Empirical Economics Letters

Conference Proceedings: September 2025

https://doi.org/10.5281/zenodo.17117289

International Business Analytics Conference for Academic and Industry Professionals 2025

Managing Editor

Mohammad A Wadud

Guest Editors

Kaustav Misra
Justin Mindzak

Empirical Economics Letters

ISSN 1681-8997

A Monthly International Journal of Economics

Managing Editor

Mohammad A. Wadud

University of Rajshahi, Bangladesh

Board of Editors

Aman Agarwal, *GGS Indraprastha University, India*
Clem A. Tisdell, *University of Queensland, Australia*
Damodar N. Gujarati, *US Military Academy, U.S.A.*
George E. Battese, *University of New England, Australia*
William F. Bowlin, *University of Northern Iowa, U.S.A.*

Publication Information

Empirical Economics Letters (ISSN 1681 - 8997) is published twelve times a year.

Publisher

Dr. Mohammad A. Wadud, Managing Editor, *Empirical Economics Letters*, Rajshahi University, Rajshahi 6205, Bangladesh

Publication Ethics and Malpractice Policy

Empirical Economics Letters *(EEL)* maintains the standard ethical and malpractice policy. Authors would ensure that the manuscript has not been published before or submitted simultaneously to any other journals and does not violate the existing copyright. If the manuscript is accepted, the copyright is automatically transferred to the publisher.

Plagiarism and any unethical behavior are prohibited. Referees are requested to inform the *EEL* office about any misconduct like plagiarism, copyright violation or any other unethical activities. Referees are also requested to maintain confidentiality regarding review process. *EEL* makes every effort to deliver publication complying with ethical and malpractice standards. For details, please visit the journal website at **www.eel.my100megs.com**

Indexing Information

This journal is listed in the American Economic Association (AEA) electronic indexes, which are included in *JEL* on CD, *e-JEL,* and *EconLit*. It is also included in *Cabell's Directory* of Publishing Opportunities in Economics and Finance. Further, this journal is listed in *ABDC journal quality* ranking and accredited in *Excellence in Research for Australia (ERA)*. Moreover, it is endorsed in Publons which is a part of Web of Science Group.

Subscription

Yearly subscription for twelve issues is US$ 400.00 or its equivalent for institutions and US$ 300.00 for individuals for rich countries (Per capita income greater than US$ 2000) and US$ 300.00 or its equivalent for institutions and US$ 200.00 for individuals for other countries. All issues of the journal are sent to the subscribers by airmail.

Subscriptions are to be paid through **Wire Transfer** (Bank Transfer) in favour of ***Empirical Economics Letters*** and account number to which subscriptions should be sent is available on request.

Copyright: The Publisher. Visit our website at http://www.eel.my100megs.com

FREDONIA
STATE UNIVERSITY OF NEW YORK

International Business Analytics Conference for Academic and Industry Professionals 2025

Fredonia, New York

Thursday – Friday, May 8 – 9

Conference Website: www.fredonia.edu/ibac
Conference Email: ibac@fredonia.edu

FREDONIA INTERNATIONAL BUSINESS ANALYTICS CONFERENCE
IBAC

Conference at a Glance

The International Business Analytics Conference (IBAC) successfully hosted its inaugural event at SUNY Fredonia on May 3-4, 2024. This landmark conference established a unique platform designed to bridge the gap between academic scholarship and industry practice in the field of business and data analytics through meaningful Academia-Industry partnerships. By convening industry professionals and academic experts, IBAC fostered valuable dialogue and collaboration within the dynamic and rapidly evolving discipline of business analytics.

The conference welcomed submissions across a broad spectrum of topics related to both research and industry applications in business and data analytics. Areas of focus included, but were not limited to, Accounting, Business and Management, Computer Science/Management Information Systems, Economics, Finance and Insurance, Education, Healthcare, Marketing, Music and Entertainment, and Sport Management. In its first year, IBAC attracted 189 participants, demonstrating strong interest and confirming the demand for a forum of this kind. Encouraged by this positive response, the conference is set to become an annual event.

The 2025 IBAC conference is scheduled for May 8-9, with plans to further expand participation and engagement. We are pleased to witness growing support from local and regional businesses, whose involvement continues to strengthen the conference's impact and reach. Additionally, IBAC 2025 will introduce two new and highly anticipated components: Undergraduate Poster Sessions and an Analytics Career Fair, both of which have generated considerable enthusiasm from students and industry partners alike.

For example, submission specifically addressed research or current business practices in:

- Data-driven decision-making
- Predictive and prescriptive analytics
- Artificial Intelligence and Machine Learning in business
- Big data analytics and its applications
- Business Intelligence and data visualization
- Data mining and pattern recognition
- Text and sentiment analysis in business
- Supply chain analytics and logistics optimization
- Marketing and customer analytics
- Risk management and fraud detection
- Ethics and privacy in business analysis

Proceedings and Publication Opportunity: We are pleased to announce that this year, IBAC is collaborating with *Empirical Economics Letters* (indexed by the Australian Business Deans Council, Rating: C, http://www.eel.my100megs.com/) to publish all conference proceedings. In addition, authors will have the opportunity to submit their articles for publication in this journal under a special issue edited by Drs. Kaustav Misra and Justin Mindzek.

 ISSN 1681 8997

Conference Executive Committee

Organizing Committee

Conference Chair
Kaustav Misra (Dean, School of Business)

Conference Vice Chair
Justin Mindzak (School of Business - Accounting)

Program Chairs
Syed Haider (Computer and Information Sciences)
Adam Cook (School of Business - Economics)

Undergraduate Program Chairs
Lisa Walters (School of Business - Management)
Megan Johnson (Mathematical Sciences)

Career Fair Chair
Christopher LaGrow (Career Development)

Administrative Support
Mindy Ostrander and Kristie Bobik
(School of Business)

Program Committee

Publicity
Charles Cornell (CIED)
Fredonia Marketing and Communication Team

Finance Chair
Neepa Gaekwad Babulal (School of Business - Economics)

Event Management
Mark Delcamp (Facilities Services), Jeff Walter and Katie Thies (FSA), Mark Suida (Campus Life), John McCune (Information Technology), Kathy Forster (Residence Life), Mark Mackey (ITS),

Advisory Member
Mohammad Nasim

Conference Proceedings

Track 1: Accounting and Election Analytics

Room: WC S204A (9:30 am - 10:50 am)
Moderator: Abdelghani Mehailia *(Yorkville University)*

1.1 Bridging the Skills Gap: Integrating Data Analytics and Artificial Intelligence into Accounting Education

Authors: Nazik Roufaiel *(SUNY-ESC Center for Distance Learning)*
Magdy Roufaiel *(SUNY-ESC-CDL-MBA)*

As the role of data analytics in business continues to expand, the accounting profession must adapt to meet the evolving demands of both industry and regulatory bodies. This paper explores the crucial role of data analytics in accounting, focusing on its application in business decision-making and its increasing importance in the education and licensure of Certified Public Accountants (CPAs). Despite the industry's growing reliance on data-driven insights, a significant gap remains between the skills demanded by employers and those possessed by recent accounting graduates.

This presentation will examine the benefits of integrating data analytics into accounting curricula, highlighting the competencies required for modern CPAs to thrive in the industry. The author will review current trends in business analytics, providing insights into how these tools are applied in practice, from predictive financial modeling to real-time fraud detection. Additionally, statistics will illustrate the widening divide between academic preparation and industry expectations, highlighting data from recent employer surveys. For instance, a recent report revealed that over 70% of accounting firms prioritize data analytics skills, yet less than 40% of business and accounting graduates feel proficient in this area.

The session will also offer recommendations for aligning academic programs with industry needs, ensuring that students and future professionals are equipped with the analytical skills required for licensure and beyond. Attendees will leave with actionable strategies to bridge this skills gap, ultimately enhancing both education and professional practice.

1.2 Detecting Anomalous Trading Patterns in the Cryptocurrency Market Using Graph Neural Network

Author: So-Jin Yu *(SUNY Fredonia)*

This study examines how to detect anomalous trading patterns in the cryptocurrency market using Graph Neural Network (GNN), an emerging technique compared to

traditional machine learning. One such anomaly, wash trading, frequently takes place and damages investors' trust. Further, the complex structure of block-chain based transactions makes it difficult for market participants to detect manipulative behaviors, especially on Ethereum's on-chain data. This study provides practical implications to better protect investors from manipulation, and to improve the credibility of the cryptocurrency market.

1.3 Filling in the Blanks - Estimating Vote Choice of Undecided Voters
Travis Brodbeck *(SUNY Albany, USA & Siena College)*

In pre-election surveys, pollsters are challenged by respondents who refuse to answer a question or say they ""don't know"". Rather than discarding this response, pollsters typically include this respondent and report out on findings despite their unknown potential voting behavior. Poll results are generally reported in this fashion, for example with 46% supporting candidate A, 48% supporting candidate B, and 6% of respondents saying they don't know or have not made up their mind yet. On election day when actual results come in, these voters either make up their mind and skip the ballot box altogether and then pollsters are evaluated against their polls composed of some number of undecided or private voters. This challenge impacts how the media frames the story, how the public interprets polls, and campaign decision making. Using other questions besides the generic horserace ballot question, can pollsters provide additional insights and more nuanced snapshots of the race when surveys contain undecided voters?

In an analysis of interviews conducted by the New York Times/Siena College poll during the 2024 pre-election season, this paper provides insights on creating vote choice estimates for undecided voters. The data demonstrates that a combination of demographic and issue questions were most effective in creating estimates for undecided voters that more accurately reflect actual election results compared to the originally published survey results. Findings from this research can be applied retroactively to older survey data and future election cycles to enhance public understanding of the electorate.

Track 2: Industry Insights

Room: WC S204B (9:30 am - 10:50 am)
Moderator: Shazad Mohammed *(State University of New York at Fredonia)*

2.1 An Empirical Analysis of the Nexus Among Inclusive Leadership, Team Members' Psychological Factors and Project Performance
Author: Muhammad Yousuf Khan *(Pakistan Agricultural Research Council)*

Leadership role in project success is irrefutable yet the most appropriate leadership style is open to debate, which triggers the need for examining different leadership styles that effectively lead the team and attain greater project performance (PP). Relevant literature prescribes that the influence of inclusive leadership (IL) and project team members'

psychological factors should be reviewed in combination to provide a much clearer understanding of the predictors of PP. This research aims to examine the effect of IL on PP together with parallel mediational roles of psychological empowerment (PE) and work meaningfulness (WM) between these relationships. Using structural equation modeling (SEM) analysis technique, research hypotheses were tested with a sample of 249 respondents working in different social sector projects in Pakistan. The study's results suggest that IL, besides having a positive direct effect on PP also positively impacts project team members' PE and WM. In addition, perceived PE and WM significantly mediate the link between IL and PP. The results broaden our understanding of underlying direct and indirect paths and detailed mechanisms from IL and team members' psychological factors to PP. Implications, limitations, and avenues for future research are provided.

2.2 Transforming Project Management Through Data Analytics and Generative AI

Authors: Srikanth Srinivasa *(Northwood University & Harvard Business School)* **Itauma Itauma** *(Northwood University)*

The integration of Artificial Intelligence (AI) and Machine Learning (ML) is reshaping the landscape of project management, offering transformative opportunities to enhance efficiency, decision-making, and outcomes. This presentation examines the intersection of data analytics and AI in project management practices, focusing on how generative AI can revolutionize project planning, resource allocation, stakeholder communication, and performance measurement. By leveraging cutting-edge AI technologies, project managers can automate repetitive tasks, mitigate risks, and optimize workflows to drive success.

Through a combination of academic research and real-world applications, this session highlights practical strategies for adopting AI in project management. Case studies will demonstrate the tangible benefits of AI-driven tools in improving key performance indicators such as error reduction, time savings, cost optimization, and risk mitigation. The importance of effective prompt engineering for maximizing generative AI's potential will be emphasized, along with step-by-step guidance to craft actionable prompts.

The presentation also addresses ethical considerations and challenges associated with AI, focusing on ensuring transparency, fairness, and reducing bias in data-driven decision-making. Attendees will gain insights into state-of-the-art AI tools for real-time tracking, risk analysis, and performance reporting, enabling them to harness data analytics to transform project management practices.

Key Takeaways: - Understanding AI Tools: Gain familiarity with AI technologies that assist project managers in real-time tracking, risk analysis, and performance reporting.

Practical Techniques: Learn effective prompt engineering strategies to maximize the benefits of generative AI. Real-World Use Cases: Discover specific scenarios where generative AI optimizes project processes and performance measurement. Ethical AI Practices: Understand the importance of ethical considerations in leveraging AI for business and project management.

This session is designed for a global audience of professionals, academicians, and students, bridging research and practice to showcase actionable knowledge and demonstrable examples of AI's profound impact on project management and business analytics.

2.3 Transforming Business Data into Stunning Visualizations

Author: Lisa Jo Romas-Elliott *(Pennsylvania State University)*

Transforming business data into compelling visual representations is a significant aspect of data analytics. A primary challenge lies in taking crucial business data, extracting its essential information, and conveying it to stakeholders in a significant manner. While tools such as Excel, Tableau, and Alteryx may appear to simplify this process, the reality is more complex. There are numerous steps involved in converting raw data into meaningful graphics that may not be apparent to the casual observer. This presentation will cover the processes of data cleaning for visualization, converting data into a suitable visualization format, and choosing the appropriate visualization tool, alongside essential visualization heuristics.

Track 3: Sports Analytics

Room: WC S204C (9:30 am - 10:50 am)
Moderator: Soumik Banerjee *(Canisius University)*

3.1 Enhancing College Football Playoff Rankings: a Stochastic Multicriteria Acceptability Analysis Approach

Author: David Mahalak *(University of Scranton)*

The National Collegiate Athletic Association (NCAA) College Football Playoff (CFP) series is one of the most anticipated events of the year. The revenue that is generated from these games is staggering, with participating teams and conferences making millions of dollars.

The current CFP selection process involves a 22-member selection committee that ranks teams based on qualitative and quantitative criteria, such as strength of schedule, head-to-head competition, and comparative outcomes against common opponents.

With so many financial implications, it is critical that the CFP selection process maintain fairness and transparency in its evaluations of team rankings to protect the overall integrity

of college football's championship. However, the selection process is often criticized for its biases, vagueness, and questionable decisions. For instance, many analysts and fans did not feel that Southern Methodist University desired a playoff spot in the 2025 CFP bracket, over other teams that played a more competitive schedule, e.g. the University of Alabama or Miami.

To address the inherent flaws of the current CFP selection process, this study will employ Stochastic Multicriteria Acceptability Analysis (SMAA) to enhance transparency and fairness of the CFP ranking process. Using this novel methodology, we will be able to combine qualitative and quantitative metrics and consider multiple weighting scenarios to improve the NCAA playoff selection process. The implementation of the SMAA-based ranking system will enhance competitive fairness and overall product integrity. Furthermore, this research will contribute to the field of sports analytics and decision science.

3.2 Unveiling Esports Trends 2010 - 2024

Authors: Reneta Barneva *(SUNY Fredonia)*
Lisa M Walters *(The State University of New York Fredonia)*

Esports is an important sector within the broader sport industry, characterized by competitive video gaming at a professional level, either individually or in teams and is often watched online or in person. It is projected to bring about 4 billion in revenue in 2024, mostly from sponsorships, media rights, merchandise, and ticket sales.

This work explores the dynamic nature of esports, highlighting its swift evolution driven by technological advancements, consumer preferences, and social media influence. More specifically, we explore the dataset ""Gaming Trends 2024,"" which delves into the key trends shaping the current gaming landscape, since esports are directly connected to the video gaming industry. By analyzing data on preferred platforms, active user numbers, new registrations, average session durations, in-game purchases, social media mentions, stream viewership, revenue, and top genres, we provide a comprehensive overview of the state of gaming from 2010 to 2024.

We also use predictive modeling to forecast future gaming trends, providing substantial insight for developers, marketers, and stakeholders. Our findings seek to help these professionals make better decisions in informed ways in order to capitalize on emerging trends and opportunities within the esports and gaming industries.

Track 4: Emerging Scholars I: AI & Natural Language Applications in Business & Cybersecurity

Room: WC S204D (9:30 am - 10:50 am)
Moderator: Megan Johnson ***(State University of New York at Fredonia)***

4.1 Digital Human Avatars with Agentic AI Integration: Evaluating Effectiveness in Hybrid Business Communication Environments

Authors: Mohammad Nasim *(Northwood University & State of Rhode Island)* **Shanshan Zhu** *(Mount Holyoke College)*

Abstract— Hybrid environments create substantial obstacles for modern business communication specifically in retaining information, maintaining engagement quality, and ensuring workflow continuity. This research presents a new method that combines AI-generated digital human avatars with automated workflow systems that utilize Agentic AI technology to solve identified business communication challenges. The research establishes and tests an all-encompassing system that enables virtual business representatives to communicate within structured settings with human users while intelligent AI mechanisms record and transform their dialogue results into action plans. Using voice and facial cloning technology from HeyGen this custom system builds lifelike digital models of business staff which work together with Zoom videoconferencing functionalities and natural language processing tools for conversation assessment along with n8n/Zapier-based workflow automation to manage multi-step data processing. The research focuses on a scenario where automated digital avatars representing sales managers lead formal interviews with team members and handle documentation and follow-up activities automatically. The study evaluates three key dimensions: The research evaluates engagement quality and conversational naturalness between humans and digital avatars as well as information retention and process continuity compared to standard meetings and ethical factors including privacy concerns and algorithmic bias. The performance evaluation features quantitative metrics for interaction quality and information accuracy assessments as well as process completion rates together with qualitative participant feedback. Initial findings show that this approach produces major advantages through uniform communication methods and enhanced memory retention along with a reduced burden for management within business communication systems. The research advances our understanding of human-AI partnerships in business settings by offering practical frameworks which help organizations improve digital avatar-based communication through agentic AI systems.

Index Terms: digital human avatars, agentic AI, business communication, information retention, workflow automation

Introduction

Business environments that change quickly today face essential challenges in maintaining effective communication and information retention which become particularly difficult in

hybrid and remote work settings. Business operations now benefit from artificial intelligence integration which provides innovative solutions for communication and data management challenges. This study investigates how AI-driven digital human replicas with autonomous AI functions can improve business communication through enhanced interactions in hybrid meeting environments. We explore how digital avatars enhanced with AI technology boost efficiency and engagement while enhancing productivity during business discussions with human participants. We establish a controlled setting to enable an AI-generated digital clone, which represents a senior manager, to interact with human team members so we can study hybrid interaction dynamics and assess their effects on engagement levels, information retention and conversation quality. The system integrates cutting-edge facial and voice cloning tools along with Natural Language Processing features through autonomous agentic AI systems which handle meeting operations including conversation management and post-meeting follow-ups. Our comprehensive methodology addresses every phase of meetings to integrate human interactions and AI communication into a unified experience. This study explores the ethical challenges of using digital human clones at work which includes evaluating privacy concerns, consent requirements, potential biases and questions about authenticity. The results from this study provide organizations with essential knowledge to boost team collaboration and information retention while optimizing communication procedures within modern distributed workplace settings.

Literature Survey

In recent years digital human avatars have become a major focus point for business application development. Qu et al. (2025) performed a systematic literature review that demonstrates how important avatar marketing and human-AI interactions in virtual settings have become with a focus on building trust between users and digital representations. The study demonstrates that building trustworthiness during digital human avatar interactions is essential specifically for business meeting environments. Bechtel et al. (2022) conducted a study that extended our understanding of human-like avatars in social and professional contexts. Agentic AI represents a transformative technology because it consists of autonomous systems that bring together perception, reasoning, and action capabilities essential for modern business communication. Billings and Henderson (2024) explain agentic AI as systems that enable AI-powered agents to plan overseas trips for users and supply-chain specialists to optimize inventory in real time, demonstrating a shift from passive to active AI capabilities. Walsh & Chen (2023) examine agentic AI's transformation of business process automation by enabling autonomous AI agents to independently analyze, plan, execute, and refine business processes. Our proposed framework depends critically on effective transcription and summarization of meetings as its technical

foundation. Adams et al.(2023) release a detailed survey study. They observed major hurdles in meeting summarization due to interactions that remain unedited featuring multiple participants and exhibit low information density.

Digital human avatars and voice cloning technologies require thorough ethical evaluation. Sen et al. (2023) emphasize the importance of purpose-driven design considerations across four key areas: Sen et al. (2023) recommend purpose-led design assessments across four crucial domains which include appearance characteristics together with voice and language attributes, personality traits and contextual and functional elements. Voxia AI (2023) lists foundational ethical practices necessary for businesses deploying voice cloning technology which involve acquiring explicit consent from users and maintaining transparency about AI voice utilization along with implementing strong security protocols and showing cultural awareness. Consumer Reports' (2024) extensive evaluation of AI voice cloning tools supports these principles through recommendations for strict standards that demand speaker consent verification alongside semantic guardrails to prevent misuse and watermarking of AI-generated audio. Long-term memory features present a substantial development for AI assistants operating in business environments. Etzioni and Shah (2024) examine how "Charlie Mnemonic uses GPT-4 to simulate human-like memory processes to enable personalized and enduring user interactions by combining different memory types which allow continuous learning."

Research Design and Methodology

The research utilized a mixed-methods experimental strategy to assess how AI-powered digital human clones perform in business communication settings. The core of our methodology centered on creating a functional hybrid meeting environment with three primary components: Our system includes three main components: an AI-powered digital avatar, a real-time interaction system and an agentic workflow automation layer. HeyGen/Similar's Creator Plan we constructed a personalized avatar of Mr. Nasim who serves as Senior Sales Manager. We enabled cloud-based recording and transcription for our Zoom meeting sessions. We engaged a group of 28 sales experts to serve as human counterparts for our Sales Team Lead position.

The experimental meeting interface in action is demonstrated in Figure 1. During Zoom sessions Mr. Nasim's digital avatar appeared live while displaying steady facial expressions and natural speech patterns throughout his interactions. Participants such as Shane (right) observed that high-quality visual rendering made interactions more engaging than conventional video recordings or static images, but qualitative feedback indicated the presence of an "uncanny valley" effect.

 ISSN 1681 8997

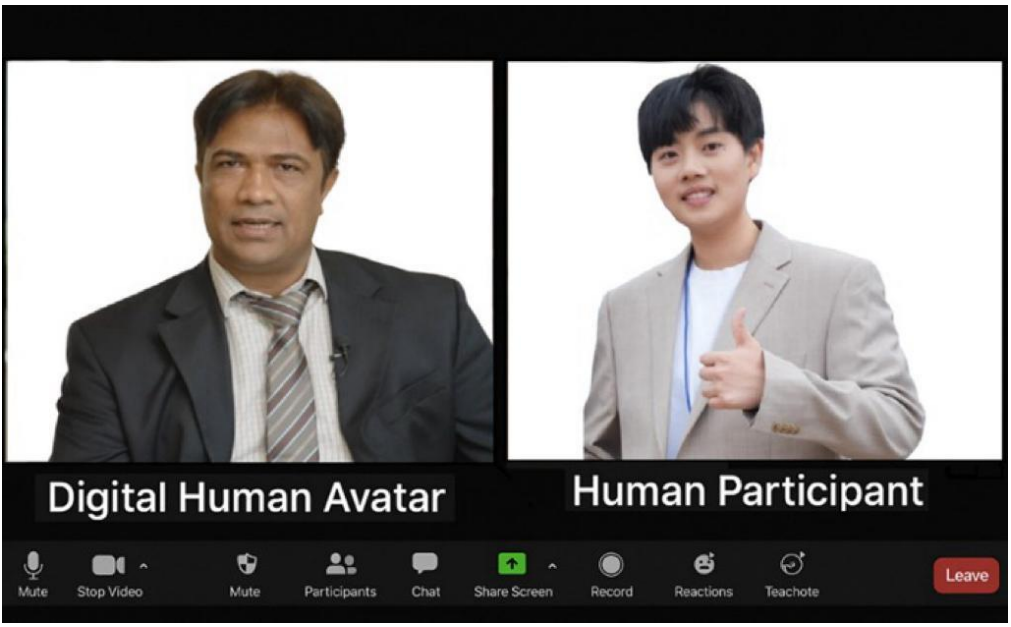


Figure 1: Experimental Zoom Meeting Environment. Screenshot of the hybrid meeting interface showing Mr. Nasim's digital avatar (left) and human participant Shane (right) during an experimental session

The complete process was managed through multi-step automation with N8N. Our data collection approach encompassed both quantitative and qualitative dimensions:

Quantitative Measures: The engagement metrics collected included speaking time ratio, turn-taking frequency and response latency measurements. Information retention scores (pre/post-meeting knowledge assessments). The system performance indicators measured transcription accuracy together with summary relevance and task completion rates. Participants completed post-interaction surveys with 7-point Likert scales to report their perceptions of naturalness, trustworthiness, and usefulness.

