Homepage

The International Conference of Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics is an annual gathering of experts, researchers, and professionals from around the world who share a passion for advancing the fields of artificial intelligence, blockchain, cloud computing, and data analytics. The conference provides a platform for knowledge exchange, networking, and collaboration in these rapidly evolving domains.Our conference is dedicated to exploring the latest research, trends, and best practices in artificial intelligence, blockchain, cloud computing, and data analytics. We seek to create an atmosphere of learning, sharing, and innovation where experts can come together to exchange ideas and collaborate on new projects. At our conference, attendees can expect to hear from a variety of thought leaders, industry professionals, and academics who are at the forefront of their fields. We offer keynote speeches, panel discussions, and technical sessions covering a wide range of topics, from machine learning and natural language processing to distributed ledgers and decentralized applications.

Media Type	Part Number	ISBN
XPLORE COMPLIANT	CFP24UK0-ART	979-8-3503-6845-1

2025 International Conference on Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics (ICoABCD) took place 1-2 December 2025 in Lombok, Indonesia.

IEEE catalog number: CFP25UK0-ART
ISBN: 979-8-3315-5749-2

Organized by Telkom University
Prepared by EDAS Conference Services.

2025 International Conference on Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics (ICoABCD) | 979-8-3315-5749-2/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICOABCD67551.2025.11470809

Time (Makassar)	Elsewhere
Monday, December 1	
10:30 am-10:50 am	
10:50 am-12:50 pm	1A: Session 1A
12:50 pm-01:10 pm	
01:50 pm-03:50 pm	2A: Session 2A 2B: Session 2B 2C: Session 2C 2D: Session 2D 2E: Session 2E
03:50 pm-04:10 pm	
Tuesday, December 2	
10:05 am-12:05 pm	3A: Session 3A 3B: Session 3B
12:05 pm-12:25 pm	3C: Session 3C 3D: Session 3D

Monday, December 1 10:30 - 1:10

1A: Session 1A

10:30 ***Multivariate Spatial Modelling of Tree Species Distribution Using Adaptive Sparse Group Lasso: Case Study in Barro Colorado Island Plot***

Refki Assadiky (Institute Teknologi Sepuluh Nopember, Indonesia); Achmad Choiruddin (Institut Teknologi Sepuluh Nopember Surabaya, Indonesia); Santi P Rahayu (Institut Teknologi Sepuluh Nopember, Indonesia)

Multivariate spatial point patterns are gaining popularity as technology enhances in large-scale data collection. This study examines tree species distribution in the Barro Colorado Island (BCI) forest plot, involving 9 tree species and 13 environmental covariates. To examine how species co-occur with their environment and potentially each other, we model the data by an inhomogeneous multivariate Poisson point process (IMPP). We use the Random Quadrature method for efficient parameter estimation, which aligns with generalized linear models while minimizing redundant computations. Given the extreme high density of covariates, we apply the adaptive sparse group lasso (ASGL) for variable selection, allowing the model to focus on the most influential environmental variables per species while retaining group structures if any among variables. Finally, hierarchical clustering is applied to cluster species according to their estimated environmental responses. This is to uncover patterns in habitat preference and ecological similarity among species within the BCI plot.

10:50 ***Development of a RISC-V Microprocessor Architecture with Integrated UART and VGA Control Modules***

Munawar A Riyadi (Diponegoro University, Indonesia); Galih Aprian Triansyah (Diponegoro University, Semarang, Indonesia); Hisyam Ahmad, Teguh Prakoso and Erwin Adriono (Diponegoro University, Indonesia)

This study proposes the design of a microprocessor based on the RISC-V architecture, integrated with Universal Asynchronous Receiver Transmitter (UART) and Video Graphics Array (VGA) control interfaces. The primary objective is to develop an open, modular, and efficient processor system. RISC-V is selected for its open-source and customizable nature, enabling flexibility in embedded system development. The UART interface facilitates low-speed serial communication between the processor and external devices, while the VGA controller module enables direct graphical output to display units. The system is implemented using Hardware Description Language (HDL) and simulated on a Field-Programmable Gate Array (FPGA) platform to evaluate its functionality, logic efficiency, and real-time performance. Experimental results demonstrate that the system successfully executes 40 base RV32I instructions and 7 multiply/divide extension instructions in accordance with the RV32IM standard. The system also supports synchronous serial communication via UART and graphical output through a VGA controller, enabling the development of integrated and adaptive embedded systems. The physical design of the microprocessor, implemented using Cadence tools, shows that at an operating frequency of 50 MHz, the processor occupies an area of 0.41 mm^2 and consumes 2.9624 mW of power, or 59 µW/MHz, on a 45 nm technology node.

11:10 ***Novel LLM-Based Detection of Cloud Attacks: Ranking Attack Techniques and Building a Log Dataset Framework***

Harun Gulec (Yildiz Technical University, Turkey & Meta Platforms Inc., United Kingdom (Great Britain)); Amac M. Guvensan (Yildiz Technical University, Turkey)

The rapid migration of information systems to cloud platforms has accelerated investments while also increasing the scale and diversity of attacks targeting these environments. Although existing studies highlight differences from traditional systems or analyze specific attack types, a key gap remains: which attack techniques are most frequently exploited in practice. To address this, we propose a reusable scoring method

for prioritizing attack techniques and a flexible dataset generation framework. We produced a dataset consisting of 316,518 log records and 42,000 labeled attack logs with the help of this framework. Furthermore, using this dataset, we developed an LLM-based anomaly detection system with tailored prompt rules, achieving 92% accuracy and 98% precision. The contributions of this study are twofold: identification and prioritization of common cloud attack techniques, and provision of a log dataset framework with a sample dataset and novel detection method for academic and industrial use. Future work will focus on validating the approach in real-world environments and expanding the dataset with new attack categories and log sources.

11:30 ***Non-Transferable Anonymous Identifiers in DID/VC Framework***

Kosuke Otsu and Takeshi Ogawa (Tokyo Denki University, Japan)

In the DID/VC framework, the holder of personal information (credentials) can only disclose the minimum number of credentials associated with the holder's DID for each verifier. However, if multiple verifiers collude to aggregate credentials associated with the same DID, a problem arises where credentials beyond the holder's intention may be identified. To prevent this, the W3C recommends that holders use multiple DIDs for each recipient of the credentials. However, the problem of transferring a DID and its corresponding private key to another person, allowing that other person to impersonate the holder and falsify credentials, remains unresolved. In this paper, we propose a method to realize anonymous DIDs where the verifier can verify that the corresponding private key is non-transferable or extremely difficult to transfer. The proposed method is realized by combining Key Blinding technology, which can generate multiple unlinkable public keys corresponding to one private key, and BBS signature scheme that allows selective disclosure. We also show the results of calculations using a sample program to demonstrate that the DID cannot be transferred.

11:50 ***Development of a Cloud-Integrated Multi-Sensor Landslide Early Warning System (LEWS)***

Muhammad Mukhlisin, Hany Windri Astuti, Roni Apriantoro and Aiun Hayatu Rabinah (Politeknik Negeri Semarang, Indonesia); Muhamad Yusuf (Politeknik Negeri Cilacap, Indonesia); Adhy Kurniawan (Vocational School, Universitas Gadjah Mada, Indonesia); Arif Sumardiono and Erna Alimudin (Politeknik Negeri Cilacap, Indonesia); Fakih Irsyadi (Chiba University, Japan & Universitas Gadjah Mada, Indonesia)

Landslides are among the most critical natural hazards in tropical and mountainous regions, often triggered by prolonged or high-intensity rainfall. This study developed and evaluated a cloud-integrated multi-sensor Landslide Early Warning System (LEWS) capable of real-time monitoring and web-based visualization by accuracy tests of the sensor and the monitoring website integration. The LEWS integrates multiple environmental sensors regarding landslides, including rainfall, pore-water pressure, soil moisture, groundwater level, and ground movement, all connected to a cloud server. Each sensor was calibrated and tested under laboratory conditions to determine its measurement accuracy. The experimental results show that the rainfall sensor achieved an average accuracy of 91.7%, the pore-water pressure sensor reached 97.95%, the groundwater level sensor attained 99%, the soil moisture sensor demonstrated 90.67% accuracy, and ground movement with 95% accuracy. These results confirm that all sensors meet the operational requirements for geotechnical and hydrological monitoring. The developed web-based platform, hosted at http://ewspolines.com, enables real-time data acquisition, visualization, and system configuration through a dashboard accessible to both administrators and general users. The integrated system successfully synchronized field data with the cloud, providing continuous access via laptops and mobile devices. Overall, the study demonstrates that the LEWS effectively combines accurate sensing hardware with cloud computing technology to deliver real-time monitoring and decision support for landslide-prone regions.

12:10 ***A Lightweight Consensus Mechanism for Preventing Fraudulent MEV in Blockchain***

Kiyora Kodama and Takeshi Ogawa (Tokyo Denki University, Japan)

In traditional blockchain systems, the contents of transactions (Tx) are publicly disclosed before finalization, allowing block proposers to exploit this information and gain unfair advantages through Tx insertion and removal attacks (commonly referred to as fraudulent maximum extractable value (MEV) attacks). Existing countermeasures such as Proposer-Builder Separation (PBS) and Shutter Network suffer from limitations including insufficient protection for transaction originators and reliance on trusted nodes. This paper proposes a lightweight and trust-minimized consensus mechanism that introduces two new transaction types: reservation transactions (Re-Tx) and reserved execution transactions (RE-Tx). Re-Tx deterministically fixes the execution order prior to content disclosure, preventing insertion attacks. RE-Tx is revealed only after Re-Tx is confirmed, and its exclusion is prevented through a novel voting-based recovery mechanism. The proposed method minimizes computational and communication overhead and maintains consensus security even if up to one-third of all voting nodes collude, and is compatible with existing block-based ledger systems. Quantitative analysis demonstrates its effectiveness in mitigating fraudulent MEVs without reducing performance or consensus stability.

12:30 ***Comparison of Process Efficiency Between GPT and Claude for Backend API Code Generation Using GitHub Copilot Agent Mode***

Tio Haidar Hanif, Dana Sulistiyo Kusumo and Nungki Selviandro (Telkom University, Indonesia)

The use of AI has become widespread, and one area that has been significantly impacted is application development. GitHub Copilot is one of the AI-based tools that can simplify application development and improve efficiency in the development process. However, the efficiency of GitHub Copilot also depends on the AI model it uses. This study aims to quantitatively compare the efficiency of the GPT 4.1 and Claude 3.7 Sonnet AI models using GitHub Copilot in Agent Mode, with a case study focused on developing a backend REST API application. The study employs a comparative approach by conducting an experiment to develop several features in parallel across each model. The metrics measured being the number of prompts and the duration required by each prompt to complete each task. The conclusion of this study shows a correlation between speed and the model used for code generation. In this research, GPT 4.1 demonstrates the efficiency of this model, as seen in the shorter duration required by the model and the fewer prompts needed. This model is appropriate for tasks that are relatively simple or want to generate code incrementally rather than all at once. However, Claude 3.7 is more effective in handling larger code generation requests in a single prompt, this model suite for complex task and want to generate code complete at once. Keywords-LLM4SE, large language model, code generation, software engineering, Rest API

Monday, December 1 10:50 - 12:50

1B: Session 1B

10:50 ***Cost-Performance Analysis of OLTP Database Workloads on ARM and x86_64 Public Cloud Environment***

Muhammad Ananta Faturrahman (Telkom University, Indonesia); Mega Pranata (Telkom University, Indonesia & Okayama University, Japan)

This paper presents a comprehensive performance evaluation of ARM-based AWS Graviton3 and x86_64-based AMD EPYC processors for Online Transaction Processing (OLTP) workloads in cloud computing environments. Using MariaDB as the database platform and SysBench for benchmarking, we conducted controlled experiments comparing AWS c7g.large (ARM) and c7a.large (x86_64) instances across varying concurrency levels from 1 to 192 threads. Our evaluation encompasses three distinct workload patterns: read-only, write-only, and mixed read/write operations. Results demonstrate that ARM-based instances consistently outperform x86_64 counterparts in low-concurrency scenarios, achieving up to 22.4% higher transaction throughput in single-threaded read operations. Additionally, ARM instances provide approximately 29.4% cost savings compared to equivalent x86_64 instances, resulting in superior cost-performance ratios for most OLTP

scenarios. These findings suggest that ARM-based cloud instances represent a viable and economically advantageous alternative for transaction-intensive database workloads.

11:10 ***Assessing DFIR Readiness for Data Breach in Cloud-Based Organizations: A Self-Assessment Instrument Aligned with Indonesia's PDP Law***

Fathoni F Irawan, Niken Cahyani and Rio Guntur Utomo (Telkom University, Indonesia)

Indonesia's Personal Data Protection (PDP) Law No. 27 of 2022, Article 46, specifically imposes strict obligations on organizations to enhance their Digital Forensics and Incident Response (DFIR) readiness in the event of a breach. Prior models, such as DFMM, DiFRI, and CR-SAT, address forensic readiness or cyber resilience separately. Still, none integrate both dimensions during the incident phase in cloud environments under explicit regulatory obligations. This study develops and validates a self-assessment instrument for evaluating DFIR maturity in cloud-based organizations, explicitly aligned with PDP Law requirements. The instrument integrates ISO/IEC 27043 and the Cloud Incident Response Framework (CIRF) into a regulation-tailored maturity model. Content validity, assessed by three subject matter experts using the Content Validity Index (CVI), achieved excellent results (S-CVI = 1.00). A pilot study at PT XYZ indicated an overall DFIR maturity of Level 2 (Managed), indicating the organization operates at a managed capability level but lacks formal standardization to reach the defined level. These findings confirm both the applicability of the instrument and the need for improved standardization, documentation, and consistency to meet regulatory expectations. Academically, the study extends the DFIR readiness literature with a PDP-aligned instrument, while practically, it offers organizations a structured approach to assess maturity and strengthen compliance with Article 46.

11:30 ***Ensemble Machine Learning Models for Intrusion Detection in Cloud Infrastructure for Cybersecurity***

Sreenivasulu Gajula (Fidelity Investments, USA)

The number of assaults is growing along with the number of computer networks. To safeguard network infrastructures against more sophisticated breaches, intrusion detection systems (IDS) have become indispensable in the era of cloud computing. This study uses the current UNSW-NB15 dataset, which covers a wide range of attack methods and shows severe class imbalance, to construct a robust and high-performance IDS that can effectively identify different cyber assaults. The suggested approach includes one-hot encoding, managing missing values, and considerable data preparation, standardization with StandardScaler, and balancing using the ADASYN technique followed by an ensemble learning approach combining Convolutional Neural Networks (CNN), XGBoost, and Random Forest (RF) classifiers. Each model is optimized via hyperparameter tuning, and their predictions are aggregated using majority voting to improve detection accuracy. The ensemble model achieves outstanding results with 99.40% accuracy, 99.64% precision, 99.31% recall, 99.47% F1-score, and 99.42% ROC-AUC, significantly outperforming individual models and other state-of-the-art methods such as LSTM, DNN, CNN-BiLSTM, LR, and DT. The significance contribution of this study lies in effectively addressing the data imbalance challenge and combining complementary models to enhance generalization and reliability in detecting both common and rare intrusion types.

11:50 ***A Comparative Analysis of Security Baselines Across Kubernetes Distributions***

Dave Andrew Nathaniel and Muhammad Edo Syahputra (Bina Nusantara University, Indonesia)

Kubernetes is the standard for container orchestration, with various distributions offering different security features. This study provides a comparative evaluation of four key distributions-Vanilla Kubernetes, Talos Linux, RKE2, and OpenShift-using the CIS Kubernetes Benchmark v1.11.0 and OS-level security assessments to analyze default security configurations and compliance. Talos Linux achieved the highest compliance at 87.2%, driven by its immutable, minimal OS and tightly integrated Kubernetes stack. RKE2 followed with 81.1%, while OpenShift and Vanilla Kubernetes scored 69.8% and 65.1%, respectively. Common issues included incomplete audit logging, weak cryptographic settings, and inconsistent admission controls. These findings emphasize the

importance of OS design and default configurations in Kubernetes security, offering practical guidance for selecting and securing distributions in production.

12:10 ***Advanced Artificial Intelligence-Based Cybersecurity Intrusion Identification System for Cloud Computing Environments***

Hari Prasad Badiginchala (VISA Inc (Cybersource), USA)

The rapid growth of cloud computing has increased reliance on scalable infrastructure but also raised significant cybersecurity challenges, particularly in intrusion detection. Real-time threat detection, handling large cloud settings, and identifying novel assaults are challenges for traditional intrusion detection systems (IDS). By increasing accuracy, automating anomaly detection, and adapting to evolving threats, machine learning (ML) has shown significant promise in strengthening IDS capability. This paper suggests a hybrid DL-based IDS that uses the advantages of bidirectional Long Short-term Memory (Bi-LSTM) networks and Convolutional Neural Networks (CNN) for temporal dependency learning. Using the CSE-CICIDS2018 benchmark dataset, supported by preprocessing and feature optimization techniques, the proposed framework achieves exceptional performance with 99.92% accuracy, 99.96% precision, 99.94% recall, and 99.98% F1-score. The CNN-BiLSTM architecture outperforms both conventional ML approaches and DL models in comparison, especially when it comes to detecting minority intrusion classes and decreasing false positives. This work highlights the novelty of combining CNN and Bi-LSTM architectures, offering a reliable and trustworthy solution for next-generation AI-driven cybersecurity systems.

12:30 ***UI/UX and Functional Design of a Mobile-Based Image and Audio Watermarking System***

Fauzan Maulana, Gelar Budiman and Linda Meylani (Telkom University, Indonesia)

The increasing demand for digital content protection highlights the need for accessible and user-friendly tools to embed and verify watermarks. This paper presents the UI/UX and functional design of a mobile application focused on image and audio watermarking. The system is designed to balance technical robustness with intuitive usability, enabling users to perform watermarking tasks efficiently on mobile devices. Through structured interface layouts, simplified user flows, and responsive design principles, the application aims to ensure ease of use for both technical and non-technical users. This study provides a detailed overview of the application's user interface, core features, and system flow, serving as a design reference for future development in mobile-based watermarking tools. The results include an objective analysis of performance and resource usage, as well as a subjective user satisfaction survey, both of which confirm the system's efficiency, responsiveness, and usability.

