

# International Conference on Data Science, Machine Learning and Statistics - 2019 (DMS-2019)\*

June 26-29, 2019

Book of Abstracts

## Conference Venue:

Faculty of Economics and Administrative Sciences,  
Van Yuzuncu Yil University, 65080 Van, Turkey.

\* This conference was supported by Scientific Research Projects Coordination Unit of Van Yuzuncu Yil University. Project number FTD-2019-7971.

# Preface

Dear Colleagues,

Welcome to the international conference on “Data Science, Machine, Learning and Statistics-2019 (DMS-2019)” held by Van Yuzuncu Yil University from June 26-29, 2019. The DMS-2019 shall create a forum to discuss recent progress and emerging ideas in areas of interest. During the conference, participants will have opportunities to discuss new theoretical and practical issues in their fields and to establish new collaborations. Presentations can cover topics such as advance soft computing, heuristic algorithms, data infrastructures and analytics which are recent advances in Data Science, Machine Learning and Statistics.

The scientific program of DMS-2019 contains eight invited lectures of renowned personalities in the field. We wish you a productive, stimulating conference and a memorable stay in Olomouc.

Editors  
H. Eray Celik  
Cagdas Hakan Aladag

### Scientific Committee

---

Abbas Mohamad Ali	Engin Avci	Mehmet Recep Minaz	Ramazan Tekin
Adem Kalinli	Erkan Aydar	Melih Kuncan	Rayimbek Sultanov
Adil Baykasoglu	Fatih Ozkaynak	Mete Celik	Resul Das
Ali Caglayan	Gratiela Dana Boca	Moayad Y. Potrus	Ridvan Saracoglu
Ali Karci	Guckan Yapar	Muhammed Baykara	Samer M. Barakat
Ali Rostami	Hamit Mirtagioglu	Murat Atan	Sedat Yerli
Alparslan A. Basaran	Hatice Hicret Ozkoc	Murat Demir	Serhat Omer Rencber
Alper Basturk	Ibrahim Kilic	Musa Atas	Sevil Senturk
Arash Kalami	Ibrahim Turkoglu	Mustafa Sevuktekin	Sinan Calik
Ashis SenGupta	Kambiz Majidzadeh	Naci Genc	Suat Ozdemir
Atilla Goktas	Kasirga Yildirak	Nasip Demirkus	Suat Toraman
Aydin Sipahioglu	Kazim Hanbay	Necmettin Sezgin	Sylwia Gwozdziejwicz
Burak Uyar	Keming Yu	Novruz Allahverdi	Sakir sleyen
Bulent Batmaz	Kenan nce	Nuh Alpaslan	Sengul Cangur
Candan Gkceoglu	M.H. Fazel Zarandi	Nuri Almali	Tahir Hanalioglu
Carlos A. Coelho	M. Kenan Dnmez	Olca Arslan	Timur Han Gur
Cetin Guler	M. Marinescu Mazurencu	Omed Salim Khalind	Ufuk Tanyeri
Davut Hanbay	M. Selim Elmali	Onur Koksoy	Veysel Yilmaz
Deniz Dal	Mahdi Zavvari	Orhan Ecemis	Yener Altun
Dervis Karaboga	Mehmet Kabak	Omer Faruk Ertugrul	Yildirim Demir
Ebru Akcapinar Sezer	Mehmet Karadeniz	Ozgur Yeniay	Yilmaz Kaya
Ebru Caglayan Akay	Mehmet Mendes	Ozgur Yilmazel	Zeki Yildiz

---

### Organization Committee

---

Birdal Senoglu	Fatma Gul Akgul	Fikriye Ataman	Serpil Sevimli Deniz
Fevzi Erdogan	Hanifi Van	Kubra Bagci	Taner Uckan
Suat Sensoy	Murat Canayaz	Onur Camli	Ali Yilmaz
Hayrettin Okut	Recep Ozdag	Seda Basar Yilmaz	Israfil Celik
Cetin Guler	Sukru Acitas	Ebubekir Seyyarer	Mesut Kapar
Hayati Cavus	Talha Arslan	Erol Kina	Aksel Akyurek
Fatma Zehra Dogru	Kadir Emir	Faruk Ayata	Hayrullah Urcan
Fuat Tanhan	Asuman Yilmaz Duva	Firat Kapar	
Sinan Saracli	Emre Bicek	Necati Erdogan	
Alper Hamzadayi	Fatih Uludag	Serbest Ziyanak	

---

## Contents

<b>Statistical Learning for Big Manifold Data</b>	<b>5</b>
<b>Data-Driven Multi-Criteria Decision Making in Decision Engineering</b>	<b>6</b>
<b>Quality Analysis Of Big Geodata Via Machine Learning</b>	<b>7</b>
<b>Machine Learning: from the Architect of Van Lake to Architecture of Neural Networks</b>	<b>8</b>
<b>Testing the “Complete Symmetrical Equivalence” of Two Sets of Variables</b>	<b>9</b>
<b>Using Deep Learning Models in Problem Solving</b>	<b>10</b>
<b>An Extreme-Value Distribution based Regression for Big Data</b>	<b>11</b>
<b>Robust and Sparse Methods for Regression and Classification in High Dimensions</b>	<b>12</b>
<b>Single Valued Triangular Neutrosophic Fuzzy c-Means for MR Brain Image Segmentation</b>	<b>13</b>
<b>An Improved Telecommunication Churn Prediction System by Enhanced Fuzzy Clustering with Ada Boosting Hybrid Model</b>	<b>14</b>
<b>Determinants of Export to Turkey’s Black Sea Economic Cooperation Member Countries: Panel Gravity Model Approach</b>	<b>15</b>
<b>A Combat Genetic Algorithm for Optimal Buffer Allocation In Unreliable Production Lines</b>	<b>16</b>
<b>Oil Prices and Exchange Rates Dynamics in Turkey</b>	<b>17</b>
<b>Comparison of Optimization Algorithms Used in Deep Learning by Using Caltech 101 Data Set</b>	<b>18</b>
<b>The Effect of Population Size on the Success of Genetic Algorithm in Optimizing the Ackley Function</b>	<b>19</b>
<b>Modeling of The Pan Evaporation Data Using Fuzzy Logic Method</b>	<b>20</b>
<b>A Statistical Comparison of Randic and Angular Geometric Randić Indices</b>	<b>21</b>
<b>On Reporting Statistical Analysis Results?</b>	<b>22</b>
<b>On Deep Learning Based Error Correction with Algebraic Codes</b>	<b>23</b>
<b>Detection of Clusters in Hierarchically Built Trees by Lifting</b>	<b>24</b>
<b>Panel Data Analysis: An Application to Dow-Jones Stock Market</b>	<b>25</b>
<b>Support Vector Machines and an Application on Natural Gas Consumptions of Power Plants in Turkey</b>	<b>26</b>
<b>Stochastic Programming: Theory, Techniques and Application</b>	<b>27</b>
<b>Optimization by Repulsive Forces Based on Charged Particles</b>	<b>28</b>
<b>The Effect of Maternal Education On The Probability of Pregnancy Termination</b>	<b>29</b>
<b>Distractor Analysis for Statistical Literacy Test</b>	<b>30</b>
<b>Benefits of Computer-Based Systems in Quality Design</b>	<b>31</b>

<b>Deep Neural Network and Its an Application</b>	<b>32</b>
<b>LDA-Based Aspect Extraction from Turkish Hotel Review Data</b>	<b>33</b>
<b>Impact of Manufacturing PMI on Stock Market Index: A Study on Turkey</b>	<b>34</b>
<b>Defective PV Cell Detection Using Deep Transfer Learning and EL Imaging</b>	<b>35</b>
<b>A Comparison of SVM Kernel Functions For Unbalanced Data</b>	<b>36</b>
<b>Comparison of Hot Deck and Regression Imputation in Multiple Imputation Methods for Missing Data Structures</b>	<b>37</b>
<b>Using Regression Analysis Methods in Biostatistics An Applied Study on A Sample of Diabetic Patients</b>	<b>38</b>
<b>Estimating the Parameters of the Bivariate Mixed Model Using Robust Method with Ordinary Method</b>	<b>39</b>
<b>Combining Forecasts for Stock Keeping Units with Intermittent Demand Pattern: An Application on Spare Parts</b>	<b>40</b>
<b>Sector-Wise Analysis of Cardinality Constraint Portfolio Optimization Problem: Selecting ISE-All Shares Based On Coefficient of Variation And Nonlinear Neural Network</b>	<b>41</b>
<b>Associative Classification For Failure Prediction in Aluminium Wheel Molding: A Case Study</b>	<b>42</b>
<b>Evaluation of Black Friday Hashtags in Turkey with Sentiment Analyse</b>	<b>43</b>
<b>Experimenting with Some Data Mining Techniques to Establish Pediatric Reference Intervals for Clinical Laboratory Tests</b>	<b>44</b>
<b>Forecasting The Industrial 4.0 Data Via Anfis Approach</b>	<b>45</b>
<b>Forecasting The Future Needs of Customers for New Products</b>	<b>46</b>
<b>Examination of Recommendation Systems and Usage Areas</b>	<b>47</b>
<b>Application of the Weighted K Nearest Neighbor Algorithm for Diabetes</b>	<b>48</b>
<b>A Study of Data Mining Methods for Breast Cancer Prediction</b>	<b>49</b>
<b>Pharmacy Students Intention towards Using Cloud Information Technologies in Knowledge Management</b>	<b>50</b>
<b>Generative Adversarial Networks Based Data Augmentation for Phishing Detection</b>	<b>51</b>
<b>Comparison of the Spliced Regression Models</b>	<b>52</b>
<b>Using Convolutional Neural Networks for Handwritten Digit Recognition</b>	<b>53</b>
<b>A Statistical Comparison of Zagreb and Angular Geometric Zagreb Indices</b>	<b>54</b>
<b>An Application of Type 2 Fuzzy Time Series Model</b>	<b>55</b>
<b>Heteroscedastic and Heavy-tailed Regression with Mixtures of Skew Laplace Normal Distributions</b>	<b>56</b>
<b>Modeling of Exchange Rate Volatility in Turkey: An Application with Asymmetric GARCH Mode</b>	<b>57</b>
<b>Comparison of Catalase and Superoxide Dismutase Enzyme Activities in Strawberry Fruit</b>	<b>58</b>

<b>Investigation of Some Antioxidant Enzyme Activities in Cherry Fruit Obtained From Various Regions</b>	<b>59</b>
<b>Prediction of Gastric Cancer Stages with Convolutional Neural Networks</b>	<b>60</b>
<b>Irony Detection in Turkish Tweets</b>	<b>61</b>
<b>A Machine Learning Sepsis Diagnosis Model for Intensive Care Units</b>	<b>62</b>
<b>Risk Classification with Artificial Neural Networks Models in Motor Third Party Liability</b>	<b>63</b>
<b>Inference in Step-Stress Partially Accelerated Life Testing for Inverse Weibull Distribution under Type-I Censoring</b>	<b>64</b>
<b>Modeling Dynamically Behavior of Users in Social Networks using Petri Nets</b>	<b>65</b>
<b>Qualitative Data: Advantages, How to Collect and Present Based on Three Examples on Health Issues</b>	<b>66</b>
<b>Research Methods In Social Pharmacy Studies</b>	<b>67</b>
<b>The Analysis of Web Server Logs with Web Mining Methods</b>	<b>68</b>
<b>An Extension of the Maxwell Distribution: Properties and Application</b>	<b>69</b>
<b>The Feasibility of Near Infrared Spectroscopy for Classification of Pine Species</b>	<b>70</b>
<b>A Comparison of Estimation Methods for The Inverted Kumaraswamy Distribution</b>	<b>71</b>
<b>Collection of Recyclable Wastes within the Scope of Zero Waste Project: A Heterogeneous Multi-Vehicle Routing Case in Krkkale</b>	<b>72</b>
<b>Analysis of Terrorist Attacks in The World</b>	<b>73</b>
<b>Analysis of the Factors Affecting the Financial Failure and Bankruptcy by the Generalized Ordered Logit Model</b>	<b>74</b>
<b>Using LSTM for Sentiment Analysis with Deeplearning4J Library</b>	<b>75</b>
<b>Network Log Analysis For Network Security By Using Big Data Technologies</b>	<b>76</b>
<b>Face Recognition Based System Input Control Application</b>	<b>77</b>
<b>Extractive Text Summarization Via Graph Partitioning</b>	<b>78</b>
<b>The Evaluation of the Effect of the Earthquake on Socio-Economic Development Level with the Cluster Analysis</b>	<b>79</b>
<b>Modelling of Photovoltaic Power Generation based on Weather Parameters Using Regression Analysis</b>	<b>80</b>
<b>Uniformly Convergence of Singularly Perturbed Reaction-Diffusion Problems on Shishkin Mesh</b>	<b>81</b>
<b>Factorial Moment Generating Function Of Sample Minimum Of Order Statistics From Geometric Distribution</b>	<b>82</b>
<b>Comparisons of Methods of Estimation Generalized Exponential Distribution</b>	<b>83</b>
<b>Interdependence Of Bitcoin And Other Crypto Money Indicators: Cd Vine Copula Approach</b>	<b>84</b>

<b>New Complex Hyperbolic Mixed Dark Soliton Solutions for Some Nonlinear Partial Differential Equations</b>	<b>85</b>
<b>Complex Solitons in the Conformable (2+1)-dimensional Ablowitz-Kaup-Newell-Segur Equation</b>	<b>86</b>
<b>Fitting Irregular Migration Data in Turkey to Itô Stochastic Differential Equation</b>	<b>87</b>
<b>Attitudes of Students Towards Coding Learning Supported With Educational Computer Games: A Case Study for Van Province</b>	<b>88</b>
<b>Design and Optimization of Graphene Quantum Dot-based Luminescent Solar Concentrator Using Monte-Carlo Simulation</b>	<b>89</b>
<b>A Review: Big Data Technologies with Hadoop Distributed Filesystem</b>	<b>90</b>
<b>Guuler and Linaro et al Model in an Investigation of the Neuronal Dynamics using noise Comparative Study</b>	<b>91</b>
<b>A Robust Confidence Interval and Ratios of Coverage to Width for the Population Coefficient of Variation: A Comparative Monte Carlo Study</b>	<b>92</b>

# Statistical Learning for Big Manifold Data

Ashis SenGupta<sup>1</sup>

<sup>1</sup> *Indian Statistical Institute, Kolkata, India, ashis@isical.ac.in*

The explosion of Big Data have attracted research from almost all areas of Science, with a panorama of approaches for solving problems of diverse nature. While fast and soft computations have been of prime concern, Data Science has been demanding scientific and objective analysis. In this backdrop, Statistical Science is playing an indispensable role. In this talk we focus on the three basic Vs of Big Data: Variety, Volatility and Volume. In the Variety aspect, we present Manifold data specifically Directional Data, where observations that can be mapped onto circles and spheres, as in astrophysics, bioinformatics, geosciences, text mining, etc. are considered and corresponding Probability models are constructed. Volatility is a prime characteristic of modern Big Data, often exhibited through multi-modality but not interpretable by mixture models. We take up the problem of modelling such data next. Volume of data, either in terms of the sheer size or in terms of high-dimensional random variables with lower dimensional sampling space, i.e. the Large  $p$  - Small  $n$  case, is a non-trivial problem to analyse. For the former, a hierarchical model-based clustering method is presented. The latter creates the nagging problem of singularity and multi-colinearity. We consider this problem through dimension reduction techniques. Novel approaches of Multivariate Statistical Inference for these problems are also briefly reviewed. Several emerging real-life examples are given to illustrate some of the above methods. It is hoped that this glimpse of the rich arena of scientific challenges and some objective and probabilistic solutions thereto as to be presented in this talk, will encourage the researchers in Big Data to explore the methods advocated and enhanced through Statistical Science.



# Data-Driven Multi-Criteria Decision Making in Decision Engineering

Adil Baykasoglu<sup>1</sup>

<sup>1</sup> *Dokuz Eylul University, Izmir, Turkey, baykasoglu@gmail.com*

A new data-driven Multi-Criteria Decision Making (MADM) model is introduced in this invited talk. The presented approach makes use of metaheuristic algorithms (Jaya algorithm) for training Fuzzy Cognitive Maps (FCMs) in order to enable learning from past data in a dynamic multi-criteria decision making scenario. Trained FCMs are used to predict future performance scores of decision alternatives by incorporating present subjective evaluations. The application of the proposed approach is also presented through an example.

# Quality Analysis Of Big Geodata Via Machine Learning

Candan Gokceoglu<sup>1</sup>

<sup>1</sup> *Hacettepe University, Ankara, Turkey, candan.gokceoglu@gmail.com*

The paradigm changes in geoscience researches have become obvious with the availability of big geodata. The new scientific concepts such as citizen science (CitSci) and crowdsourced data have started to appear in many different fields of geosciences, especially in environmental monitoring and assessment in the last decades. The benefits of volunteer geographic information (VGI), which have been evidenced by citizen supported research projects and crowdsourced geodata collected on social media platforms, have turned the topic into an essential requirement for many projects. At the same time, the use of machine learning (ML) algorithms have become unavoidable for processing of these data due to its magnitude. Although the ML algorithms can serve to the pre-processing, information extraction and interpretation of big geodata in general, the quality assessment (QA) and quality control (QC) is an active research agenda for many CitSci projects since their usability mainly depends on the reliability of the input. The CitSci and other crowdsourced geodata require specific QA&QC procedures prior to their use, since many outliers and errors may exist due mainly to: i) non-specialist data collectors and interpreters; ii) diversity of the data collection methods and the equipments; and iii) other errors caused during data transmissions. In this study, the importance, potential and existing applications of CitSci data quality analysis will be emphasized with a particular focus on ML approaches.

# **Machine Learning: from the Architect of Van Lake to Architecture of Neural Networks**

Cagdas Hakan Aladag<sup>1</sup>

<sup>1</sup> *Hacettepe University, Ankara, Turkey, chaladag@gmail.com*

Nowadays, machine learning concept is getting more and more interest. In recent years, machine learning has many application areas spread all over the life. These areas are ranging from entertainment to economy. Due to its wide usage area, there are lots of definitions for machine learning. All big organizations or governments are collecting huge data and try to analyze it efficiently to understand current state and to make inferences for the future. In recent years, one of the most efficient way to analyze data is using machine learning. Therefore, the questions “What exactly is machine learning?” and “What exactly does machine learning do” are vital questions to answer. On the other hand, it is not easy to get answers for these question in the literature. Furthermore, there are some misleading or confusing definitions for this important concept. In this study, machine learning and its links with other crucial topics such as artificial neural networks, big data, data science and statistics will be presented over real world. The important concepts are going to be explained via real world applications. In the light of these, a new artificial neural network model will be introduced to efficiently analyze real world data including outliers.

# Testing the “Complete Symmetrical Equivalence” of Two Sets of Variables

Carlos A. Coelho<sup>1</sup>, Barry C. Arnold<sup>2</sup>

<sup>1</sup> *Mathematics Department and Center for Mathematics and Applications (CMA), Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal, cmac@fct.unl.pt*

<sup>2</sup> *Statistics Department, University of California Riverside, CA, U.S.A., barnold@ucr.edu*

Let  $\underline{X} = [\underline{X}'_1, \underline{X}'_2]'$   $\sim N_{2p}(\underline{\mu}, \Sigma)$ , where both subvectors  $\underline{X}_1$  and  $\underline{X}_2$  are  $p$ -dimensional, and

$$\underline{\mu} = \begin{bmatrix} \underline{\mu}_1 \\ \underline{\mu}_2 \end{bmatrix} \quad \text{with} \quad \underline{\mu}_1 = E(\underline{X}_1) \quad \text{and} \quad \underline{\mu}_2 = E(\underline{X}_2).$$

We will say that  $\underline{X}_1$  and  $\underline{X}_2$  are “completely symmetrically equivalent” if

$$\Sigma = \begin{bmatrix} \Sigma_1 & \Sigma_2 \\ \Sigma_2 & \Sigma_1 \end{bmatrix} \quad \text{and} \quad \underline{\mu}_1 = \underline{\mu}_2, \quad (1)$$

where  $\Sigma_1$  and  $\Sigma_2$  are non-specified, but with  $\Sigma_1$ ,  $\Sigma_1 + \Sigma_2$  and  $\Sigma_1 - \Sigma_2$  being positive-definite matrices. We will be interested in testing the null hypothesis in (1), which we will call the “complete symmetrical equivalence” hypothesis.