Qualitative Methods: We performed semi-structured interviews with participants following the completion of their interactions. The research team conducted thematic analysis on both transcripts from meetings and feedback collected from participants. Communication specialists conducted expert evaluations of avatar-human conversation quality. Analysis of meeting summaries created by AI compared to those written by humans. During an eight-week span we held 45 experimental sessions where every session used the same structure to have the avatar (Mr. Nasim) ask scripted questions about sales performance while Shane answered in his natural manner. Each session lasted 22 minutes.

 ISSN 1681 8997

System Development and Results

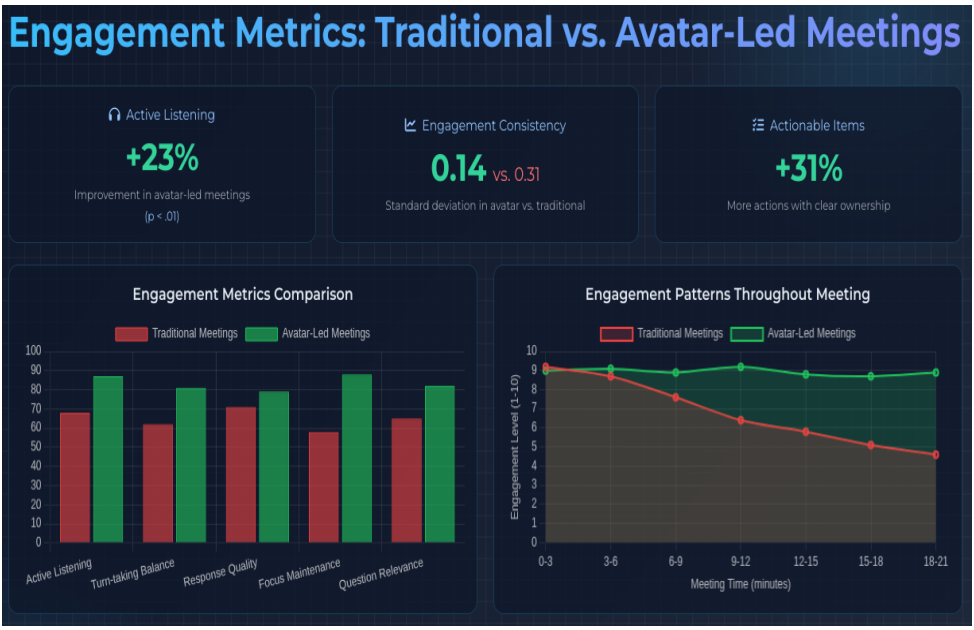


Figure 2: Engagement metrics comparison between traditional and hybrid avatar-human meeting

Engagement Metrics : The engagement analysis revealed significant distinctions when comparing conventional video conferencing to hybrid avatar-human communication. Figure 2 offers a comparative visualization of key engagement indicators. Participants showed a 23% increase in active listening behaviors during conversations led by avatars. The probability [p] being less than .01 indicates that turn-taking patterns were more balanced during these sessions compared to traditional meetings. Avatar-led questioning kept engagement levels stable throughout sessions and demonstrated a smaller variance in attention metrics (Standard deviation [SD] = 0.14) compared to traditional meetings which had higher variance (Standard deviation [SD] = 0.31). Figure 3 shows a heat map representation of engagement metrics over time segments comparing traditional meeting formats to avatar-led meeting formats. The engagement metrics of traditional meetings drop significantly after 9 minutes before reaching their minimum between 18 and 21 minutes. The four metrics of attention, participation, question response, and note-taking were evaluated using a scale ranging from 1 to 10. The heat map color patterns from avatar-led meetings align with our results that show reduced engagement variance metrics (SD=0.14 compared to SD=0.31 in traditional meetings).

 ISSN 1681 8997

Figure 3: Engagement Patterns throughout Meeting Duration

Information Retention and Processing: Assessment results demonstrated substantial benefits for hybrid meetings regarding information retention. The violin plot in Figure 4 illustrates the distribution of knowledge assessment scores taken before and after the meeting.

The avatar-led meeting condition achieved higher post-meeting knowledge scores (Mean[M] = 8.3, Standard deviation [SD] = 1.2) which outperformed traditional meetings (Mean[M] = 6.5, Standard deviation [SD] = 1.8) by 27.7% (probability [p] < .001). Meeting transcripts analysis showed avatar-led meetings generated 31% more actionable items which had clearer ownership assignments.

System Performance and Workflow Automation: The agentic AI workflow proved to be highly reliable in processing meeting outcomes. Transcription accuracy achieved 92.4% **Word Error Rate [WER]:** The summary relevance achieved an evaluation score of 83.7% from experts. The automation system managed 76.2% of follow-up tasks independently which included sending emails and setting up calendar events.

 ISSN 1681 8997

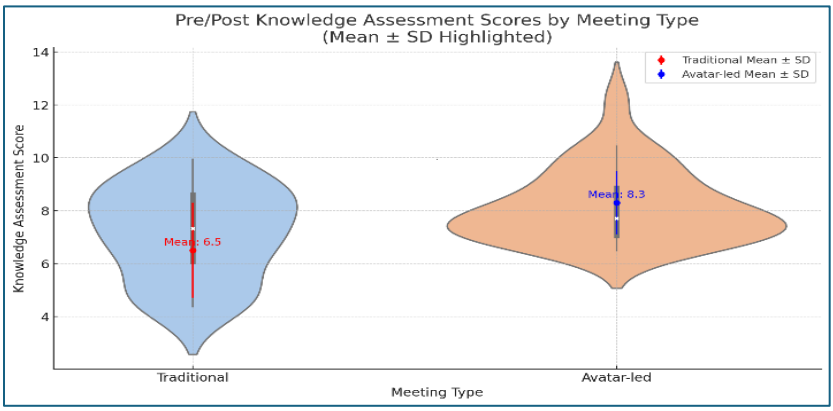


Figure 4: Information retention distribution before and after meetings

User Perception and Experience: The quality of avatar interactions as reported by participants demonstrated complex views about their digital human experience. Figure 5 illustrates the perception ratings for various important dimensions.

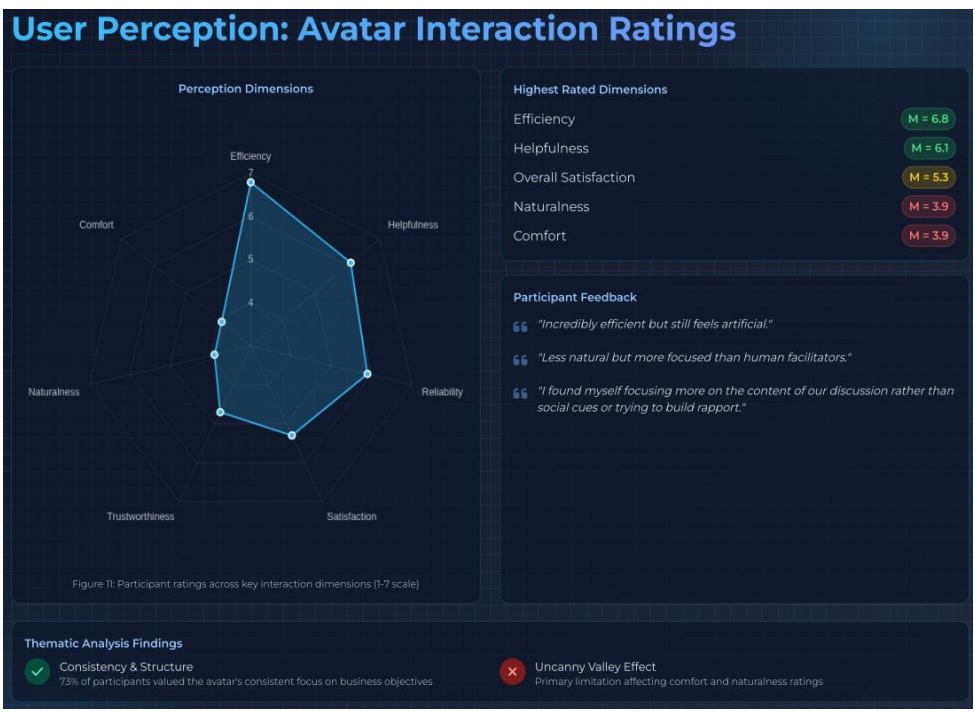


Figure 5: User perception ratings of avatar interaction

The avatar achieved the highest ratings in efficiency (Mean[M] = 6.8) and helpfulness (Mean[M] = 6.1) while naturalness (M = 3.9) and comfort (Mean[M] = 3.9) scored lower ratings. The main limitation identified through qualitative feedback was the "uncanny valley" effect since participants found the avatar to be "incredibly efficient but still feels artificial" and "less natural but more focused than human facilitators." Analysis of interview themes showed that participants appreciated the consistency and structure of meetings led by avatars and 73% of them observed that the digital human kept better attention on business objectives than traditional meetings did.

Discussion

Structured questioning techniques and fewer social distractions produced a 23% rise in active listening behaviors and a 27.7% boost in information retention. Task-oriented business environments can foster functional trust based on the avatar's efficiency rating of 6.8 and helpfulness rating of 6.1 despite limited emotional connection. The "uncanny valley" effect identified in prior research continues to explain why naturalness (M = 3.9) and comfort (M = 3.9) ratings remain low. Participants reported high satisfaction ratings of 5.3 overall while giving lower naturalness scores shows their appreciation of functional benefits over human-like simulation which provides new insights into avatar research that has mainly examined social acceptance instead of task effectiveness. Current NLP technologies can reliably automate meeting documentation because they deliver high transcription accuracy (92.4%) and summary relevance (83.7%). The combination of avatar-directed dialogue structure with automated processing generates meeting outcomes with clear ownership of actions which shows a 31% increase in accountability. The research project faced multiple ethical issues which demanded careful handling. We adhered to the protocol recommendations provided by Sen et al. (2023) We established explicit consent measures for voice cloning while providing clear information about AI applications in accordance to Sen et al. (2023) and Consumer Reports (2024). The 76.2% automation completion rate demonstrates potential but reveals areas where agentic workflow reliability must be enhanced.

Conclusion

Organizations that prioritize knowledge capture for decision-making gain substantial benefits from the 27.7% improvement in information retention. The improvements seen in consistent engagement along with the creation of 31% more actionable items with clear ownership assignments address typical challenges in traditional meetings where focus tends to wane, and accountability gets diluted. The agentic workflow automation shows effectiveness at 76.2% completion rate which confirms the potential of end-to-end communication systems that enable conversation facilitation and comprehensive outcome processing and documentation. Longitudinal studies investigating recurrent interactions

with digital avatars would ascertain whether advantages last as novelty effects diminish. Exploring a broader range of business sectors and cultural contexts might enhance the generalizability of our results. By focusing technical enhancements on reducing the uncanny valley effect through stylized designs instead of hyper-realistic models we might overcome the existing barriers to perceived naturalness.

4.2 Analytical Applications of Natural Language Processing (NLP) in Cyber Threat Intelligence

Authors: Robert G Cutlip *(Fairmont State University)*
Rebecca Giorcelli *(Fairmont State University)*
Yeabsira M Dana *(Fairmont State University)*
Teame Gidena Hiben *(Fairmont State University)*
Getachew Hailu Kinfu *(Fairmont State University)*
Gebrehiwot Tadesse Gebremariam *(Fairmont State University)*

Natural Language Processing (NLP) has revolutionized the analysis of textual data, enabling machines to extract actionable insights from human language. This paper explores the intersection of NLP and Cyber Threat Intelligence (CTI), highlighting analytical methods that enhance threat detection and mitigation strategies. Core NLP techniques such as Sentiment Analysis, Latent Semantic Indexing (LSI), Singular Value Decomposition (SVD), and Latent Dirichlet Allocation (LDA) are discussed, emphasizing their applications in identifying emerging threats, understanding adversarial patterns, and improving situational awareness in cybersecurity.

Sentiment Analysis, particularly P-N Polarity and S-O Polarity, offers insights into threat actors' intent by classifying texts as positive, negative, or neutral and distinguishing subjective opinions from objective facts. Deep learning models like LSTMs and transformers, including BERT, significantly advance the precision of such classifications. Techniques like tokenization, stemming, and lemmatization form the foundational preprocessing steps, ensuring accurate representation of textual data.

Advanced methodologies such as LSI and SVD uncover latent patterns in threat intelligence datasets, improving the detection of hidden associations and thematic clustering. LDA, a probabilistic generative model, facilitates topic extraction from cyber threat reports, enabling the identification of prevalent attack trends and vectors. The comparison of supervised and unsupervised learning models underlines the trade-offs between accuracy and the requirement for labeled data in cybersecurity applications.

This synthesis underscores the critical role of NLP in fortifying cybersecurity frameworks by transforming textual data into actionable intelligence, paving the way for proactive and informed decision making in threat management.

4.3 Enhancing Cyber Threat Intelligence Through Advanced Phishing Detection and Prevention Technologies

Authors: Robert G Cutlip *(Fairmont State University)*
Rebecca Giorcelli *(Fairmont State University)*
Getachew Hailu Kinfu *(Fairmont State University)*
Gebrehiwot Tadesse Gebremariam *(Fairmont State University)*
Teame Gidena Hiben *(Fairmont State University)*
Yeabsira M Dana *(Fairmont State University)*

Phishing remains one of the most prevalent and damaging forms of cyber threats, exploiting human vulnerabilities through social engineering tactics to achieve financial gain or data theft. This paper explores the analytical methods employed in detecting phishing attacks and their implications for Cyber Threat Intelligence (CTI). We delineate the four critical stages of a phishing attack: bait, hook, catch, and exploitation, emphasizing how cybercriminals manipulate emotions to deceive victims into divulging sensitive information. To combat these threats, we highlight the significant advancements in Machine Learning (ML), Natural Language Processing (NLP), and specifically Large Language Models (LLMs) that are revolutionizing phishing detection and prevention. These technologies enable the analysis of vast datasets to identify patterns indicative of phishing attempts, offering real-time detection capabilities that traditional methods lack. We discuss specific algorithms such as neural networks, and deep learning models that have shown promise in enhancing the accuracy of phishing identification. The integration of these advanced detection techniques into Cyber Threat Intelligence framework allows organizations to proactively monitor and respond to phishing threats. By utilizing ML and NLP, CTI can evolve from reactive measures to predictive analytics, thereby improving overall cybersecurity posture. This paper emphasizes the necessity for organizations to adopt these technologies to safeguard sensitive information and mitigate risks associated with phishing attacks.

4.4 Revolutionizing Recruitment: Enhanced Machine Learning Models for Bias Mitigation and Efficiency

Authors: Mukhesh Ravi *(Northwood University)*
Harika Yenuga *(Northwood University)*
Itauma Itauma *(Northwood University)*

This study investigates the application of machine learning models to optimize recruitment processes by predicting hiring decisions based on comprehensive candidate profiles. Utilizing a dataset of 1,500 candidates, the research incorporates diverse features such as age, gender, education level, work experience, skill scores, social media activity, and psychometric assessments. Several machine learning algorithms, including Random

Forest, Support Vector Machine, Logistic Regression, and advanced ensemble methods (CatBoost, XGBoost), were evaluated to determine the best-performing model and identify the most influential features. Results indicate that CatBoost outperformed other models, achieving an accuracy of 95%, followed by Random Forest and XGBoost. Analysis of feature importance revealed that Recruitment Strategy, Education Level, and Personality Score were the top three factors influencing hiring decisions.

This study extends existing research by implementing deep learning techniques and fairness-aware algorithms to mitigate biases in hiring decisions. Specific bias detection and correction methods, alongside fairness metrics such as demographic parity and equal opportunity, were employed to ensure equitable hiring practices. Explainable AI methods, including SHAP and LIME, were incorporated to enhance model interpretability, providing HR professionals with actionable insights. The research also addresses challenges such as bias in hiring and inefficiencies in recruitment by integrating ethical guidelines and data privacy measures to ensure transparency and accountability in AI-driven recruitment processes.

The findings demonstrate the potential of machine learning to improve recruitment efficiency, reduce biases, and promote fairer hiring outcomes. By leveraging advanced algorithms, this study highlights the real-world impact of AI in enhancing recruitment processes, offering benefits such as increased efficiency, cost reduction, and greater fairness in candidate selection.

Track 5: Healthcare Analytics
Room: WC S204A (11:00 am - 12:20 pm)
Moderator: Shazad Mohammed *(State University of New York at Fredonia)*

5.1 Analyzing Hemoglobin A1C Impact on Diabetes Care Outcomes: a Machine Learning Approach and Comprehensive Analysis of Electronic Health Records

Author: Soumik Banerjee *(Canisius University)*

Diabetes is a global endemic with rapidly increasing prevalence in both developing and developed countries. The American Diabetes Association has recommended glycated hemoglobin (HbA1c) as a possible substitute to fasting blood glucose for diagnosis of diabetes. HbA1c is an important indicator of long-term glycemic control with the ability to reflect the cumulative glycemic history of the preceding two to three months. Alongside diabetes, chronic conditions occur and can impact a patient's death which is called comorbidities. This research analyses over 100,000 hospital records to assess the relationship between comorbidities and hospital readmission rates among diabetic patients, with health records extracted from the Cerner Health Facts Database. Multivariate logistic regression, and an array of machine learning techniques were used to fit the relationship

between patient comorbidities and early hospital readmission to compare the impact of significant covariates on model performance.

Previous researchers observed that the measurement of HbA1c during patient encounters is associated with a lower rate of readmission among patients with a primary diagnosis of diabetes, compared to circulatory or respiratory diseases (Strack et al., 2014). The aim of the presented analysis is to extend these findings by investigating the interaction of high-risk comorbidities on HbA1C levels and 30-day hospital readmissions.

5.2 From Data to Decisions: AI-Powered Predictive Analytics in Breast Cancer Risk and Treatment Optimization

Authors: Mohammad Hossein Amini *(Washington University in St. Louis & Olin Business School)*
Zihang Shi *(Washington University in St. Louis)*
Arefeh Esmaeilpour *(Washington University in St. Louis)*

Abstract: The FDA Adverse Event Reporting System (FAERS) serves as a valuable resource of real-world data on drug safety, adverse reactions, and patient outcomes, offering vital insights for healthcare analytics. This study utilizes FAERS data to develop an AI-driven predictive model that assesses patient outcomes and risk factors, specifically in breast cancer treatment. Traditional methods for risk assessment often depend on clinical trials and retrospective studies, which may not effectively capture the complexities of real-world scenarios. In contrast, applying AI and machine learning techniques to FAERS data allows for the identification of patterns, prediction of adverse events, and optimization of patient management strategies. Our model predicts the likelihood of severe outcomes by integrating FAERS-reported adverse events with factors such as patient demographics and treatment regimens. Advanced algorithms, including natural language processing (NLP), enable us to extract insights from unstructured FAERS reports, enhancing early risk detection. The model's real-time risk stratification facilitates better treatment plans for healthcare providers and pharmaceutical companies, improving patient safety and reducing costs. The findings underscore the transformative potential of FAERS-based predictive analytics in refining breast cancer treatment strategies. Future developments may include expanding AI capabilities to create recommender engines that offer alternative therapies, thereby further minimizing associated risks for patients. This research highlights AI's vital role in enhancing predictive analytics for patient care in breast cancer and beyond.

1. Introduction

Breast cancer is the most common cancer among women worldwide, and there have been notable trends in its incidence and outcomes since the mid-2000s. In the United States, the incidence rates have experienced a slight increase over the past two decades, while mortality

rates have decreased

(1). Breast cancer mortality in the US has dropped significantly (about 40% decline since 1989), thanks to earlier detection and better treatments (2). As of 2021, approximately 4 million women in the US have a history of breast cancer. Current projections for 2024 anticipate over 310,000 new cases and around 42,000 deaths in the US alone. (1). Breast cancer incidence has been rising globally, particularly in developing countries. In 2020, it surpassed lung cancer to become the most diagnosed cancer worldwide. That year, there were approximately 2.3 million new cases of breast cancer, accounting for about 11.7% of all cancer cases, and it resulted in 685,000 deaths. (3).

The treatment of breast cancer typically involves a combination of methods, including chemotherapy, hormone (endocrine) therapy, targeted therapy, and immunotherapy. While these treatments have significantly improved survival rates, they tend to be expensive and can cause side effects that affect patients' quality of life.

Both the disease and its treatment can be burdensome for patients, making the cost a substantial barrier. To help reduce the expenses associated with managing the disease, we can utilize artificial intelligence to predict patient responses to available treatments. This approach can provide policymakers and healthcare providers with valuable tools to assess risks and optimize insurance pricing for patients.

2. Data

In this research, we aim to develop a predictive model to assess the associated risk of the available treatments. We customized the model to predict a risk for a specific given patient information. To do so, we gathered more than 100GBs of data sourced from the FDA Adverse Event Report System called FAERS, in short (4). FAERS is a publicly available database that collects reports on adverse drug reactions, offering valuable insights into the safety and effectiveness of disease treatments. By using AI to analyze FAERS data, businesses can identify drug safety trends, optimize pharmacovigilance, and uncover opportunities for drug repurposing or market entry.

To mine specific and relevant data, we used the standard terms for breast cancer by MedDRA (5). Using effective and efficient search methods, we could collect around 6 million records from the FAERS database spanning from 2004 up to the third quarter of 2024.

3. Methodology and Results

Data include many notable features of reports such as demographic information of the patients, such as age and gender, information about drugs like the name of the drug, dosage used and duration of usage, adverse reaction reported after using the medicine, and

 ISSN 1681 8997

outcome of the therapy. Here, we first explore the data to retrieve insightful information and better understand the pattern.

4. Data Exploration

Figure 1 illustrates the number of unique cases and unique reactions (adverse events, or AEs) reported each year. It indicates a steady increase in the overall number of reports over time. Notably, there was a spike in unique AEs in 2012, while 2021 saw a significant increase in the number of cases, followed by a subsequent decline. Other researchers (6) have observed similar trends, which can be associated with the effects of the Covid-19 pandemic. It is noted that, due to the availability of the FAERS dataset, the data for 2024 is available only up to the third quarter. This indicates a shortfall regarding AEs and unique cases in 2024.

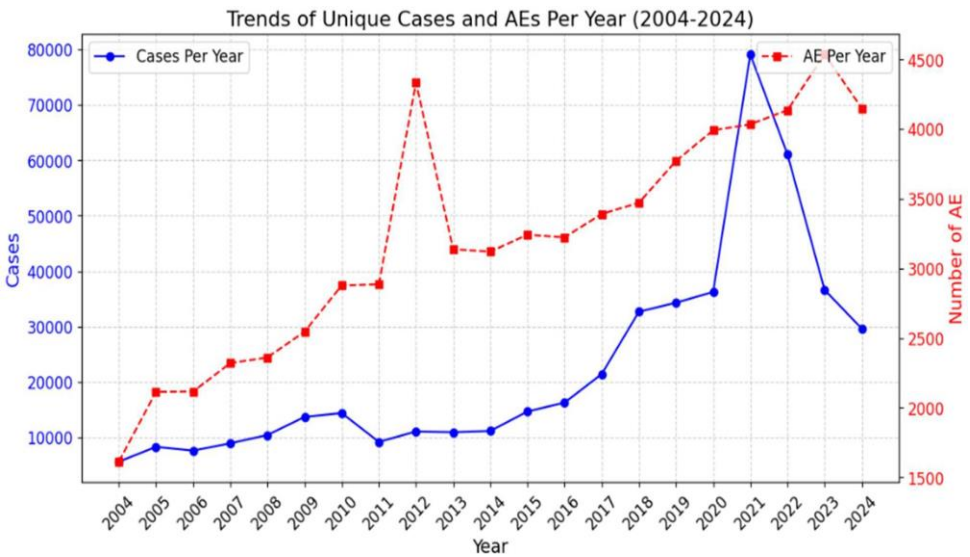


Figure 1: Number of unique cases and AEs per year

The patient's outcome (OC) plays a vital role in understanding the data as it provides an essential measure of the risks associated with each report. The outcomes are systematically categorized using a set of specific codes: death (DE), life-threatening (LT) situations, disability (DS), hospitalization (HO), congenital anomalies (CA), required intervention (RI), and other serious outcomes (OT). To conduct a more thorough analysis of the reports over the years, we carefully reviewed the total number of reports and determined the percentage of each category within the reported cases.

