ISDC 2022

ABSTRACTS BOOK

Editor

Sedat ÇAPAR

November, 2022

ISDC 2022

XIIth INTERNATIONAL STATISTICS DAYS CONFERENCE

İzmir, TÜRKİYE 2022

ABSTRACTS BOOK

ISBN: 978-975-441-576-6

Editor

Sedat ÇAPAR

November, 2022

CONTENTS

Organizing Committee

Chair	
Güçkan YAPAR	Dokuz Eylül University

Co Chairs	
Hanife TAYLAN SELAMLAR	Dokuz Eylül University
İdil YAVUZ	Dokuz Eylül University
Sedat ÇAPAR	Dokuz Eylül University
Özgül VUPA ÇİLENGİROĞLU	Dokuz Eylül University

Members	
Burcu HÜDAVERDİ	Dokuz Eylül University
Tuğba YILDIZ	Dokuz Eylül University
Begüm KARA GÜLAY	Dokuz Eylül University
Özgür DANIŞMAN	Dokuz Eylül University
Uğur BİNZAT	Dokuz Eylül University
Sinan SARAÇLI	Afyon Kocatepe University
Mustafa Cagatay KORKMAZ	Artvin Coruh University
Deniz ÜNAL	Çukurova University
Erol EĞRİOĞLU	Giresun University
Çağdaş Hakan ALADAĞ	Hacettepe University
Kasırga YILDIRAK	Hacettepe University
Aysen AKKAYA	Middle East Tecnical University
Atila GÖKTAŞ	Muğla Sıtkı Koçman University

Scientific Committee

Prof. Dr. Güçkan YAPAR	Dokuz Eylül University, TÜRKİYE
Prof. Dr. Victor LEİVA	Pontificia Universidad Católica de Valparaíso, CHILE
Prof. Dr. Saraless NADARAJAH	University of Manchester, UK
Prof. Dr. G. G. HAMEDANI	Marquette University, USA
Prof. Dr. Mir Masoom ALI	Ball State University, USA
Prof. Dr. Debasis KUNDU	Indian Institute of technology, INDIA
Prof. Dr. Nadeem Shafique BUTT	Abdulaziz University, Jeddah, KSA
Prof. Dr. S. Graça Pinto VARANDAS	Universidade de Trás-os-Montese Alto Douro, PORTUGAL
Prof. Dr. Josmar MAZUCHELI	Universitat Autònoma de Barcelona, SPAIN
Assoc. Prof. Dr. Sedat ÇAPAR	Dokuz Eylül University, TÜRKİYE
Assoc. Prof. Dr. Özgül Vupa ÇİLENGİROĞLU	Dokuz Eylül University, TÜRKİYE
Assoc. Prof. Dr. İdil YAVUZ	Dokuz Eylül University, TÜRKİYE
Assoc. Prof. Dr M. Bourguignon PEREİRA	Universidade Federal do Rio Grande do Norte, BRAZIL
Assoc. Prof. Dr. Faton MEROVCI	University of Mitrovica Isa Boletini, KOSOVO
Assoc. Prof. Dr. S.M.T.K. MIRMOSTAFEE	University of Mazandaran, IRAN
Assoc. Prof Dr. Christophe CHESNAU	Université de Caen Normandie, FRANCE
Asist. Prof. Dr. Vikas Kumar SHARMA	Banaras Hindu University, INDIA
Asist. Prof. Dr. Hanife Taylan SELAMLAR	Dokuz Eylül University, TÜRKİYE
Asist. Prof. Dr. Morad ALIZADEH	Persian Gulf University, Bushehr, IRAN
Asist. Prof. Dr. Haitham M. YOUSOF	Benha University, EGYPT
Vladimir ZAIATS	Universidade Estadual de Maringa, BRAZIL
Sylwia GWOZDZIEWICZ	University in Gorzow Wielkopolski, POLAND
Armando GONÇALVES	Universidade Federal do Rio de Janeiro BRAZIL

Program

October 13 (Thursday)

	HALL 1 (Arena)
	REGISTRATION
13:30-14:00	**OPENING CEREMONY**
14:00-14:45 INVITED SESSION 1	Judgmental forecasting: the state of play and some interesting reflections from personal research **Prof. Konstantinos NIKOLOPOULOS**
14:45-15:00	**COFFEE BREAK**
15:00-15:45 INVITED SESSION 2	"Yet" –focus on our learning slope not the intercept **Assoc. Prof. Dr.Orrin COOPER**
15:45 – 16:00	**COFFEE BREAK**
Session 1	**HALL 1**
Session Chair	**Assoc. Prof. Dr.Orrin COOPER**
16:00-16:20	A Comparison Study for the Outlier Detection Methods in Regression **Hakan Savaş Sazak, Berfin Saraçoğlu**
16:20-16:40	A Comparison Study for the Robust Time Series Methods **Işıl Diker, Hakan Savaş Sazak**
Session 1	**HALL 2 (Celsus)**
Session Chair	**Assoc. Prof.Dr. İdil YAVUZ**
16:00-16:20	Dependent Metaverse Risk Forecasts With Heteroskedastic Models and Ensemble Learning **Venansius R. Tjahjono, Arief Hakim, Khreshna Syuhada**
16:20-16:40	The VSI Tukey-Exponentially Weighted Moving Average Control Chart with Asymmetrical Limits **Selcem Adsız, Burcu Aytaçoğlu**

16:40-17:00	The Multivariate Extension of CoVaR With Higher-Order Moments for Tail Risk Network Construction **Arief Hakim, A.N.M. Salman, Khreshna Syuhada**
17:00-17:20	Value-at-Risk Forecast and Its Extensions **Khreshna Syuhada, Arief Hakim**
17:20-17:40	An Optimal Combination of Proportional and Stop-Loss Reinsurance Under Dependent Claim and Random Insurance Premium **Suci Sari, Arief Hakim, Khreshna Syuhada**
17:40-18:00	Mortality-at-Risk Forecast based on APARCH Model and Its Application to Determine Reserve Amount in Pension Fund **Darin Sabrina, Venansius R. Tjahjono, Khreshna Syuhada**

OCTOBER 14 (Friday)

	HALL 1 (Arena)	HALL 2 (Celsus)
10:00-10:45 INVITED SESSION1	Looking at meta-heuristics & a closer look at the firefly algorithm **Prof. Dr. Adil BAYKASOĞLU**	
10:45-11:15	COFFEE BREAK	
11:15-12:00 INVITED SESSION2	Useful Intersection of Statistical and Machine Learning **Prof. Dr. Ç. Hakan ALADAĞ**	
12:00-13:30	**Lunch Time**	
Session Chair	**Prof. Dr.Ç. Hakan Aladağ**	**Prof. Dr. G. Yazgı Tütüncü**
13:30-13:50	Use of Six Sigma in the Optimization of the Welding Process of Titanium Tubes: An Application in the Aerospace and Aviation Industry **Atakan Gerger**	Logistic and CSG Growth Models for the Prediction of Expected Life Expectancy **Deniz Ünal, Begüm Çığşar**
13:50-14:10	Multivariate Analysis of Variance Under Laplace Distribution **Müge Borazan, Serpil Aktaş Altunay**	Estimation of the solar plant power with PV modules subject to failure **Melek Esemen,Selma Gürler,Gözde Yazgı Tütüncü**
14:10-14:30	Genetic Algorithm Based Hybrid Model for COVID-19 Cases	Fuzzy Regression Network Functions for Time Series Prediction

	Forecasting **Ufuk Yolcu, Özge Cagcag Yolcu**	**Mehmet Raci Aktoprak, Özge Cagcag Yolcu**
14:30-14:50	Using Kullback–Leibler Divergence As Loss Function in Convolutional Neural Networks for Distribution Comparison **Cağatay Bal, Serdar Demir, Cağdaş Hakan Aladağ**	
14:50-15:10	A Hybrid Approach for Hyper-parameter Selection in Text Classification for Covid-19 tweets **Cansu Doğan, Çağdaş Hakan Aladağ**	
15:10-15:30	COFFEE BREAK	COFFEE BREAK
Session Chair	**Prof. Dr. Adil Baykasoğlu**	**Prof. Dr. Sinan Saraçlı**
15:30-15:50	On the Number of Failed Components in a Series-Parallel System **Murat Ozkut**	Bayesian Analysis of Turkish Income and Living Conditions Data, Using Clustered Longitudinal Ordinal Modelling With Bridge Distributed Random-Effects **Özgür Asar**
15:50-16:10	Robust Cryptocurrency Portfolio Optimization by Using NNN and MNN **Ilgım Yaman**	How Significant Is Chi-Square Test To Determine the Image Quality? A Retrofit in Calculation **Sinan Saraçlı, Şevkiye Babacan, Berkalp Tunca**
16:10-16:30	A Statistical Analysis of Usability of a Learning Management System **Ceylin Unal, Adil Baykasoğlu, Burcu Felekoğlu**	
16:30-16:50	Financial Performance Analysis with the TOPSIS, Fuzzy COPRAS and Entropy-COPRAS Approaches: The Case of Turkey **Yüksel Akay, Sinem Uz**	

	HALL 3 (Pagos)
Session Chair	**Prof. Dr. Adil Baykasoğlu**
13:30-13:50	Hotel Ranking and Recommendation Through Online Reviews With Integrated Weighting Approach and ARAS Method **Sedef Çalı, Adil Baykasoğlu**
13:50-14:10	Evaluation of the Success Status of University Students on the Basis of Faculties in Formal and Distance Education **Burçin Öner, Yüksel Öner**
14:10-14:30	Inflation Prediction With Artificial Neural Networks **Hasan Şen, Ömer Faruk Efe**
14:30-14:50	Machine Learning for Crop Prediction: Understanding Wheat Yield by Remote Sensing Data **Ozan Evkaya, Cansu Dogan**
14:50-15:10	HangiPixel: A Web Analytics Tool for Banking Sector **Türker Ziya Erçin, Bekir Çetintav**
15:10-15:30	COFFEE BREAK
Session Chair	**Assoc. Prof.Dr. İdil YAVUZ**
15:30-15:50	Prediction of Absorption Dose of Radiation on Thorax CT Imaging in Geriatric Patients with COVID-19 by Classification Algorithms **Adnan Karaibrahimoğlu, Ümit Kara,Özge Kılıçoğlu,Yağmur Kara**
15:50-16:10	Using Unsupervised and Supervised Machine Learning Algorithms to Evaluate Higher Education Institutions **Tuğba Söküt Açar**
16:10-16:30	Deep Fuzzy Functions Approach for Time Series Forecasting **Ali Zafer Dalar**
16:30-16:50	Ata Forecasting Method as an Alternative to Exponential Smoothing in Forecast Combination **Uğur Baran Yapar, İdil Yavuz**
16:50-17:10	Predicting Personality Traits of Employee Candidates Using Natural Language Processing **Fatma Özge Ozkok**

OCTOBER 15 (Saturday)

	HALL 1 (Arena)	HALL 2 (Celsus)
Session Chair	Assoc. Prof. Dr. Özgül Vupa Çilengiroğlu	Prof. Dr. Burcu Hüdaverdi
9:30-9:50	A Multiple Objective Optimization Model for a Novel Capability-Based University Course Timetabling Problem: A Case Study at Deu Industrial Engineering Department **Kemal Subulan, Ahmet Gürsaç**	Reliability Assessment of Maximum Likelihood Estimation on a Weighted Weibull Distribution **Yasin Altınışık, Demet Aydin**
9:50-10:10	The Mediating Effect of Hedonic and Utilitarian Value on the Relationship Between Technological Aptitude and Intention To Continue Using Wearable Technology **Ceren Börüban, Dilek Veysikarani, Ebru Özgür Güler**	A New Weighted Independence Test Based on Smooth Estimation of Kendall Distribution Function **Selim Orhun Susam, Burcu Hüdaverdi**
10:10-10:30	Bayesian Analysis of Repeated Measures: An Application **Fatma Yardibi, Mehmet Ziya Fırat**	
10:30-10:50	Bibliometric Analysis by Using CiteSpace **Fatma Yardibi**	
10:50-11:10	COFFEE BREAK	COFFEE BREAK
Session Chair	Dr. Atakan Gerger	Assoc. Prof. Dr.Tuğba Yıldız
11:10-11:30	On Weighted Random Sampling and the Computation of Participants' Chances in Turkey's Hajj Draws **Murat Güngör**	The Half-Logistic Garima Distribution and Modelling **Ayşe Metin Karakaş, Aslıhan Demir, Sinan Çalık**
11:30-11:50	The Performances of Four Quantile Estimators When Comparing Dependent Groups With a Percentile Bootstrap Method **Engin Yıldıztepe, A. Fırat Özdemir, Tuğçe Paksoy**	A Note on the Moments of Sample Minimum of Order Statistics From a Geometric Distribution **Aslıhan Demir, Ayşe Metin Karakaş,Sinan Çalık**

11:50-12:10	Forecasting Power Outages **Beyza Yapıcı, Esra Kılkıl**	Investigating Obesity Related Health Factors of Bariatric Surgery Patients by Using Machine Learning Algorithms **Sezen İnan, Özgül Vupa Çilengiroğlu, Metin Karadeniz**
12:10-12:30	Study of Artificial Neural Networks for Single Jersey Fabrics in the Textile Industry **İrem Soylu, Neslihan Demirel**	Estimating of the Distribution Function Under Ranked Set Sampling With Unequal Probabilities **Yusuf Can Sevil, Tuğba Özkal Yıldız**
12:30-13:30	Lunch Time	Lunch Time
Session Chair	**Assoc. Prof. Dr. Özgül Vupa Çilengiroğlu**	**Prof. Dr. Neslihan Demirel**
13:30-13:50	Comparing Forecast Accuracies of Exponential Smoothing and Ata Method in Demand Forecasting **Miray Betül Yılmaz, Özgül Vupa Çilengiroğlu**	Forecasting the Volatility of the Cryptocurrency Market: A Comparison of GARCH-type Models and Machine Learning Algorithms **Ömer Burak Akgün, Emrah Gülay**
13:50-14:10	Estimation of the Price Movement Direction of Cryptocurrencies With Stochastic Processes: An Application for Bitcoin **Özgür Danışman**	Classification With Bernstein Copula Through Machine Learning Process **Tolga Yamut, Burcu Hüdaverdi**
14:10-14:30	Comparison of the Fit of Cox Regression and Random Survival Forest Model With the Help of Simulation Data **Tuğçe Paksoy, İdil Yavuz**	
14:30-14:50	New Multilayer Neural Networks With NO Estimator and Winsorized Mean **Burak Dilber, A. Fırat Özdemir**	
14:50-15:10	COFFEE BREAK	
Session Chair	**Prof. Dr. Güneş Çetin Gerger**	
15:10-15:30	The Challenge of the Artificial Intelligence **Fatih Şahin**	

15:30-15:50	Proposition of A New Updated Robust Ridge Parameter Estimator for Linear Regression Model **Özge Akkuş, Selman Mermi, Atila Göktaş**	
15:50-16:10	Impact of Digital Transformation and EU Green Deal on Employment **Güneş Çetin Gerger, Atakan Gerger**	

	HALL 3 (Pagos)
Session Chair	**Dr. Hanife Taylan Selamlar**
9:30-9:50	Validity and Reliability of the Singles Stress Scale-Youth Form (SSS-YF) **K. Bahar Aydın, Nailakhanim Rustamova**
9:50-10:10	An Alternative Quantile Regression Model for the Unit Response Based on the New Unit Probability Distribution **Berrin Gültay, Mustafa Çağatay Korkmaz**
10:10-10:30	Seasonal time series analysis: decomposition and forecasting with Ataforecasting package in R **Hanife Taylan Selamlar**
10:50-11:10	**COFFEE BREAK**
Session Chair	**Prof. Dr. Burcu Hüdaverdi**
11:10-11:30	Investigation of Stress Effect in Parallel Connected Systems **Mine Doğan, Mehmet Gürcan**
11:30-11:50	Examining The Effects Of The Pandemic on Turkey's Happiness with The Dynamic Markov Switching Model **Elif Tuna**
11:50-12:10	The Effects of Under-Sampling the Dataset in Intrusion Detection with Machine Learning **Sura Emanet, Gözde Karatas Baydoğmus, Önder Demir, Sadettin Melenli**
12:30-13:30	**Lunch Time**
Session Chair	**Dr. Engin Yıldıztepe**

13:30-13:50	Similar QA Pairs Detection System to Improve Chat Quality in Turkish Conversation Groups **İzzet Kılıç, Derya Karagöz**
13:50-14:10	Comparison of Machine Learning and Traditional Approaches for Automobile Insurance Pricing **Müge Yeldan, Uğur Karabey**
14:10-14:30	The Turkish Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach **Çiğdem Lazoğlu, Uğur Karabey**
14:30-14:50	Using Alternative Prediction Methods in Portfolio Selection **Azize Zehra Çelenli Başaran, Özge Gündoğdu, Ali Zafer Dalar**
15:10-15:30	COFFEE BREAK
	Assoc. Prof. Dr. Tuğba Yıldız
15:30-15:50	Team Size Optimization for Power Restoration in Workforce Management **Esra Kılkıl, Efendi Nasiboğlu**
15:50-16:10	Determinants of Carbon Dioxide Emissions for Turkey: Quantile Regression Approach **Çiğdem Koşar Taş, Sibel Örk Özel, Dilek Veysikarani**
16:10-16:30	How Much Should We Trust the P-Value? **Oya Ekici**

Abstracts of Invited Speakers

"Yet" – focus on our learning slope not the intercept

Orrin Cooper
The Universit of Memphis

ABSTRACT

What do fundamental learning strategies, mastery-based learning, and growth mindset have in common with the term "Yet?" How is the three-letter word "Yet" a catalyst in a fundamental shift in learning and teaching?