The authors derive the likelihood ratio test statistic for the present test and show how it is possible to include the distribution of this statistic under the framework of Theorem 3.2 in [1] and as such have its exact probability density and distribution functions given by Corollary 4.2 in the same reference, in a finite closed form, in terms of an EGIG (Exponentiated Generalized Integer Gamma) distribution, with all parameters precisely defined, as simple functions of  $p$  and of the sample size. An example of the implementation of the test, using real data, is presented.

## References

- [1] Coelho, C. A., Arnold, B. C. (2019). Finite form representations for Meijer  $G$  and Fox  $H$  functions – Applied to multivariate likelihood tests using Mathematica<sup>®</sup>, MAXIMA and R, 488+xixpp. Springer Texts in Statistics (accepted for publication).

# Using Deep Learning Models in Problem Solving

Ibrahim Turkoglu<sup>1</sup>

<sup>1</sup> *Firat University, Elazig, Turkey, iturkoglu@firat.edu.tr*

The importance of artificial intelligence has been increased since the developing technology and digital era. The development of autonomous systems which can make decisions on its own serves both humanity and provides new job opportunities. Artificial intelligence can be defined as software and hardware systems that exhibit human behaviors, conduct numerical logic, and have many abilities including movement, speech and voice recognition. In general, AI can be divided into two groups: Machine learning and deep learning. Machine learning represents an algorithmic structure which learns from examples of qualitative information extracted from the data. Yet, deep learning systems learn from the data without training. In this study, problem solving approaches are given by applying artificial intelligence techniques and deep learning models used the applications. Leaf classification, colon cancer detection and epileptic seizure recognition are presented to solve problems with artificial intelligence techniques. Leaves were classified with deep learning models and the performance of deep learning models was compared. In that study, AlexNet, VGG-16, VGG-19, ResNet50, and GoogleNet were applied. At the end of the study, 32 leaf labels (classes) were considered to discriminate and all models showed over 95% classification performance. The best detection result obtained with AlexNet model with 99,72% accuracy [1]. In the second study, feature extraction was carried with deep learning models in order to determine the colon cancer risks from the FTIR signals. One of the difficulties encountered in determining the cancer from the blood is the similarity of FTIR signals between patients and healthy individuals. In this study, the feature extraction problem from FTIR signals was emphasized and a new method was proposed based on deep learning model. In the proposed method, spectrogram images of FTIR signals were obtained and features were collected with AlexNet. Then SVM was performed in order to classify the images and 90% accuracy was evaluated [2]. In the last study, epileptic seizures based on EEG signals were classified with deep learning models. EEG signals are long and nonlinear time series thus it is difficult to analyze and it takes time to comprehend the signals with traditional methods. In order to overcome that problem, deep learning model was proposed. In the first part of the study, both focal and non-focal EEG signals were converted into images. After, both VGG-16 and AlexNet deep learning models were applied to classify the images and their performance were compared. Best result obtained from AlexNet with 91,72 classification accuracy [3]. Deep learning models and techniques have been used in different engineering problems recently. In this study, the usage of deep learning models to solve different kind of engineering problems has been explained with various applications.

## References

- [1] Dogan, F., Turkoglu, I., (2018). The Comparison of Leaf Classification Performance of Deep Learning Algorithms. *Sakarya University Journal of Computer and Information Sciences*, SAUCIS-1, 10-21.
- [2] Toraman, S., Turkoglu, I., (2018). Determination of Colon Cancer Risk from FTIR Signals by Deep Learning, *Science and Engineering Journal of Firat University*, 30(3), 115-120.
- [3] Alakus, T.B., Turkoglu, I., (2019). Fokal ve Fokal Olmayan Beyin Sinyalleriyle Derin Ogrenme Kullanarak Epilepsi Nobeti Tahminin Yapilmasi, *Anadolu 2. Uluslararası Uygulamali Bilimler Kongresi, UBAK*, 458 469, 26-28 Nisan, Diyarbakir, Trkiye.

# An Extreme-Value Distribution based Regression for Big Data

Keming Yu<sup>1</sup>

<sup>1</sup> *Brunel University, London, United Kingdom, keming.yu@brunel.ac.uk*

Generalised Pareto (GP) distribution is the widely used extreme-value distribution. One of the most important types of statistical methods for gathering information for decision making is regression, when covariate information is available.

In this talk, the GP regression is introduced via the prime parameter (extreme index or shape parameter) of GP distribution. A novel least-squares type of linear regression for an easy adaptation of ‘divide and conquer’ algorithm for the GP regression to cope with analysis of massive data as well as its adaptation of lasso-type procedures in high-dimensional variable selection is provided.

# Robust and Sparse Methods for Regression and Classification in High Dimensions

Peter Filzmoser<sup>1</sup>, F.S. Kurnaz<sup>2</sup>, I. Hoffmann<sup>1</sup>

<sup>1</sup> Vienna University of Technology, Vienna, Austria, P.Filzmoser@tuwien.ac.at

<sup>2</sup> Yildiz Technical University, Istanbul, Turkey, fskurnaz@yildiz.edu.tr

The advances in technological developments led to a vast increase of statistical data. For example, it is nowadays relatively easy to install additional sensors in order to monitor a production process. This means that the datasets grow in size, particularly in dimension. Although more information can be valuable for understanding phenomena, this also results in two challenges: (a) it becomes harder or even impossible to visually clean the data from outliers which might spoil the results of the statistical analysis, (b) it is usually impossible with traditional tools like correlations to figure out which of the measured variables are relevant for the problem.

Concerning the latter issue, a lot of progress was made with so-called sparse estimators, such as the elastic net estimator [5], which allows to filter irrelevant noise variables for a range of different models, e.g. for linear or logistic regression. The goal here is to combine sparse estimators with robustness against outlying observations [2, 1]. Therefore, we propose a robust elastic net version which is based on trimming [4]. It is shown how outlier-free data subsets can be identified and how appropriate tuning parameters for the elastic net penalties can be selected. A final reweighting step is proposed which improves the statistical efficiency of the estimators. Simulations and data examples underline the good performance of the newly proposed method, which is available in the R package enetLTS on CRAN [3].

## References

- [1] Hoffmann, I., Filzmoser, P., Serneels, S., and Varmuza, K. (2016). Sparse and robust PLS for binary classification. *Journal of Chemometrics*, 30, 153-162.
- [2] Hoffmann, I., Serneels, S., Filzmoser, P., and Croux, C. (2015). Sparse partial robust M regression. *Chemometrics and Intelligent Laboratory Systems*, 149, 50-59.
- [3] Kurnaz, F. S., Hoffmann, I., and Filzmoser, P. (2018). *enetLTS: Robust and Sparse Methods for High Dimensional Linear and Logistic Regression*. R package version 0.1.0, <https://CRAN.R-project.org/package=enetLTS>.
- [4] Kurnaz, F. S., Hoffmann, I., and Filzmoser, P. (2018). Robust and sparse estimation methods for high dimensional linear and logistic regression. *Chemometrics and Intelligent Laboratory Systems*, 172, 211-222.
- [5] Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B*, 67(2), 301-320.

# Single Valued Triangular Neutrosophic Fuzzy c-Means for MR Brain Image Segmentation

A. Namburu<sup>1</sup>, S. Chakkaravarthy<sup>2</sup>, V. M. Cruz<sup>3</sup>, H. Seetha<sup>4</sup>

<sup>1</sup> Vellore Institute of Technology, Amaravati, India, namburi.anupama@vitap.ac.in

<sup>2</sup> Vellore Institute of Technology, Amaravati, India, chakkaravarthy.sibi@vitap.ac.in

<sup>3</sup> Keene State College, Keene, New Hampshire Meenalosini.vimal.cruz@keene.edu

<sup>4</sup> Vellore Institute of Technology, Amaravati, India, seetha.hari@vitap.ac.in

Medical image segmentation place a vital role in early detection and diagnosing the disease. Recent days, many researchers are working on enhancing the result of segmentation, which are crucial for treatment planning. Segmenting brain images is challenging due to the presence of noise and intensity in-homogeneity that creates uncertainty in segmenting the tissues. Neutrosophic sets are efficient tools to address these uncertainties present in the images. In this paper, a novel single valued triangular neutrosophic fuzzy c-means algorithm is proposed to segment the magnetic resonance brain images. The image is represented with triangular neutrosophic sets to obtain truth, falsity and indeterminacy regions which are further used to obtain the centroids and membership function for applying fuzzy c-means to extract the tissues of the brain. The experimental results reveal that the proposed work out performs the other relevant methods.

## References

- [1] Namburu, A, Samayamantula, S. K. and Edara, S. R. (2017). Generalised rough intuitionistic fuzzy c-means for magnetic resonance brain image segmentation. *IET Image Processing*, 11(9), 777-785.
- [2] Alsmadi, M. K. (2016). A hybrid Fuzzy C-Means and Neutrosophic for jaw lesions segmentation. *Ain Shams Engineering Journal*, 9(4), 697-706.



# **An Improved Telecommunication Churn Prediction System by Enhanced Fuzzy Clustering with Ada Boosting Hybrid Model**

J. Vijaya<sup>1</sup>, S. Srimathi<sup>2</sup>, E. Ajith Jubilson<sup>3</sup>

<sup>1</sup> VIT Univeristy, Andra Pradesh, India, vijayacsdept@gmail.com

<sup>2</sup> K.Ramakrishnan College of Engineering, Tamil Nadu, India, csuccessforsure2025@gmail.com

<sup>3</sup> VIT Univeristy, Andra Pradesh, India, ajith.jubilson@vitap.ac.in

Churn in the broadest sense indicates a quantity of people or customers moving collectively out of the subscribed services in a specific time frame under a single business organization. The objective of Churn Management is mainly to diminish the potential customer losses so that the organization becomes profitable. For the development of the clientele, the growth rate indicating the number of new customers needs to be improved than the churn rate. Hence Churn Prediction plays a crucial dynamic role that paves the way for the sustainable growth of the organization importantly in the ever challenging telecommunication industry. The technology era in which telecommunication extends as a key industry for the positive economic impact. Hence, using enhanced data mining techniques a viable churn prediction system is designed. The ensuing goal of every organization is to preserve the prevailing customer base, because accumulating new customers may involve enormous money investment, time consuming and also may pile up more of human resource. Hence it is interest of many industries and research people to focus on active churn prediction research. This paper focuses on proposing and designing a system to probable prediction of customer churn with the help of Hybrid Possible and Probabilistic fuzzy C-means Clustering methodology (PPFCM) combined in hand with Ada Boosting (PPFCM BOOSTING). This system effectively promotes the prediction accuracy in comparison with the existing PPFCM-ANN model. This paper has two modules: (1) Suggesting clustering module on the basis of PPFCM and (2) Suggesting churn prediction module on the basis of Boosting Ensemble Classifier. In the clustering module, the input dataset is grouped into clusters with the benefit of the PPFCM clustering algorithm. The clustered information obtained is deployed in the Boosting Ensemble Classifier and this hybrid creation is further deployed in the churn prediction. In the testing phase, based on the similarity measures or minimum distance the clustered data identifies the most relevant and accurate Boosting Classifier that signifies the nearby cluster of the test data. Finally to predict the customer churn the output score is used. Four different experiments are performed in which the primary experiment consists of a PPFCM clustering algorithm, the secondary experiment evaluates the classification result, and the tertiary experiment evaluates the ensemble classification result, Final experiment substantiates the hybrid proposed model. Such that the proposed Hybrid PPFCM-Ensemble Model affords maximum accuracy in comparison with any solitary models.

# Determinants of Export to Turkey's Black Sea Economic Cooperation Member Countries: Panel Gravity Model Approach

M. Sevuktekin<sup>1</sup>, B. Demirci<sup>2</sup>, I. Kiriskan<sup>3</sup>

<sup>1</sup> Bursa Uludag University, Bursa, Turkey, sevuktekin@uludag.edu.tr

<sup>2</sup> Bursa Uludag University, Bursa, Turkey, sevuktekin@uludag.edu.tr

<sup>3</sup> Giresun University, Giresun, Turkey, isin.cetin@giresun.edu.tr

In 2000, Turkey's total exports amounted to 27.5 billion dollars and in 2017 its value was close to 157 billion dollars. The Black Sea Economic Cooperation was officially established in 1992 with 11 constituent countries. 8.2% of Turkey's total exports towards the Black Sea Economic Cooperation Organization countries took place during the period from 2000 to 2017. The purpose of this study is to explain the determinants of Turkey's exports to the BSEC countries for the period from 2000 to 2017.

This study uses the extended gravity model, which is frequently used to explain the determinants of exports and imports. The variables employed to describe the gravity model are: Turkey and BSEC countries GDP, population, distance and common language (as dummy variable).

The data for the period from 2000 to 2017 were analyzed using panel data models. Only the results of the random effects model are relevant for the purpose of this study, because the fixed effects model excludes variables (distance and common language) that do not change over the years. This choice is also in accordance with the Hausman test results.

According to the random effects model results, Turkey and partner country GDP are positively related, whereas there is a negative effect on distance variable. Furthermore, Turkey's population has a positive effect on exports, but the partner country's population has negative effect on exports. The use of common language (as dummy variable) has been found to have positive effect on exports. The results of the analysis are in agreement with the literature findings on gravity model.

# A Combat Genetic Algorithm for Optimal Buffer Allocation In Unreliable Production Lines

M. U. Koyuncuoglu<sup>1</sup>, L. Demir<sup>2</sup>

<sup>1</sup> Pamukkale University, Denizli, Turkey, [ulas@pau.edu.tr](mailto:ulas@pau.edu.tr)

<sup>2</sup> Pamukkale University, Denizli, Turkey, [ldemir@pau.edu.tr](mailto:ldemir@pau.edu.tr)

A production line consists of machines connected in series. In unreliable production lines the machines may be randomly fail, and it causes disruptions in the production process. To prevent these disruptions buffers are allocated between the machines. The buffer allocation problem (BAP) is defined as finding the optimal size and location of the buffers throughout the line so as to achieve a predefined objective function. In this study, the buffer allocation problem is solved to maximize the production rate of the line under the total buffer size constraint. The mathematical formulation of the problem is given as follows:

$$\begin{aligned} & \max PR \\ & \sum_{i=1}^{K-1} N_i, \quad N_i \geq 0 \quad \text{and} \quad \text{integer} \end{aligned} \tag{1}$$

In equation (1), PR denotes the production rate of the line, K is the number of machines in the line, N is the total buffer size to be allocated and  $N_i$  depicts the buffer size for each buffer space. The buffer allocation problem is hard because of two reasons: (1) There is no algebraic relation between the buffer sizes and the production rate of the line [1], and (2) the problem is a nonlinear integer programming problem and it is in the class of NP-hard [2]. Because of these properties of the problem meta-heuristic search methods are widely used to solve BAP .

In this study, we employed the combat genetic algorithm (CGA), which is an evolutionary algorithm proposed by Eksin and Erol [3], for solving BAP. The combat genetic algorithm is based on the genetic algorithms (GAs) and the basic idea of the algorithm is to improve the convergence rate by focusing on the reproduction stage of a standard GA. The performance of the proposed combat genetic algorithm is tested on benchmark problems taken from the literature. The numerical results showed that the proposed CGA produced higher production rates than other methods used in literature for all considered benchmark cases.

## References

- [1] Chow, W.M. (1987). Buffer capacity analysis for sequential production lines with variable process times. *International Journal of Production Research*, 25 (8), 1183-1196.
- [2] Singh, A., and Smith, J.M. (1997). Buffer allocation for an integer nonlinear network design problem. *Computers & Operations Research*, 24 (5), 453-472.
- [3] Eksin, I., and Erol, O.K. (2001). Evolutionary algorithm with modifications in the reproduction phase. *IEE Proceedings - Software*, 148 (2), 75-80.

# Oil Prices and Exchange Rates Dynamics in Turkey

O. Ozturk<sup>1</sup>, E. Golveren<sup>2</sup>

<sup>1</sup> *Ataturk University, Department of Economics, ozcan.ozturk@atauni.edu.tr*

<sup>2</sup> *Ataturk University, Department of Economics, elifgolveren@gmail.com*

The link between oil prices and exchange rates has been studied frequently. The motivation is to find dynamics, causality and predictability between the variables. The literature on Turkey is rather thin and inconclusive. This study takes a comprehensive approach to examine the volatility effects between oil prices and exchange rates both in the short-run and in the long-run in Turkey. Additionally, the paper examines whether there is a causal relationship between oil prices and exchange rates and whether one can be predicted using the other.

For this study, we used monthly time series oil prices and exchange rates data obtained from the Central Bank of Turkish Republic and employed Granger causality and Auto Regressive Distributed Lag (ARDL) approach that estimates both the short and the long run parameters. Our initial results indicates that oil price movements negatively affect exchange rates both in the short-run and in the long-run with former being stronger and latter being weaker. Results also show that there is one way Granger causality running from oil prices to exchange rates. Based on these results, we discuss some policy implications.

# Comparison of Optimization Algorithms Used in Deep Learning by Using Caltech 101 Data Set

E.Seyyarer<sup>1</sup>, F. Ayata<sup>1</sup>, T. Uckan<sup>1</sup>, C. Hark<sup>2</sup>, A. Karci<sup>3</sup>

<sup>1</sup> *Van Yuzuncu Yi University, Department of Computer Technology, Van, Turkey  
eseyyarer@gmail.com, ayatafaruk@gmail.com, taner\_uckan@hotmail.com*

<sup>2</sup> *Bitlis Eren University, Department of Computer Technology, Bitlis, Turkey, cengizhark@hotmail.com*

<sup>3</sup> *Inonu University, Department of Computer Engineering, Malatya, Turkey, ali.karci@inonu.edu.tr*

In our previous study, an artificial neural network was applied to the caltech 101 data set which is an international data set, the results were analyzed and converted into a publication. In this study, image preprocessing, segmentation and feature selection were performed. 7 invariant moments applied to geometric and colorless images were used for feature selection. The success rate in the classification was observed to be about 25%.

In this study, deep learning and optimization techniques were applied on the same data set. Relu was used as activation function and cross entropy was preferred as loss function. Images are resized to 64x64. Each time the program is run, a random 6-category image is taken and 100 iterations are executed. Stochastic gradient descent (sgd), momentum, adam, adagrad, rmsprop and adadelata optimization algorithms were used for different results and these results were analyzed. The success rates in the classification were as follows: sgd: 64.5%, momentum: 85.56%, adam: 92.31%, adagrad: 71.25%, rmsprop: 40.26% and adadelata: 86.88%.

