

PROGRAMME & ABSTRACTS BOOK

The 6th International Conference on Soft Computing in Data Science (SCDS 2021)

Science in Analytics: Harnessing Data and Simplifying Solutions

Copyright © 2021 by Institute for Big Data Analytics and Artificial Intelligence (IBDAAI) Kompleks Al-Khawarizmi, & Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia.

All rights reserved. Personal use of this material is permitted. However, no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or any means, electronic, mechanical, photocopying, recording or otherwise, without prior permission, in writing, from the publisher

Table of Contents

Table of Contents	ii
About UiTM	2
Centre of Excellence Background	3
Message from the Vice Chancellor	4
Message from Rector, ITS	6
Message from the Director	8
Message from the Conference Chairs	10
Introduction to SCDS 2021	12
Keynote Speakers	13
Keynote Speaker 1: YBhg Dato' Sri Mohd Dr. Uzir Mahidin	14
Keynote Speaker 2: Professor Dr. Guan-Hua Huang	17
Keynote Speaker 3: Professor Dr.rer.pol. Heri Kuswanto	19
Keynote Speaker 4: Assistant Professor Dr. Ubydul Haque	21
Keynote Speaker 5: Associate Professor Dr. Simon Fong	24
Keynote Speaker 6: Professor Dr. Richard Millham	26
Keynote Speaker 7: Associate Professor Dr. Aiden Doherty	27
Conference Schedule SCDS 2021-The 6th International Conference on Soft Com Data Science	puting in 29
Paper Presentations Schedule	33
Proceeding Abstracts	41
Track 1: Artificial Intelligence Techniques and Applications	41
Track 2: Data Analytics and Technologies	51
Track 3: Data Mining and Image Processing	61
Track 4: Machine and Statistical Learning	65
Conference Organization	79
Reviewers	82
Acknowledgements	85





As the largest comprehensive university in Malaysia, Universiti Teknologi MARA (UiTM) has expanded its visibility through collaborations with universities worldwide. It continues to provide access to higher education, playing its role in nation building by unleashing potentials, shaping the future. Providing innovative education within reach at its 35 campuses and 26 faculties across Malaysia, UiTM offers over 500 academic programmes at Foundation, PreDiploma, Diploma, Bachelor's, Master's, and PhD level, as well as Professional Programmes. Today, UiTM is a proven premier higher education institution of choice. Amongst its long lines of accolades, in 2021 UiTM ranked 108th in the QS Asia University Rankings, placed 651-700 in the QS World University Rankings, and was in the 101-200th position in THE Impact Rankings in 2021. In 2020, the Hospitality and Leisure Management subject was at no.31 in the world, and 12 of UiTM subjects were ranked in the QS World University Rankings by Subject. UiTM also won gold for the Best International Print-ads at the QS APPLE Creative Awards 2020. With over 900,000 alumni in science, technology, humanities and entrepreneurship, UiTM offers opportunities to shape leaders at national, industry and global levels and is well poised to become a globally renowned university by 2025.

Centre of Excellence Background

Institute for Big Data Analytics and Artificial Intelligence (IBDAAI) is a Centre of Excellence dedicated for developing smart solutions to data analytics and social analytics that operate in complex environments. The centre aims to build a collaborative research team, which can contribute to high quality research outcomes and impacts. This centre of excellence will also provide expert advice through training, consultancy and commercialization related to Big Data Analytics and Artificial Intelligence. In line with the latest developments and future needs, new advances in artificial intelligence technology in empowering Big Data Analytics is very important. The combination of Big Data Analytics and Artificial Intelligence provides an optimal solution to many problems. The establishment of this centre of excellence is also in line with the direction of the Ministry of Higher Education (MoHE), Malaysia to leverage the university's expertise that will be useful and benefit to the community and industry. The vision is to become a world-renowned and internationally recognized research centre in the field of Big Data analytics and Artificial intelligence.

IBDAAI was a research centre known as Advanced Analytics & Engineering Centre (AAEC), established on 1st of June 2014 and located at the Faculty of Computer & Mathematical Sciences (FSKM), AI Khawarizmi Complex, UiTM Shah Alam, Selangor, Malaysia. Since then AAEC and FSKM have been working together to establish Big Data Lab@UiTM which were supported by Grants acquired from the Ministry of Higher Education (MOHE) and the Ministry of Trade and Industry (MITI). Total value of both grants are about RM750,000. Big Data Lab@UiTM has been used for training of students taking Data Science tracks, Postgraduate students and research projects on Data Analytics.

Since 2015, AAEC has organized the International Conference on Soft Computing in Data Science (SCDS). This annual event was co- organized with other international collaborators such as University of Tennessee, Universitas Gadjah Mada, Chulalongkorn University, University of Macau, Institut Teknologi Sepuluh Nopember, Kyushu Institute of Technology, Durban University of Technology, Liverpool John Moores University, and Charles Sturt University. This year we are honored to have Oxford University as our knowledge partner. The conference also received continuous support from Department of Statistics Malaysia, Malaysia Digital Economy Corporation (MDEC) and Quandatics (M) Sdn Bhd.

The Institute for Big Data Analytics and Artificial Intelligence aims to continue this initiative of organizing the conference with the aim of bringing together experts and researchers to share their innovative and practical applications of Big Data Analytics and Artificial Intelligence.

Message from the Vice Chancellor



It is my great pleasure to welcome all distinguished guests, esteemed keynote speakers, paper presenters and participants to the 6th International Conference on Soft Computing in Data Science 2021 (SCDS 2021). This year's annual conference is jointly organised by Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Universiti Teknologi MARA and Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia with the support of Faculty of Computer and Mathematical Sciences and ReNeU, Universiti Teknologi MARA. I am pleased to note the success of this year's event is thanks to the collaboration with several reputable universities in the region as well as industries.

Events such as SCDS 2021 are crucial in promoting impactful partnerships between the academic and industry world. We have proved that by working together we are able to withstand the challenges with the pandemic that have gripped the world over the past 18 months. This is vital to encourage a more inclusive dialogue on a global scale as the roles of universities go beyond teaching and the emphasis is now on conducting research and other impactful activities to influence society and economy. Together, each of us can do our part to step forward and contribute time and resources to provide technological advancements especially in the area of Big Data Analytics (BDA) and Artificial Intelligence (AI). These advancements have expanded opportunities in various ways such as niche marketing strategies, emergence of new career paths and industrial scale highly efficient process-flow. By harnessing data through the use of sophisticated technologies, universities can create, enhance and mobilise the transformative power of higher education through inventive teaching and learning experience, and produce graduates who can innovate and design solutions for real world problems.

With the availability of infinite data, invaluable knowledge and hidden gems are accessible. It is essential researchers representing different disciplines explore opportunities and future of BDA and AI as per the theme "Science in Analytics: Harnessing Data and Simplifying Solutions". Keynote speakers, paper presenters and

participants will share their insightful thoughts with scholars and industry experts on the way forward with Data Science. I am optimistic that this two-day event will provide opportunities for more multi-disciplinary collaborations to intensify impactful research.

On that note, I would like to congratulate the Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), UiTM, Faculty of Computer and Mathematical Sciences, ReNeU, UiTM as well as our international conference partner, Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia for jointly organising this virtual conference, and for all the assistance rendered in making this event a success. I would also like to express my appreciation to University of Tennessee, Oxford University, Universitas Gadjah Mada, Kyutech Institute of Technology, Chulalongkorn University, Liverpool John Moores University, Durban University of Technology, University of Macau ,Charles Sturt University, Department of Statistics Malaysia, Quandatics (M) Sdn Bhd, Malaysia Digital Economy Corporation (MDEC) and Women in AI (WAI) for being on board with UITM in this scholastic journey. To everyone, I wish you a rewarding and fulfilling experience at SCDS 2021. May you move forward with strategic collaborations and bring forth innovative ideas to benefit the global community.

Thank you.

PROFESSOR TS. DR HAJAH ROZIAH MOHD JANOR VICE-CHANCELLOR UNIVERSITI TEKNOLOGI MARA

Message from Rector, ITS



Assalamualaikum Warahmatullahi Wabarakatuh and warm greetings to everyone.

Praise to Allah the Almighty for all His blessing, grace and mercies that have made us possible to attend this prestigious event. On behalf of the Institut Teknologi Sepuluh Nopember (ITS), I am greatly honored and pleased to welcome you all to the 6th International Conference on Soft Computing in Data Science (SCDS2021) with the theme "Science in Analytics: Harnessing Data and Simplifying Solutions", and I would like to thank the organizing committee who have organized this event very well in a pandemic situation. This SCDS2021 is organized by IBDAAI (Institute for Big Data Analytics and Artificial Intelligence), UiTM, with the collaborations of Faculty of Computer Science and Mathematics, UiTM, Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia, and Research Nexus, UiTM from 2-3 November 2021.

I would also like to thank the seven keynote speakers: YBHG. Dato' Sri Dr. Mohd Uzir Mahidin (Chief Statistician, Malaysia); Prof. Dr. Guan-Hua Huang from Institute of Statistics National Yang Ming Chiao Tung University (NYCU), Taiwan; Prof Dr.rer.pol Heri Kuswanto from Department of Statistics ITS; Assoc. Prof. Dr. Aiden Doherty from Oxford Big Data Institute, Oxford University, UK; Assoc. Prof. Dr. Simon Fong from University of Macau, China; Prof. Dr. Richard C. Millham from Durban University of Technology, South Africa; and Assist. Prof. Dr. Ubydul Haque from University of North Texas Health Science Center, USA. The topic of these seven keynote speeches are absolutely important and timely.

The issue of big data and digital transformation is also a very interesting issue nowadays. With the Covid-19 Pandemic, we all now witness the acceleration in digital transformation in any organization.

As a large and leading university, ITS is at the forefront to foster innovation, and has done a lot with respect to digital transformation. Our innovative ideas has been transformed into useful products.

At the national level, we also need a triangle of actors to support innovation: the government, academia and business practitioners. The government provides infrastructure, defines strategic development goals and carries out research and innovation policies. Academia conducts a wide range of research and provides

training for innovation specialists, scientists and entrepreneurs. Business practitioners serve as the main producer of innovation capable of commercializing initial research results and connecting government programs with academic research.

I believe that this international conference is an excellent medium for exchanging ideas, knowledge, information, insights, and working in a collaboration with the aim of improving academic research quality and increasing contributions to the society. I am hoping that this event later will be able to generate valuable insights and implications for innovation project ideas.

I wish a very nice experience for all participants.

Prof. Dr. Ir. Mochamad Ashari, M.Eng. Rector Institut Teknologi Sepuluh Nopember (ITS)

Message from the Director



Assalamu'alaikum Warahmatullahi Wabarakatuh and warm greetings to everyone.

First and foremost, I wish a warm welcome to all keynote speakers, and delegates to The 6th International Conference on Soft Computing in Data Science (SCDS 2021) themed **"Science in Analytics: Harnessing Data and Simplifying Solutions**". SCDS 2021 focuses on the importance of Big Data Analytics and Artificial Intelligence in this data driven economy era. This year we are grateful to have Institut Teknologi Sepuluh Nopember (ITS) co-hosting this virtual conference. This conference has attracted speakers and participants from various countries, including the United Kingdom, Taiwan, USA, India, Indonesia, China, South Africa and Malaysia.

The aim of this conference is to provide a far-reaching platform for knowledge sharing and technical skills building on Big Data Analytics and Artificial Intelligence. These knowledge and skill sets are very much needed in disciplines such as bioinformatics, medical and health informatics, social sciences, manufacturing, economics, business and finance.

I would like to express my utmost appreciation to Professor Dr Yap Bee Wah and Dr Jerry Dwi Purnomo who have shown great commitment in organizing SCDS 2021. We are extremely grateful for your great efforts in preparing the venue for the conference and ensuring the success of SCDS 2021. I hope that this joint effort in co-hosting an international conference will strengthen the good relations between the two universities.

I am also very happy with the continuous support from Professor Dr Michael W. Berry, University of Tennessee, USA. I also would like to thank all our Keynote Speakers for taking time off their busy schedule to come and share with us their knowledge and experience which I am sure will benefit the conference participants.

Through this conference we have increased our international strategic partners. I would like to thank Institut Teknologi Sepuluh Nopember, Indonesia; Oxford University, UK: University of Tennessee, USA; Kyushu Institute of Technology, lizuka, Japan; Universitas Gadjah Mada, Indonesia; Chulalongkorn University, Thailand; Liverpool John Moores University, UK; Data Analytics and Collaborative Computing Group, University of Macau, China; Durban University of Technology, South Africa,

and Charles Sturt University, Australia for their support and being great academic partners in the diverse forms of scholastic endeavours. I also thank Institute for Big Data Analytics and Artificial Intelligence (IBDAAI was known as Advanced Analytics Engineering Centre (AAEC) of FSKM) for organizing this conference annually and establishing international collaborations.

My heartfelt gratitude goes to MDEC (Malaysia Digital Economy Corporation) for their support and sponsorship since the first SCDS conference in 2015. My sincere appreciation also goes to Quandatics (M) Sdn Bhd and Department of Statistics Malaysia for the conference sponsorship and materials. Last but not least, I would like to thank the conference joint-organizing committees from Universiti Teknologi MARA, Malaysia and Institut Teknologi Sepuluh Nopember, Indonesia who have shown great commitment and fantastic team efforts in ensuring that SCDS 2021 will be an unforgettable event for all participants.

I wish all of you a delightful and fruitful time in SCDS 2021 and I look forward to your participation.

Professor Dr Jasni Mohamad Zain Director Institute for Big Data Analytics and Artificial Intelligence Universiti Teknologi MARA

Message from the Conference Chairs



On behalf of Universiti Teknologi MARA (UiTM) and Institut Teknologi Sepuluh Nopember (ITS), we would like to extend a warm welcome to all our honourable guests, keynote speakers and participants of the 6th International Conference on Soft Computing in Data Science 2021 (SCDS 2021) with the theme "Science in Analytics: Harnessing Data and Simplifying Solutions". The success of the SCDS 2015, SCDS 2016, SCDS 2017, SCDS 2018, 2019 motivated the organizing of SCDS 2021. The SCDS 2021 here is a virtual conference due to the Covid-19 pandemic outbreak around the world. Knowledge sharing through virtual conferences, webinars and workshops continues to ensure teaching and learning, and research is delivered amidst the challenges faced due to Covid-19 pandemic.

SCDS 2021 continues to provide a platform for sharing on leading edge analytical methods, and also addressing challenges, problems and issues in Big Data Analytics and Artificial Intelligence. We highly appreciate the great support from Department of Statistics Malaysia, Quandatics (M) Sdn Bhd, Malaysia Digital Economy Corporation(MDEC), and Women in AI(WAI). We also thank all Honorary Chairs and International Scientific Committee for your support and commitment in working with UiTM and ITS to be in the frontiers of advancing knowledge in Big Data Analytics and AI. We hope to have strategic partnership with universities and industries to strengthen our BDA initiatives and actualize academic and research collaborations for the advancement in research to achieve the United Nation SDG (Sustainable Development Goals).

