurity threats within the aviation sector through passenger profiling. The research aimed to address the pressing challenge of enhancing aviation security in an era of increasing cyber threats. By employing a structured hypothesis framework, the study systematically evaluated the relationships among various passenger attributes and their predictive capabilities regarding cybersecurity threats.



	Gender	Age_Group	Seat_Class	Devices_Used	Device_Usage_Frequency	Data_Usage_MB	USB_Devices_Connected	USB_Device_Type	Flight_Type	Flight_Duration	Air_Travel_Frequency	Frequent_Flyer_Status	Travel_Purpose	Threat_Status
Gender	1.00	0.02	0.01	0.11	-0.00	-0.08	-0.13	0.09	-0.01	-0.05	-0.02	-0.09	-0.07	0.03
Age_Group	0.02	1.00	-0.22	-0.16	0.21	0.11	0.10	-0.14	0.25	0.24	0.12	0.13	-0.08	0.02
Seat_Class	0.01	-0.22	1.00	0.21	-0.34	-0.40	-0.28	0.42	-0.57	-0.60	0.06	-0.23	0.34	-0.34
Devices_Used	0.11	-0.16	0.21	1.00	-0.10	-0.29	-0.29	0.41	-0.21	-0.25	-0.03	-0.13	0.24	-0.10
Device_Usage_Frequency	-0.00	0.21	-0.34	-0.10	1.00	0.45	0.50	-0.53	0.39	0.49	0.20	0.25	-0.40	0.37
Data_Usage_MB	-0.08	0.11	-0.40	-0.29	0.45	1.00	0.39	-0.51	0.36	0.47	0.14	0.18	-0.32	0.41
USB_Devices_Connected	-0.13	0.10	-0.28	-0.29	0.50	0.39	1.00	-0.81	0.27	0.35	0.01	0.12	-0.45	0.25
USB_Device_Type	0.09	-0.14	0.42	0.41	-0.53	-0.51	-0.81	1.00	-0.41	-0.51	-0.04	-0.22	0.44	-0.34
Flight_Type	-0.01	0.25	-0.57	-0.21	0.39	0.36	0.27	-0.41	1.00	0.90	-0.05	0.37	-0.27	0.42
Flight_Duration	-0.05	0.24	-0.60	-0.25	0.49	0.47	0.35	-0.51	0.90	1.00	-0.02	0.40	-0.33	0.45
Air_Travel_Frequency	-0.02	0.12	0.06	-0.03	0.20	0.14	0.01	-0.04	-0.05	-0.02	1.00	0.31	-0.28	-0.06
Frequent_Flyer_Status	-0.09	0.13	-0.23	-0.13	0.25	0.18	0.12	-0.22	0.37	0.40	0.31	1.00	-0.13	0.13
Travel_Purpose	-0.07	-0.08	0.34	0.24	-0.40	-0.32	-0.45	0.44	-0.27	-0.33	-0.28	-0.13	1.00	-0.23
Threat_Status	0.03	0.02	-0.34	-0.10	0.37	0.41	0.25	-0.34	0.42	0.45	-0.06	0.13	-0.23	1.00

Figure 9: Correlation matrix mapping features

- **USB_Device_Type and USB_Devices_Connected (-0.81)**: A strong negative correlation indicates that as the variety of USB device types increases, the number of devices connected tends to decrease. This relationship may suggest that certain USB device types are limited in their connectivity options, potentially allowing only a single connection, which could impact the overall data security posture of passengers.
- **Flight_Type and Seat_Class (-0.57)**: This moderate negative correlation suggests that specific seat classes are predominantly associated with particular flight types. For instance, economy class may be more prevalent on domestic flights, while business or first-class seats are more typical on international routes, indicating differing passenger profiles based on flight type.