# The Effect of Population Size on the Success of Genetic Algorithm in Optimizing the Ackley Function

M. N. Qamari<sup>1</sup>, A. O. Kizilcay<sup>2</sup>, R. Saracoglu<sup>3</sup>

<sup>1</sup> *Van Yuzuncu Yi University, Van, Turkey, noor93noor93noor93@gmail.com*

<sup>2</sup> *Van Yuzuncu Yil University, Van, Turkey, oguzkizilcay@yyu.edu.tr*

<sup>3</sup> *Van Yuzuncu Yil University, Van, Turkey, ridvansaracoglu@yyu.edu.tr*

Genetic algorithm is a population-based method widely used in optimization problems. In this method, the choice of population size is very important. Increased population size can increase success but adversely affect working time. In this study, the Ackley function was chosen as a test function to determine the population size of the genetic algorithm. In addition, in order to make a fair comparison between population sizes, the total number of individuals dealt with in the genetic algorithm process for each alternative was fixed. Significant results have been obtained in terms of revealing the effect of population size on success.

# Modeling of The Pan Evaporation Data Using Fuzzy Logic Method

N. Ucler<sup>1</sup>

<sup>1</sup> *Van Yuzuncu Yil University, Van, Turkey, nadireucler@yyu.edu.tr*

In this study, it is aimed to model the evaporation data, which is one of the important parameters of the hydrological cycle, using the Fuzzy Logic Method, which depends on both the meteorological factors and the characteristics of the evaporation surface, such as temperature, humidity, precipitation, air pressure, solar radiation, water temperature and salinity.

In order to set the model, average daily temperature (oC), average daily relative humidity (%), average daily actual pressure (hPa), average daily wind velocity (m/s) were selected as input parameters and total daily pan evaporation (mm) was selected as output parameter.

Daily data for the year between 2013-2018 of the Van Local Station numbered 17172 were used as normalized. Data for 2017 are used for training purposes and 2018 data for testing purposes. In Sugeno type of Fuzzy Logic approach, sub clustering method is used, and 10 rules are written being created 10 membership functions for each input and output.

For the purpose of evaluating the performance of the model, the Van Local Station's daily data in 2018 and the Konya Airport Station which has similar meteorological features with Van and the data of the Kocaeli Station with different meteorological features were used.

The average error value was determined as 0,1 at Van Local Station, 0,15 at Konya Airport Station and 0,33 at Kocaeli Station.

# A statistical comparison of Randić and Angular Geometric Randić Indices

S. Ediz<sup>1</sup>, M. S. Aldemir<sup>2</sup>, M. Cancan<sup>3</sup>

<sup>1</sup> Van Yuzuncu Yil University, Van, Turkey, [suleymanediz@yyu.edu.tr](mailto:suleymanediz@yyu.edu.tr)

<sup>2</sup> Van Yuzuncu Yil University, Van, Turkey, [msaldemir@yyu.edu.tr](mailto:msaldemir@yyu.edu.tr)

<sup>3</sup> Van Yuzuncu Yil University, Van, Turkey, [mcancan@yyu.edu.tr](mailto:mcancan@yyu.edu.tr)

Topological indices have important role in theoretical chemistry for QSPR researches. Among the all topological indices the Randić index [1] has been used more considerably than any other topological indices in chemical and mathematical literature. Most of the topological indices as in the Randić index are based on the degrees of the vertices of a connected graph. Recently a novel degree concept has been defined in graph theory; geometric degree [2]. In this study, angular geometric Randić index is defined by using geometric degree concept as parallel to their corresponding classical degree version. This new angular geometric Randić index is compared with the Randić index by correlation efficient of some physicochemical properties of octane isomers. Also the exact values of the angular geometric Randić index for the well-known graph classes such as; paths, cycles, stars and complete graphs are given.

## References

- [1] Randić, M. (1975). Characterization of molecular branching. *Journal of the American Chemical Society*, 97, 6609-6615.
- [2] Ediz, S. (2019). A note on geometric graphs. *International Journal of Mathematics and Computer Science*, 14 (3), 631-634.



# On Reporting Statistical Analysis Results?

M. Mendes<sup>1</sup>, H. Mirtagioglu<sup>2</sup>

<sup>1</sup> *Canakkale 18 Mart University, Canakkale, Turkey, mmendes@comu.edu.tr*

<sup>2</sup> *Bitlis Eren University, Bitlis, Turkey, hamit\_1968@hotmail.com*

Statistics is one of the most important tools for all branches of sciences. It is because, there is no other tool, except statistics, to make sense the results of scientific studies. Since statistical analysis results serve a bridge between the researchers and readers, the results of the statistical analysis should be reported as informative as possible. In this study, it has been focused on how to report statistical analyses properly and informatively. For this purpose, results of four different studies have been used. At the end of this study, the researchers and scientists have been given the message it is also highly important that reporting the statistical analyses results properly alongside choosing an appropriate statistical test(s) in analyzing data sets and correct interpretation of the results.

# On Deep Learning Based Error Correction with Algebraic Codes

S. Bedir<sup>1</sup>

<sup>1</sup> *Yildiz Technical University, Istanbul, Turkey, sbedir@yildiz.edu.tr*

Error correcting codes have been used for decades in terms of reliable information transmission through noisy channels. Algebraic coding theory deals with topics varying among finding new codes, building new code construction methods, exploring optimal parameters and examining new criss-cross relationships between known construction methods and existing algebraic structures [3, 2, 1, 4]. Involving practical applications related to Electronics, Communication and Computing, this area of research have been widely inter-disciplinary, rather than being of pure Mathematics or Algebraic interests.

Deep learning techniques with the power of Artificial Intelligence(AI) have found applications everywhere and decoding of error correcting codes is one of this areas. In its very brief version, prediction and recognition techniques, after its early machine learning days, have nowadays reached to a human-level status with the power of neural networks [5].

Recently, channel decoding with deep learning techniques have attracted many researchers. [6, 7, 8, 9]. Specifically, it has been proposed that codes with structure, show better performance compared to random codes [7]. In this study we review some recent results on neural network decoding with algebraic codes and address new codes that are promising for deep learning based error correction in this view. We discuss the pros and cons of using deep learning techniques in decoding procedures.

## References

- [1] MacWilliams, F. J., and Sloane, N. J. A. (1977). *The Theory of Error-Correcting Codes*, NorthHolland.
- [2] Peterson, W. W., and Weldon, E. J. (1972). *Error Correcting Codes*, 2nd edn, MIT Press.
- [3] Xing, C., and Ling, S. (2003). *Coding Theory: A First Course*. Cambridge University Press.
- [4] Richardson, T., and Urbanke, R. (2008). *Modern Coding Theory*. Cambridge University Press.
- [5] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*, MIT Press.
- [6] Nachmani, E., Beery, Y., and Burshtein, D. (2016, September). Learning to decode linear codes using deep learning, 54th Annual Allerton Conf. On Communication, Control and Computing, Mouticello, arXiv:1607.04793.
- [7] Gruber, T., Cammerer, S., Hoydis, J., and Brink, S. (2017). On deep learning based channel decoding, CISS 2017, arXiv:1701.07738.
- [8] Nachmani, E., Bachar, Y., Marciano, E., Burshtein, D., and Beery, Y. (2018). Near maximum likelihood decoding with deep learning, CoRR 2018, arXiv:1801.02726.
- [9] Bennatan, A., Choukroun, Y., and Kisilev, P. (2018). Deep learning for decoding of linear codes - a syndrome-based approach, CoRR 2018.

# Detection of Clusters In Hierarchically Built Trees By Lifting

N. Bozkus<sup>1</sup>, S. Barber<sup>2</sup>

<sup>1</sup> *Giresun University, Giresun, Turkey, nebahat.bozkus@giresun.edu.tr*

<sup>2</sup> *University of Leeds, Leeds, UK, S.barber@leeds.ac.uk*

An important question in clustering, where the aim is to keep related objects in the same group, is how to detect the number of groups in a data set. If the groups are well separated and regularly shaped, available methods in the literature detect true clusters with a high performance. However, if groups overlap or have unusual shapes, the performance of these methods deteriorates. We propose a new method based on lifting which has recently been developed to extend the denoising abilities of wavelets to data on irregular structures. By checking all possible clustering patterns in a hierarchically built tree, our method seeks for the best representation of the clustering scheme in the tree. After denoising the tree, if the leaves under a node are all close enough to their centroid for the deviations to be explained as noise, we label those leaves as forming a cluster. The proposed method automatically decides how much departure can be allowed from the centroid of each cluster. Using some real and artificial data sets, we will illustrate the behaviour of our method.

# Panel Data Analysis: An Application to Dow-Jones Stock Market

S. Donmez<sup>1</sup>, O. Ozaydn<sup>2</sup>

<sup>1</sup> *Eskisehir Osmangazi University, Eskisehir, Turkey, sdonmez3@gmail.com*

<sup>2</sup> *Eskisehir Osmangazi University, Eskisehir, Turkey, oozaydin@ogu.edu.tr*

There has been 1066 articles with the keyword “Dow-Jones” until the month of April 2019. One of the recent articles working on the Dow-Jones market data is Novotny and Urga [2]. Novotny and Urga [2], suggested a method to model the meaningful price jumpings in every 5 minutes. Their data was obtained between first of January 2010 and 30th of June 2012. Their method was successful in predicting the price jumpings and suggested that taking the correlation of the price jumpings into account would make their model more successful. In another study Eckernkemper [1], tried to analyze the marginal expected fallings in the Dow-Jones market with a copula-based model. The model was successful in modelling nonlinear dependency and was worked on several variations of it. The model was compared to other models in the literature and was proven to be superior. In our study, we used the weekly Dow-Jones data from the UCI repository that was used by Brown et al. [3]. The data was comprised of two different timelines one of them being between January 2011 to March 2011 and the other being between April 2011 and June 2011. We modelled volumes of every stock in Dow-Jones market with a methodology that consists of panel data analysis, time series analysis and panel clustering. The panel clustering method we used was based on McNicholas [4]. The aim of our study was to set up a model that provided knowledge about volumes and then use this knowledge for price changes. We set up a mutual model for the two timelines and succeeded to produce this knowledge.

## References

- [1] Eckernkemper, T.(2018). Modelling systemic risk: Time-Varying tail dependence when forecasting marginal expected shortfall. *Journal of financial economics*, 16(1), 63-117.
- [2] Novotny, J., Urga, G.(2018). Testing for Co-jumps in Financial Markets, *Journal Of Financial Econometrics*, 16(1), 118-128.
- [3] Brown, M.S., Pelosi,M., Dirkska,H. (2013). Dynamic-radius Species-conserving Genetic Algorithm for the Financial Forecasting of Dow Jones Index Stocks. *Machine Learning and Data Mining in Pattern Recognition*, 7988, 27-41.
- [4] McNicholas, P.D. (2017). Mixture model-based classification, CRC press.

# Support Vector Machines and an Application on Natural Gas Consumptions of Power Plants in Turkey\*

G. Meral<sup>1</sup>, S. Saracli<sup>2</sup>

<sup>1</sup> Afyon Kocatepe University, Afyon, Turkey, [gizem.meral.1994@gmail.com](mailto:gizem.meral.1994@gmail.com)

<sup>2</sup> Afyon Kocatepe University, Afyon, Turkey, [ssaracli@gmail.com](mailto:ssaracli@gmail.com)

\* This study is a part of Gizem Meral's MSc thesis supervised by Sinan Saraccli at Afyon Kocatepe University Institute of Science and it is founded by Afyon Kocatepe University, Scientific Research Project Council. Project No: 18. FEN.BIL.15.

The aim of this study is to forecast the natural gas consumptions of power plants in Turkey via support vector machines regression method. With this aim the data set is obtained from Turkey's Energy Market Regulatory Authority and Energy Affairs General Directorate between the years 2013-2018.

In this study, first of all, the place in Turkish market, ratio within the primary energy sources, production, consumption, import and export values of natural gas, as a power supply is examined. Because of the differences in measurements of these values, the related data set is standardized before the statistical analysis. While, the consumption in energy plants is considered as a dependent variable, industrial consumption, city consumption, production, import and export values are considered as independent variables. All types of Kernel functions (Linear, Polynomial, Radial Basis Function (RBF) and Sigmoid) in Support Vector Regression are tested. RBF is chosen as the forecasting Kernel function because of having the minimum Mean Square Error (MSE). Then, support vectors, weights and decision constants are determined. By multiplying weights with support vectors and adding the bias, the final model is obtained.

By the help of final model, forecasts of natural gas consumption of power plants in Turkey, for May-December 2018 are obtained. The results are given in related tables and figures.

# Stochastic Programming: Theory, Techniques and Application

A. Sehitoglu<sup>1</sup>, S. Isleyen<sup>2</sup>

<sup>1</sup> *Mus Alparslan University, Mus, Turkey, a.sehitoglu@alparslan.edu.tr*

<sup>2</sup> *Van Yuzuncu Yil University, Van, Turkey, sakirisleyen@yyu.edu.tr*

A lot of real systems are subjected to the influence of random disturbances which cannot be ignored. Although many ways have been proposed to model uncertain quantities, stochastic models have proved their flexibility and usefulness in diverse areas of science.

Optimization problems involving stochastic models occur in almost all areas of science and engineering, so diverse as telecommunication, medicine, or finance, to name just a few. This stimulates interest in rigorous ways of formulating, analyzing, and solving such problems. Stochastic programming is one of the many specializations of Optimization [1, 2, 3]. Moreover, in recent years the theory and methods of stochastic programming have undergone major advances [4, 5].

In this review, we provide an overview on the main themes, methods and areas of application of this subject.

## References

- [1] Dantzig, G.B. (1955). Linear programming under uncertainty. *Management Science*, 1, 197-206.
- [2] Hillier, F.S., and Lieberman, G.J. (1995). Introduction to Mathematical Programming, second Edition.
- [3] Charnes, A., and Cooper, W.W. (1959). Chance-constrained programming, *Management Science*, 5, 737-9.
- [4] Stein, W.W., and William, T. Z. (eds.) (2005). Applications of Stochastic Programming. MPS-SIAM Book Series on Optimization 5.
- [5] Tamiz, M., and Jones, D.F. (1997). Interactive Framework for Investigation of Goal Programming Models: Theory and Practice. *Journal of Multi-Criteria Decision Analysis*, 6, 52-60.

# Optimization by Repulsive Forces Based on Charged Particles<sup>\*</sup>

G.Z. Oztas<sup>1</sup>, S. Erdem<sup>2</sup>

<sup>1</sup> Pamukkale University, Denizli, Turkey, gzeynepa@pau.edu.tr

<sup>2</sup> Dokuz Eylul University, Izmir, Turkey, sabri.erdem@deu.edu.tr

*\* This study is derived from the thesis called Evolutionary Algorithms for the Nonlinear Optimization by Sabri Erdem and remodeled.*

In the last decades, especially in the field of evolutionary optimization, most of the re- searches have focused on nature-based algorithms that are inspired by interactions of living and non-living objects. The main idea behind this is a belief that nature solves its problem instinctively like finding shortest path between foods and nests for ants and bees. Similarly, magnets with same pole will repulse each other while magnets with opposite pole will attract each other inherently. In the state-of-the-art, there are many algorithms that imitate these behav- iors and interactions for solving optimization problems in applied and social sciences such as: traveling salesman problems, assignment, transportation problems, scheduling, layout, conflict resolution, optimum policy making, portfolio optimization etc. In this study, we have imitated the behaviors of charged particles; see Erdem [1] for solving appropriate popular benchmark cases and have compared results with other ones generated by different algorithms. We have performed whole operations in Python coding environment and libraries. It is concluded that results are satisfactory and applicable for later usage in different areas.

## References

- [1] Erdem, S. (2007). Evolutionary Algorithms for the Nonlinear Optimization. Unpublished PhD Thesis, Dokuz Eylul University, The Graduate School of Natural and Applied Sciences, Izmir.

# The Effect Of Maternal Education On The Probability Of Pregnancy Termination

H. M. Karatas<sup>1</sup>, B. Harman<sup>2</sup>

<sup>1</sup> *Giresun University, Giresun, Turkey, haci.karatas@giresun.edu.tr*

<sup>2</sup> *Giresun University, Giresun, Turkey, berat.harman@giresun.edu.tr*

In this study, the causal effect of maternal education on the prevalence of pregnancy termination is examined using the 1997 compulsory schooling change and the corresponding regional variation in the number of middle school class opening. The effect of maternal education on several birth and pregnancy outcomes has been studied extensively. There are sound evidences on the positive effects of maternal education on fertility preferences, contraception use and infant health. This study focuses on establishing a causal link from maternal education to pregnancy termination or pregnancy loss. First, the causal link between pregnancy termination and education is examined in a broad-brush approach. Then, the effect is explored in details by the type of pregnancy termination. It is found that one additional year of maternal education reduces the probability of termination of pregnancy by 7 percentage points using two stage least squares method. Further examination shows that the one additional year of maternal education reduces the probability of having abortion by 5 percentage point. However, there is no significant effect of maternal education on miscarriage or stillbirth.



# Distractor Analysis for Statistical Literacy Test

B. Hasancebi<sup>1</sup>, Y. Terzi<sup>2</sup>, Z. Kucuk<sup>3</sup>

<sup>1</sup> Karadeniz Technical University, Trabzon, Turkey, burcuhsncb@gmail.com

<sup>2</sup> Ondokuz Mayıs University, Samsun, Turkey, yukselt@omu.edu.tr

<sup>3</sup> Karadeniz Technical University, Trabzon, Turkey, zkucuk@ktu.edu.tr

One of the most important component that should be examined in determining the quality of a multiple choice item is the power of distractors. If the distractors of a item are strong, performance of that item in the test is high. Besides, it is also important that the distractors do not contain any information about the correct answer. In a measurement tool prepared with quality items it is necessary to investigate why subjects check the distractors. Incorrect answer options, ie distractors, are not included in the calculation when estimating item statistics. The power of these wrong answer options, which are not taken into account when analyzing the distractor, is being investigated. Thus, it is possible to have an idea about how a item works in the measurement tool. In this study, a distractor analysis was performed for the items in the Statistical Literacy Test applied to the students of the Econometrics Department of the Faculty of Economics and Administrative Sciences of Karadeniz Technical University. R software was used for the analysis. According to the results obtained from the analysis based on the data obtained, the inter-test performance of each item was compared.

## References

- [1] D'sa, J.L., and Visbal-Dionaldo, M.L. (2017). Analysis Of Multiple Choice Questions: Item Difficulty, Discrimination Index And Distractor Efficiency. *International Journal of Nursing Education*, 9, (3).
- [2] Abdulgani, H.M., and Ahmad, F., Pennomperuma, G.G., Khalil, M.S., Aldress, A. (2014). The Reliationship Between Non Functioning Distractors And Item Difficulty Of Multiple Choice Questions: A Descriptive Analysis. *Journal of Health Specialties*, 2, (4), 148-151.
- [3] Hingorjo, M.R., Jaleel, F. (2012). Analysis of One-Best MCQs:The Difficulty Index, Discrimination Index and Distractor Efficiency. *Journal of Pakistan Medical Association*, 62, (2), 142-147.
- [4] <https://www.slideshare.net/anmeredithmd/item-analysis-69564841>

# Benefits of Computer-Based Systems In Quality Design

B. Aksu<sup>1</sup>, G. Yegen<sup>2</sup>, B. Mesut<sup>3</sup>

<sup>1</sup> Altinbas University, Istanbul, Turkey, [buket.aksu@altinbas.edu.tr](mailto:buket.aksu@altinbas.edu.tr)

<sup>1</sup> Altinbas University, Istanbul, Turkey

<sup>3</sup> Istanbul University, Istanbul, Turkey

Quality by Design (QbD), a systematic approach to pharmaceutical development, is designing and optimizing of formulation and production processes in order to provide the predefined product quality. Pharmaceutical product development is an intensive process in terms of data and knowledge and many tools can be utilized to perform QbD in this process. One of these tools is using mathematical models and guidelines to provide forming of the subject knowledge easily then to use it in independent or integrated style and to create Design of Experiments (DOE). For modelling the interactions of the variables there is an urge to collect experimental data which contains right range of inputs in order to achieve that statistical methods like factorial design. Response surface method (RSM), Artificial Neural Network (ANN), Genetic Algorithm (GA) are some of the assistive technologies used to perform mathematical modelling to provide forming of the subject knowledge easily then to use it in independent or integrated style [1]. Although, traditional statistical methods are quite helpful to examine relations in an extended data, when it is come to drug development they are not so sufficient, because of the complex multivariate relations between the elements that affect the quality of the product which are mostly nonlinear. The adoption of mathematical modelling via artificial intelligence programs has increased the efficiency of the development process with better understanding of the multivariate relations between the elements that affect the quality of the product [2, 3].