We are proud and happy to have seven distinguished experts as our keynote speakers: YBhg. Dato' Sri Dr. Mohd Uzir Mahidin (Chief Statistician, Malaysia); Prof. Dr. Guan-Hua Huang (Institute of Statistics National Yang Ming Chiao Tung University (NYCU), Taiwan); Prof Dr.rer.pol Heri Kuswanto (Department of Statistics ITS, Indonesia); Assoc. Prof. Dr. Aiden Doherty (Oxford Big Data Institute, Oxford University, UK); Assoc. Prof. Dr. Simon Fong (University of Macau, China); Prof. Dr. Richard C. Millham (Durban University of Technology, South Africa); and Assist. Prof. Dr. Ubydul Haque (University of North Texas Health Science Center, USA). We hope that the keynote sessions will add value to your knowledge and career and may you establish further collaborations with our Keynote Speakers.

We are delighted that this year, UiTM-ITS joint collaboration has increased the paper submissions from a diverse group of researchers. We received 79 paper submissions, among which 31 were accepted. We would like to thank Professor Dr Michael W. Berry, Professor Dr Azlinah Mohamed and Professor Dr Jasni Mohamad Zain for their contributions as editors of SCDS 2021 proceedings which will be published by Springer in the Communications in Computer and Information Science series. We gratefully acknowledge the wonderful support provided by all the technical reviewers who generously sacrifice their time for reviewing the papers.

We would also like to thank Dr Azlan Ismail, Associate Professor Dr Sayang Mohd Deni, Associate Professor Dr Dedy Dwi Prastyo, and Dr Achmad Choiruddin for the successful pre-conference online workshops: Data engineering and supervised learning using Python, Data Management and analytics using R, Time series and machine learning using R, and Spatial Statistics using R.

We greatly appreciate the dedicated support of our SCDS 2021 committees from both universities who have worked tirelessly to ensure a successful conference. We highly appreciate all those who have contributed directly or indirectly to the success of SCDS 2021.

Last but not least, we wish everyone an enjoyable and memorable experience connecting online through SCDS 2021 virtual conference.

Professor Dr Yap Bee Wah & Jerry Dwi Trijoyo Purnomo, Ph.D Conference Chairs

Introduction to SCDS 2021

The First and Second International Conference (SCDS2015, SCDS2016) were successfully held in Kuala Lumpur Malaysia. Subsequently, SCDS2017 was successfully hosted by Universitas Gadjah Mada, Indonesia. Chulalongkorn University successfully hosted SCDS2018 from 15-16 August 2018. In 2019, Kyushu Institute of Technology (KYUTECH), lizuka, Japan successfully hosted SCDS2019 from 28-29 August 2019.

The 6th International Conference on Soft Computing in Data Science 2021 (SCDS 2021) is organized by IBDAAI (Institute for Big Data Analytics and Artificial Intelligence), UiTM and Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia with the support of Faculty of Computer and Mathematical Sciences, UiTM and Research Nexus, UiTM from 2-3 November 2021. The theme of the conference is 'Science in Analytics: Harnessing Data and Simplifying Solutions'. SCDS2021 have invited renowned international and local keynote speakers who are academia or practitioners to share their knowledge and experience in the area of soft computing in various disciplines. This conference aims to provide a platform for researchers and practitioners to share their research work and to create rigorous international research collaborations. This year, SCDS 2021 virtual conference will be hosted by Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia

SCDS2021 targets participants from universities, government agencies and industries with the ultimate aim of bridging the gap between academia and the industry. Research collaborations between academia and industry can lead to the advancement of useful analytics and computing applications for providing real time insights and solutions.

All accepted and registered papers of SCDS2021 will be published in the renowned series of "Communications in Computer and Information Science series" by Springer Verlag. Authors of accepted papers will be invited to submit extended paper to Scopus indexed journals.

Keynote Speakers

Keynote Speaker 1: YBhg Dato' Sri Mohd Dr. Uzir Mahidin

Chief Statistician Malaysia & Adjunct Professor of Universiti Teknologi MARA, Malaysia



Analytics and AI: Harnessing the Unconventional Data Sources

Abstract

The field of Analytics is broad whereby some of the most important communities include, Statistics, Operational Research (OR), Artificial Intelligence (AI), Business Intelligence (BI), Web Analytics and Data Visualization. In a wider context, Analytics is usually defined as the process of developing actionable insights through data exploration, while AI is exclusively referred to as a completely autonomous process of exploring massive datasets in developing actionable insights using elaborate computer algorithms. The rise of unconventional data usage such as website data and satellite images has been significantly beneficial to the Department of Statistics Malaysia (DOSM) as a national statistics provider to produce complementary statistics in assisting the government's policy formulation. DOSM has started the implementation of Big Data Analytics (BDA) in producing more reliable and timeliness of official statistics. Initially, there are three (3) main modules in DOSM BDA, namely Trade by Enterprise Characteristics (TEC), Price Intelligence (PI), Public Maturity Assessment on Official Statistics (PMAOS). TEC involves integration of high volume of structured data with fuzzy matching techniques, PI is the modernizing data collection method with web crawling technique, while PMAOS uses text mining technique for sentiment analysis. Additionally, DOSM has launched the Malaysian Economic Monitoring System, Data Visualization that is presented in a dashboard-style of graphical display to easily gain insights and identify patterns and trends from economic data. On top of the existing initiatives, DOSM is actively exploring in BDA and utilizing unconventional data sources; Earth Observation is the exploration of environmental insights with satellite imagery to identify the proportion of

population and living quarters in remote and inaccessible areas; Informal Sector and Micro Industry Database and Analytics serve as a single reference of database by integrating a large scale of administrative data with non-traditional data sources i.e. social media & web scraping; and Mobile Positioning Data as a dynamic proxy indicator in assessing economic activity and human commuting patterns. In a nutshell, analytics and AI harness new data sources in generating new statistics. Nonetheless, these new statistics are not intended to replace conventional statistics, but to complement each other to gain richer insights.

Biography

As a Chief Statistician of the Department of Statistics Malaysia (DOSM), effective February 14, 2017, he has been devoted to his career by serving at DOSM for the past 30 years since his first posting as a Statistician in 1990. He has an extensive experience and specialise in both economic and social statistics where he has served and contributed in various fields of statistics such as national accounts; external trade; industrial & construction; and services. His career has spanned domestically as head of national statistical office, Executive Committee of Institute Statistics Malaysia, Chairman of the Working Group on Statistical Database and the Chairman of the National Organising Committee of ISI World Statistics Congress 2019. In 2020, he was honoured with 2020 Congress of Unions of Employees in the Public and Civil Services (CUEPACS) Malaysia Labour Day Icon Award. Recently, appointed as the Sixth Census Commissioner by His Majesty the Yang Di-Pertuan Agong XVI, Al-Sultan Abdullah Ri'ayatuddin Al-Mustafa Billah Shah on 1st April 2020.

His active participation and significant contribution at the international arena is acknowledged through his appointment as the Chair of the Bureau of the Committee on Statistics of the United Nations Economic and Social Commission for Asia and the Pacific (ESCAP) 2018-2020; Chair of Bureau for the 7th & 8th Sessions of Organisation of Islamic Cooperation Statistical Commission (OIC-StatCom) 2018-2019; Executive Committee of International Association of Official Statistics (IAOS) for two consecutive terms 2017-2019 & 2020-2022; Co-Chair Working Group (WG) on the Development of New Strategic Vision of OIC-StatCom; and most recently, he is re-appointed as vice-chair of the 7th Bureau of the ESCAP Committee on Statistics, 2020- 2022 terms.

He is also a member of various expert groups globally as well as international associations such as the ASEAN Community Statistical System (ACSS), UNWTO Tourism Statistics; Expert Group on Civil Registration and Vital Statistics (CRVS) ESCAP; Member of Friends of Chair Fundamental Principles of Official Statistics (FOC-FPOS) United Nations; High-Level Group for Partnership, Coordination and Capacity-Building for Monitoring Sustainable Goals (SDG); and the ASEAN Secretariat on SNA Statistics, ISI-South East Asia Network, Regional Programme on Economic Statistics and so on.

Keynote Speaker 2: Professor Dr. Guan-Hua Huang

National Yang Ming Chiao Tung University, Taiwan



Machine Learning Classification of Functional Brain Imaging for Parkinson's Disease Stage Prediction

Abstract

In this study, we analyze a dataset containing functional brain imaging from 6 normal healthy controls and 196 patients with Parkinson's disease (PD), which can be divided into 5 stages according to the severity of illness. The goal is to predict patients' PD illness stages via their functional brain images. Used approaches include multivariate statistical methods, ensemble learning models, and deep convolutional neural network (CNN). For statistical and ensemble models, PCA is performed to extract features, and the best combination of parameters is found by grid search. For CNN modeling, we use the technique of image augmentation to increase data size and adopt transfer learning to incorporate pretrained VGG16 weights and architecture into the model fitting. It is found that the deep CNN model with pretrained VGG16 weights and architecture outperforms other approaches, which can capture significant features from imaging and reach higher classification accuracy.

Biography

Dr. Huang's research is informed by the fact that in many medical studies, the definitive outcome is inaccessible, and a valid surrogate endpoint is then measured in place of the clinically most meaningful endpoint. Therefore, his primary research interest is in developing latent variable/class models for analyzing this kind of data structure.

Dr. Huang is interested in statistical problems in genetics, genomics, and bioinformatics. He is particularly interested in statistical validation of endophenotypes, high-throughput genomic data analysis, gene-gene interaction detection, next-generation sequencing data analysis, and copy number variation identification.

Dr. Huang is deeply invested in many scientific areas, including aging, nursing, schizophrenia, diabetes, sensory impairments (hearing, vision, olfaction), and data science.

Keynote Speaker 3: Professor Dr.rer.pol. Heri Kuswanto

Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia



Challenges in Weather and Climate Research: How can we benefit from open data source?

Abstract

It is not always easy to carry out research in weather and climate modelling due to the data constraints (data availability, data quality, etc.). Furthermore, scientists in developing countries also encounter the problem of access to the observation dataset from the officials, where the data are considered as an asset which should not be easily shared to the public. Open data source offers solutions to the problem. However, in most cases the data cannot be directly analyzed and need to be pre-processed to generate insight from it. This talk presents application of pre-processing and statistical methods applied to climate open data source e.g. data assimilation, downscaling, bias correction as well as machine learning. The methods have been successfully applied to various cases in Indonesia such as extreme characterization, drought prediction, impact analysis and many others.

Biography

Prof. Heri Kuswanto is a professor in computational statistics. He is currently a permanent lecturer at the Department of Statistics, Faculty of Science and Data Analytics – Institut Teknologi Sepuluh Nopember (ITS) Indonesia. He is also an active researcher at the Center for Disaster Mitigation and Climate Change – ITS. Prof. Kuswanto obtained his PhD from the Institute of Statistics- Leibniz Hannover University, Germany in

2009, and worked as Postdoctoral Research Associate in Laval University Canada in 2010. Since then, he has been working a lot on weather and climate analysis using statistical methods and machine learning. Prof. Kuswanto also works in econometrics and Bayesian fields.

Keynote Speaker 4: Assistant Professor Dr. Ubydul Haque

University of North Texas Health Science Center, USA



Spatio-temporal Dynamics of Dengue, Chikungunya, and Zika Viruses in Mexico: A Potential Co-circulation Problem

Abstract

Spatio-temporal patterns of Aedes-borne viruses (Chikungunya, Dengue, and Zika) throughout history have been dynamic because of changes in the human host, viral genetics, mosquito vector, and the environment. Nearly 105 million dengue infections are reported globally per year. The spread of Chikungunya and Zika into the Americas increased transmission and diagnosis complexity due to co-circulation in mosquitoes and humans. We present an integrated analysis of Aedes-borne diseases, local climate, and socio-demographic profiles of 2,511 municipalities in Mexico. We used SaTScan to detect spatial clusters and utilized the Pearson correlation coefficient, Randomized Dependence Coefficient and SHapley Additive exPlanations to analyze the influence of socio-demographic and climatic factors on incidence. We compared six models including XGBoost, decision tree, SVM with RBF kernel, K nearest neighbors, random forest and neural network to predict spatial clusters. Socio-demographic and climatic factors had significant and differential influence on Aedes-borne virus transmission in different regions of Mexico.

Biography

Education & Experience:

I have received a Ph.D. in Public Health from the Nagasaki University, Japan, Master of Science in Geoinformatics from the Royal Institute of Technology, Sweden, and a Bachelor of Science from Khulna University,

Bangladesh. I was also a guest Ph.D. student in Norway for one year and completed a post-doctoral research fellowship from the Johns Hopkins University. Prior to joining the UNTHSC School of Public Health in 2018, I held a previous faculty position with the Baldwin Wallace University, as Assistant Professor.

Teaching Areas & Public Health Interests:

I have been involved with global health and community-based research for more than 10 years, in the areas of spatial epidemiology. I have taught a variety of undergraduate and graduate courses, such as Biostatistics, GIS Health Analysis and Introduction to Public Health. I have had the opportunity to supervise and mentor undergraduate and graduate students.

Professional Activities & Awards:

I am an active member of the Association of American Geographers (AAG), the American Geophysical Union (AGU) and American Association for the Advancement of Science (AAAS). I serve as an editorial board member of 12 peer-review journals. I also received several awards from different institutes across the globe.

Scholarly Interests:

I started my career studying spatial malaria epidemiology in Bangladesh and gradually expanded my research into other settings. The global fund and other international donors have invested multi-billion dollars in support of these control programs. I wanted to learn the impact of these programs. My studies identified the location-specific successes and challenges which were welcomed by international donors, policymakers and public health professionals in those countries. I have investigated factors associated with the risk of mortality and determinants of patient survival due to the Ebola virus outbreak in West Africa and identified the spatiotemporal distribution and topographic risk factors for Zika virus in Colombia and Mexico for targeted interventions. More recently I contributed to analyzing five years of individually reported national databases on arbovirus infections in Mexico. Here we used machine learning techniques to analyze nearly quarter million laboratory confirmed cases on Chikungunya, Dengue, and Zika infections and identified the location-specific risk factors including climatic parameters. Looking to the future and to aid in preparedness, I investigated the role of climate change in the spread of deadly landslides in 130 countries over the last 20 years. Now I am studying the effects of climate change, community vulnerability and exposure to dengue in Laos and Thailand, a four years project sponsored by the Research Council of Norway.

As a spatial epidemiologist, using geospatial and informatics tools, I investigate the underlying factors that affect human health, particularly in less wealthy areas which are quite vulnerable to the impacts of

environmental variability and change infectious disease epidemiology as it relates to time, space and climate change. In addition, I examine the relationship between climate change and deadly landslides and its impact on public health globally. I also have a keen interest to contribute to the ongoing and unforeseen challenges in global health.