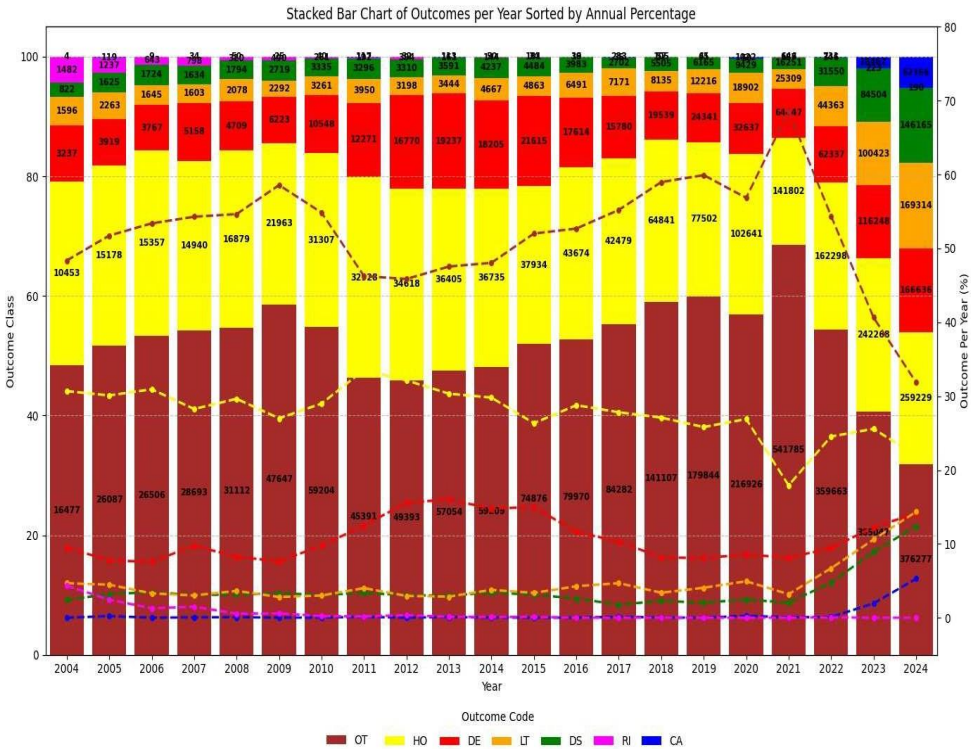


Figure 2: Outcome trends over the years

This structured approach provides a clearer view of trends, illustrated in Figure 2. It shows a significant rise in case severity from 2004 to 2015, followed by a decline. Notably, 2021 marks a crucial inflection point, with an increase in severe cases, including fatal and life-threatening outcomes.

There have been more than 8,300 unique AEs reported by patients where diarrhea and fatigue are among the most common AEs reported more than 60,000 times. Figure 3 illustrates a word cloud of AEs, highlighting the more frequent AEs.

In order to provide a clearer depiction of the AEs, we standardized the counts. This approach prevents Fatigue, Nausea, Diarrhea, and Vomiting from overwhelming the representation. Among the AEs, Alopecia (hair loss), Neutropenia, and Dyspnoea are notably frequent occurrences.

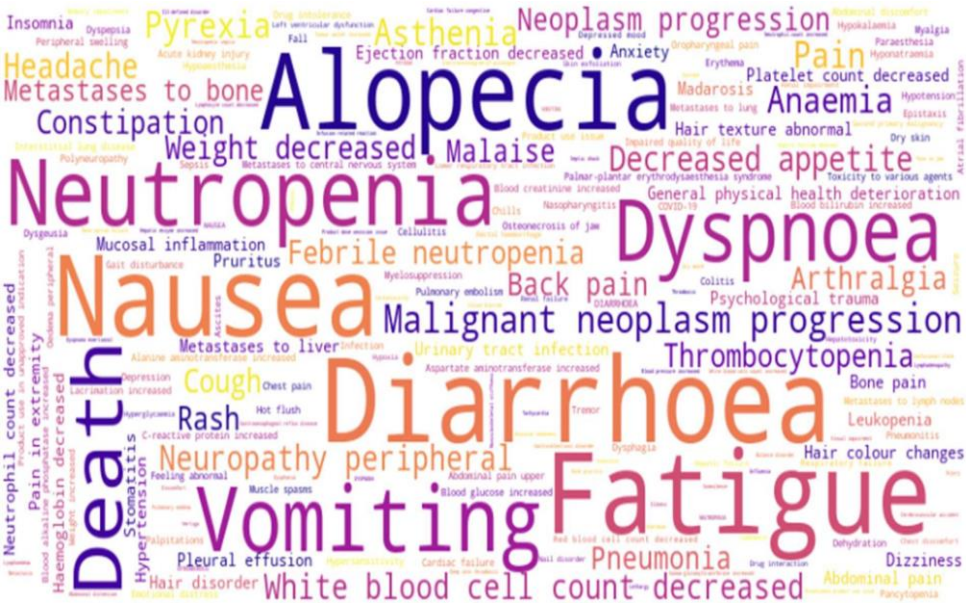


Figure 3: Word Cloud of AEs

5. Predictive Modeling

To assess the risks associated with various treatments, we selected a limited number of features. This included patient information such as country, gender, age, and weight, as well as therapy- related details like the drug name, reported reactions, and the outcomes of those reports. After cleaning the data and removing outliers, we ended up with approximately 166,000 rows of data, which included more than 8,000 different reactions reported by patients from 27 countries. As expected, the United States was the most represented country in the data, followed by Denmark and China. To quantify the treatment risk, we designated fatal cases as severe and all other cases as non-severe. Due to the significant class imbalance in the data, we employed a random under- sampling method to equalize the number of cases in each class (severe vs non-severe). This process further reduced the number of cases to approximately 36,000. To develop a predictive model, we trained multiple supervised classification models on 80% of the data while reserving the rest for testing purposes. Among the models, KNN performed the best with 80% accuracy (considering balanced data).

The developed model indicates that by utilizing fundamental patient information, such as age and weight, we can evaluate the risk associated with drug toxicity with an accuracy of 80%. This assessment is based on comprehensive toxicity data available for various

medications. This approach aims to improve patient safety by identifying potential adverse effects before drug administration.

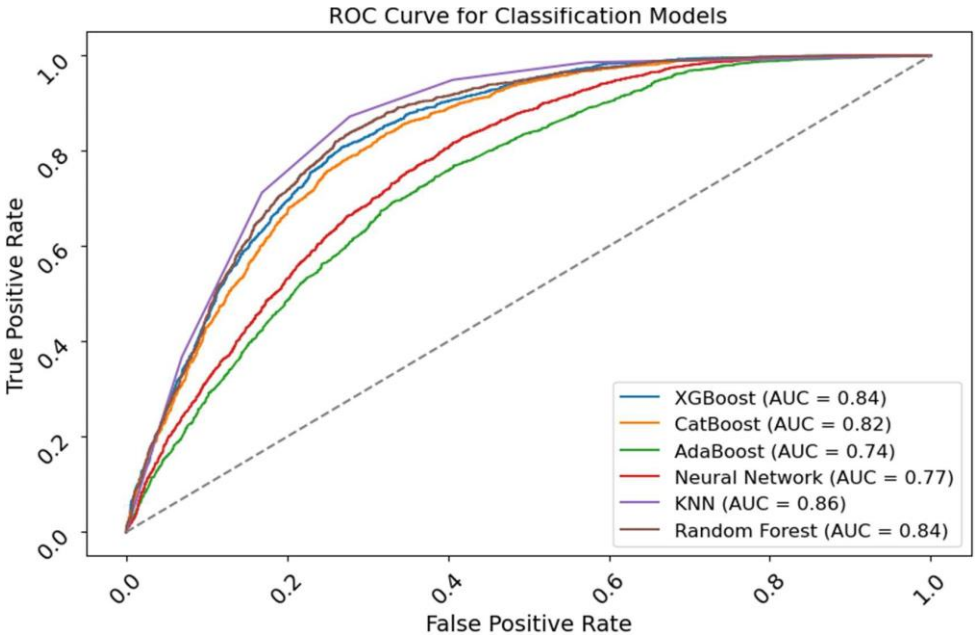


Figure 4: ROC Curve for classification models

6. Discussion

According to Figure 4, the model shows promise in assessing the risks associated with available therapies. This developed model can assist policymakers and insurance companies in evaluating and adjusting therapy pricing more effectively. In the next phase, we plan to create a recommendation system that will suggest alternative medications for specific patients, aiming to minimize associated risks.

References

1) **cancer.gov.** Cancer Stat Facts: Female Breast Cancer. *seer.cancer.gov.* [Online] 2020. https://seer.cancer.gov/statfacts/html/breast.html#:~:text=2004%20128,92.19.
2) **Society, American Cancer.** Breast Cancer Facts & Figures 2019-2020. *cancer.org.* [Online] 2019. https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast- cancer-facts-and-figures/breast-cancer-facts-and-figures-2019-2020.pdf.

3) *Global and regional trends in incidence and mortality of female breast cancer and associated factors at national level in 2000 to 2019.* **Luo, C, et al.** 2022, Chin Medical Journal, pp. 42-51.
4) **FDA.** FDA Adverse Event Reporting System. *FDA.* [Online] 7 12, 2023. https://www.fda.gov/drugs/fdas-adverse-event-reporting-system-faers/fda-adverse-event- reporting-system-faers-public-dashboard.
5) **MedDRA.** MedDRA. *MedDRA.* [Online] 2025. https://www.meddra.org/.
6) *Breast cancer mortality during the COVID-19 pandemic.* **C, Mattiuzzi and G., Lippi.** 2024, Journal of Medical Screen.

5.3 Financial Health vs. Care Quality in Skilled Nursing Facilities: Assessing the Impact of COVID-19

Authors: Ai Ren *(SUNY New Paltz)*
Pavlo Mysak *(SUNY New Paltz)*
Svetlana Doronkina *(SUNY New Paltz)*
Qi Li (*SUNY New Paltz)*

The COVID-19 pandemic placed unprecedented financial and operational strain on skilled nursing facilities (SNFs) in the United States, exacerbating long-standing challenges in healthcare funding and quality of care. This study examines the relationship between financial health and care quality in SNFs, with a particular focus on the impact of COVID-19. Our findings reveal that better financial health correlates with higher quality ratings, while increased reliance on Medicaid funding is associated with lower operating margins. For-profit facilities demonstrate higher profitability, often at the expense of staffing levels. However, facilities with a higher number of COVID-19 cases experienced increased operating margins. These insights highlight the fragile balance between financial stability and care quality in the SNF sector and underscore the need for policy interventions to ensure equitable and sustainable long-term care.

Track 6: Supply Chain and Operations

Room: WC S204B (11:00 am - 12:20 pm)
Moderator: Joseph Porter, Jr *(Nazareth University, USA)*

6.1 Exploring the Role of Generative AI in Enhancing Supply Chain Integration

Authors: Abdullah Oguz and Hassan A. Ahmed *(Cleveland State University)*
Carlo Gabriel Porto Bellini *(Universidade Federal da Paraíba (UFPB))*

This study investigates the potential for the transformative potential of generative artificial intelligence (GenAI) as a supply chain integration (SCI) enabler. Taking the legacy of ICTs and current progress in AI use in supply chain management, this paper highlights the strategic use of GenAI for real-time data sharing, collaborative decision-making, and risk

mitigation. Businesses can address inefficiencies and improve supply chain resilience using GenAI's natural language, predictive analytics, and scenario modeling tools.

This research defines the opportunities and challenges based on the existing literature on cultural enablers, technology adoption models, and supply chain AI integration challenges. Cultural, technological, and organizational integration is the key to GenAI deployment. This research underscores the need for data-driven culture, employee training, and inter-organizational cooperation to unleash GenAI's potential fully.

While conceptual in nature, this study opens the path for future empirical work to confirm and extend these results. By providing an integrated framework for GenAI-driven supply chain integration, this study will aid academic research and practical solutions to enable more robust, efficient, and sustainable supply chains.

6.2 Unveiling the Role of AI in Project Scheduling: Opportunities and Challenges

Authors: Kasun Wijayagurusinghe *(Cleveland State University)*
Chris Wimer *(Cleveland State University)*
Abdullah Oguz *(Cleveland State University)*

Much attention has recently been given to the potential for AI to revolutionize many aspects of work due to its capabilities. Despite this interest, however, a collective view of the broad ways in which AI can influence aspects of a specific area of work has yet to receive much study in the existing literature. In this study, a qualitative thematic analysis approach is used to better understand how adopting and implementing AI technologies in the field of project scheduling are perceived as leading to major changes in this area of work. From this analysis, AI is shown to have an impact on project scheduling through its abilities to both enhance and optimize processes within this area, but further research is needed to better understand how to successfully integrate AI into a workplace as well as the various implications which arise from incorporating AI into a given area of work.

6.3 The Role of Food Donations and Recycling in Reducing Food Waste in Supply Chains

Authors: Shahryar Gheibi *(Siena College)*
Jennifer Pazour *(Rensselear Polytechnic Institute)*

Do donations help reduce food waste? Indeed, they do, conventional wisdom may suggest. Our paper takes an in-depth look at the impact of food reuse (donations and recycling) on food waste and attempts to provide an answer grounded in empirical data and informed by an analytical model.

We find, interestingly, that the majority of food types feature supply chains with evidence of ""oversupply"" in the sense that the food surplus growth significantly exceeds the consumption growth.

Our paper examines the importance of accounting for the impact of donation incentives and capacity on the stakeholders' behavior and decisions. We hypothesize that a dynamic feedback loop underlies the oversupply phenomenon and construct an analytical model to provide insights into it. Our paper shows that food reuse incentives may have counterintuitive effects on food waste.

Track 7: Emerging Scholars II: Cybersecurity, Threat Detection, and Using Machine Learning

Room: WC S204C (11:00 am - 12:20 pm)
Moderator: Lisa Walters *(State University of New York at Fredonia)*

7.1 The Use of Neural Networks in Cyber Threat Intelligence

Authors: Robert G Cutlip *(Fairmont State University)*
Rebecca Giorcelli *(Fairmont State University)*
Teame Gidena Hiben *(Fairmont State University)*
Gebrehiwot Tadesse Gebremariam *(Fairmont State University)*
Getachew Hailu Kinfu *(Fairmont State University)*
Yeabsira M Dana *(Fairmont State University*)

Neural networks have emerged as a cornerstone of artificial intelligence, offering unparalleled capabilities in Cyber Threat Intelligence (CTI). These models mimic biological neural systems and employ multilayered architectures to process vast and complex datasets, enabling real-time detection and analysis of sophisticated cyber threats. From feedforward networks to advanced frameworks like convolutional neural networks (CNNs) and recurrent neural networks (RNNs), neural networks have revolutionized anomaly detection, intrusion identification, and phishing prevention. CNNs excel in image recognition tasks, such as identifying malware through pixellevel analysis, while RNNs, particularly Long Short-Term Memory (LSTM) networks, are adept at handling sequential data for log monitoring and threat pattern recognition. Generative Adversarial Networks (GANs) further enhance CTI by simulating potential attack scenarios, enabling proactive defence strategies. Despite these advancements, challenges remain, including the interpretability of "black-box" models, data imbalances, and vulnerability to adversarial attacks. Addressing these requires robust data preprocessing, regularization techniques, and the integration of explainable AI methods. This paper highlights the transformative impact of neural networks in CTI, emphasizing their role in improving threat detection accuracy, enabling automated response systems, and fostering resilience against evolving cyber threats. As these technologies continue to evolve, their integration into cybersecurity

ecosystems promises enhanced protection for critical infrastructures and digital assets. Future research should focus on scalability, ethical considerations, and developing hybrid models to address the dynamic nature of cyber threats.

7.2 Application of Anomaly Detection Approaches in Cyber Threat Intelligence

Authors: Robert G Cutlip *(Fairmont State University)*
Rebecca Giorcelli *(Fairmont State University)*
Gebrehiwot Tadesse Gebremariam *(Fairmont State University)*
Yeabsira M Dana *(Fairmont State University)*
Getachew Hailu Kinfu *(Fairmont State University)*
Teame Gidena Hiben *(Fairmont State University)*

Anomaly detection, a critical data mining technique, plays a vital role in cyber threat intelligence by identifying deviations from normal patterns in data, often signaling security breaches, fraud or other malicious activities. This study examines machine learning (ML) methods including supervised, unsupervised and semi-supervised approaches and their effectiveness in detecting anomalies. Techniques such as k-nearest neighbors, support vector machines, isolation forests, autoencoders and generative adversarial networks (GANs) with their ability to analyze high-dimensional and complex datasets are explored. Machine learning enhances anomaly detection by enabling systems to learn from evolving data patterns, improving accuracy and adaptability. For instance, deep learning models like autoencoders excel in identifying subtle and dynamic cyber threats while traditional methods like random forests and DBSCAN can be effective in less complex scenarios. Integrating these techniques into cybersecurity frameworks strengthens the ability to detect unauthorized access, malware, and data exfiltration by providing proactive threat mitigation. However, challenges like imbalanced datasets, computational demands and scalability issues remain barriers to implement Anomaly detection approaches for accurate cyber threat intelligence. Addressing these challenges through hybrid models like explainable AI tools and domain-specific optimization is critical for future progress. This study underscores the importance of anomaly detection in cybersecurity, emphasizing its ability to enhance threat intelligence, reduce false positives and support real-time decision-making. By leveraging advanced ML techniques, organizations can better protect their digital assets against increasingly sophisticated cyber adversaries which indicates that anomaly detection is a cornerstone of modern cybersecurity strategies.

7.3 Leveraging Behavior-Based Anomaly Detection for Protecting Internet of Things (IoT) Devices

Authors: Saumya Ranjan Padhi *(Northwood University & DevOS Graduate School)*
Adeeda Mukhtar *(Northwood University & DevOS Graduate School)*
Itauma Itauma (Northwood University & DevOS Graduate School)

Abstract: The widespread use of Internet of Things (IoT) devices has led to more exposure to cyber-attacks with their limited processing and security capabilities. This paper

investigates the use of supervised machine learning (ML) models in detecting anomalies in IoT environments from the Edge-IIoTset dataset, including labeled network traffic such as botnet activity, DoS attacks, and unauthorized access. The data was also preprocessed using one-hot encoding, normalization, and correlation-based feature selection. We trained and tested different models such as Logistic Regression, Naïve Bayes, Random Forest and AdaBoost. Among these, tree- based ensemble models such as XGBoost, Random Forest, and CatBoost performed better, with XGBoost having highest accuracy (98.46%) and F1-score. These models effectively detected various threats on the basis of irregular packet lengths, abnormal traffic intervals, and malicious IP accesses. The findings support previous research and validate the scalability and efficiency of ensemble models for edge-based real-time deployment. This study emphasizes the importance of lightweight, interpretable intrusion detection systems for resource-constrained IoT devices.

Index Terms—Internet of Things (IoT), Intrusion Detection System (IDS), Anomaly Detection, Machine Learning, Ensemble Learning, Random Forest, AdaBoost, XGBoost, Edge Computing, Cybersecurity, Supervised Learning, Network Traffic Analysis, IoT Security, Real-time Detection, Feature Selection.

1. Introduction

The Internet of Things (IoT) is transforming industries like healthcare, manufacturing, and smart infrastructure by enabling better connectivity and automation. However, these benefits come with security risks. Since IoT devices are always connected, vary in design, and require little human supervision, they are vulnerable to cyberattacks (Mukherjee et al., 1994; Buczak & Guven, 2015). Many of these devices have limited resources, run outdated software, and lack a common security framework, making it hard for traditional intrusion detection systems (IDS) to protect them (Northcutt & Novak, 2002). These older IDS mainly rely on fixed rules or known attack patterns, making them ineffective against new or unknown threats.

To solve this problem, machine learning (ML)-based intrusion detection systems (B-IDS) are becoming more popular. Instead of depending on predefined attack signatures, these systems learn from network traffic patterns to detect unusual behavior (Aburomman & Reaz, 2017; Al-A'araji et al., 2021). Advanced ML techniques, such as Random Forest and AdaBoost, improve accuracy, reduce errors, and help make models more understandable (Alotaibi & Ilyas, 2023; Al-Omari et al., 2021). This research focuses on using these ML models to analyze an IoT network dataset and detect cyber threats like malware, denial-of-service (DoS) attacks, and unauthorized access. The goal is to develop security solutions that are scalable, easy to understand, and efficient enough to work on IoT devices with limited computing power.

2. Literature Review

Traditional intrusion detection systems (IDS) struggle in IoT environments, where resource-limited devices operate in distributed networks. Signature-based IDS detect known threats but fail against evolving attack patterns (Mukherjee et al., 1994; Northcutt & Novak, 2002; Panigrahi et al., 2022). This has led to the rise of machine learning (ML)-based IDS, which analyze normal network behavior and detect anomalies.

Ensemble learning models have shown strong potential. Aburomman and Reaz (2017) found that combining classifiers improves detection by reducing bias and variance. Al-A'araji et al. (2021) showed that mixing clustering and classification enhances detection rates, while Alotaibi and Ilyas (2023) demonstrated ensemble learning's adaptability in real-time IoT security.

Tree-based models like XGBoost, Random Forest, and CatBoost are widely used due to their scalability and minimal preprocessing needs. Al-Omari et al. (2021) developed a tree-based IDS with high accuracy. Balyan et al. (2022) proposed a hybrid ensemble with Random Forest, achieving state-of-the-art results. Das et al. (2022) confirmed that boosting techniques like XGBoost and AdaBoost outperform traditional ML models.

Our study applied preprocessing techniques such as one-hot encoding and scaling to an IoT dataset with labeled attack traffic. The models performed as follows:

These findings align with Dhaliwal et al. (2018), who successfully applied XGBoost for scalable IoT intrusion detection.

Recent studies emphasize explainable and lightweight IDS models. Zhang et al. (2021) introduced a stacking ensemble that improves detection and interpretability. While we did not fully explore LightGBM, Dorogush et al. (2018) showed these models efficiently handle categorical features, reducing preprocessing.

Overall, ensemble learning, and behavior- based models balance accuracy, efficiency, and explainability for IoT security. Our study reinforces their effectiveness and suggests future work in explainable AI and lightweight IDS for real-world use.

3. Research and Methodology

Research Objective, this study aims to evaluate the effectiveness of various supervised machine learning models, particularly ensemble methods, in accurately detecting anomalies within the Edge-IIoTset dataset for enhanced IoT device protection.

Furthermore, it seeks to explore the impact of data preprocessing and class imbalance handling techniques on the performance of these models for scalable and efficient edge-based intrusion detection.

Data Collection, the dataset utilized in this study, is the Edge-IIoTset cyber security dataset of IoT and IIoT devices, publicly available on Kaggle. This dataset contains

labeled network traffic data relevant to IoT environments, specifically including instances of normal network activity and various cyber- attacks such as botnet activity, Denial-of-Service (DoS) attacks, and unauthorized access attempts. The dataset utilized in this study was derived from a realistic and layered IoT/IIoT traffic simulation environment that captures a comprehensive array of both benign and malicious network behaviors. It comprises 1,909,671 rows and 63 initial columns, where each row represents a distinct network event, transaction, or communication instance. The dataset spans multiple layers of the OSI model and includes metadata and protocol-specific fields commonly observed in IoT and industrial control systems, making it highly suitable for supervised machine learning tasks such as intrusion detection.

Feature selection, following Table 1 represents the select features out of 97.

Protocol Layer	Feature Name
IP Layer	ip.src_host
IP Layer	ip.dst_host
Temporal	frame.time
Network (ARP)	arp.opcode
Network (ARP)	arp.hw.size
Network (ICMP)	icmp.checksum
Network (ICMP)	icmp.seq_le
Network (ICMP)	icmp.transmit_timestamp
Application	mqtt.proto_len
Application	mqtt.topic
Application	mqtt.ver
Application	mbtcp.len
Application	mbtcp.trans_id
Target Variable	Attack_type
Target Variable	Attack_label

Table 1: List of features

Model Selection, in this study, we evaluated a diverse suite of machine learning classifiers to address the challenge of behavior-based anomaly detection in an IoT/IIoT environment. The selected models include both baseline and advanced models to assess effectiveness across various attack types:

Logistic Regression (LR) is a linear baseline model with poor minority class detection, while the non-linear and interpretable Decision Tree (DT) also struggled with these classes. The Random Forest (RF), an ensemble using bagging, achieved the best overall precision, recall, and F1-scores. In contrast, the boosting technique AdaBoost performed poorly with class imbalance, leading to high false positives. The computationally efficient probabilistic model Naive Bayes (NB) had the lowest overall performance. Both XGBoost, an optimized gradient boosting method, and CatBoost, a gradient boosting variant for categorical data, demonstrated high accuracy and strong class-wise performance, though

slightly behind RF and XGBoost overall.