## References

- [1] Aksu, B., Gokce, E., Rencber, S., and Ozyazici, M. (2014). Optimization of solid lipid nanoparticles using gene expression programming (GEP). *Lat Am J Pharm*, 33(4), 675-84.
- [2] Aksu, B., Yegen, G., Purisa, S., Cevher, E., and Ozsoy, Y. (2014). Optimisation of ondansetron orally disintegrating tablets using artificial neural networks. *Trop J Pharm Res*, 13(9), 1374 -1383.
- [3] Aksu, B., De Matas, M., Cevher, E., Ozsoy, Y., Guneri, T., and York, P. (2012). Quality by design approach for tablet formulations containing spray coated ramipril by using artificial intelligence techniques. *Inter J Drug Del*, 4(1), 59-69.

# Deep Neural Network and Its an Application

S. Elasan<sup>1</sup>, S. Keskin<sup>2</sup>

<sup>1</sup> Yuzuncu Yil University, Van, Turkey, [sadielasan@yyu.edu.tr](mailto:sadielasan@yyu.edu.tr)

<sup>2</sup> Yuzuncu Yil University, Van, Turkey, [skeskin@yyu.edu.tr](mailto:skeskin@yyu.edu.tr)

Because there are more than one hidden layer between the input and output layers in the neural network algorithm, it is called “Deep Neural Networks”. In a classical neural information is only transferred from the previous layer to the next layer or exit. In the deep neural networks, the neurons affect each other with various activation values from two successive layers. The effect of each layer on the model and thus the neuron in each layer has an effect on the model [1, 2]. In the study, the Deep Neural Networks algorithm; different input (number of layers, epoch, error rate) and evaluation of performance of the model being practiced an application are intended.

“The rule of learning in deep neural networks” is a generalized version of the Delta Learning Rule based on the least-squares method. Generalized Delta Learning Rule consists of two stages. *Feed Forward*: This step start with the presentation of a learning instance to the network at the input layer. Incoming inputs are sent to the intermediate layer without any changes. The output of the k. element in the input layer is shown as  $C_k^i \cdot G_k$ . Firstly, Net (N) input the processing elements in the intermediate layer. The formula is calculated using  $N_j^a = \sum_{k=1}^n A_{kj} A_i^k$ . Sigmoid function is used, output is:  $C_j^a = 1 / 1 + \exp(- (N_j^a + \beta_j^a))$ . Back Propagation: The error for the m processing element in the output layer is;  $E_M = B_m + C_m$ . The sum of errors for the error (TH) occurring in the output layer is:  $TH = 1/2(\sum_m^2)$ . *Chaning Weights Between Interlayer and Output Layer*: If the amount of change is  $\Delta A^a$  in the weight of the connection connecting the processing element j-th in the intermediate layer to the processing element m-th output layer, the amount of change in weight at any time t is:  $\Delta A_{jm}^a(t) = \lambda \delta_m C_j^a + \delta \Delta A_{jm}^a(t-1)$  [2, 3, 4].

The most important feature that distinguishes the deep neural network method from the classical neural networks is the number of layers that provide good results in complex problems. As a result of the study, the non-layer artificial neural network model was classified with 66.9% accuracy and 32.9% MAPE, whereas the deep neural network model was classified with 95.5% accuracy and 4.9% MAPE ratio. The study showed that the model of deep neural networks had a higher accuracy rate.

## References

- [1] LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. *Nature*, 521, pp, 436-44.
- [2] Wang Y., Mao H., Yi Z. (2017). Protein secondary structure prediction by using deep learning method. *Knowledge-Based Systems*, 118, 115-123.
- [3] Schmidhuber J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, pp. 85117.
- [4] Guzel Y (2018). Cok Katmanli Yapay Sinir Agi. <https://medium.com/@yasinguzel/yapay-zeka-ders-notlar%C4%B1-5> Accessed: 25.03.2019.

# LDA-Based Aspect Extraction from Turkish Hotel Review Data

K. Bayraktar<sup>1</sup>

<sup>1</sup> Gazi University, Ankara, Turkey, bayraktarkivanc@gmail.com

Thoughts are the most important element affecting human life and enabling institutions and businesses to shape their future plans. As technology improves, we find opportunity to employ user data acquired through web resources to determine our daily basis, habits and decisions. Along with the rapidly increasing data sizes, data processing has become notably challenging. Therefore, the concept of sentiment analysis has emerged. Sentiment analysis has divided into three as document level, sentence level and aspect- based sentiment analysis [1]. Aspect-based sentiment analysis consists of two stages: target extraction and target classification [1]. In this study, LDA-based (Latent Dirichlet Allocation) aspect extraction methods have been proposed to identify single and multi- word aspects (MWA) for Turkish datasets, automatically. Introduced methods have been tested on a fragment of hotel dataset obtained via TripAdvisor.

LDA is a topic modeling method utilizing bag of words (BoW) to uncover hidden topics within the dataset [2]. BoW increases the efficiency, however, causes loss in inter- word semantic relationship information. Proposed LDA-based models consider these relationships without the need of human annotation. For preprocessing, the spelling correction infrastructure provided by the Zemberek and Yandex.XML were adopted. Suggested models benefit from C-value [3] and PMI (Pointwise Mutual Information) [4] techniques in order to address LDAs lack of bag of words and preserve relationships between words. In the first model, using linguistic and statistical information with C-value method has improved sensitivity on multi-word and nested terms. Candidate MWA terms with C-value above a threshold are filtered and selected as MWA. In the second model, the candidate MWAs were pointed out by the linguistic filtering method which detects the noun groups. PMI score between candidate MWAs and the data domain is calculated, and as in the first technique, candidates above threshold are replaced in the corpus and aspect extraction of two models are finalized with LDA. Proposed models provided more successful outcomes than classical LDA. Both models have increased accuracy, precision, recall and f-score values by approximately 20%, 15%, 15%, 15%, respectively.

## References

- [1] Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167
- [2] Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.
- [3] Frantzi, K., Ananiadou, S., and Mima, H. (2000). Automatic recognition of multi-word terms: the c-value/nc-value method. *International Journal on Digital Libraries*, 3(2), 115-130.
- [4] Turney, P. D. (2002, July). Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In Proceedings of the 40th annual meeting on association for computational linguistics (pp. 417-424). Association for Computational Linguistics.

# Impact of Manufacturing PMI on Stock Market Index: A Study on Turkey

R. Yanik<sup>1</sup>, A.B. Osman<sup>2</sup>, O. Ozturk<sup>3</sup>

<sup>1</sup> *Ataturk University, Erzurum, Turkey, ramazan.yanik@atauni.edu.tr*

<sup>2</sup> *Ataturk University, Erzurum, Turkey, asfiasoh@gmail.com*

<sup>3</sup> *Ataturk University, Erzurum, Turkey, ozcan.ozturk@atauni.edu.tr*

Purchasing Managers Index (PMI) is considered as an important factor to the policy makers and related bodies as it is found as an influential indicator for macro economy, especially for GDP growth and Industrial Added Value. PMI is one of the indicators to measure the health of an economy. This study examines whether the PMI has any influence on the stock market index of Turkey or vice versa. We use secondary data collected from the official website of BIST, Turkey. The study covers monthly data ranging from April 2015 to February 2019. We employ Granger Causality Test and Co-integration approach to examine causality and dynamics between the variables (Manufacturing PMI and BIST All Index Data). Our initial results show that there is uni-directional causality running from stock market index to Manufacturing PMI.

# Defective PV Cell Detection Using Deep Transfer Learning and EL Imaging

M. Y. Demirci<sup>1</sup>, N. Besli<sup>2</sup>, A. Gumuscu<sup>3</sup>

<sup>1</sup> *Harran University, Sanlurfa, Turkey, mydemirci@usr.aku.edu.tr*

<sup>2</sup> *Harran University, Sanlurfa, Turkey, nbesli@harran.edu.tr*

<sup>3</sup> *Harran University, Sanlurfa, Turkey, agumuscu@harran.edu.tr*

In order to achieve high efficiency in solar energy systems proper functioning of solar panels and cells is critical. There are several techniques that can be used to determine solar cell defects in PV modules both in the manufacturing process and in the field. Electroluminescent (EL) Imaging is highly efficient technique for detecting various cell defects such as micro cracks, finger interrupts and broken cells. Nevertheless, interpreting EL images for each cell can be quite challenging and time consuming because of the cell structures and excess pattern types. Therefore, it can be useful to inspect cell images automatically. In this study, PV cell images are classified using a public EL image dataset which contains 2624 individual cell images and deep convolutional neural networks. Transfer learning method was chosen due to small dimension of the dataset. AlexNet, GoogleNet, MobileNetv2 and SqueezeNet architectures were chosen for transfer learning and networks are trained on the GPU. Using transfer learning each training session completed under an hour and over 75% validation accuracy reached. Results indicated that convolutional neural networks and transfer learning could be easily used for PV cell defect detection.

## References

- [1] Evans, R. (2014). Interpreting module EL images for quality control. Proceedings of the 52nd Annual Conference, *Australian Solar Energy Society*.
- [2] Buerhop-Lutz, C., Deitsch, S., Maier, A., Gallwitz, F., and Brabec, C. J. (2018). A Benchmark for Visual Identification of Defective Solar Cells in Electroluminescence Imagery. 35th European PV Solar Energy Conference and Exhibition, 12871289.
- [3] Deitsch, S., Buerhop-Lutz, C., Maier, A., Gallwitz, F., and Riess, C. (2018). Segmentation of Photovoltaic Module Cells in Electroluminescence Images. Retrieved from <http://arxiv.org/abs/1806.06530>
- [4] Deitsch, S., Christlein, V., Berger, S., Buerhop-Lutz, C., Maier, A., Gallwitz, F., and Riess, C. (2019). Automatic classification of defective photovoltaic module cells in electroluminescence images. *Solar Energy*, 185(February), 455468.
- [5] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). 2012 AlexNet. *Advances In Neural Information Processing Systems*, 19.
- [6] Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., and Keutzer, K. (2017). SqueezeNet : AlexNet-Level Accuracy With 50 X Fewer Parameters and 0.5 Mb Model Size. *Iclr*, 113.
- [7] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... Rabinovich, A. (2015). Going deeper with convolutions. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 07-12 June, 19.
- [8] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L. C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 45104520.

# A Comparison of SVM Kernel Functions For Unbalanced Data

P. Akin<sup>1</sup>, Y. Terzi<sup>2</sup>

<sup>1</sup> *Ondokuz Mayıs University, Samsun, Turkey, pelin.akin@omu.edu.tr*

<sup>2</sup> *Ondokuz Mayıs University, Samsun, Turkey, yukselt@omu.edu.tr*

Support vector machines are supervised learning models with associated learning algorithms that analyze data used for categorization and regression analysis. The main task of the algorithm is to find the most correct line, or hyperplane, which divides the data into two classes. SVM is basically a linear classifier that classifies linearly separable data, but in general, the feature vectors might not be linearly separable. To overcome this issue, kernel trick is used. This article presents a comparative study of different kernel functions (linear, radial and sigmoid) for unbalanced data. The classification categories are not equally distributed. Three different re-sample methods are used for balancing the dataset. Rose sampling and smote are generated Synthetic balanced samples. The last using method is Oversampling, by adding more of the minority class. The myocardial infarction dataset which was taken from the Github were classified by 10 fold cross validation to increase the performance. Accuracy, AUC, sensitivity, specificity and F measure are used for comparing the methods. The analysis is carried by R software. As a conclusion, the results of performance metrics for the original data have increased by using Rose re-sampling methods for linear kernel functions.

## References

- [1] Lunardon, N., Menardi, G., and Torelli, N.. (2014). ROSE: a Package for Binary Imbalanced Learning. R Journal. 6. 79-89. 10.32614/RJ-2014-008.
- [2] Chawla, N.V., Bowyer, K.W., Hall, L.O., and Kegelmeyer, W.P. (2011). Smote: synthetic minority over-sampling technique. arXiv preprint arXiv:1106.1813, 2011.
- [3] Hussain, M., Wajid, S., El-Zaart, A., and Berbar, M. (2011). A Comparison of SVM Kernel Functions for Breast Cancer Detection. Proceedings - 2011 8th International Conference on Computer Graphics, Imaging and Visualization, CGIV 2011. 145-150. 10.1109/CGIV.2011.31.

# Comparison of Hot Deck and Regression Imputation in Multiple Imputation Methods for Missing Data Structures

S. Inan<sup>1</sup>, S. Isleyen<sup>2</sup>, S. Aydin<sup>3</sup>

<sup>1</sup> Van Yuzuncu Yil University, Van, Turkey, radres73@gmail.com

<sup>2</sup> Van Yuzuncu Yil University, Van, Turkey, sakirisleyen@yyu.edu.tr

<sup>3</sup> Saglik Bilimleri University, Turkey, saadettinaydin@gmail.com

Missing data is a common problem in almost all research areas based on human collected data. Lack of data usually means the situation that occurs when certain cases in a data set have missing data. Missing data handling methods depend on the number of samples, the amount of missing data, the package program used, and the method used. Among these methods, the hot deck (estimated mean match), regression safety and multiple implants are the most commonly used methods.

The aim of this study is to compare the effects of hot deck imputation and regression imputation in the multiple imputation on arithmetic mean and correlation coefficients.

The data of 537 patients who had cholesterol and glucose tests were used in our study. From this data, the glucose data were completely randomized to be completely randomized to 5%, 10% ,20% and 30% in the R Language environment. This missing data structure has been imputed to a hot-deck based multiple imputation and regression imputation based multiple imputation.

As a result of the analyzes, it was found that the use of hot deck in the multiple imputation method brought the parameters much closer to the actual value than the regression imputation in the multiple imputation method.

## References

- [1] Buuren, S.V., 2012. Flexible Imputation of Missing Data. Boca Raton, FL: Chapman & Hall/CRC.416.
- [2] Enders, C. K., 2010. Applied Missing Data Analysis. The Guilford Press, New York,401.
- [3] Little, R. J. A., Rubin, D. B., 2002. Statistical Analysis with Missing Data (2nd ed.), John Wiley & Sons, New York. 381.



# Using Regression Analysis Methods in Biostatistics An Applied Study on A Sample of Diabetic Patients

A.Y. Zebari<sup>1</sup>, S. Isleyen<sup>2</sup>, S. Inan<sup>3</sup>

<sup>1</sup> Van Yuzuncu Yil University, Van, Turkey, yehya85@yahoo.com

<sup>2</sup> Van Yuzuncu Yil University, Van, Turkey, sakirisleyen@yyu.edu.tr

<sup>3</sup> Van Yuzuncu Yil University, Van, Turkey, radres73@gmail.com

Biostatistics is one of the important approaches for decision makers in the health sciences by analyzing the indicators and finding the mathematical modeling and predictions. The topic of diabetes has been chosen to be applied in this study due to the importance of finding a cure of this disease and analyzing the indicators of increasing the rate of the incidence in the last years. The reasons of the incidence increasing rate and the types are investigated by the researcher, to demonstrate the effects of some variables such as weight and age on the diabetes incidence rate.

The study was conducted on a sample of 1385 patients with diabetes, randomly selected from the diabetics community in the Diabetics Center province of Duhok/ Iraq of a total 10,083 patients with diabetes, and applying the theories of linear regression on this data to create a mathematical equation helps to anticipate future incidence rates. The Statistical Package for Social Sciences (SPSS) was used in this study to obtain accurate results, reduce the time and effort. Basically, this study is an application of linear regression modeling on the diabetics cases. In addition, a regression function was constructed to predict diabetes incidence rate in the future. As a result, the exponential model was fitted the data under study.

## References

- [1] Ozdamar, K. (2001). SPSS ile Biyoistatistik, Kaan Kitabevi, Eskisehir.
- [2] Weisberg, S., (2005). Applied Linear Regression. John Wiley & Sons, Inc. Hoboken, New Jersey.
- [3] Chen, M., Ibrahim, J.G., Shao, Q., (2009). Maximum Likelihood Inference for the Cox Regression Model with Applications to Missing Covariates. *Journal of Multivariate Analysis* 100, 2018-2030.

# Estimating the Parameters of the Bivariate Mixed Model Using Robust Method with Ordinary Method

R.Y. Masiha<sup>1</sup>, S. Isleyen<sup>2</sup>, S. Inan<sup>3</sup>

<sup>1</sup> Van Yuzuncu Yil University, rogash.babiri@gmail.com

<sup>2</sup> Van Yuzuncu Yil University, sakirisleyen@yyu.edu.tr

<sup>3</sup> Van Yuzuncu Yil University, radres73@gmail.com

This paper is a condensed study that has been done by the researcher to compare among ordinary estimators. Mainly, the robust estimator and the maximum likelihood estimator methods used to estimate the bivariate mixed model called BARMA(1,1). Simulation experiments were conducted for different types of BARMA (1,1), by using large, moderate and small sample size, where some new results obtained. Its important to mention that MAPE used as a statistical criterion for comparison.

## References

- [1] Maronna, R.A., Martin, R.D., and Yohai, V.J. (2006). Robust Statistics. Theorey and Methods. Wiley.
- [2] Gill, P.S. (2000). A robust mixed linear model analysis for longitudinal data. *Statistics in Medicine*, 19, 975–987.
- [3] Hartley, H. O. and Rao, J. N. K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika*, 54, 93–108.
- [4] Yao, W., Wei, Y., and Yu, C. (2014). Robust Mixture Regression Using t-Distribution. *Computational Statistics and Data Analysis*, 71, 116–127.

# Combining Forecasts for Stock Keeping Units with Intermittent Demand Pattern: An Application on Spare Parts

G. Halil<sup>1</sup>, A. Kapucugil Ikiz<sup>2</sup>

<sup>1</sup> *Izmir University of Economics, Izmir, Turkey, gizem.halil@ieu.edu.tr*

<sup>2</sup> *Dokuz Eylul University, Izmir, Turkey, aysun.kapucugil@deu.edu.tr*

Managing the inventory effectively is crucial for success in all industries. Even though most of the SKUs in inventory are finished goods and work-in-process items with regular demand, some of them have irregular and intermittent demand (ID). ID pattern has many periods with zero demand with infrequent demand arrivals. When a demand occurs, the size is highly variable. Spare parts are one of the most important examples for the SKUs with ID patterns. These parts may be used for after-sales purposes or for preventive and corrective maintenance processes. Although they do not play an important role in the sales of a company, they may establish up to 60% of the total stock value. Thus, small improvements in spare parts inventory may result in significant cost savings. These SKUs have strategic importance for operations and their absence may affect the processes directly. Also, excess inventory may cause obsolescence. To avoid such problems, accurate forecasting is required for spare parts. However, forecasting is a challenging task for SKUs with such patterns since the irregularity of the demand causes traditional forecasting methods to perform poorly on ID.