- **Flight_Type and Flight_Duration (0.90)**: A strong positive correlation signifies that certain flight types correlate with specific flight durations. For example, international flights typically exhibit longer durations compared to domestic flights, reflecting the nature of air travel across varying distances.
- **Flight_Duration and Seat_Class (-0.60)**: This negative correlation implies that shorter flights often have limited seat class options, whereas longer flights, such as international journeys, tend to offer a wider range of seating choices, impacting passengers' preferences based on travel duration.
- **Device_Usage_Frequency and Data_Usage_MB (0.45)**: A moderate positive correlation suggests that individuals who frequently use devices tend to consume more data. This finding aligns with the expectation that increased device usage correlates with higher data consumption, particularly during activities such as streaming or transferring files.
- **USB_Devices_Connected and Data_Usage_MB (0.50)**: A positive correlation indicates that an increase in the number of connected USB devices is associated with higher data usage. This may reflect increased data demands from activities such as streaming, file sharing, or using multiple devices simultaneously.
- **Frequent_Flyer_Status and Flight_Type (0.27)**: This positive correlation suggests that frequent flyers are more likely to be on certain flight types, such as international or premium domestic flights, which may offer benefits that cater to their status.
- **Device_Usage_Frequency and USB_Devices_Connected (0.45)**: This positive correlation implies that individuals who frequently use devices are also inclined to connect multiple USB devices, highlighting a potential pattern of multitasking behavior among passengers.
- **Age_Group and Seat_Class (-0.22)**: This negative correlation indicates that age may influence seat class preferences, with older passengers potentially favouring higher classes (e.g., business or first class) over economy seating.
- **Threat_Status and Device_Usage_Frequency (0.25)**: A weak positive correlation suggests that higher device usage frequency may be associated with increased security concerns, indicating that more frequent users might perceive a greater risk of cybersecurity threats.
- **Air_Travel_Frequency and Frequent_Flyer_Status (0.37)**: A moderate positive correlation demonstrates that passengers who travel frequently are more likely to hold frequent flyer status, reinforcing the idea that regular travel is necessary for achieving loyalty rewards.

- **Travel_Purpose and Seat_Class (-0.45)**: This moderate negative correlation suggests that the purpose of travel may influence seat class selection, with business travellers typically opting for higher classes, while leisure travellers are more likely to choose economy seats.

  The analysis yielded significant findings that led to the acceptance of all alternative hypotheses. Specifically, the results demonstrated that machine learning can effectively analyse various attributes of flight data to identify potential cybersecurity threats. The outcomes for each hypothesis were as follows:

- **H1a:** Accepted. This hypothesis indicates that machine learning analysis of flight data can predict cybersecurity threats in aviation through passenger profiling.
- **H2a** to **H14a** were all validated, these hypotheses collectively confirm that factors such as **Gender**, **Age_Group**, **Seat_Class**, **Devices_Used**, **Device_Usage_Frequency**, **Data_Usage_MB**, **USB_Devices_Connected**, **USB_Device_Type**, **Flight_Type**, **Flight_Duration**, **Air_Travel_Frequency**, **Frequent_Flyer_Status**, and **Travel_Purpose** can be effectively analysed through machine learning to identify potential cyberthreats, even in the absence of capable guardianship.

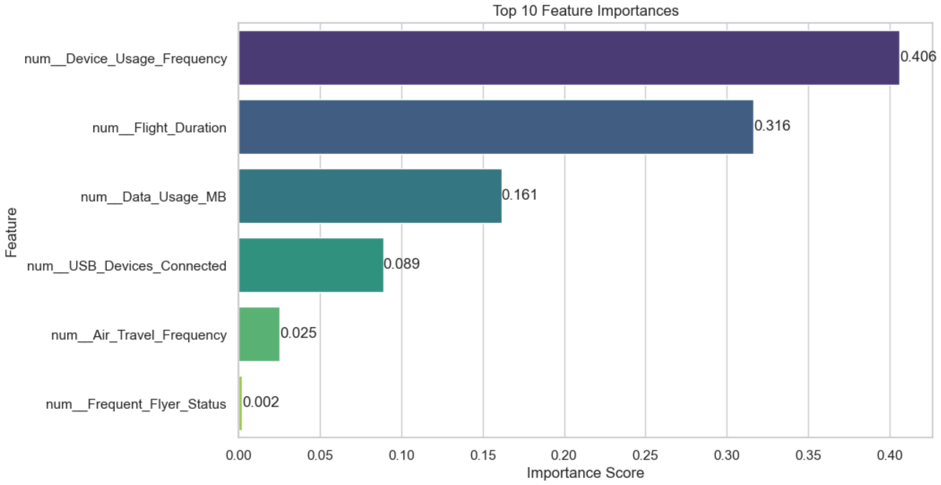


Figure 10: Bar graph showing feature importance

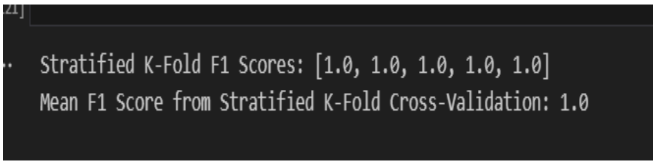
The feature importance scores (Figure 10) highlight the relative contribution of various attributes in predicting cybersecurity threats through passenger profiling. The top features identified in the analysis and their corresponding importance scores are as follows:

- **Device Usage Frequency (0.406)**:This feature emerged as the most significant predictor, indicating that the frequency with which a passenger uses devices is strongly associated with their likelihood of posing a cybersecurity threat. Frequent use of devices may correlate with behaviors that increase exposure to vulnerabilities or risks, making this a critical area for monitoring
- **Flight Duration (0.316)**: Flight duration was the second most crucial factor. Longer flights may present unique challenges or risks, such as prolonged exposure to potential cyber threats. This feature's prominence suggests that assessing flight duration could be crucial in evaluating passenger risk profiles
- **Data Usage (MB) (0.161)**: The amount of data used by passengers is another relevant feature. High data usage may indicate more active online engagement, which could lead to greater susceptibility to cybersecurity incidents. This finding supports the need for monitoring data usage patterns to identify potential threats.
- **USB Devices Connected (0.089)**: This feature reflects the number of USB devices connected by a passenger. The presence of multiple USB devices can introduce additional vulnerabilities, particularly if passengers connect potentially compromised devices to flight systems. As such, monitoring connected devices can enhance security protocols.
- **Air Travel Frequency (0.025)**: While this feature had a lower importance score, it still contributes to understanding passenger behavior. Frequent travelers may exhibit different patterns of risk, and monitoring this feature can aid in developing tailored security measures.
- **Frequent Flyer Status (0.002)**: This feature had the least importance among the top attributes. It suggests that frequent flyer status alone does not significantly correlate with cybersecurity threats. However, this does not diminish its potential relevance; rather, it may indicate that other behavioral factors hold more predictive power.

These findings highlight the robustness of the machine learning models employed, establishing a strong correlation between passenger attributes and the identification of cybersecurity threats in the aviation sector. The perfect accuracy achieved by both the initial and balanced Random Forest models reinforced the assertion that these passenger attributes are significant predictors of cybersecurity threats. The models’ performance

(Figure 5), evaluated through various metrics—including accuracy, precision, recall, and F1 scores—all at **1.0**, demonstrated a robust capacity to generalize findings from the training data to the test set.

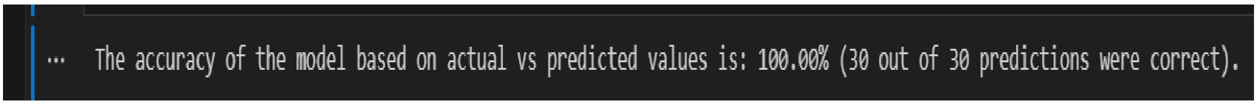
Statistical analyses further indicated high reliability and validity in the model predictions (Figure 11). The consistent performance across stratified K-fold cross-validation underscores the robustness of the machine learning approaches applied, ensuring that the findings are not merely artifacts of a specific dataset but reflect a deeper understanding of the relationships within the data.



```
Stratified K-Fold F1 Scores: [1.0, 1.0, 1.0, 1.0, 1.0]
Mean F1 Score from Stratified K-Fold Cross-Validation: 1.0
```

Figure 11: Report showing stratified k-fold scores

The evaluation of the model's predictions revealed an impressive accuracy of **100%**. This indicates that the model correctly classified all instances in the test dataset. While such a result may suggest excellent performance, it is important to consider the balance of classes in the dataset and the model's ability to generalize to new, unseen data.



```
...  The accuracy of the model based on actual vs predicted values is: 100.00% (30 out of 30 predictions were correct).
```

Figure 12: Report showing the accuracy of the model based on actual vs predicted

This research effectively achieved its objectives by demonstrating the powerful role of machine learning in predicting cybersecurity threats through passenger profiling. The rigorous analytical framework employed throughout the study—encompassing meticulous data preparation, careful model selection, hyperparameter tuning, and comprehensive evaluation—has produced compelling evidence that machine learning techniques can deliver reliable and actionable predictions in the context of aviation security.

The thorough data analysis underscored the significance of key passenger attributes in identifying potential threats, thereby addressing the central research problem. By leveraging robust statistical tests and employing a range of evaluation metrics, the study ensured the validity and reliability of its findings. This research contributes essential insights into the integration of machine learning within aviation security, highlighting its potential not only to enhance safety measures but also to streamline and improve the efficiency of passenger screening processes.