Correlation & Variance Analysis, the correlation heatmap reveals key feature redundancies. Strong correlations were found between arp.opcode and arp.hw.size (r = 0.94), mqtt.proto_len and mqtt.ver (r ≈ 1.0), and mbtcp.len and mbtcp.trans_id (r = 0.92), indicating overlapping information. Moderate correlation between icmp.checksum and icmp.seq_le (r = 0.70) also suggests redundancy. Exclusion of one feature from each pair can reduce complexity and enhance model efficiency. Figure 1 represents a correlation heatmap between selected features.

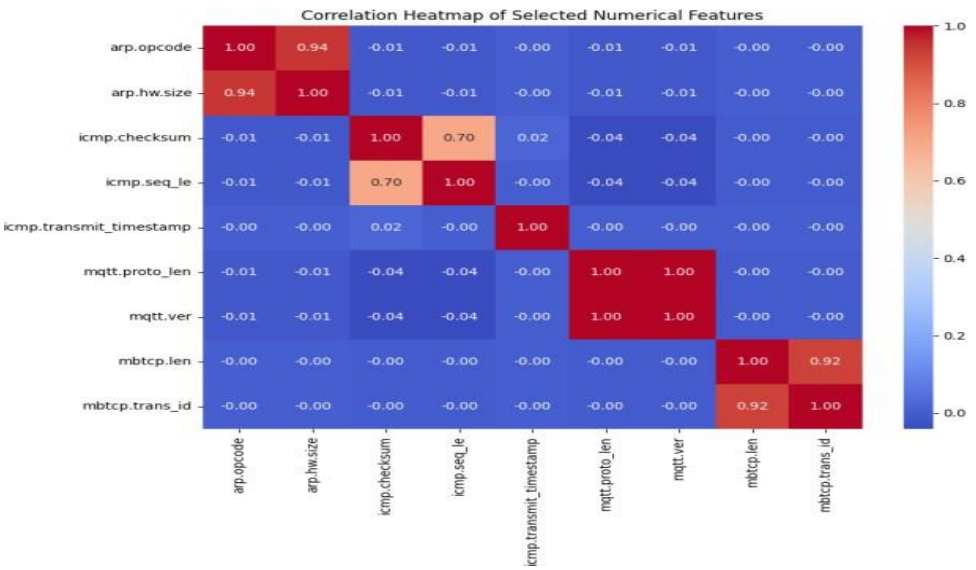


Figure 1: Correlation heatmap

4. Handling Class Imbalance

Since following Figure 2, Shows data is more biased towards Normal class attack type, Handling class balance for identifying rare class attack types was needed. This handling of class Imbalance Boosted Minority Detection Class imbalance in the dataset was a critical challenge.

The "Normal" class dominated the dataset with over 1.3 million entries, while rare classes like XSS, Fingerprinting, and SQL_Injection had under 4,000 entries. Techniques like SMOTE and Tomek Links helped oversample minority classes or remove overlapping samples, respectively, thus improving performance on rare classes.

- XGBoost and Random Forest retained high F1-scores even on minority classes.

- AdaBoost and Logistic Regression failed to generalize under severe class imbalance, exhibiting zero recall for some classes like Fingerprinting.
- Cost-sensitive training allowed models to assign more weight to underrepresented classes, improving detection without synthetic data.

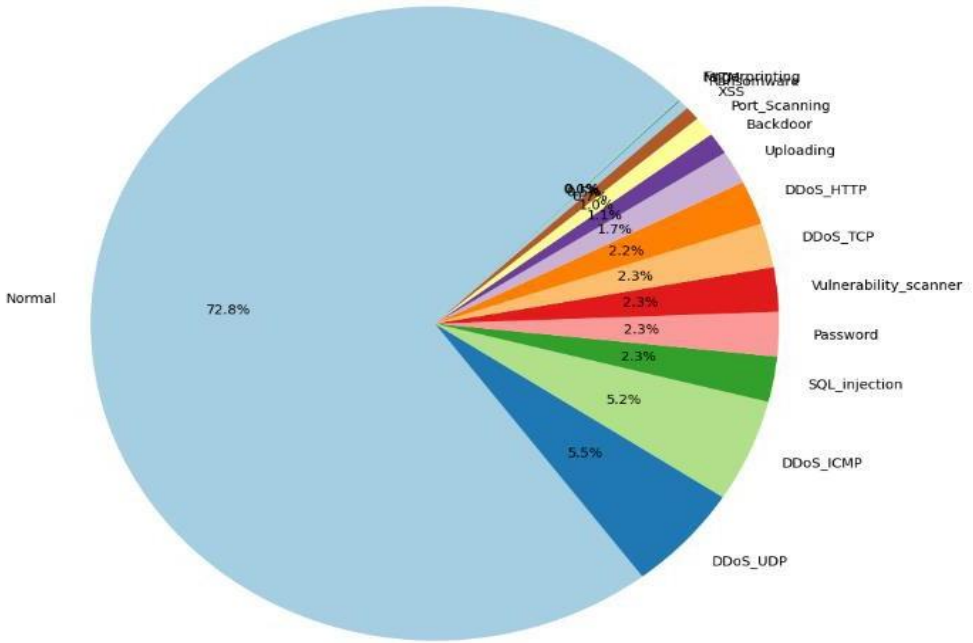


Figure 2: Distribution of Attack Types

5. ML Model Prediction

The Initial Model trained with raw dataset later followed by resampling. As per Fig-3 XGBoost achieved the highest accuracy (0.9846) and F1-score (0.9846). Random Forest followed closely (accuracy: 0.9762, F1- score: 0.9772). CatBoost performed well (accuracy: 0.9725, F1-score: 0.9695). Logistic Regression and Naïve Bayes had moderate accuracy (0.9310 and 0.9153, respectively). Decision Tree and AdaBoost performed the worst, with AdaBoost scoring the lowest (accuracy: 0.8401, F1-score: 0.8072).

 ISSN 1681 8997

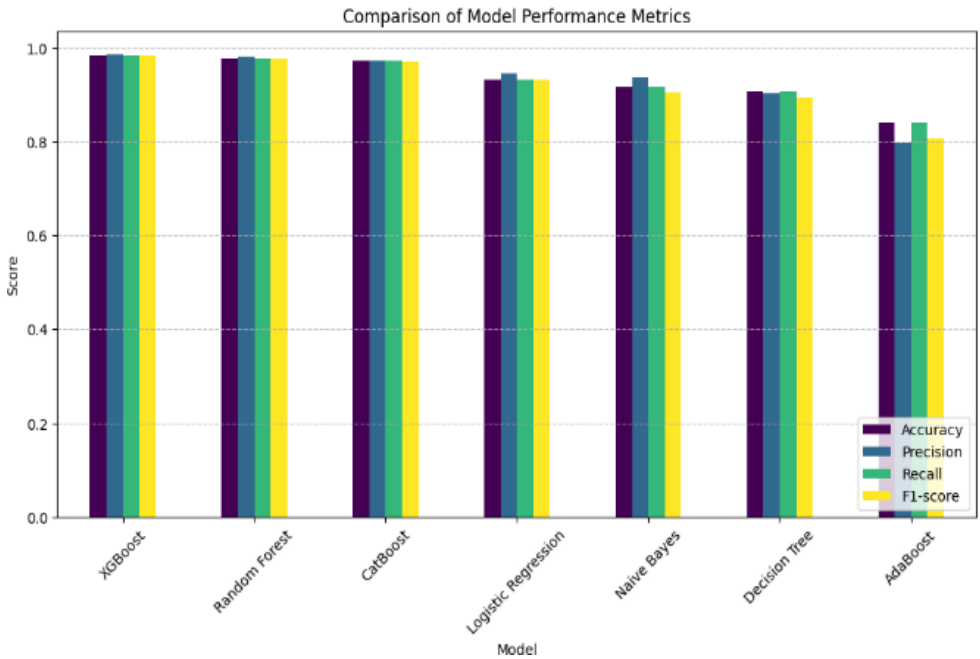


Figure 3: ML Model Comparison Raw Dataset

Since we have observed class Imbalance and biasness towards Normal Attack Type Label, We considered ML Model fit, train and predict using resampling techniques. Our experiments evaluated seven machine learning classifiers such as Logistic Regression, Naïve Bayes, Decision Tree, AdaBoost, Random Forest, XGBoost, and CatBoost. Each model was tested under various resampling strategies—SMOTE, Random Undersampling, Tomek Links, and Cost-Sensitive Learning—to mitigate class imbalance. Following Table 2 best sampling techniques across model comparison. depicts the classification report for each original and different 4 sampled data.

Model	Accuracy	Precision	Recall	F1-Score	Best Resampling Technique
XGBoost	0.9846	0.95	0.92	0.93	Tomek Links
Random Forest	0.9813	0.93	0.91	0.92	Cost Sensitive
CatBoost	0.9738	0.86	0.91	0.87	Cost Sensitive
Logistic Regression	0.941	0.75	0.71	0.71	Cost Sensitive
Decision Tree	0.9211	0.59	0.62	0.57	Random Undersampling
AdaBoost	0.8794	0.38	0.51	0.41	SMOTE
Naïve Bayes	0.9326	0.72	0.65	0.61	Random Undersampling

Table 2: Best Resampling Technique Above is a summary of the top findings

Ensemble Models Outperform Traditional Learners

XGBoost achieved the highest performance overall, with an accuracy of 98.46%, precision of 0.95, recall of 0.92, and an F1-score of 0.93 under Tomek Links sampling. Random Forest was a close second, achieving accuracy of 98.13% and balanced precision-recall under cost-sensitive learning. CatBoost also performed well, with an F1-score of 0.87 using cost-sensitive methods.

6. Feature Impact and Correlation

As shown in Figure 3 (correlation heatmap), strongly correlated features were identified and selectively pruned. For instance:

- icmp.checksum and icmp.seq_le were highly correlated, but both were retained due to their distinct temporal behavior patterns during DDoS attacks.
- Features such as arp.opcode and icmp.transmit_timestamp, which lacked variation across classes, were deprioritized or removed.

This correlation-aware feature selection helped models like XGBoost and CatBoost achieve higher performance without inflating model complexity or runtime.

Comparison to Literature our findings are consistent with previous studies that advocate ensemble learning for IDS applications. According to Aburomman & Reaz (2017) emphasized that combining classifiers reduces bias and variance, Das et al. (2022) confirmed that XGBoost and Random Forest offer superior performance on network traffic data and Alotaibi & Ilyas (2023) demonstrated the adaptability of ensemble methods for real-time IoT security, especially when integrated with lightweight deployment techniques.

7. Deployment Implications

The success of models like XGBoost and Random Forest demonstrates that behavior-based anomaly detection is highly feasible even in edge environments, provided preprocessing and class imbalance are well-managed. Moreover, the models' ability to generalize across varied attack types makes them suitable for real-time deployment in smart cities, healthcare, and industrial IoT systems.

8. Conclusion and Future Work

This research examined the effectiveness of various supervised machine learning algorithms in identifying behavioral anomalies within IoT network environments, with a focus on tree-based ensemble methods. Using the Edge-IIoTset dataset, which captures realistic IoT and IIoT traffic, we explored the application of feature selection, correlation analysis, and class imbalance correction techniques to improve model performance.

Among the tested models, XGBoost, Random Forest, and CatBoost consistently outperformed traditional models such as Logistic Regression, Naïve Bayes, and AdaBoost. XGBoost emerged as the top performer, achieving an accuracy of 98.46% and F1-score of 0.93 with TOMEK links and SMOTE, proving its robustness in handling imbalanced, high-dimensional IoT traffic data. Random Forest closely followed, showing excellent balance between precision and recall across multiple resampling methods. These results align with findings by Dhaliwal et al. (2018), who confirmed XGBoost's scalability for IoT anomaly detection, and by Alotaibi and Ilyas (2023), who validated the efficiency of ensemble models in securing real-time IoT operations.

The results also demonstrated that class imbalance remains a significant challenge. While cost-sensitive learning and hybrid resampling techniques like SMOTE and TOMEK links improved performance, rare attacks such as XSS, SQL Injection, Fingerprinting, and Password attacks still experienced lower recall and precision. Ensemble learners, particularly those using boosting strategies, proved more resilient under these constraints compared to single estimators and probabilistic models (Aburomman & Reaz, 2017; Das et al., 2022).

References

Aburomman, A. A., & Reaz, M. B. I. (2017). A survey of intrusion detection systems based on ensemble and hybrid classifiers. Computers & Security, 65, 135-152.

Al-A'araji, N. H., Al-Mamory, S. O., & Al-Shakarchi, A. H. (2021). Classification and clustering based ensemble techniques for intrusion detection systems: A survey. Journal of Physics: Conference Series, 1818(1), 012106.

Al-Omari, M., et al. (2021). An intelligent tree-based intrusion detection model for cybersecurity. Journal of Network and Systems Management, 29(2), 1-18.

Alotaibi, Y., & Ilyas, M. (2023). Ensemble-learning framework for intrusion detection to enhance Internet of Things' devices security. Sensors, 23(12), 5568.

Arisdakessian, S., et al. (2022). A survey on IoT intrusion detection: Federated learning, game theory, social psychology, and explainable AI as future directions. IEEE IoT Journal.

Balyan, A. K., et al. (2022). A hybrid intrusion detection model using EGA-PSO and improved random forest method. Sensors, 22(16), 5986.

Buczak, A. L., & Guven, E. (2015). A survey of data mining and machine learning methods for cybersecurity intrusion detection. IEEE Communications Surveys & Tutorials, 18(2), 1153- 1176.

Das, S., et al. (2022). Network intrusion detection and comparative analysis using

ensemble machine learning and feature selection. IEEE Transactions on Network and Service Management, 19(4), 4821-4833.

Dhaliwal, S. S., et al. (2018). Effective intrusion detection system using XGBoost. Information, 9(7), 149.

Dorogush, A. V., et al. (2018). CatBoost: Gradient boosting with categorical features support. arXiv preprint arXiv:1810.11363.

Mukherjee, B., et al. (1994). Network intrusion detection. IEEE Network, 8(3), 26-41.

Northcutt, S., & Novak, J. (2002). Network Intrusion Detection. Sams Publishing.

Panigrahi, R., et al. (2022). Intrusion detection in cyber–physical environment using hybrid Naïve Bayes—Decision Table and multi-objective evolutionary feature selection. Computer Communications, 188, 133–144.

Vinayakumar, R., et al. (2019). Deep learning-based intrusion detection for IoT networks. Computer Communications, 143, 304–313.

Zhang, H., et al. (2021). Multi-dimensional feature fusion and stacking ensemble mechanism for network intrusion detection. Future Generation Computer Systems, 122, 130-143.

Mohamed Amine Ferrag, Othmane Friha, Djallel Hamouda, Leandros Maglaras, Helge Janicke (2022). Edge-IIoTset: A New Comprehensive Realistic Cyber Security Dataset of IoT and IIoT Applications: Centralized and Federated Learning. IEEE Dataport. https://dx.doi.org/10.21227/mbc1-1h68

7.4 Real-Time Video Frame Similarity Measurement on Raspberry Pi Using SSIM

Authors: Jackson L Osborne-Coy *(State University of New York at Fredonia)*
Shahin Mehdipour Ataee *(State University of New York at Fredonia)*

Monitoring dynamic scenes efficiently is crucial in various applications, such as security, automation, and research on periodic scene detection. In this project, we developed an application that runs on a Raspberry Pi computer, utilizing an HD camera to continuously monitor a scene by comparing each video frame to a reference frame. The application employs the Structural Similarity Index (SSIM) to measure similarity, logging similarity scores in real-time. Implemented in Python, this system is part of a larger research initiative aimed at detecting periodic patterns in video streams or recorded footage.

The motivation for this work stems from the need to create an embedded, cost-effective, and efficient solution for video analysis. Existing methods for frame similarity measurement often rely on computationally expensive algorithms or require high-performance hardware, making them unsuitable for embedded systems. Our approach seeks to optimize the SSIM computation so that it runs efficiently on low-power devices

like the Raspberry Pi. This optimization is essential for real-time processing, particularly when transitioning to even more resource-constrained hardware such as the Raspberry Pi Pico microcontroller.

Our results indicate that SSIM provides an adequate level of accuracy for similarity measurement within the constraints of the Raspberry Pi. The performance, while acceptable at this stage, suggests potential for further optimizations to improve computational efficiency. Future work will focus on refining our implementation to operate on less expensive platforms, specifically the Raspberry Pi Pico, ensuring that the system remains both accurate and energy-efficient. The insights gained from this research will contribute to the broader objective of automated periodic scene detection in video streams using embedded systems.

Track 8: Analytics in Business Education
Room: WC S204A (2:50 pm – 4:10 pm)
Moderator: Nazik Roufaiel *(SUNY-Empire State University)*

8.1 Empowering Education: the Transformative Role of Business Analytics in Decision-Making and Operational Efficiency

Author: Abdelghani Mehailia *(Yorkville University)*

Over the past few years, data analysis has become critical for enhancing educational practices and outcomes at all educational levels. Educational institutions can use the extensive amount of data produced by educators, students and academic systems to gain insights and improve decision-making. This enables educators to gain a deeper understanding of trends in student performance and behavior. It would also help them to develop learning strategies and predict student performances. Besides, data analytics (DA) plays an important role in developing and improving the content of courses delivered to students. In addition, data analytics tools will allow improve efficiency in the allocation of resources and assist educational institutions in developing strategies to increase the retention of students. The paper discusses several applications & success stories of data analytics in the education industry namely universities and online learning websites.

8.2 Experiences with Experiential Learning in a Graduate Business Analytics Program

Authors: Necip Doganaksoy *(Siena College)*
Travis Brodbeck *(SUNY Albany & Siena College)*

Experiential learning is increasingly embraced by business schools as a means of bridging classroom knowledge with practical applications. In the context of business analytics programs, it provides students with opportunities to tackle real-world problems by

translating them into solvable technical questions and effectively communicating their findings to stakeholders.

Given the relatively recent adoption of these programs, there is no unified or universally accepted approach to integrating experiential learning into business analytics curricula. However, some of the most common methods include (Ritter, Jones-Farmer, & Faltin, 2024):

- ""Live"" Project Classes: Engaging students in real-world projects with external partners.
- Case Study-Based Courses: Applying analytics techniques to structured, real-life scenarios.
- Projects Embedded Within Methodology Courses: Allowing students to practice techniques within the specific context of a course.

This presentation shares our early experiences incorporating experiential learning into our graduate business analytics (MSBA) curriculum. While we utilize all three methods outlined above, the cornerstone of our experiential learning approach is the capstone project.

In the capstone project, we adopt a live-project approach in collaboration with functional departments on our campus, offering students a hands-on, interdisciplinary experience. Additionally, we incorporate community-engaged projects, where students apply their data analytics skills to support local organizations.

We hope these experiences will foster a forum for sharing insights with similar programs aiming to enhance experiential learning in their curricula.

8.3 Effective Planning and Implementation in Academic Institutions Through Business Analytics and Project Management

Author: Eren Akdur *(Ithaca College)*

The effective implementation of strategic plans in academic institutions requires a seamless integration of project management methodologies and data-driven business analytics. Together, these approaches help align institutional goals with practical, actionable strategies, ensuring progress and adaptability.

Achieving and maintaining accreditation further underscores the need for structured planning and implementation, emphasizing stakeholder engagement and continuous improvement. In this context, Kanban boards, widely recognized as an effective task management tool, emerge as a pivotal resource. When integrated with technologies like Power Automate and Power BI, Kanban boards enhance workflow management, facilitate better communication, and improve organizational agility.

These technologies enable the creation of a Management Information System (MIS) with dashboards that offers structured planning, real-time insights, progress tracking, proactive risk mitigation, and continuous optimization. Such an approach ensures that institutions remain agile, accountable, and aligned with their strategic objectives.

This presentation delves into the role of business analytics and project management as the driving forces behind using Kanban boards for strategic plan implementation. By leveraging task visualization, risk management, and evidence-based archives, the framework supports data-informed decision-making, streamlines operations, and facilitates the achievement of institutional goals.

Track 9: Applied Economics

Room: WC S204B (2:50 pm – 4:10 pm)
Moderator: Ai Ren *(SUNY New Paltz)*

9.1 Bridging Leadership Gaps: Unveiling a New Global North-South Assessment Tool

Authors: Richard J Muszynski III *(Wilkes University)*
Mona Pearl *(Wilkes University)*
Bruno Sergi *(Harvard University)*

*s*The disparity between the Global South and Global North has gained prominence in classifying countries according to economic, social, and governance characteristics, rather than historical links. Historically, the Global North has long been regarded as the dominant group of nations, far outperforming Global South countries; however, several Global South nations now outperform the Global North nations in multiple measures with increasing economic and political strengths that draw attention to their unique leadership challenges and opportunities. Traditional measures like GDP per capita often overlook the intricate leadership dynamics essential for understanding developments in both regions. This study introduces a novel approach to evaluating global leadership transitions by expanding on the Global South Leadership (GSL) Index. The developed GSL Index equips policymakers, business professionals, and leadership with deeper insights into governance styles and leadership dynamics, and this comprehensive approach encompasses economic performance, social mobility, and governance structures to accurately assess a nation's performance across various indicators. Scores are obtained from 18 indices across 192 nations; each index measures a unique characterization of the country. All scores are converted to a 0 to 100 percent index scale, then averaged across all indices for each nation to create the expanded GSL Index. Relationships, patterns and trends are identified through scatter plots between Global North and South countries and among the regions within. The study extrapolates why some Global South countries are outperforming and

others are underperforming, and the ones that have overcome obstacles and become role models for other Global South and North countries.

9.2 Share-the-Road: Exploring the Relationship Between Bicycle-Sharing and Ridesharing Platforms

Author: Ayush Sengupta *(Alfred University)*

In this paper, we explore the relationship between bicycle-sharing platforms such as Citi Bike and ridesharing platforms such as Lyft. The two groups of platforms are both relatively new business models in the transportation service industry and are driven by innovative online technologies. It is debated in academic research and in industry whether the two types of platforms have a complementary or substitution effect on each other. Some argue that the bicycle-sharing platforms may help alleviate the last-mile barrier in urban transportation and benefit ridesharing platforms by providing additional customers from their untapped market and by improving their financial situation. Others suggest that an adverse impact of bicycle-sharing platforms on ridesharing platforms may arise through a reduction in the ridership of the latter, especially over short trips, due to a substitution effect. We focus on answering this question empirically by examining the impact of the entry of Citi Bike on the ridership of all ridesharing platforms, in the context of New York City. We analyze data on ridesharing-platform ridership, measured by the number of their pick-ups. Through our analyses, we examine how Citi Bike's entry influences ridesharing's popularity in a large metropolis and how the impact varies across trips with different travel distances. We also analyze how Citi Bike's entry influences pooled ridesharing trips and ridesharing trips during rush hours. Our findings offer insights for bicycle-sharing and ridesharing-platform companies to make business decisions and for policymakers to devise policies to improve traffic congestion and reduce vehicular emissions.

9.3 A Functional Covid Recovery: Applying Functional Data Analysis to the Post-Covid Economic Recovery

Authors: Justin Petrovich *(Saint Vincent College)*
Zachary G Davis *(Saint Vincent College)*

The economic recovery after the response to the COVID-19 pandemic varied considerably across counties. Using data from Opportunity Insights, we employ functional data techniques to model the recovery of employment, consumer spending, and small businesses among counties as a function of time. Specifically, we regress the functional recovery data on pre-2020 American Community Survey data to estimate the time-varying effects of pre-pandemic county demographics and industry distribution on the course of recovery. From this we learn, for instance, that counties with a higher 2019 share in six different industries---agriculture, extraction, construction, transportation, finance, and

accommodations--saw positive overall recovery in employment, with an increasing rate of recovery. Other industries had more mixed, or even negative effects on the change in employment. We also use cluster analysis on the functional recovery data to identify four different recovery patterns across counties. Counties in these four clusters are distinguished primarily by a) the magnitude of their initial declines in employment and b) the rate of increase in employment following that initial decline. The results from these analyses can be used to help municipalities plan and anticipate the economic impact of future similar disasters.