Single exponential smoothing (SES) is one of the most widely used forecasting methods both for regular and ID. Yet, it is found to be biased for intermittent demand case resulting in high replenishment and excessive stock levels. To overcome such problems and generate both accurate and unbiased forecasts, several methods have been proposed in the literature. The first method, Crostons method (CR), estimates demand size and the interval between non-zero demands separately. The final estimation is calculated by averaging the separate estimates. After CR, many modifications on it such as Syntetos-Boylan Approximation (SBA), Teunter-Syntetos-Babai (TSB) and new methods such as Artificial Neural Networks and Bootstrapping are proposed. Besides the statistical methods, judgment can also be used for generating forecasts. These methods are generally based on expert opinions and can be used especially when the historical data is absent, there are significant changes in the environment, or the time series are highly variable. In practice, companies frequently use judgment in forecasting.

This study aims to suggest a model that combines statistical forecasts with judgmental forecasts in order to achieve a higher level of accuracy. First, the SKUs are categorized according to their demand patterns by using Syntetos-Boylan-Croston (SBC) categorization scheme. SBC suggests that for smooth demand pattern CR should be used while for erratic, intermittent and lumpy demand SBA should be used. From the statistical forecasting models which are proposed in literature especially good for ID patterns, SES, CR and SBA methods are applied and their accuracies are evaluated. Best performing method and parameters for each SKU are chosen and applied. With experts, judgmental forecasts are generated for these SKUs. These forecasts are combined with the forecasts of best performing statistical model by using weighted averages. Final conclusions are made based on the improvement in accuracy measures. The study demonstrates this proposed model on a real dataset having SKUs with irregular demand pattern.

# Sector-Wise Analysis of Cardinality Constraint Portfolio Optimization Problem: Selecting ISE-All Shares Based On Coefficient of Variation And Nonlinear Neural Network

I.Yaman<sup>1</sup>, T. Erbay Dalkilic<sup>2</sup>

<sup>1</sup> Giresun University, Giresun, Turkey, ilgimyaman@gmail.com

<sup>2</sup> Karadeniz Technical University, Trabzon, Turkey, tedalkilic@gmail.com

Standard portfolio optimization had proposed by Harry Markowitz in 1952 which is the benchmark problem of finance world. Many optimization methods have suggested to solve this problem to date. Actually, investors have a unique goal to get optimal portfolio which is maximizing expected return and minimizing the risk of portfolio. Portfolio optimization problem is a quadratic optimization problem. Moreover, cardinality constrained optimization problem is mixed-integer quadratic optimization problem that makes it NP-hard. Heuristic methods can solve NP-hard problems in reasonable time period. In this study, cardinality constraint portfolio optimization is solved by sector-wise. Cardinality constraint is solved by using coefficient of variation for different sectors. The main algorithm composed of mainly two parts. In the first part, coefficient of variation of each stock is calculated. Then, last quarter of ordered coefficient of variation of stocks are selected. In the second part, combinations of reduced stocks are covered to determine proportion of K stocks by using nonlinear neural network. The expected return, risks and Sharpe ratio of portfolio were calculated for different sectors. Indeed, this study reveals which sector provides better returns in the next three months. In order to analyze the proposed algorithm efficiency, ISE all shares data (Istanbul Stock Exchange all shares data was used between 10.05.2018-14.05.2019).

## References

- [1] Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1), 77-91.
- [2] Markowitz, H. (1959). Portfolio selection efficient diversification of investment. Newyork Wiley
- [3] Yan. Y. (2014). A new nonlinear neural network for solving QP problems. Springer international Publishing Switzerland, 347-357

# Associative Classification For Failure Prediction in Aluminium Wheel Molding: A Case Study

I. Kabasakal<sup>1</sup>

<sup>1</sup> *Ege University, Izmir, Turkey, inanc.kabasakal@gmail.com*

In the verge of the Industry 4.0 transformation, the digitalization is a primary concern for manufacturers. As a popular concept, Smart Manufacturing emphasizes the collection of data captured through various processes for further use. Analysis of process data with machine learning algorithms might provide useful information for root cause analysis and failure prediction. As a supervised learning approach, Associative Classification (AC) is often used for prediction tasks.

In our study, molding process data obtained from a global industry-leading wheel rim manufacturer was analyzed for fault prediction. The study might be described as a preliminary model developed for failure prediction on a real-time setting. The proposed approach for prediction consists of two steps. Firstly, the process data reported for subsequent steps were individually inspected with individuals control charts. The values out of control limits were marked as events that might correspond to potential causes of failures. With such assumption, an event dataset was organized where each event is linked to a part-code along with a class label that denotes the type of failure, if any.

Among the classification techniques in the data mining context, Associative Classification was adopted due to the descriptive nature of rules. Accordingly, when the model rises a fault prediction, the classifier rule also involves a root cause for the problem. The event data constructed after the process control was analyzed to extract rules with Association Rule Mining (ARM) technique. RuleGenerator, an implementation of the Apriori Algorithm has been employed for the discovery of rules and classifier rules. Among the 7295 rules discovered, 91 classifier rules were presented in the results. With a minimum lift value of  $\sim 22.3$  among the classifiers, the model might be promising despite the limited size of the dataset analyzed.

# Evaluation of Black Friday Hashtags in Turkey with Sentiment Analysis

I. Budak<sup>1</sup>, G. Kilic<sup>2</sup>, B.S. Kilic<sup>3</sup>

<sup>1</sup> Pamukkale University, Denizli, Turkey, [ibrahimbudak04@gmail.com](mailto:ibrahimbudak04@gmail.com)

<sup>2</sup> Pamukkale University, Denizli, Turkey, [gkilic@pau.edu.tr](mailto:gkilic@pau.edu.tr)

<sup>3</sup> Pamukkale University, Denizli, Turkey, [bskilic@pau.edu.tr](mailto:bskilic@pau.edu.tr)

Social media is a very popular communication tool to share peoples activities, ideas and feelings with others. Twitter is one of the most popular of these social media platforms. Firms and consumers are engaged in campaigns on social media platforms. These campaigns can be given as an example on Black Friday, which is accepted as the day the Christmas shopping begins in the United States. Firms in Turkey as well as in the whole world is trying to attract the attention of customers with social media hashtags Black Friday. In this study, tweets of various Black Friday hashtags were evaluated in 2018 in Turkey. Tweets were analyzed by Sentiment analysis. To evaluate the hashtags, the total number of tweets and the number of retweets were included in the analysis. With the help of the obtained results, the hashtags for Black Friday are listed.

# Experimenting with Some Data Mining Techniques to Establish Pediatric Reference Intervals for Clinical Laboratory Tests

D. Eraslan<sup>1</sup>, O. Yildirim<sup>2</sup>, D. Orbatu<sup>3</sup>, D. Sobay<sup>4</sup>, A.R. Sisman<sup>5</sup>, A. Aydin<sup>6</sup>, S. Sevinc<sup>7</sup>

<sup>1</sup> Dokuz Eylul University, Izmir, Turkey, denizeraslan148@gmail.com

<sup>2</sup> Dokuz Eylul University, Izmir, Turkey, oktay@labenko.com

<sup>3</sup> Saglik Bilimleri University, Izmir, Turkey, dilekorbatu@yahoo.com

<sup>4</sup> Dokuz Eylul University, Izmir, Turkey, sobaydoruk@gmail.com

<sup>5</sup> Dokuz Eylul University, Izmir, Turkey, aliriza.sisman@deu.edu.tr

<sup>6</sup> Dokuz Eylul University, Izmir, Turkey, adem.aydin@deu.edu.tr

<sup>7</sup> Dokuz Eylul University, Izmir, Turkey, suleysevinc@gmail.com

Reference interval studies are performed according to Clinical and Laboratory Standards Institute (CLSI) guidelines [1]. While these rules determined by CLSI can be applied for adults, it is difficult and inconvenient to apply these rules for pediatric patient groups. In order to resolve this need, this study has been performed to make reference interval establishment process easier with data mining techniques.

To apply data mining techniques, the data must go through certain stages. These stages include techniques such as filtering, which is for eliminating outliers, and analyzing the statistical distribution of the data according to criteria such as age, gender, diagnosis. In order to meet these requirements, a tool has been developed to allow specialists to easily load laboratory test data and to perform rapid data mining operations. 12 different biochemistry test data, which The Canadian Laboratory Initiative on Pediatric Reference Intervals (CALIPER) used for establishing pediatric reference intervals in their study, used for the experiments on developed tool and the results were compared with the reference intervals published by CALIPER [2]. In the experiments carried out using data mining techniques with machine learning algorithms, we found similar results with the reference intervals published by CALIPER, which were determined by using conventional methods.

By presenting this tool to specialists, they will be able to do reference interval study based on hospital or device and publish it after reviewing their clinical accuracy. With this developed tool, specialists can easily perform reference interval studies by loading laboratory test results, examine the distribution of age, gender, and diagnosis-related data and apply machine learning algorithms.

## References

- [1] Horowitz, G.L., Altaie, S., Boyd, J.C., Ceriotti, F., Garg, U., Horn, P., Pesce, A., Sine, H.E., and Zakowski, J. (2008). Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory; Approved Guideline Third Edition (Vol.28, No.30).
- [2] Colantonio, D.A., Kyriakopoulou, L., Chan, M.K., Daly, C.H., Brinc, D., Venner, A.A., Pasic, M.D., Armbruster, D., and Adeli, K. (2012). Closing the Gaps in Pediatric Laboratory Reference Intervals: A CALIPER Database of 40 Biochemical Markers in a Healthy and Multiethnic Population of Children, *Clinical Chemistry* 58:5

# Forecasting The Industrial 4.0 Data Via Anfis Approach

B. Ipek<sup>1</sup>, B. Berber<sup>2</sup>

<sup>1</sup> *Karadeniz Technical University, Trabzon. buketipek5@gmail.com*

<sup>2</sup> *Karadeniz Technical University, Trabzon. tberber@ktu.edu.tr*

In this study, information about fuzzy logic concept, historical development, purposes, advantages and disadvantages are given. Classical time series and fuzzy time series concepts are defined. The Adaptive-Network-Based Fuzzy Inference System (ANFIS) was used to model fuzzy systems and to estimate chaotic time series. Moreover, ANFIS method was applied to the data received from a factory using Industrial 4.0 technology with the help of sensor. The first degree Sugeno fuzzy inference system method has been used and suitable membership functions have been determined for the data sets. The Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) were calculated. ARIMA models were determined by Box-Jenkins method and RMSE and MAPE values were calculated. The RMSE and MAPE values obtained by using ANFIS method and ARIMA models were compared. As a result of the study, it was observed that the RMSE and MAPE obtained by first degree Sugeno fuzzy inference system method gave the best results on Industrial 4.0 data. In addition, the fuzzy inference system method produced the closest estimation values to actual values.



# Forecasting The Future Needs of Customers for New Products

A. Altinata<sup>1</sup>, A. Kapucugil İkiz<sup>2</sup>

<sup>1</sup> Dokuz Eylul University, Izmir, Turkey, anil.altinata@ogr.deu.edu.tr

<sup>2</sup> Dokuz Eylul University, Izmir, Turkey, aysun.kapucugil@deu.edu.tr

The way of doing business and being successful in it evolves within time. Status quo in business has changed greatly with the development of Internet, later WEB 2.0, e- marketing, social media and this change goes on with each new technological development that is integrated into daily life and gained acceptance from large customer groups. Globalization enabled logistic chains to cover almost all-around world. The cost to acquire information significantly decreased due to rising Internet usage. Product life cycles have become shorter. Innovative products started to get customers attention. This all resulted in competition getting very tough in markets and market leaders could only make differences due to small advantages they have. As a result of these developments, business environment demands high customer satisfaction rate for success in long term. A sector in which the technology is changing very fast, also needs faster adaptation or for a better chance, requires companies to be a leading innovator. So, companies need not only understand their customers needs but they should also anticipate the change in their needs in future. Thus, there is a need for combining a forecast system that is able to detect the changes in customer needs, interpolated from a QFD study or other analysis.

This study primarily focuses on finding a conceptual framework which can be used to predict the future customer requirements (CR) of the target market segment for new product development. The lack of historical data is a problem for forecasting when it comes to new products, so existing forecasting methods are carefully examined. QFD methodology will be used in the first step to understand the different categories and importance of customer requirements; using Kanos categories to modify weights and predict the changes of states for each CR. A modified version of Kano questionnaire will be conducted; that can be analyzed to find out transition probabilities between Kano categories. With the help of Markov Chain, the probability of states for each CR will be predicted to generate four data points. At this point Grey Theory Forecasting is a suitable tool, as it only requires four data points for a robust forecast. GM (1,1) methodology will be applied to the data to predict the change in weight of customer requirements.

The output of this model can be much valuable for management or decision makers, in the process of design for engineers; serving as a step to bullet proof the product/service for changes in customer perceptions and what they want by purchasing this product/ service in future. This will also help preventing unnecessary R&D efforts and budget spending on features which can become obsolete in future; while giving a chance to canalize the energy and time for getting a competitive advantage in an area which will be more valuable in the eyes of the customers.

## References

- [1] Akao, Y. (1988). QFD: Integrating Customer Requirements into Product Design. Portland: Productivity Press.
- [2] Wu, H. H. (2006, August). Applying grey model to prioritise technical measures in quality function deployment. *The International Journal of Advanced Manufacturing Technology*, 1278-1283. doi:10.1007/s00170-005-0016-y
- [3] Wu, H. H., and Shieh, J.I. (2008). Applying a markov chain model in quality function deployment. *Qual Quant*, 665678. doi:10.1007/s11135-007-9079-1
- [4] Wu, H. H., Liao, A., and Wang, P. C. (2004). Using grey theory in quality function deployment to analyse dynamic customer requirements. 12411247. doi:10.1007/s00170-003-1948-8

# Examination of Recommendation Systems and Usage Areas

V. Turk<sup>1</sup>, I. B. Aydilek<sup>2</sup>

<sup>1</sup> Harran University, Sanliurfa, Turkey, veyselturk484@gmail.com

<sup>2</sup> Harran University, Sanliurfa, Turkey, berkanaydilek@harran.edu.tr

Nowadays, huge amounts can be reached in data sizes due to the fact that internet is included in every aspect of life, rapid growth in internet technologies and increase in data storage areas [1]. The processing of large amounts of data has brought various problems. In the solution of these problems, solutions can be obtained by accessing valuable, interesting and unexplored information from raw data with data mining methods. The recommendation systems are one of the data mining subheadings.

Recommendation systems are the approaches that can provide the most suitable suggestions for the user based on the information of a user. Usually e-commerce, film, news, music, such as sites and applications are used [2].

In this study, a hybrid recommendation system was proposed using the MovieLens dataset. In conclusion, it was seen that the recommended hybrid recommendation system [3] was sufficiently successful.

## References

- [1] Bulut, H., and Milli, M. (2016). Isbirlikci filtreleme icin yeni tahminleme yontemleri. *Pamukkale Univ Muh Bilim Derg*, 22(2), 123-128.
- [2] Utku, A., and Akccayol, M. A. (2017). Ogrenabilen ve Adaptif Tavsiye Sistemleri Icin Karsilastirmali ve Kapsamli Bir Inceleme. *Erciyes Universitesi Fen Bilimleri Enstitusu Dergisi*, 33(3), 13-34.
- [3] Uluaymur, M. (2012). Hibrit Film Oneri Sistemi, Thesis (M.Sc.), Istanbul Technical University, Institute of Informatics.

# Application of the Weighted K Nearest Neighbor Algorithm for Diabetes

S. Koc<sup>1</sup>, L. Tomak<sup>2</sup>

<sup>1</sup> Ondokuz Mayıs University, Samsun, Turkey, kocsenem79@hotmail.com

<sup>2</sup> Ondokuz Mayıs University, Samsun, Turkey, lemant@omu.edu.tr

Machine Learning (ML) provides new methods, techniques, and tools that can help solving diagnostic and prognostic problems in a variety of medical area. K-Nearest Neighbor (KNN) is a basic classification algorithm of ML and is often used in the solution of classification problems. The aim of this paper is to use different kernel functions to weight the neighbors according to their distances and to compare the classification performance of KNN algorithms for diabetes data.

KNN algorithm is best-known, simple and a successful classification method for machine learning. KNN is a non-parametric, lazy learning algorithm. It uses a database in which the data points are separated into several classes to predict the classification of a new sample point. A query is labelled by a majority vote of its k-nearest neighbors in the training set. This paper uses kernel functions to weight the neighbors according to their distances. Kernel functions are rectangular, triangular, epanechnikov, biweight, triweight, cosine, inverse, gaussian, rank and optimal. Also two different distance were compared. The dataset is split into training (80%) and test (20%) sets as well in training (70%) and test (30%) sets. The analysis carried out using R software. The Pima Indian Diabetes dataset consist of 768 observations and 8 several medical predictor variables and one target variable, Diabetes. Predictor variables include the number of pregnancies the patient has had, their BMI, insulin level, age, blood pressure, glucose, skin thickness, diabetes pedigree functions.

Results show that when using gaussian kernel and  $k = 10$  the accuracy took a peak value of %76. If we decrease the k value the accuracy decreased too. This study determined that using different kernel functions to weight the neighbors and the higher value of k can improve the classification accuracy.

## References

- [1] Samworth, R. J. (2012). Optimal weighted nearest neighbour classifiers. *The Annals of Statistics*, 40(5), 2733-2763.
- [2] Hechenbichler, K., and Schliep, K. (2004). Weighted k-nearest-neighbor techniques and ordinal classification. *Sonderforschungsbereich 386*, p. 399.
- [3] Song Y., Huang J., Zhou D., Zha H. and Giles C. L.(2007). IKNN: Informative k- nearest neighbor pattern classification, Springer-Verlag Heidelberg, LNAI 4702, pp. 248 264.

# A Study of Data Mining Methods for Breast Cancer Prediction

Y. Gultepe<sup>1</sup>, T. Kartbaev<sup>2</sup>

<sup>1</sup> *Kastamonu University, Kastamonu, Turkey, yasemingultepe@kastamonu.edu.tr*

<sup>2</sup> *Almaty University of Power Engineering and Telecommunications, Almaty, Kazakhstan, t.kartbaev@aes.kz*

Obtaining and using biomedical data is increasing with developing information technology. At this point, different systems are needed to analyze the biomedical data quickly and accurately. Some of these systems help doctors and clinicians by analyzing and classifying data. In this paper, Breast Cancer Coimbra dataset [1] taken from Machine Learning Repository web site of the University of California Irvine (UCI ML) [2] was used. This dataset includes features that can be gathered in routine blood analysis. These features are Age (years), BMI (kg/m<sup>2</sup>), Glucose (mg/dL), Insulin (U/mL), HOMA, Leptin (ng/mL), Adiponectin (g/mL), Resistin (ng/mL) and MCP-1 (pg/dL). According to these input features, predicted data can be assigned to an unhealthy or healthy. These features were observed for 64 patients with breast cancer and 52 healthy people [3]. Filtering the information from this dataset through traditional inquiry methods and presenting this information in reports does not lead to the emergence of important secret rules hidden in the information. Therefore, it is inevitable to use the data mining algorithms used in the biomedical field for data discovery from datasets. WEKA is an open source data mining program with a functional graphical interface that combines machine learning algorithms [4]. WEKA includes various data preprocessing, classification, regression, clustering, association rules and visualization tools.

In this paper, data mining classification algorithms are examined and a prediction model is developed in Breast Cancer Coimbra dataset for early detection of breast cancer. Four algorithms were selected and applied to the data set after considering the popularity and similar studies in the literature while selecting the algorithms to be compared among the data mining algorithms in WEKA. These algorithms are J48, Multilayer Perceptron (MLP), K-Nearest Neighbor (K-NN) and Support Vector Machine (SVM) algorithm. Accuracy, Mean Absolute Error (MAE), Root Error Squares (RMSE) and Relative Absolute Error (RAE) values were considered when determining the most successful algorithm. As a result, overall performance rates of data mining classification algorithms were obtained as 76.92% with J48, 69.23% with Multilayer Perceptron, 69.23% with k- NN and 66.38% with SVM. The J48 algorithm, which has the highest accuracy rate in the diagnosis of the disease, plays an important role in the early diagnosis of an important disease such as breast cancer.

