8th International Conference on Information System and Data Mining
ICISDM 2024

7th International Conference on Knowledge Management Systems
ICKMS 2024

Los Angeles, CA, USA

June 24-26, 2024

Co-Sponsored by

Table of Contents

Welcome Message

Thank you for your attendance at the 8th International Conference on Information System and Data Mining (ICISDM 2024), 7th International Conference on Knowledge Management Systems (ICKMS2024). It is such excited that we bring the event to "The City of Angels" this year through much preparation.

ICISDM and ICKMS are co-organized by South Asia Institute of Science and Engineering (SAISE), University of North Dakota, USA, University of South Carolina Aiken, USA and Illinois State University, USA. It is poised to be a melting pot of participants from academia, industries and government agencies to share and exchange experiences, new ideas and enabling technologies. Those technologies will address the enormous challenges confronting the new landscape of science, engineering and technology pertinent to information systems, data mining, knowledge management systems. We would like to express our sincere thanks to the organizers, Illinois State University, USA, University of North Dakota, University of South Carolina and all committee members, for their devotion to the conference.

After several rounds of review, the program committee has finalized those papers accepted into our conference proceedings. We wish to express our sincere appreciation to all the individuals who have contributed to conferences. Special thanks are also extended to our colleagues in the program committee for their thorough review of all the submissions, and also to the members in the organizing committee and the volunteers who had dedicated their time and efforts in planning, promoting, and organizing the conferences. Last but not least, our special thanks go to invited keynote speakers as well as all the delegates for their contribution to the conferences. This conference program is highlighted by two keynote speaker and two invited speakers. Moreover, to inspire the young, we set two parallel sessions to encourage detailed exchanges and one poster session for free discussion.

Enjoy your participation at ICISDM, ICKMS2024 and we wish you have a fruitful encounter during the conferences in Los Angeles. Most of all, please take good care of yourself during the participation. Looking forward to meeting you next year in our conferences!

ICISDM & ICKMS Organizing Committee

The E-Central Downtown Los Angeles Hotel

Address: 1020 South Figueroa Street
Los Angeles, California
Phone: 213.748.1291
E-Mail Address: reservations@ecentraldtla.com

LAX

-

US

From Airport to the E-Central DTLA Hotel
By LAX-T7-ll (32 minutes)

☐ 1. LAX Terminal 7 Lower Level FlyAway Stop

☐ 2. Union Station FlyAway Patsaouras Bus Plaza

Metro J Line 910/950 San Pedro - Downtown LA J – Line (16 minutes, 7 stops)

☐ 3. Union Station Patsaouras Bus Plaza (Metro J Line)

☐ 4. Flower / Olympic Stop

Oral Presentation Guidelines
English, 15 minutes

Presentation Tips

1. Timing: a maximum of 15 minutes total, including speaking time and discussion. Please make sure your presentation is well timed. Please keep in mind that the program is full and that the speaker after you would like their allocated time available to them.

2. You can use USB flash drive (memory stick), make sure you scanned viruses in your own computer. Each speaker is required to meet her/his session chair in the corresponding session rooms 10 minutes before the session starts and copy the slide file(PPT or PDF) to the computer.

3. It is suggested that you email a copy of your presentation to your personal inbox as a backup. If for some reason the files can't be accessed from your flash drive, you will be able to download them to the computer from your email.

4. Please note that each session room will be equipped with a LCD projector, screen, point device, microphone, and a laptop with general presentation software such as Microsoft PowerPoint and Adobe Reader. Please make sure that your files are compatible and readable with our operation system by using commonly used fronts and symbols. If you plan to use your own computer, please try the connection and make sure it works before your presentation.

5. Videos: If your PowerPoint files contain video clips, please make sure that they are well formatted and connected to the main files.

Poster Presentation

The poster must not exceed A0, 841*1189 mm, portrait style preferred. We suggest the delegate print the poster beforehand, bring it along to the conference. The font size shall be readable by the participants from 1.5 meter away. The contents should be understandable even without the presenter's details explanation.

Dress Code

Please attend the conference in formal attire.

Conference Photos

All the conference photos will be available for download through conference website within one/two week(s) after the conference.

Safety Reminder

We remind you to secure your personal belongings at all times.
Please remember to:
* Wear your Conference Identification Badge at all times. Do not throw away Badge.
* If you are using a laptop computer, do not leave it unattended at any time.
* Keep your purse, wallet and other valuables with you at all times.
The conference organizer will not be responsible for the loss or damage to any personal belongings.

Committee

Conference Chair

Anu Gokhale, St. Augustine's University, USA

Program Committee Chairs

Emanuel Grant, University of North Dakota, USA

Jiangjiang Liu, Lamar University, USA

Program Committee Co-Chairs

Mohamed Zakaria Kurdi, Lynchburg College. USA

Muhammad Abid, Florida Polytechnic University, USA

Publicity Chair

María Olmedilla Fernández, SKEMA Business School – Paris Campus, France

Technical Committee Chairs

John Burris, Technology Southeastern Louisiana University, USA
Sikh Bagui, University of West Florida, USA
Subhash Bagui, University of West Florida, USA
Fu-Shing Sun, Ball State University, USA
Farhan Aadil, COMSATS University Islamabad, Pakistan
Sanjay Tyagi, Kurukeshetra University, India
Narameth Nananukul, Sirindhorn International Institute of Technology, Thailand
Liaqat Ali, University of Science and Technology of Fujairah UAE, United Arab Emirates
Bahjat Fakieh, King Abdulaziz University, Saudi Arabia
Puntani Pongsumpun, King Mongkut's Institute of Technology Ladkrabang, Thailand
K. Poulose Jacob, Cochin University of Sc & Tech., India
Tathagata Mukherjee, University of Alabama in Huntsville, USA
Chaity Banerjee Mukherjee, University of Alabama in Huntsville, USA
Thepul Ginige, Universal College Lanka, Sri Lanka
Munam Shah, COMSATS Institute of Information Technology, Pakistan
Sharon Andrews White, University of Houston - Clear Lake, USA
Genlin Ji, Nanjing Normal University, China
R Venkateswaran, Persistent Systems Ltd,, India
Ying Xie, Kennesaw State University, USA
Iluminada Vivien R. Domingo, Polytechnic University of the Philippines, Philippines
Uzair Iqbal, National University of Computer and Emerging Sciences, Pakistan
Muhammad Umar Aftab, National University of Computer and Emerging Sciences, Pakistan
Naveed Ahmed, Prince Sultan University, Saudi Arabia
Yasir Javed, Prince Sultan University, Saudi Arabia
Amjad Gawanmeh, University of Dubai, UAE
Sara Sutton, Grand Valley State University, USA

Thank you for your contributions to ICISDM&ICKMS2024.
We appreciate all your efforts to the success.