Track 10: Human Resources

Room: WC S204C (2:50 pm – 4:10 pm)
Moderator: Mary Han *(Toronto Metropolitan University)*

10.1 The Effects of Human Resource Management on Performance: the Mediating Role of Artificial Intelligence

Authors: Alexandra Cuero *(Alfred University)*
Halil Zaim *(Alfred University)*

This study explores the impact of Human Resource Management (HRM) practices on organizational performance and examines the mediating role of Artificial Intelligence (AI). A field study was conducted in New York State, involving 614 participants employed in various companies across different roles. Using the PROCESS Model 4 in SPSS, we analyzed data from 559 valid responses. The results reveal a significant direct effect of HRM on performance ($\beta = 0.5239$, $p < 0.001$), indicating that well-implemented HRM practices enhance organizational outcomes. Furthermore, HRM positively influences AI adoption ($\beta = 0.3458$, $p < 0.001$), which in turn contributes to improved performance ($\beta = 0.2069$, $p < 0.001$). The mediation analysis confirms that AI partially mediates the relationship between HRM and performance, with a significant indirect effect ($\beta = 0.0715$, 95% CI [0.0406, 0.1065]). These findings underscore the growing role of AI as a facilitator in HRM-driven performance improvements, highlighting the need for organizations to integrate AI strategically within HRM frameworks.

10.2 Assessing Employee Emotions at Work to Predict Organizational Citizenship Behavior

Author: Lipika Arif *(SUNY-Fredonia)*

Nowadays, an organization faces the great challenge of maintaining competitive advantages in a rapidly changing environment. The influence of globalization, changing political and geopolitical relationships, economic restructuring, and transforming communication and information technologies are shaping the modern business setting

(McGuire, Cross, & O'Donnell, 2005). Many organizations recognize the importance of employees 'positive attitudes and behavior, such as organizational citizenship behavior (OCB) for beneficial organizational outcomes. As a discretionary behavior, OCB plays a positive role in supporting organizational benefits in terms of productivity, efficiency, and employee performance evaluations and promotions (Podsakoff et al., 2009). Although studies show the positive association of OCB with beneficial organizational outcomes, limited research focuses on the psychological aspects of employees for enhancing OCB. This study aims to investigate the role of positive affect in predicting OCB following affective organizational commitment as a mediator and leader-member exchange (LMX) relationship as a moderator. Time-lagged data were collected from 243 (n=243) teachers and staff from elementary, middle, high schools, and other schools in a public school system to test the hypothesized relationship. The result shows that there are significant positive relationships between positive affect and affective organizational commitment and affective organizational commitment and organizational citizenship behavior. This study also proposed that the relationship between affective organizational commitment and OCB is further stronger with the condition of LMX. This study offers meaningful theoretical and practical implications along with future research.

Track 11: Emerging Scholars III: Data Science and Public Safety

Room: WC S204D (2:50 pm – 4:10 pm)
Moderator: Lisa Walters *(State University of New York at Fredonia)*

11.1 Data Analysis of Crime Incidents in Buffalo, NY

Authors: Andriy Martynyshyn *(Buffalo State University)*
Reneta Barneva *(Fredonia State University of New York)*

Finding crime patterns and predicting crime are important for crime prevention and resource allocation. This helps law enforcement agencies deploying forces in the high-risk areas, potentially preventing crimes, which results in safer communities.

In this work, we explore the open data source https://data.buffalony.gov/Public-Safety/Crime-Incidents/d6g9-xbgu/about_data provided by Buffalo Police Department to better understand the crime patterns in Buffalo, NY. The database contains detailed records of crime incidents, including attributes such as incident type, date, time, location, and neighborhood, enabling a comprehensive analysis of crime distribution and trends.

We analyze crime incidents from multiple years (2022, 2023, and 2024) to uncover patterns and trends. The study focuses on identifying clusters across several dimensions.

The research involves three key phases:

1. Data Preprocessing: XML or JSON crime data is parsed to extract relevant fields, the coordinates are normalized, and the missing data is handled. Features such as ""time period"" and ""season"" are derived to enable a multidimensional analysis.

2. Clustering Techniques: Using a combination of K-Means, Hierarchical Clustering, and DBSCAN clustering techniques, the analysis reveals spatial, temporal, and seasonal patterns.

3. Results Analysis: The findings include:

- High-crime areas with recurring seasonal and temporal patterns, such as increased assaults in the summer evenings.
- Variations in crime types and frequencies across neighborhoods and months, emphasizing the need for targeted law enforcement.

The outcome of this project is a detailed understanding and visualization of the crime patterns in Buffalo, offering actionable insights to stakeholders.

11.2 Spatial Analysis Beyond Numbers: Demystifying Geographically Weighted Regression (GWR) and Multiscale Geographically Weighted Regression (MGWR) for Policy and Research

Authors: Krupa B Shah *(Ontario Tech University)*
Gabby Resch *(Ontario Tech University)*

In today's day and age, where the world is more connected than ever, it is easy to overlook the influence of local nuances in geospatial data. A one size fits all model can miss critical details hidden in spatial variability. Therefore, it is important to understand the weight of each independent variable affected by its geographical significance, especially in decision making for policy makers and researchers. The goal is to analyze spatial heterogeneity by tailoring the regression coefficients to geographic contexts. Geographically Weighted Regression (GWR) and its evolution, Multiscale Geographically Weighted Regression (MGWR), enable professionals to move beyond static, global models and embrace region-specific dynamics, ensuring more informed, actionable insights. This paper simplifies and demystifies the mechanics of GWR and MGWR, contrasting their methodological underpinnings and practical applications. By bridging theory with practice, to enhance the relevance to diverse fields like immigration policies and migration studies, urban planning, environmental management, and public policy.

11.3 Predictive Analytics for Accident Prevention in Autonomous Cars

Author: Sanika T Desai *(Adelphi University)*

Abstract

Traffic accidents, predominantly caused by human error, remain a global concern despite advancements in autonomous vehicles (AVs). While AVs report fewer crashes per mile than human-driven vehicles, limited datasets and narrow analytical approaches have left key risk factors underexplored. This research utilizes a comprehensive dataset (2014-May 2024) to build predictive models aimed at accident prevention. Multiple machine learning algorithms were applied, with Random Forest achieving the highest accuracy and lowest false negatives. The study identifies vehicle damage severity, collision type, and AV manufacturer as significant predictors of crash-related injuries. These findings support safer AV development and informed policy-making aligned with global safety goals.

Keywords: autonomous vehicles, accident prediction, machine learning, random forest, safety, crash analytics, transportation policy

1. Introduction

Vehicle accidents account for over 1.35 million deaths annually, with the United States experiencing one of the highest accident rates worldwide. The United Nations has set Sustainable Development Goal 3 (SDG-3) to reduce these fatalities by 50% by 2030. Autonomous vehicles promise to drastically reduce human-error-related accidents, yet research is often constrained by data limitations and lack of system-wide insight. This study bridges that gap using a diverse and rich dataset encompassing weather, manufacturer, damage extent, and sensor information. Through comprehensive predictive modeling, the goal is to identify major factors contributing to accidents and inform AV design and safety regulations.

2. Methodology

This study follows a structured data science pipeline:

Data Collection: Sourced from a variety of AV manufacturers across multiple states covering 2014-2024.

Data Cleaning: Removed duplicates and handled missing values with mean/mode substitutions.

Exploratory Data Analysis (EDA): Identified patterns like the correlation between increased speed and accident severity.

Balancing: Used SMOTE and class weighting to handle class imbalance.

Model Selection: Tested logistic regression, decision trees, random forest, gradient boosting, and ensemble methods.

Training & Evaluation: Models were validated with 5-fold cross-validation, optimizing hyperparameters using GridSearchCV.

Feature Engineering: Feature importance scores were computed to identify the most influential factors such as speed, traffic, and braking behavior.

3. Key Results

Performance evaluation used accuracy, recall, precision, F1 score, and ROC-AUC.

- No Alert: Precision 0.95, Recall 0.94, F1 Score 0.94
- Warning Alert: Precision 0.92, Recall 0.90, F1 Score 0.91
- Critical Alert: Precision 0.93, Recall 0.92, F1 Score 0.92

The model was especially strong at minimizing false negatives for critical alerts. Feature importance from the Random Forest model indicated traffic density and speed as dominant factors. Secondary features like brake temperature, power usage, and steering angle contributed meaningfully, while categorical variables like radar and lidar status were less predictive. This implies a shift in AV system design emphasis toward dynamic real-time variables.

4. Contributions & Implications

Theoretical: Demonstrates a scalable methodology using balanced datasets and ensemble machine learning, advancing the literature on AV crash prediction.

Transportation Professional: Insights into high-risk conditions can help design better road infrastructure and signage.

AV Manufacturers: Findings suggest improvements to ADAS systems, focusing on real-time variables like speed regulation.

Policymakers: Supports legislation for V2X (Vehicle-to-Everything) adoption and stricter AV testing across varied environments.

5. Conclusion

This work offers a data-driven framework for accident prevention in autonomous vehicles. Through model performance evaluation, speed, traffic density, and damage severity emerged as primary crash predictors. The Random Forest model not only demonstrated strong classification ability but also provided insight into operational safety factors. These findings will aid manufacturers, urban planners, and policymakers in proactively reducing

AV-related injuries. Future research can integrate dynamic data sources like real-time weather and live sensor feeds to improve prediction accuracy even further.

References

1. Hassan, M. M., et al. (2018). A survey on intelligent transportation systems. IEEE Transactions.
2. Chawla, N. V., et al. (2002). SMOTE: Synthetic minority over-sampling technique. JAIR.
3. Mandala, V., & Kommisetty, P. D. N. K. (2023). Performance evaluation metrics in accident prediction. Journal of Applied Analytics.
4. Sun, W., et al. (2018). An ensemble deep learning approach for short-term traffic flow forecasting. IEEE Transactions.
5. Xu, R., Li, Z., & He, H. (2018). IoT for predictive maintenance. IEEE Access.

11.4 Retaining Value: Segment-Based Predictive Models for Customer Churn in Banking

Authors: Priyal Rawat *(Northwood University)*
Abhishek Rajeev Sharma *(Northwood University)*
Itauma Itauma *(Northwood University)*

Customer churn is an everyday problem for the banking sector, particularly in a rapidly evolving marketplace where there is greater customer expectations and an increase in digital competition. This study establishes a machine learning framework for predicting customer churn on a segment basis for High Net Worth (HNW) and Regular customers. The study used models based on Logistic Regression, Decision Tree, Random Forest and Gradient Boosting, focusing on interpretability, alignment to the business, and ability to be scalable. A cost-sensitive learning approach was employed to deal with the class imbalance and a churn risk scoring system was introduced that enabled model results to be converted into prioritized intervention strategies for banks to retain customers. Feature importance analysis indicated that the drivers of churn were different between segments, validating the modeling (HNW and Regular) on a segment basis. The framework developed here compromises a balance between predictive performance and explainability and is feasibly scalable to implement in practice for use in CRM systems. This study expands the literature addressing criteria for predicting customer churn in the banking context to include a replicable, interpretable, and data-driven approach to understand customer churn using a proactive approach and further, the approach taken considers personal retention strategies in the modern banking for customer relationship management.

Track 12: Advanced Business Analytics

Room: WC S204A (9:30 am - 10:50 am)
Moderator: Linda Hall *(State University of New York at Fredonia)*

12.1 Who is the Boss? from Agency Theory to Autonomous AI

Authors: Mary Han *(Toronto Metropolitan University)*

Literature on agency theory has documented monitoring, incentive and rewards to enhance agents' performance, but principal agent complexity persists. The rise of artificial intelligence seems to have added to this dynamic with much lower cost, higher accuracy and speed (which is good) but what if AI 'wants' to be the boss. While agents are rewarded for their managerial role on behalf of their principals, evidence exists that some even manipulate with fraud. The advancement of AI seems to have gradually shifted the agent's role as well as their ability to manipulate. AI capability has shifted from requiring data to be input for learning; to generative AI, meaning AI can generate outputs from data; to collaborative AI, where AI makes decisions alongside management; to autonomous AI where AI can decide and act without human supervision, input or intervention. In such cases, where do agents stand? While there are economic advantages, but who is the boss when AI can decide without the principal? We review traditional agency theory and situate extant AI capability to theoretically hypothesize how agency theory can explain the new phenomenon of principal agent dynamics in the new AI era. We conclude with theoretical and practical implications.

12.2 Data Analytics (DA) and Artificial Intelligence's (AI's) Impact on Jobs

Author: Joseph Porter, Jr *(Nazareth University, USA)*

In this session, I will share how Data Analytics (DA) and Artificial Intelligence (AI) are being used to transform jobs in industries like automotive, finance, and healthcare. During the presentation, I plan to discuss:

1. Productivity Trends: I will highlight data from multiple sources including the Federal Reserve to shows that productivity in companies have been increasing since World War II.
2. Automotive Industry: I will highlight the impact that DA and AI are having on jobs in the automotive industry: Robotics operating 24/7, self-driving trucks, etc.
3. Financial Industry: I will highlight the impact that DA and AI are having on jobs in the financial industry: data from AI Facial Recognition programs, mortgage approval, fraud detection on credit cards, etc.
4. Medical Industry: I will highlight the impact that DA and AI are having on jobs in the medical industry: detecting cancer using data from Mammograms, MRI, etc.

Optional Topics (If time permits):

• Education Industry: I would like to address the need for educators at the college/university level to teach how technologies will affect careers.

By discussion components above, I aim to equip educators with insights and methodologies to enhance analytics education."

12.3 Simplifying Survey Reporting: an R Solution for Calculating and Visualizing Key Metrics

Authors: Travis Brodbeck *(SUNY Albany & Siena College)*
Mohammad Wasim Shekh *(Siena College)*

The American Association for Public Opinion Research's (AAPOR) Transparency Initiative (TI) highlights the importance of reporting response rates for evaluating the quality of survey research studies. The process to calculate a survey's response rates, especially in research studies using multiple data collection methods, is labor intensive and at-risk of inconsistencies when using the industry provided Excel spreadsheet template. We developed an R-script to streamline the process of calculating response rates, for one or multiple surveys.

Using the AAPOR Response Rate Calculator V4.1, a Business Analytics graduate student and a survey researcher at the Siena College Research Institute (SCRI) collaborated to enhance the process of reporting various survey research metrics. Using a data file of case dispositions, the R script calculated and visualized contact, refusal, cooperation, and response rates, for one or multiple surveys simultaneously. To ensure accuracy and replicability, this script was tested in parallel with the industry provided Excel template. The R script's output matched the calculations from the Excel template across various survey disposition datasets. While differences in disposition code standards between pollsters may necessitate a translation map, this limitation is manageable and does not diminish the tool's overall utility.

This open-source R script can assist in reducing barriers to improve transparency and efficiency in the survey research industry. By enabling pollsters and researchers to automate response rate calculations, response rates can be calculated retroactively and for projects in the future enhancing the industry's ability to identify trends in response rates that ultimately affect survey representativeness.

12.4 Assessing the Appropriateness of Videometrics in Strategy Research on Firm Executives

Author: Andre Havrylyshyn *(Binghamton University)*

I am currently working on a paper which will test the validity of the videometrics approach in strategic leadership/ governance research. Scholars interested in how traits of

CEOs influence firm-level outcomes have a major practical challenge: how do we get the access to high-level CEOs needed to do such research? CEOs of major firms are extremely busy and unable to fill out the surveys traditionally used in psychometrics research in organizational behavior research. To address this, some scholars have utilized a videometrics approach, where videos of prolific CEOs conducting interviews with news reporters, are evaluated by third-party raters. While such an approach has been used in empirical papers in recent years, serious questions remain about whether this approach is appropriate and when. Are the behaviors of the CEOs in a public setting appropriate ways to assess their personality traits like narcissism, or their leadership styles when interacting with their employees within their firm? To address this, we utilize a unique survey we have obtained where CHROs of dozens of SP500 firms gave us evaluations of the personality traits and leadership styles of their CEOs, as they behave 'behind closed doors'. We have hired raters to then evaluate those same CEOs for the same traits, using a conventional videometrics approach. From here, we will compare where the videometrics approach aligns with the reports of colleagues of the CEOs, generating practically useful insights for scholars in the space looking for convenient ways to conduct nevertheless scientifically rigorous research.

Track 13: Finance, Marketing and Retail Analytics

Room: WC S204B (9:30 am - 10:50 am)

Moderator: Shazad Mohammed *(State University of New York at Fredonia)*

13.1 Do Banks' Self-Reported Stress Test Disclosures Convey Critical Financial Information? A Textual Sentiment Analysis

Author: Yi Zheng *(SUNY New Paltz)*

In this study, we conduct a textual sentiment analysis of hand-collected U.S. banks' stress test disclosure reports. Our findings reveal that sentiment factors, such as positive wording and strong modals, convey significant financial information about banks. Specifically, both factors are positively associated with banks' projected minimum and end Tier 1 capital ratios, as well as their equity-to-assets ratio. Additionally, positive wording sentiment reflects banks' financial standing, as indicated by z-scores. These results remain consistent across various robustness checks and endogeneity mitigation analyses. Furthermore, we find that textual sentiment provides informative insights into banks' performance and operations, including loan loss rate, net pretax income rate, dividend payout ratio, and cost of debt. Overall, the textual sentiment of banks' self-reported stress test disclosures is informative regarding their capital adequacy, financial strength, and operations.

13.2 The Discrepancy between Green Reputations and Environmental Performance

Author: Joon Yong Seo *(SUNY Brockport)*

As an increasing number of consumers display heightened concern over sustainability issues and firms' environmental impacts, companies are turning their attention toward environmental sustainability and looking forward to taking advantage of the green trend in favor of environmentally superior products. As a result, claims of green businesses and products have become universal. At the same time, consumer skepticism toward firms taking opportunistic advantage of the sustainability movement has been growing (Pomering and Johnson, 2009). In an effort to build up positive corporate image, firms sometimes selectively disclose positive information about their environmental performance without fully revealing negative information (Lyon and Maxwell, 2011). Consequently, there is an increasing concern over greenwashing which refers to firms' deliberate misleading of consumers regarding the firm's environmental practices or the environmental benefits of a product or service (Parguel et al., 2011). From a green marketing perspective, consumer perceptions of greenwashing can erode the market demand of environment-friendly products by confusing consumers and making them unsure and reluctant to buy products with environmental claims (Pomering and Johnson, 2009).

Reflecting the growing public interest in various sustainability issues, corporate green branding has assumed increasing importance in the popular press. The recent release of reports by market researchers, such as Interbrand's Best Global Green Brands and the highly publicized Newsweek Green Rankings of the largest US corporations, has raised questions about whether they represent a sophisticated form of greenwashing. This is partly because consumer perceptions of the firm's greenness may not always line up with reality, that is, the firm's actual sustainability efforts and their consequences. Newsweek recently partnered with Landor Associates, a brand consulting and design firm, and Penn Schoen Berland, a research and communication firm, to survey consumer perceptions of green brands and firms' sustainability practices. The survey results spotlight a significant gap between consumer green brand recognition and the firms' actual environmental performance, using four brand categories of unsung heroes, free passers, losers, and winners (Landor Associates, 2012).

- *Unsung Heroes*: Brands with strong green practices, but little public awareness
- *Free Passers*: Brands with limited green practices, but strong brand halos that drive their green reputation
- *Losers*: Brands with limited green practices that are publicly recognized for these limitations
- *Winners*: Brands with strong green practices that are publicly acknowledged

This classification reveals that consumer perceptions of firms' greenness are often inconsistent with the corresponding firms' sustainability practices/outcomes, causing consumer confusion in the marketplace with regard to firms and products that make environmental claims. The lack of close connection between consumer perceptions and firm behaviors underscores the importance of aligning consumer knowledge with firms' sustainability initiatives, particularly when the firm's actual green performance outweighs consumer perceptions of the firm's greenness. It also calls for the need for additional research on what drives the gap between perceived and actual sustainability. The present research is an attempt to meet this need.

In the current research, we distinguish corporate green branding from the practice of eco-labeling. Corporate green branding focuses on the overall corporation brand's reputation for greenness with consumers, while eco-labeling is associated with specific products and services. The eco-labeling phenomenon has generated a rich and growing body of research; in contrast, green branding has received very limited academic attention except only few studies (e.g., Chen, 2010), a situation we are attempting to address in the present research. Specifically, the focus of our research is *the extent to which consumer perceptions of firms' greenness match the reality of corporate environmental performance*.

A better understanding of this issue will aid firms that are challenged with delivering both sustainable business practice and effective communication to stakeholders. For instance, studying this topic will help "unsung heroes" find out where and how to overcome the unfairly low public awareness of their superior sustainable practices. On the other hand, better insight about the disparity between green brand reputations and environmental performance may help "free passers" avoid the threat of being labeled as a greenwasher or facing lawsuits for engaging in false environmental claims by disclosing proper information about their products or practices. Overall, a deeper understanding of this topic will help firms build and maintain a green brand image which can be defined as "a set of perceptions of a brand in a consumer's mind that is linked to environmental commitments and environmental concerns" (Chen, 2010). Firms with green brand image will not only better meet customers' environmental needs, but also avoid the trouble of environmental protests (Chen, 2010), resulting in stronger customer relationships. In addition, environmentalists, investors, and policy makers will benefit from making more precise judgments of green brand reputations.

Although a number of consumers are willing to buy green products offered by environmentally responsible companies, many of them suspect that businesses' environmental claims are just a marketing strategy that they do not trust (Lyon and Maxwell, 2011). This implies that there are increasing needs of consumers, investors, and

policy makers for enhanced knowledge about green brand reputations and environmental claims. Having a better comprehension of the potential disparity between consumer perceptions of green brands and the firm's actual environmental performance will help these parties accurately assess whether a firm's environmental claim is a case of greenwashing or a true reflection of environmentally responsible business practices.

To address our key question about firms' perceived sustainability and real sustainability, we develop two models by drawing on concepts related to the natural resource-based view of the firm (Hart, 1995), environmental disclosure, and information asymmetry. Our first model investigates whether corporate green brand reputation as perceived by consumers is significantly related to objective measures of corporate environmental performance. Our second model follows up by exploring the gap between corporate green brand reputation and corporate environmental performance, testing whether the gap can be explained by variations in information asymmetry, corporate environmental disclosure, and "halo" effects from general corporate reputation. The results of these analyses reveal a positive relationship between green brand reputations and corporate environmental performance, mitigating the rapidly increasing public concerns about greenwashing to some degree. More important, we identify two factors that explain the gap between green brand reputations and corporate environmental performance, that is, the extent of firms' information asymmetry and the extent of firm-level environmental disclosure. We also find these two factors are significant predictors of "free passers" and "unsung heroes", respectively. In the following section, we present our models along with results and brief discussion. We conclude with contributions of our research, practical implications of our findings, and limitations

Model 1: Green Brand Reputation = f(Corporate Environmental Performance, controls for firm size, advertising intensity, environmental governance proxies, error)

The first empirical model investigates whether firms' green brand reputations (i.e., perceived sustainability) are systematically related to the corresponding firms' actual environmental efforts and outcomes (i.e., real sustainability), after appropriate controls. As key stakeholders, consumer appreciation of corporate sustainable practices and outcomes, most directly expressed as a willingness to prefer the products and services of greener firms, even if they command premium prices, is an important aim of firms striving to be environmentally responsible. This aim drives the firms' efforts to secure, configure, and effectively utilize resources and capabilities to achieve superior environmental performance (Bhattacharya, Korschun, and Sen, 2009; Du, Bhattacharya, and Sen, 2007; Maignan, Ferrell, and Ferrell, 2005). The successful deployment of these resources and capabilities is a major focus of the natural resource-based view of the firm (NRBV) (Hart, 1995; Hart and Dowell, 2011). Supporting the NRBV, Sharma and Vredenburg (1998)

find that firms that see environmental responsibility as fundamental to increasing shareholder value tend to take proactive environmental stances, while those that see environmental responsibility and shareholder value as separate or even conflicting goals take reactionary perspectives. These scholars further show that those firms with the proactive environmental standpoints see their environmental initiatives as requiring stakeholder integration. These findings suggest the possibility that firms with higher environmental performance will have more favorable green brand reputations among consumers, their key stakeholders. Following this logic, we predict that the green brand reputations will be positively related to the corresponding firms' actual environmental performance.