In effectively illustrating the critical role that machine learning can play in bolstering aviation cybersecurity, this study tackles the initial research problem head-on while providing empirical evidence that substantiates the theoretical foundations laid out in existing literature. The findings open the door for further exploration and implementation of these methodologies in real-world aviation security frameworks, fostering a proactive approach to threat identification and management.

Ultimately, this research not only advances our understanding of how machine learning can transform aviation security practices but also lays the groundwork for future innovations in the field. By embracing these insights, aviation authorities can better prepare for emerging cyber threats, significantly enhancing the overall safety and security of air travel for passengers worldwide.

Chapter 5: Conclusion

Scholarly Contributions

This research made a significant scholarly contribution by addressing a notable gap in the literature regarding the application of machine learning in the aviation sector. By integrating theoretical insights, particularly from Routine Activity Theory, with empirical findings, the study enriched the existing discourse on cybersecurity in aviation. It demonstrated the potential of machine learning to refine passenger profiling techniques, paving the way for further academic exploration into how behavioural patterns can inform and enhance cybersecurity strategies.

The research focused on developing predictive models that analysed flight data collected directly from surveys of students at The IIE's Varsity College in South Africa. By relying solely on primary data from this specific population, the study offered fresh insights into the behaviors and device usage patterns of students, which could be leveraged to identify potential cybersecurity threats. The systematic exploration of various passenger attributes—such as ticket class, device type, and internet usage metrics—and their correlation with cybersecurity risks enhanced the understanding of the dynamics involved in aviation security. This contribution proved particularly relevant in an era where data-driven decision-making is essential for improving security measures.

Moreover, the study addressed a critical gap in the literature by illustrating how predictive models could identify vulnerabilities and propose proactive security strategies in the aviation sector. By refining passenger profiling techniques through firsthand data related to passenger behavior and device usage, the research enabled aviation authorities and airlines to gain a better understanding of the behaviors and characteristics associated with potential cybersecurity risks. The practical insights and guidelines derived from the primary data facilitated the implementation of machine learning models tailored to the unique needs of stakeholders, ultimately improving the efficiency and accuracy of threat detection.

The integration of Routine Activity Theory with machine learning methodologies provided a novel perspective on how behavioural patterns could inform cybersecurity strategies, effectively bridging the gap between theoretical frameworks and real-world data. This research not only contributed to the existing body of knowledge but also fostered ongoing discourse in the field of aviation cybersecurity. By aligning empirical findings with theoretical insights, the study facilitated a deeper understanding of the factors influencing cybersecurity threats and underscored the importance of adaptive security measures.

This study has contributed and laid the groundwork for future research to build upon these findings, further investigating the interplay between passenger behavior and cybersecurity risks across diverse demographics and contexts. The study highlighted the value of primary data collection in developing robust predictive models and encouraged further exploration into how data-driven approaches could enhance security practices in aviation and beyond. Ultimately, this ongoing research into the application of machine learning techniques in the aviation sector promised to enrich the dialogue surrounding effective cybersecurity strategies and contribute to fostering a more secure aviation environment.

Implications of findings for future practices

The findings from this study carry significant implications for future practices within both the aviation and cybersecurity sectors. The successful application of machine learning models highlights the potential to enhance security protocols through data analytics. Stakeholders in the aviation industry should consider integrating these predictive models into their existing security frameworks to improve the identification and mitigation of cybersecurity threats. This integration can not only enhance threat detection but also streamline security operations, allowing personnel to concentrate on high-risk scenarios.

The insights gained about passenger behavior are invaluable for developing customized security measures tailored to diverse traveller profiles. Implementing adaptive security strategies that take individual differences into account can enhance the passenger experience while upholding stringent security standards. Additionally, organizations should invest in ongoing staff training to ensure personnel are equipped to interpret analytical outputs effectively and make informed decisions based on data-driven insights.

This study also emphasizes the necessity for continuous monitoring and updating of predictive models to keep pace with the rapidly evolving nature of cybersecurity threats. Future practices should include regular assessments of model performance and recalibrations based on new data to ensure that security measures remain effective and relevant. The focus on primary data collection underscores the importance of context-specific information in addressing complex security issues, reinforcing the need for continuous adaptation in response to emerging threats.

