UPM
UNIVERSITI PUTRA MALAYSIA
BERILMU BERBAKTI

SEMINAR KEBANGSAAN ISM-XVIII DAN MESYUARAT AGUNG 2024

Memacu Pembangunan Industri dan Masyarakat melalui Pemerkasaan Data Sains dan Data Analitik

07 MAC 2024
8.00 am - 5.00 pm
BILIK SEMINAR AL-FARABI
INSPEM, UPM

BUKU PROGRAM & ABSTRAK

UniPutraMalaysia @uputramalaysia uniputramalaysia Putra TV uniputramalaysia

PERTANIAN • INOVASI • KEHIDUPAN

BERILMU BERBAKTI
WITH KNOWLEDGE WE SERVE

www.upm.edu.my

ISI KANDUNGAN

UCAPAN PENGARAH INSPEM

اَلسَّلَامُ عَلَيْكُمْ وَرَحْمَةُ اللهِ وَبَرَكَاتُهُ dan Salam Sejahtera.

Alhamdulillah, bersyukur kita ke hadrat Ilahi kerana dengan kurniaNya, Seminar Kebangsaan Institut Statistik Malaysia kali ke-18 (SKISM XVIII) dapat diadakan seperti yang telah dirancang. Bagi pihak Universiti Putra Malaysia (UPM), saya mengalu-alukan kehadiran kesemua pembentang dan para peserta ke kampus hijau ini.

Adalah difahamkan bahawa seminar anjuran bersama INSPEM, UPM dan Institut Statistik Malaysia (ISMy) ini merupakan aktiviti tahunan Institut Statistik Malaysia. Penganjuran bersama ini akan memberi faedah dan kebaikan kepada semua pihak, khususnya kepada semua pengamal statistik. Sesuai dengan tema yang dipilih "Memacu Pembangunan Industri dan Masyarakat melalui Pemerkasaan Data Sains dan Data Analitik", seminar ini bertujuan untuk memperkasa data sains melalui statistik kerana salah satu elemen penting dalam data sains adalah analisis data yang memerlukan pengetahuan statistik. Ini menunjukkan betapa pentingnya kita memantapkan aplikasi statistik dalam industri dan seterusnya menyumbang kepada masyarakat.

Dengan itu, saya amat berharap agar semua ilmu dan maklumat yang diperoleh daripada seminar ini dapat dimanfaatkan oleh para peserta SKISM XVIII. Akhir sekali, saya merakamkan setinggi-tinggi tahniah kepada ahli Jawatankuasa SKISMXVIII atas segala komitmen dan kerjasama sehingga seminar pada kali ini menjadi suatu kenyataan.

Sekian, terima kasih.

Wassalam.

PROF. IR. DR. ADUWATI SALI
Pengarah
Institut Penyelidikan Matematik (INSPEM)
Universiti Putra Malaysia

UCAPAN PRESIDEN ISMY

اَلسَّلَامُ عَلَيْكُمْ وَرَحْمَةُ اللهِ وَبَرَكَاتُهُ dan Salam Sejahtera.

Saya dengan sukacitanya mengalu-alukan kehadiran para pembentang dan peserta Seminar Kebangsaan Institut Statistik Malaysia (ISMy) ke-18 (SKISM XVIII) yang buat kali keenam dianjurkan secara bersama di antara ISMy dan Universiti Putra Malaysia (UPM). Saya amat berbangga dengan budaya kerjasama seumpama ini yang telah dipupuk sejak sekian lama dan masih terus menjadi amalan para ilmuan statistik di Malaysia. Saya juga ingin mengucapkan tahniah kepada semua ahli Jawatankuasa penganjur UPM dan semua pihak yang terlibat dalam menjayakan seminar pada kali ini. Saya juga mengambil peluang ini untuk menyampaikan setinggi penghargaan kepada YBhg. Prof. Dr. Aduwati Sali, Pengarah Institut Penyelidikan Matematik, UPM yang telah memberikan sokongan padu dengan penuh ikhlas dalam penganjuran bersama seminar ini.

Penglibatan dan penyertaan ahli akademik, pengamal statistik dan pelajar di dalam seminar ini adalah sebagai landasan pengetahuan dan pertemuan idea bagi membudayakan pemikiran statistik. Pembudayaan yang bercambah pada seminar kali ini akan dapat dimanfaatkan secara efektif dalam aplikasi pemikiran statistik dalam industri dan masyarakat untuk kegunaan di masa hadapan.

ISMy sentiasa berusaha untuk mengadakan aktiviti yang memberi faedah dan kebaikan kepada semua pengamal statistik. Justeru, penganjuran seminar kali ini dapat menunjukkan kesefahaman di antara semua ahli dalam membulatkan tekad dan melakukan rejuvenasi bagi mengangkat bidang statistik di negara ini. Saya amat berharap agar semua ahli ISMy akan sentiasa memberi sokongan padu yang berterusan bagi memastikan aktiviti yang dirancang berjalan dengan lancar dan mencapai objektif yang ditetapkan. Kerjasama dan komitmen semua pihak secara berterusan dapat merealisasikan objektif penubuhan ISMy.

Sekian, terima kasih.

Wassalam.

PROF. DR. HABSHAH MIDI
Presiden, Institut Statistik Malaysia 2022/2024

UCAPAN PENGERUSI SKISM

اَلسَّلَامُ عَلَيْكُمْ وَرَحْمَةُ اللهِ وَبَرَكَاتُهُ dan Salam Sejahtera.

Bagi pihak ahli jawatankuasa penganjur, saya mengalu-alukan kehadiran para pembentang dan peserta ke majlis tahunan Seminar Kebangsaan ISMy ke 18 (SKISM XVIII) ISM bagi tahun 2024, anjuran bersama Institut Penyelidikan Matematik, UPM dan Institut Statistik Malaysia (ISMy). Saya ingin merakamkan ucapan penghargaan kepada semua pihak yang terlibat terutamanya ahli Jawatankuasa yang telah bertungkus lumus bagi memastikan kelancaran seminar pada kali ini.

Bertemakan "Memacu Pembangunan Industri dan Masyarakat melalui Pemerkasaan Data Sains dan Data Analitik", saya mengharapkan akan adanya percambahan ide baharu di kalangan semua peserta seminar untuk menerokai penggunaan sains statistik dan aplikasinya dalam pelbagai sektor industri di negara ini. Penglibatan beberapa wakil dari institusi pengajian tinggi dan industri pada seminar kali ini juga dapat menunjukkan keperluan dan kepentingan statistik dan aplikasinya.

Akhir kata, saya mengucapkan terima kasih yang tidak terhingga kepada pihak ISMy di atas kepercayaan yang diberikan sebagai penganjur bersama. Segala sumbangan dan sokongan daripada semua pihak amatlah dihargai. Semoga kita semua mendapat manfaat daripada seminar ini.

Sekian, selamat berseminar dan terima kasih.

Wassalam

PROF. DR. HABSHAH MIDI
Pengerusi
Seminar Kebangsaan ISMy ke-18

JAWATANKUASA SKISM 18

Penaung	:	Dato' Prof. Dr. Ahmad Farhan Mohd Sadullah (Naib Canselor, UPM)
Penasihat	:	Prof. Ir. Dr. Aduwati Sali
Pengerusi	:	Prof. Dr. Habshah Midi
Setiausaha	:	Pn. Iszuanie Syafidza Hj. Che Ilias
Bendahari	:	En. Mohamad Asyraf Arshad Pn. Noor Suhana Mohd Azahari
Saintifik	:	Prof. Madya Dr. Jayanti Arasan Dr. Farid Zamani Che Rose Dr. Wendy Lim Shin Yie Dr. Hani Syahida Zulkafli
Buku Program, Abstrak & Penerbitan	:	En. Mohamed Faris Hj. Laham Cik Rabiatul Adawiyah Rosli
Pendaftaran & Protokol	:	Pn. Zurita Ismail Pn. Nor Hasmimi Baharudin
Hebahan & Laman Sesawang	:	En. Zahari Mahad Pn. Nazirah Mahat En. Shahrul Azzim Abd. Aziz
Makanan & Lokasi	:	Pn. Nur Raidah Salim Pn. Syarifah Hasanah Syed Kamaruddin
Sijil & Cenderahati	:	Dr. Malathi Letchumanan Cik Norhashila Mod Hashim
Logistik & Teknikal	:	Pn. Zahratun Nur Yosminar En. Muhammad Feziulnida Abdul Manan En. Mohd Fayzal Jaafar

LEMBAGA AHLI MAJLIS ISMY 2022-2024

Presiden : Prof. Dr. Habshah Midi (UPM)
Timbalan Presiden : Prof. Dr. Ibrahim Mohamed (UM)
Setiausaha Kehormat : Dr. Nur Haizum Abd Rahman (UMPSA)
Bendahari Kehormat : Prof. Madya Dr. Wan Zawiah Wan Zin (UKM)

Ahli Majlis:

i) Prof. Madya Dr. Arifah Bahar (UTM)
ii) Prof. Gs. Ts. Dr. Firdaus Mohamad Hamzah (UPNM)
iii) Dr. Adzhar Rambli (UiTM)
iv) Dr. Mohd Iqbal Shamsudheen (UPNM)
v) Dr. Norli Anida Abdullah (UM)
vi) Dr. Norshahida Shaadan (UiTM)

Juruaudit (2 jawatan) dan Pegawai Pilihanraya ISMy telah dilantik iaitu:

Juruaudit 1 : Prof. Madya Dr. Shariffah Suhaila Syed Jamaludin
Juruaudit 2 : Dr. Zalina Mohd Ali

Pegawai Pilihanraya : Prof. Madya Dr. Norhaiza Ahmad

Lantikan tanpa pilihanraya:

Penasihat ISMy: Prof. Dr. Azami Zaharim

Ahli Majlis Tetap ISMy (dengan jemputan):

a) Jabatan Perangkaan Malaysia
i) En. Mohd Yazid Kasim
ii) En. Razaman Ridzuan

b) Bank Negara Malaysia
i) Pn. Zima Mazfahani Mazlan
ii) Pn. Ong Li Meng

PENGENALAN

Seminar Kebangsaan ISM (SKISM) serta Mesyuarat Agong ISM adalah merupakan program tahunan. Dengan bertemakan "Memacu Pembangunan Industri dan Masyarakat melalui Pemerkasaan Data Sains dan Data Analitik", program ini menjadi platfom bagi ahli statistik bertemu dan bertukar idea. Sains Statistik memainkan peranan yang penting dalam memacu pembangunan dan pertumbuhan ekonomi negara kerana aplikasinya yang meluas dalam pelbagai bidang.

SKism sentiasa berusaha untuk menghimpunkan para penyelidik dari pelbagai bidang statistik di satu platform yang sama supaya perkongsian ilmu dan pengalaman dapat dilaksanakan dengan lancarnya. Ini juga akan memberi pendedahan dan pengalaman kepada ramai penyelidik yang baharu berkecimpung dalam bidang statistik.