Program Overview

Day 1	The E-central Downtown Los Angeles Hotel **Lobby, First Floor on June 24**[th]

14:00-17:00   Conference Kits Collection and Attendees Signage

Day 2	**Olympic Room, Second Floor on June 25**[th]

09:00-18:00   Conference Keynote Speech and Author Presentations.

Time	Activity	Facilitator
09:00-09:05	Opening Remarks Prof. Anu Gokhale, St. Augustine's University, USA	
09:05-09:50	Keynote Speech **Applying AI in Software Development Education** Prof. Emanuel S. Grant, University of North Dakota, USA	Prof. Anu Gokhale
09:50-10:10	Group Photo & Coffee Break	
10:10-10:55	Keynote Speech **AI in Information Systems: Algorithm Design and Applications** Prof. Anu Gokhale, St. Augustine's University, USA	Prof. Emanuel S. Grant
10:55-11:30	Invited Speech **Ranking Alternatives by an Extension to Fuzzy VIKOR** Prof. Ta-Chung Chu, Southern Taiwan University of Science and Technology, Taiwan	Prof. Emanuel S. Grant
11:30-12:05	Invited Speech **Automatic Diagnosis of Alzheimer's Disease using Lexical Features extracted from Language Samples** Assoc. Prof. Mohamed Zakaria Kurdi, University of Lynchburg in Virginia, USA	Prof. Anu Gokhale
12:05-13:30	Lunch Break	
13:30-15:30	Session 1 **Data Based System Model And Intelligent Computing**	Dr. Yasir Javed
15:30-16:00	Poster Session and Coffee Break	
16:00-18:00	Session 2 **Digital Image Analysis And Processing**	Assoc. Prof. Amjad Gawanmeh
18:00-20:00	Dinner	

Day 3	June 26th

One-day Tour in Los Angeles

Keynote Speaker

Prof. Anu Gokhale, St. Augustine's University, USA

Keynote Speech Start Date

June 25th, 2024

Start Time

10:10-10:55

Venue

Olympic room, Second Floor

Biography

Dr. Anu A. Gokhale is currently a Professor and Chair of the Department of Computer Information Systems at Saint Augustine's University. She has been selected as a 2023-24 Convergence Fellow by the Association of American Colleges & Universities. Gokhale visited Cairo University in Egypt in August 2022 as Fulbright Specialist in Data Analytics. Formerly, she was a Distinguished Professor and Coordinator of the Computer Systems Technology program at Illinois State University (ISU). Gokhale has completed thirty years as faculty and has received several College and University research, teaching and service awards. Having earned certifications in online delivery, she was recruited to mentor colleagues in online teaching beginning March 2020. Gokhale was named Fulbright Distinguished Chair in STEM+C at the University of Pernambuco, Brazil, 2016-17; was a Faculty Fellow in Israel and Fulbright Specialist in Cybersecurity at Gujarat Technological University, India in summer 2017. As a Visiting Professor in College of Business at Shandong University in Jinan, China in spring 2017, her focus was on e-commerce. Her achievements encompass extensively cited refereed publications; groundbreaking externally funded research supported by a continuous 20-year stream of grants from state and federal agencies including the National Science Foundation; and elevation of the ISU student experience through excellence in teaching, mentorship, and the creation of opportunities for students to get involved in research. Originally from India, she has a master's in physics−electronics from the College of William & Mary, and a doctorate from Iowa State University. Dr. Gokhale authored a second edition of her book Introduction to Telecommunications, which has an international edition in Chinese. She continues to be an invited keynote speaker at various conferences. As an active volunteer in IEEE, she has served as R4 Educational Activities Chair, Women in Engineering Coordinator, and MGA representative to the Educational Activities Board. She was honored with the IEEE Third Millennium Medal and 2019 Region 4 Outstanding Professional Award. She consults for business and industry to increase productivity using data analytics and business intelligence while leveraging e-technologies. She has delivered multiple workshops focusing on hybrid teaching & learning, inclusion & diversity, as well as on algorithms and data analytics.

Abstract

Enterprise information systems combined with latest developments in data mining strategies have created unprecedented opportunities for enhancing competitive advantage. Executives seek to leverage Artificial Intelligence (AI), the biggest driver of technological change, to inform decision-making. Corporate data environments include both structured and unstructured information and there exists tremendous potential to glean key insights for business advantage from the vast data that is available today and new data that is being constantly generated. Algorithms used in analyzing big data vary significantly based on the problem of study and its goals and objectives. The talk will address the issues and processes associated with analyzing big data in business information systems, applicable algorithms to enhance functionality and predictive analytics, and discuss how data-driven decisions support product/service innovation.

Keynote Speaker

Prof. Emanuel S. Grant, University of North Dakota, USA

Keynote Speech Start Date

June 25th. 2024

Start Time

09:05-09:50

Venue

Olympic Room, Second Floor

Biography

Emanuel S. Grant received a B.Sc. from the University of the West Indies, MCS from Florida Atlantic University, and a Ph.D. from Colorado State University, all in Computer Science. Since 2008, he is an Associate Professor in the Department of Computer Science (August 2002 – June 2018) and the School of Electrical Engineering and Computer Science (June 2018 – present) at the University of North Dakota, USA, where he started as an Assistant Professor in 2002. He currently serves as the Associate Director of the School of Electrical Engineering and Computer Science (SEECS) and SEECS Graduate Program Director. His research interests are in software development methodologies, formal specification techniques, domain-specific modelling languages, model-driven software development, software engineering education, and ethics for software engineering. Emanuel Grant has conducted research in software engineering teaching with collaborators from Holy Angel University, Philippines; HELP University College, Malaysia; III-Hyderabad, India; Singapore Management University, Singapore; Montclair State University, and University of North Carolina Wilmington of the USA; and the University of Technology, Jamaica. He is affiliated with the SEMAT (Software Engineering Method and Theory) organization, as a member of the Essence - Kernel and Language for Software Engineering Methods (Essence) group. Emanuel is a member of the Association for Computing Machinery (ACM), Upsilon Pi Epsilon (UPE), and the Institute of Electrical and Electronics Engineers (IEEE).