I have investigated factors associated with the risk of mortality and determinants of patient survival due to the Ebola virus outbreak in West Africa and identified the spatiotemporal distribution and topographic risk factors for Zika virus in Colombia and Mexico for targeted interventions.

I am also contributing to mapping, determining risk factors of COVID-19 transmission.

Keynote Speaker 5: Associate Professor Dr. Simon Fong

University of Macau, China



XAI Early Cancer Detection Using Swarm Intelligence Marker-Based Watershed Segmentation Algorithm and Evolving Graph Convolutional Network: Case of Lung Cancer Deep Analysis from Pulmonary Nodules and Cancer Images

Abstract

By merely relying on image processing and AI algorithms for early cancer detection, is insufficient. This talk covers the latest works we are now dealing with by fusing longitudinal big data and personalized data from biomarkers and tumour growth dynamic simulation for early cancer detection. All are just algorithms that do prediction. The efficacy of prediction depends very much on the reliability of the input data. In most cases of early cancer detections, the input data are fuzzy images of nodules that are too tiny and lack unique characteristics. So technically, using just AI, it is hard to achieve an effective solution. There are four components in our proposed early cancer detection methodology which will work cooperatively in the system: 1) specialized imaging equipment with super high resolution; 2) traditional imaging + nanotechnology biosensors; 3) traditional imaging + blood test; and 4) traditional imaging + tumour growth dynamic simulation / prediction. This research project involves a composite methodology comprising 2, 3 and 4, as well as some novel 4D deep learning which learns the time scale development of tumour growth. Approaches 2, 3, 4 will work in parallel, then information fusion is applied, 4D graph-based evolution deep learning. Finally, XAI is built using our previously published model namely White Learning Methodology for Cancer Evolution Analysis on IEEE

IT Professional Magazine. Preliminary results will be shared at the end of the talk.

Biography

Simon Fong graduated from La Trobe University, Australia, with a 1st Class Honour BEng. Computer Systems degree and a PhD. Computer Science degree in 1993 and 1998 respectively. Simon is now working as an Associate Professor at the Computer and Information Science Department of the University of Macau. He is a co-founder of the Data Analytics and Collaborative Computing Research Group in the Faculty of Science and Technology. Prior to his academic career, Simon took up various managerial and technical posts, such as systems engineer, IT consultant and e-commerce director in Australia and Asia. Dr. Fong has published over 500 international conference and peer-reviewed journal papers, mostly in the areas of data mining, data stream mining, e-health, meta-heuristics optimization algorithms, and their applications. He serves on the editorial boards of the IEEE JSAC Journal, IEEE IT Professional Magazine, and various special issues of SCIE-indexed journals. Simon is also an active researcher with leading positions such as Vice-Chair of IEEE Computational Intelligence Society (CIS) Task Force on Business Intelligence & Knowledge Management. Vice-President of Overseas Chinese International Biomedical Engineering Society, and Vice-Director of International Consortium for Optimization and Modelling in Science and Industry (iCOMSI).

Keynote Speaker 6: Professor Dr. Richard Millham

Durban University of Technology Kwazulu Natal, South Africa



Use of Bio-inspired Algorithms for Intelligent Energy Management of Renewable Energy Resources

Abstract

The impact of climate change, based primarily on the overuse of fossil fuels, has increased awareness of and interest in renewable energy. This is compounded by recent problems in South Africa, among other nations. In this talk, we will first outline how the use of renewable energy resources may mitigate climate change. Then, we will discuss how bio-inspired algorithms may play a role in smart energy management of both these renewable energy resources (the production side) and smart energy consumption (the load side) through dynamic optimization of load shifting and through optimal utilization of available renewable energy resources respectively.

Biography

Richard Millham is currently an Associate Professor at Durban University of Technology in Durban, South Africa and adjunct professor of computer science at the University of Energy and Natural Resources in Ghana. After thirteen years of industrial experience, he switched to academia and has worked at universities in Ghana, South Sudan, Scotland, and Bahamas. His research interests include software evolution, aspects of cloud computing with m-interaction, big data, data streaming, fog and edge analytics, and aspects of the Internet of Things. He is a Chartered Engineer (UK), a Chartered Engineer Assessor, and Senior Member of IEEE.

Keynote Speaker 7: Associate Professor Dr. Aiden Doherty

Oxford University, United Kingdom



Reproducible Machine Learning of Movement Behaviours in Terrabytes of Wearables Data

Abstract

My group has worked closely with UK Biobank to measure physical activity status in ~100,000 participants who agreed to wear a wrist-worn device for seven days. These measurements are now actively used by epidemiologists worldwide to demonstrate associations between physical activity, sleep, circadian rhythms and disease outcomes. In this talk I will share my group's work on reproducible machine learning of sleep and physical activity behaviours; and how they are facilitating new genetic and epidemiological insights. I will also highlight the opportunities for collaboration between clinicians and data scientists.

Biography

I am an Associate Professor at the University of Oxford and lead Health Data Research UK's national implementation project on reproducible machine learning. My research group at Oxford develops reproducible methods to analyse wearable sensor data in very large health studies to better understand the causes and consequences of disease. For example, we have developed methods to objectively measure physical activity in UK Biobank which are now actively used by researchers worldwide to demonstrate new associations with cardiovascular disease, depression, mood disorders, and others. These have allowed us to show that the lowest risk for cardiovascular disease in the UK Biobank cohort is seen at the highest level of accelerometer-measured physical activity, whether total, moderate-intensity, or vigorous-intensity. find that physical activity has a very strong association with incident cardiovascular disease. We have also developed machine learning methods to identify sleep and functional physical activity behaviours such as walking. In addition, we have discovered the first genetic variants associated with machine-learned sensor phenotypes. This work shows the first genetic evidence that physical activity might causally lower blood pressure.

Conference Schedule SCDS 2021-The 6th International Conference on Soft Computing in Data Science

Day 1: 2 November 2021 (Tuesday)

Time	Activities
8:00-8:45 am IDT 9:00-9:45 am MYT	Online Registration
8:45-9:00 am IDT 9:45-10:00 am MYT	National anthem UiTM + ITS song
9:00-9:10 am IDT 10:00-10:10 am MYT	Welcome Speech Dr Jerry Dwi Purnomo Conference Chair
9:10-9:20 am IDT 10:10-10:20 am MYT	Welcoming Remarks Prof. Dr. Mochamad Ashari, MEng Rector of ITS
9:20-9:30 am IDT 10:20-10:30 am MYT	Opening Remarks Professor Ts Dr Roziah Binti Mohd Janor Vice-Chancellor of UiTM
9:30-9:40 am IDT 10:30-10:40 am MYT	Online Exchange of Souvenirs between UiTM and ITS
9:40-9:50 am IDT 10:40-10:50 am MYT	Montage SCDS 2021
9:50-10:00 am IDT 10:50-11:00 am MYT	Online Photography Session
10:00-10:15 am IDT 11:00-11:15 am MYT	Break (play video)
10:15-11:00 am IDT 11:15-12:00 pm MYT	Keynote Speaker 1 YBhg Dato' Sri Dr Mohd Uzir Mahidin Chief Statistician, Department of Statistics, Malaysia Title: Analytics and Al: Harnessing the Unconventional Data Sources Moderator: Prof Dr Jasni Mohamad Zain, UiTM, Malaysia

	Keynote Speaker 2	
11:00-11:45 am IDT 12:00-12:45 pm MYT	Professor Dr. Guan-Hua, Huang National Yang Ming Chiao Tung University, Taiwan	
	Title: Machine Learning Classification of Functional Brain Imaging for Parkinson's Disease Stage Prediction	
	Moderator: Jerry Dwi Trijoyo Purnomo, ITS, Indonesia	
11:45-12:45 pm IDT 12:45-1:45 pm MYT	Lunch Break	
12:45-1:30 pm IDT 1:45-2:30 pm MYT	Keynote Speaker 3 Professor Dr.rer.pol Heri Kuswanto Institut Teknologi Sepuluh Nopember, Indonesia	
	Title: Challenges in Weather and Climate Research: How can we benefit from open data source?	
	Moderator: Jerry Dwi Trijoyo Purnomo, ITS, Indonesia	
1:30-2:30 pm IDT	Paper presentations (Session 1) Track 1: Al Techniques and Applications Track 2: Data Analytics and Technology	
	Zoom (Breakout rooms)	
2:30-3:30 pm MY I	Moderator:	
	Track 1: Prof. Dr Sri Hartati, UGM, Indonesia Track 2: Prof. Dr Adel Al-Jumaily, CSU, Australia	
2:30-2:45 pm IDT 3:30-3:45 pm MYT	Break	
2:45-3:45 pm IDT 3:45-4:45 pm MYT	Paper presentations (Session 2) Track 1: Al Techniques and Applications Track 2: Data Analytics and Technology	
	Zoom (Breakout rooms)	
	Moderator: Track 1: Prof. Dhiya Al-Jumeily, LJMU, UK Track 2: Dr. Azlan Ismail, UiTM, Malaysia	

Day 2: 3 November 2021 (Wednesday)

Time	Activities
8:00-8:45 am IDT 9:00-9:45 am MYT	Online Registration
9:00-9:45 am IDT 10:00-10:45 am MYT	Keynote Speaker 4 Assistant Professor Dr. Ubydul Haque University of North Texas Health Science Center,USA
	Title: Spatio-temporal Dynamics of Dengue, Chikungunya, and Zika Viruses in Mexico: A Potential Co-circulation Problem
	Moderator: Dr. Achmad Choiruddin, ITS, Indonesia
9:45-10:30 am IDT 10:45-11:30 am MYT	<i>Keynote Speaker 5</i> Associate Professor Dr Simon Fong University of Macau, China
	Title: XAI Early Cancer Detection Using Swarm Intelligence Marker-Based Watershed Segmentation Algorithm and Evolving Graph Convolutional Network: Case of Lung Cancer Deep Analysis from Pulmonary Nodules and Cancer Images
	Moderator: Dr.rer.pol Dedy Dwi Prastyo, ITS, Indonesia
10:30-10:45 am IDT 11:30-11:45 am MYT	Break
	Paper presentations (Session 3) Track 3: Data Mining and Image Processing Track 4: Machine and Statistical Learning
10:45-11:45 am IDT 11:45-12:30 pm MYT	Zoom (Breakout rooms)
	Moderator: Track 3: Dr. Ahmad Nazim Aimran, UiTM, Malaysia Track 4: Dr. Santi Wulan Purnami, ITS, Indonesia
11:45-12:45 pm IDT 12:45-1:45 pm MYT	Lunch Break
12:45-1:30 pm IDT 1:45-2:30 pm MYT	Keynote Speaker 6 Professor Dr. Richard C Millham Durban University of Technology, South Africa
	Title: Use of Bio-inspired Algorithms for Intelligent Energy Management of Renewable Energy Resources
	Moderator: Dr. Irhamah, ITS, Indonesia

1:30-2:15 pm IDT 2:30-3:15 pm MYT	Keynote Speaker 7 Associate Professor Dr. Aiden Doherty Oxford University, UK Title: Reproducible Machine Learning of Movement Behaviours in Terrabytes of Wearables Data Moderator: Professor Dr Yap Bee Wah, UiTM, Malaysia
2:15-4:30 pm IDT 3:15-5:30 pm MYT	Parallel Paper presentations (Session 4) Track 4: Machine and Statistical Learning Zoom (Breakout Rooms) Moderator: Dr. Santi Wulan Purnami, ITS, Indonesia
4:30-5:00 pm IDT 5:30-6:00 pm MYT	Best Paper Awards (10 papers) <i>Closing</i> Professor Dr Yap Bee Wah, UiTM, Malaysia

Note: IDT: Indonesia time (GMT+7) MYT: Malaysia time (GMT+8)

Paper Presentations Schedule

Track 1: Artificial Intelligence Techniques and Applications Venue : Zoom Date : 2 November 2021 (Tuesday) Session 1 : 2:30 pm – 3:30 pm MYT Session Chair : Prof. Sri Hartati, Universitas Gadjah Mada, Indonesia				
NO	TIME	ID	TITLE	AUTHORS
1.	2.30 pm – 2.45 pm MYT	1570741977	Comparison Performance of Long Short-Term Memory and Convolution Neural Network Variants on Online Learning Tweet Sentiment Analysis	Marina Yusoff; Syamil Ali
2.	2.45 pm – 3.00 pm MYT	1570742104	Performance Analysis of Hybrid Architectures of Deep Learning for Indonesian Sentiment Analysis	Theresia Gowandi; Hendri Murfi; Siti Nurrohmah
3.	3.00 pm – 3.15 pm MYT	1570742267	Machine Learning Based Biosignals Mental Stress Detection	Adel Al-Jumaily; Nafisa Matin; Azadeh Noori Hoshyar
4.	3.15 pm – 3.30 pm MYT	1570743602	Sentences Prediction Based on Automatic Lip-Reading Detection with Deep Learning Convolutional Neural Networks using Video- Based Features	Khalid Mahboob; Hafsa Nizami; Fayyaz Ali; Farrukh Alvi

Track 1: Artificial Intelligence Techniques and Applications

Venue : Zoom Date : 2 November 2021 (Tuesday) Session 2 : 3:45 pm – 4:45 pm MYT Session Chair : Prof. Dhiya Al-Jumeily, Liverpool John Moores University, UK

5.	3.45 pm – 4.00 pm MYT	1570743916	Unsupervised Learning Approach for Evaluating the Impact of COVID-19 on Economic Growth in Indonesia	Marieta Monica; Nadiah Ulfa Ayuningtiyas; Harun Al Azies; Muhammad Riefky; Hidayatul Khusna; Santi P Rahayu
6.	4.00 pm – 4.15 pm MYT	1570745178	Rainfall Prediction in Flood Prone Area using Deep Learning Approach	Siti Zuhairah Ramlan; Sayang Mohd Deni
7.	4.15 pm – 4.30 pm MYT	1570748307	Auto-DL: A Platform to Generate Deep Learning Models	Aditya Srivastava; Tanvi Shinde; Raj Joshi; Sameer Ansari; Nupur Giri
8.	4.30 pm – 4.45 pm MYT	1570751151	A Smart Predictive Maintenance Scheme for Classifying Diagnostic and Prognostic Statuses	Revi Palembiya; Muhammad Nanda Setiawan; Elnora Gultom; Adila Sekarratri Dwi Prayitno; Nani Kurniati; Mohammad Iqbal
Track 2: I	Data Analy	ytics and	Technologie	s
------------	------------	-----------	-------------	---
------------	------------	-----------	-------------	---

3.15 pm – 3.30 pm

MYT

4.