Faedah menyertai SKISM XVIII:

- Memahami peranan dan cabaran Sains Statistik di era milenium;
- Meneroka peluang kolaborasi penyelidikan di antara penganjur dan panel jemputan;
- Memupuk kerjasama dalam bidang Sains Statistik melalui perkongsian idea dan perbincangan.

ATUR CARA

Masa	Program	Tempat
08:00 - 08:30	Pendaftaran dan Ketibaan Tetamu Jemputan	Foyer @ INSPEM
08:30 - 09:30	Sesi Selari 1A Sesi Selari 1B Sesi Selari 1C	Bilik Seminar Bilik Serbaguna Bilik Mesyuarat
09:40 - 09:55	**Perasmian SKISMXVIII** Nyanyian Lagu Negaraku dan Putra Gemilang Bacaan Doa Ucapan Presiden ISMy - **YBhg. Prof. Dr. Habshah Midi** Ucapan Perasmian oleh Pengarah INSPEM, UPM - **YBhg. Prof. Ir. Dr. Aduwati Sali**	Bilik Seminar Al-Farabi
09:55 - 10:00	Sesi Bergambar	Bilik Seminar Al-Farabi
10:00 - 10:20	Kudapan Pagi	Foyer @ INSPEM
10:20 - 11:00	Ucaptama I Penceramah : **En Mohd Hafeez Nazri** *General Manager Network Analytics,* *Telekom Malaysia Berhad* **DATA SCIENCE IN THE INDUSTRY: APPLICATION & CHALLENGES** Pengerusi: Prof. Madya Dr. Jayanthi Arasan	Bilik Seminar Al-Farabi
11:00 - 11:40	Sesi Selari 2A Sesi Selari 2B Sesi Selari 2C	Bilik Seminar Bilik Serbaguna Bilik Mesyuarat
11:40 - 13:00	Mesyuarat Agung ISMY 2024	Bilik Mesyuarat Radin Umar
13:00 - 14:20	Makan Tengah Hari dan Rehat	Foyer @ INSPEM
14:20 - 15:00	Ucaptama II Penceramah : **Prof. Dr. Habshah Midi, UPM** *Presiden, Institut Statistik Malaysia* **REMEDIAL MEASURES FOR SOME COMMON MISCONCEPTION OF STATISTICAL TECHNIQUES IN DATA ANALYSIS** Pengerusi: Prof. Dr. Ibrahim Mohamad, UM	Bilik Seminar Al-Farabi
15:00 - 15:40	Sesi Selari 3A Sesi Selari 3B Sesi Selari 3C	Bilik Seminar Bilik Serbaguna Bilik Mesyuarat
15:40 - 17:00	Anugerah ISMy dan Penamaan Ahli Kehormat	Bilik Seminar Al-Farabi
17:00	Minum Petang dan Bersurai	Foyer @ INSPEM

UCAPTAMA

PENCERAMAH UCAPTAMA I

En Mohd Hafeez Nazri
General Manager Network Analytics,
Telekom Malaysia Berhad

DATA SCIENCE IN THE INDUSTRY: APPLICATION & CHALLENGES

Hafeez Nazri currently holds the position of General Manager Network Analytics at Telekom Malaysia Berhad. Before this role, he led the Data Science team at Touch N' Go and Media Prima Berhad, where he successfully delivered multiple Data Science and Machine Learning projects to transform data into valuable business insights. He served as the Lead Data Scientist for the CoronaTracker project, which gained global recognition with over 40 million pageviews. The sharing will summarize the essence of data science in industry, how companies leverages data to make decisions in business. Also, the sharing will highlight the challenges that data scientists encounter in producing outcomes that will help organisations.

PENCERAMAH UCAPTAMA II

Prof. Dr. Habshah Midi
Presiden,
Institut Statistik Malaysia 2022/2024
Felo Perunding,
Institut Penyelidikan Matematik,
Universiti Putra Malaysia

REMEDIAL MEASURES FOR SOME COMMON MISCONCEPTION OF STATISTICAL TECHNIQUES IN DATA ANALYSIS

In today's society, statistical techniques are being used with increasing rate in education, medicine, social sciences, and applied sciences such as engineering. They are crucial in interpreting data and making decisions. Based on my experience and observation through seminars, conferences and consultations, I noticed some statistics practitioners have misconceptions in using some of the statistical techniques in their researches. The easy availability of the statistical packages such as SAS, SPSS, has driven more researchers who have no sound knowledge in statistics, to use the packages in analysing their data. However, many are not aware that packages do not know the correct statistical technique to use. They just follow instruction of researchers and produce results accordingly. This led to greater abuse of statistics in data analysis. Consequently, less accurate results and misleading conclusions are obtained from an incorrect analysis. Therefore, it is imperative for the researchers to be aware of using the right statistical techniques so that a valid and objective conclusion can be made. In this presentation, some incorrect practices on some selected topics will be discussed. Appropriate suggestions are offered to tackle this problem.

JADUAL PEMBENTANGAN
SESI SELARI

SESI SELARI 1
08:30 - 09:30

Masa	Sesi Selari 1A Bilik Seminar **Pengerusi :** *Dr. Shazlyn Milleana Shaharudin*	Sesi Selari 1B Bilik Serbaguna **Pengerusi :** *Dr. Sahimel Azwal Sulaiman*	Sesi Selari 1C Bilik Mesyuarat **Pengerusi :** *Dr. Ummu Atiqah Mohd Roslan*
08:30 - 08:40	**BACKWARD ELIMINATION TECHNIQUE IN MULTIVARIATE TIME SERIES MODELLING OF CO2 EMISSIONS IN MALAYSIA** Mahayaudin M. Mansor Muhammad Ehsan M. Sukury Nur Alya Adlina M. Nasir Ath Thabrani M. Fauzi Norshahida Shaadan Shariffah Suhaila Syed Jamaludin	**OPTIMIZING IMAGE FEATURE SELECTION METHOD THROUGH THE INTEGRATION OF BIO-INSPIRED ALGORITHMS WITH META-HEURISTIC SEARCH FOR COVID-19 CLASSIFICATION PROBLEM** Mohammad Aizat Basir Mohamed Saifullah Hussin Mohd Helmy Abd Wahab	**PATENT TREND ANALYSIS IN ADVANCED CONSTRUCTION MATERIALS FOR CARBON EMISSION REDUCTION: A TOPIC MODELING APPROACH** Hafizah Farhah Saipan Saipol Syarifah Zyurina Nordin Aizul Nahar Harun Amir Syafiq Syamin Shah Amir Hamzah
08:40 - 08:50	**HYBRID ARIMA AND LS-SVM FOR BETTER SEA SURFACE TEMPERATURE (SST) FORECASTS FOR THE EAST COAST OF PENINSULA MALAYSIA** Muhamad Safiih Lola Siti Hajar Yusof Wan Imanul Aisyah Wan Mohamad Nawi Mohd Tajuddin Abdullah	**ENSEMBLE IMAGE FEATURE SELECTION METHOD BASED ON BIO-INSPIRED ALGORITHMS FOR COVID-19 CLASSIFICATION PROBLEM** Mohammad Aizat Basir Mohamed Saifullah Hussin Mohd Helmy Abd Wahab	**USING A PANEL DATA MODEL TO I NVESTIGATE THE RELATIONSHIP BE-TWEEN STOCK MARKET AND MACROECONOMIC VARIABLES IN THE PRESENCE OF STRUCTURAL BREAKS: ASEAN-5 COUNTRIES** Diana Hassan Assis Kamu Ricky Chia Chee Jiun Ho Chong Mun
08:50 - 09:00	**ENHANCING ASSET SECURITY IN MALAYSIA: A MULTIVARIATE REGRESSION AND TIME SERIES ANALYSIS APPROACH** Sahimel Azwal Sulaiman Nurul Azian Mohd Basharudin Hafiz Asyraaf Ghani Mohammad Zaharudin Ahmad Darus	**A COMPARATIVE MACHINE LEARNING BY USING K-FOLD VALIDATION OF ACUTE KIDNEY INJURY WITH LIMITED LEARNING DATA** Mohd Noor Azam Nafi Baitul Husna Salihin Siti Nurhafizah Mohd Shafie Nasuhar Ab. Aziz Mardhiah Kamarudin Azzah Amran	**INTEGRATING AMOS AND SMARTPLS 4 IN CB-SEM: A COMPARATIVE STUDY ON HIGHER ORDER CONSTRUCT MODELING** Nurul Raudhah Zulkifli (pelajar) Shahazwan Mat Yusoff
09:00 - 09:10	**FORECASTING LOCATIONS OF FOREST FIRES IN INDONESIA THROUGH NONPARAMETRIC PREDICTIVE INFERENCE WITH PARAMETRIC COPULA: A CASE STUDY** Noryanti Muhammad Amirah Hazwani Roslin Evizal Abdul Kadir Warih Maharani Hanita Daud	**BIBLIOMETRIC ANALYSIS ON PREDIC-TIVE ANALYTICS FOR FRESH AGRO-FOOD PRICE USING MACHINE LEARNING** Sahimel Azwal Sulaiman Nor Azuana Ramli	**LOGISTIC REGRESSION MODELLING FOR CONSUMER INTENTION TO PURCHASE PLANT FACTORY PRODUCTS** Aimi Athirah Ahmad Nik Rahimah Nik Omar Nadiah Sa'at Nur Hidayah Md Noh Nurul Syahida Abu Bakar
09:10 - 09:20	**COMPARATIVE ANALYSIS OF RECURRENT AND VECTOR FORECASTING IN SOLAR RADIANCE PREDICTION: A SINGULAR SPECTRUM APPROACH** Shazlyn Milleana Shaharudin Nur Amirah Shafie Mou Leong Tan	**ANALYZING TRENDS AND KEY TERMS IN CONSTRUCTION MATERIALS PATENT DOCUMENTS THROUGH DATA PRE-PROCESSING AND TEXT MINING TECHNIQUES** Syarifah Zyurina Nordin Hafizah Farhah Saipan Saipol Aizul Nahar Harun Amir Syafiq Syamin Shah Amir Hamzah	**BEYOND STATISTICS: USING MATHEMATICAL MODELS TO ASSESS THE IMPACT OF COVID-19 ON EMPLOYMENT, POVERTY, AND INCOME GROUPS IN MALAYSIA** Ummu Atiqah Mohd Roslan