Abstract

Artificial intelligence (AI) is being applied at a pace that challenges the verification of its suitability to the domains of application. This situation may arise from the proliferation of AI development being conducted from a data science point of view rather than a software engineering approach. This observation leads to the question of whether software development course curricula are addressing the necessary educational needs for graduates to respond to the challenges of applying AI development in emerging domains. The challenge has two parts: the first is the use of AI in developing software systems, and the second is the development of AI systems. The work addressed in this paper looks at the first part of this challenge by proposing a pedagogy for introducing AI tool usage in software engineering education as a springboard for structuring a methodology for AI application development that incorporates established software engineering principles. This work is exploratory in nature and first reviews existing works in this area of interest, with the aim of identifying commonalities of approaches towards selecting a required set of topics, course outcomes, and structure for a curriculum on AI in software development.

Invited Speaker

Prof. Ta-Chung Chu, Southern Taiwan University of Science and Technology, Taiwan

Invited Speech Start Date

June 25th. 2024

Start Time

10:55-11:30

Venue

Olympic Room, Second Floor

Biography

Ta-Chung Chu received his Ph.D. degree from the Department of Industrial Engineering at the University of Texas at Arlington, USA. Dr. Chu is a professor in the Department of Industrial Management and Information, Southern Taiwan University of Science and Technology, Taiwan; he served as Department Chair from August 2010 to July 2016. His research interests are in fuzzy multiple criteria decision-making, fuzzy number ranking and their applications. Dr. Chu has published 53 refereed journal papers and 88 conference papers. He served as a reviewer for various international journals and has reviewed 243 journal papers. He was the Chairman of the Conference of Industrial Management and Information Application Innovation, Taiwan (IMIAI 2013/2014). Dr. Chu was listed in 「Career-Long Impact List」 and 「Year Impact List」 of the World's Top 2% Scientists 2021/2022, and was listed in 「Career-Long Impact List」 in 2023. Currently, he serves as Academic Editor of PLOS ONE (SCIE/IF:3.7/Q2) and Guest Editor of the Special Issue of Axioms (SCIE/IF:2.0/Q2): Advances in Fuzzy MCDM, Hybrid Methods, Fuzzy Number Ranking and Their Applications (https://www.mdpi.com/journal/axioms/special_issues/7F551OEE24).

Abstract

An extension to fuzzy VIKOR is proposed, where membership functions of fuzzy weighted weightings can be derived. The proposed extension can resolve the limitation of using approximation to represent multiplication of two positive triangular fuzzy numbers in existing methods. The total integral value is used to rank fuzzy numbers and formulas of ranking procedure are presented to help complete the proposed model. Finally, a numerical example is used to demonstrate the feasibility of the proposed extension.

Invited Speaker

Assoc. Prof. Mohamed Zakaria Kurdi, University of Lynchburg in Virginia, USA

Invited Speech Start Date

June 25th. 2024

Start Time

11:30-12:05

Venue

Olympic Room, Second Floor

Biography

Dr. Kurdi is an Associate Professor of Computer Science at the University of Lynchburg in Virginia, USA. In addition to his Ph.D. in CS, he has an interdisciplinary background in Software Engineering, Cognitive Science, and Linguistics. Before joining the University of Lynchburg, he worked in several institutions in North America and Europe. His research interests are in text and data mining and their applications to areas like intelligent computer-assisted language education, authorship attribution, bioinformatics, and Social Network Analysis (SNA). He authored a two-volume textbook about Natural Language Processing (NLP) that was published in French and English. His recent work on text mining won two best paper awards and a nomination from three different international conferences.

Abstract

Background: this study has a twofold goal. First, it aims to improve the understanding of the impact of Dementia of type Alzheimer's Disease (AD) on different aspects of the lexicon. Second, it aims to demonstrate that such aspects of the lexicon, when used as features of a machine learning classifier, can help achieve state-of-the-art performance in automatically identifying language samples produced by patients with AD.

Methods: the main dataset used is derived from the ADDreSS challenge, which is a part of the DementiaBank data set. This dataset consists of transcripts of descriptions of the Cookie Theft picture, produced by 54 subjects in the training part and 24 subjects in the test part. The number of narrative samples is 108 in the training set and 48 in the test set. First, the impact of AD on 99 selected lexical features is studied using both the training and testing parts of the main dataset. Then some machine learning experiments were conducted on the task of classifying transcribed speech samples with text samples that were produced by people with AD from those produced by normal subjects. Several experiments were conducted to compare the different areas of lexical complexity, identify the subset of features that help achieve optimal performance, and study the impact of the size of the input on the classification. To evaluate the generalization of the models built on narrative speech to new tasks two generalization tests were conducted using written data from two British authors, Iris Murdoch and Agatha Christie, and the transcription of some speeches by former President Ronald Reagan.

Results: using lexical features only, state-of-the-art classification, F1 and accuracies of over 91% were achieved in categorizing language samples produced by individuals with AD from the ones produced by healthy control subjects. This confirms the substantial impact of AD on lexicon processing. The task generalization tests show that the system scales well to a new task.

Conclusions: AD has a substantial impact on the lexicon. Hence, with lexical information only, one can achieve state of the art classification of language samples produced by AD patients vs normal control patients.

Keywords: Alzheimer's Disease automatic diagnosis, lexical complexity, lexical diversity, lexical density, lexical sophistication

Session 1 - Data Based System Model and Intelligent Computing

✧ Please attend the whole session until the end. The session chair will announce the excellent presentation winner and we will have session group photo.

✧ Please enter the conference room at least 10 minutes earlier to copy your slides to the Laptop and check its availability.

✧ The certificate will be given by session chair at the end of the session.

