ComURS2024

Computing Undergraduate Research Symposium 2024

ABSTRACTS

"Bridging the future with innovative minds"

21st of February 2024
Faculty of Computing
Sabaragamuwa University of Sri Lanka

Abstracts of the ComURS2024 Computing Undergraduate Research Symposium 2024

ComURS2024 Computing Undergraduate Research Symposium 2024

Organized by

**Faculty of Computing
Sabaragamuwa University of Sri Lanka**

ORGANIZING COMMITTEE

Symposium Chair
Ms. UP Kudagamage

Symposium Secretary
Ms. PMAK Wijeratne

Symposium Treasurer
Mrs. KGL Chathumini

Scientific Committee Chair
Prof. BTGS Kumara

Technical Programme Committee
Mr. P Vigneswaran   Chair

Information Systems Track
Dr. LS Lekamge   Chair
Ms. KGL Chathumini   Co-Chair

Software Engineering Track
Prof. S Vasanthapriyan   Chair
Mrs. N Ravikumar   Co-Chair

Data Science Track
Prof. BTGS Kumara   Chair
Mrs. S Wasalthilake   Co-Chair

Open Track
Dr. KPN Jayasena   Chair
Ms. S Adeeba   Co-Chair
Mrs. RMKK Wijerathna   Co-Chair

Publication Chair
Mrs. R Nirubika

Publicity Chair
Mrs. WVSK Wasalthilaka

Logistic Chair
Mr. P Vigneswaran

ADVISORY PANEL

Prof. S Vasanthapriyan   Professor,
Sabaragamuwa University of Sri Lanka

Prof. BTGS Kumara   Professor,
Sabaragamuwa University of Sri Lanka

Mr. RL Dangalla   Senior Lecturer,
Sabaragamuwa University of Sri Lanka

Dr. LS Lekamge   Senior Lecturer,
Sabaragamuwa University of Sri Lanka

Dr. KPN Jayasena   Senior Lecturer,
Sabaragamuwa University of Sri Lanka

Dr. UAP Ishanka   Senior Lecturer,
Sabaragamuwa University of Sri Lanka

EDITORIAL BOARD

Editor-in-Chief

Prof. BTGS Kumara

Editors

Prof. S Vasanthapriyan

Prof. BTGS Kumara

Mr. RL Dangalla

Dr. LS Lekamge

Dr. KPN Jayasena

Dr. UAP Ishanka

Ms. UP Kudagamage

Mr. P Vigneswaran

Mrs. KGL Chathumini

Ms. PMAK Wijerathna

Mrs. S Wasalthilake

Mrs. R Nirubikaa

Editorial Assistants

Mrs. RMKK Wijerathna

Abstract Proceeding Cover Designing

Mr. GGRSN Wimalasooriya

PANEL OF REVIEWERS

Dr. MAC Akmal Jahan	Senior Lecturer, Faculty of Applied Sciences, South Eastern University Of Sri Lanka
Dr. RAHM Rupasinghe	Senior Lecturer, Faculty of Social Sciences and Languages, Sabaragamuwa University of Sri Lanka
Mr. HMKT Gunawardhana	Lecturer (Probationary), Faculty of Computing, Sabaragamuwa University of Sri Lanka
Mr. GACA Herath	Lecturer (Probationary), Faculty of Computing, Sabaragamuwa University of Sri Lanka
Mrs. HN Gunasinghe	Lecturer (Probationary), Faculty of Computing, Sabaragamuwa University of Sri Lanka
Mrs. K Ranasinghe	MSc (Reading), University of Aizu, Japan
Mrs. YP Manawadu	Lecturer (Probationary), Faculty of Technology, Sabaragamuwa University of Sri Lanka
Mr. S Prashanth	Lecturer (Probationary), Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka
Ms. PBSN Ariyathilake	Lecturer (Probationary), Faculty of Geomatics, Sabaragamuwa University of Sri Lanka
Mr. Dimuth Pathirana	Lecturer (Probationary), Faculty Applied Sciences, Uva Wellassa University
Mr. LL Gihan Chathuranga	Lecturer (Probationary), Faculty of Science,

University of Ruhuna

Mr. M Janotheepan

Lecturer (Probationary),
Faculty of Applied Sciences,
Uva Wellassa University

Mr. Thilina Gunarathna

Lecturer (Probationary),
Faculty of Applied Sciences,
University of Uva Wellassa

Mr. Umesh Ranassoriya

Lecturer (Probationary),

Faculty of Applied Sciences,
Uva Wellassa University

Mr. WA Sanjeewa

Lecturer (Probationary),
Faculty of Applied Sciences,
Uva Wellassa University

Mr. J Charles

MSc (Reading),
Friedrich-Alexander-University
Erlangen-Nuremberg (FAU)

Mr. MA Suja Ameer

Lecturer (Probationary),
Faculty of Technology,
South Eastern University Of Sri Lanka

Ms. Akila Brahmana

Lecturer (Probationary),
Faculty of Technology,
University of Ruhuna

Ms. AMCM Aruppola

Lecturer (Probationary),
Faculty of Applied Sciences,
Sabaragamuwa University of Sri Lanka

Ms. Buddika Gayashani

Lecturer (Probationary),
Faculty of Technology,
University of Ruhuna

Ms. HKSK Hettikankanama

Lecturer (Probationary),
Faculty of Applied Sciences,
Uva Wellasa University

Ms. KAA Chathurangi	Lecturer (Probationary), Faculty of Applied Sciences, Uva Wellassa University
Ms. P Shorubiga	Lecturer (Probationary), Faculty of Applied Science, University of Vavuniya
Ms. R Wickraarachchi	Lecturer (Probationary), Faculty of Technology, University of Ruhuna
Ms. SM Shareef	Lecturer (Probationary), Faculty of Applied Sciences, South Eastern University Of Sri Lanka
Ms. A Sivaanpu	PhD Student, University of Albetra
Ms. A Chathurangi	Lecturer (Probationary), Faculty of Applied Sciences, Uva Wellassa University
Ms. K Mathangi	Lecturer (Probationary), University of Colombo School of Computing
Ms. P Amrithaa	Lecturer (Probationary), Faculty of Technology Studies, University of Vavuniya
Ms. P Premisha	Lecturer (Probationary), Faculty of Technology Studies, University of Vavuniya
Ms. P Rukshani	Lecturer (Probationary), Faculty of Technology Studies, University of Vavuniya
Ms. T Kiriharan	Lecturer (Probationary), Faculty of Applied Sciences, Eastern University of Sri Lanka

Ms. T Sukirthan   Assistant Lecturer,
                                   Faculty of Computing,
                                   SLIIT

Ms. T Marambe   MSc (Reading),
                                   Finland

TABLE OF CONTENTS

MESSAGE FROM THE VICE-CHANCELLOR

Sabaragamuwa University of Sri Lanka

As the Vice Chancellor of Sabaragamuwa Univerity of Sri Lanka, it gives me a great pleasure to convey this message for the Computing Undergraduate Research Symposium (ComURS2024) which is organized by the Faculty of Computing, Sabaragamuwa Univerity of Sri Lanka, themed on "Bridging the future with innovative minds". They determine the future of every society. I personally am a staunch believer that it is only the education which determines the progress of an individual and a nation.