SESI SELARI 2
11:00 - 11:40

Masa	Sesi Selari 2A Bilik Seminar **Pengerusi :** *Dr. Norshahidan Shaadan*	Sesi Selari 2B Bilik Serbaguna **Pengerusi :** *Dr. Mohd Shafie Mustafa*	Sesi Selari 2C Bilik Mesyuarat **Pengerusi :** *Dr. Syaiba Balqish Ariffin*
11:00 - 11:13	**A COMPARISON OF VARIOUS CONTROL CHARTS FOR DETECTING AND MONITORING ABNORMAL RIVER WATER LEVEL** Norshahidan Shaadan Siti Nur Atiqah Mohd Shafie	**ALTERNATIVE MODEL ADEQUANCY AND DIAGNOSTICS FOR PARALLEL EXPONENTIATED EXPONTENTIAL REGRESSION MODEL WITH CENSORED DATA** Jayanthi Arasan Anwar Fitrianto	**LOGISTIC INFLUENTIAL OUTLIER NOMINATOR** Syaiba Balqish Ariffin Habshah Midi
11:13 - 11:33	**CLIMATE CHANGE DETECTION AND INVESTIGATION TOWARDS GREEN ENVIRONMENT SUSTAINABILITY IN IPOH, KUALA KRAI , MERSING AND TERMERLOH PENINSULAR MALAYSIA** Norshahida Shaadan Noor Kesuma Mohd Yazid Firdaus Mohammad Hamzah Mahayaudin M. Mansor Nurain Ibrahim	**A COMPARISON OF IMPUTATION METHOD FOR A BATHTUB HAZARD MODEL WITH RIGHT AND INTERVAL-CENSORED DATA** Idari Ismail (Pelajar) Jayanthi Arasan Mohd Shafie Mustafa Muhammad Aslam Mohd Safari	**A DIAGNOSTIC METHOD OF HIGH LEVERAGE POINTS BASED ON IMPROVISED ROBUST PRINCIPAL COMPONENT ANALYSIS BASED ON MINIMUM REGULARIZED COVARIANCE DETERMINANT IN HIGH DIMENSIONAL DATA** Habshah Midi Jaaz Suhaiza Jaafar
11:33 - 11:40	**STATISTICAL EXTREME VALUE MODELLIING OF ENVIRONMENTAL DATA** Syafrina Abdul Halim Nina Allyza Kepol Xue Haoying	**IMPROVEMENT OF INITIAL SOLUTIONS FOR METAHEURISTICS USING SIMULATED ANNEALING** Mohamed Saifullah Hussin Mohammad Aizat Basir	**COMPARATIVE STUDY OF CLUSTERING LINKAGE METHODS USING ROBUST DISTANCE AS SIMILARITY MEASURE IN DETETCTING MULTIVARIATE OUTLIERS** Syarifah Sakinah Syed Abd Mutalib Siti Zanariah Satari Wan Nur Syahidah Wan Yusoff

SESI SELARI 3
15:00 - 15:40

Masa	Sesi Selari 3A Bilik Seminar **Pengerusi :** *Dr. Muhamad Safiih Lola*	Sesi Selari 3B Bilik Serbaguna **Pengerusi :** *Dr. Nor Azuana Ramli*	Sesi Selari 3C Bilik Mesyuarat **Pengerusi :** *Dr. Noryanti Muhammad*
15:00 - 15:12	**ENHANCING BEV CHARGING PREDICTION USING LSTM NETWORKS ON FEATURE-ENGINEERED 1-MINUTE RESOLUTION START-STOP CHARGING DATA** Roslinazairimah Zakaria Syahrizal Salleh, Siti Roslindar Yaziz	**EXPLORING THE PATTERN OF HOUSEHOLD MONTHLY INCOME & EXPENDITURES BY STATE IN MALAYSIA: STATISTICAL APPROACH.** Syerrina Zakaria Nor Fatimah Che Sulaiman Siti Madhihah Abdul Malik	**STATISTICAL ANALYSIS ON CYBER ATTACK EVENTS THAT RELATED TO MALAYSIA** Hafiz Asyraaf Ghani
15:12 - 15:24	**CONSUMER ADOPTION AND PERCEPTIONS OF ELECTRIC VEHICLES IN MALAYSIA USING SENTIMENT ANALYTICS** Nur Haizum Abd Rahman Nor Farawahida Abdullah	**DATA LAKEHOUSE ARCHITECTURE FOR SELF-SERVICE DATA ANALYTICS** Nor Azuana Ramli Soon Kien Yuan Mohd Zaid Waqiyuddin Mohd Zulkifli	**DEVELOPING A NEW FEATURE FOR VULNERABILITY RISK SCORING MODEL FOR ENHANCED CYBERSECURITY** Noryanti Muhammad Lim Joey
15:24 - 15:36	**A SIMULATION BASED OPTIMIZATION APPROACH IN DETERMINE LOYALTY REWARDS POINT VALUE FOR LOYALTY PROGRAM** Wan Nuraini Fahana Wan Nasir Muhamad Safiih Lola	**FACTOR ANALYSIS OF INTERNET USE – A CASE STUDY AMONG UNIVERSITI MALAYSIA TERENGGANU STUDENTS** Siti Madhihah Abd Malik Muhamad Safiih Lola	**SCALABILITY AND PERFORMANCE IN DUPLICATE DETECTION: RELATIONAL VS. GRAPH DATABASE** Noryanti Muhammad Muhammad Farhad Khaharruddin Mohd Izhar Firdaus

ABSTRAK
SESI SELARI
1A, 1B DAN 1C

08:30 - 09:30

SESI SELARI 1A

Backward Elimination Technique in Multivariate Time Series Modelling of CO_2 Emissions in Malaysia

Muhammad Ehsan M. Sukury[1, b)], Nur Alya Adlina M. Nasir[1, c)], Ath Thabrani M. Fauzi[1, d)], Norshahida Shaadan[1, e)], Shariffah Suhaila Syed Jamaludin[2, f)] and Mahayaudin M. Mansor[1, a)]

[1]*School of Mathematical Science, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia.*
[2]*Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia.*

[*]Corresponding author: maha@uitm.edu.my

Abstract. It is well noted that the concentration of carbon dioxide (CO_2) in the atmosphere has increased since the Industrial Revolution, and CO_2 is the main greenhouse gas contributor to climate change. The negative consequences of climate change include extreme heat waves, intense rainfall, and droughts that endanger human life; affecting human safety, health, and food security. In this study, the annual CO_2 emissions in Malaysia over 33 years beginning in 1990 are plotted against time aiming at identifying any time series patterns for modelling purposes. The time series plot shows an upward trend that is trending at an alarming rate and is expected to continue in the future. Socio-economic and environmental factors are examined in this study to describe the variability the CO_2 emissions. A well-specified model for the response variable is achieved via a backward elimination procedure guided by the modelling diagnostics for multicollinearity, goodness-of-fit that includes checking the estimated symbol of the fitted coefficient against the priori, and the underlying assumptions of normality, homoscedasticity, and the absence of autocorrelation for the residuals. A Variance Inflation Factor (VIF) table was used to show the modelling transition where the procedure eliminated the total population, annual mean temperature, and urban population to obtain the well-specified model. An additional unit of fossil fuel consumption and Gross Domestic Product (GDP) per capita increases the CO_2 emissions in Malaysia by 164,169 and 5,910 tonnes, while all other factors are held constant, respectively. GDP indicates the general progress of the economy and has been linked to industrialization. Hence, we recommend that the government intensify its commitment towards developing facilities to facilitate the use of renewable or green energy, that is mostly unharmful to the atmosphere, as an alternative source to fossil fuel to reduce CO_2 emissions in the atmosphere.

Keywords: Time series, backward elimination, carbon emission, climate change, SDG 13

HYBRID ARIMA AND LS-SVM FOR BETTER SEA SURFACE TEMPERATURE (SST) FORECASTS FOR THE EAST COAST OF PENINSULA MALAYSIA

Siti Hajar Yusof[1,a)], Muhamad Safiih Lola[2,b)*], Wan Imanul Aisyah Wan Mohamad Nawi[3,c)], Mohd Tajuddin Abdullah[4,d)]

[1,2,3,a)] *Faculty of Computer Science and Mathematics, Universiti Malaysia Terengganu, Kuala Nerus 21030, Terengganu, Malaysia*
[4,d)] *Fellow Academy of Sciences Malaysia, Level 20, West Wing Tingkat 20, Menara MATRADE, Jalan Sultan Haji Ahmad Shah, Kuala Lumpur 50480, Malaysia*

*****Corresponding author: safiihmd@umt.edu.my

Abstract. The sea surface temperature (SST) is an environmental indicator due to its strong relationship to global climate change, weather, pollution, productivity, and atmospheric events worldwide. The increasing of ocean heat content will contribute to sea level rise and thinning of ice shelves and sea ice. This situation may endanger the health of corals due to lack of oxygen in a deep water. Therefore, prediction of SST can be a great help for decision making in many operational applications. Thus, this study aims to predict the SST based on a hybrid ARIMA and Least Square-Support Vector Machines (LS-SVMs) models. Besides that, this study evaluates the performance of SST prediction model based of that hybrid models using 1452 monthly SST datasets for the East Coast of Peninsular Malaysia. The empirical results revealed that the performance of proposed model generated compared with ARIMA, SVMs and hybrid ARIMA LS-SVMs models, the proposed models generate smaller values of MSE, RMSE, MAE, and MAPE for both training and testing datasets. In other words, the proposed hybrid models are better than those that we compare with. Their forecasting values are closer to the actual values. Thus, we conclude that the hybrid models can be used to generate better forecasting values with higher degree of accuracy, efficiency and, precision in forecasting time series results becomes a priority.

Keywords: Forecasting, Hybrid ARIMA LS-SVMs, Sea Surface Temperature, Accuracy and Efficiency, climate change.

SESI SELARI 1A

Enhancing Asset Security in Malaysia: A Multivariate Regression and Time Series Analysis Approach

Nurul Azian Mohd Basharudin[1, b)], Hafiz Asyraaf Ghani[1, c)], Mohammad Zaharudin Ahmad Darus[2, d)] and Sahimel Azwal bin Sulaiman[1, a)]

[1]*Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuh Persiaran Tun Khalil Yaakob, 26300 Gambang, Pahang, Malaysia.*
[2]*Level 7 Tower 1, Menara Cyber Axis, Jalan Impact, 63000 Cyberjaya, Selangor Malaysia.*

[a)]Corresponding author: sahimel@umpsa.edu.my

Abstract. This research is conducted to gain insight by fully utilizing multivariate, simple exponential smoothing (SES), and influence diagrams. The primary aim of this research is to utilize multivariate linear regression to analyze the relationship between independent and dependent variables and discern the model that best fits the data. Furthermore, this study also incorporates time series analysis employing SES to forecast the quantity of assets exposed at various risk levels (Low, Medium, High, Critical) over the subsequent four days, with the overarching goal of enhancing security measures and addressing cyber threats. In addition, an influence diagram is used as a visual aid for decision analysis and probabilistic modelling, clarifying the complex interrelationships between variables, decisions, uncertainties, and objectives relevant to decision-making situations. Data for this study is sourced from the CrowdStrike application, which provides comprehensive insights into an organization's external assets categorized by risk levels. The findings are evaluated using key metrics, including Mean Absolute Error, Root Mean Square Error, and Mean Absolute Percentage Error. As a result, this research has the potential to significantly improve asset security strategies and risk management practices, making organizations safer. This will enhance the security of external assets and help achieve the goals outlined in RMK12, which include strengthening the security of national assets and promoting technological modernization in Malaysia.