Intercepts, or starting points, have long been used to predict individuals' potential. However, focusing too much on anyone's starting point can define them and/or constrain their learning slope. The most important predictor of potential and success is not where we start (intercept) but our slope. One of the most powerful principles for both teachers and learners to understand is our innate capacity for lifelong learning. Research about how learning occurs and how "experts" become "experts" teaches a healthy perspective and mindset about deep learning. Next, it is important to know and apply fundamental learning strategies that have been proven effective across varying starting points (intercepts). These are powerful tools that elevate everyone's learning slopes. Many of the fundamental learning strategies on the surface can seem counterintuitive like: desirable difficulties, calculated risk taking and failure, and testing as a learning tool. One of the biggest obstacles in yoking our potential with fundamental learning strategies was user and designer accessibility to technology. COVID accelerated both the access to and comfort level of user and designer alike in using powerful technological tools that can implement fundamental learning strategies. We should also look to remove other obstacles, even hidden features, that limit and constrain learning and may even perpetuate inequities.

Finally, what if an individual with a lower starting point (intercept) exhibits more self-discipline, grit, and persistence than individuals with higher starting points (intercept) and by the end or a learning period because of a more elevated slope surpasses her counterparts in terms of mastery of the material? Mastery-based learning is an effective approach that relies heavily on the principle about our innate learning potential. Mastery-based learning also requires a paradigm shift about when learning occurs, how it is measured, and when it is rewarded. It can help everyone achieve a higher standard of learning, independent of their starting point (intercept). Intentional design can leverage the power in the word "Yet" and by elevating everyone's learning slope provide a fundamental shift in learning and teaching.

Judgmental Forecasting: the State of Play and Some Interesting Reflections from Personal Research

Konstantinos Nikolopoulos
Durham University Business School

ABSTRACT

In this keynote we present the state of play in the field of judgmental forecasting and behavioural operations, giving the necessary background to the area of predictive analytics as always so reasonable comparisons can be made on the merits and importance of these two anti-diametric ends of the forecasting discipline. We also present some recent findings on herding behavior and forecasting, on judgmentally forecasting the success of megaprojects, and on long term projections and scenarios for the distant future.

Useful Intersection of Statistical and Machine Learning

Cağdaş Hakan Aladağ
Hacettepe University

ABSTRACT

Statistical approaches are at the core of machine learning. In spite of this fact, sometimes these two important approaches can be competed with each other in a very wrong way. Or, it can be entered into a very wrong discussion as to which one is better in data analysis. According to the data examined, both approaches may offer some advantages or disadvantages according to themselves. It is possible to achieve a more powerful data analysis process by using these two fundamental approaches together, rather than competing them with each other. By using the advantages of both approaches, each other's disadvantages can be overcome as much as possible so data analysis can be performed more effectively. Thus, the way to reach more reliable and robust predictions will be opened. In this study, current examples of how statistical and machine learning can be used together effectively will be given, and the gains will be revealed.

Looking at Meta-Heuristics and A Closer Look at The Firefly Algorithm

Adil Baykasoğlu
Department of Industrial Engineering
Dokuz Eylül University, İzmir, Türkiye
adil.baykasoglu@deu.edu.tr

ABSTRACT

Meta-heuristics are problem independent stochastic, blind search algorithms. They are applied to diverse problems from many scientific disciplines with success. Due to their popularity, extensive number of meta-heuristics are proposed in the literature. There are more than 300 meta-heuristic algorithms in the literature that are called as new, high performing, promising etc. This inflation of meta-heuristic algorithms start to create some serious criticisms in the related literature. The main criticisms are about low level of novelty (i.e. some of the proposed algorithms are very similar to each other); Improper validation and justification of the proposed meta-heuristic algorithm (i.e. hiding lack of effectiveness of the proposed algorithms); using confusing metaphor based language without establishing a clear relationship with classical paradigms (not using the classical optimization terminology etc.) etc. for example, Camacho-Villalón et al. discussed in detail that Intelligent Water Drops algorithm of Shah-Hosseini is essentially a simplification of ideas and mechanisms of ant colony optimization based algorithms. Still many researchers are constantly looking for new natural phenomena to mimic in order to propose new meta-heuristic algorithms rather than improving the field of optimization and existing meta-heuristic algorithms. One of the motivations behind continuously presenting new meta-heuristic algorithms may also be related to the No Free Lunch Theorem (NFLT). NFLT is a theoretical finding that suggests all meta-heuristic algorithms perform equally well when their performance is averaged over all possible problems. This endorsement may supposedly encouraged researchers to search for competitive meta-heuristic algorithms for tackling complex optimization problems. However, there are many possibilities to improve existing meta-heuristic algorithms. Some possible improvement areas can be listed as follows: developing effective/intelligent step-size determination heuristics; developing effective/intelligent search direction determination mechanisms; developing effective/intelligent constraint handling procedures; devising new solution coding/encoding procedures for tricky/complicated discrete optimization problems; devising new, practical and effective procedures for handling stochastic and fuzzy optimization problems; developing new, practical and effective procedures for handling dynamic optimization problems; developing novel and effective procedures to incorporate learning algorithms into meta-heuristics; devising novel and effective procedures to incorporate memory management procedures into meta-heuristics; devising novel and effective procedures for parallelization of population based meta-heuristics; developing novel

and effective procedures for integrating meta-heuristics with mathematical programming algorithms etc. Research on meta-heuristics algorithms should focus more on improving and/or elaborating the existing/promising meta-heuristics. In this study we show a possible way to improve a well-known meta-heuristic algorithm that is known as Firefly Algorithm (FFA). FFA is one of the most popular swarm intelligence based meta-heuristic algorithm. In a literature review paper it is listed as the 5th mostly highly cited meta-heuristic algorithm. FFA was first proposed by Xin-She Yang in 2008 that is based on the flashing patterns and behaviors of fireflies. Complexity of the standard FFA is high. It is O(n2*t). So, it is possible focus on reducing this complexity without worsening its performance. Such an attempt is presented in this study by incorporating superposition principle of Weighted Superposition Algorithm of Baykasoglu into attraction mechanism of FFA. Superposition is determined by making use of the superposition principle from physics such that the weighted sum of the agents' position vectors forms the superposition vector, where weights are proportionally determined according to agents' current position fitness's. It is known from physics that artificial agents can be attracted towards superposition while choosing their own move directions. Superposition can be considered as a single point (may be more) that should be determined to direct search agents (solution vectors) towards this point during the exploration of the search space in an optimization procedure. By incorporation superposition principle into FFA, complexity of FFA can be reduced to O(n*t) without worsening its performance. The CEC'2020 benchmark functions are used to compare the performance of the classical FFA and FFA with superposition, S-FFA. In all test problem S-FFA provided better solutions in considerably shorter computational time. In larger size problems computational time is five times shorter.

Keywords: *meta-heuristics, firefly algorithm, weighted superposition attraction algorithm, analysis of meta-heuristics*

Abstracts

A Comparison Study for the Outlier Detection Methods in Regression

Hakan Savaş Sazak
Ege University
hakan.savas.sazak@ege.edu.tr

Berfin Saraçoğlu
Vestel Company
berfin.saracoglu@vestel.com.tr

ABSTRACT

Outliers are observations that deviate significantly from other observations in the same sample. They affect statistics that represent the data and the results obtained from the models. This can cause a negative impact on the decision-making process. Therefore, it is very important to detect outliers in regression analysis. In this study, it is aimed to compare four methods that are frequently used for the detection of outliers in regression models with simulation and real life data applications. Cook's distance, Studentized residuals, DFFITS and Hadi's measure methods were compared in terms of the percentages of correct and false detection of outliers. A very extensive simulation study was performed for this comparison for various models and sample sizes. From the simulation outputs, it was concluded that the compared methods were more successful in detecting outliers in the x direction than outliers in the y direction. It was seen that Hadi's measure was successful in detecting outliers in the x direction, DFFITS and Hadi's measure were successful in detecting outliers in the y direction compared to other methods. As a result of the two real life applications given at the end of the study, it was seen that Hadi's measure is the only method that can detect border outliers as well as extreme outliers in outlier detection. In addition, it was observed that DFFITS was the method least affected by the masking effect among the four methods compared while when the swamp effect was examined, it was concluded that all four methods were unsuccessful.

Keywords: *general linear regression model, influential observation, masking and swamping effects, outlier, outlier detection*

A Comparison Study for the Robust Time Series Methods

Işıl Diker
Ege Üniversitesi
isildikerr@gmail.com

Hakan Savaş Sazak
Ege Üniversitesi
hakan.savas.sazak@ege.edu.tr

ABSTRACT

Parameters are usually estimated by the least squares (LS) method in the pure autoregressive (AR) process. The LS method can be used when some assumptions are fulfilled. The mean and covariance of the random variables of the random process over time should be zero and the variance should be constant. It is not always possible to provide these assumptions in real-life data. If these assumptions are not met, the use of the LS method in parameter estimation does not yield reliable results. There may also be outliers in a real-life data set. If there are outliers in the time series data, there may be big amount of losses in the efficiency of the estimators provided by the LS method. In this study, the efficiencies of the LS estimators, M (Huber) estimators, M (Tukey) estimators, S estimators and MM estimators in the parameter estimation of AR(1) model were compared using Monte Carlo simulation, for two cases, when the error term is standard normally distributed without outliers and for the Dixon's outlier model. In the simulations, the efficiency of the aforementioned estimators for the varying values of phi and sample size is examined. Determining which parameter estimation method is more robust and efficient against deviations from model assumptions is of great importance for the accuracy of parameter estimation. Accurate estimation of the parameters ensures that the predictions to be made for the parameters of the established model will be more reliable. We also provide real-life data applications and the comparisons of the aforementioned estimators by using several criteria.

Keywords: *least squares, robust, time series*

A Hybrid Approach for Hyper-parameter Selection in Text Classification for Covid-19 tweets

Cansu Doğan
Hacettepe University
cansudoogan@gmail.com

Çağdaş Hakan Aladağ
Hacettepe University
aladag@hacettepe.edu.tr

ABSTRACT

Coronavirus (Covid-19), which is infectious disease cause of SARS-CoV-2, has affected around the world in record time. It emerged in 1 December 2019 in Wuhan, China. Covid-19 came upon all over the world. It is aimed to analyze that the effect on people in this study. It was analyzed tweets about Covid-19 from Kaggle. It was used to text classification method Natural Language Processing (NLP) which is a branch of Artificial Intelligence (AI) For this analysis. Text classification is process of categorizing which the texts assign to pre-defined tags automatically. The dataset consist of 44955 observations and 6 features. NLP is a field of common interest of computer science and artificial intelligence. It is aimed that human languages and mimics are understood by computers. Computers cannot understand the words. Therefore, it processes data by converting to vectors in the form of numbers. Text is convert to numerical values to can be analyze to unstructured data. In this reason, data pre-processing is important in NLP applications. Converting uppercase to lowercase, punctuation removal, stopwords removal, frequent words removal, rare words removal, stemming, lemmatization, emoji, hashtags and URLs removal, removal of HTML tags, spelling correction, tokenization, normalization and parts of speech tagging are pre-processing steps for extracting features. The step followed after the text preprocessing process is the model classification. It was used 3 classification methods for classification of text. for classification. In the literature, a very important problem in machine learning algorithms is selection of hyper-parameters. There are many combinations that can be tried hyper-parameter. There are no general rules for making selection of hyper-parameters but there are some approaches. Grid search, random search can be given as an example. Aladağ(2019) proposed a hybrid approach based on statistical and machine learning to solve the architectural selection problems of ANN in his study. In 2021, Doğan extended this approach to hyper-parameter selection in deep neural networks and used it by applying both linear regression analysis and nonlinear regression analysis. In this study, hyper-parameter selection was made using both linear regression analysis and nonlinear regression analysis based on Doğan (2021) study for hyper-parameter selection.In this study, hyper-parameter selections of these algorithms were made with 3 machine learning algorithms.And by comparing the results, the model that gives the best performance value was determined.

Keywords: *covid-19, hyper-parameter selection, machine learning, nlp, statistical learning, text classification*

A Multiple Objective Optimization Model for a Novel Capability-Based University Course Timetabling Problem: A Case Study at Deu Industrial Engineering Department

Kemal Subulan
Dokuz Eylul University
kemal.subulan@deu.edu.tr

Ahmet Gürsaç
Dokuz Eylul University
gursac.ahmet@ogr.deu.edu.tr

ABSTRACT

The educational content of the undergraduate departments in all of the universities all over the world is mainly based on the knowledge, skills, and capabilities that the graduates will require in their business life. Many graduates will use these skills or capabilities, which are usually gained from the compulsory and elective courses in the universities to meet the requirements and specifications in their business life. Therefore, the university course timetabling which is NP-hard and a class of scheduling problems is of great importance to training well-equipped individuals in both business and daily life. However, none of the available studies in the literature have considered the optimal distribution of the capabilities/skills over the curriculum that are essentially gained from compulsory and opened elective courses. Based on this motivation, this research first introduces a novel capability-based course timetabling approach, which provides a wide variety of courses with a maximal capability set to the students during their university education. To do this, several learning outcomes and course contents are first clustered to obtain basic capabilities that the students should acquire. Then, a mixed-integer non-linear programming model with both hard and soft constraints is developed to provide appropriate distribution of these capabilities over the whole curriculum and also to gain the maximal capability for the students in different classes. By making use of the proposed capability-based course timetabling approach, in addition to gaining the maximum number and variety of capabilities, it is also intended to assign the opened courses to the proper time slots by minimizing the distance between time differences of the given courses by each lecturer. Moreover, another objective is also formulated as an aggregated penalty function to minimize the weighted deviations in all the soft constraints. To handle these conflicting objectives simultaneously and to produce compromise course timetables for the university department managers, a fuzzy goal programming approach with different importance and priorities is applied. Finally, the validity and applicability of the proposed capability-based course timetabling approach are also demonstrated by a real-life case study at DEU Industrial Engineering department.

Keywords: *university course timetabling problem, capability-based scheduling, mixed-integer non-linear programming, multi-objective optimization*

A New Weighted Independence Test Based on Smooth Estimation of Kendall Distribution Function

Selim Orhun Susam

Burcu Hüdaverdi
Dokuz Eylul University
burcu.hudaverdi@deu.edu.tr

ABSTRACT

The independence assumption of the random variables provides a valuable inference for further statistical analysis of the involved variables in many of the research areas. In this study, we propose a weighted test of independence for d-variate Archimedean Copula families. The proposed test statistic is based on a weighted Cramér-von Mises distance that is constructed using Bernstein empirical Kendall distribution function and arbitrary weight function. We examine the power and the size of the new weighted test and we compare it with the non-weighted nonparametric tests of independence through a Monte Carlo simulation study. The simulation results indicate that both the choices of the weights and the polynomial degree have a significant effect on the test power. The procedure is applied to a real dataset on chemical elements.