I take this opportunity to congratulate the organizing committee for taking the challenge and organizing the 1st Computing Undergraduate Research Symposium in a very professional manner. The effort, encouragement and dedication of the undergraduates had on the research work would be highly appreciated. I have no doubt that symposium will offer our university undergraduates to showcase their experience, latest research findings and insight with their colleagues in the different fields of interest.

I would like to congratulate the Faculty of Computing, Sabaragamuwa University of Sri Lanka for their commitment and superb drive in organizing this symposium. I am very certain that this occasion will be able to provide a platform towards strengthening the relationships in sharing knowledge while at the same time to provide the necessary thrust in joint research collaborations and product commercialization within the research society.

Last but not least, I would also like to thank all those who have contributed to ComURS2024 with their research findings, all track chairs for their continued support and interest on us, I am sure that the quest of making Sabaragamuwa University of Sri Lanka a top class university is not going to be impossible to achieve.

I wish them success in all endevours.

Professor M. Sunil Shantha
Vice-Chancellor
Sabaragamuwa University of Sri Lanka

21st February 2024, Sabaragamuwa University of Sri Lanka, Belihuloya

MESSAGE FROM THE DEAN

Faculty of Computing
Sabaragamuwa University of Sri Lanka

As the Dean of the Faculty of Computing, Sabaragamuwa University of Sri Lanka, it is my great pleasure to pen this message to the Abstract Proceedings of the Computing Undergraduate Research Symposium (ComURS2024) which is organized by the Faculty of Computing, Sabaragamuwa University of Sri Lanka. It is with much proud to announce that this symposium will mark as the First Computing Undergraduate Research Symposium with the theme "Bridging the future with innovative minds" providing a great opportunity for the undergraduates to expose their knowledge and competencies gained from their undergraduate research projects, mini projects and so on.

Thus, this symposium will be a great avenue for undergraduates to share and publish their latest research findings under four Computing aspects; Information Systems, Software Engineering, Data Science and Open.

I would like to take this opportunity to extend my warmest congratulations to the organizing committee for their excellent efforts in hosting the ComURS2024 in a highly professional manner. I am confident that this event will provide an excellent platform for our university's undergraduates to present and discuss their latest research findings and insights in a multidisciplinary forum with their colleagues and experts.

Therefore, as the Dean of the Faculty of Computing, Sabaragamuwa University of Sri Lanka, I wish to express my sincere gratitude to the Vice-Chancellor, Sabaragamuwa University of Sri Lanka, Evaluators, Heads of the Departments, Organizing Committee of the ComURS2024 for their great effort towards making this event a success. Your dedication and hard work have not gone unnoticed.

Finally, I convey my wishes to all the presenters of the ComURS2024.

Professor S Vasanthapriyan
Dean
Faculty of Computing

MESSAGE FROM THE CHAIR
Computing Undergraduate Research Symposium
Faculty of Computing
Sabaragamuwa University of Sri Lanka

On behalf of the Organizing Committee of the Computing Undergraduate Research Symposium (ComURS2024), I'm very pleased and delighted to convey this message to the Abstract Proceedings of the Computing Undergraduate Research Symposium (ComURS2024) which is organized by the Faculty of Computing, Sabaragamuwa University of Sri Lanka. ComURS2024 is marking a milestone as the First Computing Undergraduate Research Symposium of the Faculty of Computing, Sabaragamuwa University of Sri Lanka. The symposium thrives with the theme of "Bridging the future with innovative minds" that fosters knowledge sharing across a wide range of Computing aspects in one platform.

Moreover, the symposium offers a valuable opportunity for the undergraduates of the Faculty of Computing to present their undergraduate research work and projects and get comments and suggestions from senior academics and Industry Experts to make further improvements in their research.

With a diverse range of research topics spanning various Computing aspects, attendees can expect a stimulating intellectual environment where the latest advancements are shared and discussed. ComURS2024 is comprised of presentations under four tracks; Information Systems, Software Engineering, Data Science and Open. The "Best Presenter" is selected from each track to admire the scholarly contributions of the undergraduates.

As the Chair, I am grateful to Vice-Chancellor, Sabaragamuwa University of Sri Lanka, Dean of the Faculty of Computing, Heads of the Departments and all the staff members of the Faculty of Computing, Advisory Panel, Supervisors, Evaluators of the Poster Session, Reviewers of Abstracts, Session Coordinators, Volunteers, Non-Academic Staff members, and Organizing Committee of the ComURS2024. Finally, I would like to thank all those who have worked extremely hard to organize this event. Their teamwork is deeply appreciated.

I wish all the participants an enjoyable and inspiring Research Session.

Ms. UP Kudagamage
Chair
ComURS2024

ABSTRACTS

OF THE

ComURS2024 COMPUTING UNDERGRADUATE RESEARCH SYMPOSIUM 2024

Technical Sessions of the ComURS2024 are organized as follows;

- Abstracts of the **Undergraduate Poster Presentation Session** – (pp. 04–31)
    - o Undergraduate Poster Presentations – Information System Track
    - o Undergraduate Poster Presentations – Software Engineering Track
    - o Undergraduate Poster Presentations – Data Science Track
    - o Undergraduate Poster Presentations – Open Track

Abstracts

Information Systems Track

Region Recognition in Masked Faces: A Deep Learning Based Computer Vision Approach for South and East Asian Populations

Thanistas P.E.[1*], Kumara B.T.G.S.[1], and Banujan K.[2]

[1]Department of Computing and Information Systems, Faculty of Computing, Sabaragamuwa University of Sri Lanka.
[2]Independent Researcher, Australia
[*]patrickemmanuelthanistas@email.com