For green brand reputations, we use the Newsweek (Green) Reputation Score that was developed in 2012 by Landor Associates and Penn Schoen Berland through online survey of 251 companies with a sample of 8,743 American adults (Newsweek, 2012). Each respondent evaluated a random selection of 13-15 major brands with questions regarding the firms' various sustainability attributes such as "the firm is a green company," "the firm is an environmental leader," and "the firm has environmentally friendly products." Respondents also rated their green purchasing habits, their perception of brand packaging, the impact of green advertising, etc. (Landor Associates, 2012). An objective measure of corporate environmental performance is adopted from the 2012 Newsweek Green Ranking Score, which is a composite measure of (a) actual environmental management practices, (b) actual environmental outcomes as measured by emissions and other harmful outputs, and (c) corporate environmental communication and disclosure efforts. We also control for the impact of firm size (proxied by the ln of sales), advertising expenses (scaled by dividing by sales) using data from Compustat, and several environmental governance variables obtained from Bloomberg, which may impact consumer's perceptions of the firm's commitment to environmental causes (de Villiers, Naiker, and Van Staden, 2011). These include a dummy variable indicating the presence of a CSR or sustainability committee, the percentage of women directors on the board, and the percentage of independent board directors. Finally, since our firms are situated in twenty industry sectors, we cluster firms by industry when conducting the regression analyses.

To test the model, we ran a multiple regression analysis (with firm's clustered by industry sector), using the Newsweek Green Reputation Score as the dependent variable, and the rest variables as independent variables. As illustrated in column (a) of Table 1, we found a positive relationship between the Newsweek Green Reputation Score (i.e.,perceived sustainability) and the Newsweek Green Ranking Score (i.e., real sustainability). Assuming that the third-party information intermediaries like Newsweek are doing their jobs diligently (Ramchander, Schwebach, and Staking, 2012; Lyon and Shimshack, 2012),

we interpret this as supporting the argument that higher green brand ratings by consumers are, in general, significantly associated with corporate environmental performance.

Note that, as mentioned earlier, the Newsweek Green Ranking Score consists of three components: a firm's environmental management score, environmental impact score, and environmental disclosure score. In order to test the relative impact of each component on the green reputation score, in our next analysis we broke down the Newsweek Green Ranking Score into these three components and entered them into a multiple regression as separate independent variables. We found that the relationship between harmful environmental impacts (as measured by emission outcomes) is significantly negatively related to green brand reputations, whereas environmental management efforts are significantly positively related to green brand reputations. The regression coefficients are shown in column (b) of Table 1. Next, we repeated the analysis using a factor of the presence (captured by dummy variables) of environmental policies on green supply chain management, green quality management, green packaging, and green building, using data from Bloomberg. As shown in column (c) of Table 1, the presence of these environmental governance policies has a positive impact on green brand reputations among consumers. Overall, the results of these supplementary analyses remain similar as in the original model. We interpret evidence of a systematic positive relationship between green brand reputation and corporate environmental performance as implying a lower likelihood that green brands are effectively related to corporate attempts at greenwashing.

The analyses so far have focused on the effect of corporate environmental performance on green brand reputations. In our second empirical model, we attempt to explain the "green brand gap" measured as the difference between the Newsweek Green Reputation Score (i.e., perceived sustainability) and the Newsweek Green Ranking Score (i.e., real sustainability). We posit that this gap may be systematically related to the level of information asymmetry between the firm's management and outsiders. A higher degree of information asymmetry between managers and outsiders implies less information availability about the firm's actual operations and strategies; hence, consumers would have a harder time deciding on whether the firm is really as green as claimed or the claim is a case of greenwashing. This approach is consistent with recent research spotlighting the role of explicit CSR performance information in reducing the impact of emotion on investors' CSR perceptions (Elliott, Jackson, Peecher, and White, forthcoming).

Table 1: Multiple Regression Analysis of Determinants of Corporate Green Brand Reputation

	(a) Newsweek Green Ranking Score	**(b) Components of Newsweek Green RankingScore**	**(c) Bloomberg Environmental Policies Factor**
Newsweek (Overall) Green Ranking Score	.108 * (.043)		
Newsweek Environmental Management Score		.159 *** (.036)	
Newsweek Environmental Impact Score		-.037 + (.021)	
Newsweek Environmental Disclosure Score		-.016 (.019)	
Bloomberg Environmental Policies Factor			1.414 *** (.340)
Firm Size (Ln of Sales)	-.276 (.581)	-.473 (.594)	-.373 (.598)
Advertising Intensity	15.935* (7.083)	14.370 * (5.477)	8.731+ (4.584)
Presence of CSR/Sustainability Committee	-.559 (1.261)	-.958 (1.09)	-1.100 (1.281)
Percentage of Women Directors on Board	.014 (.040)	-.025 (.039)	-.006 (.032)
Independent Director Ratio	.051 (.039)	.044 (.034)	.037 (.031)
Constant	49.400*** (5.368)	52.776 *** (4.314)	57.051*** (5.121)
F-value (prob)	6.92 (.000)	16.35 (.000)	10.50 (.000)
Adjusted R^2	0.070	0.182	0.131

N= 180, Significance levels: *** indicates p < .001, ** indicates p<.01, * indicates p <.05, + indicates p <.1 Standard errors reflect clustering by industry sector

The level of corporate's CSR, and especially environmental disclosure and reporting (both print and on-line) has increased dramatically over the past decade or so (Clarkson, Li, Richardson, and Vaswari, 2008; Clarkson, Fang, Li, and Richardson, 2013). We test for the impact of level of the firm's environmental disclosure and reporting on the green brand gap. Since this disclosure should help reduce the level of information asymmetry (Kim and Lyon, 2011), we expect higher level of environmental disclosure to be associated with smaller green brand gap. Finally, we test the impact of general corporate reputation (using ratings data for a subset of firms from Fortune's World Most Admired Corporations survey in 2012) as the firm's general reputation may provide a "halo" effect that influences the firm's green brand reputation.

Model (2): Green Brand Reputation Gap = f(Firm-level information asymmetry, firm-level environmental disclosure and reporting, general firm-level reputation, controls for firm size, advertising intensity, environmental governance proxies, error)

We calculated the green brand gap by subtracting firms' actual environmental performance from the same firms' green brand reputation score. Importantly, before we calculate the gap, we modified the actual corporate environmental performance to include only environmental impact and management subcomponents of the Newsweek Green Reputation Score and to exclude the third component, that is, corporate environmental communication and disclosure efforts. In addition, we added two potential determinants of the green brand gap: 1) the number of security analysts following the firm and the "forecast error" as measured by (i.e. the gap between the mean earnings per share (eps) forecast generated by analysts and the actual eps) as measures of information asymmetry and 2) the Newsweek environmental disclosure score as the measure of corporate environmental disclosure. As in our earlier model, we controlled for firm size, advertising intensity, the governance variables as well as the percent of shares owned by institutions, and (separately) by insiders. To test the model, we regressed all the potential determinants on the green brand gap with industry sector clustered.

Column (a) of Table 2 illustrates the results of this multiple regression and shows that the brand gap is in fact significantly related to the level of information asymmetry but not to the level of environmental disclosure. This means firms with higher levels of information asymmetry are likely to have a greater gap between its greenness reputation and its actual environmental performance.

These results immediately above have focused on how the overall green brand gap can be explained by the level of information asymmetry, but not the extent of environmental disclosure. Our next analysis tests the impact of general corporate reputation on the green brand gap, since the firm's general reputation may create a "halo" effect that influences the firm's green brand reputation. For this purpose, we used a subset of firms from Fortune's World Most Admired Corporations in 2012. We simply added the General Corporate Reputation Score as an independent variable onto the multiple regression we tested earlier. The results in column (b) of Table 2 show that firms with more positive reputations tend to have a greater disparity between their green brand reputations and actual environmental performance. This result means that firms with positive overall reputations enjoy more favorable green brand reputations than should be warranted by their environmental performance.

Table 2: Multiple Regression Analysis of Determinants of Green Brand Gap

	(a)	**(b)**
Number of Analysts Following Firm	-.526 *** (.168)	-.519* (.198)
Analyst Forecast Error	.431 *** (.074)	.395*** (.101)
Corporate Environmental Disclosure Score	-.033 (.033)	-.034 (.032)
General Corporate Reputation Score from Fortune		1.634 + (.874)
Firm Size (Ln of Sales)	-2.007 + (1.161)	-1.891 (1.592)
Advertising Intensity	31.003 (33.306)	30.277 (47.444)
Presence of CSR/Sustainability Committee	4.866 + (2.471)	4.758* (1.947)
Percentage of Women Directors on Board	-.149 (.089)	-.179 (.107)
Independent Director Ratio	.157 + (.0082)	.135 (.110)
Constant	17.656 (13.293)	6.88 (17.172)
F-value (prob)	99.54 (.000)	52.78
Adjusted R^2	0.274	0.278

Note: N= 143 for model (a) and 113 for model (b)Significance levels: *** indicates p < .001, ** indicates p<.01, * indicates p <.05, + indicates p <.1 Standard errors reflect clustering by industry sector.

In our final analyses, we focus on two sub-groups of particular interest: “free passers” and “unsung heroes” (i.e. firms whose brand reputation is much higher (lower) than warranted by their actual environmental performance). Here, our interest lies in understanding determinants of these two groups of firms, since a significant gap between perceived and real sustainability may represent a risk to the firm or an underutilized resource. Specifically, free passers may be faced with the threat of being labeled a greenwasher while unsung heroes are struggling with unfairly poor green brand reputations. Median splits were used to determine “high” and “low” categories of green brand reputation and of environmental performance. “Free passers” (“unsung heroes”) were identified using a dummy variable for those firms whose brand reputation was above (below) the brand reputation median but whose environmental performance was below (above) the median performance. We used similar regression analyses as in Model 2 with a couple of modifications. First, we conducted separate logit regressions for “free passer” dummy, and, separately, the “unsung heroes” dummy as the dependent variable. Second, we augmented the model with the inclusion of two additional measures of the percentage of stock held by institutional owners and the percentage of stock held by insiders, both of which are known to represent the level of firm’s information asymmetry to a certain degree.

The results showed that the level of information asymmetry (as proxied by the number of

analysts following the firm) is a significant predictor of "free passers", suggesting that firms with higher levels of information asymmetry are more likely to be in this group. We also find that corporate environmental disclosure score is a significant predictor of "unsung heroes", suggesting that firms with lower levels of environmental disclosure are more likely to be in this category. Additionally, the "free passer" effect is heightened and the "unsung hero" effect reduced by a more favorable general corporate reputation (as proxied by the Fortune reputation score). Overall, these results are in accordance with and confirm the findings from Model 2.

General Discussions

Our findings contribute to the sparse literature on green corporate branding, usefully augmenting the initial conceptual contributions by Chen (2011) in terms of empirical findings. Corporate brand reputations are important for marketers, because brand reputations are related to CSR beliefs held by consumers. Recent research shows that positive CSR beliefs held by consumers are associated both with greater purchase likelihood and longer-term loyalty and advocacy behaviors (Du et al. 2007); consumer satisfaction with CSR has also been shown as possible mediators of the relationship between corporate CSR initiatives and firm market value (Luo and Bhattacharya, 2006). Our first contribution is demonstrating that at least in the case of "green" brand reputations, the brand reputation across our sample of large US firms is plausibly driven by objective corporate environmental performance, thus downplaying the role of greenwashing as an important alternative determinant of brand reputation. In addition to the consumer benefits noted above, this finding should help forestall perceptions and concerns over greenwashing from powerful stakeholders such as NGOs, environmental activists, and policy makers. While the results above were "on average" results for the sample as a whole, it remains the case that some firms benefit from apparently unwarranted green brand reputations ("free passers") while others suffer from lack of deserved recognition ("unsung heroes"). Our second contribution lies in explaining the gap that has intrigued recent media commentators on corporate green branding. We examine a prominent recent example of this gap, using the disparity data from Newsweek and demonstrate that the gap is related to the extent of information asymmetry and general corporate reputation (which evidently benefits the "free passers") and also the extent of general corporate reputation as well as firm-level environmental disclosure (lack of which evidently hurts the "unsung heroes"), two plausible and interesting explanations for this phenomenon. Thus, firms that have an image perception of "unsung heroes", for instance, might well choose to deploy more extensive communications about their environmental performance in an effort to reduce the gap. Consumer advocacy groups, NGOs, and environmental activists, on the other hand, should be wary of firms where levels of

information asymmetry is high, either due to structural reasons or because insiders at these firms choose to keep outsiders in the dark, possibly due to self-interested motives.

Our research is not without limitations, the most serious being its cross-sectional approach. With this in mind, we are careful to stress that our findings, although interesting are associations at this point, with no implication of causality intended. Extensions to panel data are hampered at this point by lack of available data, but we anticipate that future researchers will be able to capitalize on greater data availability to remedy this deficiency. Investigations within specific industry contexts, replication in other national domains and investigation of the impact of green brand reputation on key corporate outcomes are also warranted to advance our understanding of green brand determinants as an interesting green marketing phenomenon.

13.3 Bridging the Divide: Industry Insights on Retail Location Analytics and AI

Authors: Tony Hernandez *(Toronto Metropolitan University)*
Joseph Aversa *(Toronto Metropolitan University)*

This paper examines the locational decision-making practices of major retail and service firms operating in Canada. The paper explores the changing nature of location decision-making based on i0n-depth interviews with key decision-makers within leading corporations. The interviews highlight the increasing challenges organizations face in integrating diverse big data sets within traditional decision-support approaches. The findings reveal a growing divide between rapidly emerging data science-based methods and legacy spatial analytical approaches. While the emerging techniques provide opportunities to enhance decision-making, there are many challenges to fully leveraging these new approaches. The paper critically explores the inertia in location decision-making cultures. It provides insights into the organizational and data infrastructural elements that need to be in place to promote the adoption, use and development of data science and AI-enhanced decision support. The implications of the findings to higher education are also discussed, with educators faced with a growing challenge in keeping pace with industry and maintaining the relevancy of the curriculum and currency of skill sets.

13.4 Valuing Commercial Real Estate and Its Real Options

Authors: Robert Jarrow *(Cornell University)*
Crocker Liu *(Cornell University)*
Motoyuki Yoshihara *(Cornell University & US Army)*

Using an extension of standard option pricing theory for sporadically traded assets, this paper presents a new methodology for valuing commercial real estate (CRE) and real options on them. We apply this method to value multifamily apartments in Los Angeles,

California over the 2001 - 2019 time period and compare it to a standard hedonic model. Our model performs better both in- and out-of-sample, maintaining robustness despite changing market conditions. We also illustrate how to use our approach to value a European call option to purchase a CRE. Thus, our approach can be useful to practitioners, both those that transact in commercial real estate, as well as its derivatives.

Track 14: General Analytics

Room: WC G103A (9:30 am - 10:50 am)

Moderator: Lisa Walters *(State University of New York at Fredonia)*

14.1 Exploring AI Power in Setting Prices: Testing the Design, Application, and Outcomes of PriceGPT Using Live Business Simulation Environments

Author: Joseph T Kuvshinikov *(Gannon University)*

This research study explores the design, application, and outcomes of PriceGPT, an AI-driven pricing optimization tool tailored to help product managers make smarter pricing decisions. PriceGPT utilizes advanced natural language processing and predictive analytics to recommend data-driven pricing strategies. Key features of PriceGPT include real-time analysis of decision maker inputs such as market demand, cost structures, competitive pricing, supply and demand, and business goals. These capabilities enable users to make informed pricing decisions aligned with both market conditions and organizational objectives.

To evaluate its effectiveness, PriceGPT was tested using the Income|Outcome (IO) Business Simulation (Andromeda Simulations International), a highly interactive team-based platform that models real-world business competitive decision making, operations, and strategy. The IO simulation allows participants to consider and make decisions based on key variables such as supply and demand, cost structures, pricing strategies, and competition. By simulating the impacts of strategic decisions in a competitive marketplace, it provides a rigorous test environment for tools like PriceGPT.

Results demonstrated that participants using PriceGPT were able to make more informed decisions and were more effective in setting prices than competitors who did not use AI. PriceGPT's ability to synthesize complex variables and offer actionable recommendations highlights its potential to enhance decision-making in competitive and volatile markets. This study showcases the role of AI in bridging the gap between theoretical pricing models and real-world application. Future testing aims to incorporate broader external market indicators, such as customer sentiment and macroeconomic trends, and internal user price setting experience for even greater accuracy.

14.2 Enhancing Telecom Customer Retention: Advanced Predictive Modeling Approaches

B. Rahmani, L. Yulei, P. K. Thodupunoori, K. R. Syamala, K. A. Bachu, P. Bollepalli, K. Sowmya, V. Sharma, P. Norouzzadeh, S. Tutun, E. Snir
[1]Saint Louis University, School of Medicine, Saint Louis, MO
[2]Washington University in Saint Louis, Olin Business School, Saint Louis, MO
[3]Saint Louis University, School of Professional Studies, Saint Louis, MO

Abstract

Customer churn remains a critical challenge in the highly competitive telecommunications industry, directly affecting financial stability and customer retention. Retaining existing customers is not only more cost-effective than acquiring new ones but also ensures long-term revenue growth and brand loyalty. This study aims to enhance churn prediction by leveraging a comprehensive dataset of 7,043 telecom customers, incorporating demographic details, service usage patterns, and financial interactions. We employ advanced predictive modeling techniques like Seasonal Moving Averages (SMA), to forecast churn likelihood and identify key determinants. Through data visualization techniques like scatter plots and time series decomposition, we uncover crucial trends influencing customer retention. Our findings indicate that factors such as monthly charges, total charges, and customer tenure are the strongest predictors of churn. Additionally, we explore the challenges posed by dynamic changes in revenue and customer preferences, emphasizing the need for adaptive predictive models. This research provides valuable insights to help telecom companies implement proactive retention strategies and improve customer loyalty.

1. Introduction

Customer churn is a major challenge in the telecommunications industry, impacting financial performance and customer relationships. Since retaining existing customers is more cost-effective than acquiring new ones, churn prediction is essential for improving customer loyalty and revenue. This study addresses this issue by applying time series modeling and machine learning techniques to a dataset of 7,043 telecom customers.

We use models such as XGBoost, Random Forest, and Naïve Bayes, along with visualization tools like scatter plots and time series decomposition, to identify churn patterns. Key features include monthly charges, total charges, and tenure, which strongly correlate with churn. Forecasting models like regression and seasonal moving averages (SMA) provide valuable insights into retention trends, even under shifting revenue and preference dynamics.

Churn prediction has been extensively researched. Ahmad et al. [1] used SNA and XGBoost, achieving high AUC scores, while Sana et al. [2] showed the benefits of data transformation. Óskarsdóttir et al. [3, 4] emphasized social network analysis, and Yang et al. [5] developed a T+2 forecasting framework. Nguyen et al. [6] enhanced SVM-based models with oversampling techniques.

Comparative studies have assessed various algorithms, with Prabadevi et al. [7] favoring gradient boosting and Wagh et al. [8] stressing the value of up-sampling and survival analysis. Deep learning models have also emerged: Wu et al. [9] introduced a self-attention-based neural network; Ahmed et al. [10] leveraged transfer learning; and Khattak et al. [11] combined BiLSTM and CNN for sequence analysis.

Data processing innovations have improved results further. Sikri et al. [12] proposed ratio-based balancing with XGBoost, and Bhuse et al. [13] showed that fine-tuned Random Forests performed best. Chang et al. [14] combined ensemble learning with SHAP for interpretability.

Together, these works show how engineering, algorithm optimization, and explainability tools have significantly advanced telecom churn prediction.

2. Data Description

Telecommunications Customer Churn dataset unveils key insights into demographics, behavior, and factors impacting customer retention. With details on 7,043 customers in California in Q2 2022, this analysis aims to decipher patterns, uncover trends, and provide valuable understanding for strategic decision-making. This dataset could be found in: https://www.kaggle.com/datasets/shilongzhuang/telecom-customer-churn-by-maven-analytics/data. The selected features are:

Tenure in months, which refers to the number of months a customer has been subscribed to the telecom service. Meanwhile short-tenure customers (0-6 months) consider it to be a higher churn rate. Mid-tenure customers (6-24 months) are moderate churn risk. Long-tenure customers (24+ months) are lower churn but may churn due to contract expiration.

Monthly charges refer to the amount a customer is billed each month for using telecom services. It includes charges for voice calls, internet, TV subscriptions, and any additional services.

3. Methodology

3.1 Correlation

Correlation measures the strength and direction of the relationship between two variables. The coefficient correlation quantifies this linear relationship, ranging from -1 (strong negative correlation) to +1 (strong positive correlation). In the case of a positive

correlation, when one variable increases by one unit, the other variable increases proportionally based on the correlation coefficient, and vice versa.

3.2 Decomposition Graph

Decomposition helps us understand the trend changes, seasonal variations, and residual fluctuations within a time series. For example, trend represents the long-term variation in a time series, describing the overall direction of the data over a period. Seasonality refers to the recurring patterns within the time series, reflecting periodic fluctuations at fixed intervals. Residual represents the random variations that remain after removing trend and seasonal components.

3.3 Regression Model

Regression models are statistical methods that model the relationship between a dependent variable (target) and one or more independent variables (predictors). These models make predictions and identify the impact of different factors on the target variable.

3.4 Smoothing Moving Average (SMA)

A Smoothing Moving Average (SMA) is a widely used technique in time series analysis that smooths out short-term fluctuations and highlights long-term trends. It is computed by taking the arithmetic means of a fixed number of past observations within a defined window. By reducing short-term volatility, SMA helps reveal underlying patterns, though it reacts slower to recent changes due to its reliance on past values.

$$\boldsymbol{SMA_t} = \frac{X_t + X_{t-1} + \cdots + X_{t-(n-1)}}{n}$$

$\boldsymbol{SMA_t}$ is the simple moving average at time t, representing the average of the past n observations. $\boldsymbol{X_t}$ is the data point in the time series at the time t. n is the number of periods (window size) over which the average is calculated.

3.5 Second Order fitting process (FIT2)

The Second Order Fitting Process (FIT2) refers to the process of fitting a second-order (quadratic) regression model to a given dataset. This method is used when the relationship between the dependent variable and independent variables is nonlinear. When the true relationship between the dependent and independent variable is nonlinear, a quadratic model can improve prediction accuracy compared to a simple linear model. And it is easier to interpret than complex machine learning models.

$$\boldsymbol{Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon}$$

Y is the dependent variable (response variable), X is the independent variable (predictor), $\mathbf{X^2}$is the second order (quadratic) term, and capturing curvature in the data, $\boldsymbol{\beta_0, \beta_1, \beta_2}$are regression coefficients,$\boldsymbol{\epsilon}$ is the error term.

4. Results

4.1 Correlation

Based on Table 1:

- "Tenure in Months" and "Total Charges" have a strong positive correlation of 0.826.
- "Monthly Charge" and "Total Charges" exhibit a positive correlation at 0.623.
- "Tenure in Months" and "Monthly Charge" have a moderate positive correlation of 0.239.

	Tenure in Months	Monthly Charge	Total Charges
Tenure in Months	1.0000000	0.2390652	0.8260735
Monthly Charge	0.2390652	1.0000000	0.6228098
Total Charges	0.8260735	0.6228098	1.0000000

Table 1: Correlation between tenure in months, monthly charges, total charges

These correlation values indicate the direction and strength of linear relationships between "Tenure in Months" and "Total Charges".

4.2 Decomposition graph

Decomposition graph demonstrated in figure 1 indicates:

Trend: The trend line's mild upward slope after 30(30*200=6000 from the row 6000 in data) suggests that the Tenure in Months will gradually rise over time.