Γ

Hau	Track 2. Data Analytics and recimologies				
Venu Date Sess Sess	ue : Zoom : 2 November 2021 (1 sion 1 : 2:30 pm – 3:30 sion Chair : Prof. Adel /	⁻ uesday) pm MYT Al-Jumaily, Charle	s Sturt University, Aust	ralia	
NO	TIME	ID	TITLE	AUTHORS	
1.	2.30 pm – 2.45 pm MYT	1570755514	Optimal Portfolio Construction of Islamic Financing Instruments in Malaysia	Muhammad Firdaus Hussin, Siti Aida Sheikh Hussin, Zalina Zahid	
2.	2.45 pm – 3.00 pm MYT	1570728255	Analytics-Based on Classification and Clustering Methods for Local Community Empowerment in Indonesia	Dyah Yuniati; Kristina Pestaria Sinaga	
3.	3.00 pm – 3.15 pm MYT	1570743535	Two-step estimation for modeling the earthquake occurrences in Sumatra by Neyman-Scott Cox	Achmad Choiruddin; Tabita Susanto; Rahma Metrikasari	

1570749722

point processes

Construction of

Market Portfolios

Using Outlier Detection Algorithm

Optimal Stock

Gee Kok Tong; Keng-Hoong Ng; Wun-She Yap; Kok

Chin Khor

Track 2: Data Analytics and Technologies

Venue : Zoom Date : 2 November 2021Tuesday) Session 2 : 3:45 pm – 4:45 pm MYT Session Chair : Dr. Azlan Ismail, UiTM, Malaysia

5.	3.45pm – 4.00 pm MYT	1570749949	Simulating The Upcoming Trend of Malaysia's Unemployment Rate Using Multiple Linear Regression	Ahmad Asyraf Mohd Ibrahim; Muhammad Hazrani Abdul Halim; Nur Sarah Yasmin Mohamad Adib
6.	4.00 pm – 4.15 pm MYT	1570750176	Time series forecasting using a hybrid Prophet and Long Short-Term Memory model	Yih Hern Kong; Khai Yin Lim; Wan Yoke Chin
7.	4.15 pm – 4.30 pm MYT	1570752416	Financial Analytics on Malaysia's Equity Fund Performance and Its Timing Liquidity	Siti Farizah Saad; Siti Meriam Zahari; Muhammad Azri Mohd
8.	4.30 pm – 4.45 pm MYT	1570752418	An Autoregressive Distributed Lag (ARDL) Analysis of the Relationships between Employees Provident Fund's Wealth and its Determinants	Haidah Syafi Parly; Siti Meriam Zahari; Muhammad Asmu'i Abdul Rahim; S. Sarifah Radiah Shariff

Track 3: Data Mining and Image Processing

Venue : Zoom Date : 3 November 2021 (Wednesday) Session 3 : 11:45 am – 12:30 pm MYT Session Chair: Dr. Ahmad Nazim Aimran, UiTM, Malaysia

NO	ТІМЕ	ID	TITLE	AUTHORS
1.	11.45 am – 12.00 pm MYT	1570743254	Iris Segmentation Based on an Adaptive Initial Contour and Partly- Normalization	Shahrizan Jamaludin; Nasharuddin Zainal; W Mimi Diyana W Zaki; Ahmad Faisal Mohamad Ayob
2.	12.00 pm – 12.15 pm MYT	1570749339	Identifying the Important Demographic and Financial Factors Related to the Mortality Rate of COVID-19 with Data Mining Techniques	Nur Sara Zainudin; Keng- Hoong Ng; Kok Chin Khor
3.	12.15 pm – 12.30 pm MYT	1570751067	Local Image Analysis of Malaysian Herbs Leaves using Canny Edge Detection Algorithm	Zuraini Othman; Sharifah Sakinah Syed Ahmad; Fauziah Kasmin

Track 4: Machine and Statistical Learning

Venue : Zoom Date : 3 November 2021 (Wednesday) Session 3 : 11:45 am – 12:30 pm MYT Session Chair : Dr. Santi Wulan Purnami, ITS, Indonesia.

NO	ТІМЕ	ID	TITLE	AUTHORS
1.	11.45 am – 12.00 pm MYT	1570740699	Amniotic Fluids Classification Using Combination of Rules- Based and Random Forest Algorithm	Putu Desiana Wulaning Ayu; Sri Hartati; Aina Musdholifah; Detty S Nurdiati
2.	12.00 am – 12.15 pm MYT	1570741688	A Modified Inverse Gaussian Poisson Regression with an Exposure Variable to Model Infant Mortality	Selvi Mardalena; Purhadi Purhadi; Jerry Dwi Trijoyo Purnomo; Dedy Dwi Prastyo
3.	12.15 am – 12.30 pm MYT	1570742248	Poisson and Logistic Regressions for Inhomogeneous Multivariate Point Processes: A Case Study in the Barro Colorado Island Plot	Ahmad Husain; Achmad Choiruddin

Track 4: Machine and Statistical Learning

Venue : Zoom Date : 3 November 2021 (Wednesday) Session 4 : 3:15 pm – 5:30 pm MYT Session Chair : Dr. Santi Wulan Purnami, ITS, Indonesia.

4.	3.15 pm – 3.30 pm MYT	1570742883	Entropy-based Fuzzy Weighted Logistic Regression for Classifying Imbalanced Data	Ajiwasesa Harumeka; Santi Wulan Purnami; Santi P Rahayu
5.	3.30 pm – 3.45 pm MYT	1570743049	Quantifying the Impact of Climatic Factors on Dengue Incidence using Generalized Linear Mixed Model with Spatio- Temporal Bayesian Poisson Random Effects Approach	Nik Nur Fatin Fatihah Sapri; Wan Fairos Wan Yaacob; Yap Bee Wah
6.	3.45 pm – 4.00 pm MYT	1570743427	The Effect of Cancers Treatment in Quality of Life of the Patient: Meta- Analysis Approach	Santi Wulan Purnami; Prilyandari Dina Saputri; Bambang Otok
7.	4.00 pm – 4.15 pm MYT	1570743428	Hybrid of Time Series Regression, Multivariate Generalized Space-Time Autoregressive, and Machine Learning for Forecasting Air Pollution	Hendri Prabowo; Dedy Dwi Prastyo; Setiawan Setiawan
8.	4.15 pm – 4.30 pm MYT	1570743523	Hybrid Machine Learning for Forecasting and Monitoring Air Pollution in Surabaya	Achmad Choiruddin; Suhartono Suhartono; Hendri Prabowo; Muhammad Hisyam Lee
9.	4.30 pm – 4.45 pm MYT	1570743552	Survival Analysis of Diabetes Mellitus Patients Using Semiparametric Approach	Jerry Dwi Trijoyo Purnomo; Santi Wulan Purnami; Febry Hilmi Anshori; Albertus Kurnia Lantika

10.	4.45 pm – 5.00 pm MYT	1570743675	Application of machine learning in credit risk scorecard	Choon Yi Lee; Siew Khew Koh; Min Cherng Lee; Wei Yeing Pan
11.	5.00 pm – 5.15 pm MYT	1570743745	Multivariate Analysis to Evaluate the Impact of COVID-19 on the Hotel Industry in Indonesia	Prilyandari Dina Saputri; Arin Angrenani; Dinda Guminta; Fonda Leviany; Ika Fitriana; Santi P Rahayu; Hidayatul Khusna
12.	5.15 pm – 5.30 pm MYT	1570749367	Identifying Sequential Influence in Predicting Engagement of Online Social Marketing for Video Games	Joseph Chia Wei Chen; Nurulhuda Firdaus Mohd Azmi

Proceeding Abstracts

Track 1: Artificial Intelligence Techniques and Applications

NO	ID	TITLE	AUTHORS
1.	1570741977	Comparison Performance of Long Short-Term Memory and Convolution Neural Network Variants on Online Learning Tweet Sentiment Analysis	Marina Yusoff; Syamil Ali
2.	1570742104	Performance Analysis of Hybrid Architectures of Deep Learning for Indonesian Sentiment Analysis	Theresia Gowandi; Hendri Murfi; Siti Nurrohmah
3.	1570742267	Machine Learning Based Biosignals Mental Stress Detection	Adel Al-Jumaily; Nafisa Matin; Azadeh Noori Hoshyar
4.	1570743602	Sentences Prediction Based on Automatic Lip-Reading Detection with Deep Learning Convolutional Neural Networks using Video-Based Features	Khalid Mahboob; Hafsa Nizami; Fayyaz Ali; Farrukh Alvi
5.	1570743916	Unsupervised Learning Approach for Evaluating the Impact of COVID- 19 on Economic Growth in Indonesia	Marieta Monica; Nadiah Ulfa Ayuningtiyas; Harun Al Azies; Muhammad Riefky; Hidayatul Khusna; Santi P Rahayu
6.	1570745178	Rainfall Prediction in Flood Prone Area using Deep Learning Approach	Siti Zuhairah Ramlan; Sayang Mohd Deni

7.	1570748307	Auto-DL: A Platform to Generate Deep Learning Models	Aditya Srivastava; Tanvi Shinde; Raj Joshi; Sameer Ansari; Nupur Giri
8.	1570751151	A Smart Predictive Maintenance Scheme for Classifying Diagnostic and Prognostic Statuses	Revi Palembiya; Muhammad Nanda Setiawan; Elnora Gultom; Adila Sekarratri Dwi Prayitno; Nani Kurniati; Mohammad Iqbal

Comparison Performance of Long Short-Term Memory and Convolution Neural Network Variants on Online Learning Tweet Sentiment Analysis

Muhammad Syamil Ali¹ and Marina Yusoff^{[1,2*}

¹Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia

²Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Al- Khawarizmi Complex, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia.

marina998@uitm.edu.my

Abstract.

Sentiment analysis can act as an assisted tool in improving the quality of online teaching and learning between teachers and students. Twitter social media platform currently has more than 500 million tweets sent each day which is equal to 5787 tweets per second. Therefore, it is hard to track users' overall opinions on the topics contained in social media. To catch up with the feedback on online learning, it is crucial to detect the topic being discussed and classify users' sentiments towards those topics. Even though there are many approaches in developing sentiment analysis models, DL models prove to provide the best performance in the sentiment analysis field. Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) are two mainstream models in DL used for sentiment analysis classification. Therefore, we evaluate CNN, LSTM, and its hybrids to classify sentiment or an online learning tweet from 2020 until 2021 of 23168 tweets. CNN-LSTM, LSTM-CNN, Bidirectional LSTM, CNN-Bidirectional LSTM models were designed and evaluated based on random hyperparameter tuning. We explain the proposed methodology and model design illustration. The outcome assesses the superiority of all models with a remarkable improvement of accuracy and a reduction loss when applying the random oversampling technique. Specifically, the LSTM-CNN model with random oversampling technique outperformed the other six models with an accuracy of 87.40% and loss value of 0.3432. However, the computational time has increased when with random oversampling technique. Thus, in the future, the performance can be improved on computational time and hyperparameter selection with the employment of nature-inspired computing for fast and optimal results.

Keywords: CNN, LSTM, Online learning, Random Oversampling, Sentiment Analysis

Performance Analysis of Hybrid Architectures of Deep Learning for Indonesian Sentiment Analysis

Theresia Gowandi^{1,} Hendri Murfi^{1,} and Siti Nurrohmah¹

¹ Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Depok 16424, Indonesia

hendri@ui.ac.id

Abstract. Sentiment analysis is one of the fields of Natural Language Processing that builds a system to recognize and extract opinions in the form of text into positive or negative sentiment. Nowadays, many researchers have developed methods that yield the best accuracy in performing analysis sentiment. Three particular models are Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU), which have deep learning architectures. CNN is used because of its ability to extract essential features from each sentence fragment, while LSTM and GRU are used because of their ability to memorize prior inputs. GRU has a more straightforward and more practical structure compared to LSTM. These models have been combined into hybrid architectures of LSTM-CNN, CNN-LSTM, and CNN-GRU. In this paper, we analyze the performance of the hybrid architectures for Indonesian sentiment analysis in e-commerce reviews. Besides all three combined models mentioned above, we consider one more combined model, which is GRU-CNN. We evaluate the performance of each model, then compare the accuracy of the standard models with the combined models to see if the combined models can improve the performance of the standard. Our simulations show that almost all of the hybrid architectures give better accuracies than the standard models. Moreover, the hybrid architecture of LSTM-CNN reaches slightly better accuracies than other hybrid architectures.

Keywords: CNN, Deep Learning, GRU, LSTM

Machine Learning Based Biosignals Mental Stress Detection

Adel Ali Al-Jumaily ^{1,2}, Nafisa Matin³, Azadeh Noori Hoshyar⁴

¹School of Computing, Mathematics and Engineering, Charles Sturt University aal-jumaily@csu.edu.au ²College of Computing, Fahad Bin Sultan University. aal-jumaily@fbsu.edu.sa ³ School of Biomedical Engineering, University of Technology Sydney Nafisa.Matin@student.uts.edu.au ⁴School of Science, Engineering & Physical Sciences, Federation University Australia a.noorihoshyar@federation.edu.au

Abstract. Mental Stress can be defined as a normal physiological and biological reaction to an incident or a situation that makes a person feel challenged, troubled, or helpless. While dealing with stress, some changes occur in the biological

function of a person, which results in a considerable change in some biosignals such as, Electrocardiogram (ECG), Electromyography (EMG), Electrodermal Activity (EDA), respiratory rate. This paper aims to review the effect of mental stress on mental condition and health, the changes in biosignals as an indicator of the stress response and train a model to detect stressed states using the biosignals. This paper delivers a brief review of mental stress and biosignals correlation. It represents four Support Vector Machine (SVM) models trained with ECG and EMG features from an open access database based on task related stress. After performing comparative analysis on the four types of trained SVM models with chosen features, Gaussian Kernel SVM is selected as the best SVM model to detect mental stress which can predict the mental condition of a subject for a stressed and relaxed condition having an accuracy of 93.7%. These models can be investigated further with more biosignals and applied in practice, which will be beneficial for the physician.

Keywords: Mental Stress, ECG, EMG, Machine learning, Support Vector Machine, Gaussian Kernel SVM.

Sentences Prediction based on Automatic Lip-Reading Detection with Deep Learning Convolutional Neural Networks using Video-Based Features

Khalid Mahboob¹, Hafsa Nizami², Fayyaz Ali³, Farrukh Alvi⁴

Department of Software Engineering

Sir Syed University of Engineering & Technology, Karachi, Pakistan kmahboob@ssuet.edu.pk¹

Abstract. Lip-reading is the process of deciphering text from a speaker's visual interpretation of facial, lip, and mouth movements without using audio. The challenge is traditionally divided into two stages: creating or learning visual characteristics and prediction. End-to-end techniques for deep lip-reading have been popular in recent years. Existing work on end-to-end models, on the other hand, only does word classification rather than sentence-level sequence prediction. Longer words improve human lip-reading ability, suggesting the relevance of characteristics that capture the temporal context in an inconsistent communication channel. In this study, an end-to-end model based on deep learning convolutional neural network shave been employed to develop an automated lip-reading system that uses a recurrent network spatio temporal convolutions, and the connectionist temporal classification loss to translate a variable-length series of video frames to text. The accuracy of the trained lip-reading process in predicting sentences was evaluated using video-based features.