Session Chair: Dr. Yasir Javed, Prince Sultan University Riyadh, Saudi Arabia

Solomon Buke Chudo	MS24-501-A	13:30-13:45

Time Series Models with Multiple Seasonal Periods: An Evaluation of their Forecasting Performance

Solomon Buke Chudo

University of Debrecen, Hungary

Currently, many time series possess complex seasonality. Integer, non-integer, multiple nested, non-nested, and non-integer multiple seasonality intervals are some of the most common complex types of seasonality. This article presents a comprehensive evaluation of time series forecasting performance using models that accommodate multiple seasonal periods. The study focuses on four prominent models: BATS, TBATS, STL+ETS, and DSHW. The objective is to assess the efficacy of these models in capturing and forecasting time series data exhibiting intricate seasonal patterns. American Electric Power's (AEP) hourly consumption data was used to endorse the method of statistical analysis. Parameter estimations, forecasting, and performance comparisons were made based on these four complex seasonality period models. R software version 4.1.2 was used to analyze the data and fit the models. There was clear daily and weekly seasonality in the time series. So, the multiple time series class models with multiple seasonality periods are preferable. A Box-Cox logarithmic transformation has been applied to adjust the variability and non-linearity. The models were fitted for hourly energy power consumption forecasts based on statistical significance. The model combinations are evaluated by considering models with the least values of: ME, RMSE, MAE, MPE, MAPE, Autocorrelation Function (ACF1), and Theil's U statistic. The Multiple seasonal and trend decomposition using the Loess model (STL) has the lowest measurement values compared to the other methods. So, according to the least values, the STL model shows the best forecasting performance over the TBATS, BATS, and DSHW models. Results from the evaluation aim to provide insights into the parameter estimation, fitting the best model, strengths and limitations of each model, aiding practitioners and researchers in selecting the most suitable model for their specific time series forecasting needs. The findings contribute to advancing the understanding of multiple seasonal period modeling, offering valuable guidance for practitioners seeking optimal forecasting solutions in various domains.

Yasir Javed	MS24-5009	13:45-14:00

Cyber-Threat-Intelligence: Detection of Selective Packet Drops and Fake Packet Attacks in Opportunistic Networks (DOSFAN)

Yasir Javed, Waqar Khalid, Naveed Ahmed, Arshad Alam, Muhammad Arshad

Prince Sultan University Riyadh, Saudi Arabia

An Opportunistic Network is a revolutionary technology making a huge impact on human society. It gained significant attention from researchers, due to its potential application domains like Vehicular Networks, Delay Tolerant Networks, or Internet of Things, etc. In case of an Opportunistic Network, whenever the source node transfers packets to destination node in the presence of malicious nodes, the malicious nodes drops few legitimate packets and inject malicious(fake) packets to launch attacks. Several techniques have been proposed by researchers to mitigate the issue, however hackers are ways to overcome these new approaches and still find a ways to compromise the communication. Due to this research, this research propose a modified algorithm based on our previously proposed algorithm FAPMIC. We used Cyber-Threat-Intelligence Strategy to take preventive measure against malicious nodes which launch selective packet drops and fake packet attacks simultaneously. The proposed modified algorithm enhances the detection accuracy, detection delay, false positive/false negative rates, and resources consumption and also improves packet delivery ratios/packet loss ratios.

Kayanna A Morgan	MS24-5260	14:00-14:15

Survey of Soft Computing Methods used to Predict Win/Loss Probability in Soccer

Kayanna A Morgan, Emanuel S Grant, Eunjin Kim

University of North Dakota, United States

Worldwide sports activities have one of the largest supporter/fan bases of all the many areas of entertainment. The sport of soccer is one of the richest sports in the world and commands the position of being the most popular sport in the world. That single data point forms the basis for an interesting set of data analytical research questions. First on that list would be: How is soccer determined to be the most popular sport in the world? How do the factors of advertising/marketing, televising/streaming, teams' win/loss, players, location, etc. contribute to the popularity of the sport? What impact does the popularity of soccer have on the players' performance? Can data analytics predict teams' performance? The report presented herein is a documentation and review of prior research efforts that have applied soft computing methods to predict teams' win/loss probability in the sport of soccer. This review comments on the success of the data analytics applied and makes an assessment of how this can affect the popularity of the sport.

| Chia-Wen Kao | MS24-5033 | 14:15-14:30 |

Forecasting Seasonal Sales in Supermarket Chains Using FFT-GA and LightGBM Models

Chia-Wen Kao, Jui-Chung Hung

University of Taipei, Taiwan

This paper presents a two-stage prediction model designed to effectively address the fluctuations in sales volume experienced by supermarket chains due to seasonal patterns, sales cycles, holidays, and irregular variations. In the initial phase, the model combines raw sales data with Fast Fourier Transform (FFT) to uncover seasonal characteristics and then employs a Genetic Algorithm (GA) with its binary encoding capabilities for feature selection.In the subsequent stage, the model leverages the Light Gradient Boosting Machine (LightGBM) as the prediction model, using weak learning strategies to manage irregular variations in sales data. Empirical simulations using three publicly available datasets of supermarket chains from Kaggle validate the effectiveness of our proposed method. This approach holds significant practical implications for understanding and predicting fluctuations in supermarket sales data and offers a new direction for future research. Moreover, the design of this model provides novel solutions for predictive problems in other domains. The source code will be publicly available at GitHub.

| Amjad Gawanmeh | MS24-5041 | 14:30-14:45 |

Behavioral Economics Influences On Decision-making In Online Electronics Retail

Fares Al Ahbabi, Hamzah Ali Alkhazaleh, Amjad Gawanmeh

University of Dubai, UAE

This work is intended to address the relationship between behavioral economics and decision-making in the realm of online electronics retail. The work examines the influences of key behavioral economics principles such as Loss Aversion, Social Proof, Endowment Effect, and Status Quo Bias on consumer behavior through empirical analysis. Utilizing a Transactional Retail Dataset and employing a quantitative approach, the study investigates the nuanced cognitive biases and social factors that shape consumers' perceptions of product value, pricing, and overall satisfaction. By addressing gaps in the understanding of consumer decision-making in the digital age, this research offers valuable insights for businesses aiming to navigate the evolving electronic retail sector effectively. The insights gained from this research have practical implications for businesses operating in the online electronics retail sector. By understanding how behavioral economics principles influence consumer behavior, businesses can optimize their marketing strategies, pricing models, and customer experiences. For example, businesses may leverage Loss Aversion by offering discounts or promotions that emphasize potential savings, capitalize on Social Proof by highlighting positive customer reviews and product popularity, and mitigate the effects of Status Quo Bias by providing personalized recommendations and introducing novelty in product offerings. Overall, this research provides actionable insights that can enhance businesses' competitiveness and success in the dynamic landscape of online electronics retail.