Monday, December 1 10:50 - 1:10

1C: Session 1C

10:50 ***A Three-Stage Microflow-to-EFSM Transformation Approach for Test Coverage Enhancement in Mendix***

Jaatsiyah Maulidina Astrianto, Rosa Reska Riskiana and Dana Sulistiyo Kusumo (Telkom University, Indonesia)

Automated testing is essential for ensuring reliability in Low-Code platforms such as Mendix, which emphasize rapid and visual development. However, Mendix's existing Microflow-based testing remains limited to unit-level validation. This study introduces a three-stage transformation from Microflow to Extended Finite State Machine (EFSM) to improve automation scalability. The stages include core logic conversion, user interface (UI) integration, and rule optimization, all implemented using TestOptimal. Results show an increase in test coverage from 15 to 44 cases with high precision and recall for backend logic, while achieving full traceability in the optimized model. Although UI and database-related operations remain partially unsupported, the findings

confirm that EFSM-based Model-Based Testing (MBT) effectively extends Mendix testing toward integration-level automation.

11:10 ***Cloud-Enabled Human Resouce and Payroll System to Enhance Operational Efficiency in Micro and Small Enterprises***

Sinung Suakanto and Fikri Haykal (Telkom University, Indonesia); Joko Siswanto (Bandung Institute of Indonesia, Indonesia); Jan M Pawlowski (Ruhr West University of Applied Sciences, Germany)

Micro, Small, and Medium Enterprises (MSMEs) often face challenges in managing human resources and payroll due to limited administrative capacity, reliance on manual processes, and the absence of affordable and context-appropriate digital solutions. This gap results in inefficiencies such as inaccurate attendance tracking, errors in salary calculation, and low transparency, which may reduce employee satisfaction and operational effectiveness. To address these issues, this study develops a web-based Human Resource and Payroll Information System specifically tailored to the operational needs of MSMEs. The research employs the Design Science Research (DSR) methodology, encompassing problem identification, objective definition, system design and development, demonstration, evaluation, and communication. The proposed system incorporates features including leave management, attendance dispute resolution, and automated reporting, providing an integrated and scalable solution. Evaluation using Black Box Testing and User Acceptance Testing (UAT) demonstrated high functional accuracy (100%) and user satisfaction (84.6%). This study contributes to the digital transformation of MSMEs by offering a validated, user-centered, and scalable system that enhances workforce administration, reduces manual workload, and improves operational efficiency.

11:30 ***Blockchain-Based NFTs for Digital Preservation and Global Knowledge Dissemination of Shu Brocade Cultural Heritage***

Lei Xu (Universiti Teknologi MARA, Malaysia & Chengdu University, China); Muhamad Fairus Kamaruzaman and Rusmawati Ghazali (Universiti Teknologi MARA, Malaysia)

Traditional weaving crafts such as Shu brocade face challenges of inheritance disruption and limited global dissemination, while conventional preservation methods struggle to meet the demands of the digital era. This study aims to explore the application model of blockchain-based NFT technology in the digital preservation and global knowledge dissemination of Shu brocade cultural heritage. Employing a case study methodology, the research focuses on the Scrolling Grass and Butterfly Pattern of Shu brocade and constructs a technical framework of "cultural investigation → digital preservation → NFT creation → global dissemination." Digital archives were established through high-resolution acquisition, image processing, and vectorization reconstruction. NFT products with copyright protection and provenance authentication were created on the Polygon blockchain and OpenSea platform based on the ERC-1155 standard. Educational effectiveness was validated through a pilot study involving 40 students from creative disciplines. Results indicate that this approach effectively achieves permanent digital preservation of Shu brocade cultural heritage. The decentralized architecture of blockchain and global platform deployment significantly enhance international accessibility to cultural knowledge resources. The educational pilot demonstrates positive potential in improving learning engagement, promoting cultural knowledge acquisition, and fostering critical thinking. This study provides a replicable technical solution for the digital preservation and knowledge dissemination of Shu brocade and similar cultural heritage, offering technical support for achieving the United Nations Sustainable Development Goals.

11:50 ***Smart Contract Implementation in Web-Based E-Voting Systems: Blockchain Security and Transparency Analysis***

Bryan Liem, Vincent Nicholas Hie and Zulfany Rasjid (Bina Nusantara University, Indonesia)

This study addresses the critical challenges of security, transparency, and integrity in traditional e-voting systems by proposing a novel, blockchain-based solution. We design and implement a web-based e-voting prototype using smart contracts on the Ethereum blockchain, leveraging technologies including React, MetaMask, and Hardhat. While previous research has explored blockchain for e-voting, many studies lack a comprehensive security analysis of the underlying smart contracts. Our key contribution lies in a rigorous, multi-faceted security analysis that validates the system's robustness. We employed static analysis with Slither, achieving a 100% statement, function, and line coverage through unit testing, and conducted extensive fuzz testing with Echidna, which reported no invariant violations. The results confirm the system's ability to ensure voter anonymity, prevent double-voting, and maintain the immutability of the ballot ledger, directly addressing the vulnerabilities often found in conventional systems. This research not only provides a practical, verifiable solution for secure e-voting but also establishes a robust methodology for auditing the security of smart contracts in high-stakes applications.

12:10 ***Design and Evaluation of Blockchain-Enabled Payment System Using API Integration for Enhanced User Experience***

Jerry D Lucas (National University, Philippines); Joseph R. Del Carmen (Nueva Ecija University of Science and Technology, Philippines); Edward Allen B. Manaloto (National University & NU Clark, Philippines); Lester G. Loyola, Carl Joseph R David, Harwin C. Mendoza, Ronielle B. Antonio and Mark Lagman (National University, Philippines)

This study designed and tested an online payment system based on blockchain technology with a built-in REST application programming interface (API). The main objective of this research was to enhance the user experience, security, and transparency of the Nueva Ecija University of Science and Technology (NEUST) institutional payment processes. In the iteration of the system development and testing, the Agile Software Development Methodology was utilized. This methodology allowed the system to satisfy the users' specified requirements. The functional adequacy, usability, reliability, security, maintainability, and portability of the system were tested based on the ISO/IEC 25010:2011 Software Product Quality Standards. The results were high acceptability and efficiency, and therefore the implementation of REST API technology and blockchain technology reduces the number of user interactions without sacrificing security and transparency. This study provides a model that may be used to suit the requirements of organizations that would like to adopt user-driven blockchain-based online payment systems.

12:30 ***Comparative Analysis of AI Models for Detecting Online Gambling Promotion from Trending YouTube Shorts' Comments in Indonesia***

Aliyah Jasmine Saliano, Bryan Santosa, Risya Syafira, Diana Diana and Shania Priccilia (Bina Nusantara University, Indonesia)

Advertisements for online gambling have become increasingly prevalent in the comment sections of YouTube shorts, often using obfuscated language and special characters to evade automatic detection. This poses risks to users and violates the platform's policies, highlighting the urgent need for efficient and proper detection to solve this issue especially within trending Indonesian Youtube content. This study aims to compare the performance of various classical machine learning and transformer-based models in identifying online gambling promotion comments in trending Indonesian YouTube Shorts. A dataset of trending Indonesian YouTube Shorts comments was collected using the YouTube Data API v3. Preprocessing included normalization, character mapping, and removal of noise and the comments were then labeled using human-in-the-loop approach, assisted by Mistral-7B-Instruct-gguf for faster labeling. The Feature extraction was done using IndoBERT-base-uncased embeddings and was evaluated by a collection of models namely Bernoulli Naive Bayes and Random Forest Classifier representing machine learning approach and IndoBERT and RoBERTa representing the transformer-based approach. From this study, IndoBERT achieved the highest

performance with F1-score of 94%, outperforming both traditional models and RoBERTa, likely due to its pretraining on Indonesian corpora. These findings suggest that language-specific deep learning models are more effective for detecting region-specific gambling content. Future work can explore extending the model to other platforms like YouTube videos (non-shorts), Tiktok, and Instagram Reels.

12:50 ***ViT-GPT2 Model in an Audio-Based Image Captioning System for Visually Impaired Accessibility***

Muthia Rihadatul Aisyi, Aji Gautama Putrada and Ikke Dian Oktaviani (Telkom University, Indonesia)

Digital innovation systems for visually impaired accessibility have been widely developed. However, there is a research opportunity to implement image captioning and text-to-speech (TTS) that are more optimal than the state-of-the-art. This study aims to evaluate the audio-based vision transformer-generative pre-trained transformer 2 (ViT-GPT2) for an image captioning system in visually impaired accessibility with the aid of TTS. We utilize the COCO dataset as training data for image captioning. Performance is evaluated using the bilingual evaluation understudy (BLEU), the metric for evaluation of translation with explicit ordering (METEOR), the Recall-Oriented Understudy for Gisting Evaluation-Longest common sequence (ROUGE-L), and the consensus-based image description evaluation (CIDEr). Then, the evaluation results are compared with the bootstrapped language-image pretraining (BLIP) baseline model. The test results show that ViT-GPT2 excels in five key metrics: BLEU-1, BLEU-2, BLEU-3, METEOR, and SPICE. The values are 0.69, 0.51, 0.36, 0.24, and 0.18, respectively. Then, BLIP outperforms ViT-GPT2 in BLEU-4, ROUGE-L, and CIDEr metrics, namely 0.26, 0.51, and 1.05, respectively. BLEU-1, which demonstrates superiority in terms of vocabulary accuracy, suggests that ViT-GPT2 is more optimal than BLIP, as accuracy is considered more important than fluency. The results of the WER and latency comparison indicate that GTTS is more optimal than Coqui TTS for the speech generator in our visually impaired accessibility solution, with values of 0.011 and 0.039 seconds, respectively.

Monday, December 1 10:50 - 12:50

1D: Session 1D

10:50 ***A Random Forest-Based Model for Customer Churn Prediction in a Telecommunication Company***

Yutika Amelia Effendi (Universitas Airlangga, Indonesia); Nania Nuzulita (Airlangga University, Indonesia); Wahyu Damar Iswara (Universitas Airlangga, Indonesia)

In today's fast-paced digital landscape, telecommunication companies generate massive volumes of data at high velocity. The abundance of competing service providers offering more affordable and attractive packages has intensified market competition, making customer retention critical for business sustainability. Customer churn, often driven by service dissatisfaction, can significantly impact revenue, particularly in service-oriented sectors like telecommunications. To mitigate this risk, the ability to accurately predict churn is essential. This study investigates the use of the Random Forest algorithm for customer churn prediction by analyzing historical customer data to identify churn-related patterns. Random Forest is selected for its strong classification performance, resistance to overfitting, and capability to assess feature importance. A structured data preprocessing pipeline was employed, including class balancing, feature reduction, categorical encoding, and hyperparameter tuning. Among four training configurations, the model trained on a reduced dataset without hyperparameter tuning achieved the highest accuracy of 84.8%. These findings confirm the effectiveness of Random Forest for churn prediction and highlight the importance of feature selection in improving model performance.

11:10 ***Chain-of-Memory: Taming Hallucinations in Million-Token Models with Verifiable Retrieval and Causal Tracing***

Jasmin Jarsania and Vivek Patel (USA)

Large language models (LLMs) with million-token context windows exhibit severe hallucination when processing extensive documents, fabricating information that contradicts source material. We introduce Chain-of-Memory (CoM), a novel framework that combines verifiable retrieval with causal tracing to maintain factual accuracy in long-context scenarios. CoM employs three key mechanisms: (1) memory-anchored reasoning that explicitly grounds each inference step to retrieved passages, (2) causal intervention probes that identify hallucination-prone attention patterns, and (3) a verification protocol that validates generated content against source material. Through extensive experiments on the LongBench and InfiniteBench datasets, we demonstrate that CoM reduces hallucination rates by 73.2% compared to standard prompting while maintaining 94.1% of baseline performance on accuracy metrics. Our analysis reveals that hallucinations primarily emerge from attention dilution in the middle regions of long contexts, which CoM effectively mitigates through structured memory retrieval. We release our implementation and evaluation suite to facilitate further research in trustworthy long-context language understanding.

11:30 ***Image-to-Image-Based Wave Forecasting Using Pix2pix Approach, Case Study in Pacitan, Indonesia***

Hendri Bhirowo Dwi Hananto and Indra Ardhanayudha Aditya (PT PLN, Indonesia); Didit Adytia (School of Computing, Telkom University, Indonesia); Andry Alamsyah (Telkom University, Indonesia)

Accurate wave forecasting is essential for supporting maritime operations, particularly in coastal regions where high waves often disrupt activities such as shipping, loading and unloading, and port accessibility. The southern coast of Java, including the waters around Pacitan, is especially vulnerable to high wave conditions influenced by wind seas and swells, posing challenges for operations such as coal transportation to PLTU Pacitan. To address these challenges, this study proposes a wave downscaling framework based on an image-to-image translation approach using the Pix2pix model. The model is trained on high-resolution SWAN simulations performed with a nested grid system, enabling the transformation of coarse-resolution global model outputs into high-resolution local forecasts. The experimental results demonstrate that the choice of grid resolution and training data length significantly affects model performance. Among the tested configurations, the 3×3 degree grid yielded the most accurate results overall, achieving the highest correlation coefficient, the lowest RMSE, superior CSI values, and the smallest FAR. Furthermore, extending the training dataset to 5 years provided the lowest RMSE and balanced predictive performance, despite only marginal differences in other evaluation metrics compared to shorter training periods. Overall, the findings confirm the effectiveness of the Pix2pix framework in enhancing coastal wave forecasting, offering improved accuracy and spatial representation in the Pacitan coastal domain. This study highlights the potential of image-to-image deep learning approaches in coastal oceanography and suggests that further research involving model tuning, hybrid architectures, or integration with additional physical constraints could further improve forecasting accuracy and reliability.

11:50 ***Integrated Monitoring System with LSTM-Based Energy Load Prediction for Smart Building Management***

Naufal A Rajabi and Sinung Suakanto (Telkom University, Indonesia); Ahmad Musnansyah (Telkom University Bandung, Indonesia); Suhono Harso (Indonesia)

Modern building management requires high operational efficiency and security but is often hindered by fragmented monitoring systems. In many newly constructed buildings, energy and Closed-Circuit Television (CCTV) systems operate separately, limiting the operational team's ability to perform holistic monitoring and respond promptly to incidents. This study addresses this gap by developing a comprehensive framework with two main components. First, a reliable RESTful API architecture integrates energy consumption data (voltage, ampere, kWh) and CCTV activity logs into a single platform, solving the problem of data silos. Second, a Long Short-Term Memory (LSTM) model is designed and trained on historical energy consumption data to enable proactive energy load forecasting. Evaluation results indicate that the proposed model achieves a Root Mean Squared Error (RMSE) of 3.45, demonstrating high predictive accuracy to support operational planning. In conclusion, this integrated and predictive framework transforms building management from a reactive

monitoring approach into an intelligent, proactive process. The main contribution of this research lies in providing a practical, end-to-end solution that combines real-world system integration with advanced predictive analytics, offering a replicable roadmap for smart building management.

12:10 ***Trustworthy Multimodal Fraud Detection with Federated Learning and Computer Vision***

Dendy Pramudito and Jufriadif Naam (Universitas Nusa Mandiri, Indonesia); Ferda Ernawan (Universiti Malaysia Pahang Al-Sultan Abdullah, Malaysia)

Fraud detection in financial systems is still very difficult to do due to a vast class imbalance, how quickly fraudsters change their tactics and stringent privacy laws. This study interprets the XClever framework which is based on gradient fatigue and that embraces multi-modality and interpret-ability. These modifications include gradient boosting (XGBoost), attention-based gated recursive unit (GRU-Attention), cost-sensitive thresholding, federated learning (FL) as well as computer vision (CV) extensions. This study employed a benchmark credit card dataset in this experiment. While GRU-Attention had a higher recall and better accuracy, XGBoost had the best precision and ROC-AUC (0.985). A late-fusion strategy can find a middle ground between these trade-offs, which could lead to the lowest predicted loss in costly calibration. The use of federated simulations shows that decentralized training with FedAvg preserves recall with only a slight decrease in ROC-AUC (0.01-0.02), while differential privacy further protects sensitive data. Convolutional Neural Networks (CNNs) trained on feature grids and recurrence plots attain a ROC-AUC of approximately 0.96, with visual representations elucidating supplementary structural anomalies. SHAP explanations identify V14, V12, and V4 as primary predictors, ensuring transparency and compliance with regulatory standards. The results indicate that an effective fraud detection system necessitates a comprehensive workflow that incorporates imbalance-aware learning, multimodal representation, privacy-preserving collaboration, and explainability to guarantee a reliable and scalable implementation within the financial services ecosystem.

12:30 ***Region-Aware (RAw) 3D U-Net for the Segmentation of Small and Boundary-Sensitive Multiple Sclerosis Lesions***

Umayal Venkatasamy (The University of Auckland, New Zealand); Alan Wang (Auckland Bioengineering Institute, New Zealand); William Schierding (Ophthalmology, New Zealand); Eryn Kwon (Auckland Bioengineering Institute, New Zealand); Helen V Danesh-Meyer (Ophthalmology, New Zealand); Samantha Jane Holdsworth (Medical and Health Sciences, New Zealand); Catherine Shi (Data Science and Artificial Intelligence, New Zealand)

Multiple Sclerosis is an autoimmune condition that leads to lesion formation in the brain and spinal cord, which can be visualized and assessed using MRI. Recently, deep learning models have demonstrated substantial improvements in the automated segmentation of MS lesions from MRI scans. This work introduces a novel approach featuring multi-scale feature extraction, both spatial and channel attention modules, and region-specific processing streams to more accurately detect a wide range of lesion types and anatomical distributions. The model is trained using a combination Dice-Focal loss and extensive data augmentation, while inference leverages region-adaptive post-processing, including tissue-based masking and minimal lesion size filtering. In evaluations on an internal dataset (n=245, FLAIR images with expert lesion masks), a held-out test cohort (n=25) achieved, after removal of clear outliers, a mean Dice coefficient of 0.6826, mean sensitivity of 0.652, and mean positive predictive value (PPV) of 0.8524; the corresponding medians were Dice 0.7668, sensitivity 0.683, and PPV 0.944, with 18 of 25 cases exhibiting Dice scores ≥0.6. Training curve analysis showed stabilization around Dice ≈0.77. These results reflect high precision with balanced sensitivity, supporting the value of learning region-dependent lesion patterns from FLAIR images alone.