Keywords: Lip-reading, Convolutional Neural Networks, Model.

Unsupervised Learning Approach for Evaluating the Impact of COVID-19 on Economic Growth in Indonesia

Marieta Monica¹, Nadiah Ulfa Ayuningtiyas², Harun Al Azies³, Muhammad Riefky⁴, Hidayatul Khusna⁵ and Santi Puteri Rahayu⁶

1,2,3,4,5,6 Institut Teknologi Sepuluh Nopember, Kota Surabaya, Indonesia

Abstract. The spread of COVID-19 that occurred in several parts of Indonesia resulted in the economy getting worse. Almost all business fields in Indonesia experienced a contraction. Each province has a different impact from one another so that the policies taken cannot be generalized. Therefore, this study was conducted to group provinces based on the value of Gross Regional Domestic Product (GRDP) in 2019 and 2020 using unsupervised learning. The vear 2019 represents the economic conditions before COVID-19 and 2020 represents the conditions during the COVID-19 pandemic. The unsupervised learning method used in this research is the K-Means, K-Medoids, SOM, as well as Hybrid SOM-K-Means methods. From the grouping results obtained, then a comparison of the results of the four methods will be carried out to obtain the best method based on the Silhouette Index. The results show that based on 2019 data, the grouping of provinces using the K-Medoids, SOM and hybrid SOM-K-Means methods is three clusters. While the results of grouping using the K-Means method are as many as two clusters. On the other hand, based on 2020 data, both the K-mean, K-Medoids, SOM, and hybrid SOM-K-Means methods show the same results, namely the grouping is carried out in two clusters. The best method based on 2019 GRDP data is K-Means with two groups. Meanwhile, in 2020, the best method obtained is the K-mean, K-Medoids, and SOM methods with two groups. In addition, all economic growth indicators have contracted or decreased due to the impact of the COVID-19 pandemic.

Keywords: Unsupervised Learning, COVID-19, Gross Regional Domestic Product (GRDP).

Rainfall Prediction in Flood Prone Area using Deep Learning Approach

Siti Zuhairah Ramlan¹ and Sayang Mohd Deni²

¹ Faculty of Computer and Mathematical Sciences, UiTM, 40450 Shah Alam, Selangor, Malaysia

² Center for Statistics and Decision Science Studies, Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Faculty of Computer and Mathematical Sciences, UiTM, 40450 Shah Alam, Selangor, Malaysia

zuhairah87@gmail.com, sayang@tmsk.uitm.edu.my

Abstract. Flood is a catastrophic event that contributes to the impact on socioeconomic of a developing country. Flood prone area is also known as the location at risk as heavy rainfall attribute to flood events. This circumstance leads to the effective flood mitigation phase. One of the critical problems facing by responsible government agencies is to minimize future uncertainties. This study explores four deep learning methods with univariate rainfall temporal data in gauging station near to flood prone area. Four models tested are Multi-layer Perceptron MLP, Long Short Term Memory LSTM, Stacked-LSTM, and hybrid model Convolutional Neural Network CNN-LSTM. The aim of this paper is to compare and determine the best method for rainfall prediction of next day event. Model comparison is conducted by comparing the correlation coefficient, Root Mean Square error (RMSE) and Mean Absolute Error (MAE). Based on the selected locations, the results showed at Kuantan station generally underfitting. meanwhile the Kuala Krai station does not showed discrepancies between training and testing dataset. It could be concluded that by adding the complexity to the model, will not significantly improved the model prediction. The LSTM model with 16 memory blocks was outperformed in both locations. The potential of deep learning methods should be considered for rainfall amount prediction due to less complexity and much easier to be applied into dataset for model fitting purposes. It may assist to accommodate strategic precautions for flood mitigation phases in flood prone areas.

Keywords: Floods Mitigation, Long Short Term Memory (LSTM), Rainfall Prediction, Univariate Time Series.

Auto-DL: A Platform to Generate Deep Learning Models

Aditya Srivastava¹, Tanvi Shinde¹, Raj Joshi¹, Sameer Ahmed Ansari¹ and Dr. Nupur Giri¹

¹Vivekanand Education Society's Institute of Technology, Mumbai, India

{2017.aditya.srivastava, 2017.tanvi.shinde, 2017.raj.joshi, 2017.sameer.ansari, nupur.giri}@ves.ac.in

Abstract. Deep Learning (DL) model building is a tedious and taxing process. The number of prerequisites is high and a lot of time is invested. Hence, there is a scope of Automation. Code to build DL models follows a standard structure, broadly classified into four categories (Imports, Data Input, Model Creation, and Evaluation). The work in this paper proposes to automate this core structure and build a Graphical User Interface (GUI) based tool/platform called "Auto-DL" which, on defining the task and training data, generates code in python for the specified deep learning model. The paper then discusses the platform capabilities and evaluates it and the generated code against various quantitative and qualitative parameters.

Keywords: Deep Learning (DL), Code Generator, Neural Networks (NN), Keras, PaaS.

A Smart Predictive Maintenance Scheme for Classifying Diagnostic and Prognostic Statuses

Revi Asprila Palembiya¹, Muhammad Nanda Setiawan², Elnora Oktaviyani Gultom³, Adila Sekarratri Dwi Prayitno⁴, Nani Kurniati⁵, and Mohammad Iqbal⁶

Department of Mathematics, Institut Teknologi Sepuluh Nopember, Sukolilo, Surabaya, 60111, Indonesia, ¹reviasprila@gmail.com, ²m.nanda98@hotmail.com, ³elnoragultom@gmail.com, ⁴adilasekarratri@gmail.com, ⁵nanikur@ie.its.ac.id Corresponding Author: ⁶igbal@ matematika.its.ac.id

Abstract. This study attempts to propose a smart predictive maintenance method to classify manufacturing machines' diagnostic and prognostic statuses. The main goal of this study is to reduce the manual predictive maintenance budgets of manufactures in Indonesia. In the proposed method, we perform feature maps to obtain the binary states of sensor data, which is further clustered into the machine's error states (diagnostic status) and the machine' useful life states (prognostic status). Moreover, the proposed method comprises the two states predictions of machines based on *Deep Long Short Term Memory*. The proposed method is demonstrated on the Rawmill and Kiln machines of a cement factory in Indonesia for evaluation performances. Without labelling manually, we investigated the annotation of both states, which are similar to the ground truth. In addition, the proposed method can achieved high accuracy and outperformed to another baseline method.

Keywords: Smart Predictive Maintenance, Cement Factory, Diagnostic State Prediction, Prognostic State Prediction, Deep Learning.

Track 2: Data Analytics and Technologies

NO	ID	TITLE	AUTHORS
1.	1570755514	Optimal Portfolio Construction of Islamic Financing Instruments in Malaysia	Muhammad Firdaus Hussin, Siti Aida Sheikh Hussin, Zalina Zahid
2.	1570728255	Analytics-Based on Classification and Clustering Methods for Local Community Empowerment in Indonesia	Dyah Yuniati; Kristina Pestaria Sinaga
3.	1570743535	Two-step estimation for modeling the earthquake occurrences in Sumatra by Neyman-Scott Cox point processes	Achmad Choiruddin; Tabita Susanto; Rahma Metrikasari
4.	1570749722	Construction of Optimal Stock Market Portfolios Using Outlier Detection Algorithm	Gee Kok Tong; Keng-Hoong Ng; Wun-She Yap; Kok Chin Khor
5.	1570749949	Simulating The Upcoming Trend of Malaysia's Unemployment Rate Using Multiple Linear Regression	Ahmad Asyraf Mohd Ibrahim; Muhammad Hazrani Abdul Halim; Nur Sarah Yasmin Mohamad Adib

6.	1570750176	Time series forecasting using a hybrid Prophet and Long Short-Term Memory model	Yih Hern Kong; Khai Yin Lim; Wan Yoke Chin
7.	1570752416	Financial Analytics on Malaysia's Equity Fund Performance and Its Timing Liquidity	Siti Farizah Saad; Siti Meriam Zahari; Muhammad Azri Mohd
8.	1570752418	An Autoregressive Distributed Lag (ARDL) Analysis of the Relationships between Employees Provident Fund's Wealth and its Determinants	Haidah Syafi Parly; Siti Meriam Zahari; Muhammad Asmu'i Abdul Rahim; S. Sarifah Radiah Shariff

Optimal Portfolio Construction of Islamic Financing Instrument in Malaysia

Muhammad Firdaus Hussin Author¹ Siti Aida Sheikh Hussin² and Zalina Zahid³

^{1,2,3} Universiti Teknologi MARA, Shah Alam 50400 Selangor Malaysia

sitiaida@tmsk.uitm.edu.my.

Abstract. Conventional financing and Islamic financing have the same mechanism of operation, although their nature is different. This difference creates different risk, which requires a proper investigation and observation. Thus, the conventional portfolio manager cannot use the same set of proportion or strategy when they want to invest in Islamic financial product. This study determines the optimal portfolio combination and its proportion for Islamic bank financing which include several contracts (Murabahah, Mudharabah, Musharakah, Bai Bithaman Ajil, Ijarah, Ijarah Thumma Al Bai, Istisna'). We apply single index model (SIM) since SIM enables precise calculation of the composition of each asset (financing) by identifying the value of Excess Return to Beta (ERB) as well as the cut-off point based on the acquisition of equivalent rate of profit sharing for each financing. The results show that the optimal composition of portfolio consist of Ijarah or leasing (59.93%), Musharakah or joint venture (29.18%) and Murabahah or sales (10.89%). The portfolio expected return is 1.20 percent with portfolio risk of 1.67 percent.

Keywords: Financial Analytics, Portfolio Optimisation, Islamic Finance.

Analytics-Based on Classification and Clustering Methods for Local Community Empowerment in Indonesia

Dyah Yuniati^{1,2} and Kristina Pestaria Sinaga²

 ¹ Dkatalis, Jakarta, Indonesia
 ² Dept. Master in Information System Management Bina Nusantara University

¹dyah.yuniati@binus.ac.id; ²kristina.sinaga@binus.edu

Abstract. West Papua is reportedly the second-most populous province in Indonesia. The United Nations International Children's Emergency Fund (UNICEF) highlights Papua's performance in selecting the Sustainable Development Goals (SDG) indicators compared to other provinces in the country. The data shows that food, nutrition, health, education, housing, water, sanitation, and protection are defined as multidimensional child poverty. Population statistics and poverty figures show that inter-provincial equity in Indonesia needs to be re-measured. In 2008, the Regional Governments of Papua and West Papua Provinces implemented a Community Empowerment Program called "PNPM RESPEK", which provided direct community assistance for IDR 100 million per village. To determine the people's level of understanding and perception towards this program, PNPM RESPEK, in collaboration with the Central Statistics Agency, conducted an integrated PNPM RESPEK Evaluation Survey in July 2009. Based on the survey results, this paper identifies a model (pattern) of understanding the people of Papua and West Papua towards the program and finds the best method to build this model through classification techniques. Then the data model was also tested using unsupervised learning. the clustering method. The experimental results show that the J48 decision tree produces the highest accuracy compared to the others. As for clustering, the clustering hierarchy provides the best accuracy. Decision Tree J48 has the best accuracy with an accuracy of 97.31%. In this case, 97.31% of the people of Papua and West Papua who receive direct community assistance meet the level of understanding and perception of the PNPM RESPEK Program.

Keywords: SDG, Poverty, Equality, Machine Learning, J48 Decision Tree, hierarchical clustering.

An efficient algorithm for modeling the earthquake occurrences in Sumatra by Neyman–Scott Cox point processes

Abstract. The Cox point process is highly considered for earthquake modeling. However, the complex earthquake data which involves a large number of occurrences and geological variables often requires expensive computation. This study aims to propose an efficient algorithm based on the two-step procedure by constructing the first and second order composite likelihoods. We apply the proposed procedure to model the earthquake distribution in Sumatra.

Keywords: Big data problem · cluster point process · disaster risk reduction · earthquake modeling · spatial point pattern.

Construction of Optimal Stock Market Portfolios Using Outlier Detection Algorithm

Gee-Kok Tong¹, Keng-Hoong Ng¹, Wun-She Yap² and Kok-Chin Khor²

¹ Faculty of Computing & Informatics, Multimedia University, 63100, Cyberjaya, Selangor, Malaysia

² Lee Kong Chian Faculty of Engineering Science, Universiti Tunku Abdul Rahman, Kajang 43000, Malaysia

{gktong;khng}@mmu.edu.my;{yapws;kckhor}@utar.edu.my

Abstract. It is not easy for investors to trade in stock markets as building stock portfolios requires financial knowledge and consumes much time. Thus, this study aims to construct optimal stock market portfolios for investors using a LOF-based methodology. We used an outlier detection algorithm called Local Outlier Factor (LOF) to identify outperforming stocks from a stock pool. We then constructed two portfolios using these outperforming stocks, namely, tangency and equal-weighted portfolios and compared their performance against the benchmark portfolios, namely, the market portfolio and the cash market. It was followed by using Mean-Variance Portfolio Optimisation (MVPO) to measure the performance of the portfolios and determined whether they were efficient. To identify the most efficient portfolio, we used the Sharpe ratio. In general, the results showed that both tangency and equal-weighted portfolios gave better returns than the benchmark portfolios. To conclude, the LOF-based methodology helps to build and identify profitable stock portfolios for investors.

Keywords: Stock portfolios, local outlier factor, mean-variance portfolio, Sharpe ratio.

Simulating The Upcoming Trend of Malaysia's Unemployment Rate Using Multiple Linear Regression

Nur Sarah Yasmin Mohamad Adib¹, Ahmad Asyraf Mohd Ibrahim¹ and Muhammad Hazrani Abdul Halim¹

¹ Universiti Teknologi MARA (UiTM) Shah Alam, 40450 Shah Alam, Selangor, Malaysia

sarahyasmin7117@gmail.com,asyrafibrahim0609@gmail.com, hazrani@uitm.edu.my

Abstract. This research paper presents the approach of applying the simulation technique to predict the upcoming trend of Malaysia's unemployment rate. The recent Malaysia's unemployment rate has fluctuated at quite a high rate ever since the COVID-19 pandemic occurred. Population growth, Growth Domestic Product (GDP), inflation rate, interest rate, exchange rate, investment, government expenditure and most importantly the number of COVID-19 cases act as the independent variables in this paper. The Multiple Linear Regression (MLR) is used to determine the significance of each variable to be included in the model and also to simulate the upcoming trend of Malaysia's unemployment rate. The result of the analysis shows that the upcoming five years trend of Malaysia's unemployment rate will continue to increase in the future based on the average value of the simulations conducted.

Keywords: Unemployment Rate, Multiple Linear Regression, Monte Carlo Simulation, COVID-19, Prediction.