People have been identified by their physical characteristics. Body characteristics have been used by biometric systems to identify individuals, and identify races. Nowadays, nationality detection is a key component in improving security measures across a variety of contexts. This technology can help in border control systems by precisely determining a person's nationality. Since December 2019, the COVID-19 epidemic has left a significant mark on numerous nations throughout the world. To stop the virus from spreading during the COVID-19 coronavirus epidemic, practically everyone wears a mask. The typical face recognition technique essentially fails in situations involving face-based identity verification, such as security checks, community visit check-ins, etc. The VGG-16 model is chosen for its proven efficacy in image classification tasks and its ability to capture intricate features within images. The research methodology involves training the VGG-16 model on a carefully curated dataset comprising diverse masked-wearing facial images from South and East Asian populations. This ensures representative coverage of gender, age, and ethnic variations. The dataset includes individuals wearing different types of masks to simulate real-world scenarios, ranging from surgical masks to traditional cultural face coverings. The evaluation of the proposed approach involves comprehensive testing on an independent dataset, measuring the model's accuracy, precision, and recall in region recognition. When the model was used to identify faces wearing masks, it achieved an impressive 91.04% accuracy rate. Comparative analyses are conducted against facial recognition methods such as MobileNetV2, and ResNet18 to highlight the strengths and improvements achieved by our tailored approach. The findings contribute to the advancement of computer vision technologies in accommodating diverse global populations, especially in the era of increased mask usage.

Keywords: *East Asian, Ethnicity, Masked Face Recognition, South Asian, VGG-16*

21st February 2024, Sabaragamuwa University of Sri Lanka, Belihuloya

Analysis and Prediction of Task-Related Mental Health Effects on Employees in Sri Lankan IT Organizations

Franklin S.V.[1]*, Kudagamage U.P.[2], and Prasanth S.[3]

[1,2]Department of Computing & Information Systems, Faculty of Computing, Sabaragamuwa University of Sri Lanka, Belihuloya.
[3]Department of Physical Sciences and Technology, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Sri Lanka.
*subasfrank1997@gmail.com

In recent years, global instances of mental health issues have surged. As Sri Lanka's IT sector experiences rapid expansion, the well-being of its employees has emerged as a pivotal concern. This study delves into the specific impact of organizational tasks on employee mental health, excluding external factors like family issues or health conditions. Utilizing advanced machine learning techniques applied to online questionnaires, the research aims to explore and address task-related mental health effects within Sri Lankan IT organizations. A comprehensive dataset was gathered through 626 online questionnaires distributed among IT professionals in various roles across Sri Lanka. The study employs eight machine learning techniques to predict task-induced mental health issues, including stress, anxiety, and depression. A questionnaire was designed based on six key factors influencing the mental health of Sri Lankan IT employees: workload, deadlines, job demands, technical responsibilities, IT-related stressors, and adaptation to new technologies. Using machine learning algorithms, the study predicts five levels of severity for stress, anxiety, and depression. Among the techniques utilized, radial basis function network, random forest, and gradient boosting, all neural network methods, yielded the highest accuracy. Notably, random forest and gradient boosting achieved the highest accuracy, with results ranging from 87.8% to 89.25% across depression, anxiety, and stress prediction. These findings have the potential to shape organizational policies and initiatives aimed at enhancing workplace well-being within the IT industry.

Keywords: Stress, Anxiety, Depression, Prediction, IT-Sector

Revolutionizing Tea Quality Measurement using Multi Method Approach: An Innovative Mobile Application Approach

Dilka H.W.C.[1*], Ilangasinghe I.M.S.C.[1], Surendra S.S.[1], Samaraweera J.D.C.[1], Pamoda G.H.M.[1], Wasalthilaka W.V.S.K.[2], and Wickramasooriya P.M.T.A.[3]

[1]Department of Computing and Information Systems, Faculty of Computing, Sabaragamuwa University of Sri Lanka, Sri Lanka
[2]Department of Software Engineering, Faculty of Computing, Sabaragamuwa University of Sri Lanka, Sri Lanka
[3]Impact IT Solutions
*chamaalidilka@gmail.com

Tea is a globally cherished beverage which requires meticulous attention to quality. Sri Lanka is one of the best high quality Tea exporters in the world. The quality of the tea is the major factor which improves the customer satisfaction as well as the income. Even Though many researches were carried out to test the tea quality, few investigations used computer based technologies. But most of those methods have many issues and drawbacks including complex methodologies, limited user awareness and accessibility issues, hindering effective quality assessment, etc. To overcome all those issues this research proposes a ground-breaking solution; a mobile application focused on a Tea Quality Measuring System. This application aims to bridge the awareness gap, empower users with knowledge about tea attributes, and enhance their ability to make informed choices. Descriptive research methodology was used for observational studies, data gathering, and making conclusions. Collected data was analysed and made the machine learning model to identify tea grades and suggest suitable tea products according to users' selection of moods by backend technology. The main feature in the mobile application is the getting tea grades and details about each tea grade by capturing the tea dust. And also, aware tea customers about tea by tea news, and help customers to select tea products according to their requirements. Finally, the proposed system achieved 85% success rate. This research endeavours to a user-friendly mobile application that not only measures tea quality but also educates users about tea attributes, promoting an informed and enjoyable tea-drinking experience.

Keywords: *Machine Learning, Mobile Application, Multi Method Approach, Tea, Tea Quality*

6

Utilizing data-driven techniques to guide Advanced-level course selection: personalized course recommender system for high school students.

Shaahithya K.[1*], Kumara B.T.G.S.[1], and Banujan K.[1]

[1]Department of Computing and Information Systems, Faculty of Computing, Sabaragamuwa University of Sri Lanka.
*kanaharshaahithya@gmail.com

Education is the most powerful tool that can facilitate students' achievement of their goals and objectives, strengthen their lives, and contribute to society. When considering education, higher secondary education plays a crucial role in determining an individual's future career. But nowadays, most students fail to select the most suitable course for them in the higher secondary education stage. There are many influencing factors affecting students' course selection such as student's personal interests, extracurricular activities, their educational background, their individual personality profile, various learning approaches and patterns, the student's academic performance, and so on. So, this paper aims to identify the various factors influencing advanced-level stream selection in Sri Lanka and design a personalized course recommender system for advanced-level students in Sri Lanka. The data was collected from students from the advanced-level, and the data was gathered through the online survey, which includes multiple-choice questions and open-ended questions. After data collection was completed, there are some techniques such as Data cleaning, mapping the categorical variables and handling text encoding were carried out throughout the data preprocessing stage. Multiple machine learning and deep learning techniques such as Decision tree, Artificial Neural Network and Convolutional Neural Network were used to develop the best prediction model. The ANN technique provided better results compared to the other methods with the 83 percentage of accuracy. Especially we used peer review method for better evaluation. We calculated the similarity between expert opinion and model recommendation results. So, this model provides 84 percentage of similar results in the model evaluation stage. In conclusion, the final outcome of this research contributes to increasing retention and graduation rates among students of advanced level in Sri Lanka.