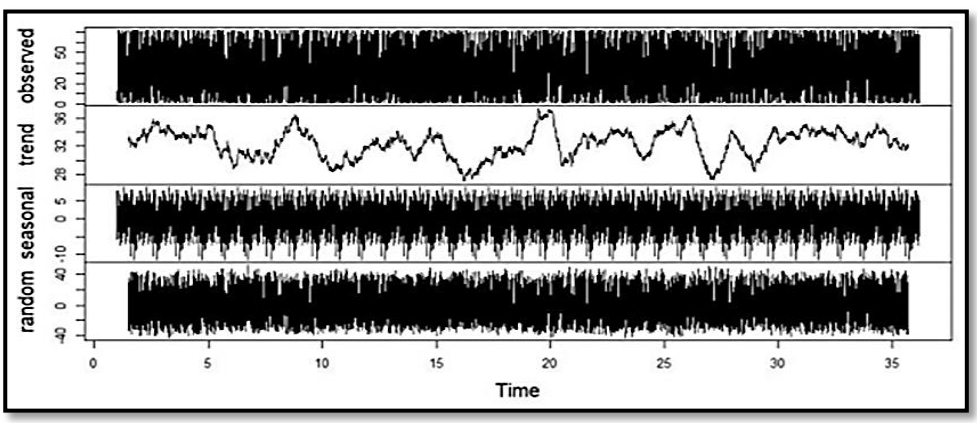


Figure 1: Decomposition graph of tenure in months

Seasonality: There are two peaks in the pattern, one around month six and the other around month eighteen. This implies that there is a greater chance during these months.
Random: There are unpredictable fluctuations in the random component, which fluctuates without appearing to follow a prevailing pattern.

4.3 Prediction of Total Revenue

4.3.1 Regression Model

The regression model shown in Figure 2 predicts the total charges based on age and monthly charges. A broad distribution of residuals is shown by the substantial difference (12576.6) between the minimum and maximum residuals. 25% of the residuals are less than -1332.3 (1Q) while 75% of the residuals are less than 1143.7 (3Q). The remaining standard error shows how much the model's predictions typically differ from the actual total charges. F-Statistic, the model's overall significance is shown by the F-statistic (2236) and the corresponding p-value (< 2.2e-16) which suggests that the model is statistically significant in explaining the variability in total charges. Based on figure 2 age and total charges show a positive correlation.

Figure 3 shows the predicted values using linear regression model. This is in line with predictions because older clients may have had higher tier plans or been with the firm longer, accumulating more expenses. While the overall trend shows a positive slope, there's also scatter around the trend line. This indicates that not all older customers have higher charges, and not all younger customers have lower charges. There are exceptions based on individual circumstances.

4.3.2 Smoothing Moving Average (SMA)

Figure 4 shows the prediction of total revenue using smoothing moving average (SMA(12)) for 12 months indicating a year. The next predicted values would be around 4952.752.

```
Call:
lm(formula = Total.Charges ~ Age + Monthly.Charge, data = selected_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-3993.6 -1332.3   -72.4  1143.7  8583.0 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)     -457.0006    71.7233  -6.372 1.99e-10
Age               -3.3193     1.2725  -2.609  0.00911
Monthly.Charge    45.4707     0.6831  66.568  < 2e-16
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1772 on 7040 degrees of freedom
Multiple R-squared:  0.3885,	Adjusted R-squared:  0.3883 
F-statistic:  2236 on 2 and 7040 DF,  p-value: < 2.2e-16
```

Figure 2: Regression Model to predict total charge based on age and monthly charge

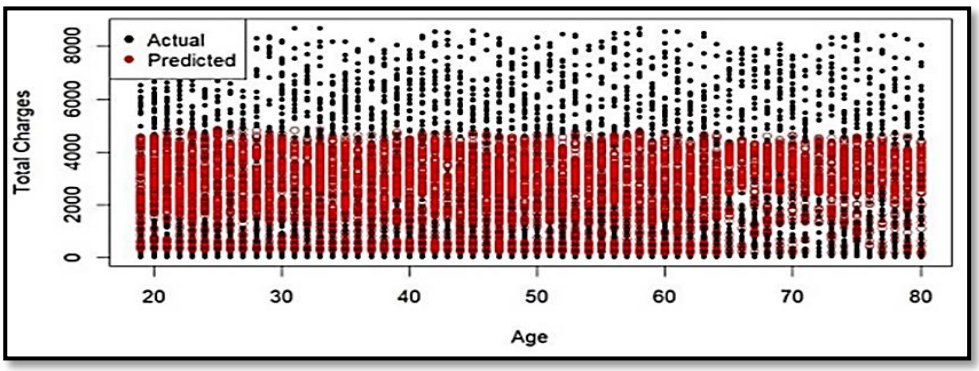


Figure 3: Predicted total charges with linear regression

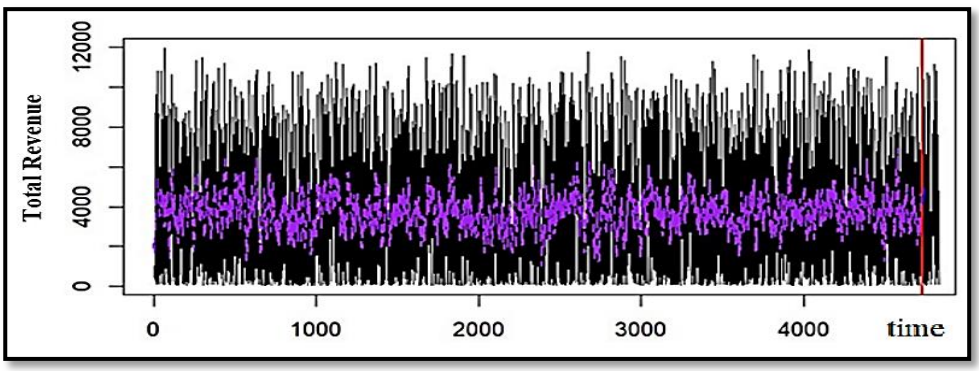


Figure 4: Predicted total revenue with SMA(12)

4.3.3 Second Order fitting process (FIT2)

Figure 5 represents a second-degree polynomial fit, nonlinear regression, or iterative model improvements. The shaded regions show the range within which the true total revenue is expected to lie with a certain level of confidence of -1.618 and 1.647. The line in the graph indicates how the variable (X) is expected to change over time or across different conditions. The highlighted area suggests a specific interval where the predictions are more concentrated or of particular interest.

5. Conclusion

The developed models demonstrate significant accuracy and effectiveness in predicting churn, providing valuable insights for decision-making in customer retention strategies.

Trends and Patterns: Fluctuations in tenure over time are observed, providing insights into tenure trends and data patterns. While no clear seasonal pattern is evident, the dense plot illustrates variability in tenure across numerous observations.

Data Distribution: The dataset's deviation from a theoretical normal distribution is evident, indicating non-normality in the data distribution.

Prediction Modeling: A prediction model is developed to forecast future values of a variable, providing insights into variable fluctuations and focused prediction intervals.

Model Evaluation: Both the Logit and Probit models achieve a high accuracy of 79% in predicting customer churn. ROC curve analysis indicates comparable performance between the two models, suggesting their effectiveness in classification tasks.

Probability Distributions: Predicted probability distributions for both models offer insights into how they estimate the likelihood of churn for each customer, aiding in understanding the models' confidence in their predictions.

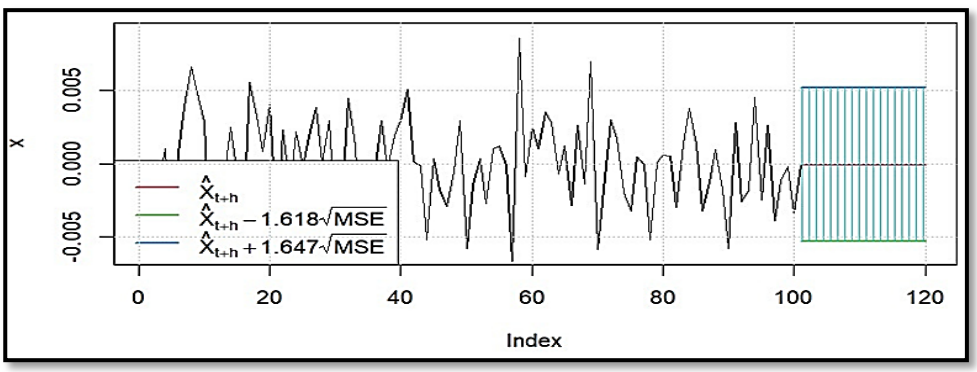


Figure 5: Predicted total revenue with FIT2

References

[1] A. K. Ahmad, A. Jafar, and K. Aljoumaa, "Customer churn prediction in telecom using machine learning and social network analysis in big data platform," *J. Big Data*, vol. 6, p. 28, 2019.

[2] J. K. Sana, M. Z. Abedin, M. S. Rahman, and M. S. Rahman, "Data transformation-based optimized customer churn prediction model for the telecommunication industry," *arXiv preprint arXiv:2201.04088*, 2022.

[3] M. Óskarsdóttir, C. Bravo, W. Verbeke, C. Sarraute, B. Baesens, and J. Vanthienen, "Social network analytics for churn prediction in telco: Model building, evaluation and network architecture," *arXiv preprint arXiv:2001.06701*, 2020.

[4] M. Óskarsdóttir, C. Bravo, W. Verbeke, C. Sarraute, B. Baesens, and J. Vanthienen, "A comparative study of social network classifiers for predicting churn in the telecommunication industry," *arXiv preprint arXiv:2001.06700*, 2020.

[5] L. Yang, D. Li, and Y. Lu, "Prediction modeling and analysis for telecom customer churn in two months," *arXiv preprint arXiv:1911.00558*, 2019.

[6] N. Y. Nhu, T. Van Ly, and D. V. Truong Son, "Churn prediction in telecommunication industry using kernel support vector machines," *PLoS ONE*, vol. 17, p. e0262722, 2022.

[7] B. Prabadevi, R. Shalini, and B. R. Kavitha, "Customer churning analysis using machine learning algorithms," *Int. J. Intell. Netw.*, vol. 4, pp. 1–8, 2023.

[8] S. K. Wagh and K. S. Wagh, "Customer churn prediction in telecom sector using machine learning techniques," *SSRN Electron. J.*, 2022.

[9] H. Wu, "A high-performance customer churn prediction system based on self-attention," *SSRN Electron. J.*, 2022.

[10] U. Ahmed, A. Khan, S. H. Khan, A. Basit, I. U. Haq, and Y. S. Lee, "Transfer learning and meta classification based deep churn prediction system for telecom industry," *arXiv preprint arXiv:1901.06091*, 2019.

[11] A. Khattak, Z. Mehak, H. Ahmad, M. U. Asghar, M. Z. Asghar, and A. Khan, "Customer churn prediction using composite deep learning technique," *Sci. Rep.*, vol. 13, p. 17294, 2023.

[12] A. Sikri, R. Jameel, S. M. Idrees, and H. Kaur, "Enhancing customer retention in telecom industry with machine learning driven churn prediction," *Sci. Rep.*, vol. 14, Article 13097, 2024.

[13] P. Bhuse, A. Gandhi, R. Muni, and N. Katre, "Machine learning based telecom-customer churn prediction," in *Proc. Int. Conf. Inf. Sci. Syst. (ICISS)*, Dec. 2020, pp. 1-6, doi: 10.1109/ICISS49785.2020.9315951.

[14] V. Chang, K. Hall, Q. A. Xu, F. O. Amao, M. A. Ganatra, and V. Benson, "Prediction of customer churn behavior in the telecommunication industry using machine learning models," *Algorithms*, vol. 17, no. 6, p. 231, 2024.

14.3 Examining the Pathways of Consumer Animosity: a Structural Equation Modeling Study Across Japan and the Philippines

Authors: Reynaldo Bautista *(De La Salle University)*
Luz T Suplico Jeong *(De La Salle University)*
Takanori Osaki (*Kagawa University)*
Michele Stewart *(University of Windsor)*

Introduction

The global consumer landscape is increasingly influenced by geopolitical developments, with international conflicts shaping perceptions of nations and their associated brands. One of the most pronounced conflicts of the 21st century, the Russia-Ukraine war, has not only disrupted political alliances but also impacted consumer sentiment and behavior on a global scale. Media content originating from or related to the aggressor country can influence consumers' emotional and cognitive responses, often culminating in anti-consumption behavior. With digital media becoming a primary channel for news dissemination and public discourse, understanding its psychological and behavioral ramifications is crucial.

This study investigates the pathways through which digital media intrusiveness influences consumer animosity and its subsequent emotional and behavioral outcomes—fear, product boycott, and negative word-of-mouth (NWOM)—among consumers in Japan and the Philippines. These countries present contrasting cultural orientations—Japan leans toward collectivism and structured societal norms, while the Philippines is characterized by relational collectivism and high social engagement, especially through digital media. By employing Structural Equation Modeling (SEM), this research extends the Uses and Gratifications (U&G) theory and offers comparative insights into how cultural and psychological factors mediate the effects of conflict-related media exposure.

Theoretical Framework

The theoretical underpinning of this study is the Uses and Gratifications (U&G) theory, which posits that media consumption is driven by individuals' needs and gratifications but can also produce unintended negative effects. Originally designed to explain why individuals engage with various media platforms, U&G has evolved to encompass outcomes of media usage, especially in the context of unsolicited or intrusive content.

In the context of international conflict, media perceived as intrusive—particularly when originating from or related to an aggressor country—can disrupt cognitive processes and evoke strong emotional reactions. Perceived intrusiveness leads to consumer animosity, which refers to hostility towards a nation and its products due to political or military actions (Harmeling et al., 2015). This animosity may further provoke fear, defined as a

psychological retreat response to perceived threats (Nesse, 1990). These emotional responses, in turn, trigger behavioral outcomes such as product boycott and NWOM, both of which are coping and protest mechanisms.

Hypotheses

Perceived intrusiveness is understood as a disruption in cognitive processing caused by media content (Li et al., 2002; Smink et al., 2020). When ads from a politically aggressive country invade digital spaces, users may develop resentment. This is particularly relevant in conflict reporting, where media from the aggressor may exacerbate consumer animosity (Ducoffe, 1995).

H1: Perceived intrusiveness positively affects the Japanese and Filipino consumers' animosity towards Russian brands.

Consumer animosity—negative feelings toward a country and its products due to political behavior—can foster fear, especially as individuals perceive threats to peace and economic stability (Harmeling et al., 2015; Nesse, 1990). According to the Uses and Gratifications (U&G) theory, this fear may be triggered by unwanted media exposure and serve as a psychological retreat mechanism (Ducoffe, 1995).

H2: Consumer animosity of Filipino and Japanese consumers is positively related to fear.

Fear, as a retreat-oriented emotion, may lead individuals to avoid or reject products associated with a threatening country. This form of consumer withdrawal, such as a boycott, aligns with the U&G perspective that media-triggered distress can result in behavioral disengagement (Harmeling et al., 2015; Ducoffe, 1995).

H3: The fear felt by Japanese and Filipino consumers is positively related to product boycott.

Fear may also stimulate coping responses like spreading warnings through negative word-of-mouth (NWOM). Consumers fearing harm from an aggressor country might discourage others from purchasing its products by sharing negative opinions (Bougie et al., 2003; Siddiqi et al., 2020). U&G theory supports this as a negative consequence of disruptive media exposure (Ducoffe, 1995).

H4: The fear felt by Japanese and Filipino consumers is positively related to negative word-of-mouth.

Results and Discussion

Results from the SEM analysis using SmartPLS indicate that all four hypotheses were statistically supported for both Japanese and Filipino samples. However, differences emerged in the strength of the relationships.

For Japanese consumers, the path from perceived intrusiveness to consumer animosity (H1) was particularly strong ($\beta = 0.683$, $p < 0.001$), reflecting the culturally rooted aversion to cognitive disruption and preference for harmony. Filipino consumers, while also responsive ($\beta = 0.468$, $p < 0.001$), showed relatively lower sensitivity, possibly due to higher digital media engagement and adaptive coping styles.

H2 (consumer animosity to fear) was robust across both groups but stronger for Filipino consumers ($\beta = 0.579$) than Japanese ($\beta = 0.423$), suggesting that animosity may trigger more intense emotional responses in socio-politically expressive cultures like the Philippines.

H3 and H4 (fear to boycott and NWOM) were confirmed in both contexts. Interestingly, Filipinos exhibited a slightly higher tendency to engage in NWOM ($\beta = 0.541$) compared to Japanese consumers ($\beta = 0.436$), consistent with the Filipino cultural emphasis on social communication and peer guidance.

The MGA (Multi-Group Analysis) showed invariance across groups, indicating that while effect sizes vary, the directionality of relationships is consistent. These findings highlight both universal and culturally specific pathways in how consumers process media content during conflicts, thereby validating the extended U&G model.

From a theoretical standpoint, this study refines the U&G framework by incorporating conflict-based consumer behavior. Practically, it alerts marketers and policy strategists to the consequences of media intrusiveness, suggesting that brand origin visibility and content tone must be managed cautiously during global tensions.

This study affirms that media intrusiveness, particularly from aggressor nations during times of conflict, triggers a cascade of consumer reactions beginning with animosity, leading to fear, and culminating in both avoidance (boycott) and prosocial advocacy (NWOM). While the theoretical pathways were stable across both cultural settings, the Philippines exhibited stronger behavioral responses (boycott, NWOM), likely due to heightened media engagement and expressive emotional tendencies. Japan, in contrast, displayed higher cognitive sensitivity to intrusiveness but comparatively lower behavioral reactivity.

From a theoretical standpoint, the study validates the utility of U&G theory in explaining unintended consequences of media exposure in international conflict scenarios. It expands our understanding of how media saturation in politically volatile times can influence consumer sentiments and actions globally.

Practically, marketers must be cautious when operating in conflict-sensitive regions. Ad content, especially that which seems intrusive or politically loaded, may backfire and result in reputational harm or product rejection.

Conclusion

This study investigated how perceived intrusiveness of media content related to the Russia-Ukraine conflict affects consumer attitudes and behaviors in Japan and the Philippines. Grounded in the Uses and Gratifications (U&G) theory (Ducoffe, 1995), the findings confirm a sequential psychological pathway: from perceived media intrusiveness to consumer animosity, then to fear, and subsequently to behavioral responses such as product boycott and negative word-of-mouth (NWOM).

Results affirmed all four hypotheses. Perceived intrusiveness significantly increased consumer animosity toward Russian brands (Li et al., 2002; Smink et al., 2020). This animosity, in turn, positively influenced fear (Harmeling et al., 2015; Nesse, 1990), which led to two forms of behavioral disengagement—boycotting (Harmeling et al., 2015) and NWOM (Bougie et al., 2003; Siddiqi et al., 2020). Notably, the strength of relationships between constructs varied across countries. Filipino consumers reported higher levels of perceived intrusiveness and fear, suggesting greater emotional sensitivity to foreign political aggression. Japanese consumers, while similarly responsive, exhibited relatively lower levels of boycott behavior, possibly reflecting differences in cultural collectivism and conflict-related consumer behavior (Triandis, 1995).

The U&G theory remains a robust framework for understanding how media exposure can trigger emotional and behavioral responses, particularly in geopolitical contexts. Media perceived as intrusive does not only disrupt cognitive processing but can also amplify consumer hostility, making national identity and international relations critical elements of global marketing strategy.

Recommendations

For marketers and advertisers: Avoid placing advertisements from or about politically aggressive countries in conflict-sensitive digital spaces. Intrusive ads from such contexts can provoke strong negative emotional reactions and damage brand reputation, particularly in collectivist societies like the Philippines and Japan (Triandis, 1995).

For policymakers: There is a need to regulate international advertising content, especially during times of geopolitical tension. Media from countries involved in conflicts can unintentionally stoke animosity and influence public perception beyond political arenas.

For future researchers: This study highlights the importance of culturally nuanced approaches to consumer sentiment analysis. Future work could expand the model to other conflict zones (e.g., Israel–Palestine) and test cross-cultural validity using structural equation modeling (SEM) in diverse contexts. Incorporating longitudinal data may also reveal how media-induced animosity evolves over time.

For educators and institutions: Understanding media intrusiveness and consumer reactions should be part of the digital literacy curriculum. Educating users to critically evaluate international content may reduce impulsive fear or animosity responses that lead to misinformation or biased consumer behavior.

References

Bougie, R., Pieters, R., & Zeelenberg, M. (2003). Angry customers don't come back, they get back: The experience and behavioral implications of anger and dissatisfaction in services. Journal of the Academy of Marketing Science, 31(4), 377–393. https://doi.org/10.1177/0092070303254412

Ducoffe, R. H. (1995). How consumers evaluate advertising: The case of magazine ads. Journal of Current Issues & Research in Advertising, 17(1), 1–18. https://doi.org/10.1080/10641734.1995.10505022

Harmeling, C. M., Magnusson, P., & Singh, N. (2015). Beyond anger: A deeper look at consumer animosity. Journal of International Business Studies, 46(6), 676–693. https://doi.org/10.1057/jibs.2015.2

Li, H., Edwards, S. M., & Lee, J.-H. (2002). Measuring the intrusiveness of advertisements: Scale development and validation. Journal of Advertising, 31(2), 37–47. https://doi.org/10.1080/00913367.2002.10673665

Nesse, R. M. (1990). Evolutionary explanations of emotions. Human Nature, 1(3), 261–289. https://doi.org/10.1007/BF02692193

Siddiqi, M., Khan, I., & Alshammari, S. D. (2020). The impact of animosity and consumer ethnocentrism on Pakistani consumers' purchase intentions. Journal of Islamic Marketing, 11(3), 793–811. https://doi.org/10.1108/JIMA-04-2018-0067

Smink, A. R., van Reijmersdal, E. A., van Noort, G., & Neijens, P. C. (2020). Understanding the effects of ad intrusiveness in digital video advertising. Journal of Interactive Marketing, 50, 99–115. https://doi.org/10.1016/j.intmar.2019.09.003

Triandis, H. C. (1995). Individualism and collectivism. Westview Press.

Track 15: Emerging Scholars IV: Economics and Policy

Room: WC G103B (9:30 am - 10:50 am)
Moderator: Megan Johnson *(State University of New York at Fredonia)*

15.1 A Data Driven Analysis of Consumer Heterogeneity in the Causal Effects of Interest Rate Changes in Economic Behavior

Authors: Phalguni Unmesh Mahajan *(State University of New York at Buffalo)*

When the central bank announces changes in the interest rate, not all households ride the wave uniformly. Instead, the adjustment in monetary policies represents a heterogeneous mosaic in consumer responses influencing activities like consumption, debt management, savings and investments. This study aims to evaluate the disparity in the impact of the interest rate tweaks on the economic behavior of different consumer segments. Previous research has investigated how rate changes affect economic behavior at an aggregate level. This study bridges the gap by performing a granular analysis of the wide-ranging reactions to better aid in the targeted cognizance of policymakers and businesses. We combine the macroeconomic time series data with the consumer level datasets to assess how monetary policy shifts affect investments, savings and consumption in diverse segments. The study employs econometric models like Difference-in-differences and Instrumental Variable to separate any exogenous impact on the interest rates. To enhance the stakeholder engagement, interactive dashboards from Power BI were employed to display the causal linkage between pre and post policy trends. To address the non-linearities in the data and enhance our feature selection we augment the econometric models with machine learning techniques. Our findings demonstrate significant disparities in consumer responses. For example, middle aged households tend to save on their disposable income and delay their investments during rate hikes as compared to the younger customers. These findings can therefore enable businesses to optimize their offerings per consumer and policymakers can better design targeted interventions.

15.2 Report on Crude Death Rate from Drug Poisoning and Related Variables

Authors: Alexandra Gillispie *(University at Buffalo)*
Alexandra Stathopoulos *(University at Buffalo)*
Janelle Valentine *(University at Buffalo)*

This study explores the relationships between how time, age, and gender affect the crude death rate due to drug poisoning mortality within the United States. The dataset being examined contains records of mortality due to drug poisoning across the United States from 1999-2018. The analysis employs two-sample t-tests, a correlation test, and an ANOVA test to examine the incidence of deaths from drug poisoning and how crude death rates differ across the factors of time, gender, and age. Results indicate that more recent

years (2009-2018) are significantly associated with higher crude death rates due to drug poisoning. Similarly, the results when age was tested against crude death rates indicated that there is a significant difference in overall crude death rates across age groups, specifically the population of ages 35-54. Lastly, the results affirmed our hypothesis that males have significantly higher death rates from drug poisoning than females. These findings indicate the importance of understanding factors that contribute to higher rates of deaths from drug poisoning, and ultimately the importance of this issue. Future research could benefit from a dataset that includes the type of drug poisoning when considering the overall risk of death from drug poisoning.