Keywords: *cramér-von-mises distance, bernstein polynomial, kendall distribution, archimedean copula*

A Note on the Moments of Sample Minimum of Order Statistics From a Geometric Distribution

Aslıhan Demir
Firat University
a.bilici128@gmail.com

Ayşe Metin Karakaş
Bitlis Eren University
aysekarakas5767@gmail.com

Sinan Çalık
Firat University
scalik@firat.edu.tr

ABSTRACT

In this study, the moments of sample minimum of order statistics from a geometric distibution are obtained with the help of factorial moment generating function, characteristic function and cümülant generating fuction. Using these moments, the means and variance are given as algebraic and numerical.

Keywords: *order statistics, sample minimum, expected value, variance, geometric distribution, moment, cumulant generating function.*

A Statistical Analysis of Usability of a Learning Management System

Ceylin Unal
Department of Industrial Engineering The Graduate School of Natural and Applied Sciences Dokuz Eylül University
ceylin.unal@ogr.deu.edu.tr

Adil Baykasoğlu
Department of Industrial Engineering Dokuz Eylul University
adil.baykasoglu@deu.edu.tr

Burcu Felekoğlu
Department of Industrial Engineering Dokuz Eylul University
burcu.felekoglu@deu.edu.tr

ABSTRACT

With the spread of the concept of distance education after Covid-19 pandemic worldwide, the usage of Learning Management Systems (LMS) has become more widespread. There are many LMS that have been adopted by many educational organizations and cover a large part of their educational activities. However, selection of the most appropriate LMS is a complex Multiple Criteria Decision Making (MCDM). Usability is one of the most important concepts in evaluating this selection. Real life case studies using statistical analyzes and MCDMs in the usability evaluation of LMS are very rare. Based on this inference, a systematic methodology is developed for the usability evaluation of a LMS, using statistical analyzes and MCDMs. First of all, a hierarchical structure consisting of main criteria and sub-criteria connected to these criteria is established. Based on these criteria, a questionnaire is created in line with statistical rules and directed to 3 types of system users namely; graduate students, undergraduate students and faculty members in order to evaluate these usability criteria. Data obtained with the questionnaire are analyzed statistically, using One-Way-ANOVA and Games-Howell post-hoc tests. According to this analysis, it is determined that the collected data are reliable and it is concluded that the usability evaluation of each evaluation group should be done separately. Criterion weights are found with the Analytic Hierarchy Process, and then usability evaluation of the LMS is done with the Axiomatic Design Procedure. With this study, it is aimed to bring an example case to the literature about usability evaluation of a software system through integrating the statistical analysis with multi-criteria decision-making methods. It is concluded that the proposed approach is easy to apply to practical circumstances and it is able to quantify usability of the LMS.

Keywords: *learning management systems, usability evaluation, statistical analysis, axiomatic design, decision making methods*

An Alternative Quantile Regression Model for the Unit Response Based on the New Unit Probability Distribution

Berrin Gültay
Çanakkale Onsekiz Mart University, Department of Statistics
berringultay@comu.edu.tr

Mustafa Çağatay Korkmaz
Department of Measurement and Evaluation, Artvin Çoruh University
mustafacagataykorkmaz@gmail.com

ABSTRACT

The aim of this study is to propose a new unit distribution and its quantile regression modeling to model the percentages and proportions with their applications. In this context, this work has introduced a new modified distribution defined on the (0,1) unit interval. It has two model parameters. The proposed model has been obtained by a transformation of another unit distribution. The applied transformation has supplied the new shapes to ordinary unit distribution model in terms of both probability density and hazard rate functions. Some distributional properties such as density shape, the shape of the hazard rate function, ordering, quantile function, moments, and order statistics of the newly defined distribution are derived. The different estimation methods have been pointed out for the model parameters' estimations. An associated unit quantile regression model has been presented based on the proposed distribution. The residual analysis has been discussed about the model's error via two special residual formulations. Two simulation studies have been separately performed for the proposed distribution and regression model. Based on the different quantile levels, Real data applications have been applied in order to see the applicability of the proposed regression and unit distributions. Applications show that the proposed models have better modeling abilities than competitive models under some comparison criteria. The results also indicate that covariates are statistically significant at any usual significance level for the unit response. The residual analysis has been worked to check whether the regression model is suitable. Finally, we say that the proposed models are strong alternative models to ordinary models.

Keywords: *quantile regression, unit distribution, estimation*

An Optimal Combination of Proportional and Stop-Loss Reinsurance Under Dependent Claim and Random Insurance Premium

Suci Sari
Statistics Research Division - Institut Teknologi Bandung, Indonesia
sucifratama@s.itb.ac.id

Arief Hakim
Statistics Research Division - Institut Teknologi Bandung, Indonesia
30121003@mahasiswa.itb.ac.id

Khreshna Syuhada
Statistics Research Division - Institut Teknologi Bandung, Indonesia
khreshna@math.itb.ac.id

ABSTRACT

This research investigates an optimal reinsurance policy using the risk model under dependent claim and insurance premium. We assume that the insurance premium is random. Moreover, we also assume that an insurance company has only one line of business and the dependence structure is described by bivariate distribution and copulas. We use Sarmanov's bivariate exponential distribution and the FGM copula-based bivariate exponential. The reinsurance policy considered in this paper is the combination of proportional and stop-loss reinsurance. Since the reinsurer shares the risk, the insurer should pay an additional cost in the form of a reinsurance premium to the reinsurer. We assume that the reinsurance premium is fixed and is charged by the expected value principle. The structure of optimal reinsurance is from the insurer's perspective. The objective of this research is to formulate optimization criteria used for determining an optimal form of the combination of proportional and stop-loss reinsurance for the insurer. Specifically, with a constrained reinsurance premium, we use the minimization of the risk exposure of the insurer quantified by Value-at-Risk (VaR). The quantification of the insurer's risk exposure is carried out using a risk measure VaR of the insurer's total cost, which should be as low as possible for choosing the proportion and the retention limit. Finally, in order to determine the proportion and the retention limit, some numerical examples are given to illustrate the theoretical results by using Toolbox Nminimize in Mathematica. We also show the effects of copula parameter and the probability of claim occurrence on the VaR of the insurer's total cost.

Keywords: *stop-loss, proportional, random insurance premium, copula, retention*

Ata Forecasting Method as an Alternative to Exponential Smoothing in Forecast Combination

Uğur Baran Yapar
Dokuz Eylül University
baranyapar@hotmail.com

İdil Yavuz
Dokuz Eylül University
idil.yavuz@deu.edu.tr

ABSTRACT

An easy way to improve forecast accuracy is to use several different methods on the same time series, and to average the resulting forecasts. Forecast combination have been found in empirical studies to produce better forecasts than methods based on the best individual forecasting model. Combinations of some basic methods such as ARIMA, ETS, THETA have been used widely in many empirical studies. In this study, the combination performance of the Ata forecasting method will be investigated as an alternative to exponential smoothing.

Keywords: *time series, forecast combination, ata method, forecast accuracy*

Bayesian Analysis of Repeated Measures: An Application

Fatma Yardibi
Akdeniz University
fatmayardibi@gmail.com

Mehmet Ziya Fırat
Akdeniz University
mzfirat@akdeniz.edu.tr

ABSTRACT

Repeated measurements consist of measurements taken on the same subject at different times or by measuring the values of the same subject determined under different situations. The strength of repeated measurement analysis is that it is the only design form in which it is possible to obtain information on individual variations. For this reason, it is often preferred. However, the most important feature is a variance-covariance structure between observations obtained from the same units over time. Disregarding the variance-covariance structure between these observations from the same unit can lead to erroneous and inappropriate interpretations. Therefore, it is important to model repeated measures' variance-covariance structure adequately. For this reason, there is a need to represent the repeated measures data with special statistical models and their complementary analysis methods. General linear mixed models provide more flexibility in modeling variance-covariance structures for repeated measurement data and adequately explain the time-dependent correlations of units. Bayesian Monte Carlo methods offer a more effective alternative to statistical methods such as Maximum Likelihood (ML) and Restricted Maximum Likelihood (REML). In this study, live weight data of lambs were used. Body weight data were modeled by the Bayesian method for three different covariance structures, and their solutions were obtained. In Bayesian analysis, noninformative normal priors with large variance were determined for fixed effects such as regression coefficients. A normal prior distribution with a mean of 0 and a variance was assigned for random effects. Noninformative general distribution was determined for the variance component. The Wishart distribution is assigned for the variance-covariance matrix. Bayesian analyses were performed using the SAS program's PROC MCMC procedure. The model selection from the models created for unstructured (UN), compound symmetry (CS), and first-order autoregressive AR(1) covariance structures were made with the Bayesian model comparison criterion Deviance Information Criterion (DIC). This criteria, which uses the posterior distributions of the models obtained by the MCMC method, is useful in Bayesian model selection and used in many studies. Posterior statistics are based on 5000 Gibbs sampling. The Bayesian method is computationally applicable and gives a more reasonable solution to interpretation-based problems than probability-based estimation methods.

Keywords: *bayesian analysis, repeated measures, sas, gibbs sampling*

Bayesian Analysis of Turkish Income and Living Conditions Data, Using Clustered Longitudinal Ordinal Modelling With Bridge Distributed Random-Effects

Özgür Asar
Afiniti
ozgurasarstat@gmail.com

ABSTRACT

This paper is motivated by the panel surveys, called Statistics on Income and Living Conditions (SILC), conducted annually on (randomly selected) country-representative households to monitor EU 2020 aims on poverty reduction. We particularly consider the surveys conducted in Turkey within the scope of integration to the EU. Our main interests are on health aspects of economic and living conditions. The outcome is self-reported health that is clustered longitudinal ordinal, since repeated measures of it are nested within individuals and individuals are nested within families. Economic and living conditions have been measured through a number of individual- and family-level explanatory variables. The questions of interest are on the marginal relationships between the outcome and covariates that we address using a polytomous logistic regression with Bridge distributed random-effects. This choice of distribution allows us to directly obtain marginal inferences in the presence of random-effects. Widely used Normal distribution is also considered as the random-effects distribution. Samples from the joint posterior densities of parameters and random-effects are drawn using Markov Chain Monte Carlo. Interesting findings from the public health point of view are that differences were found between the sub-groups of employment status, income level and panel year in terms of odds of reporting better health.

Keywords: *bridge distribution, latent variables, multi-level data, repeated measures, self-reported health*

Bibliometric Analysis by Using CiteSpace

Fatma Yardibi
Akdeniz University
fatmayardibi@gmail.com

ABSTRACT

The number of academic studies, such as articles and reviews, is increasing day by day. It is getting harder to extract a qualified literature summary from these many scientific studies. Bibliometric studies are developed to provide a comprehensive overview of the literature on the identified field and its applications to academicians and researchers. In other words, through bibliometric analysis and visualization, we can explore the academic domain of a field of knowledge and understand what questions researchers are trying to answer and what methods they have developed for this purpose. Bibliometrics is the quantitative analysis of research publications using mathematical and statistical methods. This method uses a defined set of metrics to assess published research's output, impact, and trends. Bibliometric analysis is used to qualitatively and quantitatively analyze the effects of journals, institutions, research groups, individual researchers, or countries. One of the main components of bibliometric studies is co-citation analysis. Researchers use the co-citation analysis to recognize if any new theory stems from existing literature. With the development of information technology, bibliometric analysis can create network images, allowing for more visualized information than keywords. CiteSpace is one of the most popular software tools in bibliometric studies to analyze co-citation networks. Thanks to CiteSpace, it becomes easier to understand evolution, development, and trends in a particular scientific field by defining intellectual milestones and dynamically visualizing citation networks. Visualization information maps used for bibliometric analysis in CiteSpace consist of nodes and links. Structural measures such as mean silhouette score, modularity Q index, and betweenness centrality value are used to evaluate the network's structural quality. Citation burst is another crucial metric that shows hot points in a field of research. A burst term is defined as a keyword, author, or institution that appears with an abrupt change in frequency within the literature during a specific period. A general evaluation is completed over the network structure, which results from metrics and visualization analysis. Thus, it is thought that bibliometric studies will guide researchers in evaluating the importance and impact of published academic studies and pave the way for new studies.

Keywords: *bibliometric analysis, citespace, network analysis*

Classification With Bernstein Copula Through Machine Learning Process

Tolga Yamut
Dokuz Eylul University

Burcu Hüdaverdi
Dokuz Eylul University
burcu.hudaverdi@deu.edu.tr

ABSTRACT

In machine learning, modeling dependence holds an essential role. It is often difficult to construct a joint environment for the distributions, however, copulas simplify the process by separating the learning of the marginal distributions from the learning of the multivariate dependence structure. In this study, a Bernstein copula model is proposed as a discrimination function to classify the given data in machine learning process. The dependence structures among features are constructed by the Bernstein copulas. The performance of the Bernstein copula models on supervised learning algorithm is investigated via a comprehensive simulation study. Some improvements for distribution calibration are also proposed to obtain efficient and flexible solutions. The method is compared with the parametric copula approaches. Also, an empirical application is considered where the classification performance is additionally investigated with the convex Bernstein density functions. It is observed that the combination of distributional information proves to be a useful tool in the discrimination process and the convex Bernstein density approach improves the discrimination ability.

Keywords: *machine learning, copula, bernstein distribution*

Comparing Forecast Accuracies of Exponential Smoothing and Ata Method in Demand Forecasting

Miray Betül Yılmaz
Dokuz Eylül University, The Graduate School of Natural And Applied Sciences

Özgül Vupa Çilengiroğlu
Dokuz Eylül University, Faculty of Science
ozgul.vupa@deu.edu.tr

ABSTRACT

Today, time series methods are frequently used in demand forecasting in production planning. When the performances of the time series approaches used in recent years are compared, it is observed that especially the demand trends of the customers change depending on the various components of the market and seasonality. Especially in today's conditions, since the use of resources is important for cost and product planning in production planning of enterprises, it is important to organize and structure them with determined targets. Thanks to the use of resources with the determined targets, the financial and commercial values of the enterprises in the sector become stable or increasing. In this study, the performance of the exponential smoothing method, which is frequently used among the time series approaches for demand forecasting in production planning in the refrigeration sector, and the performances of the newly developed Ata methods in recent years have been compared. For performance comparison, the sales volumes of refrigerator refrigerators in the refrigeration sector in Manisa between the years 2003-2022, as well as dollar, TUFE, UFE and consumer confidence index data were used.

Keywords: *time series, ets, ata method, farecast accuracy*

Comparison of Machine Learning and Traditional Approaches for Automobile Insurance Pricing

Müge Yeldan
Hacettepe University
mugeyeldan@hacettepe.edu.tr

Uğur Karabey
Hacettepe University
ukarabey@hacettepe.edu.tr

ABSTRACT

Pricing of automobile insurance policies is an important challenge for many non-life insurance companies. Ineffective premium pricing means not only that premium rates are too high or too low, but also that premiums are not differentiated between different risks or that the distinctions are not appropriate. This can lead to a large number of adverse selections. Therefore, insurance companies need to carry out fair and comprehensive pricing, which can effectively prevent the problem of adverse selection, keep the insurance industry in healthy competition and promote the development of the insurance industry. Since the early 1990s, Generalised Linear Models (GLMs) have been regarded as the traditional method of ratemaking. GLMs extend linear regression techniques by allowing a distribution from the exponential family for the response variable. In other words, GLMs assume a linear relationship between the known variables and the unknown variable. In some cases, linear relationship may not be sufficient to model complex behaviors. In this case, nonlinear transformations and interactions between variables should be considered to reflect reality. For this reason, machine learning technics become popular recently. In this study, we compare the traditional approach and machine learning approaches used in pricing non-life insurance. Random forests, a decision tree based model, and neural network models are used as the machine learning approach, while generalized linear models are used as the traditional approach.