Time series forecasting using a hybrid Prophet and Long Short-Term Memory model

Kong Yih Hern¹, Lim Khai Yin², Chin Wan Yoke³

^{1,2,3} Tunku Abdul Rahman University College, Setapak, Kuala Lumpur, 53300, Malaysia

¹kongyh-wa15@student.tarc.edu.my ²limky@tarc.edu.my ³chinwy@tarc.edu.my

Abstract. Forecasting analysis is a common research topic these days. The development in this area has allowed organizations to retrieve useful information and make important decisions based on the forecast results. Different forecasting models are used to model data with different characteristics as each of the forecasting model has its own strength and weakness. As such, Hybrid Prophet-LSTM that combines Long Short-Term Memory (LSTM) and FBProphet (Prophet) is introduced. This study aims to examine the effectiveness of the hybrid model and the influence of holiday effect to the forecast result. Weighted Mean Absolute Percentage Error (WMAPE), Mean Absolute Deviation (MAD), R^2 value, and Root mean square error (RMSE) were used to evaluate the performance of the proposed hybrid model. The proposed Hybrid Prophet-LSTM is found to outperform both the standalone LSTM and Prophet, and holiday effect shows high attitude of influence to the forecast result.

Keywords: Prophet; LSTM; Hybrid forecasting; Time series forecasting; Combined forecast; Holiday effect

Financial Analytics on Malaysia's Equity Fund Performance and Its Timing Liquidity

Siti Farizah Saad¹, Siti Meriam Zahari² and Muhammad Azri Mohd³

¹ Standard Chartered Bank, Malaysia ^{2,3} Universiti Teknologi MARA, Shah Alam, Malaysia

¹fariezahsaad@yahoo.com,²mariam@tmsk.uitm.edu.my, ³azri@tmsk.uitm.edu.my

Abstract. This study focuses on equity fund performance in the Malaysian market with three different time frames (daily, monthly, and yearly) and examines the relationship between its expected return and liquidity timing. Traditional financial ratios such as Sharpe, Jensen Alpha, Treynor index, and Capital Asset Pricing Model (CAPM) help in analyzing the performance of the equity fund whereas Trading Volume and Turnover methods were utilized to measure the fund liquidity. Deduced from the analysis, the equity fund performs differently within the group itself, depending on the time frames stated. This study also found that the liquidity timing affects the expected return where the fund manager can use the beta values from the fund analysis to increase their market exposure, prior to market timing liquidity.

Keywords: Financial Ratios, Liquidity Timing, Equity Fund

An Autoregressive Distributed Lag (ARDL) Analysis of the Relationships between Employees Provident Fund's Wealth and its Determinants

Haidah Syafi Parly¹, Siti Meriam Zahari², Muhammad Asmu'i Abdul Rahim³ and S.Sarifah Radiah Shariff 4

¹Bancassurance Associates, RHB Bank, Malaysia ^{2,3,4}Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia

> ¹syafihaidah@gmail.com, ²mariam@tmsk.uitm.edu.my, ³asmui@tmsk.uitm.edu.my, ⁴radiah@tmsk.uitm.edu.my

Abstract. This study empirically examines the potential determinants of the wealth of Malaysian Employees Provident Fund (EPF). The auto-regressive distributed lag (ARDL) bounds test approach was employed to determine the existence of a long-run relationship between EPF's wealth and its determinants, namely inflation rate (INF), gross domestic product (GDP), life expectancy (EXP), and the Gini coefficient (GINI) as a proxy for income inequality. Data for this study consisted of a set of annual time series data from 1980 to 2018. The findings indicate the existence of a long-run equilibrium relationship between the total EPF wealth and GDP, income inequality, INF, and EXP. Further, there is a relatively guick adjustment in the total EPF wealth when all of the determinants change. GDP, EXP, and GINI were found to be significantly important drivers of EPF wealth. A 10% change in GDP will result in a long-run change of 25.8% in the total EPF wealth. Meanwhile, a 1% increase in income inequality and life expectancy will reduce the total EPF wealth by 11.8% and 20.4%, respectively. One of the implications of this study's findings is that EPF wealth would grow continuously with an increase in economic prosperity and improvements in income inequality and demographic structure. Thus, this study suggests that EPF wealth can be improved by making positive changes to economic and demographic factors in ensuring the sustainability of the fund.

Keywords: Employee Provident Fund, ARDL, Long-run Relationship, Shortrun Dynamics, Income Inequality

Track 3: Data Mining and Image Processing

NO	ID	TITLE	AUTHORS
1.	1570743254	Iris Segmentation Based on an Adaptive Initial Contour and Partly-Normalization	Shahrizan Jamaludin; Nasharuddin Zainal; W Mimi Diyana W Zaki; Ahmad Faisal Mohamad Ayob
2.	1570749339	Identifying the Important Demographic and Financial Factors Related to the Mortality Rate of COVID-19 with Data Mining Techniques	Nur Sara Zainudin; Keng- Hoong Ng; Kok Chin Khor
3.	1570751067	Local Image Analysis of Malaysian Herbs Leaves using Canny Edge Detection Algorithm	Zuraini Othman; Sharifah Sakinah Syed Ahmad; Fauziah Kasmin

Iris Segmentation Based on an Adaptive Initial Contour and Partly-Normalization

Shahrizan Jamaludin¹, Nasharuddin Zainal², W Mimi Diyana W Zaki² and Ahmad Faisal Mohamad Ayob¹

¹ Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia

² Department of Electrical, Electronic and Systems Engineering, Faculty of Engineering and Built Environment,

Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

Abstract. Active contour is accurate for iris segmentation on the non-ideal and noisy iris images. However, understanding on how active contour reacts to the motion blur or blurry iris images is presently unclear and remains a major challenge in iris segmentation perspective. Moreover, studies on the initial contour position in the blurry iris images are infrequently reported and need further clarification. In addition, convergence or evolution speed is still a major drawback for active contour as it moves through the boundaries in the iris images. Based on the above issues, the experiment is conducted to obtain an accurate and fast iris segmentation algorithm for the blurry iris images. The initial contour is also investigated to clarify its positioning for the blurry iris segmentation. To achieve these objectives, the Wiener filter is used for preprocessing. Next, the morphological closing is applied to eliminate reflections. Then, the adaptive Chan-Vese active contour (ACVAC) algorithm is designed from the adaptive initial contour (AIC), δ and stopping function. Finally, the partlynormalization is designed where only prominent iris features near to the inner iris boundary are selected for normalization and feature extraction. The experimental results show that the proposed algorithm achieves the highest segmentation accuracy and the fastest computational time than the other active contour-based methods. The accurate initial contour position in the blurry iris images is clearly clarified. This shows that the proposed method is accurate for iris segmentation on the blurry iris images.

Keywords: Iris Segmentation, Adaptive Chan-Vese Active Contour, Adaptive Initial Contour, Segmentation Accuracy, Computational Time.

Identifying the Important Demographic and Financial Factors Related to the Mortality Rate of COVID-19 with Data Mining Techniques

Nur Sara Zainudin¹, Keng-Hoong Ng¹ and Kok-Chin Khor²

¹ Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia
² Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Bandar Sungai Long, Kajang 43000, Malaysia

nursaraz@gmail.com; khng@mmu.edu.my; kckhor@utar.edu.my

Abstract. The whole world has been greatly affected by the recent emergence of the COVID-19 pandemic since December 2019, and the death toll has reached millions. Thus, this problem needs to be addressed and mitigated immediately. In this study, the primary objective is to determine the factors affecting the mortality rate of COVID-19 in demographic and financial factors. This study utilised supervised learning methods with feature selection methods: filter and wrapper, to identify factors attributed significantly to the Case Fatality Ratio (CFR), a measure for mortality. The result showed that the wrapper method running K-Nearest Neighbour with the Sequential Forward Selection produced the feature subset that gave the best result. The feature selection results also suggest that the factor - household debt is the key to affecting the mortality rate of this infectious disease.

Keywords: COVID-19, Mortality Rate, Classification, Regression, Feature Selection

Local Image Analysis of Malaysian Herbs Leaves using Canny Edge Detection Algorithm

Zuraini Othman^{1*}, Sharifah Sakinah Syed Ahmad¹, Fauziah Kasmin¹, Azizi Abdullah² and Nur Hajar Zamah Shari³

¹ Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya,

76100 Durian Tunggal, Melaka, Malaysia ² Center for Artificial Intelligence Technology, Faculty of information Science and Technology, Universiti Kebangsaan Malaysia 43600 Bangi, Selangor Darul Ehsan, Malaysia ³ Forestry and Environment Division, Forest Research Institute Malaysia (FRIM), 52109 Kepong, Malaysia

zuraini@utem.edu.my

Abstract. Machine vision helps a lot with the latest recognition technology. To get the best recognition results, the initial phase of image processing should be done as best as possible. This phase involves the production of an image map of the resulting image using edge detection. The Canny method is often used because of its performance of producing meticulous strong edges but this method is sensitive to changes in image intensity when involving complex images such as image leaves and results in a lot of noise at the edges of the resulting image. This is because in this method the threshold value is selected empirically on the image globally. In this study, local image analysis will be discussed along with its impact on the resulting image edge results. In addition, a set of data from herbal leaves in Malaysia has also been produced by containing ground truth images for each herbal image available. The results from this study found that the locally analysing image approach has outperform the global approach of the conventional Canny method. Findings from this study may help the identification system in the future.

Keywords: Machine Vision, Edge Detection, Canny Method, Local Image Analysis, Malaysian Herbs Leaves Images.

Track 4: Machine and Statistical Learning

NO	ID	TITLE	AUTHORS
1.	1570740699	Amniotic Fluids Classification Using Combination of Rules-Based and Random Forest Algorithm	Putu Desiana Wulaning Ayu; Sri Hartati; Aina Musdholifah; Detty S Nurdiati
2.	1570741688	A Modified Inverse Gaussian Poisson Regression with an Exposure Variable to Model Infant Mortality	Selvi Mardalena; Purhadi Purhadi; Jerry Dwi Trijoyo Purnomo; Dedy Dwi Prastyo
3.	1570742248	Poisson and Logistic Regressions for Inhomogeneous Multivariate Point Processes: A Case Study in the Barro Colorado Island Plot	Ahmad Husain; Achmad Choiruddin
4.	1570742883	Entropy-based Fuzzy Weighted Logistic Regression for Classifying Imbalanced Data	Ajiwasesa Harumeka; Santi Wulan Purnami; Santi P Rahayu
5.	1570743049	Quantifying the Impact of Climatic Factors on Dengue Incidence using Generalized Linear Mixed Model with Spatio-Temporal Bayesian Poisson Random Effects Approach	Nik Nur Fatin Fatihah Sapri; Wan Fairos Wan Yaacob; Yap Bee Wah
6.	1570743427	The Effect of Cancers Treatment in Quality of Life of the Patient: Meta-Analysis Approach	Santi Wulan Purnami; Prilyandari Dina Saputri; Bambang Otok
7.	1570743428	Hybrid of Time Series Regression, Multivariate Generalized Space-Time Autoregressive, and Machine Learning for Forecasting Air Pollution	Hendri Prabowo; Dedy Dwi Prastyo; Setiawan Setiawan
8.	1570743523	Hybrid Machine Learning for Forecasting and Monitoring Air Pollution in Surabaya	Achmad Choiruddin; Suhartono Suhartono;

			Hendri Prabowo; Muhammad Hisyam Lee
9.	1570743552	Survival Analysis of Diabetes Mellitus Patients Using Semiparametric Approach	Jerry Dwi Trijoyo Purnomo; Santi Wulan Purnami; Febry Hilmi Anshori; Albertus Kurnia Lantika
10.	1570743675	Application of machine learning in credit risk scorecard	Choon Yi Lee; Siew Khew Koh; Min Cherng Lee; Wei Yeing Pan
11.	1570743745	Multivariate Analysis to Evaluate the Impact of COVID-19 on the Hotel Industry in Indonesia	Prilyandari Dina Saputri; Arin Angrenani; Dinda Guminta; Fonda Leviany; Ika Fitriana; Santi P Rahayu; Hidayatul Khusna
12.	1570749367	Identifying Sequential Influence in Predicting Engagement of Online Social Marketing for Video Games	Joseph Chia Wei Chen; Nurulhuda Firdaus Mohd Azmi

Amniotic Fluids Classification Using Combination of Rules-Based and Random Forest Algorithm

Putu Desiana Wulaning Ayu^{1,4}, Sri Hartati^{2*}, Aina Musdholifah², Detty S Nurdiati³

¹Doctoral Program Department of Computer Science and Electronics, Faculty of Mathematics and Natural Science, Universitas Gadjah Mada, Sekip Utara, Yogyakarta 55281, Indonesia

² Department of Computer Science and Electronics, Faculty of Mathematics and Natural Science

Universitas Gadjah Mada, Sekip Utara, Yogyakarta 55281, Indonesia ³Department of Obstetrics and Gynaecology, Faculty of Medicine, Universitas Gadjah Mada, Sekip Utara, Yogyakarta 55281, Indonesia ⁴ Department of Information Technology, Faculty Computer and Informatics, Institut Teknologi dan Bisnis STIKOM Bali, Denpasar, Indonesia wulaning_ayu@stikom-bali.ac.id, {shartati,aina_m,detty}@ugm.ac.id

 $\ast \ Corresponding \ author$

Abstract. One of the studies of fetal anatomy is to measure the volume and echogenicity of the amniotic fluid. This study categorizes amniotic fluid into six, such as Oligohydramnion Clear, Oligodramnion Echogenic, Polygohydramnion Clear, and Polygohydramnion Echogenic, as well as Normal Clear and Normal Echogenic. Meanwhile, the current condition in determining the category of amniotic remains a perception difference among doctors, especially in identifying volume and echogenicity study, which is always conducted manually and visually. Therefore, this research proposed a model for the classification of amniotic fluid by combining the rule-based of the Single Deep Pocket (SDP) method and the Random Forest algorithm. The rulebased used was based on the feature value obtained by extracting the Single Deep Pocket (SDP) feature. Also, the Random Forest algorithm was formed to classify amniotic fluid based on the condition of echogenicity, which includes clear and echogenic based on texture features using First Order Statistical (FOS) and Gray Level Co-occurrence Matrix (GLCM) methods. The average value performance of the proposed model showed an accuracy of 90.52%, a precision of 95.72%, a recall of 75.57%, and an F-measure of 81.51%. Considering this result, the proposed model showed an average increase in accuracy performance of 9.12%, precision of 14.92%, and recall of 0.51% value of the model in previous studies.

Keywords: Amniotic fluid classification, Feature extraction, Rule-based, Random Forest.