The primary aim of this research was to explore the application of machine learning techniques in predicting cybersecurity threats within the aviation sector, specifically through the analysis of passenger behavior. The empirical findings derived from data collected directly from students at The IIE’s Varsity College revealed significant correlations between various passenger attributes—such as device type, ticket class, and internet usage metrics—and the likelihood of potential cybersecurity threats. These results indicated that specific behavioural patterns could effectively inform targeted security measures, thereby enhancing threat detection capabilities in aviation security frameworks.

Furthermore, the systematic examination of passenger profiling techniques using primary data provided fresh insights into leveraging behaviors and device usage patterns to identify cybersecurity vulnerabilities. By bridging the gap between theoretical frameworks, such as Routine Activity Theory, and real-world data, this study demonstrated the practical application of machine learning methodologies in aviation security.

Despite its contributions, the study faced certain limitations, particularly concerning the representativeness of the sample and potential biases in model outcomes. The reliance on data collected solely from students at a private higher education institution may restrict the generalizability of the findings to broader populations. Future research should aim to expand the dataset to capture a more diverse spectrum of passenger demographics and behaviors, thereby enhancing the robustness and applicability of the predictive models developed.

Conclusion

This research successfully explored the application of machine learning techniques to predict cybersecurity threats in the aviation sector, yielding critical insights into the interplay between passenger behavior and security risks. The analysis of primary data collected from students at The IIE's Varsity College revealed significant correlations between various passenger attributes—such as device type, ticket class, and internet usage metrics—and the likelihood of potential cybersecurity threats. These findings illuminate how specific behavioural patterns can inform targeted security measures, thereby enhancing threat detection capabilities within the aviation industry. The predictive models developed in this study provide a promising foundation for integrating machine learning into aviation security frameworks. By refining passenger profiling techniques and addressing a notable gap in the literature on the application of machine learning in aviation, this research contributes not only to theoretical understanding but also offers practical insights for stakeholders. However, the study also identified limitations concerning the representativeness of the sample and potential biases in model outcomes. These challenges underscore the need for further research to expand the dataset and capture a broader spectrum of passenger demographics and behaviors. Such efforts would enhance the robustness and generalizability of the predictive models, allowing for more comprehensive threat detection capabilities. The implications of this research extend beyond academic contributions; they signal a shift towards data-driven security practices in aviation. The study emphasizes the necessity for continuous monitoring and refinement of predictive models to adapt to the evolving landscape of cybersecurity threats. Stakeholders in the aviation sector should prioritize the implementation of adaptive security measures informed by ongoing research, ensuring they remain proactive in the face of emerging risks.

This research has significantly advanced the understanding of cybersecurity in aviation, providing actionable insights and a methodological framework for future studies. As the cybersecurity landscape continues to evolve, ongoing investigations should focus on validating and building upon these findings, exploring the interplay between passenger behavior and cybersecurity across diverse demographics and contexts. The integration of Routine Activity Theory with machine learning methodologies offers rich avenues for deeper exploration into behavioural patterns and their implications for cybersecurity strategies. Future research should refine predictive models, explore additional variables, and assess the practical effectiveness of implemented security strategies.

By engaging with the dynamic landscape of cybersecurity, researchers can enhance the development of effective practices that safeguard the integrity of passenger travel and ensure a more secure aviation environment. Through these ongoing efforts, we can foster an aviation sector that not only effectively mitigates risks but also adapts to the changing nature of cybersecurity challenges.