Keywords: *automobile insurance pricing, generalized linear models, machine learning, random forests, neural networks*

Comparison of the Fit of Cox Regression and Random Survival Forest Model With the Help of Simulation Data

Tuğçe Paksoy
Dokuz Eylul University
paksoytugce35@gmail.com

İdil Yavuz
Dokuz Eylul University
idil.yavuz@deu.edu.tr

ABSTRACT

Survival analysis, which is a widely used and important subject in applied statistics, is related to the analysis of the time until a certain event occurs. An important feature that distinguishes this analysis from other statistical analyzes is that it contains censored observations. Censored data is frequently observed in real data sets and for this reason artificial data produced with the help of simulation for survival analysis generally involves censoring. Various simulation scenarios have been tried in the literature to generate artificial data for survival analysis. The most commonly used approaches either assume a cox proportional hazards model [1] or parametric distributions. When data is produced assuming parametric distributions, the most important assumption of the Cox regression is broken. This assumption is related to the fact that the shape of the baseline hazard function should not conform to a particular distribution. Therefore, Kropko et al. proposed a method to generate survival data via simulation which they provided in the "coxed" package in R [4]. The sim.survdata function [3] in this package uses cubic splines for the baseline hazard function and the data was produced in accordance with the proportional hazards assumption. In this study, survival data were obtained by using the sim.survdata function in the "coxed" package. Different simulation scenarios were employed in the study. Different sample sizes and censorship rates were used in the simulations, and trials where the proportional hazards assumption was not provided were also included. The aim of the study is to compare the concordance values obtained by using various Random Survival Forest models (RSF) [2] and Cox regression. It was observed that the Cox regression fits the model better than RSF if the Cox proportional hazards assumption is met however the performance of RSF improves as the sample size increases and it appears to behave more robust to higher levels of censorship.

Keywords: *sim.survdata, cox regression, random survival forest*

Deep Fuzzy Functions Approach for Time Series Forecasting

Ali Zafer Dalar
Giresun University
ali.zafer.dalar@giresun.edu.tr

ABSTRACT

In the literature, fuzzy set theory-based approaches are used as an alternative to probabilistic models for time series forecasting. Unlike probabilistic models, there is a fuzzy approach to uncertainty in methods based on fuzzy set theory. Fuzzy inference systems based on fuzzy set theory work with linguistic variables similar to the inference mechanism of the human brain. The purpose of methods based on fuzzy sets for time series forecasting is to produce better interval and point estimates than probabilistic models. Fuzzy set theory-based inference systems require to use rule-based systems. The advantage of the type-1 fuzzy functions approach proposed by Türkşen (2008) over other fuzzy inference systems is that it does not require rule-based systems. The type-1 fuzzy functions approach, which has been used in time series forecasting in the last ten years, has not yet been used for forecasting in the literature as a system based on deep learning. In recent years, it has been observed that deep artificial neural networks provide better forecasting performance than other artificial neural networks. Although long short-term memory (LSTM) method is one of the deep artificial neural network methods, it is a method that has recently started to be used in the time series literature. In this study, the fuzzy functions approach based on LSTM neural network will be proposed in time series forecasting, and the forecasting performance of the approach will be compared with other time series forecasting methods on real-world time-series datasets as BIST 100, FTSE 100, NASDAQ 100 and TAIEX.

Keywords: *forecasting, fuzzy functions approach, lstm, deep learning*

Dependent Metaverse Risk Forecasts Using Heteroskedastic Models and Ensemble Learning

Venansius R. Tjahjono
Institut Teknologi Bandung
20821026@mahasiswa.itb.ac.id

Arief Hakim
Institut Teknologi Bandung
30121003@mahasiswa.itb.ac.id

Khreshna Syuhada
Institut Teknologi Bandung
khreshna@math.itb.ac.id

ABSTRACT

The metaverse is an interconnected world, or artificial platform, between our actuality and our digital reality. Metaverse allows users to purchase digital items in any games platform and online services through metaverse cryptocurrency. Behind this practicality, crypto assets in the metaverse could cause a risk. The risk comes from the negative return that leads us to a loss. This study provides the dependent metaverse risk forecast between Decentraland (MANA) and Theta Network (THETA). MANA is one of the leading metaverse cryptocurrencies that reshapes how the world sees virtual assets. Furthermore, there is a fact that the Theta Network supports Decentraland with its security system so that they have an implicit association. To model such dependency, we propose the use of copula through two classes of predictive models, that is, heteroskedastic models and ensemble learning-based models. The heteroscedastic models that are employed are GARCH(1,1), EGARCH(1,1), and GJR-GARCH(1,1). The ensemble learning-based model incorporates XGBoost, LightGBM, and CatBoost algorithms. We also present the use of modified aggregate risk (AggM) obtained from the linear combination of aggregate Value-at-Risk (AggVaR) and aggregate expected shortfall (AggES) involving a specific weight to determine the optimal risk forecast. In the simulation, we discovered that heteroskedastic models perform better on AggM than other risk measures. Additionally, the ensemble learning-based model has the capability to lower the predicted risk measure from AggES such that it is suitable to do the risk forecast. Other statistical analyses, such as Christoffersen's test and coverage probability, are also provided in the final section to evaluate the validity and accuracy of the risk forecasts, respectively.

Keywords: *metaverse cryptocurrency, risk forecast, heteroskedastic models, ensemble learning, modified aggregate risk*

Determinants of Carbon Dioxide Emissions for Turkey: Quantile Regression Approach

Çiğdem Koşar Taş
Cukurova University
ckosar@cu.edu.tr

Sibel Örk Özel
Cukurova University
sork@cu.edu.tr

Dilek Veysikarani
Cukurova University
dveysikarani@cu.edu.tr

ABSTRACT

With the increase in the world population and accordingly the increase in the energy consumption of the countries, the ecological balance is deteriorated and problems such as environmental pollution arise. Especially the use of fossil fuels, which are non-renewable energy sources, cause high levels of carbon dioxide (CO2) emissions, which negatively affected the environment. This situation, which is very threatening for the world, has made the factors affecting CO2 emissions one of the important research topics, especially in recent years. In this context, the variables that are thought to have an impact on CO2 emissions have been analyzed using various methods for different countries and country groups on the basis of different periods. From this point of view, in this study, the non-renewable energy (fossil fuel) consumption, renewable energy consumption, population density and gross domestic product (GDP) variables, which are thought to have an impact on CO2 emissions, are discussed and their relationship with CO2 emissions is examined for Turkey. For this purpose, quantile regression is applied using the data set for the years 1990-2020, which is the widest data range available in the variables discussed. Quantile Regression, proposed by Koenker and Basset (1978), is a method that can also be used when the assumptions of the linear regression model are violated. This method, which is developed based on the median regression, gives robust results. By applying the quantile regression method to the data discussed in the study, the results for different quantile slices are revealed and the variables affecting (increasing and decreasing) CO2 emissions are determined and compared with the relevant literature. The obtained results are considered to be important as they will contribute to the environmental policy pursued by our country.

Keywords: *carbon dioxide emissions, quantile regression, energy, environment*

Estimating of the Distribution Function Under Ranked Set Sampling With Unequal Probabilities

Yusuf Can Sevil
The Graduate School of Natural and Applied Sciences, Dokuz Eylul University
yusufcansevil92@gmail.com

Tuğba Özkal Yıldız
Department of Statistics, Faculty of Science, Dokuz Eylul University
tugba.ozkal@deu.edu.tr

ABSTRACT

Ranked set sampling (RSS), suggested by McIntyre (1952), is a popular sampling strategy when the measurements of the sample units are relatively difficult (expensive and/or time-consuming). The estimation of distribution function has received considerable attention in the literature of RSS. Because many practical problems involve estimation of distribution function from experimental data. Many authors have proposed empirical distribution functions (EDFs) based on RSS and its modifications (see, for example, Stokes and Sager, 1988; Samawi and Al-Sagher, 2001, Nazari et al., 2016 and Zamanzade, 2019). For finite population setting, there are a few studies on estimation of distribution function (see, for example, Sevil and Yildiz, 2017; Yildiz and Sevil, 2018, 2019). In this study, design-based estimators for distribution function have been developed using RSS designs (level-0, level-1 and level-2). Some of their asymptotic properties have been investigated. Theoretical and numerical results show that the level-2 sampling design provides a more efficient EDF estimator than its counterparts of level-0, leve-1 and simple random sampling.

Keywords: *ranked set sample, design-based estimators, empirical distribution function*

Estimation of the Price Movement Direction of Cryptocurrencies With Stochastic Processes: An Application for Bitcoin

Özgür Danışman
Dokuz Eylül Üniversitesi
ozgur.danisman@deu.edu.tr

ABSTRACT

Stochastic processes are frequently used in many different areas such as stock market, time series, audio and video processing. They are defined as random processes over time whose outcomes cannot be predicted in advance. Markov processes are a special case of stochastic processes with different assumptions. The process makes transitions between the states and, if the conditional transition probability of the next state depends only on the current state, it is called the Markov process. The states of Markov processes are observable and the observer can record the state sequence. In some real life problems, the states that occur during the process cannot be observed from outside (hidden). If a stochastic modeling is aimed in this assumption, "hidden Markov models" are often used. Hidden Markov models can be considered as a two-dimensional stochastic process. The first stochastic process is related to the states and it is modeled by finite-state Markov chains. The second stochastic process, on the other hand, is related to the observations emitted from the states of the Markov chain. In this study, a hidden Markov model is presented on the probabilistic estimation of the movement directions of cryptocurrency prices relative to the previous day, week and month. An application has been presented on the Bitcoin cryptocurrency, which ranks first in terms of market volume and popularity. In the cryptocurrency market, it is generally not known in advance what kind of trend there is compared to the previous period (day, week or month) or the speculative movements that may affect the market. Therefore, these factors are modeled with a hidden discrete Markov chain, called the “market condition”. The market condition is classified into three classes as “good”, “bad” and “average” and these classes constitute the states of the Markov chain. In addition, the observations which are emitted after each state transition are defined as "increase", "decrease" and "no movement" compared to the previous period. Models are constructed considering the daily, weekly and monthly changes. In addition, probabilistic estimation of the movement direction of Bitcoin cryptocurrency prices according to the previous day, month and year are presented.

Keywords: *stochastic processes, markov processes, cryptocurrency, estimation, statistical modeling*

Estimation of the Solar Plant Power with PV Modules Subject to Failure

Melek Esemen
Dokuz Eylul University
melek.esemen@gmail.com

Selma Gürler
Dokuz Eylul University
selma.erdogan@deu.edu.tr

Gözde Yazgı Tütüncü
İzmir University of Economics
yazgi.tutuncu@ieu.edu.tr

ABSTRACT

The power produced by a solar power plant consisting of PV modules depends on solar radiation. In this study, the theoretical distribution function of the power generated by a solar plant is derived. Then, the characteristics of the power produced by the solar power plant are estimated using the data obtained from the PVGIS interactive tool for a specific location in Izmir, Turkey. The typical meteorological year (TMY) data of solar irradiation is used in the analysis for the given location for certain months of the year, covering the period (2007-2016). Under the assumption that the PV modules are subject to failure, the mean power generated by a PV system is estimated using the distribution function of the power. Also, the optimum number of PV modules to be installed in the plant was obtained.

Keywords: *solar power, distribution function, reliability*

Evaluation of the Success Status of University Students on the Basis of Faculties in Formal and Distance Education

Burçin Öner
Bilecik Seyh Edebali University Bozuyuk Vocational School Department of Finance, Banking and Insurance
burcin.oner@bilecik.edu.tr

Yüksel Öner
Ondokuz Mayis University Faculty of Arts and Sciences Department of Statistic
yoner@omu.edu.tr

ABSTRACT

The Covid 19 epidemic has affected the whole world, and therefore our country, in every aspect. In the last few years, almost every segment, especially the health sector, has experienced this process in the deepest way and continues to do so. It is seen that the execution of education and training activities is also affected by this process and that countries in the world carry out educational activities with different applications. During the Covid 19 epidemic in Turkey, the execution of education and training activities in higher education continued according to the decisions taken by the Higher Education Council. Until this period, including the Fall Semester of the 2019-2020 academic year, formal education was being implemented in higher education institutions. Due to the spread of the Covid 19 epidemic and the aggravation of the effects of the epidemic, universities started the distance education process in our country, as in almost all countries, as of the Spring Semester of the 2019-2020 academic year, and this process continued throughout the 2020-2021 academic year. As of the 2021-2022 academic year, the hybrid system (40% distance education and 60% formal education) has started to be implemented. In this study, it was aimed to evaluate the success status of the students who enrolled in different faculties in the 2017-2018, 2018-2019, 2019-2020 and 2020-2021 academic years at two state universities affiliated to the Council of Higher Education, taking into account their student transcripts. For this purpose, the real data created from the student achievement certificates obtained from the universities in question according to the sampling plans created for the faculties discussed in the study were used. The data were analyzed according to the faculties by applying the repeated measures analysis by using the periodic weighted averages of the students for the specified academic years. Based on the findings, the success of the students was evaluated comparatively both within the universities in question and between universities in terms of formal education and distance education.

Keywords: *formal education, distance education, hybrid education, repeated measurement analysis, student transcript*

Examining The Effects Of The Pandemic on Turkey's Happiness with The Dynamic Markov Switching Model

Elif Tuna
Yıldız Teknik Üniversitesi
elfztrk@gmail.com

ABSTRACT

There are many different definitions of happiness, often defined as positive emotions and life satisfaction. The emergence of a country or worldwide disaster such as Covid-19 has adversely affected people's mental and physical health and social relations, and these negative effects have reduced people's happiness levels. My aim in this study is to compare the results on the determinants of happiness before and after the pandemic in Turkey and to determine what is most important for happiness under these changing conditions. The Markov Regime Switching Regression model includes multiple equations that can characterize time series behavior in different regimes, so this method were used to compare the pre-pandemic and post-pandemic periods. In addition, the model allows happiness variables to take different coefficients in different periods, as well as determining the transition probabilities between happiness states and estimating the duration in these states. The results show that Turkey is in a long unhappy state.

Keywords: *happiness, covid-19, dynamic markov switching model*

Financial Performance Analysis with the TOPSIS, Fuzzy COPRAS and Entropy-COPRAS Approaches: The Case of Turkey

Yüksel Akay Ünvan Sinem Uz Cansu Ergenç

ABSTRACT

The evaluation of financial performance and profitability is always very important for the banking sector. With the technological and economic developments in recent years, the increasing competitive environment and globalization, it has become more important to determine the financial performance in the banking sector. Banks which play a significant role in the economy as they operate the payment system, are the important source of credit for large swathes of the economy, and act as a safe haven for depositors' funds. Banks are one of the most important factors in the formation of the economic structures of countries. Measuring and evaluating performance is very important for determining the position of banks in the sector. In addition, banks can identify their own shortcomings through performance evaluation. By measuring their performance, banks can see their position in the market and evaluate the situation needed to create competitive advantage. In this study, a financial performance analysis was examined in order to determine the criteria and situations that affect the financial performance of banks. According to the reports received from the Banks Association of Turkey for the 2015-2020 financial year, the top ten banks are ranked in total assets according to the relevant criteria and performance evaluation is made using TOPSIS, fuzzy COPRAS and Entropi-COPRAS (complex proportional evaluation). The criteria weights are determined by the entropy method and banks are ranked according to their financial performance by TOPSIS, fuzzy COPRAS and Entropy-COPRAS methods. Fuzzy approaches are preferred due to the fact that risk, uncertainty and competition continue to increase in the banking sector.