Keywords: Personalized Course Recommender System, Higher Secondary Course Selection, Influencing factors for course Selection

A Machine Learning Framework Guided by Expert Insight for Team Player Classification in the IT Industry

Priyadarshani T.D.A.S.[1*], Banujan K.[2] and Kumara B.T.G.S.[3]

*[1,3]Department of Computing & Information Systems, Faculty of Computing,
Sabaragamuwa University of Sri Lanka, Sri Lanka
[2] Independent Researcher, Australia
[*]anne.sudari@gmail.com*

In the face of intense competition in the IT industry, most organizations focus only on technical skills to build teams that improve their performance. But team building goes beyond this attitude where human traits play a prominent role, and neglecting them during team building can lead to problems among team members. The main purpose of this research is to prioritize human traits and provide an approach to achieving fairer team formation without bias. Another objective is to create a comfort zone to reduce excessive stress in individuals, which is currently a significant problem in the IT industry. As the first step in this study, the traits most reflected in the team player role were selected. The specialty of this study is that those traits are extracted in two main ways, literature-based and expert-based, and the most five influential traits are identified using separate weight analysis processes. A survey was designed to assess individuals' possession of the key traits identified in the study and distributed among IT professionals, resulting in a data set of 817 rows. This data set was then converted to numerical values to assess trait levels, which were classified by experts as novice, intermediate, and advanced based on predefined criteria. Grid Search CV identified optimal parameters for Support Vector Machine, Random Forest, and Neural Network algorithms. Focusing on random forests, the framework achieved high performance with precision (0.81), ROC curves (area = 0.93), recall (0.81), and F1-score (0.81) metrics, with an 80:20 train-test ratio for model validity. Based on that framework, it can be determined on what level the team player behavior of individuals depends. The study aims to stimulate fair and unbiased team formation by effectively identifying team player levels through a machine learning approach based on long-term IT industry experience and insightful research considerations.

Keywords: *Expert-guided, IT Industry, Machine Learning, Team Player, Traits*

Software Engineering Track

GemmoFTIR for Gem Stone Identification Using FTIR Spectroscopy

Weerakutti G.D.W.C.[1], Wimalasooriya G.G.R.S.N[1], Surendra S.[1], Chathumini K.G.L.[1], Wijesekara S.S.R.M.D.H.R.[2], Wanniarachchi D.N.S.[2], Udawatte C.P.[3]

[1]*Department of Computing and Information Systems, Faculty of Computing, Sabaragamuwa University of Sri Lanka, Sri Lanka*
[2]*Department of Natural Resources, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka*
[3]*Department of Physical Sciences and Technology, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Sri Lanka*
[*]*wikum@linuxdeveloper.space*

The Gemstone identification and characterization using Fourier Transform Infrared spectroscopy (FTIR) is a widely used technique. The existing challeng lies in the limited availability of comprehensive FTIR libraries for users. This forces individuals to build their own libraries, often requiring extensive expertise and access to a vast collection of gemstone samples. Based on the demand, we have developed a user-friendly web application named GemmoFTIR to address this issue. This application uses Pearson Correlation to assess the similarity between FTIR histograms. We have mainly used Spring Boot and Firebase technologies to implement the web application. This application consists of two distinct interfaces, an admin interface and a user interface. The admin interface serves as a robust tool empowering administrators to effortlessly augment the database by uploading FTIR spectra of known gemstones. On the other hand, the user interface is tailored for end-users, allowing them to identify gemstone types by uploading corresponding FTIR spectra. The functionality overview of the system is, that users should generate a CSV file containing the FTIR spectrum of the gemstone in question. This CSV file can then be uploaded through the application interface. Then the application meticulously compares the uploaded FTIR spectrum with the database of known gemstone spectra. Upon completion of the analysis, the application generates a comprehensive report for the user. This report includes a similarity percentage, quantifying the resemblance between the unidentified gemstone and potential gem types stored in the database. This application simplifies the gemstone identification process through an intuitive interface while leveraging advanced technologies to ensure security and efficiency. This development contributes to the field by offering a user-friendly solution to the challenges associated with FTIR spectroscopy in gemstone analysis.

Keywords: *Gemstone Identification, Fourier Transform Infrared Spectroscopy (FTIR), Pearson Correlation, GemmoFTIR*

Non-Functional Software Requirements Classification Using LSTM Model

Saroth M.A.F.[1], Wijeratne P.M.A.K.[2], and Kumara B.T.G.S.[1]

[1]Department of Computing and Information Systems, Faculty of Computing, Sabaragamuwa University of Sri Lanka, Sri Lanka
[2]Department of Software Engineering, Faculty of Computing, Sabaragamuwa University of Sri Lanka
* mafsaroth@std.appsc.sab.ac.lk

Non-functional requirements (NFRs), which are important to identify early in the software development life cycle, define the quality attribute of a software application. Not only is the early identification and categorization of non-functional requirements a difficult and time-consuming procedure, but it is also essential for the assessment of architectural alternatives from the very beginning of design decisions. For this reason, a NFR classification method is needed to make it easier for software developers to classify NFR from requirement documents. This research proposed a model for classifying NFR based on product, process, and external. The Kaggle website and documentation from some Sri Lankan software companies served as the authors' two main data sources for this research. It is made up of around 2950 distinct NFR records. The authors aim to clean and structure the data by performing pre-processing techniques like stop word removal, normalizing, tokenizing, and lemmatizing, making it suitable for machine learning classification. The dataset was divided into training and testing subsets with a 4:1 split. Mainly, 80% of the data was used for training, and the remaining 20% was reserved for testing the models' performance. The authors constructed four machine learning-based algorithms named Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), and K Nearest Neighbor Classifier (KNN). Long Short-Term Memory (LSTM) model to classify the NFR. Out of the five classifiers that were used in this research, LSTM generated the most accurate findings overall which is 99.69%. A confusion matrix is used to calculate recall and precision. Measures of recall and precision enhance classification accuracy and offer a more thorough understanding of model assessment. Ultimately, the researchers revealed that the LSTM classifier is the most suitable for this study