Bibliography

1. Anon. n.d. Anomaly Detection. AVI Networks. [Online] Available at: https://avinetworks.com/glossary/anomaly-detection/#:~:text=Anomaly%20detection%20is%20the%20identification,noise%2C%20novelties%2C%20and%20exceptions. [Accessed 09 April 2024]
2. Anon. n.d. Definitions of Aviation. Vocabulary.com. [Online] Available at: https://www.vocabulary.com/dictionary/aviation#:~:text=The%20aviation%20industry%20is%20the,synonyms%3A%20airmanship [Accessed 09 April 2024]
3. Anon. n.d. What is a cyberattack? IBM. [Online] Available at: https://www.ibm.com/topics/cyber-attack [Accessed 09 April 2024]
4. Anon. n.d. Cybercriminals. Trendmicro. [Online] Available at: https://www.trendmicro.com/vinfo/us/security/definition/cybercriminals [Accessed 09 April 2024]
5. Anon. n.d. What is Cyber Security? Definition and Best Practices. IT Governance. [Online] Available at: https://www.itgovernance.co.uk/what-is-cybersecurity#:~:text=Cyber%20security%20is%20the%20application,systems%2C%20networks%2C%20and%20technologies. [Accessed 09 April 2024]
6. Anon. n.d. What is IoT? Oracle. [Online] Available at: https://www.oracle.com/za/internet-of-things/what-is-iot/#:~:text=The%20Internet%20of%20Things%20(IoT)%20describes%20the%20network%20of%20physical,and%20systems%20over%20the%20internet. [Accessed 09 April 2024]
7. Anon. n.d. Predictive Modeling overview. OutSystems. [Online] Available at: https://www.outsystems.com/tech-hub/ai-ml/what-is-predictive-modeling/#definition [Accessed 09 April 2024]
8. Anon. n.d. What is a threat actor? IBM. [Online] Available at: https://www.ibm.com/topics/threat-actor#:~:text=IBM-,What%20is%20a%20threat%20actor%3F,to%20digital%20devices%20or%20systems . [Accessed 09 April 2024]
9. Anon. n.d. Ransomware. TrendMicro. [Online] Available at: https://www.trendmicro.com/vinfo/us/security/definition/ransomware [Accessed 09 April 2024]
10. Anon. n.d. Case Study: Cyberattacks in the Aviation Industry. TechForce. [Online] Available at: https://techforce.co.uk/blog/2023/case-study-cyberattacks-in-the-aviation-industry---risksandremedies#:~:text=Impact%20and%20Future%20Outlook%3A%20The,and%20damage%20to%20brand%20reputation. [Accessed 27 April 2024]
11. Anon. 1998. Air Traffic Control: Weak Computer Security Practices Jeopardize Flight Safety. Office of Justice Programs. [Online] Available at: https://www.ojp.gov/ncjrs/virtual-library/abstracts/air-traffic-control-weak-computer-security-practices-jeopardize [Accessed 27 April 2024]
12. Anon. 2019. Protecting aviation from hackers: Aircraft and airports are increasingly vulnerable to cyber-attacks, writes Morand Fachot. Medium. [Online] Available at: https://medium.com/e-tech/protecting-aviation-from-hackers-9d361a4f1bce [Accessed 06 June 2024]

13. Anon. 2021. Risk Insights. Jamaican Civil Aviation Authority. [Online] Available at: https://www.icao.int/Security/Security-Culture/State%20and%20Industry%20Promotional%20Material/Risk%20Insights%20%20The%20Jamaica%20Civil%20Aviation%20Authority's%20Risk%20Management%20Newsletter,%20vol.%204%20(July%202021).pdf [Accessed 27 April 2024]
14. Anon. 2023. Cyberattacks Are on The Up: What Are the Risks & Remedies for Aviation? Resilinc. [Online] Available at: https://www.resilinc.com/in-the-news/cyberattacks-are-up-the-risks-remedies-for-aviation/ [Accessed 09 April 2024]
15. Anon. 2023. Aviation Cybersecurity. IATA. [Online] Available at: https://www.iata.org/en/iata-repository/pressroom/fact-sheets/fact-sheet--cyber-security/ [Accessed 27 April 2024]
16. Anon. 2024. The Aviation and Aerospace Sectors Face Skyrocketing Cyber Threats: Cyber Threat Intelligence. Resecurity. [Online] Available At: https://www.resecurity.com/blog/article/the-aviation-and-aerospace-sectors-face-skyrocketingcyberthreats#:~:text=By%20conducting%20regular%20assessments%2C%20airports,and%20safeguarding%20employees%20and%20passengers. [Accessed 09 April 2024]
17. Alqushayri, D. 2020. Cybersecurity Vulnerability Analysis and Countermeasures of Commercial Aircraft Avionic Systems. Scholarly Commons. [Online] Available At: Https://Commons.Erau.Edu/Edt/519 [Accessed 09 April 2024]
18. Aljohani, A. 2023. Predictive Analytics and Machine Learning for Real-Time Supply Chain Risk Mitigation and Agility. ResearchGate. [Online] Available at: https://www.researchgate.net/publication/374849416_Predictive_Analytics_and_Machine_Learning_for_Real-Time_Supply_Chain_Risk_Mitigation_and_Agility [Accessed 01 November]
19. Andresen, G and Farell, G. 2015. The Criminal Act: The Role and Influence of Routine Activity Theory. Google Books. [Online] Available at: https://books.google.co.za/books?id=0vS3BgAAQBAJ&pg=PT38&lpg=PT38&dq=crime+equation,+crime+%3D+(offender+%2B+target+%E2%88%92+guardian)+(place+%2B+time)&source=bl&ots=ZIRzI-DKxM&sig=ACfU3U12e1ReLrOgnBmnz3JRWa6-zb7arQ&hl=en&sa=X&ved=2ahUKEwjrucGQk8CGAxVD3QIHHfVkGZE4ChDoAXoECAMQAw#v=onepage&q&f=false [Accessed 28 April 2024]
20. Benova, L and Hudec, L. 2024. Comprehensive Analysis and Evaluation of Anomalous User Activity in Web Server Logs. *Sensors,* 24(3):746. [Online] Available at: https://doi.org/10.3390/s24030746 [Accessed 28 April 2024]
21. Bhandari, P. 2023. Construct Validity | Definition, Types, & Examples. Scribbr. [Online] Available at: https://www.scribbr.com/methodology/construct-validity/ [Accessed 26 April 2024]
22. Bock, K., Cukier, M., Shannon, S., Movahedi, Y. 2017. Application of Routine Activity Theory to Cyber Intrusion Location and Time. *European Dependable Computing Conference, 139-146.* [Online] Available at: https://www.jstor.org/stable/2094589 [Accessed 26 April 2024]
23. Cavusoglu, H., Kwark, Y., Mai, B., Raghunathan, S. 2013. Passenger Profiling and Screening for Aviation Security in the Presence of Strategic Attackers. Decision Analysis. ResearchGate. [Online] Available at: https://www.researchgate.net/publication/262160720_Passenger_Profiling_and_Screening_for_Aviation_Security_in_the_Presence_of_Strategic_Attackers#:~:text=Developi