Tuan Tran	MS24-5480	14:45-15:00

Unveiling Valuable Skills and Credentials in the Information Technology Job Market through Social Media Mining

Tuan Tran, Dong Nguyen

Sullivan University, USA

In the dynamic realm of information technology, aligning educational curricula with market demands is crucial for producing employable graduates and supporting industry growth. This research addresses the significant lag in the integration of rapidly evolving technologies, such as artificial intelligence and machine learning, into academic programs and identifies discrepancies in skill requirements across various sectors, including cybersecurity. Utilizing a novel approach, we developed a software tool to extract data from extensive job postings on a leading job advertisement platform, www.highered.com. Our analysis provides a detailed examination of the IT job market, highlighting geographical distribution of opportunities, required technical skills, experience levels, and necessary credentials across different sectors. The findings reveal skill mismatches and offer critical insights for educators to update course offerings and tailor educational programs to better meet industry needs. This paper also serves as a guide for policymakers in resource allocation and infrastructure development, aiming to enhance graduate employability and address sector-specific demands effectively. The study not only highlights the importance of timely curriculum updates but also demonstrates the benefits of leveraging real-time labor market data to inform educational strategies.

John W. Burris	MS24-5049	15:00-15:15

Bridging Theory and Practice: Preparing Future IT Service Leaders through Experiential Learning in Cybersecurity Policy and Governance

John W. Burris

Southeastern Louisiana University, United States

This paper proposes implementing online, multi-disciplinary tabletop exercises as an experiential learning approach to better prepare future IT service leaders in cybersecurity policy and governance. The interactive simulations will present scenarios spanning regulatory compliance, incident response, risk mitigation, and strategic planning across various industries. Participants will assume diverse roles (e.g. CISOs, legal counsel, operations managers) and apply critical thinking, problem-solving, and communication skills to address the evolving situations. This innovative pedagogy leverages cloud-based tools to facilitate remote, team-based learning experiences that bridge the gap between theoretical concepts and pragmatic application while fostering crucial virtual coordination abilities. The paper outlines the design principles, learning objectives, and anticipated outcomes of these cyber tabletop exercises, aiming to cultivate a holistic cybersecurity mindset that empowers future leaders to develop robust governance frameworks aligning people, processes, and technologies.

| Chih-Hsiang Ko | MS24-5420 | 15:15-15:30 |

User Acceptance and Quality Requirements of Digital Twins

Chih-Hsiang Ko

National Taiwan University of Science and Technology, Taiwan

Digital twins provide conditions for cyber-physical integration, as a bridge connecting the physical world and the cyber world, providing a new way for the manufacturing industry to conduct smart production and precise management. Data from the real world is transmitted to the virtual model through sensors to complete the simulation, verification, dynamic adjustment, and feedback. By improving the designer's ability to extract knowledge from large and complex data, it can speed up design and development. Therefore, it is necessary to have a deep understanding of how to provide data to designers and how they interact with the data. Its quality will determine the usability and efficiency of digital twins. This study emphasized on the benefits that digital twins could provide to designers in data processing and integration, and accelerated the realization of design innovation through the interactive design model of virtual reality. The results of this study could help enhance the designer's ability to control the data and strengthen the way the designer interacts with the data.

Session 2 - Digital Image Analysis and Processing

- ✧ Please attend the whole session until the end. The session chair will announce the excellent presentation winner and we will have session group photo.
- ✧ Please enter the conference room at least 10 minutes earlier to copy your slides to the Laptop and check its availability.
- ✧ The certificate will be given by session chair at the end of the session.
  Session Chair: Assoc. Prof. Amjad Gawanmeh, University of Dubai, UAE

Hany Al Ashwal	MS24-5043	16:00-16:15

Alzheimer's Disease Identification through Learning from Missing values : Insights from ADNI Merged Datasets

Hamzah Dabool, Hany Al Ashwal, Ahmed Moustafa

United Arab Emirates University, United States

Researchers often encounter significant hurdles when dealing with datasets that contain a vast number of missing data points. This predicament forces them to make a tough choice: either discard a substantial amount of data, which could drastically undermine the accuracy of the machine learning (ML) model or attempt to fill these gaps in sensitive medical datasets, a method that is far from ideal. This paper proposes an approach to this issue, suggesting that bypassing the traditional path of data imputation in favor of a model that learns from the missing values themselves could paradoxically improve the accuracy and predictive capabilities of Alzheimer's Disease (AD) identification models. We introduce a comparison between state-of-the-art ML models and the XGBoost algorithm, which is designed to integrate the learning of missing values into its training cycle, using the official ADNI datasets with extensive missing data. The experiment further evaluates these models on the same datasets post-imputation. The results strikingly indicate that this unconventional strategy not only bridges the gap created by missing data but also surpasses the accuracy of traditional methods that rely on filling in incomplete samples. This discovery opens up new avenues for research in medical diagnostics for conditions like AD, where data scarcity and imperfections are common. By rethinking how we handle incomplete data, we unlock a new potential for refining ML applications in healthcare, particularly in enhancing the precision of diagnoses in complex diseases such as AD.

Sakher Khalil Alqaaidi	MS24-540	16:15-16:30

Relation Prediction in Knowledge Graph Completion using Large Language Models

Sakher Khalil Alqaaidi, Krzysztof Kochut

University of Georgia, United States

Knowledge Graphs have been widely used to represent facts in a structured format. Due to their large scale applications, knowledge graphs suffer from being incomplete. The relation prediction task obtains knowledge graph completion by assigning one or more possible relations to each nodes pair.

In this work, we make use of the knowledge graph node names to fine-tune a large language model for the relation prediction task. By utilizing the node names only we enable our model to operate sufficiently in the inductive settings. Our experiments show that we accomplish new scores on a widely used knowledge graph benchmark.

Iluminada Vivien R. Domingo	MS24-E01	16:30-16:45

Nude Content Detection on Video Using Haar Cascade Classifier

Erick Evener G. Tamondong, Shairhem Evans R. Fernandez, Iluminada Vivien R. Domingo, DBA

Polytechnic University of the Philippines, Philippines

Nudity is also an element in pornography. There are individuals that are conservative about what they are watching. There are children that are too exposed in nudity and it becomes a normal scene and it leads to changes of their attitudes. That is why this study developed a system that can detect frontal view of female breast, frontal view of female genital and frontal view of male genital in videos. The researchers tested the accuracy of the system in detecting nude objects in different resolutions. The researchers used experiment method for this research and used formula of Precision, Recall and F-Measure to obtain the accuracy of the system in detecting frontal view of female breast, frontal view of female genital and frontal view of male breast in different resolutions. The system's accuracy got different average precision, average recall and average f-measure in each model depending on their resolution. The higher the resolution is, the higher the accuracy of the model it gets. The researchers recommend the use of relative position of the human body in detecting the breast, female genitals, and male genitals improve the accuracy of the system. For example, the detected breast and genital candidate outside of Human body ROI will be marked as negative. Finally, use other object detection algorithms that can recognize different angles to improve robustness, and other classification algorithms such as Support Vector Machine and Artificial Neural Network.