Keywords: Multivariate Analysis, Simple Exponential Smoothing, Influence Diagram, Time Series Analysis, Forecasting

SESI SELARI 1A

Forecasting Locations of Forest Fires in Indonesia Through Nonparametric Predictive Inference with Parametric Copula: A Case Study

Amirah Hazwani Roslin[1, a)] , Noryanti Muhammad[1, b),*], Evizal Abdul Kadir[2, c)], Warih Maharani[3, d)] & Hanita Daud[4, e)]

[1]*Centre for Mathematical Sciences, Universiti Malaysia Pahang, Lebuh Persiaran Tun Khalil Yaakob, 26300, Kuantan, Pahang, Malaysia.*
[2]*Department of Informatics Engineering, Universitas Islam Riau, Pekanbaru, 28284, Indonesia.*
[3]*School of Computing Telkom University, Bandung, 40257, Indonesia.*
[4]*Department of Applied Mathematics, Universiti Teknologi Petronas, 32610, Perak, Malaysia.*

*Corresponding author: noryanti@umpsa.edu.my

Abstract. Forest fire causes major ecological, economic, and anthropological damage to a country. These problems incurred high redevelopment and restoration costs, a huge loss. Despite numerous predictive studies in this field, there were still uncertainties in forest fire management, mainly in small fires. The machine learning (ML) technique was the most popular choice among researchers on this topic, but it portrayed some gaps. The common issues in ML approaches were non-generalizable and exposed to inaccurate possibilities. Hence, this study intended to apply nonparametric predictive inference (NPI) with a parametric copula to predict the next forest fire location. Also, it aims to introduce new parameters for forest fire prediction – longitude and latitude. The NPI quantifies the uncertainties in imprecise probabilities to provide confidence for the forest fire management plan. Meanwhile, the copula integration considers the dependence structure between the past coordinates in predicting the next forest fire location. The suggested method was hypothesized to enhance the prediction accuracy and promote generalizability in other countries with the same parameters. For this study, the NPI with the parametric copula method was implemented in Indonesia's forest fire datasets due to its severe forest fire history in Southeast Asia. The proposed method was validated using simulation and shows a robust conclusion. Besides, NPI with parametric copula was highly accurate because of the low differences within its imprecise probabilities. Overall, the Indonesia archipelagoes have different optimal copula in predicting the next forest fire location. Clayton and Gumbel are the best copulae to be integrated with NPI to predict the next forest fire location in Sumatra, while Normal copula showcased the smallest deviation for the Kalimantan archipelago forest fire occurrence. Therefore, the NPI is a valid alternative for forest fire location prediction.

Keywords: copula, forest fire location, imprecise probability, Indonesia, nonparametric predictive inference

SESI SELARI 1A

Comparative Analysis of Recurrent and Vector Forecasting in Solar Radiance Prediction: A Singular Spectrum Approach

Shazlyn Milleana binti Shaharudin[1,a,*], Nur Amirah binti Shafie[1, b] , Mou Leong Tan[2,c]

[1]*Department of Mathematics, Faculty of Science and Mathematics,*
Universiti Pendidikan Sultan Idris
Tanjong Malim, Perak, Malaysia
[2]*Geoinformatic Unit, Geography Section, School of Humanities,*
Universiti Sains Malaysia,
Gelugor, Malaysia

*Corresponding author: shazlyn@fsmt.upsi.edu.my

Abstract. The objective of this study is to predict the solar radiance in Langkawi by using Recurrent Forecasting -Singular Spectrum Analysis (RF-SSA) and Vector Forecasting-Singular Spectrum Analysis (VF-SSA). The study also investigates to present the comparison between two forecasting algorithms. The data used in this study is solar data for year 2020. As part of its methodology, this study had used an imputation method for handling missing data. Mean imputation, median imputation and mode imputation are three model used to obtain the missing value. The proposed model was analysed by using the performance measures that are Root Mean Square Error (RMSE). The result of imputation methods shows mode imputation is having the lowest RMSE value. The complete data sets of solar radiation then conducted by using Singular Spectrum Analysis(SSA). In the conclusion of this study, the comparison of results from the RF- SSA and VF-SSA reveals that recurrent forecasting reproduces the most reasonable hourly solar radiation and making it a perfect candidate for solar prediction research. The implication of this study is to establish the best model in Singular Spectrum Analysis(SSA) in order to obtain accuracy predicted value for solar radiance.

Keywords: solar radiance, Singular Spectrum Analysis (SSA), Recurrent Forecasting (RF),
Vector Forecasting (VF), forecasting, imputation

SESI SELARI 1B

Optimizing Image Feature Selection Method through the Integration of Bio-Inspired Algorithms with Meta-Heuristic Search for COVID-19 Classification Problem

Mohammad Aizat Basir[1, a)], Mohamed Saifullah Hussin[1, b)] and Mohd Helmy Abd Wahab[2, c)]

[1]*Faculty of Computer Science and Mathematic*
Universiti Malaysia Terengganu,
21030 Kuala Nerus, Terengganu
Malaysia.

[2]*Department of Electronic Engineering*
Faculty of Electrical and Electronic Engineering
Universiti Tun Hussein Onn Malaysia
86400 Pt Raja, Batu Pahat, Johor
Malaysia.

[a)]Corresponding author: aizat@umt.edu.my
[b)]saifullah@umt.edu.my
[c)]helmy@uthm.edu.my

Abstract. Achieving the selection of optimal image features, particularly for image classification tasks, is a challenging and crucial endeavor. The conventional method of identifying independently functioning image features often results in the selection of unrelated image features, thereby degrading the consistency of classification accuracy. The primary objective of this article is to optimize Meta-heuristic algorithms, specifically Harmony Search (HS) and Tabu Search (TS), by leveraging the capabilities of bio-inspired search algorithms in conjunction with the wrapper. The essential stages involve refining the HS and TS combination with appropriate bio-inspired methods and incorporating the creation of various image feature subsets. Subsequently, a subset evaluation is conducted to confirm the optimum image feature set. The evaluation criteria are based on both the number of image features utilized and the image classification accuracy. For testing purposes, 1000 online images of COVID-19 chest x-ray datasets from three (3) different categories (normal, lung opacity and viral pneumonia) were meticulously chosen. Extensive testing has demonstrated that the optimal combination of the selected bio-inspired algorithm and meta-heuristics algorithms, particularly HS and TS, holds the promise of providing a superior optimum solution. This solution entails fewer image features with greater image classification accuracy for the selected image datasets. Consequently, this research suggests that the ability of bio-inspired algorithms, when used in conjunction with wrapper/filtered methods, enhances the efficiency of HS and TS in selecting and identifying characteristics.

Keywords: bio-inspired, feature, harmony search, tabu search, image classification, COVID-19

SESI SELARI 1B

Ensemble Image Feature Selection Method based on Bio-Inspired Algorithms for COVID-19 Classification Problem

Mohammad Aizat Basir[1,a)], Mohamed Saifullah Hussin[1,b)] and Mohd Helmy Abd Wahab[2,c)]

[1]*Faculty of Computer Science and Mathematic*
Universiti Malaysia Terengganu,
21030 Kuala Nerus, Terengganu
Malaysia.

[2]*Department of Electronic Engineering*
Faculty of Electrical and Electronic Engineering
Universiti Tun Hussein Onn Malaysia
86400 Pt Raja, Batu Pahat, Johor
Malaysia.

[a)]Corresponding author: aizat@umt.edu.my
[b)]saifullah@umt.edu.my
[c)]helmy@uthm.edu.my

Abstract. Feature selection poses a considerable challenge, particularly in achieving an optimal solution. This difficulty arises from the challenge of selecting the most suitable feature selection methods, which often work independently, resulting in the selection of incorrect features and subsequently impacting classification accuracy. The primary objective of this research is to harness the potential of ensemble methods, specifically boosting, in conjunction with bio-inspired techniques to enhance the performance of image feature selection model in terms of the optimum image feature set. A crucial stage in this research involves optimizing both algorithms using appropriate bio-inspired search algorithms combined with ensemble methods. The subsequent step is to validate the optimal selected image feature set by conducting an image classification task. Evaluation metrics are determined based on the number of selected image features exhibiting good image classification accuracy. A selection of 1000 chest x-ray images related to COVID-19, sourced from online datasets, was carefully curated. These images span three distinct categories: normal, lung opacity, and viral pneumonia. Experimental results revealed that both algorithms, when utilizing selected bio-inspired search algorithms with ensemble methods, successfully achieved superior solutions. Specifically, they demonstrated an optimum set of image features, comprising fewer image features with greater image classification accuracy, across the selected COVID-19 datasets. This discovery implies that the combination of bio-inspired algorithms with ensemble methods, particularly boosting, has the potential to enhance the performance of both algorithms for image feature selection and classification tasks.

Keywords: bio-inspired, ensemble, feature, image classification, COVID-19

SESI SELARI 1B

A Comparative Machine Learning by using K-fold Validation of Acute Kidney Injury with Limited Learning Data

Baitul Husna Salihin[1, a)], Mohd Noor Azam Nafi[1, b)], Siti Nurhafizah Mohd Shafie[1, c)], Nasuhar Ab. Aziz [1, d)], Mardhiah Kamarudin[2, e)] and Azzah Amran[3, f)]

[1]*College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Kelantan Branch, Kota Bharu Campus, 15050 Kota Bharu, Kelantan.*
[2]*Faculty of Entrepreneurship and Business, Universiti Malaysia Kelantan, 15050 Kota Bharu, Kelantan.* [3]*Faculty of Business and Management, Universiti Teknologi MARA Kelantan Branch, Kota Bharu Campus, 15050 Kota Bharu, Kelantan.*

[b)]Corresponding author: norazam05@uitm.edu.my [a)]2020456096@student.uitm.edu.my [c)]sitinurhafizah@uitm.edu.my [d)] nasuha978@uitm.edu.my [e)]mardhiah.k@umk.edu.my [f)]azzahamran@uitm.edu.my

Abstract. This study evaluates the machine learning algorithm by using K-fold cross-validation for Acute Kidney Injury from learning data. In order to simulate a lack of training data, the dataset was randomly divided into an 80% training set and a 20% test set. The main objective is to compare the performance for model evaluation, which is Misclassification Rate (MR), Average Square Error (ASE) and Receiver Operating Characteristic (ROC) index. About 121 samples have been used by applying a randomized 5-fold cross validation process to validate the findings. Three machine learning algorithms, which are regression (HP Regression), random forest (HP Tree), and gradient boosting versus the default model in SAS Enterprise Miner, were used to assess the best model. The findings indicate that gradient boosting is the best model since it gives the smallest value for MR and ASE, and the largest value for the ROC index. Additionally, the accuracy, sensitivity, and specificity values from gradient boosting also give the highest value, respectively.