Keywords: *Non-Functional Requirements, Machine Learning, Long Short-Term Memory, Classification*

Source Code Recommendation System to Recommend the Source Codes Based on Design Patterns using Machine Learning

Thennakoon A.G.D.S[1*], and Wijeratne P.M.A.K.[2]

[1] *Department of Computing & Information System, Faculty of Computing, Sabaragamuwa University of Sri Lanka*
[2] *Department of Software Engineering, Faculty of Computing, Sabaragamuwa University of Sri Lanka*
*agdsthennakoon@std.appsc.sab.ac.lk

As a result of the fact that design patterns offer templates that are necessary for addressing common problems and improving the communication, productivity, and maintainability of projects consistently, design patterns are of great importance in the process of developing software. Machine Learning (ML) techniques such as Support Vector Machine (SVM), Decision Trees (DT), K-Nearest Neighbors (KNN) and Long Short-Term Memory (LSTM) can be utilized to simplify the process of recognizing software design patterns in source code. The research was carried out to employ ML to produce predictions about design patterns that are present in code and provide five code suggestions base on prediction design patterns. Initially, the procedure began by collecting datasets from professionals in the field of software engineering and preparing a dataset with a size of nearly 1550 samples. As the scope of this study is very wide, we collected only data related to three design patterns and eventually classified them into three categories: Abstract Factory, Factory Method, and Singleton. We enhanced the data quality by applying preprocessing techniques such as tokenization, stopword removal, character elimination, integer encoding, and padding. We applied three different ML models and one deep learning model to make predictions. KNN, SVM, DT and LSTM models, respectively. For the goal of creating such predictions, SVM which has been demonstrated to correctly recognize design patterns with an impressive 98.67% accuracy, is the suitable algorithm for this purpose. We use the degree of cosine similarity among the input code snippets and the group code snippets to determine the next five most relevant code suggestions.

Keywords: *Machine Learning, Design Patterns, Support Vector Machine, Recommendation System, Cosine Similarity*

Data Science Track

Integrative Approach on Fake news Text Analysis: A Multimodal Framework for Fake News Detection, Categorization and User Profiling

Abishethvarman V.[1*], Kumara B.T.G.S.[1], Banujan K.[1], and Nirubikaa R.[2]

[1]*Department of Computing and Information Systems, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka.*
[2]*Department of Software Engineering, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka.*
[*]*abishethvarman@gmail.com*

Fake news is intentionally misleading or misinformation delivered as reliable news, often with the aim of manipulating the audience's opinion. The proliferation of misinformation is needed to eradicate. Due to the complexity of fake news composition and subjective sense, it is hard to identify in a single direction. This research presents a multifaceted supervised learning approach, encompassing fake news detection, categorization and user profiling. A diverse dataset about 1 million data by combining various real and fake news dataset. Preprocessing is same for both models including stop words removal, custom word removal, duplication and feature extraction mainly. Evaluation splitting is 80:20 for training and testing. Detection models, employed in different deep learning models. Recurrent Neural Network Long Short-Term Memory outperformed well by providing 99.63% accuracy score as maximum and 94.99% as least. It is better than the other algorithms due to its ability to understand the semantic features. Categorization models, leveraging models trained to classify content into distinct categories such as politics, entertainment and sports through algorithms employed in ensemble approaches. Random Forest and ADA Boost gave promising accuracy 96% and 92% respectively. The process of User Profiling began with a robust fake news detection mechanism to filter out misinformation. Subsequently, the identified fake news undergoes a categorization process to classify it into distinct categories as mentioned. Then the study delves into user profiling, where individuals' behaviour in sharing social media. It analysed the content sharing behaviour of a user regarding how much fake news that the user shared and what category shared. 77% of the users sharing fake news related to politics without their knowledge. The multimodal approach provided a comprehensive framework for tackling the complex challenges of fake news detection, categorization and user profiling in digital age.

Keywords: *Fake News Categorization, Fake News Detection, Multimodal Approach, Semantic Text Analysis, User Profiling*

Cyst Detection in Kidney Using Machine Learning Algorithm

Lakmali W.G.N.[1*], and Chathumini K.G.L.[2]

[1, 2]*Department of Computing and Information Systems, Faculty of Computing, Sabaragamuwa University, Sri Lanka*
*nirodha41@gmail.com

The kidney is one of the most vital organs in the human body, and it can suffer from a number of diseases. This research focuses solely on cystic kidney disease and other kidney diseases include Kidney Stones, Tumors, and Chronic Kidney disease. Kidney disease is also one of the primary concerns encountered by the majority of the population in Sri Lanka, and Cystic kidney disease, which is the formation of fluid-filled sacs in or around the kidneys, compromising their filtering function, is one of the kidney diseases. Kidney problems must be detected as early as possible, and cutting-edge methods and emerging technologies such as machine learning are required for this. The primary goal of this study is to use 2D images from human kidney CT scans to identify kidney cyst disease. This study uses CT scan images to develop a machine learning hybrid model for the categorization and diagnosis of cyst disease utilizing the VGG16, ResNet50, and DenseNet121 architectures of Convolutional Neural Networks. This hybrid deep learning technique offers a robust approach to medical picture categorization by demonstrating the significance of integrating many architectural paradigms to enhance model performance. The robustness of the approach lies in its ability to achieve high accuracy and fast detection, ensuring reliable performance across varied conditions. This hybrid model identified Cysts with an accuracy of 90.37% and its precision, recall, and F1-score values were 85.64%, 95.69%, and 90.39%, respectively. The hybrid model's robustness stems from its incorporation of three different CNN architectures and achieving a high accuracy rate of 90.3%.