1E: Session 1E

10:50 ***Hybrid Blockchain-Based Presidential Election System to Increase Vote Efficiency and Security***

Inayatulloh Inayatulloh (Bina Nusantara University, Indonesia)

The presidential election is a process of voting by the people to elect the president and vice president of a country. This process involves the active participation of citizens in determining the national leader who will lead the government for a certain period. The purpose of the presidential election is to realize democracy and ensure that state power comes from the sovereignty of the people by electing leaders through majority votes. Several countries have used information technology to support presidential elections to make the election process easier. The problems that arise from conventional information technology-based election systems are data manipulation and non-transparent vote counting processes that raise doubts about the presidential election system. Several studies offer solutions using blockchain technology. Blockchain technology has several advantages, such as peer-to-peer validation, which guarantees transparency of the vote counting process, and distributed storage that guarantees the security of vote counting data. However, research on the application of existing blockchain technology is limited to conceptual models that are difficult to understand and implement. Therefore, this research aims to build a presidential election prototype using Ganache. The selection of hybrid blockchain in research allows the presidential election system already owned by the government to be used for blockchain-based presidential elections. Research uses a qualitative approach through literature review to study the weaknesses of conventional information technology-based and blockchain technology-based presidential election systems. The latest research is a prototype of a presidential election using a hybrid blockchain and simulation with Ganache.

11:10 ***MetaCropZ: Microservices-based Architecture for AI-blockchain Convergence in Smart Agriculture***

Gopi K Akella (CQUniversity, Australia); Santoso Wibowo (Central Queensland University, Australia); Srimannarayana Grandhi (CQUniversity, Australia); Sameera Mubarak (University of South Australia, Australia); Fariza Sabrina (Central Queensland University, Australia)

Smart agriculture increasingly requires solutions that deliver real-time analytics alongside verifiable data integrity. Artificial Intelligence (AI) enables rapid insights for tasks such as crop disease detection and yield prediction, while blockchain technology provides transparency, auditability, and decentralized trust. Yet, most existing systems treat these technologies separately, failing to leverage the complementary capabilities of these technologies. This paper introduces MetaCropZ, a microservices-based architecture that systematically integrates AI-driven insights and blockchain-based traceability through layered components: presentation, microservices, intelligence, blockchain, and storage, linked by a unified API gateway. Analytical results such as disease outbreak classifications are processed through secure microservices, allowing authorized users to initiate on-chain transactions; corresponding event metadata is reliably stored off-chain via the InterPlanetary File System (IPFS). A working prototype demonstrates the feasibility of structured AI-blockchain convergence through secure microservices orchestration, role-based smart contract execution, and seamless dual-token lifecycle management. These architectural innovations confirm that modular integration enables scalable, real-time, and trustworthy analytics for next-generation agri-food systems.

11:30 ***Tokenized Credit System for Edible Food Waste Redistribution and Fiscal Incentives***

Gilbert Christian Caesario Siburian (School of Economics and Business; Telkom University; Bandung-Indonesia); Herry Irawan and Andry Alamsyah (School of Economics and Business; Center of Excellence SAKTI; Telkom University; Bandung-Indonesia)

Efficient management of edible food waste (EFW) requires systems that are transparent and auditable to safeguard social and fiscal accountability while supporting food security and sustainable development goals. Yet current practice remains fragmented and reliant on manual reporting, which hampers traceability, sustains risks of miscalculation and double-counting, and leaves non-refundable tax credits illiquid when liabilities fall below the incentive value. This study designs and validates a blockchain-based EFW tokenization model that

automates policy-compliant issuance, circulation, and retirement of credits with verifiable audit trails, while measuring transaction confirmation performance as an operational benchmark. The method combines business process mapping, an integrative literature review, stakeholder interviews, and expert validation, followed by model design and proof-of-concept evaluation in a local EVM environment to measure transaction stability. Results from 200 sequential mint transactions show an average confirmation time of 4.1906 seconds (σ = 0.0171s) without degradation, validating performance consistency under load. In parallel, FIFO allocation and on-chain event logging yield a deterministic audit trail that prevents double-counting. These findings provide an operational basis for tax credit reconciliation and accountable redistribution governance, indicate readiness for deployment in low-cost EVM ecosystems such as Polygon, and offer practical guidance for policy design and broader field trials.

11:50 ***DeFi 3.0 Risk Taxonomy and Pathways to Financial Resilience***

Gerry Nevile Kurnia, Irni Yunita and Andry Alamsyah (Telkom University, Indonesia)

Blockchain-based decentralized finance (DeFi) is a major financial innovation, enabling transparency and inclusion through programmable rails. The transition to DeFi 3.0-defined by cross-chain interoperability, multichain ecosystems, and tokenized real-world assets (RWAs)-broadens functionality yet introduces potential systemic vulnerabilities. Prior research often treats protocol exploits or single risk families in isolation, leaving no unified lens connecting DeFi risks to financial resilience. This study develops a unified DeFi 3.0 risk taxonomy and maps it to resilience capacities. Using a three-lane systematic literature review (peer-reviewed, grey literature, preprints; 2021-2025; 42 sources), we identify twelve risk domains in three categories: technology and data infrastructure; market and economic; and governance, legal, and operational. We then assess resilience along three capacities-absorptive (stablecoins, automated market makers/AMMs, insurance), adaptive (regulatory alignment, RWA tokenization, AI integration), and transformative (transparency, inclusion, ESG alignment). The resulting framework operationalizes resilience theory via this taxonomy, providing a structured reference for regulators, developers, and scholars to support innovation while strengthening systemic stability.

12:10 ***Integrating Post Quantum Cryptography into Bitcoin Sidechains: A Simulation Based Study***

Chol Hyun Park and Leopoldo Misael Ocas Olguin (University of the Ozarks, USA)

The arrival of quantum computing poses a huge threat to conventional public key cryptography used in blockchain systems such as Bitcoin. To address this challenge, we proposes and evaluates a quantum resistant sidechain framework that integrates post quantum digital signature schemes and quantum key distribution (QKD) protocols with the Bitcoin mainchain. Using a Python simulation environment, the performance of multiple signature algorithms. ECDSA as a baseline measurement, Falcon, CRYSTALS-Dilithium, and SPHINCS+ were analyzed in combination with quantum communication protocols BB84, E91, and SARG04. Experimental results show that ECDSA remains the fastest baseline scheme but lacks quantum resistance, while SPHINCS+ provides the highest security with an expected overhead of 93.5%. Among quantum protocols, BB84 achieved the best overall efficiency across transaction sizes. The optimal integration pairs were BB84 and CRYSTALS-Dilithium combination for speed, security balance and SPHINCS+ and E91 for maximum quantum resilience. These findings demonstrate the practical feasibility of deploying post quantum cryptographic components in Bitcoin compatible sidechains, paving the way for future blockchain networks that remain secure in the post quantum era.

12:30 ***Digitizing Sharia-Based Cooperative Loan Management: A Framework Integrating Sharia Compliance, Ethical Audits, and Smart Contract***

Sinung Suakanto, Syfa Nur Lathifah, Muharman Lubis, Fauzan Victorino Jais and Akbar Affaruk Khuzaimi Ahmadani (Telkom University, Indonesia)

Sharia-based savings and loan cooperatives play a vital role in ethical financial inclusion, but their operations are hampered by inefficiencies, manual processes, and a critical lack of integrated digital governance to ensure verifiable ethical compliance and transparency. This operational gap undermines member trust and hinders the realization of higher ethical objectives. This study proposes a comprehensive framework for digitizing Sharia-based cooperative loan management that integrates Sharia compliance mechanisms, ethical auditing processes, and smart contract process. The research employs the Design Science Research (DSR) methodology to design, develop, and evaluate a Sharia-compliant digital loan management framework that enhances operational efficiency while reinforcing accountability and transparency. The core artifact includes the Compliance Engine, Smart Contracts, and an Ethical Audit Dashboard designed to automate checks and secure data integrity. The system successfully digitized loan management processes, significantly improving efficiency and compliance monitoring while reducing manual intervention and processing time. Furthermore, the integrated framework demonstrated systemic validation of cooperative financial activities against Sharia principles and ethical audit standards, safeguarding both compliance and financial protection for members. The system was validated with a "Good" System Usability Scale (SUS) score of 77.5. This research contributes a practical, technologically advanced model for digitizing Sharia-based cooperative loan management, explicitly aligning operational efficiency with Sharia compliance, ethical auditing, and transparent governance, providing a blueprint for sustainable and accountable cooperative practices.

Monday, December 1 1:50 - 3:50

2A: Session 2A

1:50 ***Implementation of the Long Short-Term Memory (LSTM) Algorithm for Web-Based Book Recommendations: A Case Study on the Google Books API***

Bella Wahmilyana Asril, Tegar Arifin Prasetyo, Maria Pangaribuan, Carlos Marpaung and Cecilia Situmorang (Institut Teknologi Del, Indonesia)

Reading is an effective way to enhance knowledge and broaden perspectives, yet the rapid growth of digital information often results in an overwhelming number of book options, making it difficult for users to identify relevant content. Recommender systems address this challenge by providing personalized suggestions based on user preferences. Traditional methods such as content-based and collaborative filtering, however, are limited in capturing sequential dependencies and dynamic changes in user behavior. To overcome these limitations, this study proposes a web-based book recommendation system that integrates the Long Short-Term Memory (LSTM) algorithm with the Google Books API. Cosine Similarity was employed to evaluate the quality of recommendations, yielding an average similarity score of 0.9930, which demonstrates the effectiveness of the proposed system. The implementation of this system on a web platform ensures real-time access to diverse and up-to-date book data, thereby improving user experience.

2:10 ***An Edge AIoT Framework with Temporal Convolutional Networks for Site-Specific Fertilizer Recommendation in Rice Farming***

Moh Khairudin Khairudin (Universitas Negeri Yogyakarta, Indonesia); R. Hafid Hardyanto (Universitas PGRI Yogyakarta, Indonesia); Rustam Asnawi (Universitas Negeri Yogyakarta, Indonesia)

Efficient and sustainable fertilizer management is crucial for enhancing rice productivity while reducing environmental degradation caused by excessive nitrogen, phosphorus, and potassium usage. Conventional fertilization practices often rely on intuition and fail to account for site-specific variations in soil nutrients, crop phenology, and local weather conditions. To address this gap, this paper presents an Edge Artificial Intelligence of Things (Edge AIoT) framework that leverages Temporal Convolutional Networks (TCN) for site-specific fertilizer recommendation in rice farming. The framework integrates multimodal IoT sensor data-including soil moisture, nutrient content, pH, and ambient weather parameters-processed directly on low-

power edge devices to ensure real-time operation in rural areas with limited connectivity. The TCN model, implemented with dilated causal convolutions and an attention mechanism, captures both short-term nutrient fluctuations and long-term seasonal trends. To enable deployment on constrained edge hardware, the trained model is optimized through quantization and pruning and converted into TensorFlow Lite for efficient on-device inference. Experiments using simulated IoT sensor datasets reflecting rice field conditions demonstrate that edge-based TCN inference is feasible, achieving competitive accuracy while reducing model size and inference latency compared to baseline methods. These results support the integration of AIoT, advanced time-series modeling, and prescriptive nutrient strategies for sustainable rice production, with future research aimed at large-scale field implementation.

2:30 ***Resilient Routing for Machine Learning Inference During Scale Events***

Ihor Novoseltsev (Lytx, USA)

Machine learning (ML) inference services often scale out/in rapidly on Kubernetes, creating brief availability gaps between actual backend membership and what clients, the Domain Name System (DNS), and load balancers see. These "scale events" cause short bursts of 503/UNAVAILABLE errors and tail-latency spikes. We study resilience for HTTP and gRPC routing under two paths: direct service routing and Application Load Balancer (ALB)/ingress-assisted routing. Our method instruments per-pod traffic share, CPU, and p50/p95/p99 latency during controlled traffic ramps and spikes. For numeric payloads, gRPC with Protobuf and transport compression reduces on-wire bytes and latency. For image payloads, keeping inputs compressed and Base64 encoding the original JPEG/PNG outperforms client-decoded raw tensors packed into Protobuf and yields lower tail latency. On the routing side, ALB/ingress provides fast removal of unhealthy/draining pods, while client freshness determines fast addition of new pods; together they smooth transients and spread load to new replicas within seconds. These techniques are lightweight and practical for production MLOps teams aiming to mitigate scale-event error spikes while maintaining low p95/p99.

2:50 ***Propeller Selection in Electric Propulsion Using Multi-Criteria Sugeno Fuzzy Inference System***

Erwan Eko Prasetiyo, Erwhin Irmawan, Gaguk Marausna, Hendriana Helda Pratama, Rian Adrian and Aulia Dea Fadzila (Sekolah Tinggi Teknologi Kedirgantaraan, Indonesia)

The selection of an appropriate propeller is a critical factor in ensuring the efficiency, endurance, and safety of Unmanned Aerial Vehicles (UAVs). Conventional approaches often rely on single metrics such as maximum thrust or efficiency, which may neglect essential trade-offs involving current draw, aerodynamic constraints, and stability. To address this limitation, this study proposes a systematic propeller selection framework based on a Sugeno-type fuzzy inference system. The methodology integrates three key stages: (1) experimental identification of Brushless DC (BLDC) motor parameters to ensure accurate propulsion modeling, (2) definition of four propeller performance criteria-thrust margin, efficiency, current ratio, and tip Mach number, and (3) fuzzy inference using a 32-rule knowledge base to generate a unified suitability score (FuzzyScore). Fourteen commercially available propellers were evaluated using this framework. The fuzzy system successfully produced a clear ranking of candidates, where propellers exhibiting balanced trade-offs across the defined criteria achieved the highest scores. In particular, the results demonstrated that propellers meeting a thrust factor of 2× hover, while maintaining safe current levels and avoiding excessive tip Mach values, were consistently identified as optimal. The discussion highlighted the robustness of empirical BLDC modeling, the necessity of multi-criteria evaluation, and the advantages of Sugeno fuzzy logic in UAV propulsion design. Overall, the proposed approach offers a practical and computationally efficient decision-support tool, providing UAV designers with a comprehensive framework to select propellers that balance performance, efficiency, and reliability.

3:10 ***AI-Enhanced Wearable Devices for Real-Time Health Monitoring and Early Disease Detection***

Afm Rafid Hassan Akand (Westcliff University, Bangladesh); Abdul Quddus Mozumder (Stanton University, Bangladesh); Afia Fairooz Tasnim (California State University Long Beach, Bangladesh); Abu Hanif (International American University, Bangladesh); Md Abdur Rob (Ohio University, USA); Durga Shahi (Westcliff University, USA); Mohammad Hossain (Graduate Teaching Assistant, USA); Hasan M Sozib (Ahsanullah University of Science and Technology, Bangladesh)

Cardiovascular diseases (CVDs) remain the leading global cause of mortality, yet conventional diagnostic pathways are episodic and clinic-bound. We propose an AI-enhanced wearable framework that bridges static clinical predictors and streaming physiological signals to enable continuous, real-time cardiac risk monitoring and early disease detection. Using a publicly available cardiovascular dataset (1,025 patients, 13 features), we evaluate three baseline classifiers-Logistic Regression, Random Forest, and Gradient Boosting-and two hybrid ensembles-Soft Voting and Stacking-under stratified train/test splits. The held-out set is accurately predicted by Random Forest and Stacking models with a performance of 100 %, and with 99.0 % performance on Gradient Boosting and Voting classifiers, reflecting good predictive quality. Edges embedded deployment on ultra-low power microcontrollers (< 150 ms inference, < 512 kB) memory with preprocessing, feature extraction, and ensemble inference is then described. Anonymized risk scores are pushed into a secure cloud through a bidirectional pipeline to be re-trained periodically, adapting the model to the thresholds to solve the problem of model drift and personalization. We add a complete comparative study on the ensemble methods of heart disease prediction, a deployable wearable device architecture, and look into practical challenges, energy limits, computational efficiency, and real-time alerts in clinical settings. This research paves the way to wearable AI, showing how integrated static and dynamic data, with the edge cloud shared processing, may revolutionize indefinite cardiovascular health monitoring.

3:30 ***Comparison of Machine Learning Models on Balanced and Imbalanced Dataset of Rodent Tuber***

Iwan Binanto (Sanata Dharma University, Indonesia); Nesti Sianipar (Bina Nusantara University, Indonesia); M. Rizky Fajar Mali, Basilius Arilla Dimas N and Ajitama Jaya (Sanata Dharma University, Indonesia)

In this study, we investigate the performance of various machine learning models on both balanced and unbalanced datasets. The data processing pipeline begins with loading and preprocessing the dataset, followed by two parallel approaches: handling balanced and unbalanced data. For the balanced data approach, we employ random undersampling to achieve class balance. The Rodent Tuber's unbalanced dataset, which includes chemical data useful for cancer diagnosis, was used in this study. We use and evaluate Random Forest classifiers and Support Vector Machines (SVM) with three different kernels: sigmoid, linear, and radial basis function (RBF) on balanced and unbalanced datasets. Based on our research, the Random Forest algorithms generated the best classification results on balanced datasets, achieving a total 8.9-second runtime with 99% accuracy, precision, recall, and F1 score. On balanced datasets, however, SVM performance differed greatly between kernels; linear kernels produced 50% accuracy, RBF kernels 61%, and sigmoid kernels 45%. The models' ineffectiveness is shown by their poor performance on unbalanced datasets, which mostly identified the majority class. This emphasize the negative impact of class imbalance. Furthermore, running the algorithm on the imbalanced dataset took significantly longer than on the balanced dataset.