Keywords: Acute Kidney Injury, Machine Learning, Random Forest, Regression.

Bibliometric Analysis on Predictive Analytics for Fresh Agro-Food Price using Machine Learning

Sahimel Azwal Sulaiman[1, a)] and Nor Azuana Ramli[1, b)]

[1]*Centre For Mathematical Sciences,*
Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuh Persiaran Tun Khalil Yaakob, 26300 Gambang, Pahang, Malaysia.

[a)]Corresponding author: sahimel@umpsa.edu.my
[b)]azuana@umpsa.edu.my]

Abstract. Predictive analytics leveraging machine learning techniques has emerged as a powerful tool in forecasting fresh agro-food prices, offering valuable insights for stakeholders in the agricultural and food industries. This paper presents a comprehensive bibliometric analysis aimed at synthesizing the existing literature on the application of machine learning in predicting agro-food prices. This study adopted a bibliometric analysis based on documents obtained from the Scopus database. To achieve the objectives of the study, various tools have been employed, such as VOSviewer, and Microsoft Excel. This study presents the results using standard bibliometric indicators. The analysis of relevant scholarly articles, conference proceedings, and patents, this study identifies trends, patterns, and gaps in research pertaining to predictive analytics in the agro-food domain. Key aspects such as methodologies employed, types of data utilized, performance metrics, and application domains are examined to offer a holistic understanding of the current state of research in this field. Furthermore, this analysis provides insights into the evolution of techniques, interdisciplinary collaborations, and emerging research directions. The findings of this bibliometric study contribute to the advancement of knowledge in predictive analytics for agro-food pricing and offer valuable guidance for future research endeavors and practical applications in the agricultural and food sectors.

Keywords: Predictive Analytics, Machine Learning, Agro-Food Prices, Bibliometric Analysis, Forecasting

SESI SELARI 1B

Analyzing Trends and Key Terms in Construction Materials Patent Documents through Data Pre-Processing and Text Mining Techniques

Syarifah Zyurina Nordin [1, 2, a)], Hafizah Farhah Saipan Saipol[1, 2, b)], Aizul Nahar Harun [1, 2, c)], and Amir Syafiq Syamin Shah Amir Hamzah [1, 2, d)]

[1]*Department of Management of Technology, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, 54100 Kuala Lumpur, Malaysia.*
[2]*Intellectual Property and Innovation Management iKohza, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, 54100 Kuala Lumpur, Malaysia.*

[a)]Corresponding author: szyurina@utm.my
[b)]hafizah.farhah@utm.my
[c)] aizulnahar.kl@utm.my
[d)] syamin@utm.my

Abstract. Analyzing massive patent documents in construction material has become important in recent years as they contain a lot of information that is extremely difficult to deal with from huge numbers and various forms. Important documents such as patent data on a huge scale has become a major concern owing to time constraints and enormous costly work. In this paper, our objectives are to examine the fundamental concept of pre-processing and data analyzing, describe the similarities, distances, and frequencies between several patent documents of the construction material, analyze the patent trends of patent publications using the DWPI abstract in text mining, and find key terms that can differentiate between the two periods of patent application. Our methodology includes pre-processing and analyzing patent data from various construction materials, dividing patents into two time periods, employing text mining techniques, comparing correlations between terms, and using visual representation and quantitative analysis methods. We also conduct a computational experiment on the data patent to analyze the co-occurrence network and correspondence analysis for construction material and visualized it in TF-DF plot from KH Coder.

Keywords: construction material, text mining, data analyzing, patent analysis

SESI SELARI 1C

Patent Trend Analysis in Advanced Construction Materials for Carbon Emission Reduction: A Topic Modeling Approach

Hafizah Farhah Saipan Saipol[1, 2, a)], Syarifah Zyurina Nordin[1, 2, b)], Aizul Nahar Harun [1, 2, c)], and Amir Syafiq Syamin Shah Amir Hamzah [1, 2, d)]

1Department of Management of Technology, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, 54100 Kuala Lumpur, Malaysia.
2Intellectual Property and Innovation Management iKohza, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, 54100 Kuala Lumpur, Malaysia.

a)Corresponding author: hafizah.farhah@utm.my

Abstract. The increasing concern over carbon emissions and climate change has led to great interest in research and innovation for construction materials aimed at reducing environmental impact. This study employs topic modeling technique to analyze patent trends in the field of advanced construction materials focusing on carbon emission reduction. The methodology involves the application of topic modeling algorithms, which is Latent Dirichlet Allocation (LDA), to categorize the latent thematic clusters within the patent corpus. The resulting topics identify key areas of technological advancement, providing insights into emerging trends and innovations. In addition, this study explores the interconnectivity between topics that will help to the development of solutions for carbon emission reduction. The analysis highlights the advanced materials technologies and their development to achieve su tainable solutions. This patent trend analysis can serve as a guide for researchers, and industry professionals and contributes to the understanding of the global efforts in advanced construction materials for carbon emission reduction.

Keywords: advanced material, carbon emission reduction, topic modeling, climate change, patent analysis

Using a Panel Data Model to Investigate the Relationship between Stock Market and Macroeconomic Variables in the Presence of Structural Breaks: ASEAN-5 Countries

Diana Binti Hassan[1,a)] Assis Kamu[2], Ricky Chia Chee Jiun[3] and Ho Chong Mun[4]

[1,2,4] *Faculty of Science and Natural Resources, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia*
[3] *Labuan Faculty of International Finance, Universiti Malaysia Sabah Labuan International Campus, Jalan Sungai Pagar, 87000 Labuan F.T. Malaysia.*

[a)] Corresponding author: dianabintihassan@gmail.com

Abstract. The goal of this research was to investigate to investigate the relationship between stock market and macroeconomic variables in the presence of structural breaks using a panel data model approach. The selected macroeconomic variables, namely exchange rate, money supply, producer price index, and industrial production index that might affect the stock market's performance during a pandemic COVID-19 is a great approach to represent the impact of unknown factors especially when the Movement Control Order was implemented by governments worldwide. The relationship between the variables was estimated using pooled ordinary least square model, fixed effects model, and random effects model. A structural break will most probably cause any existing time series model to be inaccurate, and consequently, policy recommendations can be misleading or worse. Therefore, the outcome of this study will be crucial to the stakeholders, such as investors, fund managers, policy makers, etc. The confirmation of the existence of structural breaks, especially in the daily returns of the selected ASEAN-5 stock markets, will change their understanding of the true mechanisms driving changes in recent data. This good understanding will lead to valid conclusions and accurate forecasts in the stock markets.

Keywords: Structural break, Panel data model, Stock market performances, COVID-19, ASEAN-5

SESI SELARI 1C

Integrating AMOS and SmartPLS 4 in CB-SEM: A Comparative Study on Higher Order Construct Modeling

Nurul Raudhah Zulkifli[1, a)] and Shahazwan Mat Yusoff[2, b)]

[1]*School of Mathematical Sciences, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia.*
[2]*Department of Curriculum and Instructional Technology, Faculty of Education, Universiti of Malaya, 50603, Kuala Lumpur, Malaysia.*

[a)]Corresponding author: nurulraudhahzulkifli@gmail.com

Abstract. Recently, covariance-based structural equation modeling (CB-SEM) functionality has been integrated into SmartPLS 4.0.9.2 or later versions, yielding results from common factor analysis (CFA) and structural equation modeling (SEM) through the maximum likelihood (ML) estimator. However, its effectiveness has not been firmly established yet and demands further practical applications to validate its utility. Hence, this research investigates how the SmartPLS 4 software and the commonly used software, Analysis of Moments Structure (AMOS) can be used in CB-SEM, specifically with higher-order construct models. It presents a comprehensive comparison of these tools by highlighting their respective strengths and limitations in modeling complex relationships. The study employs a methodical approach to demonstrate practical applications in various research scenarios, thus offering insights into best practices for CB-SEM implementation. Key findings include nuanced differences in model accuracy and efficiency and providing valuable guidance for researchers in selecting appropriate software. This research makes a substantial contribution to the field by suggesting recommendations for the effective utilization of AMOS and SmartPLS 4 in advanced statistical modeling. This, in turn, enhances the accuracy and precision of research methodologies in the social sciences and other related disciplines.

Keywords: Covariance, Structural Equation Modeling, SmartPLS 4, AMOS, Higher Order Constructs

SESI SELARI 1C

Logistic Regression Modelling for Consumer Intention to Purchase Plant Factory Products

Aimi Athirah Ahmad[1, a)], Nik Rahimah Nik Omar[1, b)] , Nadiah Sa'at[2, c)], Nur Hidayah Md Noh[3,d)] and Nurul Syahida Abu Bakar[4, e)]

[1]*Socio Economic, Market Intelligence and Agribusiness Research Centre, Malaysian Agricultural Research and Development Institute, 43400 Serdang, Selangor, Malaysia.*
[2]*School of Medical Sciences, Universiti Sains Malaysia Kampus Kesihatan, Kelantan, Malaysia,*
[3]*College of Computing, Informatics and Media, Universiti Teknologi Mara Cawangan Terengganu, Kuala Terengganu Campus, Terengganu, Malaysia*
[4]*STEM Foundation Centre, Universiti Malaysia Terengganu, Terengganu, Malaysia*

[a)]Corresponding author: aimiathirah@mardi.gov.my

[b]**Abstract.** In response to increased food demand and decreasing cropland, innovative farming systems such as plant factories have arisen to boost food production sustainably. A plant factory is a closed-growing system that enables farmers to control growing conditions and can ensure safer vegetables without using pesticides. However, to maximise the success of the plant factory, it is critical to discover those aspects regarded as relevant by consumers when purchasing veggies and fruits. This study aims to explore the market potential of curly kale and strawberries cultivated through a sustainable plant factory production system by determining the factor that influences the consumer intention to purchase plant factory products using the logistic regression model. Data were collected through a structured face-to-face survey in two phases in 2022 (curly kale) and 2023 (strawberry). A total of n = 558 in phase one and n = 438 in phase two were randomly selected in urban areas across Malaysia. The result of this study explains that socio-demographic factors such as household income, plant factory consumer, lifestyle diets and consumer concerns about pesticide-free positively influence the consumer intention to purchase plant factory curly kale. Meanwhile, the logistic regression model also revealed that socio-demographic factors such as age and household size, selected attributes (taste, colour) and consumer attitudes towards plant factory significantly influenced consumer intention to purchase plant factory strawberries. This study has practical consequences, particularly for individuals or businesses interested in introducing the notion of a plant factory in Malaysia. These models' significant factors can be employed to establish the value proposed to the client as well as the marketing strategy for plant factory producers.