ng%20effective%20inspection%20processes%20at,ongoing%20debates%20about%20its%20usefulness. [Accessed 09 April 2024]

24. Cohen, L and Felson, M. 1979. Social Change and Crime Rate Trends: A Routine Activity Approach. *American Sociological Association*, 44(4): 588-608. [Online] DOI:10.1109/EDCC.2017.24 [Accessed 26 April 2024]
25. Chen, C., Yuantian, M., Lei, P., Han, Q., Yang, X., Zhang, J. 2021. Machine Learning–based Cyber Attacks Targeting on Controlled Information: A Survey. *ACM Computing Surveys.* 54. 1-36. [Online] Available at: DOI:10.1145/3465171 [Accessed 27 April 2024]
26. Department of Homeland security. 2021. Secure Cyberspace and Critical Infrastructure. Homeland Security. [Online] Available at: https://www.dhs.gov/secure-cyberspace-and-critical-infrastructure [Accessed 27 April 2024]
27. Dutta, S., Sinha, A and Basu, D. 2021. AI-Enabled Solutions in Indian Agriculture: The Way Forward. ResearchGate [Online] Available at: https://www.researchgate.net/figure/Subset-of-Artificial-Intelligence-Dutta-et-al-2020_fig1_348565838 [Accessed 01 November 2024]
28. European Aviation Safety Agency. 2019. Guidance for Level 1 & 2 machine learning applications. EASA. [Online] Available at: https://www.easa.europa.eu/en/downloads/139504/en [Accessed 27 April 2024]
29. Gabrijelčič, D., Kaur, R., Klobučar, T. 2023. Artificial intelligence for cybersecurity: Literature review and future research directions. *Information Fusion,* 97. [Online] Available at: https://doi.org/10.1016/j.inffus.2023.101804 [Accessed 27 April 2024]
30. Gillis, A. n.d. What is augmented reality (AR)? Tech Target. [Online] Available at: https://www.techtarget.com/whatis/definition/augmented-reality-AR/ [Accessed 09 April 2024]
31. Grossman, L. 2021. Before The United States House of Representatives Committee on Transportation and Infrastructure: The Evolving Cybersecurity Landscape: Federal Perspectives on Securing the Nation's Infrastructure. Federal Aviation Administration. [Online] Available at: https://www.faa.gov/testimony/united-states-house-representatives-committee-transportation-and-infrastructure-evolving [Accessed 27 April 2024]
32. Ham, J and Macnish, K. 2020. Ethics in cybersecurity research and practice. ScienceDirect. [Online] Available at: https://www.sciencedirect.com/science/article/pii/S0160791X19306840 [Accessed 01 November 2024]
33. Hu, Z., Yang, I., Salakhutdinov, R and Xing, P. 2017. On Unifying Deep Generative Models. Cornell University. [Online] Available at: https://arxiv.org/abs/1706.00550 [Accessed 01 November 2024]
34. International Civil Aviation Organization. 2020. Aviation cybersecurity. International Civil Aviation Organization. [Online] Available at: https://www.icao.int/aviationcybersecurity/Pages/default.aspx [Accessed 27 April 2024]
35. Kanade, V. 2022. What Is Machine Learning? Definition, Types, Applications, and Trends for 2022. Spiceworks. [Online] Available at: https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-ml/ [Accessed 09 April 2024]
36. Kabashkin I, Misnevs B, Zervina O. 2023. Artificial Intelligence in Aviation: New Professionals for New Technologies. *Applied Sciences*, 13(21):11660. [Online] Available at: https://doi.org/10.3390/app132111660 [Accessed 28 April 2024]