15.3 An Analysis of Energy Sector Allocations within the U.S. Powergrid

Authors: Robert G Cutlip *(Fairmont State University)*
Rebecca Giorcelli *(Fairmont State University)*
Saunders *(Fairmont State University)*
M Goff *(Fairmont State University)*
X Lopez *(Fairmont State University)*
D Messner *(Fairmont State University)*
Isaac Guentert *(Fairmont State University)*

The purpose of this inquiry was to analyze the United States powergrid to characterize the current state of energy sector allocations and projected changes for the near term. Historical sector data was analyzed, and forecast models were generated in Coal, Nuclear, Wind, Hydroelectric, Petroleum, Fuel Cell, Geothermal, Solar, and Natural Gas sectors. Descriptive analysis showed the percentages of the primary sectors contributing to the powergrid were Natural Gas (43.1%), Nuclear (18.6%), Coal (16.2%), and Hydroelectric (10.2%). Forecasting models were generated for each energy sector based on historical data through the year 2035. Prescriptive analytics were generated to support recommendations for a more resilient and climate friendly grid. Recommendations include expanding nuclear power and equipping nuclear facilities to revert to coal-fired operation in case of uranium supply disruptions; expanding renewable energy sources, such as wind and solar, to reduce natural gas-fired generation; optimizing existing hydropower assets through infrastructure retrofitting and operational enhancements; exploring alternative hydropower solutions, such as in-stream turbines and pumped storage facilities, to expand capacity; promoting domestic production of renewable energy components through tax incentives and grants; and, strengthening trade policies to prioritize domestic sourcing. The grid should be modernized by upgrading aging infrastructure and implementing smart grid technologies. Lastly, cybersecurity measures should be enhanced to protect critical infrastructure assets from cyber threats and attacks. Final recommendations for a more

balanced grid suggest energy sector re-allocations to Nuclear (30%), Natural Gas (28%), Hydroelectric (15.2%), Coal (10%), and Solar (9%) by year 2035.

Undergraduate Poster Session

Poster 1: Testing the Golden Ticket Syndrome: are Pronatal Financial Policies Effective?

Authors: JaeRyoung Lee *(State University of New York at Fredonia)*
Mentor: Dr. Adam Cook *(State University of New York at Fredonia)*

In a report released on January 6th 2023 by the Korean Economics Institute of America, economist Randall Jones addressed the low youth employment rate in the Republic of Korea. His analysis focused on the competitive nature of secondary education and tertiary education enrollment, and the cultural pressure for the youth to secure employment in a company for social status, coining the term "Golden Ticket Syndrome". In this paper, I will test if the Golden Ticket Syndrome affects fertility rates. Indicators commonly associated with the syndrome, such as youth unemployment rates, tertiary education attainment rates, prevalence of a caste system, and composition percentage of services in total GDP. Upon building the model to estimate fertility rates, I will then use data from Japan and Korea to test the effectiveness of government sponsored pronatalist financial aid policies. The results of this analysis will give insights into the non-medical factors that affect the motivators behind child birth.

Poster 2: A Model Framework for Automating the Recruitment Process

Authors: Jyoti Singh Rojariya *(Alfred University)*
Mentor: Ayush Sengupta *(Alfred University)*

The primary aim of this paper is to propose a framework for evaluating the compatibility between a candidate's personality and a company's culture. The framework is designed to assess the alignment of an individual's traits with both the broader company culture and specific team cultures. This evaluation is based on the interaction between three key dynamics: company culture, team culture, and personality. We conceptualize company culture as a set of metrics that are organized across various levels to provide a detailed and multifaceted framework. In this framework, we distinguish between general company culture-reflecting organizational-wide characteristics-and team culture, which is tailored to the unique dynamics of specific teams within the organization. The personality dynamic encompasses individual traits, linked to the aforementioned cultural metrics to enable a quantitative assessment of their compatibility. This paper builds on existing literature on organizational culture and personality traits to develop a more precise and measurable approach to aligning candidates with company values and team dynamics. The core

objectives of this study are to (1) establish measurable metrics for different types of company culture and team culture, (2) propose a flexible framework for exploring the relationship and assessing the fit between the dimensions of company culture and team culture, and individual personality traits.

Poster 3: The Overall Effect of Education on Income Inequality

Authors: Jaynul Abedin *(Alfred University)*
Mentor: Ayush Sengupta *(Alfred University)*

Income inequality in the US has been increasing as the population grows rapidly. With the growth of the population, the economic imbalance is also increasing, which is impacting the national budget as well. According to economic researchers, one of the main reasons behind this is the lack of proper education or not having the willingness to get an education, leading to income inequality, and poverty. Some other potential factors that can contribute to income inequality are health conditions, physical activity, and the pay gap based on gender. My paper will conduct thorough research on the effects of education on income inequality where the independent variable is going to be education and income inequality as the dependent variable. In addition, other control variables or independent variables that I am going to take into account are unemployment, physical inactivity, alcohol consumption, and the gender pay gap. Throughout this paper, I was able to demonstrate that there is a significant impact of education on income inequality. However, the control variables that I selected for this paper were needed to make the impact statistically significant which means that those variables significantly affect income inequality as well.

Poster 4: Order Accuracy Through Order Number Sticker Systems

Authors: Kloee Yanni *(State University of New York at Fredonia)*
Prince Ntor *(State University of New York at Fredonia)*
Mentors: Dr. Lisa Walters, Dr. Lipika Arif *(State University of New York at Fredonia)*

This research project examines how technological innovations can improve order accuracy at Company T, addressing its competitive disadvantage relative to its direct competitor, Company D. By implementing an order number ticket/sticker system, this study explores how such technology can streamline the order fulfillment process, reduce human error, and minimize revenue loss due to inaccurate orders. A comparative SWOT analysis of Company T and Company D reveals key areas where Company T lags behind and how technology can enhance operational efficiency, improve employee satisfaction, and boost revenue generation. Additionally, quality management analytics are used to identify significant quality complaints and root causes. Prioritization matrices guide the

development of targeted recommendations to address these root causes and improve the overall quality control processes. Ultimately, this research highlights the potential financial and operational benefits of adopting technological advancements, including cost savings and improved customer satisfaction, through more efficient order processing and reduced errors.

Poster 5: Decoding Crime: a Data-Driven Investigation of in Los Angeles
Author: Molly Sheehan *(State University of New York at Fredonia)*
Mentor: Dr. Megan Johnson *(State University of New York at Fredonia)*

Does a love of math correlate with a love of true crime? In this paper-it does! Inspired by a love of true crime shows, we sought to investigate trends in criminal activity, in an attempt to predict future crime patterns. This research was initiated as part of an honors thesis course, focusing on criminal activity reported in Los Angeles between 2020-2023. Utilizing a comprehensive public dataset composed of various types of crime, this project aims to identify key features correlated with the solvability of criminal cases. Based on analysis in R, findings suggest a significant relationship between when a crime occurs and its solvability, as well as other variables such as location, victim characteristics, severity, crime type, and length of time until report. By using data analytics, this project contributes to a deeper understanding of crime, providing insights that law enforcement can use to enhance solvability strategies in Los Angeles and beyond.

Poster 6: The Impact of TikTok and Short-Form Videos on Gen Z Communication Styles in the Workplace

Author: Jude Lord (State University of New York at Fredonia)
Mentor: Dr. Lisa Walters (State University of New York at Fredonia)

This independent study examines the impact of TikTok and short-form video platforms on Gen Z communication styles within the workplace. As the first generation to grow up entirely in the digital era, Gen Z brings unique communication preferences shaped by platforms like TikTok-favoring brevity, informality, visual elements, and trend-based storytelling. These emerging styles often diverge from traditional workplace norms, presenting both opportunities and challenges for organizational communication. The study aims to explore which elements of TikTok's format-such as hashtags, humor, visual storytelling, and rapid information delivery-are being integrated into Gen Z's workplace interactions. It also investigates whether these styles foster or hinder effective communication, especially in multigenerational settings where differing expectations can lead to misunderstandings. The research is conducted through a multi-method approach, including a literature review on generational communication differences and digital media influence, a content analysis of at least 30 TikTok videos related to workplace

communication, and sentiment analysis of 30 critical incidents found in public forums like LinkedIn, Reddit, and Glassdoor. Key tools include Pareto charts and visual sentiment graphs to identify dominant themes and emotional responses. Findings from this study will offer insight into the most prevalent TikTok-inspired communication behaviors in professional environments and how they are perceived across generations. Recommendations will include strategies for organizations to leverage Gen Z's communication strengths-such as clarity, creativity, and engagement-while also ensuring professionalism and mutual understanding across teams. These insights aim to help employers foster inclusive, adaptable communication cultures that bridge generational gaps and enhance collaboration.

Poster 7: Innovations in Data Analytics for Accessible Communication: Leveraging Computer Vision for Real-Time ASL Fingerspelling Recognition

Author: Bex Piede *(State University of New York at Fredonia)*
Mentor: Dr. Megan Johnson *(State University of New York at Fredonia)*

According to the Americans with Disabilities Act (ADA), equal access to public services is required without additional charge. For members of the Deaf/Hard of Hearing (HoH), accessibility services are an afterthought. Coupling ignorance with scarcity, either minimal interpreting services are provided or available on a regular basis. Furthermore, deaf students and employees are not typically given accessibility services including interpreters nor offered alternative communication options. According to language consulting firm Nimdzi, "This limits the deaf person's access and hinders their ability to work at their full potential." Developing an open source computer vision (CV) program which outputs accurate English translations given video input of American Sign Language (ASL) fingerspelling will greatly improve accessibility, ease of use, and increase utilization of interpretation services. In conceiving of the program, it became clear that there are insufficient, and frankly inadequate, open source datasets available for CV training. Hence, our first step will be to collect handshape images from a large, diverse group of individuals. Using those images, we will train a computer model to recognize ASL handshapes. In further research, we will produce Python code to generalize the model to one which accepts video inputs with the ultimate goal of reversing the translation process with use of generative AI.

Poster 8: Sales Analysis of Coffee Shop

Author: Lauren Green *(State University of New York at Fredonia)*
Mentor: Dr. Reneta Barneva *(State University of New York at Fredonia)*

The goal of this project is to examine a coffee shop's sales data in a way that addresses the type of coffee sold and seasonality of consumer preferences. Looking at sales data helps to

define trends with respect relating to coffee and key factors that drive consumption of coffee. Understanding consumer behavior and market dynamics gives useful insights. In practice, these insights serve as a guide for an entrepreneur or a small shop's owner thinking and exploring to start up a coffee shop business.

Poster 9: Analysis of Nigerian Car Sales

Author: Habib Yusuff *(State University of New York at Fredonia)*
Mentor: Dr. Reneta Barneva *(State University of New York at Fredonia)*

This project analyzes car sales in Nigeria, specifically with respect to several factors, including make and model, color preference, and price points. By understanding these factors, the project should allow us to identify important trends and factors affecting the automotive market in Nigeria. The information gathered from the analysis will benefit car dealers interested in exporting cars to Nigeria. The study will help them re-align their sourcing strategies to meet market demands and choose from inventory selections to be more competitive.

Poster 10: Digital vs. Physical Music Sales in the U.S.

Ashlyn Dugdale *(State University of New York-Fredonia)*
Mentor: Dr. Reneta Barneva *(State University of New York at Fredonia)*

The project offers a comprehensive analysis of both physical and digital music sales, examining diverse questions to uncover trends and insights. It evaluates which music genre is most popular in sales and examines if digital downloads or physical formats, like CDs and records, are more profitable for artists. The study identifies the top five streaming sources in the US and states with the highest digital and physical music sales in 2024. It also highlights the top three selling US artists and the most streamed daily tracks worldwide, offering valuable guidance for artists. Ultimately, this analysis empowers artists to strategically succeed.

Poster 11: Cosmetology Customer Service Analysis

Author: Brayton Tripi *(State University of New York at Fredonia)*
Mentor: Dr. Reneta Barneva *(State University of New York at Fredonia)*

This project provides a comprehensive analysis of a cosmetic salon’s customer service data. It investigates several key aspects, including contact information, service type and timing, assigned employee, pricing details, service duration for kept appointments, customer feedback, ratings, and product recommendations. Additional factors such as membership status, loyalty points, customer preferences, appointment reminders, and the clients’ geographic state are also examined. The insights derived from this multifaceted

analysis equip prospective entrepreneurs with valuable guidance on customer management, service optimization, and strategic planning, ultimately aiding in the successful launch and sustainable operation of a salon business.

Poster 12: Spotify Global Streams Analysis

Author: Jash Parekh *(State University of New York at Fredonia)*
Mentor: Dr. Reneta Barneva *(State University of New York at Fredonia)*

This project utilizes Spotify Global Streams data to gain actionable insights into the ever-evolving music streaming industry. It analyzes the top five artists based on monthly listeners while separating out the top countries by streaming volume, and by counting downloads from free or paid subscribers. It identifies the artists and albums that are best suited for targeted promotional activity, compares genres by total streams, and the genres with the highest percentage of skips. It also looks to identify genres likely to experience growth based on the latest trending activity. A breakdown of listener activity over time is evaluated to provide a detailed overview of changing consumer behavior, which ultimately facilitates insight into promotional strategies in today's music market.

Poster 13: AI Market Worldwide 2020–2030

Author: Billie Coddington *(State University of New York at Fredonia)*
Mentor: Dr. Reneta Barneva *(State University of New York at Fredonia)*

This project analyzes the AI market as one of the fastest growing markets available, with applications in all aspects of life, from speech recognition to image processing and self-driving cars. New technology development and investment have resulted in a rapid growth rate. The AI market is a complete ecosystem which packages and combines software, hardware, and various services that empower all organizations to build and develop modern AI applications. Therefore, by analyzing the market landscape, this project provides information that will help stakeholders think strategically about how to best position their organizations to take advantage of the changes and diverse industry potential applications of AI development.

Poster 14: Analyzing Construction Project Specifications

Author: Abdullah Yusuf *(State University of New York at Fredonia)*
Mentor: Dr. Reneta Barneva *(State University of New York at Fredonia)*

The goal of this project is to provide an analysis of construction data and useful insights for the design process which helps with the construction documentation processing.

Poster 15: Behind a Business

Author: Carissa Shanahan *(State University of New York at Fredonia)*
Mentor: Dr. Reneta Barneva *(State University of New York at Fredonia)*

Poster 16: We study a dataset, which offers comprehensive data on different transactions in a business from 2017 to 2023 and answer a number of research questions. The data set contains commissions expense, consulting expense, cost of goods sold, marketing expense, payroll expense, R&D expense, sales, software/hardware expense, and travel & entertainment expense.

Poster 17: Online Retail Store Analysis

Author: Taylor Ingrao *(State University of New York at Fredonia)*
Mentor: Dr. Reneta Barneva *(State University of New York at Fredonia)*

The goal of this study is to provide company-wide sales analysis, offer strategic insights into product performance, geographic trends, consumer behavior and profit distribution. Using over 1,000 transactions across various cities and states, this analysis supports decision-making for marketing, inventory planning, and operational optimization.

Poster 18: Analysis of Chocolate Sales Worldwide

Author: Donevan Rotmans *(State University of New York at Fredonia)*
Mentor: Dr. Reneta Barneva *(State University of New York at Fredonia)*

This project analyzes the chocolate sales that differ worldwide, including most popular sales by country and what amounts they are purchasing at. The idea of this project is to point out the factors of the business side of chocolate to educate and allow businesses and distributors to take advantage of. This includes marketing and targeting for certain countries and products of chocolate they should be selling.

 ISSN 1681 8997

Analytics Career Fair 2025

Thursday, May 8th, 1:00pm – 4:00pm
Williams Center G138 Blue Lounge

This year the IBAC is introducing the Analytics Career Fair. The Career Fair is an opportunity for participants, particularly current students, to network with potential employers recruiting candidates in the analytics and data science fields. The Career Fair will run from 1:00 to 4:00 p.m. on Thursday in the Blue Lounge, a side room off of the Williams Center MPR.

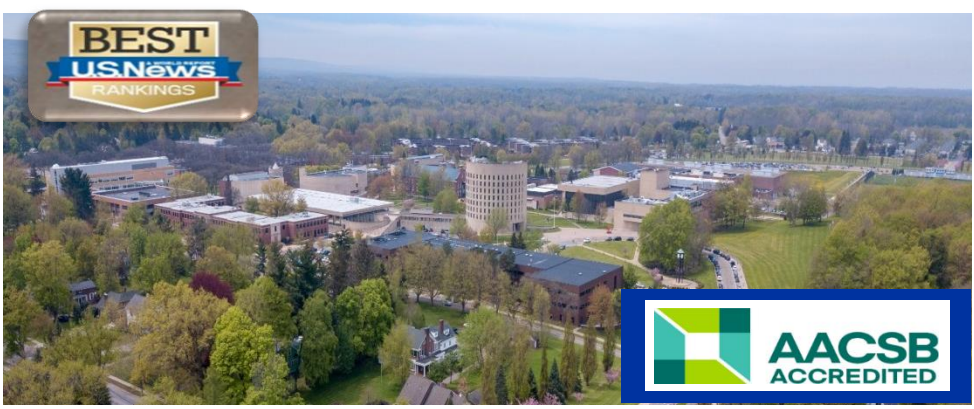
Candidates are encouraged to bring resumes and dress professionally. This year, industries such as health care, education, finance, and logistics are represented, as well as graduate level programs in the field.

Fredonia School of Business

The Fredonia School of Business prepares future business leaders by providing the knowledge, skills, and real-world experience necessary to compete in a global business environment. We offer a wide range of business programs to prepare you for success in Accounting, Public Accountancy, Finance, Management, Marketing, Business Administration, Music Industry, Sport Management, and Economics. All programs in the School of Business will introduce you to the latest technology, make you aware of the global and cultural components of business, and provide you with experiential learning and practical internship opportunities. Our curriculum emphasizes critical thinking, communication skills, business ethics, and professionalism.

The School of Business sustains the tradition of excellence at Fredonia, which is consistently ranked among the finest public universities in the Northeast -- and among the most affordable options in higher education. Small class sizes lead to close interaction with our faculty and the opportunity to conduct research at the undergraduate level - a hallmark of a Fredonia education. In addition, all students have opportunities for internships, available locally, regionally and across the nation.

Our students have access to the Center for Innovation & Economic Development (CIED), which supports start-up companies committed to Western New York. The Center is a New York State certified business incubator operated by SUNY Fredonia. Formerly called the Fredonia Technology Incubator (FTI), the CIED provides entrepreneurs with access to work-ready space, business consulting, mentoring and training, professional services, assistance with business formation, networking opportunities, and connections to University resources and student interns. Students work with entrepreneurs on market research, business plan development, operations and record keeping, and unique problem-solving initiatives. The Center expands the entrepreneurial aspects of our curriculum, enhances internship and employment possibilities, and inspires students to start businesses of their own.



 ISSN 1681 8997

Thank you for your support and participation.
Visit us again!

 ISSN 1681 8997

Volume 24 Number 1 January 2025 ISSN 1681-8997

Empirical Economics Letters

Eric Brice Ngatsi Fosso and Donal Ajoumessi Houmpe
Effect of Sanitary and Phytosanitary Measures on Cocoa Exports: Does the Quality of Governance Matter?

Shreya Roy and Bibek Ray Chaudhuri
Artificial Intelligence and its Impact on Skilled-Unskilled Wage Gap

Haifa Hammami and Wissem Bouaziz
Do Good Corporate Governance Practices and Financial Performance in Banks Matter in Reducing the Occurrence of Stock Market Crises? Empirical Evidence from Tunisia

Mark Funk
Liquidity, Thresholds, and R&D Investment

Nishi Malhotra and Priya Malhotra
Relationship between Green Finance and Climate Change: An Empirical Study

Nabaneeta Biswas and Poulomi Dasgupta
Untapped Potential: The Spillover Effects of College Merit Aid Programs on Secondary School Outcomes

J Vineesh Prakash
Profitability Determinants of Manufacturing Firms in India: The Case of Pharmaceutical and Electronics Industry

Yue-Duan Guan
Characteristics of Audit Committee Members and Audit Fees: Empirical Evidence from Taiwan

Shyafina Soleha Mohd Afendi, Sarimah Surianshah and Ho Chong Mun
Effects of Macroeconomic Factors on Household Savings: Evidence from Different Income Level Development

Prabuddha Sanyal and Mark A Ehlen
Bi-Directional Causality between Home Prices and Financial Stresses using Bayesian Methods: Evidence from the United States

Bishal Lamichhane, Joseph Palardy and Aniket Kumar Singh
The Nuances of Large-Language-Model-Agent Performance in Simple English Auctions

Yen-Ju Chiang
Corporate Social Responsibility and Stock Return Predictability

Sri Mulyati
The ESG-ROA Nexus within Enterprise Value in ASEAN Countries

Li-Yu Chen, Juei-Chi Chang and Chi-Feng Wang
Board Tenure Diversity and the Performance of IMA: The Moderating Effects of Director Ownership

Sunil Sapra
Dynamic Semiparametric Regression Modelling with an Econometric Application

Shanxia Sun, Haidi Han, Kangjuan Lv and Wuyang Hu
Hybrid or Full Electric? Choices of Chinese Consumers with Different Environmental Concerns, Openness to Lifestyle Changes, and Pro-technological Attitudes

Managing Editor

Mohammad A Wadud

Board of Editors

Aman Agarwal

George E Battese

William F Bowlin

Damodar N Gujarati

Clem A Tisdell.

Empirical Economics Letters

ISSN 1681 - 8997

Papers on all aspects of empirical economics and business are invited for publication in ***Empirical Economics Letters (EEL).*** Authors are requested to adhere to the instructions that follow.

Instructions to Authors*

1. Papers for publication must be in English and should be sent to Dr. Mohammad A. Wadud, Managing Editor, by **e-mail as attached files in MS Word** to the email address: **wadud68@yahoo.com**
2. Papers should contain original unpublished work. All papers will be refereed and the decision of the Board of Editors on acceptability of papers will be final.
3. Authors must supply their email address. All correspondences are made through **email only.**
4. The first page of the paper should contain (a) title, (b) name(s) of the author(s), (c) affiliation(s), and (d) an abstract of not more than 100 words followed by 3 to 5 keywords.
5. The paper should not normally exceed 2500 words.
6. Footnotes should be limited to explanatory purpose and kept to a minimum. Diagrams and tables should be put at the right places.
7. All references should be alphabetically arranged according to the following style:

Published papers

Aigner, D.J., C.A.K. Lovell and P. Schmidt, 1977, Formulation and Estimation of Stochastic Frontier Production Function Models, Journal of Econometrics, 6, 21-37.

Research papers

Ajibefun, I.A., Battese, G.E. and Kada, R., 1996, Technical Efficiency and Technological Change in the Japanese Rice Industry: A Stochastic Frontier Analysis, CEPA Working Papers, No. 9/96, Department of Econometrics, University of New England, pp. 22.

Book

Greene, W.H., 1997, Econometric Analysis, International Edition, Third Edition, Prentice-Hall International, Inc.

---

* For details, please visit the journal's website at: http://eel.my100megs.com

Editorial Policy

Empirical Economics Letters (EEL) invites papers on all aspects of empirical economics. The *EEL* is a fully refereed international journal. There is no specific deadline for submission of papers. Papers can be sent any time and will be considered for the subsequent issues. All papers must have an empirical content and the theoretical models on which the empirical results are based should be clearly, but briefly stated.

The *EEL* is published every month and has format of a letters journal, that is, papers will be short and have a definite size restriction.

In deciding to publish this journal, we have taken into account the difficulties faced by both the authors and the readers, and have carefully tried to remove them. Our approach is outlined below.

Quick acknowledgement The *EEL* sends acknowledgement within 48 hours of the receipt of a paper.

Expeditious evaluation The *EEL* expeditiously completes evaluation process and communicate to the author its decision within eight weeks from the receipt of the paper.

Quick publication The *EEL* publishes all accepted paper within six months of their acceptance.

Minimum sending cost The *EEL* receives papers only through an email attachment thus bringing the cost to a minimum.

Submission fee The *EEL* has decided to require **no submission fee,** but we request our authors of accepted papers to contribute a small amount of production cost to defray the processing, reviewing, maintaining website, printing and other expenses depending on the size of the papers.

Subscription fee Most journals have high subscription rates. The *EEL* charges a yearly subscription of US$ 400.00 or its equivalent for institutions and US$300.00 for individuals for rich countries (per capita income greater than US$ 2000) and US$ 300.00 or its equivalent for institutions and US$200.00 for individuals for other countries. This is much lower than those charged by journals of the same standard and frequency as ***Empirical Economics Letters.***

Because of the author friendly approach of the journal it is expected that publishing in ***Empirical Economics Letters*** becomes very competitive. **Clarity, brevity and nontriviality** are the three things that determine acceptability of a paper.

We urge all scholars to remember the ***Empirical Economics Letters*** if they have any clear, short and nontrivial research work to submit. We also urge all readers of the *EEL* to give any suggestion that may help it.