Keywords: Cysts, CNN, VGG16, DenseNet121, ResNet50

Social Media Platform for Deaf and Dumb People through Sign Language Recognition Using Media Pipe and Machine Learning

Randika J.I.[1*], Randeni R.M.S.D.[1], Deshan G.A.A.[1], Perera W.P.K.I[1],
Wasalthilaka W.V.S.K.[2], and Samarajeewa C.M.A.[3]

[1*]*Department of Computing and Information Systems, Faculty of Computing,
Sabaragamuwa University, Sri Lanka*
[2]*Department of Software Engineering, Faculty of Computing, Sabaragamuwa
University of Sri Lanka*
[3]*Wiley Global Technology (PVT) LTD, Sri Lanka*
*jishanrandika@gmail.com

People in the world who are deaf/dump their entertainment is limited because they are embarrassed when expressing their ideas with normal people. Since they typically utilize sign language to connect with regular people, regular people ought to be proficient in sign language as well. Nevertheless, it is quite challenging to locate individuals in the community who are proficient in sign language. Even Though many research studies were carried out to address this issue, there is still a large room for improvement. Many of those approaches cannot afford poor people, and it is hard to use wearable devices with the disturbance to normal communication. Therefore, as a solution, this study has utilized sign language recognition to introduce a social media platform for deaf/dumb people to enhance their global communication. This will help them to mingle with different cultures in the world and make social bonds among different countries while avoiding the social barriers for deaf/dumb people. This approach used Google's "MediaPipe" Library to identify human motions and Machine Learning (ML) to train the model. To collect the data used a 720p Logitech camera. The data was collected from 5 persons with 50 actions for each expression in US English. These data are stored as numpy arrays with 30 frames for each action to train ML model. In conclusion, the proposed approach has produced 96% training accuracy and 89% testing accuracy. In order to produce the subtitles in accordance with the trained ML model, it was integrated with the social media app's front end and hosted on Google App Engine. If this API can be used in every social media, it will help to reduce the social distance between deaf/dumb people and normal people by sharing the same rights with all human beings.

Keywords: *Sign Language Recognition, Deaf, Dumb, Mediapipe, Machine Learning*

Deep Unsupervised Learning for Motion Blurred Image Restoration

Munasinghe A.P.[1*], and Chathumini K.G.L.[2]

[1,2]*Department of Computing and Information System, Faculty of Computing,*
University of Sabaragamuwa, Belihuloya, Sri Lanka
*amilaprasanga97@gmail.com

The quality of the image is very important in image processing. Image blurring is a most common issue, which caused to reduce the quality of the image. Blurred images often occur due to camera shake, object motion, or long exposure times, leading to a loss of critical details and visual fidelity. Deblurring of an image is a challenging task, and there are very few researches available in this field. This proposed study is focusing on introducing a novel approach to detect and deblur motion-blurred images using autoencoder mechanism. This proposed study will be a remarkable era in image quality restoration in various real-world applications, such as surveillance, medical imaging, autonomous vehicles, and photography. The inability to recover sharp images from blurred ones can significantly impact the accuracy and reliability of image-based systems in these sectors. The primary objective of this research is to develop an advanced and practical solution to restore motion-blurred images with high accuracy and speed, enabling improved decision-making and analysis in critical domains. The motion deblurring model makes use of an autoencoder architecture, which includes an encoder and a decoder. The encoder reduces spatial dimensions gradually, forming a condensed representation in latent space. To generate deblurred images, the decoder, its mirror counterpart, upsamples the latent space using transposed convolutional layers. Over 100 epochs, the end-to-end autoencoder learns the difficult mapping between motion-blurred and clear images, attaining over 80% accuracy.

Keywords: *Image Processing, Blurred Images, Image Deblurring, Deep Learning*

Evaluating the Influence of Auto-Correct Technology on Spelling Proficiency among Computer Science Undergraduates

Archchaya U.[1*], Banujan K.[2], Kumara B.T.G.S.[1], and Prasanth S.[3]

[1]*Department of Computing and Information Systems, Faculty of Computing, Sabaragamuwa University of Sri Lanka, Sri Lanka*
[2]*Independent Researcher, Australia*
[3]*Department of Physical Sciences and Technology, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Sri Lanka*
*uthyakumararchchaya@gmail.com

The impact of auto-correction on technology is dual-faceted, manifesting in both positive and negative effects, particularly evident in students' academic writing proficiency. While it can enhance academic writing abilities and potentially reduce workload, it may also lead to a diminished capacity to recognize errors, a loss of focus, and decreased contextual comprehension. This study assesses the English spelling proficiency of computer science undergraduates, focusing on homophones, phonetic errors, and difficult-to-spell words. Initially, Notepad, lacking auto-correction by default, was employed for data collection and organization. Presently, natural language processing (NLP) techniques are predominantly utilized for text analysis tasks, necessitating comprehensive pre-processing for accurate analysis. Specific techniques such as tokenization and lowercasing were applied for this investigation. CountVectorizer, a Python Sci-kit-learn module, was utilized to create word frequency-based vectors, serving as novel input for machine learning algorithms like Support Vector Machine (SVM), Naive Bayes (NB), and Random Forest (RF) to determine the most accurate classification algorithm. This study categorizes various types of errors and provides a fresh perspective on the impact of auto-correct technology on spelling proficiency. Among the tested models, SVM demonstrates superior performance, achieving 81% accuracy in identifying and assessing errors. Furthermore, this study explores the potential correlation between spelling proficiency assessment and machine learning techniques. By elucidating the effects of auto-correct technologies on language skills in Computer Science, this research holds significant implications for teaching methodologies.

Keywords: *Auto-Correction, Spelling Proficiency, CountVectorizer, Spelling Error, Machine Learning*

Incremental Learning for Tomato Leaf Disease Detection

Roopasinghe R.U.N.[1*], Vasanthapriyan S.[2], and Chathumini K.G.L.[3]

[1, 2, 3]*Department of Computing & Information Systems, Faculty of Computing, Sabaragamuwa University of Sri Lanka*
*umeshanirmani@gmail.com

Early recognition and classification of tomato leaf diseases is vital for high-yield production. With the advancement of technology, automatic disease detection has been integrated, and Convolutional Neural Networks (CNNs) have emerged as a widely spread approach today. However, these architectures use batch learning which they train on fixed datasets, and lack the ability to get updated in a changing environment due to catastrophic forgetting. To address this limitation, our study presents a comparison of three popular CNN architectures, MobileNetV2, EfficientNetV2B0, and ResNet50, by retraining models with previously unseen data using incremental learning. The images have been taken from the PlantVillage dataset, which belongs to eight categories along with seven disease classes and one healthy class. According to the results, EfficientNetV2B0 outperforms the other architectures with a validation accuracy of 99.72% with the first unseen disease set. MobileNetV2 and ResNet50 have validation accuracies of 94.98% and 99.00%, respectively. In subsequent iterations, involving three new disease classes, MobileNetV2, EfficientNetV2B0, and ResNet50 have validation accuracy of 97.68%, 99.19%, and 98.09% at the final iteration, respectively. But considering the model size and execution time, MobileNetV2 stands out among the other two architectures. In this research, we evaluated our model performance using metrics like accuracy, precision, recall, F-Score, and also average accuracy and forgetting measure in the context of incremental learning. Our work highlights the effectiveness of CNNs in dynamic agricultural settings and underscores the critical role of incremental learning in enhancing plant disease detection capabilities for future implementations.