A Modified Inverse Gaussian Poisson Regression with an Exposure Variable to Model Infant Mortality

Selvi Mardalena¹, Purhadi^{1,2}, Jerry Dwi Trijoyo Purnomo¹ and Dedy Dwi Prastyo¹

¹ Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia

²purhadi@statistika.its.ac.id

Abstract. Infant mortality has generally been increasing and has become an issue that urgently needs to be addressed. As the number of infant deaths is count data, a Poisson regression model is needed to determine the causal factors. However, the assumption of equidispersion in Poisson regression is rarely satisfied. The overdispersion issue is frequently found in real data. Thus, this research employs mixed Poisson distribution modeling to overcome the overdispersion issue, namely, the inverse Gaussian Poisson regression (IGPR) model. In this study, a simple IGPR model, a modified IGPR model, and the negative binomial regression (NBR) model are compared. The results show that the modified IGPR model and the NBR model with an exposure variable outperform the benchmark, based on the global deviance and Akaike Information Criteria (AIC) value, to model the number of infant deaths in East Nusa Tenggara, Indonesia. The significant predictors that affect the number of infant mortalities are the percentage of complete basic immunization, the percentage of low birth weight (LBW), the percentage of babies under six months who receive exclusive breastfeeding, the percentage of infants who receive vitamin A, and the percentage of births assisted by health workers in the district.

Keywords: Poisson Inverse Gaussian, Negative Binomial, Overdispersion, Exposure, infant mortalities.

Poisson and Logistic Regressions for Inhomogeneous Multivariate Point Processes: A Case Study in the Barro Colorado Island Plot

Ahmad Husain¹ and Achmad Choiruddin²

¹ Department of Statistics, Institut Teknologi Sepuluh Nopember, Indonesia ahmadhusain.018@gmail.com
² Department of Statistics, Institut Teknologi Sepuluh Nopember, Indonesia

choiruddin@its.ac.id

Abstract. This study aims to extend the estimating equations based on the Poisson and logistic regression likelihoods to model the intensity of a multivariate point process. The proposed approaches result in a framework equivalent to the estimation procedure for generalized linear model. The estimation is different from the existing methods where repetition independently with respect to the number of types of point process is obliged. Our approach does not require repetition and hence could be computationally faster. We implement our method to analyze the distribution of 9-species of trees in the Barro Colorado Island rainforest with respect to 11-environmental variables.

Keywords: Logistic regression · Multivariate point pattern · Poisson regression.

Entropy-based Fuzzy Weighted Logistic Regression for Classifying Imbalanced Data

Ajiwasesa Harumeka^{1, 2}, Santi Wulan Purnami¹, Santi Puteri Rahayu¹

¹Department of Statistics, Institut Teknologi Sepuluh Nopember, Indonesia ²Central Bureau of Statistics-East Java, Indonesia

aharumeka@gmail.com or ajiwasesa@bps.go.id

Abstract. Logistic regression is a popular classification method that has disadvantages when it is applied to large data. Truncated Regularized Iteratively Reweighted Least Square (TR-IRLS) is a method that overcomes this problem. This method is similar to Support Vector Machine (SVM) because both of them have similar loss functions and parameters that can adjust the bias and variance. Both methods were designed with the assumption of balanced data, so that they are not suitable to be applied on imbalanced data. Both methods were developed to overcome problem on imbalanced data. TR-IRLS was developed into Rare Event Weighted Logistic Regression (RE-WLR) and SVM was developed into Fuzzy Support Vector Machine (FSVM). Both RE-WLR and FSVM use weights based on class differences, so that RE-WLR had better performance than TR-IRLS on imbalanced data whereas FSVM was better than SVM. Then, Entropy-based Fuzzy Support Vector Machine (EFSVM) was developed by obtaining weighting values not only based on class differences, but also based on entropy. EFSVM further enhanced minority class interest in imbalanced data than SVM and even FSVM. Therefore, Entropy-based Fuzzy Weighted Logistic Regression (EFWLR) is proposed by adopting the success of Entropy-based Fuzzy Membership (EF) as weight on SVM. This study applied EF as weight on Weighted Logistic Regression for binary classification. Experiments on 20 simulation data and 5 benchmark data with various rarity schemes validated that the EFWLR outperformed TR-IRLS and RE-WLR based on AUC. EFWLR had more efficient AUC than RE-WLR on imbalanced data.

Keywords: Binary Classification, Entropy-based Fuzzy, Imbalanced Data, Weighted Logistic Regression.
Quantifying the Impact of Climatic Factors on Dengue Incidence using Generalized Linear Mixed Model with Spatio-Temporal Bayesian Poisson Random Effects Approach

Nik Nur Fatin Fatihah Sapri¹, Wan Fairos Wan Yaacob^{2*} and Yap Bee Wah^{1,3}

 ¹ Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
 ² Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Kelantan, Lembah Sireh, 15050 Kota Bharu, Kelantan, Malaysia
 ³ Institute for Big Data Analytics and Artificial Intelligence (IBDAAI) Kompleks Al-Khawarizmi, Universiti Teknologi MARA, 40450 Shah Alam Selangor, Malaysia

niknurfatinfatihah@gmail.com
{wnfairos, yapbeewah}@uitm.edu.my

Abstract. Dengue fever is a global life-threatening vector-borne disease which is mainly distributed by the vector Aedes Aegypti mosquito. It is known that the development and survivorship of this vector depends on surrounding climate. The dengue outbreak in Kelantan. Malavsia is alarming. The aim of the study was to compare the fixed effects Negative Binomial GLM and Bayesian Poisson GLMM in prediction of dengue incidence in Kelantan. The data involved daily number of reported dengue cases (1st January 2013 - 31st December 2017) in ten districts of Kelantan which was collected from Ministry of Health Malaysia. The climate variables, average daily temperature, relative humidity and rainfall (climatic data) were obtained from NASA's Global Climate Change website, while the population data were from Department of Statistics Malaysia. Statistical modeling results revealed that the fixed effects Negative Binomial GLM failed to fit the daily dengue incidence when serious epidemic occurred. The spatio-temporal Bayesian Poisson GLMM model improved the prediction of dengue incidence. Relative humidity at lag 7 days and 21 days and average temperature at lag 21 days were found to be significant contributing factors of dengue incidence in Kelantan. The findings of the study are significant to respective local authorities in providing vital information for early dengue warning systems in a particular area. This is important for authorities to monitor and reduce dengue incidence in endemic areas and to safeguard the community from dengue outbreak.

Keywords: Dengue, Generalized Linear Model. Generalized Linear Mixed Model, Spatio-Temporal Bayesian Poisson Model, Climate

The Effect of Cancers Treatment in Quality of Life of the Patient: Meta-Analysis Approach

Santi Wulan Purnami, Prilyandari Dina Saputri, and Bambang Widjanarko Otok

Department of Statistics, Institut Teknologi Sepuluh Nopember, Kampus ITS Keputih Sukolilo, Surabaya 60111, Indonesia

Abstract. Cancer treatments are expected to improve the quality of life of the patients. Thus the comparison of the quality of life before and after treatment is important to be present. Several dimensions are used in measuring the quality of life, i.e., global quality of life, physical functioning, role functioning, emotional functioning, cognitive functioning, and social functioning. We applied metaanalysis to identify the comparison based on the included study material from Google Scholar. The duration in measuring the quality of life after treatment is at least two months. The treatments included in this study are chemotherapy, radiotherapy, and surgery. There are 12 suitable study materials from 1150 available articles. The models used are a random-effects model since there exists heterogeneity between studies for all variables. In general, the guality of life of patients after treatment has increased, which is indicated by the negative mean difference estimation for all variables. However, the improvement of the quality of life in the dimensions of role functioning and cognitive functioning is not significant. Meanwhile, in the other four dimensions, the improvement was significant. Thus, cancer treatments are able to improve the patient's quality of life in the long term.

Keywords: Cancer, Meta-Analysis, Quality of Life, Treatment.

Hybrid of Time Series Regression, Multivariate Generalized Space-Time Autoregressive, and Machine Learning for Forecasting Air Pollution

Hendri Prabowo¹, Dedy Dwi Prastyo¹, and Setiawan¹

¹ Department of Statistics, Institut Teknologi Sepuluh Nopember Kampus ITS - Sukolilo, Surabaya 60111, Indonesia

dedy-dp@statistika.its.ac.id

Abstract. The purpose of this study is to propose a new hybrid of space-time models by combining the time series regression (TSR), multivariate generalized space-time autoregressive (MGSTAR), and machine learning (ML) to forecast air pollution data in the city of Surabaya. The TSR model is used to capture linear patterns of data, especially trends and double seasonal. The MGSTAR model is employed to capture the relationship between locations, and the ML model is used to capture nonlinear patterns from the data. There are three ML models used in this study, namely feed-forward neural network (FFNN), deep learning neural network (DLNN), and long short-term memory (LSTM). So that there are three hybrid models used in this study, namely TSR-MGSTAR-FFNN, TSR-MGSTAR-DLNN, and TSR-MGSTAR-LSTM. The hybrid models will be used to forecast air pollution data consisting of CO, PM₁₀, and NO₂ at three locations in Surabaya simultaneously. Then, the performance of these three-combined hybrid models will be compared with the individual model of TSR and MGSTAR, two-combined hybrid models of MGSTAR-FFNN, MGSTAR-DLNN, MGSTAR-LSTM, and hybrid TSR-MGSTAR models based on the RMSE and sMAPE values in the out-of-sample data. Based on the smallest RMSE and sMAPE values, the analysis results show that the best model for forecasting CO is MGSTAR, forecasting PM_{10} is hybrid TSR-MGSTAR, and forecasting NO_2 is hybrid TSR-MGSTAR-FFNN. In general, the hybrid model has better accuracy than the individual models. This result is in line with the results of the M3 and M4 forecasting competition.

Keywords: Air Pollution, Forecast, Hybrid, Machine Learning, Space-Time.

Hybrid Machine Learning for Forecasting and Monitoring Air Pollution in Surabaya

Suhartono¹, Achmad Choiruddin^{1,} Hendri Prabowo¹, and Muhammad Hisyam Lee²

 ¹ Department of Statistics, Institut Teknologi Sepuluh Nopember JI. Raya ITS Sukolilo, Surabaya 60115, Indonesia
 ² Department of Mathematical Science, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia

¹choiruddin@its.ac.id

Abstract. This research aims to propose hybrid machine learnings for forecasting and monitoring air pollution in Surabaya. In particular, we introduce two hybrid machine learnings, i.e. hybrid Time Series Regression - Feedforward Neural Network (TSR-FFNN) and hybrid Time Series Regression - Long Short-Term Memory (TSR-LSTM). TSR is used to capture linear patterns from data, whereas FFNN or LSTM is used to capture non-linear patterns. Fifteen halfhourly series data, i.e. CO, NO₂, O₃, PM₁₀, and SO₂ in three SUF stations at Surabaya, are used as the case study. We compare the forecasting accuracy of these hybrid machine learnings with several individual methods (i.e. TSR, ARIMA, FFNN, and LSTM), and combined methods (i.e. TSR with AR error and TSR with ARMA error). The identification step showed that these air pollution data have double seasonal patterns, i.e. daily and weekly seasonality. The comparison results showed that no superior method that yields the most accurate forecast for all series data. Moreover, the results showed that hybrid methods gave more accurate forecast at 8 series data, whereas the individual methods yielded better results at 7 series data. It supported that methods that are more complex do not always produce better forecasts than simple methods, as shown by the first result of the M3 competition. Additionally, the results of the forecast of air pollution index for monitoring air pollution in Surabaya show that the air quality is in good and moderate air pollution levels for duration of 19.30 to 03.00 and 0.30 to 19.30, respectively.

Keywords: Air Pollution, Forecasting, Hybrid, Machine Learning, Monitoring.

Survival Analysis of Diabetes Mellitus Patients Using Semiparametric Approach

Jerry Dwi Trijoyo Purnomo^(⊠), Santi Wulan Purnami, Febry Hilmi Anshori, and Albertus Kurnia Lantika

Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

jerrypurnomo@gmail.com

Abstract. The disease that attacks the human body and often gets special attention is diabetes. Diabetes mellitus is a non-communicable disease that is most commonly suffered by the world's population. Diabetes is a condition that interferes with the body's ability to process glucose in the blood, otherwise known as blood sugar. So, most patients have survival of only a few months. Therefore, research was conducted on the survival of people with diabetes mellitus and factors that affect it during the event. The method used in this study was the cox regression model. The results obtained from this study are three variables that significantly affect the survival of diabetes mellitus patients, namely Genetics, Age, and Diet. Then the variables Genetic, Age, and Diet became part of *Cox Proportional Hazard* (PH) modeling.

Keywords: Cox Proportional Hazard, Diabetes Mellitus, Survival.

Application of machine learning in credit risk scorecard

Choon Yi Lee^{a,1*}, Siew Khew Koh^b, Min Cherng Lee^b, Wei Yeing Pan^b

 ^a LKC Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, 43000 Kajang, Selangor, Malaysia
 ^b Department of Mathematics and Actuarial Sciences, LKC Faculty of Engineering and Science.

University Tunku Abdul Rahman, 43000 Kajang, Selangor, Malaysia

choonyi@lutar.my

Abstract. Machine Learning models have been extensively researched in the area of credit scoring. Banks have put in substantial resources into improving the credit risk model performance as improvement in accuracy by a fraction could translate into significant future savings. Given the lack of interpretability in machine learning models, it is often not used for capital provisioning in banks. This paper uses the Taiwan Credit Card dataset and illustrates the use of machine learning techniques to improve construction of credit scoring models. In factor transformation for a credit scorecard, Decision Tree technique showed the ability to produce quick and predictive transformation rule as compared to traditional approach. The resulting scorecard also showed high predictive power on Test sample. Given the ability of machine learning to produce predictive result, banks should explore on the techniques to improve their overall credit risk management framework. Credit underwriting scorecard could be built using higher discriminatory power techniques, as more good customers are likely to be better than score cut-off and thus accepted by banks.

Keywords: Data mining, Credit risk scorecard, Machine classifier, Bank credit underwriting scorecard, Predictive modeling.

Multivariate Analysis to Evaluate the Impact of COVID-19 on the Hotel Industry in Indonesia

Prilyandari Dina Saputri^[∞], Arin Berliana Angrenani, Dinda Galuh Guminta, Fonda Leviany, Ika Nur Laily Fitriana, Santi Puteri Rahayu, and Hidayatul Khusna

> Department of Statistics, Institut Teknologi Sepuluh Nopember, Kampus ITS Keputih Sukolilo, Surabaya 60111, Indonesia

Abstract. Pandemic has a significant impact on many sectors, especially for the hotel industry sector in Indonesia. To find out the impact of the pandemic on the hotel industry sector, we conducted an inferential statistic using a nonparametric location test to determine the significant differences between variables in 2019 and 2020. Then, we conducted cluster analysis using K-Means and Self-Organizing Map (SOM) methods. We also create the perceptual mapping by Biplot. Using the paired-fisher test for multivariate nonparametric location test, we found that the differences between variables relating to the occupancy rate of hotel rooms in 2019 and 2020 have been significantly decreasing. According to the biplot analysis, in 2019, the characteristics between provinces were quite different. While, in 2020, almost all provinces have identical characteristics. The result shows that SOM and K-Means have the same performance. In 2019, there are 4 clusters, and in 2020 there are 3 clusters. There has been a change in cluster members before and during the COVID-19 pandemic. Bali is the province that most affected by the COVID-19 incident because the tourism sector is the primary regional income. We found that the small and medium hotel industry is severely affected by COVID-19 outbreaks.