2B: Session 2B

1:50 ***A Sparse Logistic Regression Approach for Lightweight Android Malware Detection***

Siti Nurfajarina Asyifa, Deny Haryadi and Rana Zaini Fathiyana (Telkom University, Indonesia)

The widespread adoption of the Android operating system has made it a primary target for sophisticated malware attacks, posing risks such as data exfiltration, unauthorized surveillance, and persistent background activity. Traditional detection methods, including signature-based and deep learning approaches, often face limitations in scalability, computational cost, and interpretability, especially in resource-constrained mobile

environments. Permissions such as INTERNET, READ_EXTERNAL_STORAGE, WRITE_EXTERNAL_STORAGE, and RECEIVE_BOOT_COMPLETED is frequently exploited for malicious activities, making it a critical indicator for detection. To address these challenges, this study applies Sparse Logistic Regression (SLR), a machine learning algorithm that leverages L1 regularization to automatically select the most relevant features and build a lightweight, effective detection model. The methodology involved data collection, dataset splitting, model training with feature space alignment, and performance evaluation using standard metrics. Six experiments were conducted using different training-to-testing ratios (80:20, 70:30, and 90:10) on both the full permission set and a top 20 permission subset to assess model performance. The best configuration, using a 90:10 split with 17 selected permissions, achieved 94.6% accuracy, 92.9% precision, 98.0% recall, and a 95.4% F1-score. The results show a strong correlation between selected permissions and malicious behaviors, enabling balanced learning without overfitting. These findings demonstrate that combining SLR with sparse feature selection provides an accurate, interpretable, and efficient solution for lightweight Android malware detection.

2:10 ***A Two-Stage Predictive Model for Securities Firms Profitability in Indonesia Based on Market Volatility Using Extreme Gradient Boosting (XGBoost)***

Alfarhad Maulana, Deny Haryadi and Desi Nurnaningsih (Telkom University, Indonesia)

Accurate Net Profit Margin (NPM) forecasting is essential for securities firms preparing initial public offerings (IPOs) and investors evaluating investment opportunities in Indonesia's capital market, which now comprises 955 listed firms. Existing approaches fail to capture non-linear relationships between market volatility and profitability, provide inadequate early distress detection, and do not integrate macro-level volatility indicators with firm-level financial characteristics. This study introduces a novel two-stage XGBoost framework that addresses these limitations by combining distress classification with NPM regression. Stage 1 employs binary classification to predict revenue and net profit negativity using temporal features and volatility indicators derived from IHSG and LQ45 indices. Stage 2 integrates these distress predictions with firm-specific characteristics to forecast quarterly NPM. Using data from 71 Indonesian securities firms spanning 2022 Q1 to 2025 Q2, the methodology includes data preprocessing through winsorizing and Z-score normalization, as well as hyperparameter optimization via GridSearchCV across 5,400 configurations using a composite objective function (60% R-squared, 40% stability metric). The framework achieves 100% test accuracy for revenue negativity prediction (F1-score: 0.9803) and 71.83% for net profit negativity (F1-score: 0.7987) at Stage 1. At Stage 2, it attains an R-squared value of 0.7749 with an RMSE of 0.4478, exhibiting minimal overfitting. This volatility-aware framework offers early warning signals for IPO readiness assessments and risk-adjusted profitability forecasts to inform investment decisions, surpassing traditional econometric and single-stage machine learning models while providing novel insights into profitability forecasting for emerging capital markets.

2:30 ***Bot Account Detection on Twitter-X using Random Forest Algorithm with Feature Selection Based on Binary Particle Swarm Optimization***

Naufalul Fajri, Deny Haryadi and Syifa Nurgaida Yutia (Telkom University, Indonesia)

The rapid growth of social media, particularly Twitter, has led to an increase in the number of bot accounts involved in spreading spam, manipulating public opinion, and other inauthentic activities. This highlights the need for a reliable detection system that can accurately distinguish bot accounts from real users. This study aims to apply a combination of BPSO and Random Forest algorithms to improve the performance of the model in detecting bot accounts. The research stages include applying log transformation to normalize the data distribution on several features such as friends-count, average-tweets-per-day, followers-count, statuses-count, and favourites-count, selecting optimal features using BPSO, and training and testing the model with data split ratios of 70:30, 80:20, and 90:10. The results of the experiment show that logarithmic transformation effectively corrects unbalanced data and improves model stability, while BPSO efficiently selects six relevant features that significantly improve classification performance. The best performance was achieved with an

80:20 split ratio, resulting in an accuracy of 87.61%, precision of 81.76%, recall of 80.64%, and an F1 score of 81.20%. Overall, the combination of logarithmic transformation, BPSO, and Random Forest proved to be efficient in producing an accurate, stable bot detection system with good generalization for new data.

2:50 ***Ai facial recognition for Timekeeping using Yolov8 With ensemble***

Joel T Jayme (University of San Carlos, Philippines); Godwin S. Monserate (Adviser, Philippines)

This study investigates the feasibility of utilizing YOLOv8 (You Only Look Once version 8) for facial recognition with novel datasets. After the global pandemic, there is a pressing need for alternative biometric technologies that minimize physical contact and address hygiene concerns. Traditional fingerprint biometrics carry risks of disease transmission and security vulnerabilities, prompting the exploration of other biometric authentication methods.

YOLOv8, known for its real-time object detection capabilities, is evaluated for its high performance in facial detection and recognition tasks. The research examines YOLOv8's accuracy, precision, and reliability, Also the Researcher will use Stack ensembling and its potential benefits in helping the base Model Detector (YoloV8) with a relatively small dataset and classes preventing it from Generalization as well as its potential benefits in enhancing hygiene and security compared to traditional fingerprint biometrics. The researcher uses a novel dataset using approximately 1000 images in order to test whether or not the YoloV8 with ensemble is capable or better at performing by using this metric mAP ≥ 91.4%, Recall ≥ 82.6%, Accuracy ≥ 94.1% as a baseline for its performance the researcher can reliably compare the result with scientific and mathematical backings. Ultimately, this research contributes to the advancement of biometric authentication technology, offering potential solutions to concerns brought about by the pandemic and the evolving landscape of cybersecurity threats.

3:10 ***AI-Driven Governance Using Case-Based Reasoning for Strategic Project Control***

Tao An and Thacha Lawanna (Chiang Mai University, Thailand)

Strategic project control often suffers from a lack of adaptive decision-making at critical governance points, particularly in complex and dynamic environments. Traditional models such as Waterfall, PRINCE2, Stage-Gate, and Earned Value Management (EVM) rely on static rules, predefined workflows, and retrospective metrics, limiting their responsiveness and agility. This paper proposes an AI-driven governance framework based on Case-Based Reasoning (CBR) to address these limitations. The model follows a five-step CBR cycle (case representation, retrieval, adaptation, revision, and retention) and integrates this process into project governance workflows to enable contextual, real-time decisions based on past experiences. To evaluate its effectiveness, five datasets from the UCI Machine Learning Repository, including COCOMO and Online Retail, were adapted into governance cases. Comparative analysis against traditional models was conducted using four evaluation criteria: efficiency, accuracy, competency, and operational time. Results show that the AI-CBR model outperformed all traditional baselines, achieving an average accuracy of 0.91, a competency score of 0.95, and reducing operational decision time by nearly 45%. These outcomes confirm the model's effectiveness in enhancing governance quality, adaptability, and speed. The contributions of this research are threefold: (1) the development of a learning-based governance model; (2) a practical decision-support algorithm applicable across domains; and (3) empirical validation using standard datasets. This study demonstrates that AI-CBR integration holds significant promise for modernizing strategic project governance and improving decision outcomes in high-uncertainty environments.

2C: Session 2C

1:50 ***G-CRI: A Cross-Modal Framework for Quantifying and Mitigating Cultural Bias in Vision-Language Models***

Yeni Dwi Rahayu (Universitas Muhammadiyah Jember, Indonesia & Universitas Muhammadiyah Yogyakarta, Indonesia); Slamet Riyadi (Universitas Muhammadiyah Yogyakarta, Indonesia)

This paper introduces the Generalized Cultural Representation Index (G-CRI), a novel mathematical framework for quantifying cultural bias in vision-language models. Through comprehensive evaluation across five Indonesian cultural domains, we reveal significant domain-specific bias patterns and cross-modal interactions that challenge traditional additive bias models. Our cross-modal decomposition demonstrates a strong negative correlation between visual and textual bias components, while Cultural Calibration emerges as the most effective mitigation strategy. Statistical validation confirms the framework's rigor and establishes requirements for robust cultural bias detection. This work advances multimodal fairness research by providing a statistically validated approach to cultural bias assessment, addressing critical underrepresentation of Southeast Asian contexts in AI systems.

2:10 ***A Robust and Imperceptible Image Watermarking Scheme Using Polar Harmonic Transform, STDM, and BCH Codes***

Sultan Zhorgy Pratama Harahap, Gelar Budiman and Linda Meylani (Telkom University, Indonesia)

The rapid proliferation of digital media and the rising incidence of unauthorized copying, tampering, and distribution have intensified the need for robust copyright protection mechanisms. This paper proposes a hybrid image watermarking framework that integrates Polar Harmonic Transform (PHT), Spread Transform Dither Modulation (STDM), and Bose-Chaudhuri-Hocquenghem (BCH) error correction codes to achieve high imperceptibility and robustness under diverse attack conditions. The PHT domain provides inherent invariance to geometric transformations, STDM ensures stable embedding against signal distortions, and BCH coding enhances error resilience. The proposed method was implemented in MATLAB and evaluated on multiple standard host images. Performance was assessed in terms of Peak Signal-to-Noise Ratio (PSNR), Bit Error Rate (BER), and payload capacity under various configurations and common image processing attacks, including JPEG compression, rotation, scaling, additive white Gaussian noise (AWGN), and filtering. Experimental results demonstrate that the scheme achieves high imperceptibility (PSNR up to 44.46 dB) and strong robustness, with BER = 0 for most attacks and minimal degradation under severe conditions (e.g., BER ≤ 0.30 under high-level AWGN). Comparative analysis with recent state-of-the-art methods confirms that the proposed approach consistently outperforms in robustness while maintaining competitive image quality. These findings highlight the effectiveness and practicality of the PHT-STDM-BCH integration for real-world copyright protection and authentication, offering a balanced trade-off between imperceptibility, robustness, and payload.

2:30 ***Multi-Label Classification of Infiltration and Effusion in Chest X-ray Images using Hybrid CNN-Transformer***

Andre Aditya Amann and Putu Harry Gunawan (Telkom University, Indonesia)

Early detection of infiltration and effusion on chest X-ray images is crucial but challenging due to the frequent occurrence of cases where both appear simultaneously. This study aims to evaluate whether a hybrid ensemble model can improve multi-label classification performance in such cases. Two architectures, DenseNet-121 (CNN) and Vision Transformer (ViT), were trained separately using a two-phase fine-tuning strategy on the PadChest dataset, which was divided into training, validation, and test sets using a stratified split approach. The training process was optimized using Focal Loss to address class imbalance, before the final ensemble model was developed using the simple averaging technique, which averages the prediction probabilities of the two best models. In the evaluation on the test set, model performance was evaluated using Area Under the Curve (AUC) as well as Precision, Recall, and F1-Score with optimized thresholds. The ensemble model showed the best performance with an average AUC of 0.9167, surpassing the single models DenseNet-121 (0.9112) and ViT (0.9065). The success, supported by Grad-CAM analysis, shows that the synergy between

the local feature extraction capabilities of CNN and the global context understanding of Transformer is an effective and robust approach for complex medical image classification.

2:50 ***Comparison of Supervised Learning Performance and Graph-Based Methods in Detecting Fake Reviews on E-Commerce***

Aqiela Putriana Shabira, Putu Harry Gunawan and Fitriyani Fitriyani (Telkom University, Indonesia)

This study presents a comparative analysis between traditional machine learning and graph based approaches for fake review detection. Using a curated Amazon 2018 product review dataset from prior research, we evaluated the performance of the Support Vector Machine (SVM) and Graph Convolutional Network (GCN) models, with TF-IDF as the feature extraction method. Experimental results show that SVM slightly outperforms GCN, achieving an accuracy and F1-score of 0.8511, compared to 0.8498 for GCN. Both models also achieved high AUC scores above 0.93, indicating strong classification capability. These findings suggest that for text-based review detection tasks, traditional machine learning methods such as SVM can remain competitive with more complex deep learning architectures, particularly when the dataset lacks relational structures between entities. This research provides practical insights into model selection for fake review detection systems in real-world e-commerce applications.

3:10 ***Used Car Prices Prediction Using RNN***

Sara Abou Chaar (American University of Science & Technology (AUST), Lebanon); Georges Chamoun and Gaby H Abou Haidar (American University of Science and Technology, Lebanon); Michel Owayjan (American University of Science & Technology, Lebanon & IREENA - Nantes University, France); Roger Achkar (Antonine University, Lebanon)

This study looks at how Recurrent Neural Networks (RNNs), especially LSTM models, can help predict the prices of used cars. It uses a large dataset with details like the car's age, mileage, brand, fuel type, and transmission to spot patterns that affect pricing. By cleaning the data and creating smart features like how fast a car loses value, the model learns to understand how different factors impact price over time. The model's accuracy is tested using common error measures, and its settings are fine-tuned to get the best results. In the end, the RNN model does a great job at predicting prices and can be used in real-world tools for buyers, sellers, and auto experts. To make it easy to use, a simple interface was also built so users can get instant price estimates.

3:30 ***Heart Disease Classification using Neural Networks: A Performance Analysis of SCG and CGP Models***

Yessi Jusman and Rizky Insani (Universitas Muhammadiyah Yogyakarta, Indonesia); Julnila Husna Lubis (Universiti Malaysia Perlis, Malaysia & Universitas Muhammadiyah Yogyakarta, Indonesia)

heart disease is one of the most critical conditions within the spectrum of cardiovascular diseases (CVDs) and remains the leading cause of mortality worldwide. According to the World Health Organization (WHO), CVDs account for approximately 17.9 million deaths annually, with heart disease contributing the most significant proportion. Early identification of at-risk individuals is essential to support effective prevention and treatment strategies. This study presents a comparative performance analysis of two multilayer perceptron (MLP) models, Conjugate Gradient with Powell/Beale Restarts (CGP) and Scaled Conjugate Gradient (SCG), using two hidden layer settings (5 and 10 hidden nodes) to examine the effect of network complexity on classification performance. The experimental results demonstrate that the CGP model with 10 hidden nodes achieved the highest testing accuracy of 90.98%, outperforming other configurations. These findings highlight the effectiveness of CGP-based optimization in improving classification accuracy and provide essential insights for designing reliable decision support systems aimed at heart disease prediction within the broader context of cardiovascular risk assessment.

Monday, December 1 1:50 - 4:10

2D: Session 2D

1:50 ***MLP vs TabNet: A Comparative Study of Models for Detecting Cardiovascular Diseases***

Nida Anggraeni and Putu Harry Gunawan (Telkom University, Indonesia)

Cardiovascular disease (CVD) remains the leading cause of global mortality with 17.9 million deaths annually, necessitating accurate early detection systems for intelligent healthcare. This study compares Multi-Layer Perceptron (MLP) and TabNet architectures for CVD risk prediction, addressing effectiveness, efficiency, and interpretability requirements for secure clinical deployment. We integrated five datasets (UCI Cleveland, Hungarian, Switzerland, VA Long Beach, and Mendeley) comprising 1,920 patients with 13 clinical features, applying hybrid imputation and stratified validation. Four configurations were evaluated: MLP Baseline, MLP Tuned, TabNet Baseline, and TabNet Tuned. TabNet Tuned achieved superior performance (90.62% accuracy, 91.86% F1-score, 95.29% AUC-ROC) with attention-based interpretability identifying fasting blood sugar, chest pain, and ST depression as key predictors. However, computational cost was substantial (211.4s tuning, 46.6ms inference). MLP Baseline provided optimal balance (86.98% accuracy, 90.91% optimized F1-score) with minimal overhead (4.3ms inference, 0.061MB size), suitable for resource-constrained environments. Feature importance analysis validated clinical knowledge, emphasizing metabolic and stress test parameters. Results demonstrate that TabNet excels in accuracy-critical scenarios requiring transparency, while MLP offers efficient, interpretable solutions for point-of-care deployment, supporting intelligent and secure cardiovascular risk assessment systems aligned with Industry 4.0 healthcare transformation.

2:10 ***Robust Spread Spectrum-based Information Hiding on Stationary Audio using DST and QR Decomposition***

Gelar Budiman, Andika Rizki Pratama and Sofia Saidah (Telkom University, Indonesia)

A robust hybrid audio watermarking scheme is proposes in this paper, utilizing a combination of Stationary Wavelet Transform (SWT), Discrete Sine Transform (DST), QR Decomposition, and Spread Spectrum (SS) to achieve im- proved embedding performance. The watermark is embedded into selected wavelet subbands after a series of multilevel transformations and spread using pseudo-noise sequences for enhanced robustness. The system is evaluated using four mono audio signals sampled at 44.1 kHz, with a 32 × 32 binary watermark image. MATLAB is employed as the simulation platform. Performance is assessed in terms of Signal-to-Noise Ratio (SNR), Bit Error Rate (BER), and payload. Experimental results show that the proposed method achieves strong robustness under a wide range of signal processing attacks-including filtering, resampling, requantization, additive noise, time scaling, and MP3 compression-while maintaining Imperceptibility (SNR = 27.21) and supporting payloads up to 172.27 bps. Compared to existing methods, the system offers a favorable trade-off between watermark transparency, capacity, and resilience, making it suitable for real-world copyright protection applications.

2:30 ***The Role of Artificial Intelligence Technology in Managing Stress and Improving Teachers' Mental Health***

Hasan Baharun and Najiburrahman Najiburrahman (Nurul Jadid University, Indonesia); Zamroni Zamroni (UIN Sultan Aji Muhammad Idris, Indonesia); Fathor Rozi, Khodijatul Qodriah and Raudlotul Hikmah Hikmah (Universitas Nurul Jadid, Indonesia)

This study examines the role of artificial intelligence (AI) technology in managing stress and improving teachers' mental health. Teachers often face high levels of stress due to workload, interactions with students, and administrative pressures. AI technology, through stress monitoring applications, AI-based counselling programs, and coaching platforms, offers practical solutions to support teachers' mental well-being. This study employs a qualitative approach, utilising case studies in several schools that have implemented AI technology. The results demonstrate that AI is effective in monitoring stress conditions in real-time, facilitating early

intervention, and delivering personalised mental health coaching. However, challenges related to access to technology, usage skills, data privacy issues, and the acceptance of technology persist as significant barriers. The implications of this study suggest that with improved infrastructure support, teacher training, and effective privacy management, AI technology can be a valuable tool in enhancing teachers' mental health and mitigating the impact of workplace stress.