Md Mamun Bin Ibne Reaz	MS24-5010	16:45-17:00

Machine Learning Driven Feature Selection for Accurate CKD Prediction in Diabetes: A Dual Type Exploration

Md Nakib Hayat Chowdhury, Sawal Hamid Md Ali, Mamun Bin Ibne Reaz, Fahmida Haque, Md. Johirul Islam, Shamim Ahmad, Andrés Cicuttin, María Liz Crespoa

Independent University, Bangladesh

Diabetes stands as the primary cause of chronic kidney disease (CKD), emphasizing the critical need for early identification of CKD in diabetes patients for effective treatment. This research addressed the challenge of asymptomatic CKD progression in diabetes patients, who typically do not undergo CKD-related tests during routine checkups. Our study focused on developing a feature selection strategy to create CKD prediction models for type 1 diabetes (T1D) and type 2 diabetes (T2D) patients, utilizing readily available features from patients' routine checkups. Leveraging two distinct datasets, the Early Diabetes Intervention Program (EDIC) for T1D and the Chronic Renal Insufficiency Cohort (CRIC) for T2D, we employed machine learning-based prediction models, including Gradient

Boosting Machines (GBM), Extreme Gradient Boosting (XGB), Extremely Randomized Trees Classifier (ERT), Random Forest (RF), and CatBoost. An iterative backward feature selection approach was applied to optimize model performance. The XGB model showcased exceptional predictive accuracy of 95.81% in T1D, while the Extra Trees model outclassed with an accuracy of 89.11% in T2D patients. In T1D, key features for predicting CKD include hypertension, duration of insulin-dependent diabetes mellitus (IDDM), age, SBP, LDL, total cholesterol, ACE inhibitors, drinker, and triglycerides. For T2D, crucial features comprise any cardiovascular disease (CVD), age, DBP, hemoglobin A1c, BMI, SBP, HDL, total cholesterol, LDL, triglycerides, sex, and smoking. These findings offer targeted insights for accurate CKD prediction and personalized interventions in diabetes care. The robust performance of our proposed models underscores their potential in harnessing routine checkup data to significantly improve early detection and intervention strategies for CKD across diverse diabetic populations.

Zesheng Hu	MS24-5013	17:00-17:15

Keypoints-based multimodal network for robust human action recognition

Zesheng Hu, Genlin Ji, Jiaquan Gao, Bin Zhao, Xichen Yang

Nanjing Normal University, China

Skeleton-based action recognition has garnered widespread attention. However, due to the inherent limitations of skeleton sequences, existing works often confuse actions with inter-class similarities and struggle to meet the requirement for viewpoint invariance. As a solution, multimodal action recognition leverages the complementarity of information between modalities to significantly enhance the performance of unimodal models. However, effectively integrating these modalities remains an open problem. In this work, we first propose a keypoints-based multimodal data fusion method to construct images that adequately represent the crucial spatiotemporal characteristics and their variations of actions. Building upon this, we introduce the keypoints-based multimodal fusion network (KBMN), which comprehensively learns action features from skeleton, RGB, and depth data. Extensive experiments on two large-scale datasets demonstrate that our KBMN exhibits robust performance in both unimodal and multimodal action recognition tasks. As an auxiliary model for skeleton-based methods, KBMN effectively assists various baseline methods in improving their recognition accuracy.

Abdulrauf Aremu Gidado	MS24-5220	17:15-17:30

iHSPRec: Image Enhanced Historical Seq uential Pattern Recommendation

Abdulrauf Aremu Gidado, C I Ezeife

University of Windsor, Ontario, Canada

Collaborative filtering faces sparsity and cold start problems, while content-based approaches suffer from overspecialization and lack of personalization. Systems such as HSPRec18, HSPRec19, and SemRec have not explored using item images to enrich recommendations. Cross-regional e-commerce systems like Amazon and eBay have retailers from disparate backgrounds. For instance, the same items are likely to be named or described differently by different retailers on these systems, such as "Slippers", "Flipflop", and "Sandals" or "Pad" as a female hygiene product or "Pad" as a

computer mouse pad. Additionally, systems such as pRNN and Caser attempt to incorporate item images into sequential recommendation using neural network models but suffer from (i) the assumption that adjacent interactions in a sequence must be dependent, which is not always true, (ii) difficulty in training and finding optimal hyper-parameters, (iii) vanishing and exploding gradient problems, and (iv) low interoperability. To improve recommendation accuracy, this paper proposes iHSPRec (Image Enhanced Historical Sequential Pattern Recommendation), which uses items' image similarity scores to build a vectorized item-item similarity matrix. Then, integrate this with a vectorized sequential pattern of users' purchases or clicks as input data for the recommendation process. This is achieved by learning the similarity between item image vectors using a convolution neural network, mining visually similar sequential purchase patterns, and enriching the item-item matrix with the products' structural similarity score and products' purchase patterns. iHSPRec provides Top-K personalized recommendations based on image similarities between items without needing a user's rating. Experimental results and comparison with existing systems show that iHSPRec has higher recommendation accuracy than benchmark methods.

Yun-Chen Tsai Morgan	MS24-550	17:30-17:45

The Application of Machine Learning to Empowerment, Stress and Workplace Outcomes

Yun-Chen Tsai Morgan, Taylor Nichols, John Wesley Burris

Southeastern Louisiana University, United States

The field of data science revolves around uncovering patterns, extracting meaningful insights, and facilitating data-driven decision making from existing data sources. This research leverages a unique dataset designed to explore the intricate relationships between stress, empowerment, and various workplace outcomes. By harnessing modern techniques spanning classification, regression, and predictive analysis, the objective is to construct a robust model that can guide management strategies and workplace decisions. Given the distinctive nature of the data, rigorous exploratory data analysis techniques are employed to unveil valuable insights. Subsequently, a comprehensive evaluation of multiple classification models is undertaken, including Nearest Neighbors, Linear Support Vector Machines, Radial Basis Function Support Vector Machines, Gaussian Process, Decision Tree, Random Forest, Multi-Layer Perceptron, AdaBoost, and Naive Bayes. The findings provide invaluable business intelligence, elucidating how factors such as accountability, personality traits, and authority-sharing dynamics are likely to influence individual employees' workplace outcomes. Notably, the results demonstrate that Linear Discriminant Analysis, applied in conjunction with multi-dimensional reduction methods, yielded the most accurate predictions among the evaluated models.