Keywords: *Comparative Analysis, Convolutional Neural Networks, Incremental Learning, Tomato Leaf Diseases*

Predicting Preterm Birth: Machine Learning Approaches Using Maternal Factors

Gobitha S.[1*], Banujan K.[2], Kumara B.T.G.S.[1], and Prasanth S.[3]

[1]*Department of Computing and Information Systems, Faculty of Computing, Sabaragamuwa University of Sri Lanka, Sri Lanka*
[2]*Independent Researcher, Australia*
[3]*Department of Physical Sciences and Technology, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Sri Lanka*
*gobithaselvaranjan@gmail.com

Preterm birth (PTB), defined as delivery occurring prior to the completion of 37 weeks of gestation, represents a multifaceted medical condition influenced by various factors. PTB significantly contributes to neonatal mortality rates. However, advancements in specialized care have substantially improved survival rates for preterm infants. Additionally, modern medical technology has enabled the conversion of potential PTB cases into normal births. This study employs Machine Learning (ML) techniques to develop a PTB risk assessment model. Conducted at the Teaching Hospital in Jaffna, Sri Lanka, the study involved 165 mothers who delivered both term and preterm babies. Data pre-processing was initially conducted to handle incomplete data, address missing values, and encode categorical variables numerically. Furthermore, feature extraction was performed to identify the most informative factors contributing to PTB prediction. Out of 33 attributes, 19 were extracted using Pearson correlation analysis, and these attributes were then fed into Extreme Gradient Boosting (XGBoost) to assess classification accuracy. Considering the traditional and prevailing algorithms in most classification problems, XGBoost provided more precise results; that's the main reason behind this selection. The results demonstrate that XGBoost achieved a reliable accuracy rate of 81% in predicting PTB. This proposed solution holds promise for assisting Gynecology and Obstetrics departments, especially in regions with growing populations and limited resources, by enabling efficient prediction of PTB within short timeframes. Addressing current challenges in predicting PTB, the development of an accurate risk assessment model using ML techniques has the potential to significantly improve prenatal care and ultimately reduce the incidence of PTBs, thereby enhancing maternal and neonatal health outcomes.

Keywords: *Preterm Birth, Neonatal Deaths, Gestation, Gynaecology and Obstetrics, XGBoost*

BART Summarization for Closed Captions to Generate YouTube Video Titles

Senevirathne H.U.[1*], Kumara B.T.G.S.[2], and Banujan K.[3]

[1,2]*Sabaragamuwa University of Sri Lanka*
[3]*Individual Researcher, Australia*
[*]*senevirathnehu@email.com*

YouTube is a leading platform for information dissemination and entertainment. With billions of hours of video content uploaded and viewed daily, effective content organisation and discovery have become critical. BART is a transformer-based model with unique characteristics ideal for summarising textual data. By incorporating the Closed Captions (CC) of videos, this study examines BART's capabilities in summarising textual content and using the popular natural language processing library Hugging Face to generate relevant and engaging titles for YouTube videos. The methodology consists of multi-steps aiming to maximise the potential of BART summarisation. Initially, a diverse dataset of YouTube CC was gathered, encompassing a variety of genres and content types. This dataset served as the basis for training the BART model. Hugging Face's transformers library was then used to fine-tune the pre-trained BART model on a specific CC dataset, allowing for better adaptation to different YouTube contents. The fine-tuning process entailed adjusting the model's parameters to match the properties of YouTube CC. It ensured that the model accurately captures the essence of the spoken content while maintaining context. The trained and fine-tuned BART model successfully demonstrates its ability to grab the essence of the video into concise and engaging headlines. For example, "First Class Travel to Bali, Indonesia" is the generated title for the video with the title "Dining On A $30,000 Plane Ticket" and "Everything I Know About Cooking In Under a Minute" is the generated title for the video with the title "Learn To Cook In Less Than 1 Hour". The study has implications for content creators, viewers, and the YouTube platform as a whole, as it can improve content discoverability, increase user engagement, and simplify the content creation process.

Keywords: Title Generation, BART, Text Summarization

Identification of paddy leaf disease: An end-to-end training approach

Hewarathna A.I.[1*], Vigneshwaran P.[2], and Charles J.[3]

[1]Department of Computing and Informaion Systems, Faculty of Computing, Sabaragamuwa University of Sri Lanka, Sri Lanka
[2]Department of Software Engineering, Faculty of Computing, Sabaragamuwa University of Sri Lanka, Sri Lanka
[3]Department of Computer Science, Faculty of Engineering, Friedrich-Alexander-University Erlangen-Nuremberg, Germany
*asheniranga@gmail.com

Leaf infections in rice crops are a major threat to agricultural production, and prompt identification is critical for effective treatments. The purpose of this study is to investigate a more effective method for predicting illnesses in paddy leaves utilizing 3-channel 10407 pictures. In this study, we compare the effectiveness of single-task end-to-end ResNet-50 training to a multi-task EfficientNet model in terms of classification accuracy. To test the suggested approach, a dataset of 10407, 3-channel pictures of paddy leaves was used. This dataset consists of images of 10 classes including healthy class. Other classes were Bacterial leaf blight, bacterial leaf streak, bacterial panicle blight, blast, brown spot, dead heart, downy mildew, hispa, tungro. The ResNet-50 model was trained using a single-task end-to-end approach, with an emphasis on categorizing diverse leaf diseases. Concurrently, a multi-task EfficientNet model was used for comparison. Our findings show that the ResNet-50 model, trained using single-task end-to-end training, has higher generalization capabilities, beating the multi-task EfficientNet model in terms of classification accuracy. In the single-task end-to-end training approach with 80:20 train-test data ratio, it scores 0.9891 train accuracy and 0.9804 test accuracy which was able to outperform the multi-tasking fine-tuning approach which scores 0.9735 train accuracy and 0.9707 test accuracy, according to these results single-task end-to-end training approach was the best model. This study helps to advance our understanding of successful deep-learning techniques for predicting paddy leaf disease and emphasizes the need to select appropriate model architectures. Further investigation and validation of these findings are required to improve the proposed approach's robustness and applicability across a variety of paddy cropping scenarios.