2:50 ***Artificial Intelligence in Omani Public Schools: Challenges and Strategies for Effective Educational Administration***

Shaima Adil Aljuma (College of Education, Oman & Sultan Qaboos University, Oman); Yasser F. Hendawy Al-Mahdy (SQU, Oman)

This study aims to investigate the challenges associated with integrating Artificial Intelligence (AI) in public schools in the Sultanate of Oman and to propose innovative strategies to enhance the effectiveness of educational administration. A descriptive survey methodology was employed, utilizing a 22-item questionnaire that addressed two primary dimensions: obstacles to AI implementation and proposed developmental strategies. The study sample comprised 480 participants, including school principals, vice-principals, and teachers from public schools across the governorates of Muscat, North Al Batinah, South Al Sharqiyah, Al Dakhiliyah, and Dhofar. Data were analyzed using the Statistical Package for the Social Sciences (SPSS) to calculate means, standard deviations, and item rankings.

The results revealed that the primary challenges included inadequate digital infrastructure, limited financial support, and insufficient professional training programs for educators and administrators. Conversely, key strategies for development were identified, such as improving technological infrastructure, organizing continuous workshops and professional development programs, fostering coordination among relevant authorities, and adopting AI-supported interactive learning approaches.

Based on these findings, the study recommends providing advanced communication and AI-ready infrastructure in schools, allocating sustainable budgets to support AI initiatives, developing staff capacities through continuous professional training, and fostering a school-wide culture of AI adoption. Furthermore, it emphasizes the importance of collaboration between the Ministry of Education and the Ministry of Transport, Communications, and Information Technology to ensure effective support.

3:10 ***Data Center Fire Outbreak Simulation Using Machine Learning***

Mashrur Kabir, Mirza Ahmad Shayer and Sushana Islam Mim (BRAC University, Bangladesh); Nazia Parvin (Brac University, Bangladesh); Arnob Rahee and Md Motahar Mahtab (BRAC University, Bangladesh)

Data in our modern world is the most important currency. It is crucial for the data to be stored safely. Servers come to mind when large data storage is required. Many servers are required, so server racks are grouped together in buildings called data centers. Though data centers are typically safe, a disaster can strike anytime. Fires can break out of data centers which are difficult to control. All data in the racks are destroyed. As such, fire simulation in data centers is essential to stop the destruction of servers. Server racks, sources and causes of fires along with fire detection software are discussed in the background. Fire detection and suppression in data centers are elaborated with prior and modern systems. A machine learning based approach is utilized with K Nearest Neighbor (KNN), Support Vector Machine (SVM) and Random Forest Classifier (RFC) algorithms. The models are simulated. SVM achieved 96% accuracy followed by RFC at 92% and KNN at 44%. SVM achieves higher accuracy but RFC achieves better performance. Findings are concluded and plans for future research are discussed.

3:30 ***Integrating AI with Technical Analysis: A Python-Based Approach to Stock Trading***

Akshay Kamath (Symbiosis Institute of Business Management, Pune, India); Deepika Pandita (Symbiosis Institute of Business Management, Pune & Symbiosis International (Deemed) University, Pune, India); Rojarani Kanithi (Symbiosis Institute of Business Management, Pune, India & Symbiosis International (Deemed) University, Pune, India)

In financial markets, the reliability and interpretability of trading strategies are crucial for informed decision-making. A strategy becomes truly effective when its historical performance is validated through systematic back-testing. Traditional manual back-testing methods are static, error-prone, time-consuming, and often inadequate for analysing complex, multi-factor trading conditions. Given the high volatility of stock markets, where price movements can rapidly shift between bullish and bearish trends, manual analysis struggles to capture subtle patterns and rapid fluctuations. Addressing these challenges, this study proposes an interactive and automated back-testing framework that integrates technical indicators, congestion zone detection, candlestick pattern recognition, and moving average analysis to simulate realistic trade performance on the Nifty 50 index and ten selected constituent stocks. The system, developed in Python, allows traders and analysts to visualise OHLC data distributions, apply peak detection and congestion algorithms, and conduct efficient trade simulations. The proposed framework enhances strategy validation, reduces manual effort, and bridges the gap between traditional technical analysis and real-world market dynamics.

3:50 ***Efficient Machine Learning-Based Classification Model for Credit Risk Scoring in Lending Systems***

Sandeep Shivam (Tavant Technologies, USA)

Financial organizations have a major difficulty with credit risk prediction. The expansion of credit risk prediction models accessible to financial organizations is a direct outcome of the data and processing capacity made possible by advancements in machine learning. This paper presents an effective machine learning-driven MLP classification framework for credit risk scoring, including application scoring based on structured financial and personal data. Random Forest is used to select features and remove the irrelevant ones. High-risk and low-risk borrowers are taken in equal measures to curb the imbalance of classes through the use of random oversampling. The proposed MLP model can be optimized using hyperparameters such as learning rate, batch size, and dropout and is configured with three hidden layers, ReLU activation, and the Adam optimizer. The model has been found to have very high accuracy (Acc) (99.99%), precision (Prec) (99.75%), recall (Rec) (99.56%), and F1-score (99.34%) that are quite significantly better than the conventional models, such as RF and SVM. The model's stability and generalizability are shown by the accuracy and loss curves, as well as a confusion matrix. This research is useful to the lenders to make credible and informed financial decisions as it offers a scalable interpretable system to make accurate credit risk predictions.

Monday, December 1 1:50 - 3:50

2E: Session 2E

1:50 ***Global Trends and Thematic Evolution of Machine Learning in Breast Cancer Research***

Dian Yuliati (Institut Teknologi Sepuluh Nopember Surabaya, Indonesia & UIN Sunan Ampel Surabaya, Indonesia); Mohammad Isa Irawan, Sr. and Muhammad Syifaul Mufid (Institut Teknologi Sepuluh Nopember, Indonesia)

Breast cancer remains a worldwide health problem and requires new diagnostic and prognostic tools. Despite the revolutionary role of artificial intelligence (AI) in oncology, no study to date has comprehensively examined global AI research on breast cancer published between 2020 and 2025. This study presents a systematic bibliometric analysis of 5,032 Scopus-indexed articles, utilizing VOSviewer and a Python-based script to visualize scientific structures, collaboration networks, and thematic evolution. The PRISMA flow diagram was used to ensure that the literature was screened in a clear and organized way. We observe high growth rates in the number of published papers and papers cited, with a compound annual growth rate of 25.6% for 389

papers in 2020, increasing to 1,141 research articles as of 2024, accompanied by a total citation count of 94,295 and an H-index value of 119. China and India led in productivity, while the United States achieved the highest citation impact and centrality in collaboration. IEEE Access published the most significant number of papers, and Diagnostics achieved the highest average citation rate. Thematic evolution indicates a transition from image analysis to multimodal, biologically integrated models, with a growing focus on radiomics, gene expression, and understandable AI. The results suggest that this field has already matured into a more coordinated and homogeneous area of interest, with the primary challenges being clinical validation and interpretability. Future research should focus on the clinical translatability, model fairness, federated learning, and multi-modal omics integration to bridge algorithmic invention with equitable precision oncology practice.

2:10 ***Applying Vision Transformer and EfficientNet for Human Ear Identification***

Felicia Janice Pranoto (Binus University, Indonesia); Kevin Tjahjadi, Nathaniel Emmanuel Jiaw and Ivan Sebastian Edbert (Bina Nusantara University, Indonesia)

Fingerprints and facial recognition are among the most widely used biometric methods for human identification, yet both have notable limitations. To address these challenges, multimodal biometrics and alternative traits have been explored to improve accuracy and reliability. One emerging biometric is the human ear, which possesses unique shape characteristics that differ across individuals and remain relatively stable over time. In this paper, we investigate the effectiveness of deep learning approaches for ear recognition using the EarVN1.0 dataset. Specifically, we compare multiple variants of EfficientNet (B3, B4, and B7) and Vision Transformer (ViT-B16 and ViTB32) architectures. Experimental results show that EfficientNetB7 achieves the highest accuracy of 84.84%, slightly surpassing both transformer models, thereby highlighting the effectiveness of convolutional architectures with compound scaling for ear recognition tasks. These findings demonstrate the potential of ear biometrics as a reliable method for human identification

2:30 ***Hybridization of Case-Based Reasoning and AI Model for Intelligent Estimation of Building Costs***

Xinyi Tu and Thacha Lawanna (Chiang Mai University, Thailand)

curate early-stage cost estimation remains a critical challenge in construction management, particularly under dynamic conditions involving fluctuating material prices, regional labor costs, and changing regulations. Traditional methods-such as Quantity Surveying, Parametric Estimation, Regression Analysis, and Analogous Estimation-often fall short due to their reliance on linear assumptions, limited contextual adaptability, and poor generalization to complex project scenarios. To address these limitations, this study proposes a hybrid estimation framework that integrates Case-Based Reasoning (CBR) with Artificial Intelligence (AI). The proposed model follows a structured four-step cycle: Retrieve, Reuse, Revise, and Retain. Historical project cases are retrieved based on similarity metrics learned and optimized by AI algorithms. These cases are adapted using contextual parameters and continuously updated through a retention mechanism. Evaluation was conducted across five benchmark datasets from the UCI Machine Learning Repository, including Concrete Compressive Strength and Energy Efficiency. Results show that the AI-CBR model achieved a mean absolute percentage error (MAPE) between 6.0% and 6.6%, outperforming Regression (9.1%-10.2%) and Analogous Estimation (16.9%-18.4%). Computational runtime was reduced to under 18 seconds on average, and learning efficiency improved by 10-14% over iterations. These results confirm the model's superiority in accuracy, efficiency, and adaptability. The study contributes theoretically by demonstrating the effectiveness of hybrid AI-CBR models, and practically by offering a scalable solution for intelligent cost estimation in real-world construction projects.

2:50 ***Transforming Financial Engineering through AI-Enhanced Case-Based Reasoning for Digital Innovation***

Song Song, Mingyang and Thacha Lawanna (Chiang Mai University, Thailand)

Traditional models in financial engineering-such as Rule-Based Systems, Bayesian Networks, Decision Trees, and Markov Decision Processes-have long been used for tasks like credit risk assessment, fraud detection, and portfolio optimization. However, they often suffer from limited adaptability, lack of transparency, and difficulty integrating heterogeneous financial data. This study proposes an AI-enhanced Case-Based Reasoning (CBR) framework to address these limitations. The proposed model is structured around five core stages: case acquisition, case indexing using machine learning embeddings, similarity retrieval using neural metrics, solution adaptation powered by reinforcement learning, and feedback-driven learning for continual refinement. Across five benchmark datasets (Adult, Bank Marketing, Car Evaluation, Student Performance, and Online Retail), the AI-CBR system achieved an average accuracy of 91.6%, significantly outperforming traditional methods, which averaged below 78%. It also led in efficiency, competency, and operational effectiveness, with AI-CBR maintaining execution time under 3.5 seconds per transaction. These results validate its superiority in both predictive performance and system-level robustness. The model's ability to update itself based on user and market feedback ensures continual learning, making it highly suited for dynamic financial environments. In addition, its explainable architecture supports regulatory compliance and real-world implementation. Overall, this work contributes a novel hybrid model that bridges symbolic reasoning with modern AI, offering an advanced solution for digital financial innovation.

3:10 ***Aspect and Emotion Multiclass Classification of Online Loan Application Reviews Using IndoBERT***

Zahraa Sakinah Almuhibb, Andry Alamsyah and Danang Indrajaya (Telkom University, Indonesia)

The rapid growth of online lending platforms in Indonesia has transformed access to digital financial services but simultaneously created persistent challenges, including high interest rates, privacy violations, and aggressive debt collection practices. Existing studies in this area have largely concentrated on sentiment polarity analysis, leaving a gap in understanding how both objective service related issues and subjective emotional experiences jointly shape user perceptions. To address this problem, this study proposes a two-stage multiclass classification framework leveraging IndoBERT, a pre-trained transformer model specifically adapted to the Indonesian language. A dataset of 15,750 Google Play reviews was annotated into service aspects debt collection, high interest and fees, privacy, and accessibility and emotional categories anger, disappointment, joy, and fear and subsequently fine-tuned with different hyperparameter settings. Experimental results show that the proposed framework achieved macro F1-scores of 0.82 in both stages, with the categories of Privacy and Accessibility producing the strongest performance, while High Interest/Fees and Disappointed remained more difficult to classify due to semantic overlap and subtle linguistic expression. These findings provide a structured view of how Indonesian users articulate both technical and emotional concerns in their online lending experiences, offering valuable insights for improving service design, strengthening consumer protection, and guiding regulatory strategies in the fintech sector. Future research may extend this framework with larger datasets, additional categories, and lightweight transformer-based methods such as IndoBERTweet or parameter efficient fine-tuning. Multilingual and explainable approaches can further improve scalability and transparency.

3:30 ***Lightweight YOLO Architectures for Efficient Dog Emotion Classification***

Auryn Larissa Sabrina (BINUS University, Indonesia); Martin Sebastian (Bina Nusantara University, Indonesia); Viencentius Delvin Frelim (Bina Nusantara, Indonesia); Samuel Philip (Bina Nusantara University, Indonesia)

This study explores lightweight YOLO architectures (YOLOv8s-cls and YOLOv11s-cls) for classifying dog emotions-angry, happy, relaxed, and sad-derived from the Circumplex Model of Affect. A curated dataset of 3,600 images (900 per class) was preprocessed through quality filtering, duplicate removal, and feature clustering to ensure diversity and balance. Both models were trained with the same dataset split, preprocessing steps, image resolution, learning rate, batch size, number of epochs, evaluation metrics, and cross-validation method, ensuring fair comparison. The performance was evaluated using 5-fold cross-

validation, followed by testing on a separate test set that was excluded during training or validation. YOLOv8s-cls achieved the highest performance with 86.42% accuracy and a macro average F1-score of 86.43%, surpassing YOLOv11s-cls (85.28% accuracy, 85.22% F1-score) and outperforming conventional CNNs, VGG16, and ResNet50 from prior studies. Results highlight YOLO's potential as a fast, resource-efficient solution for real-time dog emotion detection, with applications in monitoring systems and mobile platforms.

Tuesday, December 2 10:05 - 12:05

3A: Session 3A

10:05 ***The Development of the Hotel Price Monitoring Dashboard for Analyzing the Competitors Pricing: A case study***

Rio Nurtantyana (National Research and Innovation Agency, Indonesia & Telkom University, Indonesia); Muhammad Fajar Jati Permana (Universitas Pendidikan Indonesia, Indonesia); Giap Van Nguyen (National Central University & Thai Nguyen University, Taiwan); Pham Thao (National Economics University, Vietnam); Pramesthi Anggoro Sekti (Gloaicloud, Bandung, Indonesia)

The hospitality industry faces challenges in competitive price for their revenue management strategy, necessitating advanced solutions for monitoring their dynamic pricing. In this case study, we developed the Hargana Hotel Monitoring (HHM) system with prototyping method. HHM is a software that leveraging machine learning to predict the hotel pricing for revenue management strategies. As a results, we choose the HHM system integrates with XGBoost which has good performance based on the RMSE, MAE, and R2 metrics compared to other four algorithms. Further, we have deployed the ML system for the backend and the developed the front-end of HHM system. The user acceptance testing results of the development of HHM system was passed from several scenarios. In addition, the user's perspectives felts that the HHM system was easy to use. Therefore, HHM system was developed well, and it could be used for helping the revenue management strategies.

10:25 ***The Development of an Interactive Programming Learning Platform with AI-Driven Feedback and Online Code Execution for the Java language***

Lukman Alfaris Hidayatullah and Gede Agung Ary Wisudiawan (Telkom University, Indonesia); Rio Nurtantyana (National Research and Innovation Agency, Indonesia & Telkom University, Indonesia)

Programming education often encounters challenges in delivering fast, easily understood feedback when students run into errors in their code. Traditional approaches that rely on manual debugging and searching for solutions on online forums which are time-consuming and can undermine learners' motivation. To address these issues, we have developed a web-based interactive platform for learning Java object oriented programming that integrates AI-generated feedback with online code execution. The system delivers slide-based materials, weekly timed quizzes with auto-grading, and a multifile workspace with visibility controls, a code editor with AI feedback. The evaluation included two weeks of developer-led functional testing and classroom settings deployment with 40 students. The functional pass rates were 100% for quizzes, coding task submission, learning progress, ratings, and authentication; 95% for AI features; and 89% for code execution, with failures traced to external API constraints. During user testing, students generated 1,019 AI-feedback interactions, which they rated 4.61 on average; the mean AI response time was 3.8 seconds. The results indicated positive acceptance of AI support with a mean of 4.08. The user testing highlighted priorities for hardening and performance, such as resolving workspace auto-refresh via better state management, implementing robust autosave, improving execution error handling, optimizing database queries, enhancing caching, and adding loading states. Additional roadmap items include refining AI feedback for complex logic,

mobile responsiveness, notifications, and offline access. These results demonstrate the platform's potential as an engaging and effective tool for programming instruction.

10:45 ***Leveraging Machine Learning for Sentiment Analysis on Indonesia's M-Paspor Application***

Evaristus Didik Madyatmadja (Bina Nusantara University & School of Information Systems, Indonesia); Fifinellametta Kurniadi, Raffaele Delon Torbik, Ryan Winata, Samuel Matthew Masman and Arya Swastika (Bina Nusantara University, Indonesia)

This study focuses on sentiment analysis of the M-Passport application, an application developed by the Indonesian government to help the process of taking care administration for things such as passport renewal, changing documents for said passport, to fill in data and upload documents. The research aims to evaluate user sentiments to provide insights for future development and/or optimization. Using 12,500 Google Play Store reviews from Indonesian users collected via web scraping, preprocessed through stages including normalization, case folding, stopword removal, tokenization, and stemming normalization. Three machine learning algorithms namely, Naïve Bayes, Support Vector Machine, and Random Forest were employed for classification. For evaluation metrics we use precision, recall, Fl-score, and accuracy. The research found that Random Forest scored better overall against the other two algorithms, establishing a solid foundation for this domain specifically. The novelty of this work lies in its timely analysis of a platform that is mainly useful for international travel, its focus on the unsatisfied Indonesian user reviews and its comparison between three machine learning algorithms for this context. This study provides data with specific actionable insights based on the reviews given. This research contributes to understanding user feedback and improving app strategies, particularly for sentiment-driven developments in the Indonesian market.