Keywords: *banks, topsis, fuzzy copras, entropy-copras*

Forecasting Power Outages

Beyza Yapıcı Esra Kılkıl

ABSTRACT

The widespread use of electrical energy in almost all sectors causes the demand for energy to increase day by day. This situation causes the existing power systems to be insufficient to meet the energy demand. In addition, the unstoppable increase in energy demand requires the system to operate at the maximum stability limit, and thus the possibility of encountering some difficulties in the power system increases. The biggest challenge that can arise is the power outages that have occurred very frequently in the world recently.Power outages, which cause serious effects on human life both socially and economically, need to be analysed well.It is extremely important to anticipate these disruptions in order to minimize the disruptions and their effects. In this study, it is aimed to make a forecasting of the number of power outages in the future in the grid centers under the responsibility GDZ Electricity and Distribution Company. The regions that distribution companies are responsible for exhibit a hierarchical structure that is differentiated according to various characteristics by nature. While GDZ Electricity and Distribution Company is responsible for the provinces of İzmir and Manisa, these provinces are divided into 4 regional directorates, and the regional directorates are divided into 29 enterprises.For this reason, the forecasting was made separately for the lowest level enterprises, for the middle level regions and provinces, and for the total number of cuts at the highest hierarchy level.The outage forecasting model was developed using historical outage data from the GDZ Electricity and Distribution Company.Time series shows intermittent and lumpy characteristics, since the interruption data is discrete and the interruptions occur irregularly and with little intervals, especially at the lowest level of the hierarchy. For this reason, statistical and machine learning techniques were investigated and the most appropriate methods were used to model intermittent and lumpy time series.The success of forecasting models evaluated with various success criteria in the literature.The outputs of the model that gives the best result as a result of the evaluation will be used as an effective tool that enables the distribution company to prepare its operational units against any possible interruption and to minimize the interruptions.

Keywords: *accuracy, electric, forecasting, power, time series*

Forecasting the Volatility of the Cryptocurrency Market: A Comparison of GARCH-type Models and Machine Learning Algorithms

Ömer Burak Akgün
Dokuz Eylül University
akgunburakomer@gmail.com

Emrah Gülay
Dokuz Eylül University
emrah.gulay@deu.edu.tr

ABSTRACT

Cryptocurrencies have attracted a great deal of investor interest and have received a lot of media coverage in recent years. Since then, there have been significant price fluctuations for cryptocurrencies. This paper provides a comprehensive review and additional explanation of the volatility behavior of the top three cryptocurrencies by using different GARCH-type models and machine learning (ML) algorithms. In this study, we forecast daily of three widely traded cryptocurrencies, i.e., Bitcoin, Ethereum, and Binance Coin by modelling volatility to select the best model. Furthermore, this study investigates the window effect, sliding or expanding, in cryptocurrency market. The reason for choosing different window size is because of the non-stationary of cryptocurrency that allows structural breaks. Besides the GARCH-type models and ML algorithms, this study proposes the hybrid and combining algorithms to improve the predictive accuracy of daily volatility forecasts. We evaluate the results under the five loss functions such as mean squared error (MSE), mean absolute percentage error (MAPE), heteroscedasticity-adjusted mean squared error (HMSE), quasi-like (QLIKE), and mean absolute scaled error (MASE). The empirical findings show that the hybrid model and combining algorithm yield promising results rather than the GARCH-type models and ML algorithms for the three datasets.

Keywords: *garch, machine learning, cryptocurrency, forecasting, volatility*

Fuzzy Regression Network Functions for Time Series Prediction

Mehmet Raci Aktoprak

Özge Cağcağ Yolcu
Marmara University
ozge.cagcag@marmara.edu.tr

ABSTRACT

The time series forecasting problem is a problem faced by decision makers in many fields. Different models have been developed to obtain predictions that play an important role in finding answers to questions such as what may be encountered in the future in the relevant sector or fields, what results the determined strategies and policies may lead to. Although traditional models are widely used in many fields, different approaches and models have also been put forward to overcome the disadvantages caused by the strict assumptions they require. These models are generally called non-probabilistic time series models and can be considered in two groups as computational-based and fuzzy based. While computational-based models include different artificial neural networks, fuzzy-based models include approaches such as fuzzy time series methods, ANFIS, and fuzzy regression functions. In time series problems, fuzzy-based models show superior performance due to their approach to uncertainty, while computation-based models show superior performance thanks to their high adaptability to nonlinear patterns. Fuzzy regression functions (FRF) approach, by using some transformation of the memberships as well as the real values of series, create a model with more information. However, it should be noted that the FRF creates a holistic model that is a combination of a series of linear function of inputs. In this study, a model is proposed having the advantages of both computational and fuzzy based. The proposed prediction tool, via fuzzy C-means clustering algorithm, produces the memberships by fuzzifying the actual observations of the time series. From a lagged observations of real time series and some transformation of these membership values, the inputs are generated. A number feed-forward neural networks, in equal to the number of fuzzy sets, produce outputs as nonlinear function of the inputs. These outputs are converted the final outputs by combining in the direction of membership values representing the degree to which the relevant time point belongs to the relevant fuzzy sets. The proposed model can be called as Fuzzy Regression Network Functions (FRNFs). The performance of FRNFs, in terms of some criteria, is put forward by applied it to different real-world time series.

Keywords: *fuzzy regression functions, artificial neural networks, fuzzy c-means, time series, prediction*

Genetic Algorithm Based Hybrid Model for COVID-19 Cases Forecasting

Ufuk Yolcu
Marmara University
ufuk.yolcu@marmara.edu.tr

Özge Cağcağ Yolcu
Marmara University
ozge.cagcag@marmara.edu.tr

ABSTRACT

The COVID-19 pandemic, which affected the whole world, has caused irreversible tragic effects in different aspects of human life. The massive loss of humans was the most devastating and dramatic of these. It would have been possible to take the necessary precautions by forecasting the number of cases accurately and reliably throughout the pandemic. Thus, confronting both an overload on public health systems and dramatic mortality rates could have been prevented. This study presents a hybrid model to forecast COVID-19 cases. The proposed hybrid model consists of the hybridization of intuitionistic fuzzy regression functions with elastic-net regularization (E-IFRFs) and long short-term memory (LSTM) neural network. There are two main objectives in the hybridization of these two methods. E-IFRFs develop a fuzzy modelling perspective to model the uncertainty in the data. A fuzzy modelling perspective is developed to model uncertainty in data using E-IFRFs. Thanks to the flexible modelling capability of LSTM, it is also possible to fully comply with the patterns in the data. For both components, fuzzy inputs are obtained by fuzzification of the actual observations with the intuitionistic fuzzy C-means clustering algorithm. Some related parameters of the E-IFRF and of the hybridization transaction in the forecasting process were determined by optimizing with the genetic algorithm. The proposed hybrid model was applied to the entire pandemic process from March 2020 to June 2022 and the forecasts were produced for 23 different periods in one-month forecasting steps. The obtained forecasts with a reasonable error were evaluated and interpreted as a measure of the feasibility of the proposed hybrid model.

Keywords: *covid-19, intuitionistic fuzzy regression functions, elastic-net regularization, lstm, genetic algorithm, forecasting*

HangiPixel: A Web Analytics Tool for Banking Sector

Türker Ziya Erçin Bekir Çetintav

ABSTRACT

In this study, the development processes of the WhichPixel Web Analytics tool, which enables instant measurement and analysis of the traffic directed to the web applications of our business partners through our own website, are included. The main purpose aimed with this tool is to collect various data and statistics, especially the data at which stage the redirected traffic is at the bank's side, in the light of these, to optimize the company's paid and organic customer traffic sources and to maximize profitability. In addition, instant reports are presented to business partners throughout the flow, thus deepening the integration is aimed.

Keywords: *web analytics tools, web metrics, data analysis, optimisation*

Hotel Ranking and Recommendation Through Online Reviews With Integrated Weighting Approach and ARAS Method

Sedef Çalı
Dokuz Eylül University
sedefcali@gmail.com

Adil Baykasoğlu
Dokuz Eylül University
adil.baykasoglu@deu.edu.tr

ABSTRACT

Multiple attribute decision making (MADM) is a field of Operations Research, which focuses on the decision problems of ranking or sorting the limited number of alternatives or proposing the most appropriate one by considering conflicting attributes. MADM process starts with defining the goal of the problem and the components such as alternatives and attributes. The performance evaluations of alternatives regarding attributes are provided by either decision maker(s) or any source, the attribute weights are obtained and the process ends with ranking or sorting alternatives or selecting the best one by following a specific mathematical procedure of the MADM method. Attribute weights are important components in MADM analysis because the assigned weights have a great role on result of analysis, and these weights can change the ranking of decision analysis. Nowadays, customers post the reviews on their experiences with the purchased products or services in online platforms, it is recognized that these platforms are valuable information sources for MADM analysis such as product ranking and recommendation. However, new adaptations in this research field are needed. MADM methodologies should evolve in such a way that they enable to extract performance evaluations and attribute weights from online data with the least loss of information. MADM analysis through online reviews necessitates appropriate weighting approaches extracting attribute importance from both text reviews and numerical ratings. In this study, hotel ranking based on Turkish online reviews is conducted through a novel MCDM methodology using text mining integrated weighting approach and ARAS method. The proposed weighting approach is a hybrid approach based on combination of text-mining and CRITIC method. CRITIC method is applied to decision matrices comprising numerical customer ratings to obtain objective weights. On the other hand, natural language processing is applied to customer comments and frequencies of words on comments are used to obtain text-based weights. These weights are aggregated to obtain hybrid weights. The alternative hotels are ranked and appropriate hotels are recommended via ARAS method by considering different types of customers and review dates.

Keywords: *big data, text mining, madm, hotel ranking, online reviews, data driven decision making*

How Much Should We Trust the P-Value?

Oya Ekici
Istanbul University
oyaekici@istanbul.edu.tr

ABSTRACT

Notably, since 2013, the problems with the hypothesis testing approach and the use of p-value have been discussed intensively in the international literature. This study aims to activate the debates of the misuse of p-value in the national literature and thus contribute to raising awareness among the researchers. In the study, the discussions on the issue of significance testing are supported with authentic quantitative analyses. We interpret the critical aspects of the significance testing results with simulations. Simulation studies are generated to examine the types of error, the power of a test and the factors affecting the power. As one of the proposed solutions to the significance testing issues, whether reporting the effect size and the confidence intervals can be informative besides p-value is evaluated based on the simulation. The overall conclusion is that there is an obvious need to accept the limitations of statistical analysis and embrace the uncertainties in scientific research. The implications of this paper in the academic studies and statistics education done in both social and fundamental sciences will contribute to extending the correct use of p-value.

Keywords: *effect size, hypothesis testing, p-value, power of a test, reproducibility*

How Significant is Chi-Square Test To Determine The Image Quality? A Retrofit in Calculation

Sinan Saraçlı
Balıkesir University
ssaracli@balikesir.edu.tr

Şevkiye Babacan
Süleyman Demirel University
sevkiyebbcn@gmail.com

Berkalp Tunca
Eskişehir Osmangazi University
berkalp0606@gmail.com

ABSTRACT

The purpose of this study is to determine the quality indicator of an image via Chi-square approach. There are many approaches to determine the image quality such as MSE, PSNR, Entropy and etc. Related with even there is a reference image, reduced reference image and non-reference image, quality measures differ. Even this statistic is used in earlier studies for this purpose, when there is a reference image, we propose a retrofit in the calculation of Chi-Square tests statistics by considering the pixel values of the reference image instead of difference from the value of white color. A raven image was used for analysis in the application part of the study. In order to process the image and make comparisons in a standard, it was converted to a gray scale image with 8 bit depth and 256 x 256 dimensions. This image is taken as a reference image, and the other three images to be used for comparison are the blurred and distorted version of this image with the help of Impulse noise. With this purpose in this study, considering a reference image, Chi-square test statistics is calculated for three noised images of the reference one. Analysis is done by MATLAB software. The results indicate that Chi-square value of the image which is much similar to original one is also the smallest one and overlap with the other measures. Detailed results are given in related tables and figures.

Keywords: *image quality measure, chi-square test, reference image*

Impact of Digital Transformation and EU Green Deal on Employment

Güneş Çetin Gerger
Manisa Celal Bayar University
gunes.cetin@hotmail.com

Atakan Gerger
Ege University
atakangerger@gmail.com

ABSTRACT

Digital transformation and the EU green deal are among the most current issues of our time. With the introduction of intelligent production systems into our lives because of digitalization, many jobs, especially heavy and dangerous work, have started to be done by robots and cobots. The EU green deal, it is aimed to reduce carbon emissions by 55% by 2030 compared to 1990 levels and eliminate them completely by 2050 to combat global climate change. For this, more environmentally friendly production systems are encouraged. However, this rapid digitalization and EU green deal raise many question marks. Undoubtedly, one of these issues is how employment will be affected. Because digitalization requires non-routine jobs and employees with analytical skills. Studies have shown that; Even if the same technologies are used, these technologies increase employment in some countries and decrease jobs in others. The results of the study indicate that while green and digital transformation will increase jobs in Germany, it will decrease them in Turkey. Looking at the employment skills in Turkey and Germany; While approximately 50% of employment in Germany is done with non-routine and analytical skills, this rate is 20% in Turkey. It is foreseen that the jobs that will be adversely affected by digitalization will be routine jobs, while non-routine analytical jobs will not be adversely affected by digital transformation. Moreover, the results of the study show a direct relationship between whether the work performed in the countries is routine or not, and the education levels of the population of that country.

Keywords: *digitalization, eu green deal, green production, digitalization and employment relationship, productivity*

Inflation Prediction With Artificial Neural Networks

Hasan Şen
Bursa Teknik Üniversitesi
hasan.sen@btu.edu.tr

Ömer Faruk Efe
Bursa Teknik Üniversitesi
omer.efe@btu.edu.tr

ABSTRACT

Inflation is caused by the growing between the amount of money actively involved and the sum of goods and services available for purchase. An increase in prices is an economic and monetary process that manifests itself as a decrease in the value of money. Inflation is a subject that keeps itself updated in the our country and in the world. Depending on the all the countries, the main purpose of the central banks is to ensure price stability. Artificial intelligence techniques have also been used more and more in recent years in order to predict inflation consistently. Today, artificial intelligence techniques can find application in many different fields. Despite the increasing application areas with artificial intelligence techniques, it is very difficult to make predictions with the estimation method in the field of finance. The high uncertainty in the financial sector and the volatility in the financal data reveal the most important reason for this situation. On the other hand, it is very important for the all country's economy and households to make an inflation forecast with high accuracy. For developing countries, keeping the inflation rate at low levels is among the one of the most important targets. The aim of this study is to estimate an inflation value with high accuracy in the Turkish economy by using the artificial neural network method, which is one of the artificial intelligence techniques, with time series analysis. In this study, a high-accuracy inflation forecast was made by using the Long Short Term Memory (LSTM) model, which is one of the models of artificial neural networks. In order to make the inflation forecast more accurate, it has been suggested to use different parameters affecting the inflation value in addition to the inflation data used in the study for subsequent studies and add them to the model.

Keywords: *artificial neural networks, time series analysis, inflation, artificial intelligence*

Investigating Obesity Related Health Factors of Bariatric Surgery Patients by Using Machine Learning Algorithms

Sezen İnan
nternational Patients Coordinator, Clinic of Dr. Metin Karadeniz, Izmir, Turkey

Özgül Vupa Çilengiroğlu
Ass. Prof. Dr., Dokuz Eylül University, Faculty of Science, Statistics Department, Izmir, Turkey
ozgul.vupa@deu.edu.tr

Metin Karadeniz
Op. Dr., Medical Doctor, Clinic of Dr. Metin Karadeniz, Izmir, Turkey

ABSTRACT

Purpose: The aim of this study is to make predictions about obesity factors and related parameters of adult patients who have undergone bariatric surgery in a private hospital in Turkey by using machine learning methods. Methods: The data used in this study were obtained from the medical reports of all persons (n = 592) who applied to a private bariatric clinic in İzmir between July 2018 and December 2020 and underwent gastric sleeve and gastric bypass operations. Patients aged 18 years and over, with a body mass index over 30 were included in the study. Multiple logistic regression analysis and decision trees (CART, CHAID) were used to examine the association of obesity levels with different risk factors. Results: Logistic regression model and CHAID algorithm were used to identify parameters associated with obese and morbid obese categories. It was detected that the most important variables on the obesity level variable were gender, age, insulin resistance and hepatosteatosis. These analyses revealed that males and those under 31 years of age were more likely to be morbidly obese. The effect of hepatosteatosis and insulin resistance variables in predicting obesity level was also supported in the same direction in logistic regression and CHAID analyzes. Conclusion: The lack of studies examining obesity levels as a dependent variable and using machine learning algorithms increases the importance of this study. The study can be strengthened by adding more data about psychological status and comorbid diseases of persons also by adding different nationalities and more record from different centers.