Keywords: purchase intention, plant factory, curly kale, strawberry, logistic regression

SESI SELARI 1C

Beyond Statistics: Using Mathematical Models to Assess the Impact of COVID-19 on Employment, Poverty, and Income Groups in Malaysia

Ummu Atiqah Mohd Roslan

Faculty of Computer Science and Mathematics,
Universiti Malaysia Terengganu,
21030 Kuala Nerus,
Terengganu, Malaysia

Corresponding author: ummuatiqah@umt.edu.my

Abstract. The COVID-19 pandemic is infecting almost all of the world's nations. Nearly every element of life has been impacted by the epidemic, including the health sector, environment, society, economy, and demography. Examining how COVID-19 has affected Malaysian socioeconomic variables is an intriguing topic. Therefore, the purpose of this study is to analyse this problem through the use of statistics and mathematical model. Several parameters, including the unemployment rate, poverty rate, and income categories, were found to have been considerably impacted by the pandemic based on the findings of the statistical analysis. Then, using ordinary differential equations, these elements were taken into account in the mathematical model. Using stability analysis method and time series plots, the dynamic behaviour of equilibrium points was examined from this model. The results demonstrated that the populations of T20 and B40 are significantly larger than those of M40 at the current rates. It is anticipated that the study's conclusions will help those engaged in providing suitable assistance and support and in creating the recovery plan.

Keywords: COVID-19

ABSTRAK
SESI SELARI
2A, 2B DAN 2C

11:00 - 11:40

SESI SELARI 2A

A Comparison of Various Control Charts for Detecting and Monitoring Abnormal River Water Levels

Siti Nur Atiqah Mohd Shafie[1, a)] and Norshahida Shaadan[1, b)]

[1]*School of Mathematical Sciences, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA,40450, Shah Alam, Selangor, Malaysia.*

[b)] Corresponding author: shahida@tmsk.uitm.edu.my

Abstract. Malaysia experiences recurring flood events, and numerous issues have surfaced annually. A key concern is the community's lack of preparedness for floods, exacerbated by the absence of early warnings, notably evident in the December 2021 incident in Taman Sri Muda. The rapid overflow of river water that causes floods triggers the need for monitoring abnormal river water levels using a visualization tool as an alarm system. This study utilized a control chart approach to assess and address this issue, aiming to identify the most reliable control chart for monitoring the Klang River's water level in Taman Sri Muda. A comparison analysis is conducted, evaluating the performance of various control charts in monitoring the Klang River's water levels at Taman Sri Muda. The study aims to assess the best control chart for the autocorrelated river water level observations. The Autocorrelation Reduction with Lag 1 (AR (1)) and a Residual control chart based on the best estimated time series model ARIMA (1,1,1) are employed in the Mean Shewhart (I-Shewhart), Moving Average (MA), and Exponential Weighted Moving Average (EWMA) with different parameters of lambda. The results have shown that the Exponentially Weighted Moving Average Residual (EWMA Residual) with λ=0.05 control chart is the most effective in detecting anomalies in river water levels, exhibiting a 7.12% sensitivity detection performance. Anomalies detected using the EWMA Residual control chart with λ=0.05 occurred dominantly in April, October, and November. The study emphasizes the efficiency of early signals produced by the control chart in revealing abnormal water level behaviour in the Klang River at the study location.

Keywords: Control Chart, Floods, River water level, Anomaly, Autocorrelation

SESI SELARI 2A

Climate Change Detection and Investigation Towards Green Environment Sustainability in Ipoh, Kuala Krai, Mersing and Temerloh Peninsular Malaysia

Noor Kesuma Mohd Yazid[1, a)], Norshahida Shaadan [1, b)], Firdaus Mohamad Hamzah[2, c)], Mahayaudin M. Mansor[1, d)] and Nurain Ibrahim[1, e)]

[1]*School of Mathematical Sciences, College of Computing, Informatics and Mathematics Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia.*
[2]*Centre for Defence Foundation Studies, Universiti Pertahanan Nasional Malaysia, Kem Sungai Besi, 57000, Malaysia*

[b)] Corresponding author: norshahida588@uitm.edu.my

Abstract. Climate change has had a detrimental impact on the ecosystem, often resulting in the extinction of many species and their habitats and harm to human health. An increasing pattern of human and industrial activity has led to a rise in greenhouse gas emissions. These have exacerbated and triggered nature. To enhance the state of the ecosystem, the green space must be preserved and carefully guarded. Thus, to increase understanding of green environment sustainability awareness, this study examines climate change in Malaysia. The methodology involved investigating the pattern of the climate variation using several visualization tools, the Mann-Kendall test, the Kruskal-Wallis test, and control charts such as the Shewhart and EWMA. Data on monthly average temperature and rainfall amount for the 30 years between 1989 and 2018 was obtained from the Department of Meteorology in Petaling Jaya and the study locations are Ipoh in Perak, Kuala Krai in Kelantan, Mersing in Johor, and Temerloh in Pahang, Peninsular Malaysia. To assess the changes, the 30 years of monthly data is divided into three cohort years; Cohort 1 (1989-1998), Cohort 2 (1999-2008), and Cohort 3 (2009-2018). Based on the data analysis conducted, the results have provided evidence that climate change occurs in all four study locations, however, among them, Mersing showed the most observed rate of change of mean temperature for both Cohort 2 and Cohort 3 from Cohort 1, by 1.01% and 1.02% while Ipoh showed the highest rate of change of mean rainfall amount by 1.20%. This study's findings support the need for green environment awareness among all Malaysians and an increased effort to educate and implement the love for nature in the general public so that our green environment can be sustained.

Keywords: Climate change, Green Environment, Temperature, Rainfall, Control Chart

SESI SELARI 2A

Statistical Extreme Value Modelling of Environmental Data

Syafrina Abdul Halim[1, a)], Nina Allyza Kepol[1, b)] and Xue Haoying[1, c)]

[1]*Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.*

[a)]Corresponding author: syafrina@upm.edu.my

Abstract. Statistical modelling of extreme values has a very practical motivation on reliability of environment data. This has significant implications for civil engineers and planners to plan and build up strategies for mitigation and adaptation risk planning. Variable interests such as temperature and rainfall are frequently used for modelling and prediction. Built-in data from R package are used in this study. Issues that may arise in fitting the model such as non-stationarity, dependency and selection of threshold are also discussed. This study will help the researchers to analyze extreme values accurately by examining the historical data and identifying patterns of extreme values. The outputs will provide valuable insights for policymakers, urban planners, and other stakeholders involved in the sustainable development and resilience of the region.

Keywords: environment; extreme values; rainfall; temperature

SESI SELARI 2B

Alternative Model Adequacy and Diagnostics for Parallel Exponentiated Exponential Regression model with censored Data

Jayanthi Arasan[1,a),*] and Anwar Fitrianto[2,b)]

[1]*Department of Mathematics and Statistics, Faculty of Science*
Universiti Putra Malaysia, 43400, Serdang,
Selangor, Malaysia.
[2]*Department of Statistics, Faculty of Mathematics and Natural Sciences,*
IPB University, Bogor, Indonesia

*corresponding author:jayanthi@upm.edu.my

Abstract. This study aims to explore several residuals based on the bias-corrected bootstrap harmonic means and influence diagnostics tools in identifying poorly fit observations, extreme values, and outliers for the parallel exponentiated exponential regression (EER) model with censored data. The effectiveness of these adjusted residuals and influence diagnostic methods will be assessed through a simulation study at different sample sizes and levels of censoring proportions. The results demonstrate that the modified Cox-Snell residuals based on the bias- corrected bootstrap harmonic mean consistently outperforms other modified residuals and influence diagnostic tools at all censoring levels and sample sizes. The proposed methods are then applied to the Diabetic Retinopathy Data with treatment as the covariate.

Keywords: extended exponential, residual, influence, censored

A Comparison of Imputation Method for a Bathtub Hazard Model with Right- and Interval-Censored Data

Idari Ismail[1,2,a)], Jayanthi Arasan[1,b)], Mohd Shafie Mustafa, [1c)], Muhammad Aslam Mohd Safari[1d)]

[1]*Department of Mathematics and Statistics, Faculty of Science, Universiti Putra Malaysia,*
43400 UPM Serdang, Selangor, Malaysia.
[2]*School of Mathematical Sciences, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Cawangan Kelantan, Kampus Machang*
Bukit Ilmu, 18500 Machang, Kelantan, Malaysia

[a)] Corresponding author: idarismail@gmail.com
[b)]jayanthi@upm.edu.my
[c)]mshafie@upm.edu.my
[d)]aslam.safari@upm.edu.my

Abstract. This paper extended a two-parameter lifetime model known as bathtub hazard model, which is widely used due to its flexibility as its failure rate function may have increasing or bathtub-shaped depending on the shape parameter. The model was expanded by including covariates in the presence of right- and interval-censored data. In order to incorporate interval-censored data into the model, imputation technique using midpoint, left point and right point approaches were used. Through a simulation study, the performance of estimators associated with this imputation techniques were assessed and compared based on the bias, standard error (SE) and root mean square error (RMSE) values. The results reveal that, midpoint imputation yields slightly smaller SE and RMSE at specific levels of censoring proportion sample sizes compared to other approaches. This suggests that, it is the best to employ midpoint imputation approach in the context of incorporating interval-censored data. Finally, a real data set of occurrence of oral lesions in children after heart transplantation was used for illustrative purposes.

Keywords: Bathtub-Shaped, Right-Censored, Interval-Censored, Simulation, Imputation

SESI SELARI 2B

Improvement of Initial Solutions for Metaheuristics using Simulated Annealing

Mohamed Saifullah Hussin[1, a)] and Mohammad Aizat Basir[1, b)]

[1]*Fakulti Sains Komputer dan Matematik (FSKM),*
Universiti Malaysia Terengganu (UMT),
21030 Kuala Nerus, Terengganu
Malaysia.

[a)]Corresponding author: saifullah@umt.edu.my
[b)]aizat@umt.edu.my

Abstract. The Quadratic Assignment Problem is an NP-Hard combinatorial optimization problem that arises in many practical applications, for example in assignment of facilities to locations, backboard wiring, turbine balancing, and many others. QAP is often solved using a class of algorithms, known as metaheuristics. Classical heuristics, such as Simulated Annealing (SA), Tabu Search (TS), and Iterated Local Search (ILS) are among the most well performing when solving the QAP. In these algorithms, the first step is to generate initial solutions that is often done randomly. An initial solution is very important in metaheuristics since it determines the starting point in the search space that affect the convergence of the algorithm. This study aims to investigate the effect of using different initialization strategies, as opposed to random initialization. In generating the initial solution, the first strategy is to generate a set of local minima and identify one with the least evaluation value to focus on solution with the most potential. Another strategy is to generate a set of local minima, but with more focus on solution with the largest evaluation value. The idea is to allow the search to begin at a location with higher potential for improvement, and hopefully will guide the search towards an area with better quality solutions. These two strategies will adopt SA algorithm, so that the search for initial solution can be done rapidly. We expect that this approach will improve the performance of algorithms compared to the ones using randomly generated initial solutions.