Mehmet Salih Goceri · **MS24-E06** · **17:45-18:00**

Quality Assessment of Mobile Banking Apps with Expert Users Using AHP Decision Making Tool

Mehmet Salih Goceri

Santa Clara University , USA

This paper presents a distinctive approach for evaluating the quality of mobile apps in Turkish banks. The assessment is conducted using the Analytic Hierarchy Process (AHP) method, setting it apart from traditional methods. The criteria for app evaluation were determined through interviews with 10 banking industry experts, resulting in a comprehensive list of features considered essential for banking apps in Turkey. Additionally, input from banking managers was gathered to determine the most important criteria and their relative importance. By applying the AHP method, the weight of each criterion was calculated, incorporating both qualitative and quantitative measures. The study stands out for its novel approach to evaluating mobile app quality, specifically in Turkish banks. By utilizing the AHP method and incorporating expert perspectives, the evaluation provides a robust and comprehensive assessment of app features. The findings have the potential to drive improvements in mobile app development and enhance user satisfaction within the Turkish banking industry.

Poster Session

15:30-16:00

- ✧ We will provide the tape and assist the author to stick it on the wall.
- ✧ Please ask the staff to help you take a photo if needed.
- ✧ The certificate will be prepared after the session.

Emanuel S Grant	MS24-5270	Poster 1

Verification and Validation of Software System Class Models

Kruti P Shah, Emanuel S Grant

University of North Dakota, United States

The verification and validation of software system class models are crucial aspects of software engineering, aiming to ensure correctness, reliability, and quality in software systems. This research explores advancements in verification and validation techniques, focusing on UML class diagram models. The study proposes a methodology to enhance the reliability and security of software models by simplifying UML class diagrams, eliminating semantic overload, and validating models using the Object Constraint Language (OCL). The practical and theoretical values of this research are discussed, highlighting benefits for developers, system administrators, end-users, cybersecurity professionals, organizations, and academic communities. The work addresses challenges and limitations in current verification and validation techniques, emphasizing the complexity of modern software systems. The proposed methodology is implemented using the USE tool, demonstrating its effectiveness in validating UML-based specifications and OCL constraints.

Paul Shapiro	MS24-E4-A	Poster 2

Teaching with Simulation and Gaming: From Development to Practice - The Project Management Game 3.0

Paul Shapiro

University of Maryland, United States

The Project Management (PM) Game is an innovative and experiential learning tool for the instruction of both undergraduate and graduate PM courses that applies course content to real world contexts. The online simulation reinforces the learning outcomes related to planning and managing projects by enabling the hands-on execution of a realistic information systems project. Student achievement is directly assessed by the simulation via a performance based ranking system of project management constraints. The tool was developed using open-source technologies including Node,js, React, and MySQL. Since development in the summer of 2020, the PM simulation has been used by over a thousand students (1500+) across 35 course offerings and 8 semesters at University of Maryland's Robert H. Smith Business School. The current version, 3.0, reflects three years of continuous improvement in system features and functions.

Di Ye	MS24-604-A	Poster 3

The Impact of Internal Processes and External Factors on Industry-University-Research Cooperation Performance in the Digital Era

Di Ye, Jun Zhang

National Huaqiao University,China

Previous studies have predominantly focused on the individual entities involved in Industry-University-Research cooperation (IURC) and the process of knowledge transfer, while lacking a comprehensive perspective on external influencing factors. This study applies knowledge transfer theory, the triple helix model, and transaction cost theory to identify seven influencing factors related to the internal processes of knowledge transfer. By analyzing 223 survey responses, this study employs necessary condition analysis (NCA) and qualitative comparative analysis based on fuzzy sets (fsQCA) to examine the mechanisms of the impact of internal processes and external factors on the performance of IURC from an interactive perspective. The study findings reveal that knowledge sharing, organizational distance, absorptive capacity, organizational unlearning, digitalization level, and government support play a key role in the performance of university-industry research cooperation, with multiple concurrent causal relationships existing among these factors. Also, six configuration paths lead to high performance in IURC, categorized as digital empowerment driven, government-enterprise synergy driven, and internal and external integration driven patterns, demonstrating that the seven influencing factors can yield equivalent result paths under different configurations. The conclusions of this study provide important recommendations for all parties involved in IURC to enhance its performance.

Shin Ando	MS24-5029	Poster 4

An Ensemble OOD Detection with Norm-enhancing Representation Learning

Takeshi Kounoike, Shin Ando

Tokyo University of Science, Japan

Diabetes stands as the primary cause of chronic kidney disease (CKD), emphasizing the critical need for early identification of CKD in diabetes patients for effective treatment. This research addressed the challenge of asymptomatic CKD progression in diabetes patients, who typically do not undergo CKD-related tests during routine checkups. Our study focused on developing a feature selection strategy to create CKD prediction models for type 1 diabetes (T1D) and type 2 diabetes (T2D) patients, utilizing readily available features from patients' routine checkups. Leveraging two distinct datasets, the Early Diabetes Intervention Program (EDIC) for T1D and the Chronic Renal Insufficiency Cohort (CRIC) for T2D, we employed machine learning-based prediction models, including Gradient Boosting Machines (GBM), Extreme Gradient Boosting (XGB), Extremely Randomized Trees Classifier (ERT), Random Forest (RF), and CatBoost. An iterative backward feature selection approach was applied to optimize model performance. The XGB model showcased exceptional predictive accuracy of 95.81% in T1D, while the Extra Trees model outclassed with an accuracy of 89.11% in T2D patients. In T1D, key features for predicting CKD include hypertension, duration of insulin-dependent diabetes

mellitus (IDDM), age, SBP, LDL, total cholesterol, ACE inhibitors, drinker, and triglycerides. For T2D, crucial features comprise any cardiovascular disease (CVD), age, DBP, hemoglobin A1c, BMI, SBP, HDL, total cholesterol, LDL, triglycerides, sex, and smoking. These findings offer targeted insights for accurate CKD prediction and personalized interventions in diabetes care. The robust performance of our proposed models underscores their potential in harnessing routine checkup data to significantly improve early detection and intervention strategies for CKD across diverse diabetic populations.