Keywords: *obesity, morbid obesity, bariatric surgery, logistic regression, decision trees*

Investigation of Stress Effect in Parallel Connected Systems

Mine Doğan
Fırat Üniversitesi
mine.dogan@firat.edu.tr

Mehmet Gürcan
Fırat Üniversitesi
mgurcan@firat.edu.tr

ABSTRACT

Theoretically, one of the most important indicators in the evaluation of the technical system's operating performance is the average uptime. This indicator can be calculated in different ways according to the variety of the mutual connections of the components. A similar field of study investigating the effect of stress on the system is stress strength models. However, the results obtained from stress-resilience models cannot be directly applied to systems operating under stress. Stress-strength models are more about comparing the effect of existing stress with the strength of stress. Briefly, the stress-strength model is related to the probability of {X>Y} event, with X being the variable showing the endurance of the system and Y being a variable showing the strength of the stress. In the stress-strength model, the reliability of the system is expressed by this probability, R=Pr{X>Y}. When the stress effect in the environment is desired to be reflected on the operation of the system, three different situations can be in question. In the first case, the stress effect is reflected in the variable representing the lifetime of the system, X^*=YX. In the second case, the stress effect can be reflected in the reliability function of the system. In this case, the reliability of the system under stress is R^*(t)=YPr{X>t}=YR(t). In the third case, the stress effect can be reflected in the periods between disruptions. In our study, the working principle of parallel connected technical systems under stress will be examined.

Keywords: *stress-strength models, distributions, reliability, parallel connected systems*

Logistic and CSG Growth Models For Predicting Life Expectancy

Deniz Ünal
Çukurova University
dunal@cu.edu.tr

Begüm Çığşar
-
begumcigsar@gmail.com

Abdel Hack Boulou
Çukurova University
bdelhack51@gmail.com

Bassel Alshahaby
Çukurova University
asselstat1984@gmail.com

ABSTRACT

Mortality/life tables are used to make estimations about and death rates. Life table is a chart that records the number of deaths of a certain number of newborns in a certain community over the period of time. Thus, using these tables, it is possible to obtain predictions about how many people at a certain age will die in a year. Life tables are used not only in humans, but also in all living things (even inanimate ones, with examples such as recording the downtime of a certain number of machines produced). The concept of life expectancy, which is a criterion that determines how long a creature will live, is also determined by mortality rates obtained from life tables. It is also possible to model the expected life time with mathematical functions. For this purpose, some nonlinear mathematical functions are used. One of the functions that is often used in modeling mortality rates is the logistic growth function. The aim of this study is to propose a model that can be used as an alternative to the logistic growth model and to interpret the mortality rates of countries both for overall data and for genders. For this purpose, the life expectancy of males and females in four countries (Turkey, Singapore, Norway and China) was modeled using the Logistics growth model and the CSG growth model, which was newly introduced to the literature. When modeling the life expectancy of countries, a model fitted graph was drawn in accordance with the data of each growth model. Then, the performances of the logistics growth model and the CSG growth model were compared with R^2, RMSE and MAPE statistical criteria. As a result of the comparison, it was revealed that the CSG growth model is more suitable to model life expectancy for over all data and for each gender than the logistic growth model. Thus, the CSG model has been proposed as a model that can be used as an alternative to the logistic growth model that is often used in the modeling of life expectancy in the literature. Acknowledgement: This study was supported by Çukurova University Scientific Research Projects Coordination Unit. Project Number: FBA-2022-14131.

Keywords: *life tables, growth models, csg model, logistic model*

Machine Learning for Crop Prediction: Understanding Wheat Yield by Remote Sensing Data

Ozan Evkaya
University of Edinburgh
ozanevkaya@gmail.com

Cansu Doğan
Hacettepe University
cansudoogan@gmail.com

ABSTRACT

In the majority of developing countries, agriculture is the principal source of income. Modern agriculture is a continually evolving system of agricultural innovations and farming practises. Meantime, it is the most vulnerable business type under the increasing treat of climate change all over the world. All of the above makes the viable crop prediction is crucial for different purposes such as agricultural risk management. Machine learning is an important decision support tool for crop production prediction, including decisions regarding which crops to produce, what to do during the growing season and what could be the expected future yield. In the literature, various machine learning techniques have been used on crop yield prediction studies. This study aims to investigate suitable machine learning tools for crop yield prediction for a specific province, Konya in Turkey. For the use of suitable machine learning methods, the available data set is pre-processed with classical time series modeling. To examine possible predictors for the Wheat Yield (WY), the remote sensing data are considered such as Vegetation Health Index (VHI), Temperature Condition Index (TCI), Vegetation Condition Index (VCI), Standardized Thermal Condition (SMT) and Smoothed Normalized Difference Vegetation Index (SMN - Smoothed NDVI), generated from environmental satellite. The primary findings indicate that the use of remote sensing data can empower the crop yield prediction accuracy under widely used techniques.

Keywords: *crop yield, prediction crop, machine learning, remote sensing data*

Multivariate Analysis of Variance Under Laplace Distribution

Müge Borazan
Department of Security Sciences, Gendarmerie and Coast Guard Academy
mugeborazan@yahoo.com

Serpil Aktaş Altunay
Department of Statistics, Hacettepe University
spxl@hacettepe.edu.tr

ABSTRACT

Many application areas have study designs derived from repeated measurements. Studies in which the measurements taken from each subject at multiple time points constitute the response variable are important and frequently used in research. The main strength of studies in which repeated measurements from different subjects are taken is that it is the only design type that allows to obtain information about individual change patterns. One of the difficulties encountered in the analysis of data obtained from repeated measurement studies is that the analysis has a complex structure due to the dependence between repeated observations from the same experimental unit. Another is the assumptions that the response variable has to provide. While there are many approaches to analyzing such data, most are limited to situations where the response variable is normally distributed and the data are balanced and complete. One of the analyzes used in this context is MANOVA. MANOVA is used to compare multiple means of multivariate normally distributed data. The main purpose of using MANOVA for repeated measures is that it is not necessary to provide the assumption of sphericity, which examines the dependency structure. In this study, the multivariate Laplace distribution, which is an alternative to the normal distribution, was used for the analysis of repeated measure data in cases where the assumption of normality was not met. An alternative test statistic to the normal distribution assumption is proposed for the analysis of repeated measures data using multivariate linear models under the multivariate Laplace distribution. By testing the hypotheses for the effect, trial and interaction effects, the power of these test statistics was calculated by the EM algorithm. Under the multivariate Laplace distribution, the results of the power values in the calculations of the proposed test statistics using linear models were quite good.

Keywords: *manova, laplace distribution, repeated measures, em algorithm*

New Multilayer Neural Networks With NO Estimator and Winsorized Mean

Burak Dilber
Dokuz Eylül University
burakdilber91@gmail.com

A. Fırat Özdemir
Dokuz Eylül University
firat.ozdemir5761@gmail.com

ABSTRACT

Multilayer feed forward neural networks have been widely used for prediction, forecasting and classification over the past few years. However, it is a known fact that the mostly preferred Mc - Culloch Pitts neuron model used in these network types does not give a successful prediction performance in data sets with outliers. Therefore, robust neuron models using median and trimmed mean aggregation functions have been proposed. However, these studies were generally focused on time series forecasting. In this study, we developed new neuron models using NO estimator and the Winsorized mean for prediction, classification, and time series forecasting. NO is a quantile estimator with weights determined by using a subsampling approach. For estimating a population quantile, it uses all order statistics in a sample and the accompanying weights of the order statistics are calculated from a Binomial Distribution. The proposed NO and Winsorized mean neuron models are not sensitive to outlying observations. Back propagation, particle swarm optimization and artificial bee colony optimization algorithms were used when training multilayer neural networks and several activation functions such as sigmoid, hyperbolic, tangent, and rectified linear unit were tried. All steps of the study were performed using statistical programming language R. The written functions of the proposed neural networks enable the prediction of new observations and observing the change of errors at each iteration by providing a dynamic plot. The developed methods were applied on the real data sets and their performances were compared with the existing ones. More successful results were achieved in terms of different performance criterions.

Keywords: *neural networks, neuron models, no estimator, winsorized mean, optimization algorithms*

On the Number of Failed Components in a Series-Parallel System

Murat Özkut
Izmir University of Economics
murat.ozkut@ieu.edu.tr

ABSTRACT

In this study, a series-parallel system with N subsystems having dependent components is addressed, and the number of failed components is taken into consideration. Copulas have modeled the dependence among the components. The number of components that have failed when the system fails is essential to track down, as it indicates how many spare components should be available to replace all failed components when the system breaks down. Once the results have been obtained, they're utilized to figure out the best amount of components for each subsystem by lowering the expected cost per unit of time when the system fails.

Keywords: *copulas, series-parallel system, optimal design, reliability*

On Weighted Random Sampling and the Computation of Participants' Chances in Turkey's Hajj Draws

Murat Güngör
İstanbul Medeniyet Üniversitesi
murat.gungor@medeniyet.edu.tr

ABSTRACT

Let there be N items partitioned into a number of groups. Group i has size n_i and weight w_i. Consider the sampling procedure of choosing n items as follows: in each round one of the groups is randomly selected to include in the sample all items therein, and the probability of selecting a group is equal to its weight divided by the yet unselected groups' total weight. The special case where groups consist of one item only is referred to as "weighted random sampling without replacement with defined weights." We abbreviate this as WRS, and the general case, which we shall call "weighted random sampling of groups without replacement with defined weights," as WRSG. Let p_i denote the probability that group i is included in the sample. Even for WRS there exists no efficient method in the literature to calculate p_i exactly. This presentation's subject matter is an investigation of these inclusion probabilities. Our motivation comes from a real-life problem related to hajj draws, the so-called "katsayılı kura sistemi" in Turkish, conducted every year by some countries including Bosnia and Herzegovina and Belgium. To summarize, we derive theoretical lower and upper bounds on inclusion probabilities in terms of item weights. We use these bounds as well as simulation to estimate applicants' chances in Turkey's 2020 hajj draw. Our results rely on a conjecture for which we provide a supportive example. It turns out that one who participates in the draws for the first time has a chance in between %0.12 and %0.13; similar bounds for one who participates for the eleventh time are %13.22 and %14.16.

Keywords: *weighted random sampling, without replacement, inclusion probabilities, hajj draws*

Predicting Personality Traits of Employee Candidates Using Natural Language Processing

Fatma Özge Özkök
Erciyes University
fozgeozkok@erciyes.edu.tr

ABSTRACT

The recruitment process and the selection of the right employees are significant for companies. Identifying talented personnel is one of the essential stages in this process. However, it is very critical that the characteristics of these personnel are compatible with the company. Recruitment of personnel incompatible with the company culture leads to different problems, such as low productivity at work and an increase in the rate of employee turnover. As a result, both financial and time losses can be experienced. The lack of attention, inexperience, or biased decisions of human resources personnel may cause this problem. Artificial intelligence-based decision support systems can solve this problem by making objective evaluations. Deep learning-based natural language processing methods have been very successful recently. In this study, the personality traits of employee candidates were classified using a statistical and deep learning-based natural language processing method. Experimental studies show that this method gives promising results in predicting personality traits by the responses of candidate employees.

Keywords: *nlp, natural language processing, recruitment process, deep learning, classification*

Prediction of Absorption Dose of Radiation on Thorax CT Imaging in Geriatric Patients with COVID-19 by Classification Algorithms

Adnan Karaibrahimoğlu
üleyman Demirel University
ıdnankaraibrahim@gmail.com

Ümit Kara
Süleyman Demirel University
ımitkara1983@gmail.com

Özge Kılıçoğlu
adiation Laboratory, University of Notre Dame
›zgekoglu@gmail.com

Yağmur Kara
University of Health Sciences, Isparta City Hospital
:ankayagmur@gmail.com

ABSTRACT

Objective: The aim of study is to predict the absorbed radiation dose on thorax CT imaging in geriatric patients with COVID-19. Materials and Methods: The SIEMENS SENSATION 64 CT scanner was performed with real protocols to patients (male/female phantom) using Monte Carlo simulation methods with the real height and weight of patients and the actual parameters CT scanner. Absorbed organ doses have been calculated based on these Monte Carlo results. These results were used to predict the optimal absorbed radiation dose by Artificial Neural Networks, Linear Discriminant Analysis, Random Forest Classification, and Naive-Bayes Classification algorithms. The dose values were clustered for genders by the Fuzzy C-Means algorithm. Results: The ages of the patients were between 60 and 70 years old. The Body Mass Index of male and female patients was 26.11±4.49 and 25.03±4.86 kg/m2 respectively. All classification algorithms were validated with approximately 100% success. The Fuzzy C-Means technique was found to be successful in clustering the dose values for gender clusters. Conclusion: While the predicted and the observed values of patients do not change in the organs/tissues around and outside of the thorax, they generally vary in the intra-thoracic organs and tissues. It can be concluded that the data-driven techniques are useful to obtain optimal radiation dose for organs/tissues in CT imaging.

Keywords: *radiation dose, covid-19, data science, classification, clustering*

Proposition of A New Updated Robust Ridge Parameter Estimator for Linear Regression Model

Özge Akkuş
Muğla Sıtkı Koçman University
ozge.akkus@mu.edu.tr

Selman Mermi
Muğla Sıtkı Koçman University
selmanmermi@mu.edu.tr

Atila Göktaş
Muğla Sıtkı Koçman University
gatilla@mu.edu.tr

ABSTRACT

The Ridge regression approach was proposed by Hoerl and Kennard (1970) and is widely used by researchers to reliably estimate the regression coefficients in a multiple linear regression model with multicollinearity. The main issue in the ridge regression method is to estimate the ridge parameter as close to the truth as possible. This study aims to propose a new robust ridge parameter estimator using the "search method" proposed by Göktaş et al. (2021). Furthermore, it is aimed to compare the performance of this robust estimator with a total of 366 different ridge parameter estimators suggested in the literature in terms of both the Mean Square Error (MSE) criterion and the criteria for the values generated by the ridge parameter estimators to have outliers with a low ratio and to have a normal distribution with a high ratio. For this reason, a Monte Carlo simulation study with N=10000 iterations was carried out using parameters varying widely in terms of the number of independent variables (p), sample size (n), the correlation coefficient between independent variables (ϱ) and standard deviation of errors (σ). According to the simulation results, the proposed robust ridge parameter estimator outperformed the other 366 estimators in terms of MSE criterion in the majority of cases. Moreover, the proposed robust ridge parameter estimator was found to produce values with a normal distribution in almost all cases and almost never generated outliers. Finally, it was revealed that there is a significant linear correlation between the performance of the ridge parameter estimators in terms of the MSE criterion and the ratios of producing outliers and having a normal distribution, respectively, in negative and positive directions. Acknowledgment: This study is supported by TUBITAK (The Scientific and Technological Research Council of Türkiye) with project number 121F257. We thank TUBİTAK for supporting our study.

Keywords: *multicollinearity, ridge regression, robust ridge parameter estimator, search method, monte carlo simulation*

Reliability Assessment of Maximum Likelihood Estimation on a Weighted Weibull Distribution

Yasin Altınışık
Sinop University, Faculty of Arts and Sciences
yasinn_alt@hotmail.com

Demet Aydın
Sinop University, Faculty of Arts and Sciences
daydin@sinop.edu.tr

ABSTRACT

The main purpose of this study is to examine the statistical properties of a two-parameter size-biased weighted Weibull (WW) distribution as an extension of the usual Weibull (W) distribution. WW distributions are used to model the data in many research fields such as survival analysis, weather forecasting, lifetime studies, industrial engineering etc. In the literature, these distributions are frequently utilized for studies in which the values of a dependent variable in the data do not have equal probabilities of being observed where the usual W distribution does not perform well. Since such data sets often contain heterogeneity, non-parametric estimation possible by relaxing the assumption that the data come from the proposed WW distribution. Therefore, along with maximum likelihood estimation (MLE) as the main estimation in this study, non-parametric bootstrapping (NB) will be used to estimate the parameters of WW distribution and their standard errors. We employ W and WW distributions on a real data using MLE and NB estimation methods based on two additional objectives: (1) to investigate whether we attain better fitting performance with the proposed WW distribution in comparison to the usual W distribution. (2) to examine whether MLE and NB provide reasonably close parameter estimates (and their standard errors) for W and WW distributions. This second objective is particularly important as attaining close results with MLE and NB methods in parameter estimation means that the results of MLE is reliable on modeling the data using W or WW distribution.