## References

- [1] Patricio, M., Pereira, J., Crisostom, J., Matafome, P., Seica, R., and Cramelo, F., (2018). Breast Cancer Coimbra Data Set, Web site: <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Coimbra>.
- [2] UCI Machine Learning Repository, Lung Cancer Data Set. (1992). Web site: <https://archive.ics.uci.edu/ml/index.php>.
- [3] Patricio, M., et al. (2018). Using Resistin, glucose, age and BMI to predict the presence of breast cancer. *BMC Cancer*, 18(1).
- [4] Witten, I. H., Frank, E., and Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques*. Elsevier, London.

# Pharmacy Students Intention towards Using Cloud Information Technologies in Knowledge Management

E. Ulutas Deniz<sup>1</sup>, M. Arslan<sup>2</sup>, N. Tarhan<sup>3</sup>, B. Sozen Sahne<sup>4</sup>, S. Yegenoglu<sup>5</sup>, S. Sar<sup>6</sup>

<sup>1</sup> Ataturk University, Turkey, elif.ulutas@atauni.edu.tr

<sup>2</sup> Van Yuzuncu Yil University, Van, Turkey, mirayarslan@yyu.edu.tr

<sup>3</sup> Izmir Katip Celebi University, Izmir, Turkey, nilay.tarhan@ikc.edu.tr

<sup>4</sup> Hacettepe University, Ankara, Turkey, bilge.sozen@hacettepe.edu.tr

<sup>5</sup> Hacettepe University, Ankara, Turkey, selen@hacettepe.edu.tr

<sup>6</sup> Ankara University, Ankara, Turkey, sevgisar98@gmail.com

Cloud computing is the basis of e-health applications and various learning processes developed today. It also offers a variety of opportunities to provide better quality health care, reduce costs, and transfer health information [1, 2]. Although there are several studies on the pharmacy students technology usage in various countries [3, 4], studies were very limited in Turkey. In this regard, the aim of this study is to determine the intention of the pharmacy students towards using cloud information technologies in knowledge management. In the study, a measurement tool developed by Arpaci [5] was applied to 4th class students of Ataturk, Hacettepe and Van Yuzuncu YI Universities Pharmacy Faculties (n=202). Confirmatory factor analysis (CFA) and two sample independent t-test were conducted via LISREL 8.80 and SPSS 22. Similarly to Arpaci [5] knowledge creation and discovery (KC), knowledge storage (KS), knowledge sharing (KSh), knowledge application (KA), innovativeness (I), training and education (T), perceived usefulness (PU), perceived ease of use (PEU), attitude (A), and continued use intentions (CI) factors were confirmed. No significant differences were found between gender and factors. However, using cloud technology differs in some factors as KC, I, PEU, PU, A and CI.

The results of the study shows that giving a training to students on this issue and supporting innovative approaches of students may be effective in increasing students intention for using cloud computing in knowledge management.

## References

- [1] Gao, F., Thiebes, S., Sunyaev, A. (2018). Rethinking The Meaning Of Cloud Computing For Health Care: A Taxonomic Perspective And Future Research Directions. *Journal of Medical Internet Research*, 20(7), e10041.
- [2] Bayn, G., Yesilaydn, G., Ozkan, O. (2016). Bulut bilisiminin saglik hizmetlerinde kullanimi. *Sosyal Bilimler Dergisi*, 48, 233-252.
- [3] Stolte, S.K., Richard, C. Rahman, A., Kidd, RS. (2011). Student pharmacists use and perceived impact of educational technologies. *Am J Pham Educ*, 75 (5), Article 92.
- [4] Siracuse, M.V. Sowell, J.G. (2008). Doctor of pharmacy students use of personal digital assistants. *Am J Pham Educ*, 72(1), Article 7.
- [5] Arpaci, I. (2017). Antecedents and consequences of cloud computing adoption in education to achieve knowledge management. *Computers in Human Behavior*, 70, 382- 390.

# Generative Adversarial Networks Based Data Augmentation For Phishing Detection

F. Uludag<sup>1</sup>, F. Kapar<sup>2</sup>, H.E. Celik<sup>3</sup>

<sup>1</sup> *Van Yuzuncu Yil University, Van, Turkey, fatihuludag@yyu.edu.tr*

<sup>2</sup> *Van Yuzuncu Yil University, Van, Turkey, firatkapar@yyu.edu.tr*

<sup>3</sup> *Van Yuzuncu Yil University, Van, Turkey, ecelik@yyu.edu.tr*

In machine learning studies, data augmentation is one of the methods used to increase the size of the training set and to avoid overfitting. In contrast to the classical approaches for data augmentation, we used different Generative Adversarial Networks (GAN) architectures in this study. Although GANs are often used for image data, we have used tabular data for this study.

Generative Adversarial Networks was first introduced by Goodfellow et al. as a powerful generative model [1]. The basic idea of GANs is to build a game between the two players called Generator and Discriminator. The Generative Adversarial Networks have been extended to a conditional model called cGAN, the Generator and the Discriminator are conditioned with extra information [2]. Wasserstein GAN (WGAN) architecture was obtained by using Wasserstein distance as GAN's loss function [3]. This approach was then extended to the conditional state called WCGAN [4].

In this study, we have created a dataset using 33 features of the websites. The XGBoost algorithm was able to distinguish phishing websites from legal websites with an accuracy of 89%. We have augmented our dataset with GAN, cGAN, WGAN and WCGAN architectures. We compared the success rates of XGBoost algorithm on augmented datasets with each architecture. According to the results, on the dataset augmented with CGAN architecture, XGBoost algorithm was able to classify with 98% accuracy rate.

## References

- [1] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (pp. 2672-2680).
- [2] Mirza, M., and Osindero, S. (2014). Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.
- [3] Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein gan. arXiv preprint arXiv:1701.07875.
- [4] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C. (2017). Improved training of wasserstein gans. In *Advances in Neural Information Processing Systems* (pp. 5767-5777).

# Comparison of the Spliced Regression Models

H. Konsuk Unlu<sup>1</sup>, A. Yigiter<sup>2</sup>, Y. Gencturk<sup>3</sup>

<sup>1</sup> Hacettepe University, Ankara, Turkey, hkonsuk@hacettepe.edu.tr

<sup>2</sup> Hacettepe University, Ankara, Turkey, yigiter@hacettepe.edu.tr

<sup>3</sup> Hacettepe University, Ankara, Turkey, yasemins@hacettepe.edu.tr

When data contains zeros and exhibits fat-tail behaviour, the well-known parametric models such as Exponential, Weibull, Burr, Pareto, Gamma, Lognormal and etc. might be inadequate or inappropriate. In this case, the composite (spliced) models made up by piecing together two (or more) weighted distributions at specified threshold(s) might provide a better fit. Composite (spliced) regression models can be used when dataset contains information about the underlying explanatory variables. There is not many work done on using spliced distributions with covariate information. One of the studies about spliced regression is carried out by Gan and Valdez [1], where Gamma-Pareto and Pareto-Type I Gumbel distributions are used to model Singapore automobile insurance data.

The aim of this study is to investigate the use of the Exponential-Pareto, Weibull-Pareto and Lognormal-Pareto regression models for this dataset and compare the results.

## References

- [1] Gan, G., and Valdez, E. A. (2018). Fat-tailed regression modeling with spliced distributions. *North American Actuarial Journal*, 22(4), 554-573.

# Using Convolutional Neural Networks for Handwritten Digit Recognition

Y. Gultepe<sup>1</sup>, A. E. Duru<sup>2</sup>

<sup>1</sup> *Kastamonu University, Kastamonu, Turkey, yasemingultepe@kastamonu.edu.tr*

<sup>2</sup> *Kastamonu University, Kastamonu, Turkey, aysemineduru15@gmail.com*

Nowadays, the internet contains many more images and videos; this urges development or search applications and algorithms that can investigate the semantic analysis of images and videos to provide the user with better search content and summarization [1]. As recently reported by different researchers, there has been great progress in image tagging, object detection, and stage classification in parallel with increasing processing power and improvements in graphics processors. This makes it possible to contribute to the solution of object detection and scene classification problems.

CNN (Convolutional Neural Networks) presents a model class that works to better understand the content in the image, thus resulting in better image segmentation and classification [2]. CNNs, which are made up of many layers, learn an attribute of each problem in the layer and these attributes are output to the next layer [3]. CNN algorithms are applied in many different fields such as natural language processing, biomedical, especially in the field of image and sound processing. In the paper, the classification success of the proposed method with CNN was measured using the MNIST (Modified National Institute of Standards and Technology Database) [4]. It consists of numbers written in handwritten and appropriately classified. MNIST is a commonly used handwritten numeric data set of 28X28 size, 60,000 training and 10,000 tests. Several different types of methods, from artificial neural networks to statistical methods, have been tested on this data set.

This paper was carried out using the Keras library in Python programming language. In this paper, the result obtained by CNN is a very low error rate (0.082) and an accuracy rate of 97%. The aim of the study, which is to minimize the loss value, was carried out successfully.

## References

- [1] Kou, F., Du, J., He, Y. and Ye, L. (2016). Social network search based on semantic analysis and learning, *CAAI Transactions on Intelligence Technology*.
- [2] Sharma, N., Jain, V. and Mishra, A. (2018). An analysis of convolutional neural networks for image classification, *Procedia Computer Science*, (132), 377-384.
- [3] Lee, S-J., Chen, T., Yu, L. and Lai, C-H. (2018). Image classification based on the boost convolutional neural network, *IEEE Access*, (6), 12755-12768.
- [4] LeCun, Y., Cortes, C. and Burges, C. (2017). MNIST handwritten digit database. Web site: <http://yann.lecun.com/exdb/mnist/index.html>.



# A statistical comparison of Zagreb and Angular Geometric Zagreb Indices

M. Cancan<sup>1</sup>, S. Ediz<sup>2</sup>, M. S. Aldemir<sup>3</sup>

<sup>1</sup> Van Yuzuncu Yil University, Van, Turkey, mcancan @yyu.edu.tr

<sup>2</sup> Van Yuzuncu Yil University, Van, Turkey, suleymanediz @yyu.edu.tr

<sup>3</sup> Van Yuzuncu Yil University, Van, Turkey, msaldemir @yyu.edu.tr

Topological indices have important role in theoretical chemistry for QSPR researches. Among the all topological indices the Zagreb indices [1] has been used more considerably than any other topological indices in chemical and mathematical literature. Most of the topological indices as in the Zagreb indices are based on the degrees of the vertices of a connected graph. Recently a novel degree concept has been defined in graph theory; geometric degree [2]. In this study, angular geometric Zagreb indices are defined by using geometric degree concept as parallel to their corresponding classical degree version. This novel angular Zagreb indices are compared with the Zagreb indices by correlation efficient of some physicochemical properties of octane isomers. Also the exact values of the angular geometric Zagreb indices for the well-known graph classes such as; paths, cycles, stars and complete graphs are given.

## References

- [1] Gutman, I., Trinajstić N. (1971) Graph theory and molecular orbitals. Total  $\Pi$ -electron energy of alternant hydrocarbons. *Chem. Phys. Lett.*, 17, 535-538.
- [2] Ediz S. (2019). A note on geometric graphs. *International Journal of Mathematics and Computer Science*, 14 (3), 631-634.

# **An Application of Type 2 Fuzzy Time Series Model**

H. Guney<sup>1</sup>

<sup>1</sup> *Ataturk University, Erzurum, Turkey, hilal.guney@atauni.edu.tr*

Classical and fuzzy methods which are used to model the time series are encountered in many areas of life. The tendency to model with fuzzy logic increases when the classical analyses are inadequate, unsatisfying and when the assumptions that classical model needs are not met. In the fuzzy time series analysis, which gives the opportunity to work with uncertainties, Type 2 models which include more information in calculations can be used in the literature. Furthermore, the applicability of a model to different data set is important. Therefore, in this study, Type 2 fuzzy time series model developed by Huarng and Yu (2005) was handled, and it was fitted in the gold price data in Turkey. The forecast results were obtained and the weaknesses of the Type 2 fuzzy time series model were also examined.

# Heteroscedastic and Heavy-tailed Regression with Mixtures of Skew Laplace Normal Distributions

F. Z. Dogru<sup>1</sup>, K. Yu<sup>2</sup>, O. Arslan<sup>3</sup>

<sup>1</sup> *Giresun University, Giresun, Turkey, fatma.dogru@giresun.edu.tr*

<sup>2</sup> *Brunel University, London, UK, keming.yu@brunel.ac.uk*

<sup>3</sup> *Ankara University, Ankara, Turkey, oarslan@ankara.edu.tr*

In regression analysis, joint modelling skewness and heterogeneity is a challenging problem. In this study, we consider the skew Laplace normal (SLN) distribution studied by [1] which is a heavy tailed and skew distribution, and we propose a joint modelling of location, scale, and skewness parameters of mixtures of SLN distributions to model heteroscedastic skew-heavy tailed data set coming from a heterogeneous population. The maximum likelihood (ML) estimators of all parameters are obtained via the expectation- maximization (EM) algorithm (see [2]), and also asymptotic properties of estimators are derived. Numerical analyses via a simulation study and a real data example are conducted to show the performance of the proposed model.

## References

- [1] Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39, 1-38.
- [2] Gómez, H.W., Venegas, O., and Bolfarine, H. (2007). Skew-symmetric distributions generated by the distribution function of the normal distribution. *Environmetrics*, 18, 395- 407.

# Modeling of Exchange Rate Volatility in Turkey: An Application with Asymmetric GARCH Models

A. Genc<sup>1</sup>

<sup>1</sup> Pamukkale University, Denizli, Turkey, [agencpau@gmail.com](mailto:agencpau@gmail.com)

Volatility refers to changes in the price of a financial asset. Volatility measures the magnitude, degree and permanence of the changes in prices. The use of conventional econometric models to measure volatility in financial time series has been to cause some shortcomings in terms of reliability due to some features of the financial series and the first autoregressive conditional variance model of ARCH (Autoregressive Conditional Heteroskedastic) was developed by R. F. Engle [1]. In this study using financial series, autoregressive conditional variance models were preferred.

Asymmetric exponential volatility models have been developed, which take into account the features of the financial series and examine the leverage effect. Asymmetric exponential volatility models make volatility modeling taking into account that the news of bad news is not symmetrical with good news, and bad news causes more volatility than good news (leverage effect); see [2]. In this way, exponential GARCH models are superior to those volatility models that are only interested in the impact of shock. Therefore, in this study in order to test whether the exchange rate volatility Turkey is the modeling of the asymmetry effect, showing the asymmetric exponential GARCH volatility was used. Thus, the good or bad news on the information shock within the foreign exchange market in Turkey will be determined which one is more dominant.

Within the framework of EGARCH, GJRGARCH and APGARCH models, the exchange rate volatility was modeled with the daily data of January 2001-December 2018 period. USD / TL and EUR / TL exchange rates were used in the study, and the best predictive performance for both exchange rates was the APARCH (Asymmetric Power ARCH) model developed by Ding, Granger and Engle [3]. The leverage parameter and power parameter in this model provide information on the asymmetric effect and persistence of the shock in the information shocks on the market. According to the results obtained from this model, the existence of asymmetric effect for both dollar and euro was determined and the leverage effect was found to be negative. In this case, the effect of good news in the dollar and euro volatility in market prices in Turkey was found to be more than bad news. In addition, the power parameter of the model showed that the information shock reaching the market for both exchange rates was persistent.

## References

- [1] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 50 (4), 987-1007.
- [2] Tsay, R. S. (2005). *Analysis of Financial Time Series*, 3rd Ed., John Wiley & Sons, New Jersey.
- [3] Ding, Z., Granger, C. W., and Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1(1), 83-106.

# Comparison of Catalase and Superoxide Dismutase Enzyme Activities in Strawberry Fruit

T. Gur<sup>1</sup>, F. Karahan<sup>2</sup>, H. Demir<sup>3</sup>, C. Demir<sup>4</sup>

<sup>1</sup> Van Yuzuncu Yil University, Van, Turkey, tugbagur@yyu.edu.tr

<sup>2</sup> Van Yuzuncu Yil University, Van, Turkey, halitdemir@yyu.edu.tr

<sup>3</sup> Van Yuzuncu Yil University, Van, Turkey, fatihkarahan2727@gmail.com

<sup>4</sup> Van Yuzuncu Yil University, Van, Turkey, canandemir@yyu.edu.tr

Strawberry, a powerful antioxidant, strengthens the immune system. Strawberry contains plenty of iron and phosphorus. It is also rich in terms of B, C and K vitamins. Reduce cholesterol and also prevents vascular occlusion. It is protective against cancer. At the same time, harmful substances remove from the body. Reduces the blood pressure and decrease stress. It is good to rheumatism and liver disturbances. The aim of this study was to determine some antioxidant enzyme activities in strawberry fruit obtained from various regions.

CAT and SOD antioxidant enzyme activities were determined by preparing the extract of strawberry fruit. Strawberry is a grassy plant with a creeping body. It is also a very fragrant, a cone shaped fruit. There are many species and types. For this purpose, it is aimed to determine some enzyme activities which are thought to be present in strawberry with these beneficial properties of strawberry fruit. In this study, antioxidant enzyme activities were determined by spectrophotometric method. The strawberry pieces that we have previously shredded were centrifuged at 8000 rpm for 5 minutes. The liquid at the top of the centrifuge tube was then receipted. The absorbance measurements were then performed with a spectrophotometer set at 240 nm.

In this study, extracts of strawberries obtained from various regions were prepared. Then, measurements were made with spectrophotometer. The obtained data was evaluated. Descriptive statistics for the features discussed; Mean is expressed as Standard Deviation. T-Test was used in cases where normal distribution condition was provided and Mann Whitney U test statistic was used in cases where normal distribution condition was not provided. Statistical significance level was taken as 5% in the calculations and SPSS statistical package program was used for the calculations.

In this study, antioxidant enzyme activities in the strawberry fruit were determined. Thus, the strawberry fruit was seen that how strong antioxidant fruit.

# Investigation of Some Antioxidant Enzyme Activities in Cherry Fruit Obtained From Various Regions

T. Gur<sup>1</sup>, F. Karahan<sup>2</sup>, H. Demir<sup>3</sup>, C. Demir<sup>4</sup>

<sup>1</sup> Van Yuzuncu Yil University, Van, Turkey, tugbagur@yyu.edu.tr

<sup>2</sup> Van Yuzuncu Yil University, Van, Turkey, halitdemir@yyu.edu.tr

<sup>3</sup> Van Yuzuncu Yil University, Van, Turkey, fatihkarahan2727@gmail.com

<sup>4</sup> Van Yuzuncu Yil University, Van, Turkey, canandemir@yyu.edu.tr

Cherry is a fruit rich in vitamin C. They do not contain fat and cholesterol. It contains essential minerals such as fiber, vitamin A, iron, calcium, protein as well as abundant potassium. Red cherries also contain melatonin, which helps combat harmful toxins. Due to its antioxidant properties, it has many benefits such as prevention of some types of cancer, reduction of inflammation, prevention of gout and removal of muscle pain. In our study, it was aimed to determine the antioxidant enzyme activities in cherry fruit. The cherry fruit extract was prepared and some antioxidant activities were determined.