11:05 ***Multilevel Confirmatory Factor Analysis for Modelling the Determinants of Malnutrition among Under-Five Children in Indonesia***

Kamaluddin Simamora (Institut Teknologi Sepuluh Nopember Surabaya, Indonesia); Bambang Widjanarko Otok (Institut Teknologi Sepeluh Nopember, Indonesia); Setiawan Setiawan (Institut Teknologi Sepuluh Nopember (ITS), Indonesia)

Malnutrition, including stunting, wasting, and underweight, remains a serious public health challenge in Indonesia. Various factors contribute to malnutrition, ranging from individual- and household-level to environmental, socioeconomic, and health policy factors. The use of data analysis as a foundation for policy formulation has been widely adopted in many fields, including healthcare. Most studies have analyzed the determinants of nutritional status using conventional approaches, without accounting for the hierarchical structure of the data or the presence of unobserved (latent) variables. This study aims to examine the factors influencing malnutrition in Indonesia using a Multilevel Confirmatory Factor Analysis (MCFA) approach. This approach allows modeling the causal relationships among latent variables at two distinct levels of analysis, the district (level one) and the province (level two) simultaneously. Based on the Intraclass Correlation test, all malnutrition indicators have values greater than 0.05, indicating variability at the group (province) level. In conclusion, most indicators of Parenting, Health and Environmental Services, and Malnutrition are significant, with Parenting and Malnutrition showing a strong impact. Socioeconomic factors at the district level show limited influence, while provincial-level factors are more significant for wasting and underweight. Both district and provincial-level factors contribute to stunting, with district-level factors having a greater impact.

11:25 ***Evaluating User Perception of Artificial Intelligence-Driven Healthy Behavior and Herbal Medicine Recommendations using Mean Opinion Score (MOS) Approach***

Rita Destiwati (Telkom University, Indonesia); Junardi Harahap (Universitas Padjadjaran, Indonesia); Devi Mulyanti (Communication Science, Telkom University, Indonesia)

This study aims to provide information regarding the role of Artificial Intelligence (AI) in the development of herbal medicine and the implications of healthy lifestyle models. This study uses qualitative methods. Data was collected through in-depth interviews with AI experts, herbal practitioners, herbal users, students who use AI, and developers of herbal product chatbots. The results of this study found that AI plays a dual role, including: 1. a scientific instrument that accelerates research on herbal ingredients, 2. an educational platform that encourages healthy behavior based on digital knowledge. AI has been proven to improve health literacy, strengthen user confidence in evidence-based medicine, and encourage safer and more measurable medical practices. However, challenges still arise in terms of data quality, scientific validation, and algorithmic transparency. This study emphasizes that AI in herbal medicine requires multidisciplinary collaboration between data scientists, herbal researchers, health institutions and regulators to ensure the effectiveness and ethical use of the technology.

11:45 ***Data Analytics Maturity Model for Digital Product Innovation in Firm: An Overview***

Noor Alamsyah (King Mongkut's Institute of Technology Ladkrabang, Thailand & University of Mataram, Indonesia); Bundit Thanasopon (King Mongkut's Institute of Technology Ladkrabang, Thailand); Pornsuree Jamsri (KMITL, Thailand)

This paper aims to report the results of the review, map the characteristics, and conduct a comparative analysis of the selected data analytics maturity models that are commonly implemented in firms engaged in digital product innovation. The maturity models examined include TDWI, Hortonworks, Gartner, and DELTA Plus, which represent some of the most recognized framework guiding organizations in assessing and improving their data analytics capabilities. Characteristics and comparative analysis were developed based on investigative literature reviews, referring to official reports from institutions that developed the model and publications that discussed the model. Each maturity model is explained concisely and comprehensively, highlighting the structure, dimensions, and progression criteria that define each maturity level and stage. The comparison indicates that although these models outline well-defined conceptual stages of analytics maturity, most provide limited methodological guidance for assessing firms or determining their placement within a spesific maturity level. In the future, it is expected there will be an increase in the availability of maturity models that are more adaptive and personalized for various business sectors. This paper contributes to present a fairly intensive comparative analysis and characterization of data analytics maturity models that are widely recognized and frequently applied in supporting digital product development within firms.

3B: Session 3B

10:05 ***XLM-RoBERTa as a Multilingual Pre-Trained Model for Question Answering in AQI Monitoring Installation Guides***

Cholilur Rohman, Aji Gautama Putrada and Ayu Qatrunnada Istiqfarri (Telkom University, Indonesia)

The Internet of Things (IoT) has become a widely diffused product in society, one of which is air quality index (AQI) monitoring. However, several studies highlight that IoT is acomplex technology, while others report that user complaints can hinder its diffusion. Our research proposes a natural language processing (NLP)-based question-answering (QA) system as a guide for AQI monitoring installations. We propose a cross-lingual language model-robustly optimized bidirectional encoder representations from transformers pretraining approach (XLM-RoBERTa) as the QA method, then benchmark it with an other QA method, namely the Indonesian version of BERT (IndoBERT). We use three metrics to compare the two: the metric for evaluating machine translation output (METEOR), the bilingual evaluation understudy (BLEU), and the recall-oriented understudy for query evaluation-longest common se-quence (ROUGE-L). METEOR can measure the semantics of a prediction, and BLEU is a metric that measures word precision in a predicted sentence. Lastly, ROUGE-L measures the fluency of the prediction. The test results show that XLM-RoBERTa outperforms IndoBERT for METEOR, BLEU-1, and ROUGE-L, with values of 0.54, 0.42, and 0.62, respectively. These results indicate that

XLM-RoBERTA exhibits good fluency and semantics, demonstrating effective synonyms. However, some understandings may be erroneous. Our research contribution presents a novel QA system for the AQI monitoring installation guide, utilizing an optimized model, namely XLM-RoBERTa.

10:25 ***Convenience Prediction Based on Coffee Consumption Patterns Using Machine Learning***

Tiari Juliani and Mindit Eriyadi (Telkom University, Indonesia); Bambang Pudjoatmodjo (Telkom University, Indonesia & Universiti Teknikal Malaysia Melaka, Malaysia); Desy Puspa Rahayu (Telkom University, Indonesia); Nor Azuana Ramli (Universiti Malaysia Pahang Al-Sultan Abdullah, Malaysia); Agus Pratondo (AILO, Indonesia & Telkom University, Indonesia)

This study aims to predict the perceived convenience or inconvenience of coffee consumption among university students using machine learning methods. A total of 200 Indonesian students participated in a survey covering demographic, behavioral, and perceptual factors related to coffee habits. Two algorithms Random Forest and XGBoost were applied and compared. The Random Forest model achieved an accuracy of 87%, while XGBoost reached 81%. Key influencing factors included self-reported health complaints, academic level, and study habits. The results show that machine learning can effectively identify behavioral patterns related to caffeine use, offering valuable insights for promoting balanced and healthy consumption practices.

10:45 ***The Impact Of Social Media Influencers Using Customer Engagement Theory On The Zoomer Generation With Partial And Simultaneous Testing***

Abdullah Billman (Bina Nusantara University, Indonesia & Raffles University Iskandar, Malaysia); Satria Budi (Universitas Bina Nusantara, Indonesia); Surjandy Surjandy (Institut Teknologi Sains Bandung, Indonesia); Bahri Bahri (Widya Mataram University, Indonesia)

This study summary examines the impact of Social Media Influencers (SMIs) on the behavior of Generation Z consumers in Indonesia through the lens of Customer Engagement Theory (CEt). This study used a quantitative methodology with the Smart Partial Least Square (SEM-PLS) technique, involving 252 Generation Z participants in Indonesia over a six-month period. Customer engagement is evaluated in three dimensions (Cognitive Engagement (CE), Emotional Engagement (EE), and Behavioral Engagement (BE)), either partially or simultaneously. The test results showed that all three categories significantly and positively influenced influencer impact, with emotional engagement being the main contributor, followed by cognitive and behavioral engagement. The concurrent analysis validated that customer interactions accounted for 60.7% of variance in influencer impact, reinforcing the importance of customer participation in increasing influencer credibility and appeal. The findings enrich the academic literature by offering multidimensional empirical evidence and filling conceptual gaps in the measurement of customer engagement. This research shows that marketers should prioritize collaborative methods with influencers that can foster emotional connections, rather than just focusing on the number of followers. This study highlights the importance of ethical standards in reducing negative effects, including excessive consumption and Negatively Valenced Influencing Behavior (NVIB). This study establishes that the effectiveness of influencer marketing in Generation Z is significantly influenced by the intensity of emotional engagement.

11:05 ***Exploring Deep Learning Pedagogy through AI-Based Sentiment and Emotion Analysis in Vocational Schools***

Pipit Utami, Nuryake Fajaryati, Yussi Anggraini, Rangga B Rinawan and Putu Sudira (Universitas Negeri Yogyakarta, Indonesia)

This study explores the implementation of Deep Learning Pedagogy, which integrates the principles of Joyful, Meaningful, and Mindful Learning, within the context of Indonesian vocational education. The research aims to understand how teachers and students experience and emotionally respond to deep learning practices. Data

were collected through reflective responses and Likert-scale items using Google Forms distributed to teachers and students in vocational schools (SMK). The qualitative text data were analyzed using AI-based sentiment and emotion analysis, while quantitative responses were examined descriptively to reveal affective patterns in learning experiences. The findings are expected to identify the dominant emotional tones (positive, negative, and neutral) and specific emotions such as curiosity, confidence, pride, anxiety, and confusion that frequently emerge during learning. This approach allows for mapping the balance between supportive and challenging emotions that contribute to engagement and learning depth. The study contributes to the growing discourse on Human-Centered AI in Education, emphasizing that effective deep learning requires not only cognitive and behavioral engagement but also a balanced emotional experience. The results are expected to inform future teacher training and AI-supported learning analytics for improving affective dimensions in vocational education.

11:25 ***Extreme Rainfall Analysis in the Surabaya Metropolitan Region using Spatial Extreme Value Model with Kumaraswamy Generalized Inverse Lomax Distribution***

Maria D.N. Ayu Bima Pebruwasti (Institut Teknologi Sepuluh Nopember, Indonesia); Sutikno Sutikno (Institut Teknologi Sepuluh Nopember Surabaya, Indonesia); Purhadi Purhadi (Sepuluh Nopember Institute of Technology, Indonesia)

Extreme weather events are natural phenomena that can cause significant losses across various sectors and have serious impacts on many aspects of life. These impacts can be minimized by analyzing the patterns and characteristics of extreme events. Extreme Value Theory (EVT) is one of the statistical techniques used to detect such events, and to capture extreme occurrences across multiple locations (multivariate), it has been extended into the Spatial Extreme Value (SEV) approach. This study applies the Peaks Over Threshold (POT) method to excessive rainfall using an extreme value model based on the Kumaraswamy Generalized Inverse Lomax (KumGIL) distribution, which combines the Kumaraswamy Generalized and Inverse Lomax distributions to provide high flexibility for modeling heavy-tailed and asymmetric data. The POT-KumGIL model is applied to monthly rainfall data from 1989-2024 in the Surabaya Metropolitan Area. Model suitability is tested using the Kolmogorov-Smirnov statistic and probability plots. If the model fits well, parameter estimation is performed using Maximum Likelihood Estimation (MLE) followed by BHHH numerical iteration. A trend surface model is then built to relate longitude, latitude, and elevation with parameter estimates, improving the robustness of spatial analysis. The best model, determined by the smallest Akaike Information Criterion (AIC), is used to calculate return levels for several future periods. This study aims to develop a POT-KumGIL spatial extreme value model and obtain return level estimates for the Surabaya Metropolitan Area to support future hydrometeorological disaster mitigation strategies

11:45 ***A Traceable UML-Based Requirements Modeling Framework for Educational Attendance Systems***

Aisha Putri Nuryan, Novian Anggis Suwastika and Isa Mulia Insan (Telkom University, Indonesia); Qori Qonita (Public Vocational High School 10 Bandung, Indonesia)

This paper proposes a traceable UML-based framework for modeling and validating system requirements in educational attendance applications. Unlike prior works that treat UML merely as a documentation tool, this study emphasizes the traceability and validation of requirements from stakeholder elicitation to model verification. The framework integrates stakeholder analysis, user story formulation, artefact mapping, and cross-diagram validation using a Requirements Traceability Matrix (RTM). Applied to the Presentia attendance system case, the study produced six functional and two non-functional requirements represented across use case, activity, sequence, class, and deployment diagrams. Validation results confirmed full traceability between stakeholder needs and UML artefacts, demonstrating model completeness and internal consistency. This

research contributes a replicable method for integrating UML with traceability-driven validation in educational software projects, enriching the body of knowledge in requirements engineering.

3C: Session 3C

10:05 ***Player Preferences for Educational and Non-Educational Games***

Zahra Putri Diva, Rio Korio Utoro and Rickman Roedavan (Telkom University, Indonesia); Duddy Soegiarto (Dept. of Multimedia Engineering, School of Applied Sciences, Telkom University, Indonesia); Kohbalan Moorthy (Universiti Malaysia Pahang Al-Sultan Abdullah, Malaysia); Agus Pratondo (AILO, Indonesia & Telkom University, Indonesia)

The rapid development of the digital sector, especially in entertainment and education, drives the need to understand user preferences in choosing the types of games they play, whether educational or non-educational. Predicting these preferences is crucial to support developers and educators in designing targeted content and enhancing user engagement. This study examines the application of machine learning techniques, specifically the Random Forest algorithm, to predict game-playing preferences among users in Indonesia. The dataset consists of 200 respondents evenly distributed between educational and non-educational game users. The Random Forest algorithm was chosen due to its ability to handle high-dimensional behavioral data and produce accurate and interpretable predictions. The developed model achieved an accuracy of 90%, with high precision and recall values in both categories. This study also identifies behavioral and contextual factors influencing user preferences, such as motivation to play, frequency of play, and visual appearance preferences. These findings provide valuable insights for the development of educational games, content personalization, and user engagement strategies. By taking the growing Indonesian digital market context into account, this study serves as a relevant reference for understanding gaming behavior in developing countries and supports data-driven decision-making for developers and educators.

10:25 ***Geotagged Tweets & Time-Stamp Sentiment Analysis of Public Reactions to the Death of Palestinian Leaders in 2024 Using BiLSTM***

Chisyana Intan Habibah and Fitriyani Fitriyani (Telkom University, Indonesia)

The prolonged conflict between Palestine and Israel has again attracted global attention following the deaths of two Hamas leaders, Ismail Haniyeh and Yahya Sinwar. Their passing triggered strong emotional reactions on social media, emphasizing the need for real time sentiment analysis using geotagged data. This research examines global sentiment dynamics through geotagged tweets and timestamps collected from platform X between July and December 2024, employing a Bidirectional Long Short-Term Memory (BiLSTM) classification model. Temporal analysis reveals a sharp rise in positive sentiment in August after Haniyeh's death, who was seen as a moderate figure. Conversely, October showed dominant negative sentiment after Sinwar's death, widely known for his militant role. This contrast illustrates how leadership images and media narratives influence public expression. The BiLSTM model achieved 84% accuracy with a peak F1-score of 94% in the neutral class. Spatial analysis indicates polarization: Muslim majority countries such as Turkey and Pakistan leaned positive, while Russia, Spain, and Portugal expressed more negative sentiment. Western nations varied, with the United Kingdom and Germany displaying relatively stronger support. These findings underscore the significance of integrating spatial-temporal approaches to interpret public reactions toward geopolitical crises and highlight opportunities for future research employing transformer-based models and multilingual datasets.

10:45 ***Prediction Comparison in Educational Data Using Hybrid Time Series SARIMA-XGBoost Model***

Gede Bramanta Pandya Wisesa and Muhammad Zarlis (Bina Nusantara University, Indonesia)

The present study proposes and assesses a hybrid time-series model for forecasting monthly new-student enrollments at the Bali Institute of Design and Business. The methodology utilizes a two-stage, residual-learning framework. Initially, a Seasonal Autoregressive Integrated Moving Average (SARIMA) model is fitted to capture the primary linear trends and seasonality in the data. Secondly, an Extreme Gradient Boosting (XGBoost) model is trained on the resulting residuals to learn the remaining non-linear patterns. Two distinct sets of engineered features were tested to optimize the XGBoost component. The model's performance was evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Symmetric Mean Absolute Percentage Error (sMAPE). The model was trained using data from 2021-2024 and tested using data from 2025. The findings substantiate that the hybrid model, particularly when endowed with a comprehensive feature set, exhibits a marked superiority in performance over the standalone SARIMA model, thereby furnishing a more precise forecasting instrument for institutional planning.

11:05 ***Identifying Key Determinants of Stunting Prevalence in East Nusa Tenggara Using Multiview Fuzzy C-Means Clustering***

Alya Fitri Syalsabilla, Nur Iriawan and Muhammad Mashuri (Institut Teknologi Sepuluh Nopember, Indonesia)

Stunting remains a pressing public health issue in East Nusa Tenggara (NTT), Indonesia, where malnutrition severely affects child development. This study investigates the multidimensional determinants of stunting using secondary data from the Ministry of Home Affairs (2024), covering health, sanitation, education, and economic indicators across districts. Data preprocessing involved imputation, normalization, and variable selection, after which the Multiview Attribute Fuzzy C-Means (MV-FCM) clustering method was applied. Cluster validity was assessed using the silhouette index, while the significance of indicators was evaluated through fuzzy membership degrees. The results reveal four key determinants are women's education level with maximum membership degree 0.99, clean water facilities 0.98, health infrastructure 0.96, and livelihood 0.73. Cluster analysis identified three distinct groups of regencies, with Cluster 2 showing consistently high prevalence (18.8-20%) requiring urgent intervention. The silhouette index of 0.72 indicates strong clustering validity, confirming the reliability of the analysis. Policy recommendations emphasize improving women's education through expanded access and community programs, strengthening sanitation via clean water infrastructure, and enhancing health services in remote areas. Livelihood improvement through sustainable local economies and social protection is also advised. In conclusion, targeted, evidence-based strategies integrating education, sanitation, health, and economic empowerment are essential to reduce stunting prevalence and achieve equitable child health outcomes in NTT.