Keywords: *Leaf Disease Detection, End-to-end Training, Classification, Deep Learning, Ensemble Learning*

Open Track

IoT-Driven Synergy: Enhancing Passenger Safety through Smart Vehicles Integrated with Centralized Home Systems in Smart Towns

Suja M.F.M.[1*], and Dangalla R.L.[1]

[1]*Department of Computing and Information Systems, Faculty of Computing, Sabaragamuwa University of Sri Lanka.*
[*]*mohamedsuja7n@gmail.com*

The integration of Internet of Things (IoT) technologies has become critical in reshaping transportation and smart living in the evolving modern world. This research introduces a novel approach to improving passenger safety by seamlessly merging IoT-enabled Smart Vehicles with a centralized home system in the context of Smart Towns. This proposes a comprehensive ecosystem that facilitates the cutting edge technologies such as real-time tracking, intelligent communication, and user-centric safety measures. It lies in the strategic deployment of WiFi modules and IoT devices comparing to the existing smart home models. Smart Vehicles are equipped with cutting-edge IoT technologies, including GPS modules, sensors, and communication devices, ensuring a continuous flow of real-time data. WiFi modules enable secure and reliable connectivity, fostering communication between Smart Vehicles and a centralized home system. The centralized home system serves as the nerve center, facilitating the seamless integration of Smart Vehicles and the smart home environment. Through an advanced backend infrastructure, the system processes and interprets data from IoT devices, facilitating timely responses to potential safety concerns. Passengers benefit from enhanced safety features, such as live tracking, emergency response mechanisms, and predictive maintenance alerts through this approach. This interconnectedness enhances user experience and augments overall safety by allowing occupants to focus on the road while maintaining control over their home environment. The system undergoes testing to validate its performance, scalability, and security, followed by deployment in Smart Towns manually. This gives a solution to make Smart Towns by presenting an innovative and holistic solution that transcends conventional boundaries, creating a safer and more efficient urban ecosystem. The seamless integration of IoT-enabled Smart Vehicles with centralized home systems not only advances passenger safety standards but also sets a precedent for the future of smart and connected urban living.

Keywords: *Centralized Home System, IoT-enabled Smart Vehicles, Passenger Safety, Smart Towns, Urban Ecosystem*

Sri Lankan Sign Language Detection Approach for Deaf People Using Human Pose Estimation Technique

Nadesha S.J.M.J.[1*], Wasalthilake W.V.S. K.[2]

[1*]*Department of Computing and Information Systems, Faculty of Computing, Sabaragamuwa University, Sri Lanka*
[2]*Department of Software Engineering, Faculty of Computing, Sabaragamuwa University of Sri Lanka*

Individuals with disabilities experience health problems related to their body, mind, senses, or behavior that restrict their ability to engage in particular activities or engage with their environment. One of the most prevalent forms of disability in society is deafness. One of the most common ways deaf people communicate with one another is through sign language, as verbal communication is not possible for them. However, a third party with sign language proficiency ought to manage the interactions between the deaf and others. Many new approaches are being developed to deal with this, but they aren't always appropriate. Sign language is unique from country by country, and Sri Lankan sign language uses hands, facial expressions, and body movements, and solutions should be appropriate. Many studies have been done to use contemporary technologies to solve this problem, but they have been done using only hand gestures. The specialty here is to use the pose estimation technique for the Sri Lankan sign language to check the movements of the entire upper body. This study suggested a Sinhala language sign language detector that estimates human pose using the Media Pipe library and the four algorithms (Logistic Regression, Ridge Classifier, Random Forest, Gradient Boosting). For this purpose, a database of 400 videos was created, as there are 100 videos of human postures for one signal using 40 commonly used signals. The suggested method extracts full-body stance landmarks from videos using pose estimation technique, which are then fed into a machine learning model that has already been trained to classify body language. This study yielded an accuracy of 86%, which is among the highest levels of accuracy recorded in Sri Lanka. The results of this study will allow deaf people to better communicate themselves with society, and that was the purpose of this grant.

Keywords: Deaf People, Human Pose Estimation, Logistic Regression, Sri Lankan Sign Language

Movie Recommendation System using NLP: Based on Critiques and Ratings

Mahesha K.[1*], Kumara B.T.G.S.[2] and Banujan K.[3]

[1]*Department of Computing and Information Systems, Faculty of Computing, Sabaragamuwa University of Sri Lanka, Sri Lanka*
[2]*Department of Computing and Information Systems, Faculty of Computing, Sabaragamuwa University of Sri Lanka, Sri Lanka*
[3]*Independent Researcher, Australia*
*kandambi.mahesha@gmail.com

Entertainment is crucial for everyone to refresh their energy and mood in this hustling world. Among these, films play a vital role in the entertainment. But the problem is the number of movies and productions from around the world has increased yearly. Individuals may become confused about which films to watch first while searching for the finest and greatest movies. The solution is the movie recommendation systems that assist us to find our favorite movies among all of these numerous genres of movies and hence decrease the bother of spending a lot of time finding our favorite movies. This research proposed an innovative approach to improve content-based movie recommendations using Natural Language Processing (NLP) techniques such as Latent Dirichlet Allocation (LDA), Term Frequency-Inverse Document Frequency (TF-IDF), and cosine similarity based on critiques and ratings. This study takes advantage of the critiques to eliminate the shortcomings of previous studies. Professional film critics provide significant critiques in movie recommendations since they give professional insights and contextual evaluation rather than audience reviews. This study applies TF-IDF to convert topics into numerical vectors, ensuring meaningful topic representation. Cosine similarity is then calculated between movies based on these vectors, measuring their directional similarity. LDA extracts relevant topics such as comedy, romance, drama, etc. from movie critiques in the Rotten Tomatoes dataset, aiding in movie categorization. Then interprets the topic labels for each topic based on the weight. Ratings are converted into a 0-5 scale and finally, movies are recommended from highest to lowest ratings within specified LDA topics. The model's 76.99% accuracy was calculated by comparing the list of recommended movies for each topic with the IMDB dataset. Finally, these findings conjecture that this approach is well suited for searching for favorite movies among a large number of movies.

Keywords: *Cosine Similarity, LDA, Movie Recommendations, TF-IDF*

AUTHOR INDEX

Best wishes for the ComURS2024 Computing Undergraduate Research Symposium 2024