Keywords: *weibull distribution, weighted weibull distribution, parameter estimation, maximum likelihood estimation, non-parametric bootstrapping, model comparison*

Robust Cryptocurrency Portfolio Optimization by Using NNN and MNN

Ilgım Yaman
Giresun University
ilgim.yaman@giresun.edu.tr

ABSTRACT

The more important the gain is for the investors, the more important the risk is. Therefore, the portfolio optimization is a major problem in the finance world. Harry Markowitz had proposed standard portfolio optimization method in 1952. Considering the fitness function of the portfolio optimization model, solving this quadratic optimization problem does not that simple. In 2019, we proposed nonlinear neural network for portfolio optimization (NNNPO) which solves that complex problem for the standard portfolio optimization problem. Nowadays crypto assets more common than stock exchange market, but cryptocurrencies do not meet Markowitz's assumption which is assets have to be normally distributed. In this study, cryptocurrency data taken in to account between July 5, 2018 to July 4, 2019. Firstly, to evaluate a portfolio return for risk, information ratio, sharpe ratio, sortino ratio calculated. Secondly, to avoid normality assumption, robust parameters calculated. Thirdly, the forecasts obtained from multiplicative neural networks (MNN). Then, forecasted data taken into nonlinear neural network algorithm for portfolio optimization problem to determine the proportion of the currencies in the selected portfolio. Finally, the return of selected portfolio compared with the real cryptocurrency data.

Keywords: *nonlinear neural network, robust, crypotocurrency, portfolio optimization*

Seasonal time series analysis: decomposition and forecasting with Ataforecasting package in R

Hanife Taylan Selamlar
Dokuz Eylul University
hanife.taylan@gmail.com

ABSTRACT

The evaluation of financial performance and profitability is always very important for the banking sector. With the technological and economic developments in recent years, the increasing competitive environment and globalization, it has become more important to determine the financial performance in the banking sector. Banks which play a significant role in the economy as they operate the payment system, are the important source of credit for large swathes of the economy, and act as a safe haven for depositors' funds. Banks are one of the most important factors in the formation of the economic structures of countries. Measuring and evaluating performance is very important for determining the position of banks in the sector. In addition, banks can identify their own shortcomings through performance evaluation. By measuring their performance, banks can see their position in the market and evaluate the situation needed to create competitive advantage. In this study, a financial performance analysis was examined in order to determine the criteria and situations that affect the financial performance of banks. According to the reports received from the Banks Association of Turkey for the 2015-2020 financial year, the top ten banks are ranked in total assets according to the relevant criteria and performance evaluation is made using TOPSIS, fuzzy COPRAS and Entropi-COPRAS (complex proportional evaluation). The criteria weights are determined by the entropy method and banks are ranked according to their financial performance by TOPSIS, fuzzy COPRAS and Entropy-COPRAS methods. Fuzzy approaches are preferred due to the fact that risk, uncertainty and competition continue to increase in the banking sector.

Keywords: *ata method, stl, stlplus, tbats, str, time series forecasting.*

Similar QA Pairs Detection System to Improve Chat Quality in Turkish Conversation Groups

İzzet Kılıç
Hacettepe University
izzet.kilic@hacettepe.edu.tr

Derya Karagöz
Hacettepe University
deryacal@hacettepe.edu.tr

ABSTRACT

Nowadays, social media applications have become one of our essential communication tools. These applications can be used to communicate individually or in groups. Retrieving information by processing the text data produced in user groups increases the quality of the chat. Consequently, creating a system that will prevent the same question from being asked multiple times in groups with thousands of users will help increase efficiency in the chat. With the help of the application developed on the Slack platform in this study, users can search for similar questions in old messages and retrieve their answers. In cases where there is no similar question in the conversation, a mechanism has been developed to direct the question that is asked by a user to a user who has knowledge of the subject. The BERTurk model is fine-tuned for sentence classification and question answering tasks, which are natural language processing techniques. Since a large amount of training data is needed, datasets created in other languages are translated to Turkish and open-source datasets are used.

Keywords: *natural language processing, deep learning, question answering, question detection*

Study of Artificial Neural Networks for Single Jersey Fabrics in the Textile Industry

İrem Soylu
Dokuz Eylül Üniversitesi
irem.soylu@hotmail.com

Neslihan Demirel
Dokuz Eylül Üniversitesi
neslihan.ortabas@deu.edu.tr

ABSTRACT

Today, with the increase in the demand for textile products and the rise in customer expectations, the development in the process, product and service factors offered to the customer in the textile sector has become a mandatory requirement. One of the research methods that can meet this need is artificial neural networks. Artificial neural networks represent algorithmic structures derived from the simplification of the human brain. Artificial neural networks involve the transformation of neurons in the human brain into mathematical models. It is extremely important to estimate the width parameter since the fabric will be produced for the first time in newly developed fabrics. In this study, it is aimed to determine the necessary factors and the levels of these factors for the correct estimation of fabric width and to model them with artificial neural networks. When the fabric code was defined for the first time, it arose from the need to give the final fabric width value automatically by the system, independent of the human factors. While defining the fabric code, the width of the fabrics is determined entirely according to the experience of the employee. This results in inconsistent and unreliable results. Fabric return, fabric repair, etc. resulting from the width parameter defined in this way. problems cause serious loss of time and cost increase in the factory. In this study, a comparison was made between the methods discussed as a result of the arrangement and visualization of the data on the data set of cotton single jersey fabrics and the most appropriate models created.

Keywords: *neural networks, supervised learning methods, multilayer perceptron model, single-layer perceptron model*

Team Size Optimization for Power Restoration in Workforce Management

Esra Kılkıl
GDZ Elektrik Dağıtım Şirketi
esra.yesilbaskilkil@gdzelektrik.com.tr

Efendi Nasiboğlu
Dokuz Eylül Üniversitesi
efendi_nasibov@yahoo.com

ABSTRACT

A scientific and reasonable power emergency supplies workforce management is essential to speed up the restoration process after damage of a power system. In this study simulation approach is proposed for annual planning of power restoration workforce related to an GDZ electricity distribution company. Workforces are employed to restore power after interruptions throughout the province on service point. According to the electricity distribution network, locations and field sizes, service points are divided into 29 operation centers, each to restorate power interruptions. The scale of the study is a discrete, stochastic and multi-criteria problem due to the nature of the electricity distribution system therefore determining the size of the team in each service point over the year is challenging. Given this background, firstly, an event definition was made for the period from the occurrence of the interruption notification to its completion. Time intervals in events have stochastic lengths because they are affected by unpredictable factors. Using the outage data of the last two years, monthly outage frequency estimations were made for each operation center. The probability distributions of each event were determined and distirubitons are simulated. Monthly labor demand is then forecasted based on an optimistic scenario where there is a large enough labor resource at the time of the outage to keep the downtime level to a minimum for the operations centers. Finally, for purpose of validation of workforce distribution multicriteria model are developed. The model is based on a survey of relevant data, especially those related to the need to perform tasks on the lines, as the number of consumers, voltage lines, transformers, so on. Multicriteria model is grouped according to theoperation centers. Correlations between model coefficients and estimates of operations center team size are examined. The presented model is expected to increase efficiencies by focusing especially on peak hours, with optimum design teams that ensure the distribution of workforce according to operation centers, minimizing customer interruptions and personnel costs.

Keywords: *decision making, mathematical programming, statistical analysis, power interruption, workforce planning*

The Challenge of the Artificial Intelligence

Fatih Şahin
İstanbul Nişantaşı University
fatih.sahin@nisantasi.edu.tr

ABSTRACT

Research on artificial intelligence in the last two decades has dramatically improved the performance of both manufacturing and service systems. There is a dire need for an article that presents a holistic literature survey of worldwide theoretical frameworks and practical experiences in the field of artificial intelligence. This paper reports the state-of-the-art on artificial intelligence in an integrated, concise, and elegantly distilled manner to show the experiences in the field. In particular, this paper provides a broad review of recent developments within the field of artificial intelligence (AI) and its applications. The work is targeted at new entrants to the artificial intelligence field. It also reminds the experienced researchers about some of the issues they have known. Artificial intelligence (AI) techniques are becoming useful as alternate approaches to conventional methods or as components of integrated systems. They have been used to solve complicated practical problems in various areas and are becoming more popular nowadays. They can learn from examples, are fault tolerant in that they can handle noisy and incomplete data, can deal with nonlinear problems, and, once trained, can perform prediction and generalization at high speed. AI-based systems are being developed and deployed worldwide in various applications because of their symbolic reasoning, flexibility, and explanation capabilities. AI has been used in different sectors, such as engineering, economics, medicine, military, marine, etc. They have also been applied for modeling, identification, optimization, prediction, forecasting, and control of complex systems. The paper outlines an understanding of how AI systems operate by presenting several problems in photovoltaic systems application. Issues raised include three areas: forecasting and modeling of meteorological data, sizing of photovoltaic systems, and modeling, simulation, and control of photovoltaic systems. Published literature presented in this paper shows the potential of AI as a design tool in photovoltaic systems.

Keywords: *artificial intelligence, nlp, expert system, genetic algorithm*

The Effects of Under-Sampling the Dataset in Intrusion Detection with Machine Learning

Sura Emanet
Türk Hava Yolları
suraemanet@gmail.com

Gözde Karatas Baydoğmus
Marmara Üniversitesi
karatas@marmara.edu.tr

Önder Demir
Marmara Üniversitesi
demir@marmara.edu.tr

Sadettin Melenli
NTT Data
sadettin.melenli@nttdata.com

ABSTRACT

It is inevitable that devices connected to the Internet become vulnerable to attacks with the increasing number of users. Ransomware, which has emerged in recent years, has given users great difficulties both financially and morally. Intrusion Detection Systems are generally preferred to deal with such network-based attacks. Intrusion Detection Systems are systems that are popularly used in the field of computer security to ensure data security and prevent unauthorized access. There are several metrics for an intrusion detection system to be considered successful; these are accuracy and error rate, efficient run time, precision and recal rate values. In recent years, Intrusion Detection using Machine Learning techniques is highly preferred among researchers. One of the problems in this regard is that the current datasets are not too many, and some of the datasets contain an unbalanced number of large data. Especially when classification is made with big data, the run time of the developed system increases a lot, and sometimes overfitting occurs. Subsampling techniques allow machine learning classifiers to run representatively on smaller subsets, allowing them to get high accuracy with shorter processing time by undersampling the dataset. In this study, the importance of subsampling in attack detection with Machine Learning was examined using the CSE-CIC-IDS2018 dataset containing 16,232,943 data. Intrusion detection System performance was evaluated with five popular Machine Learning algorithms such as Decision Tree, Random Forest, Logistic Regression, Passive Aggressive Classifier and Gradient Boosting algorithms by subsampling the CSE-CIC-IDS2018 dataset. Near-Miss algorithm was used to obtain subsamples. In the study, Decision Tree and Logistic Regression algorithms were first run with the original size of the dataset. However, since the working time was too long, operations with other algorithms were continued by reducing the dataset size by 40%, 45%, and 50%. As a result, the running time, accuracy rate, precision, recal and f1-Score values of the algorithms were obtained as evaluation criteria. The study presents a ratio in the number of subsamples for researchers who will use the CSE-CIC-IDS2018 dataset and related machine learning algorithms. In addition, as a result of the study, it was seen that the accuracy rate in the logistic regression algorithm with 50% subsampling increased compared to the original and the time was reduced by half.

Keywords: *intrusion detection, under-sampling, machine learning, near-miss*

The Half-Logistic Garima Distribution and Modelling

Ayşe Metin Karakaş
Bitlis Eren University
aysekarakas5767@gmail.com

Aslıhan Demir
Fırat University
a.bilici128@gmail.com

Sinan Çalık
Fırat University
scalik@firat.edu.tr

ABSTRACT

The Half-Logistic Garima distribution is introduced and its structural, statistical properties are investigated. This study includes the compounding representation of the distribution, the shapes of the density and hazard rate functions, the moments and quantiles as well as the limiting distributions of extreme order statistics. Afterwards, the superiority of the Half-Logistic Garima distribution over the other known distributions on the real data set was shown.

Keywords: *garima distribution, hazard rate function.*

The Mediating Effect of Hedonic and Utilitarian Value on the Relationship Between Technological Aptitude and Intention To Continue Using Wearable Technology

Ceren Börüban
Topkapi University
cerenboruban@topkapi.edu.tr

Dilek Veysikarani
Cukurova University
dveysikarani@cu.edu.tr

Ebru Özgür Güler
Cukurova University
ozgurebru@cu.edu.tr

ABSTRACT

Technological aptitude is a concept that emerges as individuals adopting and using new technologies to achieve their goals in their home and work lives. Herein, it is stated that rapid changes and innovations in information and communication technologies positively affect the individuals's lives. In addition, the demands of individuals for technological products are increasing annually. The tendency of individuals who love technology and have technological aptitude, to wearable technology among technological products, stands out in this field. The use of smart watches in wearable technology is increasing. In this context, the aim of study is to reveal the effect of technological aptitude of smart watch users on their intention to continue using smart watches. The research also aims to investigate the mediating role of hedonic and utilitarian value in this effect. The research population consists of individuals residing in Adana. Individuals using smart watches selected by simple random sampling method at the time of research constitute sample group of study. There is an information form in first part of the questionnaire, which is prepared to be applied face-to-face and on the internet, to the participants in the research. In second part of the questionnaire, there is a 10-item Technological Aptitude Scale developed by Parasuraman and Colby (2015) and tested for validity and reliability by Şenlik (2021). Also, there are Hedonic Value Scale, Benefit Value Scale, and Continuing Use Intent Scale prepared by Hong et al. (2017) and tested for validity and reliability by Yıldız and Kütahyalı (2021). At the stage of analyzing, firstly, a sample profile was presented with descriptive statistics, and then reliability coefficients of the scales were examined. By using structural equation model, it is planned to contributing the relevant field by evaluating the technological aptitude effect which is the research's main purpose, on the intention to continue use. The effect size of hedonic value and utilitarian value, which is predicted to have a mediating effect, on the intention to continue use was calculated and evaluated. It is anticipated the study will provide useful information on effects of technological aptitude on the intention to continue using smart watches.

Keywords: *wearable technology, mediation effect, structural equation modeling*

The Multivariate Extension of CoVaR With Higher-Order Moments for Tail Risk Network Construction

Arief Hakim
Institut Teknologi Bandung, Indonesia
30121003@mahasiswa.itb.ac.id

A.N.M. Salman
Institut Teknologi Bandung, Indonesia
msalman@math.itb.ac.id

Khreshna Syuhada
Institut Teknologi Bandung, Indonesia
khreshna@itb.ac.id

ABSTRACT

In a financial system, entities (e.g., institutions, countries, or markets) are exposed to systemic risk that can lead to financial instability and insolvency. Managing systemic risk is thus crucial for anticipating these severe impacts. To do this task, we require the so-called Conditional Value-at-Risk (CoVaR) and conditional tail risk network. The former is a measure of a targeted entity's tail risk conditional on another entity being under financial distress. The corresponding Delta CoVaR can be employed to construct such a network model, whose nodes represent entities connected by directed weighted edges. In this study, we aim at formulating and modifying a multivariate extension of CoVaR (MCoVaR) that takes the joint conditioning event of multiple entities into consideration. A multivariate Johnson's SU risk model is first utilized to capture the asymmetry and leptokuritic properties of entities' asset returns. This model is then expanded analytically using the Cornish-Fisher expansion by accounting for its conditional moments higher than the second order. With an application to global foreign exchange (forex) markets before and during COVID-19, we reveal that the resulting expanded MCoVaR forecast exhibits better conditional coverage performance than the unexpanded version. The former's superiority appears to be more pronounced in times of COVID-19. Using a Delta MCoVaR-based network, the nodes representing some forex markets are found to have a huge coefficient of clustering, suggesting that these nodes tend to cluster together with their neighboring nodes. Accordingly, special attention needs to be paid to systemic risk sourced from these forex markets. In summary, this study offers a more sophisticated way to formulate a more accurate systemic risk measure forecast and construct a more representative tail risk network.