11:25 ***A Dimensional Framework for Cage Environmental Sensor: Focusing on Eid Sacrifice Animals***

Deby Tetra Ariesti, Muhammad Hablul Barri and Rahmat Yasirandi (Telkom University, Indonesia)

Animal welfare for sacrificial livestock demands precise environmental management of the pen, considering the risk of suboptimal environments during temporary detention. Research addresses the issue that existing sensor-based monitoring systems are often implemented empirically, lacking a structured framework to justify sensor selection based on their functionality and relevance. Therefore, this study aims to develop a dimensional sensor framework and validate the reliability of selected sensors through calibration testing. The method involves classifying 11 key environmental parameters into four functional dimensions: (1) Managerial Intervention, (2) Animal Welfare Impact, (3) Functional Domain, and (4) Temporal Data Priority. Calibration testing demonstrated significant accuracy improvements compared to raw datasheet specifications: the calibrated Maximum Mean Percentage Relative Error (MAE) for Temperature is only 0.8%, while non-linear gas parameters (such as Ammonia) showed acceptable accuracy with a Percentage Relative Error of 15%. The implication of this dimensional framework is the transformation of the sensor selection process into a

structured, functional-based approach, essential for optimizing high-turnover sacrificial livestock monitoring systems.

11:45 ***Prediction of Javanese Language Usage Preference in Daily Life Using Machine Learning***

Rasyadiaz Yogatama (Telkom University Bandung, Indonesia); Rio Korio Utoro, Rickman Roedavan and Duddy Soegiarto (Telkom University, Indonesia); Agus Pratondo (AILO, Indonesia & Telkom University, Indonesia); Kohbalan Moorthy (Universiti Malaysia Pahang Al-Sultan Abdullah, Malaysia)

This study explores the prediction of Javanese language usage preferences in daily life through machine learning methodologies. By gathering behavioral and demographic data via a structured questionnaire, a classification model was developed using Random Forest to determine the likelihood of individuals using the Javanese language in informal settings.

Tuesday, December 2 10:05 - 12:25

3D: Session 3D

10:05 ***Design and Evaluation of a Continuous Testing Framework for Automated Front-End Validation in Student Attendance Systems***

Novita Sabila Nugraha, Novian Anggis Suwastika and Isa Mulia Insan (Telkom University, Indonesia); Qori Qonita (Public Vocational High School 10 Bandung, Indonesia)

This paper presents the design and evaluation of an Automated Continuous Testing Framework (ACTF) for validating the front-end functionalities of a student attendance system within a CI/CD environment. The framework integrates Cypress for end-to-end automation with GitHub Actions to enable continuous execution, traceability, and reproducibility of tests. The study follows the Design Science Research (DSR) methodology through five iterative stages: problem identification, framework design, implementation, demonstration, and evaluation. Functional requirements were first analyzed and mapped into test scenarios using equivalence partitioning and boundary value techniques. Each test case was linked to its corresponding requirement through a Requirements Traceability Matrix (RTM) to ensure coverage validation. A total of 48 automated test cases were executed across local and cloud environments, achieving 100% execution consistency and 95.8% requirement coverage. The results demonstrate that the proposed framework improves reliability, scalability, and early defect detection in continuous web application testing. This research contributes a validated and replicable model for integrating automated testing frameworks within CI/CD pipelines, extending the body of knowledge in software engineering and educational system development.

10:25 ***A Continuous Integration-Oriented Automated Testing for Educational SaaS Backends: Design and Evaluation***

Reza Adhie Dharmawan, Novian Anggis Suwastika and Isa Mulia Insan (Telkom University, Indonesia); Qori Qonita (Public Vocational High School 10 Bandung, Indonesia)

This paper presents the design and evaluation of a Continuous Backend Testing Framework (CBTF) integrating PHPUnit with GitHub Actions for automated testing within continuous integration (CI) pipelines. Unlike conventional implementation reports, this study emphasizes the methodological design, traceability, and evaluation of backend test automation in a multi-tenant educational SaaS context. The framework adopts the Design Science Research (DSR) methodology comprising five stages: problem identification, objective definition, framework design, demonstration, and evaluation. Test cases were systematically generated using the path coverage and boundary value techniques and mapped to functional requirements through a Requirements Traceability Matrix (RTM) to ensure completeness and traceability. A total of 69 test cases, including 64 unit and 5 integration tests were executed through the GitHub Actions CI pipeline, achieving

98.55% time reduction and 100% functional success across environments. The results demonstrate that CBTF enhances test reliability, coverage validation, and reproducibility within continuous delivery workflows. This study contributes a validated model for integrating backend test automation into CI/CD pipelines and expands the understanding of continuous testing design in software engineering.

10:45 ***Machine Learning Approach for Identifying Cyberbullying in Social Media***

Jonathan Tjahjana, Yosua Sugihartono, Said Achmad and Derwin Suhartono (Bina Nusantara University, Indonesia)

The widespread use of social media has transformed communication worldwide, enabling instant sharing of ideas and information. However, it has also facilitated negative behaviors such as cyberbullying, which can cause serious harm to individuals. While most previous studies focus only on binary classification to detect the presence of cyberbullying in a tweet, this study aims to perform multi-class prediction by identifying the specific type of cyberbullying, such as those targeting race, religion, or other personal attributes. We enhance text preprocessing by applying a TF-IDF approach with N-gram features (unigrams, bigrams, and trigrams) to capture richer linguistic patterns and contextual relationships. In addition, we implement an ensemble learning approach using stacking, with Support Vector Machine, XGBoost, and Random Forest as the base models. Experimental results show that the stacked model achieves high performance, with an accuracy of 0.9328 and an F1-Score of 0.9335, demonstrating the effectiveness of this approach in detecting and categorizing various types of cyberbullying on social media. These findings highlight the potential of advanced ensemble learning methods combined with contextual text analysis in improving automated cyberbullying detection and classification on social media platforms.

11:05 ***Classifying Regional Poverty Levels in Indonesia: A Random Forest Analysis of Key Contributing Factors***

Michelle Imanuela Winata and Muhammad Fikri Fikri Hasani (Bina Nusantara University, Indonesia)

This study applies a Random Forest classifier to classify poverty levels-low, middle, and high-across 514 Indonesian districts using nine socioeconomic indicators. Labels were generated through a fuzzy system based on HDI, Life Expectancy, and Schooling. Five-fold Group K-Fold validation prevented data leakage, and robustness was tested with noise injection. Results show the fuzzy baseline achieved 95.35% accuracy and 95.59% macro-F1, while Random Forest variants performed consistently around 93-94%. SHAP confirmed education and health indicators as most influential. These findings highlight the potential of ensemble learning for guiding poverty reduction policies.

11:25 ***An Empirical Study on Lazy Loading Optimization in ReactJS-Based Educational Single Page Applications***

Muhammad Zaky Fathurahim, Novian Anggis Suwastika and Isa Mulia Insan (Telkom University, Indonesia); Qori Qonita (Public Vocational High School 10 Bandung, Indonesia)

This study presents the design and empirical evaluation of a Lazy Loading Optimization Framework (LLOF) for ReactJS-based Single Page Applications (SPAs) in educational attendance systems. While Lazy Loading has been widely recognized for improving frontend performance, limited empirical work has examined its impact across various page complexities and user requirements in real educational contexts. Using a Design Science Research (DSR) approach, this study follows five stages: problem identification, objective definition, system design, experimentation, and evaluation. User requirements were gathered from school administrators and teachers to guide component prioritization for Lazy Loading. Six representative pages were tested using Chrome DevTools under controlled network conditions to measure Page Load Time (PLT) improvements, component complexity correlation, and perceived responsiveness. Results show that the optimized SPA reduced initial load time by an average of 20.71%, with performance gains correlating to page complexity ($r = 0.83$). The proposed framework demonstrates a systematic approach for integrating and evaluating Lazy

Loading strategies in educational SPAs, contributing to performance optimization research in web-based learning systems

11:45 ***Exploring User Engagement in Social Commerce through Social and Technology Factors: A Pilot Test in the Indonesian Market***

Hanny Juwitasary (Bina Nusantara University, Indonesia); Dr Norizan Anwar Action (Universiti Teknologi MARA, Malaysia); Mohd Nasir Ismail (Universiti Teknologi MARA Cawangan Kelantan, Malaysia); Yohannes Kurniawan and Violitta Yesmaya (Bina Nusantara University, Indonesia)

Technological advancements have driven a shift in e-commerce, where social media platforms, as venues for social interaction, play a significant role in consumer engagement, leading to the emergence of social commerce. This study investigates the influence of social factors (social support, social interaction, social presence) and technological factors (information quality, service quality, and system quality) on user engagement, with relationship quality and motivational factors acting as mediators. Based on the Information Systems Success Model, a pilot study was conducted with 37 social commerce users in Indonesia to validate the measurement model and assess the initial relationships between variables. The results indicate that social and technological dimensions significantly influence motivational and relational factors, which in turn drive user engagement across affective, cognitive, and behavioral dimensions. The novelty of this study lies in extending the Information Systems Success Model to the context of social commerce by integrating social dimensions and motivational constructs. Practically, these findings provide insights for platform providers in designing features that enhance user trust, motivation, and engagement.

2025 International Conference on Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics (ICoABCD) | 979-8-3315-5749-2/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICOABCD67551.2025.11470741

Committee

Steering Committee
Assoc. Prof., Dr. Eng. Khoirul Anwar (ASASI, IDN)
Prof. Dr. Brian Kurkoski (Japan Advanced Institute of Science and Technology, JPN)
Prof. Dr.-Ing. Kalamullah Ramli (Universitas Indonesia, IDN)
Prof. Dr. Elfahmi (ASASI, IDN)
Prof. Ir. Moh. Khairudin, MT., PhD., IPU. (Universitas Negeri Yogyakarta, IDN)

General Chair
Dr. Hilal H. Nuha (Telkom University, IDN)

Program Chair
Assoc. Prof. Dr., Nurul Hidayat (ASASI, IDN)
Indra Lukmana, M.T. (ASASI, IDN)

Secretariate

Dr.Eng. Muhammad Almaududi Pulungan (Telkom University, IDN)
Moch, Fahru Rizal, S.T., M.T (Telkom University, IDN)
Treasure

Dana Kusumo, Ph.D. (Telkom University, IDN)
Catur Prasetiawan (Telkom University, IDN)

Sponsorship

Ferdian (TDA, IDN)
Hendrik Ria Hifman (ASASI, IDN)

TPC Chair

Dr. Gelar Budiman (ASASI, IDN)
Dayat Kurniawan (IEEE Indonesia Section)

Track Chair

Dick Maryopi, Ph.D. (ASASI, IDN)

TPC Member

Mada Sanjaya WS, Ph.D. (UIN Sunan Gunung Djati Bandung, IDN)
Muhammad Reza Kahar Aziz, Ph.D. (Institut Tenologi Sumatera, IND)
Ade Irawan, Ph.D. (Pertamina University, IDN)
Dr. Erwin Budi Setiawan (Aptikom, IDN)
Dr. Norulhusna Ahmad (Universiti Teknologi Malaysia, MYS)
Prasetiyono Hari Mukti (IEEE Indonesia Section)
Fardan, M.Sc. (Telkom University, IDN)
Shibbir Ahmed Arif (Montclair State University, USA)

Publication Chair

Isman Kurniawan, Ph.D. (Telkom University, IDN)

Event and Logistic

Caca Sukma (Telkom University, IDN)
Zulfa Fauzia, S.Ds. (ASASI, IDN)

Webmaster

Yahdi Siradj, M.T. (Telkom University, IDN)

Organized by Telkom University
Prepared by EDAS Conference Services.

2025 International Conference on Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics (ICoABCD) | 979-8-3315-5749-2/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICOABCD67551.2025.11470688

Author Index

T

Organized by Telkom University
Prepared by EDAS Conference Services.

2025 International Conference on Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics (ICoABCD) | 979-8-3315-5749-2/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICOABCD67551.2025.11470845

2025 International Conference on Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics (ICoABCD)

1A (*Session 1A*)

1B (*Session 1B*)

1C (*Session 1C*)

1D (*Session 1D*)

1E (*Session 1E*)

2A (*Session 2A*)

2B (*Session 2B*)

2C (*Session 2C*)

2D (*Session 2D*)

3B (*Session 3B*)

3C (*Session 3C*)

3D (*Session 3D*)

Comparison of Machine Learning Models on Balanced and Imbalanced Dataset of Rodent Tuber

Iwan Binanto*
Department of Informatics,
Faculty of Science and Technology,
Sanata Dharma University,
Yogyakarta, Indonesia
iwan@usd.ac.id

M. Rizky Fajar Mali
Department of Informatics,
Faculty of Science and Technology,
Sanata Dharma University,
Yogyakarta, Indonesia
rizkyfadjarmali@gmail.com

Basilius Arilla Dimas N
Department of Informatics,
Faculty of Science and Technology,
Sanata Dharma University,
Yogyakarta, Indonesia
basiliusarilla@gmail.com

Ajitama Jaya
Department of Informatics,
Faculty of Science and Technology,
Sanata Dharma University,
Yogyakarta, Indonesia
anjitama38@gmail.com

Nesti F. Sianipar
Food Technology Department
Faculty of Engineering
Research Interest Group Biotechnology
Bina Nusantara University,
Jakarta, Indonesia
nsianipar@binus.edu

Abstract—In this study, we investigate the performance of various machine learning models on both balanced and unbalanced datasets. The data processing pipeline begins with loading and preprocessing the dataset, followed by two parallel approaches: handling balanced and unbalanced data. For the balanced data approach, we employ random undersampling to achieve class balance. The Rodent Tuber's unbalanced dataset, which includes chemical data useful for cancer diagnosis, was used in this study. We use and evaluate Random Forest classifiers and Support Vector Machines (SVM) with three different kernels: sigmoid, linear, and radial basis function (RBF) on balanced and unbalanced datasets. Based on our research, the Random Forest algorithms generated the best classification results on balanced datasets, achieving a total 8.9-second runtime with 99% accuracy, precision, recall, and F1 score. On balanced datasets, however, SVM performance differed greatly between kernels; linear kernels produced 50% accuracy, RBF kernels 61%, and sigmoid kernels 45%. The models' ineffectiveness is shown by their poor performance on unbalanced datasets, which mostly identified the majority class. This emphasize the negative impact of class imbalance. Furthermore, running the algorithm on the imbalanced dataset took significantly longer than on the balanced dataset.

Keywords—rodent tuber, classification, support vector machine, random forest, balanced and imbalanced data, random undersampling

I. INTRODUCTION

Cancer is one of the diseases that poses a major threat to world health. Despite many efforts to find effective medicines, various kinds of cancer still lack a really effective therapy. Medicinal plants are used for a variety of purposes in both conventional and alternative medicine. Medicinal plants include phytochemicals that may help prevent cancer and other chronic disorders [1]. The development of cells that are resistant to many medications is one of the reasons cancer treatments is not always effective. Since herbs use natural ingredients to treat cancer, there may not be any negative side effects. Rodent tuber plant is one of the plants that may have therapeutic qualities that could help prevent cancer [2], [3].

Researchers have used Typhonium flagelliforme Lodd, or rodent tuber, in alternative medicine studies. Several methodologies were employed in the research on rodent tuber plants, including in vitro, in vivo, and bioinformatics approaches [2], [4], [5]. Similarly, addressing data imbalance is a critical aspect of machine learning studies.

Imbalanced dataset complicates the interpretability of model outputs because performance metrics like accuracy may give a misleading impression of success, while important outcomes for minority classes remain poorly predicted. Addressing imbalance through methods like oversampling, undersampling, or specialized algorithms is essential to ensure that models support equitable and clinically reliable decision-making in biomedicine.

Rahmi employs random undersampling to improve classification performance on microarray datasets with class imbalances by selecting data sets, computing the difference between the majority and minority classes, and then randomly removing data from the majority class [6]. As a successful algorithm, Random Forest has been widely used in both academia and business to tackle such challenges [7].

Random Undersampling is a technique that removes samples from the majority class and iteratively reselects them until the desired class distribution is achieved, making it particularly effective for addressing cases with significant class imbalance. Random Undersampling was chosen in this study because it is computationally efficient, simple to implement, and avoids the risk of introducing synthetic or noisy samples. This technique is especially suitable for large datasets where some information loss from the majority class is acceptable, and data authenticity is important. By focusing on naturally occurring data and reducing computational demands, Random Undersampling allows for interpretable and resource-efficient analysis appropriate for the study's goals.

II. LITERATURE REVIEW

Numerous studies on rodent tuber plant and the significance of categorizing medicinal plants to optimize their

health benefits for humans have been performed [8], [9], [10], [11].

In study addressing imbalanced dataset on depression among university students, a combination of random oversampling and Tomek links undersampling was used to mitigate the data imbalance. The Random Forest algorithm, applied in this research, achieved a notable accuracy rate of 94.17% [12]. Similarly, the effectiveness of the Random Forest algorithm in handling imbalanced datasets is further exemplified in the work of Yang and Ismail [13]. They introduced the Random Forest as a crucial component of the RF-MICA approach for identifying PV errors. By combining decision trees, Random Forest significantly enhances predictive performance, achieving impressive accuracy and precision rates of 99.88% and 99.43%, respectively, with computational times of 42.20 seconds and 15.51 seconds for different scenarios. These studies collectively emphasize the robustness and efficiency of the Random Forest algorithm in various applications involving imbalanced data [13].

Meanwhile, Saffariha et al. developed an SVM model to accurately forecast hypericin concentration in the Hypericum perforatum plant species, achieving a model fit value of 0.74 [14]. Similarly, Hikmah et al. studied three anti-cancer plants in West Sulawesi using SVM with linear, polynomial, RBF, and sigmoid kernels using 80% training and 20% testing data. They achieved an accuracy of 0.92 using linear, polynomial, and RBF kernels and an accuracy value of 0.87 using the sigmoid kernel [15].