Keywords: initial solution, Simulated Annealing, Iterated Local Search, Tabu Search, Quadratic Assignment problem

SESI SELARI 2C

Logistic Influential Outlier Nominator

Syaiba Balqish Ariffin[1, a)] and Habshah Midi[2, b)]

[1]*School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia.*
[2]*Institute for Mathematical Research, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.*

[a)]Corresponding author: balqishariffin@usm.my
[b)]habshah@upm.edu.my

Abstract. This study addresses the critical need to identify influential outliers in multivariate data for logistic regression models. Influential outliers can distort model fitting and lead to misleading conclusions. Traditional visual inspections become inadequate as dimensions increase, and complicating outlier detection procedure due to masking and swamping effects. The Blocked Adaptive Computationally Efficient Outlier Nominators (BACON), known for its robustness in multivariate and linear regression settings, serves as the foundation for this research. However, the limitations of BACON in addressing y-outliers in binary logistic regression model prompted the development of a new approach termed Logistic Influential Outlier Nominator (LION). The multivariate LION integrates a reweighted scheme based on robust Mahalanobis distances to improve location and scatter matrix estimates. Furthermore, the inclusion of consistency and correction factors ensures reliability and reduces bias in scatter matrix estimates, particularly in small sample data. Similarly, the regression LION utilizes a reweighting scheme for the robust deviance component to identify y-outliers. The efficacy of LION is evaluated through simulation study and real example, demonstrating substantial improvements over BACON. The findings emphasize the significance of LION in providing more accurate and reliable outlier detection in binary logistic regression model, thus enhancing the reliability of statistical analyses in multivariate settings.

Keywords: Outliers, high leverage points, multivariate data, logistic regression, robust estimators.

SESI SELARI 2C

A Diagnostic Method of High Leverage Points Based on Improvised Robust Principal Component Analysis Based on Minimum Regularized Covariance Determinant in High Dimensional Data

Jaaz Suhaiza Jaafar[1, 2], and Habshah Midi[2, 3, a)]

[1] *Faculty of Computing & Multimedia, Universiti Poly-Tech Malaysia, 56100 Cheras, Kuala Lumpur, Malaysia.*
[2] *Institute for Mathematical Research, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.*
[3] *Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.*

a) Corresponding author: habshah@upm.edu.my

Abstract. The primary objective of this research is to propose a novel and enhanced diagnostic approach for detecting high leverage points (HLPs) in high-dimensional datasets. Our proposed method accurately identifies these points while simultaneously mitigates problems such as masking and swamping effects. Additionally, the proposed method offers a substantial reduction in computational times associated with high-dimensional data, which usually tend to escalate exponentially with the number of dimensionalities of the data. Our approach, which we called Improvised Robust Principal Component Analysis (IRPCA), integrates the techniques of the Minimum Regularised Covariance Determinant (MRCD) algorithm, a modified version of the Minimum Covariance Determinant (MCD) specifically designed for high-dimensional data with Principal Component Analysis (PCA), typically utilized as a dimensionality reduction tool. The combination of the two methods aims to achieve robust location and scatter estimates in the presence of HLPs in high-dimensional data. Our methodology involves computing the Robust Mahalanobis Distance (RMD) for principal component score distance for diagnosing HLPs. We compare IRPCA with both MRCD-PCA and ROBPCA to show its effectiveness, and the results show that our method yields comparable results in HLP detection within high-dimensional datasets. Notably, while MRCD-PCA and ROBPCA successfully identify high leverage points, MRCD-PCA requires longer computational times, and ROBPCA is more vulnerable to swamping problems. The credibility of our approach is supported by simulation outcomes and analyses of real-world data.

Keywords: Minimum regularized covariance determinant, Principal component analysis, High Leverage Point, Robust distance

SESI SELARI 2C

Comparative Study of Clustering Linkage Methods Using Robust Distance as Similarity Measure in Detecting Multivariate Outliers

Sharifah Sakinah Syed Abd Mutalib[1,a)], Siti Zanariah Satari[2, b)] and Wan Nur Syahidah Wan Yusoff[2,c)]

[1] *Fakulti Sains Komputer dan Matematik, Universiti Malaysia Terengganu, 21030, Kuala Nerus, Terengganu, Malaysia.*
[2] *Centre for Mathematical Sciences, College of Computing & Applied Sciences, Universiti Malaysia Pahang, Lebuhraya Tun Razak, 26300 Gambang, Pahang, Malaysia.*

[a)]Corresponding author: s.sakinah@umt.edu.my

Abstract. Outliers are abnormal data, and the detection of outliers in multivariate data has always been of interest. A visual inspection alone is insufficient for outlier detection in multivariate data, unlike univariate data. One of the methods to detect outliers in multivariate data is clustering-based methods. The goal of the clustering method is to group observations into clusters based on how similar and dissimilar the observations are from one another. Euclidean and Mahalanobis distances are used to find the distance for multivariate data. However, the difference between these two distances is that Euclidean does not take into account the covariance or shape of the data, whereas Mahalanobis distance does. Additionally, Mahalanobis distance can be robustified to overcome the problem of outliers. Therefore, in this study, robust distance is used as a similarity measure in three clustering linkage methods to detect outliers for multivariate data. The three linkage methods used in this study are single linkage, complete linkage and average linkage. The performance of these clustering linkage methods is investigated via a simulation study using one outlier scenario. Three performance measures are used, which are pout, pmask, and pswamp to determine the best clustering linkage method for detecting multivariate outliers.

Keywords: Outliers, Multivariate Data, Clustering Linkage Methods, Robust Distance, Similarity Measure

ABSTRAK
SESI SELARI
3A, 3B DAN 3C

15:00 - 15:40

SESI SELARI 3A

Enhancing BEV Charging Prediction using LSTM Networks on Feature-Engineered 1 -Minute Resolution Start-Stop Charging Data

Syahrizal Salleh[1, a)], Roslinazairimah Zakaria[1, b)] and Siti Roslindar Yaziz[1]

[1]*Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah,, Kuantan, Pahang, Malaysia.*

[a)] First author: pjz23006@student.umpsa.edu.my
[b)] Corresponding author: roslinazairimah@umpsa.edu.my

Abstract. The global surge in Battery Electric Vehicle (BEV) adoption has led to widespread acceptance, prompting several nations to exclusively commit to the sale of BEVs for new passenger cars from 2035 to 2040. This commitment represents a strategic initiative to mitigate the release of greenhouse gases (GHGs) into the atmosphere, as BEVs rely solely on onboard batteries for propulsion. However, the simultaneous charging of multiple BEVs places stress on the electric grid, creating challenges related to grid capacity. Electric utility companies are actively forecasting demand capacity from BEVs to mitigate losses from overgeneration. Additionally, these companies must enhance electricity generation and grid infrastructure to accommodate increasing demands that surpass current capacity levels. To address the scarcity of high-resolution BEV charging data, a novel feature engineering technique was applied, transforming start-stop electricity charging data sourced from the My Electric Avenue project into the count of concurrent charging events. Recognizing the nonlinear, dynamic, and noisy nature of BEV charging behavior, a Long Short-Term Memory (LSTM) network was chosen to model the electricity demand arising from multiple concurrent BEV charging events. The selected LSTM network comprises a single layer with 125 units of LSTM cells employing a tanh activation function and a single dense output layer. In the model optimization process, hyperparameter tuning focused solely on the number of epochs, with intervals set at 1, 10, 20, 30, 40, and 50 epochs. The LSTM network approached its global minima at epoch 10, and the lowest Mean Absolute Percentage Error (MAPE) was achieved at epoch 20. The lowest recorded MAPE was 1.19%, with a corresponding Root Mean Squared Error (RMSE) of 0.51. The LSTM model, developed for 1-minute resolution electricity demand, is suitable for electric utility companies to predict very short-term load, with forecasting ranging from a few minutes to a few hours ahead.

Keywords: Battery Electric Vehicles, Long Short-Term Memory, Feature Engineer, Charging Behavior.

Consumer Adoption and Perceptions of Electric Vehicles in Malaysia Using Sentiment Analytics

Nor Farawahida Abdullah[1, b)] and Nur Haizum Abd Rahman[1, a)]

[1]*Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuhraya Persiaran Tun Khalil Yaakob, 26300 Gambang, Pahang, Malaysia.*

[a)]Corresponding author: haizum@umpsa.edu.my

Abstract. Despite the excitement around electric automobiles, Electric Vehicles (EVs) in Malaysia are still relatively new compared to more advanced markets such as European countries. To understand Malaysian acceptance and perception of EVs, this study uses sentiment analysis to ascertain whether Malaysians view and utilise EVs positively or negatively. The dataset is scrapped data collected from comments on YouTube with count vectorisation and term frequency-inverse document frequency (TF-IDF) as the feature extraction methods. The machine learning algorithm, Logistic Regression and Support Vector Machine (SVM) are then created to identify and classify comments about EV adoption and perception automatically. Key performance indicators, including accuracy, precision, and recall percentages, are used to assess the algorithm's performance. From the collected data, the feedback and reviews significantly impacted this study topic. The result of this study will provide nuanced insights into public opinions regarding EVs in Malaysia by seamlessly integrating machine learning and sentiment analysis. The findings will have substantial implications for researchers in the sustainable transportation sector, policymakers, and industry stakeholders.

Keywords: Electric Vehicles, Sentiment Analytics, Machine Learning, Perceptions

SESI SELARI 3A

A Simulation Based Optimization Approach in Determine Loyalty Rewards Point Value for Loyalty Program

Wan Nuraini Fahana Binti Wan Nasir[1, a)], Muhamad Safiih Bin Lola[1, b)]

[1]*Faculty of Science Computer and Mathematics, Universiti Malaysia Terengganu, 21300 Kuala Nerus, Terengganu, Malaysia.*

[a)] wannuraini.fahana@umt.edu.my
[b)] safiihmd@umt.edu.my

Abstract. Loyalty program (LP) is seen as one of the marketing strategies. LP helps a company to achieve the goal that leads to strengthening the relationship between the company and the customer. One of the issues faced by LP managers is to make plans for determining effective reward points value to achieve management goals and customer satisfaction. Uncertainties in determine the loyalty rewards points value usually increase the variance of the profits or cost to the company and increasing the likelihood of decreased profit. In this paper we proposed the use of deterministic planning supply of rewards points model which incorporate customer valuation as a means of meeting customers' satisfaction in routine operation. The problem of determining the reward point value to meet the desired level of customers' satisfaction is addressed using a simulation-based optimization approach. The results reveal that the optimal solution is depend on the parameters value. The firm is able to adjust the reward point value that will obtain the maximum total profits. The proposed model is only applicable for a company that runs a LP that use points to redeem and does not compete with other companies. However, the model developed represents useful analytical tools to support LP provider to make best decision making in setting the reward point value.

Keywords: Loyalty program, rewards point, customer satisfaction, simulation, optimization problem

SESI SELARI 3B

Exploring the Pattern of Household Monthly Income & Expenditures by State in Malaysia: Statistical Approach.