Cherry, (*prunus avium*) is a fruit belonging to the family of rosaceae. Its homeland is asia minor. Many varieties are grown in Turkey. There are more than a hundred culture forms grown in north america with temperate regions of europe and asia. Its body is in the form of a flat-shell tree. For this purpose, it is aimed to determine some enzyme activities which are thought to be found in cherry plants. In this study, antioxidant enzyme activities were determined by spectrophotometric method. The cherry pieces that we have previously shredded were centrifuged at 8000 rpm for 5 minutes. The liquid at the top of the centrifuge tube was then receipted. The absorbance measurements were then performed with a spectrophotometer set at 240 nm.

In this study, extracts of cherry from various regions were prepared. Then, measurements were made with spectrophotometer. The obtained data was evaluated. Descriptive statistics for the features discussed; Mean, Standard deviation, Minimum and Maximum values are expressed. One way ANOVA was used for normal distribution conditions and Kruskal Wallis test statistic was used for cases where normal distribution condition was not provided. Statistical significance level was taken as 5% in the calculations and SPSS statistical package program was used for the calculations.

In this study, antioxidant enzyme activities in the cherry fruit were determined. Thus, the cherry fruit was seen that how strong antioxidant fruit.

# Prediction of Gastric Cancer Stages with Convolutional Neural Networks

F. Uludag<sup>1</sup>, S. Celik<sup>2</sup>, H. E. Celik<sup>3</sup>, A. Sohail<sup>4</sup>, U. H. Iliklerden<sup>5</sup>

<sup>1</sup> Van Yuzuncu Yil University, Van, Turkey, fatihuludag@yyu.edu.tr

<sup>2</sup> Van Yuzuncu Yil University, Van, Turkey, scelik@yyu.edu.tr

<sup>3</sup> Van Yuzuncu Yil University, Van, Turkey, ecelik@yyu.edu.tr

<sup>4</sup> COMSATS University, Islamabad, Pakistan, ayeshasohail81@gmail.com

<sup>5</sup> Van Yuzuncu Yil University, Van, Turkey, umitliklerden@yyu.edu.tr

Gastric cancer (GC) is one of the most common causes of cancer-related death worldwide and Turkey. Surgical resection is the only cure available and is dependent on the GC stage at presentation, which incorporates depth of tumor invasion, extent of lymph node and distant metastases [1]. It is important to determine the staging correctly for a good treatment. Multidetector computed tomography (MDCT) is the most commonly used technique for the staging of GC as it provides higher resolution scans with thin collimation that allows excellent multiplanar reconstructions [1].

Deep learning methods are representation-learning methods with multiple levels of representation, obtained by composing simple but nonlinear modules that each transform the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level [3]. Convolutional Neural Networks (CNN) are designed especially recognize visual patterns from pixel images [2]. Deep learning has been frequently used in the literature to determine the presence of cancer, but its use is limited in staging studies.

In this study, we applied deep convolutional networks of 45 gastric cancer patients CT images. We have implemented LeNet, AlexNet, GoogLeNet and ResNet architectures and compared their success rates.

## References

- [1] Hallinan, J. T. P. D., and Venkatesh, S. K. (2013). Gastric carcinoma: imaging diagnosis, staging and assessment of treatment response. *Cancer imaging*, 13(2), 212.
- [2] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.
- [3] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436.
- [4] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.
- [5] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., and Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).
- [6] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

# Irony Detection in Turkish Tweets

A. Karabas<sup>1</sup>, B. Diri<sup>2</sup>

<sup>1</sup> *Yildiz Technical University, Istanbul, Turkey, ahmetkarabas93@gmail.com*

<sup>2</sup> *Yildiz Technical University, Istanbul, Turkey, diri@yildiz.edu.tr*

The number of people who express themselves in social media is increasing day by day. Twitter, which is an important place for people to share their thoughts on a topic has 500 million tweets sent per day [1]. In such a large data manual classification is a very challenging task. Therefore, using autonomous systems, software etc. for classification is of great importance.

The irony is the expression in which the meaning of the opposite is said. The action that is said or done is intended to draw the contradiction or action under the serious image to the point of contradiction [2]. In recent years, after successful results in the sentiment analysis with tweets, studies have been conducted on irony detection as well. However, this issue is more challenging than sentiment analysis. While it is easier to determine the irony in face-to-face conversation, it can be difficult for normal people to understand it in written communication. The character limit on Twitter, typing and punctuation errors of some people hamper the direct implementation of classification methods. For this reason, it is necessary to apply the preprocessing steps in the tweets. Afterwards, it is aimed to determine the irony by extracting features.

In this study, after the preprocessing of the data by correcting the words, writing errors etc., the machine learning and deep learning algorithms were applied with different parameters and the success of the results were examined and compared. As a result the most successful models/algorithms were Pre-trained Word Embeddings and Hyperparameter Optimization on Deep Neural Networks yielding an F-score of .878 and .873, respectively.

## References

- [1] Cooper, P. (2019). 28 Twitter Statistics All Marketers Need to Know in 2019, <https://blog.hootsuite.com/twitter-statistics/>, accessed on 2014-04-02.
- [2] (2018). Ironi, <https://tr.wikipedia.org/wiki/%C4%B0roni>, accessed on 2014-04-02.



# A Machine Learning Sepsis Diagnosis Model for Intensive Care Units

G. Silahtaroglu<sup>1</sup>, Z. Canbolat<sup>2</sup>, N. Yilmazturk<sup>3</sup>

<sup>1</sup> *Medipol University, Istanbul, Turkey, gsilahtaroglu@medipol.edu.tr*

<sup>2</sup> *Medipol University, Istanbul, Turkey, zncanbolat@medipol.edu.tr*

<sup>3</sup> *Pusula, Istanbul, Turkey, nevin.yilmazturk@pusulayazilim.com.tr*

Sepsis infection, which is one of the most important causes of deaths in intensive care units in health industry, is seen as a serious global health crisis. Sepsis, which affects between 27 and 30 million people every year, kills 7 to 9 million people. If sepsis infection cannot be diagnosed early enough, it may cause septic shock, multiple organ failure and finally the story ends with death. According to the figures announced by the World Health Organization, every 6 seconds a person dies of sepsis infection. Sepsis infection causes high mortality or most survivors experience long-term morbidity. Since the incidence of sepsis is high, this remains one of the leading causes of death globally. Therefore, sepsis is seen as an important public health problem with significant economic consequences. Prediction of sepsis and an early sepsis diagnosis may lead rapid treatment and provide better results. However, the diagnosis of sepsis is hard and needs experienced caregivers. The predictive accuracy of existing instruments is poor and the diagnosis is based on expensive and time-consuming laboratory results.

In this study, we have used MIMIC III database which is provided by the Beth Israel Deaconess Medical Center hospital for researches on Intensive Care Unit health issues. The database consists of 57,000 patients records which includes information such as demographics, vital sign measurements made at the bedside laboratory test results, procedures, medications, caregiver notes, imaging reports, and mortality (both in and out of hospital). Although the main objective of the study is to predict sepsis, this work solely focuses on data preparation phase because it plays a vital role in sepsis machine learning process. For this purpose, clinicians expertise and machine learning depth and necessities have been taken into account. An artificial intelligence unit needs to be designed to collect and mould all data in a desired format. This format is dynamic and cannot be predicted beforehand. It should be designed on the route to machine learning and must be re-designable. The AI module checks data types, timestamp sequence within a given environment and makes decisions to organise dataset in the best possible way so that machine learning algorithms can learn. For this purpose, a new model or algorithm has been developed.

# **Risk Classification with Artificial Neural Networks Models in Motor Third Party Liability**

K. Yildirak<sup>1</sup>, S. Kestel<sup>2</sup>, I. Gur<sup>3</sup>

<sup>1</sup> *Hacettepe University, Ankara, Turkey, kasirga@hacettepe.edu.tr*

<sup>2</sup> *Middle East Technical University, Ankara, Turkey, skestel@metu.edu.tr*

<sup>3</sup> *Hacettepe University, Ankara, Turkey, ismailgur@hacettepe.edu.tr*

One of the most fundamental requirements in today's insurance sector is the determination of fair premium for the insured. In order for this purpose to be fulfilled, the correct risk classification is required for each insured in the portfolio. By the realization of correct risk classification, the insured can continue to be provided insurance services with more suitable pricing, while the insurance companies will have the opportunity to provide the correct person with the correct insurance and carry out financially sustainable insurance transactions. Risk classification as a common interest on both sides will ensure the existence and sustainability of the market.

In this study, risk classification is made by using the claim information about insured within the scope of motor third party liability. As a classification model, ANN (Artificial Neural Network) models used on various attributes of the tools mentioned in the insurance policies are utilized. The data provided by the Insurance Information and Monitoring Center (SBM), consists of basic policy based information about insured vehicles, from 2006 to 2010. It has been shown that ANN models respond to our problem significantly, model has reached high accuracy in classification process for both training and testing data.

# Inference in Step-Stress Partially Accelerated Life Testing for Inverse Weibull Distribution under Type-I Censoring

F.G. Akgul<sup>1</sup>, K. Yu<sup>2</sup>, B. Senoglu<sup>3</sup>

<sup>1</sup> Artvin Coruh University, Artvin, Turkey, [fm.gul.fuz@artvin.edu.tr](mailto:fm.gul.fuz@artvin.edu.tr)

<sup>2</sup> Brunel University, London, UK, [keming.yu@brunel.ac.uk](mailto:keming.yu@brunel.ac.uk)

<sup>3</sup> Ankara University, Ankara, Turkey, [senoglu@science.ankara.edu.tr](mailto:senoglu@science.ankara.edu.tr)

With technological and industrial improvements, it is difficult to obtain the information about the lifetime of the high reliability products under normal-use conditions. Therefore, to get the information about these products, a sample of them is subjected to stress. This kind of tests are called the accelerated life tests (ALT). ALT provides enough failure data in a short period of time [1]. The basic assumption of ALT is that the mathematical model relating to the lifetime of the unit and the stress are known. Nevertheless, sometimes the life-stress relations are not always known and ALT is not available [2]. In this case, partially accelerated life test (PALT) which has the products tested under normal conditions until the prefixed time and then the surviving products are changed to put in accelerated stress conditions is often used.

The stress can generally be applied in various ways, the commonly used methods are step-stress and constant-stress. In step-stress PALT (SSPALT), firstly the test item is run at normal condition. If it does not fail for a specified time, then it is run at accelerated condition until the test terminates. But in constant-stress PALT (CSPALT), each unit is run at constant stress level until the test terminates.

This study deals with the classical and Bayesian estimation of SSPALT model under Type-I censoring when the lifetime distribution is Inverse Weibull (IW). In the context of classical estimation, maximum likelihood (ML) estimates of the distribution parameters and the acceleration factor are obtained. In addition, approximate confidence intervals (ACI) of the parameters are constructed based on the asymptotic distribution of the ML estimators. Under Bayesian inference, while the approximation posterior expectation methods by Lindley and Tierney-Kadane could provide point estimates of the distribution parameters and the acceleration factors under square error loss (SEL) function, Gibbs sampling method is also used to construct credible intervals of these parameters together with their point estimates. Monte-Carlo simulations are performed to compare the performances of the different estimation methods.

## References

- [1] Wang, B.X., Yu, K., Sheng, Z. (2014). New inference for constant-stress accelerated life tests with Weibull distribution and progressively type-II censoring. *IEEE Transactions on Reliability*, 63(4):807-815.
- [2] Zheng, D., Fang, X. (2018). Exact confidence limits for the acceleration factor under constant-stress partially accelerated life tests with type-I censoring. *IEEE Transactions on Reliability*, 67(1):92-104.

# Modeling Dynamically Behavior of Users in Social Networks using Petri Nets

A. Karadogan<sup>1</sup>, A. Karci<sup>2</sup>

<sup>1</sup> *Inonu University, Malatya, Turkey, ahmet.karadogan@inonu.edu.tr*

<sup>2</sup> *Inonu University, Malatya, Turkey, ali.karci@inonu.edu.tr*

Online social networks have been developed rapidly in recent years. People can share information with each other using social networks, due to this case, social network analysis has recently been used to analyze the relationships between users. This paper aims to model and analyze the dynamic behaviors of users in social networks. Petri nets, a mathematical and graphical modeling tool that are suitable for describing the dynamic properties of a system, are used to model the behaviors of users in social networks. In this case, social networks are also modelled by using Petri nets at the same time. After modeling the networks, some mathematical information of model such as incidence matrix obtained from the model are used to analyze the system by linear algebraic system. Results show that social networks can be modeled and analyzed mathematically using dynamical properties of Petri nets.

## References

- [1] Murata, T. (1989). Petri Nets: Properties, Analysis and Applications, Proc. IEEE, 77, 4, pp. 541580.
- [2] Pinna, A., Tonelli, R., Orrú, M. and Marchesi, M. (2018). A petri nets model for blockchain analysis, *Comput. J.*, 61, 9, 13741388.
- [3] Celaya, J.R., Desrochers, A.A. and Graves, R.J. (2009). Modeling and analysis of multi- agent systems using petri nets. *J. Comput.*, 4, 10, 981996.
- [4] L. Li, W. Zeng, Z. Hong, and L. Zhou (2016). Stochastic Petri Net-based performance evaluation of hybrid traffic for social networks system. *Neurocomputing*, 204, 3–7.
- [5] Wang, T., He, J. and Wang, X. (2018). An information spreading model based on online social networks. *Phys. A Stat. Mech. its Appl.*, 490, 488496.

# Qualitative Data: Advantages, How to Collect and Present Based on Three Examples on Health Issues

S. Yegenoglu<sup>1</sup>

<sup>1</sup> Hacettepe University, Ankara, Turkiye, [selen@hacettepe.edu.tr](mailto:selen@hacettepe.edu.tr)

In this presentation first the advantages of qualitative data will be provided. Then the collection techniques will be presented and important points while collecting the qualitative data will be underlined. And then three exploratory studies will be presented one another. In the first study the data collected on the nature of substance will be shown. And in the second study the data of the university students enrolled to a pharmacy faculty will be revealed and the “reasons for smoking” will be discovered. Finally there will be another qualitative study findings collected from community pharmacists concerning their “interaction with geriatric patients”.

In the end of the presentation contributions of the audience on qualitative techniques will be requested.

## References

- [1] [www.course.ccs.neu.edu](http://www.course.ccs.neu.edu): Qualitative Research Methods: A Data Collectors Field Guide, Module 1: Qualitative Research Methods Overview, Family Health International, Accessed: 11.03.2019.
- [2] [www.blogsocialcops.com](http://www.blogsocialcops.com): The 3 Qualitative Research Methods You Should Know, 26 March, 2018, Accessed: 18.03.2019
- [3] [www.snapsurveys.com](http://www.snapsurveys.com): Whats the difference between qualitative and quantitative research?, 16 September 2011, Accessed: 20.03.2019
- [4] Aksit, B.T., Onaran, S. Istanbulda Degisik Gruplarin Madde Kullanimina Iliskin Yaklasimlari, (in Farklilikla Yasamak, Edited by Nuray Karanci), 87-111, Turk Psikologlar Dernegi Yayinlari, 1. Basim, Aralik 1997, Ankara.
- [5] Yegenoglu, S., Aslan, D., Erdener, S.E., Acar, A., Bilir, N. (2006) What Is Behind Smoking Among Pharmacy Students: A Quantitative and Qualitative Study From Turkey. *Subst Use Misuse* 41(3): 405-414.
- [6] Yegenoglu, S., Baydar, T. (2011). Information and Observations of Community Pharmacists on Geriatric Patients: A Qualitative Study in Ankara City. *Turkish Journal of Geriatrics*, 14(4): 344-351.

# Research Methods In Social Pharmacy Studies

N. Sencan<sup>1</sup>

<sup>1</sup> *Acbadem University, Istanbul, Turkey, nazli.sencan@acibadem.edu.tr*

Social Pharmacy is a multi-cultural, inter-disciplinary field that focuses on utilisation of medicine by both consumers and healthcare professionals. Social Pharmacy borrows research methods and theories from diverse disciplines of different perspectives.

Teaching, learning and sharing experiences on social research techniques means knowing social, humanistic and natural sciences research techniques. The aim is to understand;

- the use of medicines, the patient/consumer, profession, society, and drug industry,
- the costs and access to medicines has several political, economic and social factors,
- the pharmacists and pharmacies struggling with high speed transitioni,
- expectation dissimilarities of patients and consumers.

Pharmacists wherever they work (community pharmacies, hospitals, authorities, the pharmaceutical industry and etc.) have to use the social pharmacy acquisition in their daily bussiness. To find solutions for related situations, the problems need to be highlighten as the results and implemantations of social pharmacy studies has a vital importance on public health.

The research methods of social pharmacy can be summarised as;

- a qualitative narrative studies/observations; This is a kind of anthropological field work (repeated participant observations and interviews with a few participants).
- b semi-structured interviews; One to one, face to face formal interviews. (The aim is to identify medicine-related characteristics, such as behaviours, knowledge, perceptions and attitudes, of different groups, such as patients and pharmacists.)
- c focus-group interviews; Often employ the same methods of interviews as those at the group level (6-10 participants)(the study of organizational features of pharmacies, hospitals, patients, technicians etc.).
- d surveys; Mostly web based or face to face, structured and self filled formal questions. e. mix methods/triangulation; all the above methods could be used with variations.

According to the needs of problematic, the research method is preferred. The pros/cons and experinces of the techniques are discussed in the presantation.

## References

- [1] Mount J.K. Contributions of the social sciences. In: Wertheimer A.I., Smith M.C., editors. *Pharmacy Practice Social and Behavioural Aspects*. 3rd ed. Williams & Wilkins; Baltimore, MD, USA: 1989. pp. 115.
- [2] Nouri A., Hassali M.A.A. (2018). Scope and Challenges of Social Pharmacy in Training of Pharmacy Students, *Int J Drug Disc*, 2(2), 1–12.

# The Analysis Of Web Server Logs With Web Mining Methods

S. Ziyanak<sup>1</sup>, H.E. Celik<sup>2</sup>

<sup>1</sup> Van Yuzuncu Yil University, Van, Turkey, [szivanak@yyu.edu.tr](mailto:szivanak@yyu.edu.tr)

<sup>2</sup> Van Yuzuncu Yil University, Van, Turkey, [ecelik@yyu.edu.tr](mailto:ecelik@yyu.edu.tr)

While surfing on a website, all the actions taken by the visitors are actually recorded on web server in the error and access files. In process of time, the files recording all the access and error entries are deleted by the people who are the management of web servers by reason of their reaching high sizes. However; when these inputs deleted without checking out are checked out with correct methods, it not only gets statistical inferences but also it is possible to make improvements on website for its efficient usage, taking security measures, identifying cyber-attacks to the website. In this work, the conclusions are drawn to check out access logs on web servers with web mining methods, for more efficient usage of web sites and its formatting.

When the web servers are checked out with web mining methods, it is crucial to check out all the access logs. As the web access logs becomes big data in time, Apache Hadoop System which checks out big data with a different file system and methods is used in this paper. Apache Hive query language designed for Big Data is used so as to check out the data more easily and efficiently on Apache Hadoop. Within the context of thesis study, the opportunities are provided to draw significant and useful conclusions from access logs before they are deleted.

## References

- [1] Goel, N., 2013. Analyzing users behavior from web access logs using automated log analyzer tool. *International Journal of Computer Applications*, 62(2):29- 33.
- [2] Grace, L. K. J., 2011. Analyss of web logs and web user in web mining. *International Journal of Network Security & Its Applications (IJNSA)*, 3(1):99-110.
- [3] Makhecha, H., 2016. Clickstream analysis using hadoop. *International Journal of Computer Trends and Technology (IJCTT)*, 34(2):89-92.
- [4] Mayer, V., Cukier, S., Cukier, K., 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, New York.
- [5] Mehta, J., 2015. Trend analysis based on access pattern over web logs using hadoop. *International Journal of Computer Applications*, 115(8):34-37.
- [6] Mobasher, B., Cooley, R., Srivastava, J., 2000. Automatic personalization based on web usage mining, *Communications of the ACM*, 43(8), 142-151.