37. Kigerl, A. 2012. Routine Activity Theory and the Determinants of High Cybercrime Countries. Social Science Computer Review. *Sage Journals*. [Online] https://doi.org/10.1177/0894439311422689 [Accessed 26 April 2024]
38. Kost, E. 2021. What is a Cyber Threat? UpGuard. [Online] Available at: https://www.upguard.com/glossary/cyber-threat [Accessed 09 April 2024]
39. Miro, F. 2014. Routine activity theory. *The encyclopaedia of theoretical criminology*. [Online] https://doi.org/10.1002/9781118517390.wbetc198. [Accessed 26 April 2024]
40. Paper, W. 2018. AI in aviation: Exploring the fundamentals, threats, and opportunities of Artificial intelligence (ai) in the aviation industry. IATA. [Online] Available at: https://www.iata.org/contentassets/2d997082f3c84c7cba001f506edd2c2e/ai-white-paper.pdf [Accessed 27 April 2024]
41. Reed, J. 2023. Increasing Insider Cyber Threats Pose Risks to Aviation. Avionics International. [Online] Available at: https://www.aviationtoday.com/2023/06/14/increasing-insider-cyber-threats-pose-risks-to-aviation/ [Accessed 09 April 2024]
42. Rossmo, D and Summers, L. 2015. Routine Activity Theory in Crime Investigation. Springer Link. [Online] Available at: https://link.springer.com/chapter/10.1057/9781137391322_3#citeas [Accessed 26 April 2024]
43. Sadeghi, Manijeh. (2019). A Shift from Classroom to Distance Learning: Advantages and Limitations. *International Journal of Research in English Education*. 4. 80-88. [Online] Available at: 10.29252/ijree.4.1.80 [Accessed 01 November 2024]
44. Shah, V. 2022. Machine Learning Algorithms for Cybersecurity: Detecting and Preventing Threats. ResearchGate. [Online] Available at: https://www.researchgate.net/publication/378396020_Machine_Learning_Algorithms_for_Cybersecurity_Detecting_and_Preventing_Threats [Accessed 01 November 2024]
45. Statista. 2024. Average cost of a data breach worldwide from May 2020 to March 2023. Statista. [Online] Available at: https://www.statista.com/statistics/387861/cost-data-breach-by-industry/ [Accessed 27 April 2024]
46. Taherdoost, H. 2022. Data Collection Methods and Tools for Research; A Step-by-Step Guide to Choose Data Collection Technique for Academic and Business Research Projects. International Journal of Academic Research in Management (IJARM), 2021, 10 (1), pp.10-38. [Online] Available at: https://hal.science/hal-03741847/document [Accessed 01 November 2024]
47. Thomas, L. 2023. Stratified Sampling | Definition, Guide & Examples. Scribbr. [Online] Available at: https://www.scribbr.com/methodology/stratified-sampling/ [Accessed 01 November 2024]
48. Trottenberg, P. 2020. National Aviation Research Plan (NARP). Federal Aviation Administration. [Online] Available at: https://www.statista.com/statistics/387861/cost-data-breach-by-industry/ [Accessed 28 April 2024]
49. Ukwandu E, Ben-Farah MA, Hindy H, Bures M, Atkinson R, Tachtatzis C, Andonovic I, Bellekens X. Cyber-Security Challenges in Aviation Industry: A Review of Current and Future Trends. *Information*. 2022; 13(3):146. https://doi.org/10.3390/info13030146 [Accessed 01 November 2024]
50. Zetter, K. 2015. All Airlines Have the Security Hole That Grounded Polish Planes. Wired. [Online] Available at: https://www.wired.com/2015/06/airlines-security-hole-grounded-polish-planes/ [Accessed 27 April 2024]