Syerrina Zakaria[1, a)], Nor Fatimah binti Che Sulaiman[2, b)] and Siti Madhihah Abdul Malik[1, c)]

[1]*Mathematics Program, Faculty of Computer Science and Mathematics, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia.*
[2] *Faculty of Business, Economics and Social Development, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia.*

[a)] Corresponding author: syerrina@umt.edu.my
[b)] n.fatimah@umt.edu.my
[c)] madhihah@umt.edu.my

Abstract. Viewed from a long-term perspective, Malaysian households have surely seen significant improvements in their standard of living, livelihoods, and economic opportunities. This study aims to investigate the pattern and relationship between household income, household expenditure and household size among states in Malaysia. Several methods have been used to achieve the objectives such as Pearson correlation coefficient, paired t-test, Anova and multiple regression method. It was found that, there is significant difference in the mean monthly household median income across states in Malaysia for the two years. This implies a significant increase in household income over the three-year period. Similar significant differences were observed for monthly household expenses. A significant difference in the mean household expenditure was found between the years 2014 and 2019, but no statistical difference was observed between 2014 and 2016, as well as between 2016 and 2019. Despite the significant increase in household median income, households seemed unable to allocate the income increment to their expenses, possibly due to higher living costs requiring adjustments in expenditure. As result from multiple regression method, Median Income was a significant positive predictor of median expenses, but house hold size was not a significant predictor. The results suggest important policy implications that should be considered in formulating future policies, especially those related to household income and expenditure across states in Malaysia. This is crucial for improving the living standards and well-being of Malaysians. Future research could enhance the understanding of household income and expenditure by utilizing comprehensive data at the district level within each state.

Keywords: Household income, household expenditure, household size, Statistical approach.

SESI SELARI 3B

Data Lakehouse Architecture for Self-service Data Analytics

Soon Kien Yuan[1, b)], Nor Azuana Ramli[1, 2, a)] and Mohd Zaid Waqiyuddin Mohd Zulkifli[3, c)]

[1]*Centre for Mathematical Science, Universiti Malaysia Pahang Al-Sultan Abdullah, 26300 Gambang, Pahang, Malaysia.*
[2] *Centre for Data Sciences & Artificial Intelligence, Universiti Malaysia Pahang Al-Sultan Abdullah, 26300 Gambang, Pahang, Malaysia.*
[3]*Credence, 1 Jalan Damansara, Damansara Kim, 60000 W.P. Kuala Lumpur, Malaysia.*

[a)]Corresponding author: azuana@umpsa.edu.my
[b)]sd20040@student.ump.edu.my
[c)]zaidwaqiyuddin@credence.tech

Abstract. This project focuses on designing and implementing a Data Lakehouse architecture to facilitate self-service analytics. The objectives include creating a collaborative analytics environment, streamlining the management of multiple extract, transform and load (ETL) processes, adopting a cost-effective and non-proprietary architecture, integrating with business intelligence (BI) tools, ensuring high query performance for interactive visualization, enabling data warehousing capabilities, and offering a self-service data discovery and metadata platform. The project followed an iterative development methodology that involves requirement gathering and planning, design, implementation, testing, deployment, and maintenance phases. The logical design consists of six layers which are data ingestion, storage, catalog, semantic, processing, and consumption. The physical design utilized Dremio as the core component, along with Apache Iceberg and Arrow Flight Engine for data format and query processing. The project also adopted an Integrated Multi-Zone Analytics Framework to handle different data tasks and workloads. The implementation was performed using Docker, and the testing validates the achievement of the objectives. Deployment was done on an Azure Kubernetes Service (AKS) cluster, although the deployment of the Dremio cluster is hindered due to budget and resource limitations. Maintenance activities included security, backup, node monitoring, cost usage, configuration tuning, metadata management, and training and support. This project concludes that the objectives have been achieved and suggests some future enhancements, such as using Project Nessie for catalog storage, considering Dremio Enterprise Edition for advanced features, and exploring Databricks and MLflow if expecting extensive machine learning workloads.

Keywords: Data Lakehouse, data management, business intelligence, architecture, data analytics

SESI SELARI 3B

Factor Analysis of Internet Use – A Case Study among Universiti Malaysia Terengganu Students

Siti Madhihah Abd Malik[1, a)], Muhamad Safiih Lola[1, b)]

[1]*Bidang Matematik, Fakulti Sains dan Matematik, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia.*

[a)]Corresponding author: madhihah@umt.edu.my
[b)]safiihmd@umt.edu.my

Abstract. The Internet has become an influential communication tool, shaping various aspects of our lives. Understanding the factors that drive Internet usage is crucial for optimizing its benefits. However, despite its significance, there remains a gap in research regarding the dominant factors affecting Internet use. In this study, we explore these factors and also assess the level of computer skills among Universiti Malaysia Terengganu (UMT) students using factor analysis. Our research instrument consists of a well-structured questionnaire, administered to a sample of 120 randomly selected UMT students. The study's outcomes reveal six distinct factors influencing internet use and seven components characterizing the computer and internet skills of UMT students. The identified factors influencing internet use encompass references and information sources, connectivity methods, quality and service, entertainment, and location and frequency. Additionally, the components of computer and internet skills include internet applications, programming, statistical programming, Microsoft Office usage, computer center utilization, participation in courses, and engagement in other learning initiatives. The widespread popularity of internet usage is attributed to multiple factors, and factor analysis proves instrumental in discerning these underlying influences.

Keywords: Internet, factor analysis, computer skills, internet usage factors, internet applications

SESI SELARI 3C

Statistical Analysis Cyberattack Events That Related to Malaysia

Hafiz Asyraaf Ghani

[1]*Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah, 26300 UMPSA Gambang, Pahang, Malaysia*

Corresponding author: sd20033@student.umpsa.edu.my

Abstract. Cyberattacks are a very delicate subject in the field of cyber security. Both governments and businesses are devoting massive resources to protect their information systems. In an effort to steal sensitive information, adversaries are actively working to penetrate security and distribute malware including botnets, viruses, and trojans. Understanding these assaults both before and after they occured is crucial for enhancing system security. It is possible to better defend a network from future attacks if security personnel are familiar with attack models. Without understanding a network's weaknesses, it is hard to foresee a possible attack in the realm of cyber security. To effectively secure the network, it is necessary to first conduct an analysis to determine the most likely set of vulnerabilities. The network and critical data are at serious risk if an ongoing attack is not dealt with immediately. In order to grasp the network's susceptibility and the adversary's behaviour and goals, this study analysed the most common types of cyber attacks observed in a log network traffic analysis dataset, identified the best time series forecasting model and visualize the result of statistical analysis based on data by using Python. The best ARIMA model was used to perform the statistical analysis. The best model was chosen by the least values of mean absolute error (MAE) and root mean squared error (RMSE) and Mean Absolute Percentage Error (MAPE). The outcome of this research is presented in an interactive dashboard as a deployment of this project. The outcome of this study shows the potential of Distributed Denial-Of-Service attack can happen for any sector due to TcpUdp_SG2 showing the highest among the other messages.

Keywords: Python, ARIMA, MAE, MAPE, RMSE, Statistical Analysis, Predictive Modelling

Developing a New Feature for Vulnerability Risk Scoring Model for Enhanced Cybersecurity

Lim Joey[1,a)] and Noryanti Muhammad[1,b)]

[1]*Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuhraya Persiaran Tun Khalil Yaakob, 26300 Gambang, Pahang, Malaysia.*

[a)] SD21038@student.umpsa.edu.my
[b)] Corresponding author: noryanti@umpsa.edu.my

Abstract. As Malaysia becomes increasingly reliant on fast-paced technologies, the vulnerability of data to various cyber threats rises significantly. Cybersecurity companies are crucial in bolstering organisational resilience by implementing comprehensive risk-scoring models. However, given the evolving nature of cyber vulnerabilities, with diverse branches emerging over time, it becomes imperative to tailor a risk-scoring model specifically for Malaysia. This research aims to achieve two key objectives, where the focus is to develop a sample risk-scoring model and create a new feature that reflects the risk score of the situation developed using the currently existing dataset in this study. The proposed model is designed based on the most prevalent cyber threats in Malaysia, addressing the unique challenges faced by the nation. In this paper, we focus on enhancing cybersecurity measures through innovative approaches. Methods using multivariate analysis and regression modelling using machine learning are employed, and an appropriate evaluation metrics for regression models are used to measure the performance of the models developed. The results show that the new features created which correspond to the K-Nearest Neighbors regression model give a good performance based on the evaluation metrics. The risk score of cyber vulnerabilities, where higher scores indicate a greater need for urgent and priority attention for cybersecurity companies. The potential application of this research includes by considering the proposed model as an effective alternative to mitigate specific risks faced by Malaysia.

Keywords: Risk-scoring model, Multivariate analysis, Correlation, Feature Generation, Regression modelling

SESI SELARI 3C

Scalability and Performance in Duplicate Detection: Relational vs. Graph Database

Noryanti Muhammad[1,a)], Muhammad Farhad Khaharruddin[1,b)] and Mohd Izhar Firdaus[2,c)]

[1]*Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuhraya Persiaran Tun Khalil Yaakob, 26300 Gambang, Pahang, Malaysia.*
[2] *Abyres Enterprise Technologies Sdn Bhd (AET), C2-2-7, Block C2, CBD Perdana 3, Lingkaran Cyber Point Timur, Cyber 12, 63000 Cyberjaya, Selangor, MALAYSIA.*

[a)]Corresponding author: noryanti@umpsa.edu.my
[b)] sd20028@student.ump.edu.my
[c)] izhar@abyres.net

Abstract. Duplicate detection is a common problem in customer data management, as it is important to be able to identify and remove duplicate records to maintain the data's integrity and reliability. Traditional approaches to duplicate detection rely on comparing individual fields within records, such as name, address, and phone number, to determine whether two records refer to the same customer. Additionally, it has been demonstrated that complex data makes it harder to identify duplicate records for businesses, which can increase processing time and leads to raising overhead cost. This research focuses on investigating the application of graph algorithms for identifying duplicate entries in customer data. Then, the scalability and algorithm performance of the graph database systems is evaluated in identifying duplicate customer data, contrasting their efficacy with that of relational databases. The findings revealed that PostgreSQL exhibits exceptional scalability and efficiency in handling large datasets, surpassing both Neo4J and MySQL. However, Neo4J excels in exact duplicate and near-duplicate detection algorithms, highlighting its proficiency in managing complex, interconnected data structures. In conclusion, the choice between PostgreSQL and Neo4J should be made based on specific task requirements, with PostgreSQL preferred for fast identification of similar items in large datasets, while Neo4J proves more suitable for tasks involving the discovery of communities within intricate networks.

Keywords: Graph, Duplicate Detection, Community Detection, Unsupervised Operations, Complex Relationships

NOTA

Sekretariat SKISM-XVIII 2024

Institut Penyelidikan Matematik
Universiti Putra Malaysia
43400 UPM Serdang, Selangor

Tel: 03 9769 8923 / 6854
Emel: skism2024@gmail.com