Aetesam Ali Khan Ashar	MS24-5140	Poster 5

A Survey on Object Detection and Recognition for Blurred and Low-Quality Images: Handling, Deblurring, and Reconstruction

Aetesam Ali Khan Ashar, Asir Abrar, Jiangjiang Liu

Lamar University, USA

Object detection and recognition in blurred and low-quality images present significant challenges in computer vision and image processing. This survey paper provides a comprehensive overview of the state-of-the-art techniques, methodologies, and advancements in addressing these challenges. Blurred and low-quality images are encountered in various real-world scenarios, such as surveillance, medical imaging, and autonomous vehicles, making robust object detection and recognition essential. This paper reviews the key issues, datasets, evaluation metrics, and recent advancements in this field, with an emphasis on deep learning-based approaches. Particular emphasis is placed on the integration of popular object detection models such as SSD (Single Shot Multibox Detector), COCO (Common Objects in Context), and YOLO (You Only Look Once) with deblurring techniques. Through this survey, we aim to provide researchers and practitioners with valuable insights into the current landscape of object detection and recognition in challenging imaging conditions, facilitating further research and application development in this critical domain.

Asir Abrar	MS24-5340	Poster 6

A Survey on Early Stage Dementia Detection Using Natural Language Processing: Datasets and Approaches

Asir Abrar, Aetesam Ali Khan Ashar, Jiangjiang Liu,

National Huaqiao University,China

According to the World Health Organization, Dementia, a chronic, degenerative condition that affects 55 million people worldwide, is prevalent in seniors. To date, there is no cure for dementia, so recent advancements in this area emphasize the urgent need for detecting early symptoms to facilitate timely intervention strategies. As dementia starts by damaging neurons in parts of the brain responsible for memory, language, and thinking, the analysis of language sample could potentially offer a promising avenue for detecting subtle cognitive shifts that could allow the possible interventions to slow down the disease's progression and improve the life of the individuals. We document this paper to explore the potential of leveraging speech and text data for detecting early-stage dementia by leveraging semantic, syntactic, and acoustic features. This paper surveys natural language processing (NLP) techniques applied to speech, text, audio, and handwritten data in monolingual and multilingual settings, focusing on their potential to aid in the early detection and understanding of dementia. According to our study, most datasets available in this domain are in

English, with the support vector machine being the most frequently used classification method. Interest is also growing in using large language models to identify the signs of cognitive decline based on language patterns

Peyman Mashhadi	MS24-520	Poster 7

Steered Metaheuristic Optimization Approach for Vehicle Breakdown Prediction

Mohsen Tajgardan, Reza Khoshkangini, Peyman Mashhadi

Halmstad University, Sweden

Vehicle fault prediction is becoming one of the main goals in manufacturers' maintenance strategies to reduce the number and severity of quality problems in vehicles. Hundreds of vehicle sensors can be used for the early detection of component breakdowns. This work introduces a breakdown prediction approach based on vehicle usage over time. This study proposes a steered optimization system using an evolutionary algorithm called Genetic Algorithm coupled with an Elastic technique to select the most informative predictors. Then, a specific kind of ensemble technique, namely stacking, is utilized for the final prediction. The proposed system has been applied to a complex problem of predictive maintenance to forecast components' failures. The experimental evaluations on the real usage data collected from thousands of heavy-duty trucks justify the proposed approach is promising.

Shilpa Rajan	MS24-530	Poster 8

The Role of Knowledge Graph in Smart Cities

Shilpa Rajan, Enayat Rajabi, Reza Khoshkangini

Cape Breton University, Canada

With the invention of advanced technologies, there are millions of options to improve the quality of life in an urban city. Several innovative implementations transform urban cities into smart cities using new technologies to enhance urban inhabitants' efficiency, sustainability, and overall quality of life. Our study shows that knowledge graphs play an important role in smart cities for transportation, parking, traffic, and city development. They serve as significant repositories, bringing together data from various sources. Several crucial domains of smart cities use knowledge graphs to resolve challenges that hinder urban development. In this paper, we discuss the applications of knowledge graphs in various smart city areas, identify existing challenges, and propose strategies to enhance the current implementation of knowledge graphs in smart cities. We highlight a few innovations that used knowledge graphs in smart cities, showcasing their versatility. Integrating knowledge graphs into smart cities significantly enhances the efficiency of urban services by consolidating data from different sources and constructing a graph, aiding in better decision-making.

Yi Sun	MS24-5031	Poster 9

The Application of Constructing Knowledge Graph of Oral Historical Archives Resources Based on LLM-RAG

Yi Sun, Wanru Yang, Yin Liu

Shanghai Jiao Tong University, China

Oral historical archive resources are an emerging archive resource with the rapid development of modern technology. Its "bottom-up" approach to historical research has received widespread attention in the fields of history, archives, and libraries. Under the common knowledge discovery mode, oral historical archives resources are showing a dispersed state. Information technology represented by knowledge graphs can break through the data solidification of oral historical archives, reshape the information stack of oral historical archives, and achieve knowledge association and aggregation of oral historical archive resources. The article attempts to construct a knowledge graph of the oral historical archives resources on the theme of "science and art" in the collection of T.D. Lee Library of Shanghai Jiao Tong University. It uses Large Language Model - Retrieval Augmented Generation (LLM-RAG) for knowledge extraction, and then uses a semantic model for knowledge organization and management. The article attempts to empower humanities with technology, exploring the possibility of combining "digital technology" and "humanities research", extending traditional humanities research methods, breaking down barriers between technology and humanities resources, and providing a new path reference for revealing resource content characteristics, semantic deep correlation, and multi-dimensional knowledge discovery.

Sara Sutton	MS24-5045	Poster 10

A Generative AI Powered Approach to Cyberbullying Detection

Mohammad Shafiqul Islam, Sara Sutton, Rahat IBN Rafiq

Grand Valley State University, Michigan, USA

Cyberbullying poses significant risks in online social media environments, leading to severe psychological distress and societal harm. In this paper, we present a systematic approach to cyberbullying detection, aiming to mitigate these risks and protect individuals in digital spaces. Our approach leverages curated datasets from social media platforms and employs preprocessing techniques to extract vital information and calculate confidence scores for indicators of bullying and aggression. We then utilize an open-source Large Language Model (LLM) to generate specialized detection queries based on the extracted data. Responses generated by the LLM are evaluated for signs of cyberbullying using automated classification methods. Specifically, the presence or absence of key indicators within responses determines their classification. Our systematic approach enables the automated identification and categorization of cyberbullying instances, facilitating proactive intervention and prevention strategies. Through this paper, we contribute to the ongoing efforts to combat cyberbullying and promote safer online interactions.

Memo

Memo