Keywords: *systemic risk, conditional value-at-risk, cornish-fisher expansion, network model, clustering coefficient*

The Performances of Four Quantile Estimators When Comparing Dependent Groups With a Percentile Bootstrap Method

Engin Yıldıztepe
Dokuz Eylul University,
Department of Statistics
engin.yildiztepe@deu.edu.tr

A. Fırat Özdemir
Dokuz Eylul University,
Department of Statistics
firat.ozdemir@deu.edu.tr

Tuğçe Paksoy
Dokuz Eylul University, The
Graduate School of Natural
and Applied Sciences
paksoytugce35@gmail.com

ABSTRACT

Comparing two dependent groups is an essential research topic in applied statistics. When doing this, the location measures of the marginal distributions are generally used to compare before and after measurements derived from the same sample group. The location estimators, especially the mean, are generally sensitive to outlying observations. So if the point of interest is the tails of the marginal distributions, location measures might not provide deep insight, and a more specific measurement might be needed. Quantiles are reference values intended to reflect the typical observations of a given particular point of a distribution. There are three main approaches for obtaining a quantile estimator: using a single order statistic, taking the weighted average of two order statistics, and taking the weighted average of all order statistics. A common problem for all quantile estimators is getting a reasonable accurate standard error estimation and a hypothesis testing procedure for the corresponding population parameter. The percentile bootstrap approach in hypothesis testing performs reasonably well in simulations. In this study, the newly proposed NO quantile estimator, the Harrell-Davis quantile estimator, the trimmed Harrell-Davis quantile estimator, and the default quantile estimator in R function quantile() (type-7 quantile estimator) were used to compare the different quantile values of the two dependent groups using the percentile bootstrap method. The study aims to compare the performances of four different quantile estimators in terms of saving Type I error for the 0.05 level. A simulation design was conducted for varying correlations, quantile values, sample sizes, and targeted statistical distributions. All computations were performed in R 4.2. It was found that the NO quantile estimator gives better results than the other quantile estimators in most of the scenarios covered.

Keywords: *no quantile estimator, percentile bootstrap method, two dependent groups*

The Turkish Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach

Çiğdem Lazoğlu
Hacettepe University
cigdemkobal@hacettepe.edu.tr

Uğur Karabey
Hacettepe University
ukarabey@hacettepe.edu.tr

ABSTRACT

This study examines the relationship between the term structure of interest rates and macroeconomic fundamentals in Turkey. The state-space model is used to research the dynamic interaction between three latent yield-curve factors (level, slope, and curvature). Then the model is expanded to include three macroeconomic variables (real activity rate, inflation rate, and overnight lending interest rate). The model that consisted of macroeconomic variables shows better fits the yield curve in the long and short term. We found that macroeconomic variables effects the future yield curve and vice versa.

Keywords: *yield curve, state-space model, nelson siegel model*

The VSI Tukey-Exponentially Weighted Moving Average Control Chart with Asymmetrical Limits

Selcem Adsız
Ege University
selcemadsiz@gmail.com

Burcu Aytaçoğlu
Ege University
burcuaytacoglu@gmail.com

ABSTRACT

One of the most important tools in statistical process control is control charts. Conventional control charts are generally designed under the assumption of normality and samples are taken at fixed time intervals. However, in real life, the assumption of normality may be violated and the control chart may not show the desired performance. One of the control charts proposed for non-normal data is the Tukey-Exponentially Weighted Moving Average (Tukey-EWMA) control chart. It is designed by combining the features of the Tukey control chart and the EWMA control chart. The quartiles and inter-quartile range are included in the control limits. A method to strengthen the performance of the traditional control charts is to take samples with varying time intervals instead of fixed time intervals. The Variable Sampling Interval (VSI) feature adds warning limits to the current control limits of the control charts. If the last sample taken fell between the warning limits and the control limits, the next sample is taken in a shorter time. In this study, the VSI Tukey-EWMA control chart with asymmetrical limits is proposed by adding the VSI feature to the Tukey-EWMA control chart. The performance of the VSI Tukey-EWMA control chart was examined for gamma and Weibull distributions. The shift in the process may not occur at the beginning of the process and its size and direction may not be predicted. Therefore, it is necessary to take into account a certain range of shifts and the expected time from the assignable cause to the out-of-control signal. Accordingly, Average extra quadratic loss (AEQL) and Adjusted Time to Signal (AATS) is used as performance measures. To calculate AEQL and AATS, Markov chain approach is applied with the help of R software package. Additionally, the performance of the proposed chart is evaluated by comparing it with Tukey-EWMA control chart.

Keywords: *variable sampling interval, average extra quadratic loss, markov chain, adjusted time to signal*

Use of Six Sigma in the Optimization of the Welding Process of Titanium Tubes: An Application in the Aerospace and Aviation Industry

Atakan Gerger
Ege University
atakangerger@gmail.com

ABSTRACT

The Six Sigma methodology finds more and more application areas every day, as it contains very powerful statistical techniques and makes improvements measurable. One of these areas is the aerospace and aviation industries. A very strict control procedure is applied in accordance with industry standards. The materials used in this industry are high-tech and often consist of expensive components. In this article; Internal structure defect and weld seam defect quality problems that occur in the welding operation of a supplier that directly produces Titanium pipes for the aerospace main industry are discussed. The Six Sigma project was carried out using the DMAIC methodology. The company aimed to optimize the welding process of Titanium tubes and Taguchi experimental design was used in the improvement phase. As a result of the realized project; The sigma level was increased from 1.19 to 2.38.

Keywords: *six sigma, taguchi experiment design, design of experiment, attribute control chart, application of six sigma in aerospace and aviation industry*

Using Alternative Prediction Methods in Portfolio Selection

Azize Zehra Çelenli Başaran
ondokuzmayıs üniversitesi
azizezehracelenli@gmail.com

Özge Gündoğdu
Süleyman Demirel Üniversitesi
ozge5gundogdu@hotmail.com

Ali Zafer Dalar
Giresun Üniversitesi
alizaferdalar@hotmail.com

ABSTRACT

Investors tend to investment instruments with low risk and high return while determining portfolio varieties. Especially the economic fluctuations experienced during the Covid-19 period, sudden rises and falls in the markets have led investors to the stock market with the expectation of extra profit. Although the stock market is risky, it is also known as investment instruments that can provide high returns. Since investment is an action against the uncertainty of the future, the concept of risk is not destroyed. Calculating risk is very important for a correct investment. In the literature, risk is separate into two as systematic and unsystematic risk. Systematic risk is a type of risk that arises from non-stop changing parameters in the economy. Various methods have been used in the literature for portfolio optimization and estimation of stock exchange direction. In this study, linear and robust regression methods were used to estimate the beta coefficient which is a measure of systematic risk. The obtained beta values were used in the Treynor criterion to determine the portfolio performance. Treynor criterion is a reviewable criterion and beta is used when calculating Treynor criterion. Portfolio varieties obtained by using different beta estimation values are evaluated in terms of the benefit they provide to the investor. It was compared in terms of risk, return, systematic risk and Treynor criterion for each portfolio and optimum portfolio was obtained.

Keywords: *robust regression, treynor criteria, portfolio optimization*

Using Kullback–Leibler Divergence As Loss Function in Convolutional Neural Networks for Distribution Comparison

Cağatay Bal
Muğla Sıtkı Koçman University
cagataybal@mu.edu.tr

Serdar Demir
Muğla Sıtkı Koçman University
serdardemir@mu.edu.tr

Cağdaş Hakan Aladağ
Hacettepe University
aladag@hacettepe.edu.tr

ABSTRACT

Determining the distribution of the data has an important role in statistics. There have been many methods developed within years to test the similarity of data`s distribution with known distributions. Identifying the distribution of the data allows to apply statistical tests to analyze and produce predictions from it. Therefore, achieving significant results based on statistical tests would be valuable if only data`s distribution can be identified truly. Kullback-Leibler divergence is a statistical distance, a measure of how a probability distribution P is different from a reference probability distribution Q. The proposed method uses Convolutional Neural Networks with Kullback-Leibler divergence as loss function as Artificial Intelligence to identify the distribution of given data as input to provide probabilistic similarity between known distributions. Proposed method is applied simulated and real-world data sets to show using artificial intelligence can improve statistical analyzing.

Keywords: *artificial intelligence, convolutional neural networks, distribution, kullback-leibler divergence, loss function*

Using Unsupervised and Supervised Machine Learning Algorithms to Evaluate Higher Education Institutions

Tuğba Söküt Açar
Çanakkale Onsekiz Mart Üniversitesi
t.sokut@comu.edu.tr

ABSTRACT

Higher education institutions are of critical importance for the social, economic and cultural development of a country and contribute to national development by transferring specialized knowledge and skills to new generations. Higher education institutions' primary responsibilities are education and research. The Council of Higher Education (CoHE) supervises the activities of higher education institutions. In today's world, where the diversity of higher education institutions and the number of students are increasing, it has become compulsory to make this numerical growth qualified by evaluating the studies that indirectly or directly affect society carried out in institutions. The study of University Monitoring and Evaluation Criteria is one of the studies carried out in line with these ideas. These reports are prepared in the format presented by CoHE in five categories. The combined data from the reports were subjected to machine learning (ML). ML, taking importance with increasing data, is applied in many field fields such as biology, medicine, automotive, aviation, and generation of energy. ML basically has two types of algorithms: supervised and unsupervised. Supervised learning is a type of ML in which machines are trained using labeled training data and then predict the output based on that data. In unsupervised learning, the same process is done with unlabeled data. The paper describes an experimental study that uses machine learning (ML) methods to classify higher education institutions with supervised algorithms and to cluster with unsupervised algorithms according to the Monitoring and Evaluation reports of CoHE. In this context, 52 criteria in the 2020 reports of each university were included in the analyses. The results of the algorithms were compared according to performance criteria. According to the findings of supervised algorithms, the Boosting classification produced the most successful estimation results among other algorithms, with the test accuracy as 98.2% values. In unsupervised learning, the best clustering was obtained by the Random forest algorithm according to AIC and BIC criteria.

Keywords: *machine learning, classification, clustering, higher education*

Validity and Reliability of the Singles Stress Scale-Youth Form (SSS-YF)

K. Bahar Aydın
Ankara Yıldırım Beyazıt Üniversitesi
kamilebahar@hotmail.com

Nailakhanim Rustamova
Milli Spor ve Tıp Rehabilitasyon Eğitim-Araştırma Enstitüsü
nell.rustamova@gmail.com

ABSTRACT

Studies on singlehood are very limited worldwide, especially in Muslim countries. However, singlehood is a natural part of the life cycle. On the other hand, the changes that take place at national and global levels constantly make the rate of the singles population increase. For example, the divorce rate has increased rapidly as a result of the COVID-19 Global Pandemic. The term singlehood refers to individuals who have never been married, divorced, in the process of divorce living separately, and widowed. They are also referred to as singlehood status in the Singles Counseling Theory (SCT). Sub-single groups emerge when developmental periods and singlehood status are matched (Young widows/widovers, old widows/widovers, etc.). The term developmental period is used to mean late adolescence, young, middle, or advanced adulthood. The late adolescence and youth periods are critical in the SCT, because these periods affect the following periods. The professional support given during these periods can positively affect the rest of an individual's life. Developmental and contextual factors reveal problems, needs, and characteristics common to the entire singles population and specific to each subgroup. All of these point to the necessity of counseling specific to the singles population. This type of counseling is called Singles Counseling. To provide effective counseling services to single individuals, their stress, needs, and characteristics should be determined objectively. The youth period, which falls between the ages of 18-30, is of great importance in terms of preparation for healthy adult life. For this reason, in this study, the validity and reliability of The Singles Stress Scale-Youth Form (SSS-YF), which measures the stress sources of young never-married individuals, was examined. Data were collected online from 600 (90 males, 510 females) young never-married individuals living in Turkey, with at least high school education, with an average age of 20.10, during the years 2020-2021. Construct validity of SSS-YF was determined by Confirmatory Factor Analysis, Criterion Validity was determined by Beck Depression Inventory, and reliability was determined by Cronbach's Aplha technique. The validity and reliability levels of the scale were found within the limits accepted in the literature.

Keywords: *singlehood, stress, singles counseling, youth, scale, validity, reliabilty*

Value-at-Risk Forecast and Its Extensions

Khreshna Syuhada
Institut Teknologi Bandung
khreshna@itb.ac.id

Arief Hakim
Institut Teknologi Bandung
arief.rahman41@sci.ui.ac.id

ABSTRACT

Value-at-Risk (VaR) has been an important risk system for financial and insurance institutions. The VaR forecast is crucial for reserving capital as well as for avoiding worse risk. In this paper, we present VaR forecasts for various models. Their VaR calculations are carried out based on both quantile- and expectile- approaches. Furthermore, we consider some heteroscedastic processes i.e. Autoregressive Conditional Heteroscedastic (ARCH) / Generalized-ARCH class and Stochastic Volatility Autoregressive (SVAR) from which financial data are modeled. There have been some extensions of VaR which are developed from different directions. The first direction is improving VaR forecast to have better coverage property. In the second direction, the VaR is extended in capturing coherent property. The example of the former direction (we call this as probabilistic based risk measure) is improved VaR (imprVaR) whilst the latter (expected based risk measure) is conditional VaR (CVaR). In fact, the CVaR forecast may also be extended as Modified CVaR and Dependent CVaR. Some numerical simulations are demonstrated in order to have some figures on VaR and its extensions, particularly on how they are computed and backtested.

Keywords: *risk measure, stochastic volatility, coherent property, improved var*

Mortality-at-Risk Forecast based on APARCH Model and Its Application to Determine Reserve Amount in Pension Fund

Darin Sabrina
Institut Teknologi Bandung
20821009@mahasiswa.itb.ac.id

Venansius R. Tjahjono
Institut Teknologi Bandung
20821026@mahasiswa.itb.ac.id

Khreshna Syuhada
Institut Teknologi Bandung
khreshna@itb.ac.id

ABSTRACT

Mortality represents the observed number of deaths in a group of individuals in a specific time horizon. Regardless, to represent the index of mortality on an objective basis, we can calculate it through the mortality rate. The mortality rate is a measure of the proportion of deaths in a particular population amount. In insurance modeling, the mortality rate has become one of the fundamental topics. If the mortality rate follows a downward trend, it is possible to create a longevity risk. The longevity risk may produce a loss due to more expenses than expected in the pension fund. Hence, we need to predict the mortality rate for calculating the minimum reserve amount to cover the future loss. In this study, the specific mortality rate measure that we employed is the central death rate. We also contribute a novel analysis of the stochastic model accommodating the conditional heteroscedasticity effect to model the mortality rate dynamics. We rely on a stochastic model which is Asymmetric Power Autoregressive Conditional Heteroscedasticity (APARCH) in the first order. This study utilizes data from the United States population for the total number of deaths and the number of individuals exposed to the death risk at age 67 to 100 in the calendar year 1933 to 2019. Parameter estimation that we use for estimating the model is the Maximum Simulated Likelihood Estimation-Simplex Search (MSLE-SS). We then construct the Mortality-at-Risk (MaR) as a risk measurement based on Value-at-Risk (VaR) that is derived from the APARCH(1,1) model to determine the minimum reserve amount for pension fund industries.

Keywords: *mortality rate, mortality-at-risk, longevity risk, heteroscedasticity, APARCH model, maximum likelihood*

Author Index

Keywords Index