Keywords: Biplot, Clustering, COVID-19, Hotel, K-Means, SOM.

Identifying Sequential Influence in Predicting Engagement of Online Social Marketing for Video Games

Joseph Chia Wei Chen¹ and Nurulhuda Firdaus Binti Mohd Azmi Ais¹

¹ Razak Faculty of Technology & Informatics, University Teknologi Malaysia Kuala Lumpur (UTM KL), 54100 Kuala Lumpur, Malaysia

> joseph.wc.chia@gmail.com huda@utm.my

Abstract. Advancement of online social networks has seen digital marketing use platforms like YouTube and Twitch as key levers for video games marketing. Identifying key influencer factors in these emerging platforms can both deliver better understanding of user behavior in consumption and engagement towards marketing on social platforms and deliver great business value towards video game makers. However, data sparsity and topic maturity has made it difficult to identify user behavior over a sequence of different marketing videos, with a key challenge being identifying key features and distinguishing their contribution to the measure that defines sustained engagement over sequential marketing. This paper presents a method to understand sequential behavioral patterns by extracting features from marketing frameworks and develop a supervised model that takes all the features into consideration to identify the best contributing features to predicting engagement that delivers sustained interest for the next video in a series of marketing videos on YouTube. Experiment results on dataset demonstrate the proposed model is effective within constraint.

Keywords: Online Social Marketing, Regression Prediction, Video Game Trailers, Machine Learning, Sequential Pattern

Conference Organization

Patron

Professor Ts Dr Roziah Janor

Prof. Dr. Mochamad Ashari, MEng

Honorary Chairs

Mohd Nazip Suratman Jasni Mohamad Zain Haryani Haron Hamzah Fansuri Azlinah Mohamed Michael W. Berry Mahadhir Bin Aziz

Conference Chair Yap Bee Wah Jerry D.T. Purnomo

Secretary

Nurain Ibrahim Achmad Choiruddin

Secretariat

Santi Puteri Rahyu Wibawati Adatul Mukarromah Khairil Anuar Md Isa

Finance Committee

Nurbaizura Borhan Mohd Fikri Hadrawi Nur Asyikin Abdullah Vita Ratnasari Azira Mohamed Amin Vice-Chancellor, Universiti Teknologi MARA Malaysia Rectors, Institut Teknologi Sepuluh Nopember, Indonesia

Universiti Teknologi MARA, Malaysia Universiti Teknologi MARA, Malaysia Universiti Teknologi MARA, Malaysia Institut Teknologi Sepuluh Nopember, Indonesia Universiti Teknologi MARA, Malaysia University of Tennessee, USA Malaysia Digital Economy Corporation, Malaysia

Universiti Teknologi MARA, Malaysia Institut Teknologi Sepuluh Nopember, Indonesia

Universiti Teknologi MARA, Malaysia Institut Teknologi Sepuluh Nopember, Indonesia

Institut Teknologi Sepuluh Nopember, Indonesia Institut Teknologi Sepuluh Nopember, Indonesia Institut Teknologi Sepuluh Nopember, Indonesia Universiti Teknologi MARA, Malaysia

Universiti Teknologi MARA, Malaysia Universiti Teknologi MARA, Malaysia Universiti Teknologi MARA, Malaysia Institut Teknologi Sepuluh Nopember, Indonesia Universiti Teknologi MARA, Malaysia

Technical Program Committee

Jasni Mohamad Zain	Universiti Teknologi MARA, Malaysia
Azlinah Mohamed	Universiti Teknologi MARA, Malaysia
Azlan Ismail	Universiti Teknologi MARA, Malaysia
Kartika Fithiriasari	Institut Teknologi Sepuluh Nopember, Indonesia
Santi Wulan Purnami	Institut Teknologi Sepuluh Nopember, Indonesia
Dedy D. Prastyo	Institut Teknologi Sepuluh Nopember, Indonesia
Rizauddin Saian	Universiti Teknologi MARA, Malaysia

Publicity & Sponsorship Committee

<i>z</i>	
Ahmad Nazim Aimran	Universiti Teknologi MARA, Malaysia
Nooritawati Md Tahir	ReNeU, Universiti Teknologi MARA, Malaysia
Siti Haslinda Mohd Din	Jabatan Perangkaan Malaysia
Sharifah Sakinah Syed Ahmad	Universiti Teknikal Malaysia
Muhammad Asmu'i Abdul Rahim	Universiti Teknologi MARA, Malaysia
Fairoza Amira Hamzah	WAI Ambassador Malaysia
Syed Muslim Jameel	Sir Syed University of Engineering and Technology, Pakistan

Publication Committee (Program Book)

Marina Yusoff	Universiti Teknologi MARA, Malaysia
Saiful Farik Mat Yatin	Universiti Teknologi MARA, Malaysia
Sayang Mohd Deni	Universiti Teknologi MARA, Malaysia
Muhammad Ahsan	Institut Teknologi Sepuluh Nopember, Indonesia

Website Committee

Mohamad Asyraf Abd Latif	Universiti Teknologi MARA, Malaysia
Muhammad Ridhwan Muhamad Razali	Universiti Teknologi MARA, Malaysia

International Scientific Committee

Nur Iriawan	Institut Teknologi Sepuluh Nopember, Indonesia
Mario Köppen	Kyutech Institute of Technology, Japan
Sri Hartati	Universitas Gadjah Mada, Indonesia
Chidchanok Lursinsap	Chulalongkorn University, Thailand
Aiden Doherty	Oxford University, UK
Richard Millham	Durban University of Technology, South Africa
Simon Fong	University of Macau, China

Layth Sliman Hizir Sofyan Dhiya Al-Jumeily Adel Al-Jumaily Min Chen Mohammed Bennamoun Agus Harjoko Yasue Mitsukura Dariusz Krol Richard Weber Jose Maria Pena Yusuke Nojima Siddhivinayak Kulkarni Tahir Ahmad Daud Mohamed Ku Ruhana Ku Mahamud Ubydul Haque Rajalida Lipikorn Suhartono Wahyu Wibowo Edi Winarko Retantyo Wardoyo Rayner Alfred Faiz Ahmed Mohamed Elfaki Mohamed Chaouch Abdul Haris Rangkuti Sapto Wahyu Indratno Angela Kim

Efrei Paris, France Liverpool John Moores University, UK Syiah Kuala University, Indonesia Charles Sturt University, Australia Oxford University, England University of Western Australia, Australia Universitas Gadjah Mada, Indonesia Keio University, Japan Wroclaw University of Poland, Poland University of Chile, Santiago, Chile Technical University of Madrid, Madrid, Spain Osaka Perfecture University, Japan University of Ballarat, Australia Universiti Teknologi Malaysia, Malaysia Universiti Teknologi MARA, Malaysia Universiti Utara Malaysia, Malaysia University of North Texas Health Care Centre, USA Chulalongkorn University, Thailand Institut Teknologi Sepuluh Nopember, Indonesia Institut Teknologi Sepuluh Nopember, Indonesia Universitas Gadjah Mada, Indonesia Universitas Gadjah Mada, Indonesia Universiti Malaysia Sabah, Malaysia Qatar University, Qatar Qatar University, Qatar **BINUS University**, Indonesia Institut Teknologi Bandung, Indonesia Women in AI, Asia Pacific

Reviewers

Achmad Choiruddin Adam Shariff Adli Aminuddin Ahmad Farid Abidin Ahmad Nazim Aimran Ahmad Taufek Abdul Rahman Aida Mustapha Albert Guvenis Alexander Bolotov Ali Seman Angela Lee Anis Farihan Mat Raffei Azlan lobal Azlin Ahmad Azlinah Mohamed Azman Taa Bong Chih How Chidchanok Lursinsap Chin Kim On **Choong-Yeun Liong Christoph Friedrich** Dariusz Krol Daud Mohamad **Dedy Prastyo** Deepti Theng Dittaya Wanvarie Dr Eng Harish Kumar Edi Winarko Ensar Gul Ezzatul Akmal Kamaru Zaman Faroudia Abid Foued Saâdaoui Hamidah Jantan Hamzah Abdul Hamid Hanaa Ali Haris Rangkuti

Institut Teknologi Sepuluh Nopember, Indonesia Universiti Malaysia Pahang, Malaysia Universiti Teknologi MARA, Malaysia Universiti Teknologi MARA, Malavsia Universiti Teknologi MARA, Malaysia Universiti Tun Hussein Onn Malaysia Boğaziçi University, Turkey University of Westminster, UK Universiti Teknologi MARA, Malaysia Sunway University, Malaysia Universiti Malaysia Pahang, Malaysia Universiti Tenaga Nasional, Malaysia Universiti Teknologi MARA, Malaysia Universiti Teknologi MARA, Malaysia Universiti Utara Malaysia, Malaysia Universiti Malavsia Sarawak, Malavsia Chulalongkorn University, Thailand Universiti Malaysia Sabah, Malaysia Universiti Kebangsaan Malaysia, Malaysia University of Applied Science and Arts Dortmund, Germany Wroclaw University of Science and Technology, Poland Universiti Teknologi MARA, Malaysia Institut Teknologi Sepuluh Nopember, Indonesia G. H. Raisoni College of Engineering Chulalongkorn University, Thailand King Khalid University, KSA Universitas Gadjah Mada, Indonesia Maltepe University, Turkey Universiti Teknologi MARA, Malaysia CDTA, Algeria King Abdulaziz University, KSA Universiti Teknologi MARA, Malaysia Universiti Malaysia Perlis, Malaysia Zagazig University, Egypt Bina Nusantara University, Indonesia

Hasan Kahtan Heri Kuswanto Hizir Sofvan Jasni Mohamad Zain Karim Al-Saedi Khairul Anwar Sedek Khatijah Omar Khyrina Airin Fariza Abu Samah Ng Kok Haur Krung Sinapiromsaran Ku Ruhana Ku-Mahamud Lee Chinh Marina Yusoff Mario Koeppen Marshima Mohd Rosli Maslina Abdul Aziz Michael W. Berry Min Chen Mohamad Johari Ibrahim Mohamed Imran Mohamed Ariff Mohd Sulaiman Muhammad Ahsan Muhammad Firdaus Mustapha Muhammad Izzad Ramli Muhammad Svafig Mohd Pozi Mustafa Abuzaraida Nasiroh Omar Norvanti Muhammad Noor Asiah Ramli Noor Azilah Muda Nooritawati Md Tahir Norazliani Md Lazam Norhaslinda Kamaruddin Noriko Etani Norzehan Sakamat Nurazzah Abd Rahman Nurin Asmuni Omar Moussaoui Pakawan Pugsee

University of Malava, Malavsia Institut Teknologi Sepuluh Nopember, Indonesia Syiah Kuala University, Indonesia Universiti Teknologi MARA, Malaysia Mustansiriyah University, Iraq Universiti Teknologi MARA, Malaysia Universiti Malaysia Terengganu, Malaysia Universiti Teknologi MARA, Malaysia Universiti Malaya, Malaysia Chulalongkorn University, Thailand Universiti Utara Malaysia, Malaysia Universiti Tunku Abdul Rahman, Malaysia Universiti Teknologi MARA, Malaysia Kyushu Institute of Technology, Japan Universiti Teknologi MARA, Malaysia Universiti Teknologi MARA, Malavsia University of Tennessee, USA Oxford University, UK Universiti Teknologi MARA, Malaysia Universiti Teknologi MARA, Malaysia Universiti Teknologi MARA, Malaysia Institut Teknologi Sepuluh Nopember, Indonesia Universiti Teknologi MARA, Malaysia Universiti Teknologi MARA, Malaysia Universiti Utara Malaysia, Malaysia Universiti Utara Malaysia, Malaysia Universiti Teknologi MARA, Malaysia Universiti Malaysia, Pahang Universiti Teknologi MARA, Malaysia Universiti Teknikal Malaysia Melaka Universiti Teknologi MARA, Malaysia Universiti Teknologi MARA, Malaysia Universiti Teknologi MARA, Malaysia All Nippon Airways Co., Ltd., Japan Universiti Teknologi MARA, Malaysia Universiti Teknologi MARA, Malaysia Universiti Teknologi MARA, Malaysia Université Mohammed Premier, Morocco Chulalongkorn University, Thailand

Peraphon Sophatsathit Rajalida Lipikorn Raseeda Hamzah Ravner Alfred Retantyo Wardovo **Richard Millham** Rizauddin Saian Rizwan Aslam Butt Rodrigo Campos Bortoletto Rohit Gupta Roselina Sallehuddin Saiful Akbar Sakhinah Bakar Sayang Mohd Deni Ong Seng Huat Shukor Sanim Mohd Fauzi Leong Siow Hoo Siti Sakira Kamaruddin Sofianita Mutalib Sri Hartati Suraya Masrom Syaripah Ruzaini Syed Aris Syazreen Niza Shair Tahir Ahmad Taiul Rosli Razak Tri Priyambodo Wahyu Wibowo Waidah Ismail Wan Fairos Wan Yaacob Weng Siew Lam Xiaolong Jin Yasue Mitsukura Yap Bee Wah Yusuke Nojima Zainab Othman Zainura Idrus

Chulalongkorn University, Thailand Chulalongkorn University, Thailand Universiti Teknologi MARA, Malaysia Universiti Malaysia Sabah, Malaysia Universitas Gadjah Mada, Indonesia Durban University of Technology, South Africa Universiti Teknologi MARA, Malaysia NED University of Engineering and Technology, Pakistan Instituto Federal de São Paulo, Brazil IIT Delhi, India Universiti Technologi Malaysia, Malaysia Institut Teknologi Bandung, Indonesia Universiti Kebangsaan Malaysia, Malaysia Universiti Teknologi MARA, Malaysia Universiti Malaya, Malaysia Universiti Teknologi MARA, Malaysia Universiti Teknologi MARA, Malaysia Universiti Utara Malaysia, Malaysia Universiti Teknologi MARA, Malaysia Universitas Gadjah Mada, Indonesia Universiti Teknologi MARA, Malaysia Universiti Teknologi MARA, Malaysia Universiti Teknologi MARA, Malaysia Universiti Teknologi Malaysia, Malaysia Universiti Teknologi MARA, Malaysia Universitas Gadjah Mada, Indonesia Institut Teknologi Sepuluh Nopember, Indonesia Universiti Sains Islam Malaysia Universiti Teknologi MARA, Malaysia Universiti Tunku Abdul Rahman, Malaysia Institute of Computing Technology, China Keio University, Japan Universiti Teknologi MARA, Malaysia Osaka Prefecture University, Japan University of Basrah, Iraq Universiti Teknologi MARA, Malaysia

Acknowledgements

Organized by





UNIVERSITI TEKNOLOGI MARA



UNIVERSITI Research TEKNOLOGI Nexus MARA UITM

Hosted by





In co-operation with













THE UNIVERSITY OF









Technical Co-Sponsor



Supported by