Additionally, Sabat-Tomala et al. identified three invasive or expansive plant species—Rubus spp., Calamagrostis epigejos, and Solidago spp.—using hyperspectral air imaging data from the HySpex sensor. The Random Forest algorithm achieved an accuracy of 86% for classifying invasible plant speeds such as Euphorbia escula and Centaurea maculosa in Montana using HyPex hyperspektral citra and air data, while the Support Vector Machines (SVM) algorithm achieved 93% accuracy for classifying species Solanum mauritianum shrubs in the Pinus patula plantation in South Africa using the AISA Eagle image. The Support Vector Machines (SVM) algorithm also achieved an F1-Score of approximately 0.90 for classifying invasively or expansionally planted species like Rubus spp., Calamagostis epigeos, and Solidago spp. using hyperspektral air image data of HySpiex [16].

In another study, data on breast cancer from the UCI Machine Learning Repository was utilized. This study employed Random Forest (RF) and Support Vector Machines (SVM) as the machine learning techniques, achieving an accuracy of 90% for Random Forest and 95% for Support Vector Machines [17].

Moreover, a comparative analysis of Random Forest classifier, Support Vector Machine, and Artificial Neural Network was conducted to classify multiclass brain tumors based on MRI images. The analysis revealed variations in the performance of the three classification models, with the Random Forest classifier having an accuracy of 0.7, weighted precision of 0.55, weighted recall of 0.7, and weighted F1 score of 0.59. The Support Vector Machine demonstrated an accuracy of 0.71, weighted precision of 0.5, weighted recall of 0.71, and weighted F1 score of 0.59, while the Artificial Neural Network achieved an accuracy of 0.62, weighted precision of 0.6, weighted recall of 0.62, and weighted F1 score of 0.61 [18].

III. Research Method

As illustrated in Figure 1, we offer a visual breakdown of the research process to aid in a full understanding of it.

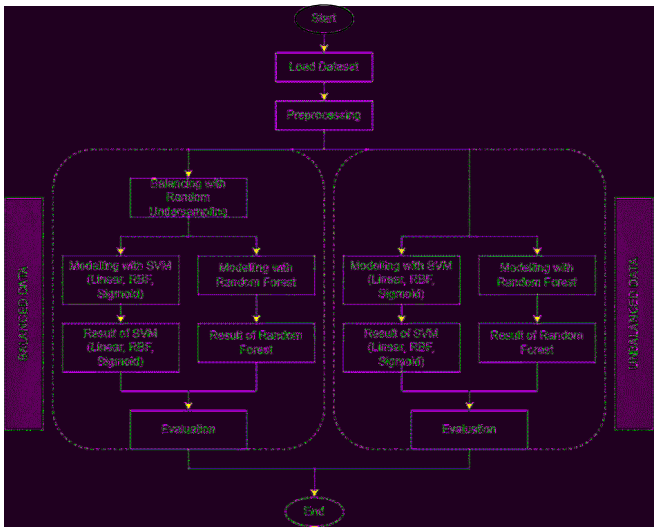

Fig. 1. Research Method Diagram

This study use the Rodent Tuber dataset, which comprises approximately 663,228 rows and 4 columns. Our calculations to determine whether the Rodent Tuber dataset is balanced or imbalanced show that it is unbalanced, with a ratio of 98 to 2 between status 0 and status 1. This ratio consists of approximately 653,398 data points for status 0 and 9,830 for status 1, as shown in Figure 2.

The dataset is significantly unbalanced, thus balancing will be done to guarantee correct classification results. Random Undersampling is the algorithm utilized to balance the dataset in this study. Random Undersampling is selected due to its simplicity, efficiency, and ability to balance classes, making it suitable for imbalanced learning problems where computational resources or time are constraints. It is the simplest way to handle imbalanced datasets by randomly removing samples from the majority class, thereby balancing the class distribution. This allows the classification model to learn without bias towards the majority class.. Figure 3 displays the outcomes the completion of the balancing process.

Stages of preprocessing and modeling were used to prepare the data. Preprocessing is done to eliminate redundant and missing data and normalize.

As an initial step, the data was separated into two dataframes: X (Retention Time, Intensity, Real_m/z) and y (Status). Retention Time and Intensity are native features of the LC-MS raw data. Retention time refers to the time a specific compound spends in the chromatographic column from injection to detection. It is characteristic for each compound and helps identify substances based on how long they are retained in the column. Intensity is the measure of the signal strength corresponding to the abundance or concentration of the detected analyte at a given retention time. While Real_m/z and Status are features derived from previous research [9].

The Standard Scaler was used to normalize the X dataframe. Random undersampling is used to correct imbalanced dataset, as shown in Figure 2, and generate balanced dataset for each class, namely status, as shown in

Figure 3. The data was split into training and testing sets in an 80:20 ratio to yield thorough results.

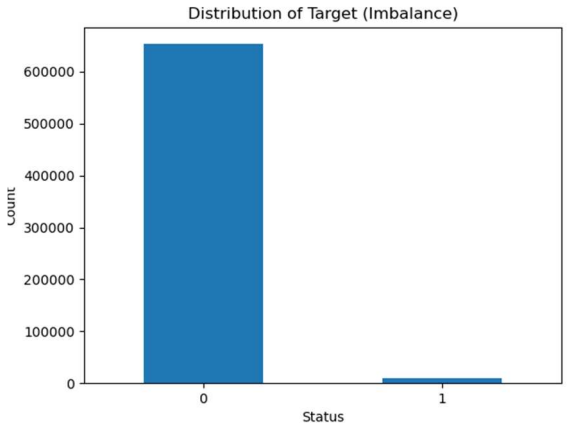

Fig. 2. Unbalanced Data

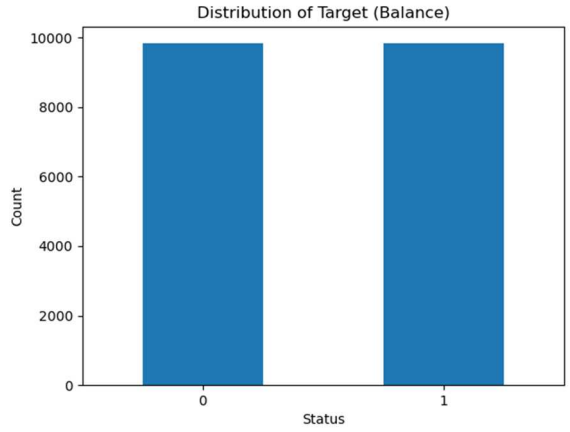

Fig. 3. Balanced Data by Random Undersampling

Eighty percent of the training data is being balanced. To get thorough results, the remaining data is left imbalanced as testing data.

All kernels of the Support Vector Machine algorithm will utilize default parameters as shown in Table 1. In a similar manner, the Random Forest algorithm uses default parameters.

TABLE I. THE DEFAULT PARAMETERS

No.	Classification Algorithm	Parameter
1	SVM (Linear Kernel)	C=1.0, tol=1e-3
2	SVM (RBF Kernel)	C=1.0, gamma='scale'
3	SVM (Sigmoid Kernel)	C=1.0, gamma='scale', coef0=0.0
4	Random Forest	n_estimators=100

We used a computer with the specification listed in Table 2 for this experiment in order to obtain objective results.

TABLE II. COMPUTER SPECIFICATION

Processor	11th Gen Intel(R) Core(TM) i7-11800H @ 2.30GHz 2.30 GHz
Memory	16.0 GB (15.7 GB usable)
Storage	512 Gb, 1x M.2 SSD slot (NVMe PCIe Gen3)

IV. RESULTS AND DISCUSSIONS

The dataset needs to be balanced because it is extremely lopsided, with a ratio of roughly 98:2 between target 0 and 1. As seen in Figure 3, balancing is accomplished by using Random Undersampling, which produces a smaller dataset. The outcome of the process is a new dataset that contains around 9,830 data points for target 0 and 9,830 for target 1, with a 1:1 ratio between the two.

Table 3 displays each classification algorithm on the balanced and imbalanced for F1 score, accuracy, precision, and recall. The table also shows the time in seconds when the program runs the algorithms on both balanced and imbalanced datasets.

TABLE III. EXPERIMENTS RESULTS

	Random Forest		SVM (Linear)		SVM (RBF)		SVM (Sigmoid)	
	Balance	Unbalance	Balance	Unbalance	Balance	Unbalance	Balance	Unbalance
Precision	0,99	0,99	0,25	0,49	0,61	0,49	0,45	0,50
Recall	0,99	0,68	0,50	0,50	0,61	0,50	0,45	0,50
F1-Score	0,99	0,76	0,33	0,50	0,61	0,50	0,45	0,50
Accuracy	0,99	0,99	0,50	0,99	0,61	0,99	0,45	0,97
Execution Time (s)	8,9	182,6	10,7	444,3	27,4	465,9	16, 2	2969,6

With an overall duration of 8.9 seconds, Random Forest algorithms produced the best classification results on balanced datasets, with accuracy of 99%, precision of 99%, recall of 99%, and score F1 of 99%. Linear kernels yield 50% accuracy, RBF kernels 61%, and sigmoid kernels 45% for Support Vector Machines (SVMs) given balanced datasets. The

problem stems from the unequal distribution within the dataset. When the model is trained on imbalanced data, it tends to recognize only the majority class, which results in poor modeling performance. It shows in the three Support Vector Machines (SVMs) kernels with 50% recall but accuracy above 97%. Additionally, running the algorithm on the imbalanced dataset takes a lot longer more than 4000% than it does on the balanced dataset .

Random Undersampling works by randomly removing samples from the majority class. This creates two main challenges for the model which are (1) Information Loss which mean Random Undersampling risks missing important samples that are at class boundaries or represent minority subpopulations within the majority class; (2) Increased Variability (Noise) which mean random sample selection can introduce new bias or noise if the remaining sample does not adequately represent the entire majority population.

Based on the findings from this study, it is showed that data balancing significantly impacts the performance of machine learning models. Random Forest classifiers demonstrated superior performance on balanced datasets, achieving high accuracy, precision, recall, and F1 scores. Support Vector Machines (SVM), on the other hand, showed inconsistent results based on the kernel that was applied; the linear kernel fared the poorest, while the radial basis function (RBF) kernel performed somewhat better. It is crucial to resolve class imbalance in datasets because the models' performance on unbalanced datasets was noticeably low and tended to favor the majority class. Moreover, the increased computational time required for processing unbalanced data further highlights the practical benefits of data balancing techniques.

V. Conclusions

Since Random Forest is an ensemble-based model composed of multiple decision trees, with the default parameter n_estimators set to 100, it incorporates built-in strategies that effectively mitigate the issues caused by Random Undersampling. This inherent robustness makes Random Forest the most suitable model in this study. In contrast, Support Vector Machines (SVM) are more vulnerable to the negative effects of Random Undersampling because they rely heavily on support vectors and clear decision boundaries, which makes them highly sensitive to the random removal of data points.

Random Forest's ensemble nature reduces the risk of information loss during undersampling and enables the model to utilize the remaining data more efficiently. Meanwhile, SVM tends to be affected by the loss of critical boundary samples, which can degrade its performance.

Overall, this study highlights the importance of using data balancing techniques to improve the predictive accuracy and efficiency of machine learning models applied to imbalanced datasets. To further enhance model reliability, researchers are encouraged to combine Random Forest with rigorous preprocessing steps such as data cleaning, normalization, and careful dimensionality reduction. Additionally, implementing stratified cross-validation can help reduce bias, while evaluating models using metrics beyond accuracy—such as recall and F1-score for minority classes—provides a more comprehensive assessment.

Exploring other data balancing methods, including synthetic oversampling techniques like SMOTE or cluster-based approaches alongside ensemble models, can offer deeper insights into data structure and inherent complexities, particularly in chemical or biomedical datasets.

For future research, applying more advanced and comprehensive strategies for handling data imbalance is recommended to achieve optimal model performance. Also, relying solely on a single train-test split can lead to biased and less representative performance estimates, further underscoring the necessity of cross-validation for more accurate and reliable evaluations.

Adopting these advanced balancing and interpretability-focused evaluation methods is crucial for translating machine learning outcomes into meaningful and trustworthy insights, especially when working with highly imbalanced real-world datasets in biomedical and chemical research.

References

[1] C. Godswill, "Medical Plants: The Medical, Food, and Nutritional Biochemistry and Uses," 2019.

[2] T. Siswanto, R. Shofiati, and I. H. Afifah, "Typhonium flagelliforme as a Cancer Prevention Plant-Based on In Vitro, In Vivo and Bioinformatics Research Method: A Systematic Literature Review," *Nusantara Science and Technology Proceedings*, pp. 39–43, Mar. 2022, doi: 10.11594/nstp.2022.2106.

[3] N. F. Sianipar, K. Assidqi, S. Yuliani, and R. Purnamaningsih, "Anticancer activity of nanoemulsion formulation of rodent tuber mutant extract (Typhonium flagelliforme) on human breast cancer cell line," *Rasayan Journal of Chemistry*, vol. 14, no. 1, pp. 535–544, 2021, doi: 10.31788/RJC.2021.1415742.

[4] N. F. Sianipar, K. Assidqi, and B. S. Abbas, "The Effects of Subculture on the Mutant Plant Regeneration of Rodent Tuber (Typhonium flagelliforme) in Vitro Mutagenesis Using Gamma-Ray Irradiation," in *IOP Conference Series: Earth and Environmental Science*, Institute of Physics Publishing, Mar. 2020. doi: 10.1088/1755-1315/426/1/012180.

[5] N. F. Sianipar, R. Reflinur, M. D. Ashan, K. Assidqi, and R. Purnamaningsih, "Identification of Specific DNA Fragments in the Superior Mutant Plant of Rodent Tuber Accession Pekalongan (<i>Typhonium flagelliforme</i>) Based on Sequencing Analysis," *Materials Science Forum*, vol. 1069, pp. 231–238, Aug. 2022, doi: 10.4028/p-78ru0f.

[6] N. S. Rahmi, "Ensemble-support vector machine-random undersampling: Simulation study of multiclass classification for handling high dimensional and imbalanced data," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Sep. 2020. doi: 10.1088/1742-6596/1613/1/012064.

[7] F. A. Rafrastara, C. Supriyanto, C. Paramita, Y. P. Astuti, and F. Ahmed, "Performance Improvement of Random Forest Algorithm for Malware Detection on Imbalanced Dataset using Random Under-Sampling Method," vol. 8, no. 2, 2023, [Online]. Available: https://orangedatamining.com/

[8] N. F. Sianipar and R. Purnamaningsih, "TROPICAL AGRICULTURAL SCIENCE Enhancement of the Contents of Anticancer Bioactive Compounds in Mutant Clones of Rodent Tuber (Typhonium flagelliforme Lodd.) based on GC-MS Analysis," *Pertanika J. Trop. Agric. Sci*, vol. 41, no. 1, pp. 305–320, 2018, [Online]. Available: http://www.pertanika.upm.edu.my/

[9] I. Binanto, H. L. H. S. Warnars, N. F. Sianipar, and W. Budiharto, "Webscraping Data Labelling System on Liquid Chromatography-Mass Spectrometry of Rodent Tuber for Efficiency of Supervised Learning Preprocessing," *ICIC Express Letters, Part B: Applications*, vol. 13, no. 1, pp. 107–114, Jan. 2022, doi: 10.24507/icicelb.13.01.107.

[10] I. Binanto, H. L. H. S. Warnars, N. F. Sianipar, and W. Budiharto, "Mining frequent pattern on liquid chromatography-mass spectrometer data of rodent tuber to find the association rules of compounds for machine learning," *ICIC Express Letters, Part B: Applications*, vol. 12, no. 12, pp. 1185–1191, Dec. 2021, doi: 10.24507/icicelb.12.12.1185.

[11] I. Binanto, N. Fronika Sianipar, and H. Leslie Hendric Spits Warnars, "Understanding LCMS Data for Identification of Chemical

Compounds Contained in Rodent Tuber: Timeseries or Not", doi: 10.31838/srp.2021.1.93.

[12] S. Sawangarreerak and P. Thanathamathee, "Random forest with sampling techniques for handling imbalanced prediction of university student depression," *Information (Switzerland)*, vol. 11, no. 11, pp. 1–13, Nov. 2020, doi: 10.3390/info11110519.

[13] N. C. Yang and H. Ismail, "Robust Intelligent Learning Algorithm Using Random Forest and Modified-Independent Component Analysis for PV Fault Detection: In Case of Imbalanced Data," *IEEE Access*, vol. 10, pp. 41119–41130, 2022, doi: 10.1109/ACCESS.2022.3166477.

[14] M. Saffariha, A. Jahani, R. Jahani, and S. Latif, "Prediction of hypericin content in Hypericum perforatum L. in different ecological habitat using artificial neural networks," *Plant Methods*, vol. 17, no. 1, Dec. 2021, doi: 10.1186/s13007-021-00710-z.

[15] H. Hikmah, N. H. A Syahrir, and P. I. Rahayu, "Application of Support Vector Machine for Class Imbalance Learning to Predict Anticancer Compounds of Medicinal Plants in West Sulawesi," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 18, no. 1, pp. 0141–0150, Mar. 2024, doi: 10.30598/barekengvol18iss1pp0141-0150.

[16] A. Sabat-Tomala, E. Raczko, and B. Zagajewski, "Comparison of support vector machine and random forest algorithms for invasive and expansive species classification using airborne hyperspectral data," *Remote Sens (Basel)*, vol. 12, no. 3, Feb. 2020, doi: 10.3390/rs12030516.

[17] C. Aroef, Y. Rivan, and Z. Rustam, "Comparing random forest and support vector machines for breast cancer classification," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 18, no. 2, pp. 815–821, Apr. 2020, doi: 10.12928/TELKOMNIKA.V18I2.14785.

[18] Amaliah Faradibah, Dewi Widyawati, A Ulfah Tenripada Syahar, and Sitti Rahmah Jabir, "Comparison Analysis of Random Forest Classifier, Support Vector Machine, and Artificial Neural Network Performance in Multiclass Brain Tumor Classification," *Indonesian Journal of Data and Science*, vol. 4, no. 2, pp. 54–63, Jul. 2023, doi: 10.56705/ijodas.v4i2.73.