# An Extension of the Maxwell Distribution: Properties and Application

N. Erdogan<sup>1</sup>, K. Bagci<sup>2</sup>, T. Arslan<sup>3</sup>

<sup>1</sup> Van Yuzuncu Yil University, Van, Turkey, [nerdogan@yyu.edu.tr](mailto:nerdogan@yyu.edu.tr)

<sup>2</sup> Van Yuzuncu Yil University, Van, Turkey, [kubrabadgi@yyu.edu.tr](mailto:kubrabadgi@yyu.edu.tr)

<sup>3</sup> Van Yuzuncu Yil University, Van, Turkey, [mstalhaarlsan@yyu.edu.tr](mailto:mstalhaarlsan@yyu.edu.tr)

In this study, we derive a new life-time distribution called as Alpha Power Maxwell (APM) by using the Alpha Power Transformation method proposed by Mahdavi and Kundu [1]. Some statistical properties of it are provided and discussed, as well. The maximum likelihood (ML) method is utilized to obtain the estimates of parameters of the APM distribution. However, the ML estimates of the unknown parameters can not be obtained explicitly. Therefore, iterative techniques should be used, e.g. Newton-Raphson. At the end of the study, the APM distribution is used to model the actual data set and the results are presented.

## References

- [1] Mahdavi, A. and Kundu, D. (2017). A new method for generating distributions with an application to exponential distribution, *Commun. Stat. – Theory Methods* 46(13), pp. 6543–6557,



# The Feasibility of Near Infrared Spectroscopy for Classification of Pine Species

F. D. Tuncer<sup>1</sup>, E. Akdeniz<sup>2</sup>, A. D. Dogu<sup>3</sup>

<sup>1</sup> *University of Istanbul-Cerrahpasa, Istanbul, Turkey, digdem.kangaloglu@istanbul.edu.tr*

<sup>2</sup> *University of Marmara, Istanbul, Turkey, esra.akdeniz@marmara.edu.tr*

<sup>3</sup> *University of Istanbul-Cerrahpasa, Istanbul, Turkey,*

In this study, utilization of near infrared spectroscopy (NIRS) for classification of two *Pinus* samples was investigated. 195 and 181 spectra of *Pinus sylvestris* and *Pinus nigra* were collected, respectively; with a resolution of 4 cm<sup>-1</sup> and a spectral range of 12 000–4 000 cm<sup>-1</sup>. Several classification models based on these spectral data were developed using partial least squares discriminant analysis (PLS-DA), Shrunken Centroid Discriminant Analysis (SCR-DA), Diagonal Linear Discriminant Analysis (DL-DA), Decision Tree (DT), Gradient Boosting (GB), Support Vector Machines (SVMs) and Artificial Neural Networks (ANNs) applied after principal components analysis (PCA). In addition to these, some pre-processing methods were compared by PLS-DA for improving the model performance. Standard Normal Variate (SNV), Multiplicative Scatter Correction (MSC), smoothing and transformation according to Savitzky-Golay (SG) algorithm and various combination of these pre-processing methods were employed. The accuracy rates of 50.79 %, 52.38 %, 53.97 %, 58.73 %, 61.38 %, 61.38 % and 87.30 % were observed based on SCR-DA, DL-DA, ANN, DT, GB, SVM and PLS-DA model in the testing set, respectively. In our study highest accuracy for raw data were found in PLS-DA model in the test set. The accuracy rates of pre-processing methods Smoothing, Autoscaling, Mean-centering, SNV, MSC, Standardization, Smoothing + Second Derivative, First Derivative, First Derivative+MSC and First Derivative+SNV for PLS-DA model were found 80.42 %, 80.42 %, 87.30 %, 88.89 %, 91.01 %, 91.98 %, 98.41 %, 98.94 %, 98.94 % and 99.47 %, respectively. The results demonstrated that Near Infrared Spectroscopy combined with pre-processing methods and multivariate data analysis could be an effective classifier of *Pinus* species.

# A Comparison of Estimation Methods for The Inverted Kumaraswamy Distribution

K. Bagci<sup>1</sup>, T. Arslan<sup>2</sup>

<sup>1</sup> Van Yuzuncu Yil University, Van, Turkey, [kubrabadgi@yyu.edu.tr](mailto:kubrabadgi@yyu.edu.tr)

<sup>2</sup> Van Yuzuncu Yil University, Van, Turkey, [mstalhaarlsan@yyu.edu.tr](mailto:mstalhaarlsan@yyu.edu.tr)

The Inverted Kumaraswamy (IKum) distribution is proposed by AL-Fattah et al. [1]. It is derived by applying transformation  $X = (1/T) - 1$  to the random variable T has Kumaraswamy distribution. It is a flexible distribution with two shape parameters along with a scale parameter. Also exhibits a longer right tail than several distributions that are widely used. This positively affects the distribution's ability to fit the rare events occurring in the right tail; see [1]. Maximum likelihood (ML) and Bayes estimators of unknown parameters of the IKum distribution are obtained by [1]. In this study, Least Squares (LS) and Maximum Product of Spacing (MPS) estimators of the parameters of the IKum distribution are obtained. A Monte Carlo simulation study is conducted to compare efficiencies of these estimators with the ML counterparts by means of the bias, and mean square error criteria. At the end of the study, actual data set is used to show the implementation of the LS, MPS and ML methodologies.

## References

- [1] AL-Fattah, A. M., EL-Helbawy, A. A., and AL-Dayian, G. R. (2017). Inverted Kumaraswamy Distribution: Properties and Estimation. *Pakistan Journal of Statistics*, 33(1), pp. 37–61.

# Collection of Recyclable Wastes within the Scope of Zero Waste Project: A Heterogeneous Multi-Vehicle Routing Case in Kirikkale

S. Kiziltas<sup>1</sup>, H.M. Alakas<sup>2</sup>, T. Eren<sup>3</sup>

<sup>1</sup> Kirikkale University, Kirikkale, Turkey, safakkiziltas7@gmail.com

<sup>2</sup> Kirikkale University, Kirikkale, Turkey, hmalagas@kku.edu.tr

<sup>3</sup> Kirikkale University, Kirikkale, Turkey, teren@kku.edu.tr

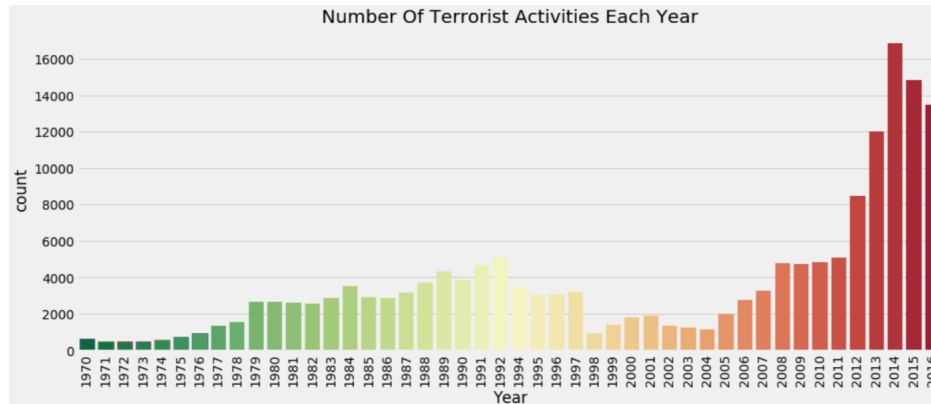
The importance of waste management services increase day by day due to the increase in the world population and urbanization. Collecting and transporting processes from waste management processes are considered as a vehicle routing problem in the literature. Vehicle routing problems are the problem of designing the routes of distribution or collection to the customers at the most appropriate cost with one or more vehicles. In this study, waste collection operations from all public institutions such as governorship, prefectures, registry offices and police departments etc. and all of public/private schools in Kirikkale Merkez and its 8 districts were discussed in line with "Zero Waste Project" initiated by Turkish Republic Ministry of Environment and Urbanization. At first, the demand forecasting was made for the collection of recyclable wastes such as paper, glass, plastic and metal and then the least cost of waste collection routes were found. The problem is considered as a heterogeneous multi-vehicle vehicle routing problem under working hour constraint.

# Analysis of Terrorist Attacks in The World

S.O. Rencber<sup>1</sup>

<sup>1</sup> Van Yuzuncu Yil University, Van, Turkey, rencber.serhatomer@gmail.com

Terrorism has spread death in many parts of the world. It is the biggest problem in the world. In this study, data sets with various terrorist attacks in the world are used. Using data mining methods, terrorist attacks and attack types around the world have been determined according to the countries.



According to data from the data sets, the most terrorist attacks were in Iraq. The attack, which has the most death, occurred in Iraq. The least terrorist attack has been in 1971. The most terrorist attack was in 2014. The least terrorist attack has been in Australia. Many such information is obtained and visualized in the results.

# Analysis of the Factors Affecting the Financial Failure and Bankruptcy by the Generalized Ordered Logit Model

M. G. Van<sup>1</sup>, S. Sehribanoglu<sup>2</sup>, M. H. Van<sup>3</sup>

<sup>1</sup> Van Yuzuncu Yil University, Van, Turkey, mgunayvan@gmail.com

<sup>2</sup> Van Yuzuncu Yil University, Van, Turkey, sehribanoglu@gmail.com

<sup>3</sup> Van Yuzuncu Yil University, Van, Turkey, hanifivan@yyu.edu.tr

It is important for companies to determine the financial and administrative problems in advance and to make the necessary arrangements without going into financial failure process. The aim of this study is to determine the factors that cause financial failure and bankruptcy of 139 industrial companies listed in the BIST (Stock Exchange Istanbul) in 2017 by the generalized ordered logit model.

In order to prevent the economic problems caused by financial failures, many methods have been developed and used from past to present. In our study, the score values of Altman (1968), which has the highest reliability, were obtained with Z score method. In this way, it can be determined which companies are within the three safe, gray and dangerous zones [1].

In the ordered logit model, the dependent variable consists of Z score discrimination zone values, while the independent variables consist of liquidity, activity, financial structure and profitability ratios obtained from company data. Among these ratios, the variables that are used most frequently and do not cause high correlation with each other are selected. The brant test was applied to check whether the ordered logit model results provide the assumption of parallel slopes. As a result of the Brant test, the generalized ordered logit model, which is a model that relaxed the parallelism assumption, is used considering the ordered structure of the dependent variable due to the violation of the parallel slope assumption. With the generalized ordered logit model, the factors affecting the financial failure and bankruptcy of the companies are estimated [2].

## References

- [1] Altman, E. I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589-609.
- [2] Williams, R. 2016. Understanding and interpreting generalized ordered logit models. *The Journal of Mathematical Sociology*, 40(1), 7-20.

# Using LSTM for Sentiment Analysis with Deeplearning4J Library

F. Ataman<sup>1</sup>, H.E. Celik<sup>2</sup>, F. Uludag<sup>3</sup>

<sup>1</sup> Van Yuzuncu Yi University, Van, Turkey, fataman@yyu.edu.tr

<sup>2</sup> Van Yuzuncu Yil University, Van, Turkey, ecelik@yyu.edu.tr

<sup>3</sup> Van Yuzuncu Yil University, Van, Turkey, fatihuludag@yyu.edu.tr@yyu.edu.tr

Sentiment analysis is a wide area of research that can be applied for many sectors. Approaches to sentiment analysis are classified in two general group. The first is lexicon based sentiment analysis while the other is machine learning based sentiment analysis. In this study we applied LSTM for sentiment analysis task. We used Deeplearning4J [1] library for LSTM algorithm. As a dataset we used Large Movie Review Dataset that is belong to IMDB includes 25000 positive and 25000 negative at the total 50000 movie reviews. We have combined Word2Vec [2] model with Recurrent Neural Network for sentiment classification. For this purpose we used Google News Vector as Word2Vec model dataset. At the end of evolution and training phase the task completed with 0.8624 accuracy rate; 0.8647 precision rate; 0.8624 recall rate; 0.8567 F1 Score rate. In this study we represent implementation of LSTM for sentiment analysis.

## References

- [1] Deeplearning4j Development Team, Deeplearning4j: Open-source distributed deep learning for the JVM, Apache Software Foundation License 2.0, deeplearning4j.org (Referenced May 2019)
- [2] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. Paper presented at the Advances in neural information processing systems.

# Network Log Analysis For Network Security By Using Big Data Technologies

S. Ziyanak<sup>1</sup>, H. E. Celik<sup>2</sup>, M. Kayri<sup>3</sup>

<sup>1</sup> Van Yuzuncu Yil University, Van, Turkey, [szivanak@yyu.edu.tr](mailto:szivanak@yyu.edu.tr)

<sup>2</sup> Van Yuzuncu Yil University, Van, Turkey, [ecelik@yyu.edu.tr](mailto:ecelik@yyu.edu.tr)

<sup>3</sup> Batman University, Batman, Turkey, [murat.kayri@batman.edu.tr](mailto:murat.kayri@batman.edu.tr)

Nowadays, the cyber-attacks which are made to government agencies, the servers of the universities and big companies, their systems and even to their internet infrastructures present big threats. So, the systems having Firewall, VPN, IPS features are used for providing network security. The purposes of these systems are to prevent the cyber-attacks from inside and outside, malwares which can strike to net infrastructure and to record all the network traffic.

Network traffic should be followed regularly and it is necessary to take precaution according to the detected threats for systems security and taking precautions to the threats which can occur in the future. This reveals the importance of network log records.

Network logs can have big sizes in time that aren't be able to read by standard software according to the network traffic density. Most of the time, network logs reach terabyte sizes and many valuable data cannot be analyzed. So, this causes to not to determine the features which can be threatful for the systems and net infrastructures and reveal the security gaps.

The data having huge sizes can be read fast and actively, analyzed and obtained important conclusions by using big data technologies. Network logs can be analyzed by examining with big data technologies and very useful results can be obtained.

In this study, network logs are analyzed in a short time by using Apache Hadoop and Apache Hive and it is shown that how useful and meaningful results can be deduced for network security.

## References

- [1] S. Lohr, "The Age of Big Data", New York Times, Feb 11, 2012, <http://www.nytimes.com/2012/02/12/sunday-review/big-datasimpact-in-the-world.html>
- [2] Voltage Security, Big Data, Meet Enterprise Security, White paper <http://www.voltage.com/solution/enterprise-security-for-big-data/>
- [3] Gartner-Research Firm, <https://www.gartner.com>
- [4] "Getting Real About Security Management And Big Data: A Roadmap for Big Data in Security Analytics", White Paper, RSAs and EMC Corporation, [www.EMC.com/rsa](http://www.EMC.com/rsa).
- [5] Arbor Networks Blog on "Next Generation Incident Response, Security Analytics and the Role of Big Data", <http://www.arbornetworks.com/corporate/blog/5126-next-generationincident-response-security-analytics-and-the-role-of-big-data-webinar>, Feb. 2014.
- [6] Matti, M. and Kvernvik, T. (2012). Applying Big-data technologies to Network Architecture, Ericsson Review.

# Face Recognition Based System Input Control Application

F. Ayata<sup>1</sup>, M. Inan<sup>2</sup>, E. Seyyarer<sup>3</sup>, H. Cavus<sup>4</sup>

<sup>1</sup> *Van Yuzuncu Yil University, Van, Turkey, ayatafaruk@gmail.com*

<sup>2</sup> *Van Yuzuncu Yil University, Van, Turkey, fubilmuh@gmail.com*

<sup>3</sup> *Van Yuzuncu Yil University, Van, Turkey, eseyyarer@gmail.com*

<sup>4</sup> *Van Yuzuncu Yil University, Van, Turkey, hayatiicavus@gmail.com*

Speed, time and safety are of great importance in many processes today. In addition to the access to information and the use of information, there are also many studies on the retention of information. Fingerprint, card reader and face recognition systems are used in the entrance and exit gates of state institutions and many large companies and in accessing system rooms of these institutions. The developed system restricts your personal computer use of foreign personalities by using face recognition algorithms. In addition to this limitation, the system automatically pulls a picture of the person who wants to use your personal computer and sends the information to the mobile phone mentioned in the system.



# Extractive Text Summarization Via Graph Partitioning

T. Uckan<sup>1</sup>, C. Hark<sup>2</sup>, E. Seyyarer<sup>3</sup>, F. Ayata<sup>4</sup>, M. Inan<sup>5</sup>, A. Karci<sup>6</sup>

<sup>1</sup> *Van Yuzuncu Yil University, Van, Turkey, taneruckan@yyu.edu.tr*

<sup>2</sup> *Bitlis Eren University, Bitlis, Turkey, cengizhark@hotmail.com*

<sup>3</sup> *Van Yuzuncu Yil University, Van, Turkey, eseyyarer@yyu.edu.tr*

<sup>4</sup> *Van Yuzuncu Yil University, Van, Turkey, ayatafaruk@gmail.com*

<sup>5</sup> *Van Yuzuncu Yil University, Van, Turkey, fubilmuh@gmail.com*

<sup>6</sup> *Inonu University, Malatya, Turkey, ali.karci@inonu.edu.tr*

Automatic text summarization is the automatic creation of a short and fluent summary of the text. There are different approaches to summarizing by selecting sentences from the main text. One of the text summarization methods used to create an automatic summary is to summarize the text by selecting a sentence from the original text. In this study, we present a graph-based extractive text summarization method. This method provides a summary of the two main steps. The first of these steps includes the representation of the input text and the steps of creating the graph. The second step includes the steps of graph partitioning and sentence scoring. A text preprocessing tool developed by us is used to protect the semantic loyalty between sentences in the representation phase of the input texts. After the representation of the texts with the graphs, the graphs partitioning perform. After this division, the sentences to be found in the abstract in terms of the number of sentences found in the subsections obtained are determined. In this method, Closeness Centrality and Degree Centrality methods were used in sentence scoring stage. The most valuable nodes obtained as a result of these methods are included in the abstract. Finally, this study, which is proposed for the purpose of text summarization, has been calculated by using ROUGE evaluation metrics in the Document Understanding Conference (DUC- 2002) data set including open access texts and summaries of these texts. As a result of the measurements, it was observed that the proposed summation system offers significant accuracy.

# **The Evaluation of the Effect of the Earthquake on Socio-Economic Development Level with the Cluster Analysis**

M. B. Gorentas<sup>1</sup>, S. Saracli<sup>2</sup>

<sup>1</sup> *Van Yuzuncu Yil University, Van, Turkey, burakgorentas@gmail.com*

<sup>2</sup> *Afyon Kocatepe University, Afyon, Turkey, ssaracli@aku.edu.tr*

Socio-economic development refers to the development of settlements in social, cultural, environmental and spatial dimensions as well as economic development. Cluster analysis, unknown groups precisely, units, similar subsets of variables with each other (group, class) is one of multivariate statistical analysis, which helps to separate. With strong mathematical foundation and is a method used in almost all branches of science. In this study, the study conducted by H. Eray CELIK, Sinan SARACLI and Sanem SEHRIBANOGLU in 2011, before the earthquake in Van province was compared with our study in 2017, after the earthquake in Van province. Thus, the effect of the earthquake on the socio-economic development level of the Van province was observed.

# Modelling of Photovoltaic Power Generation based on Weather Parameters Using Regression Analysis

E. Bicek<sup>1</sup>, H.E. Celik<sup>2</sup>, N. Genc<sup>3</sup>, S.O. Rencber<sup>4</sup>, E. Kina<sup>5</sup>, F. Kapar<sup>6</sup>

<sup>1</sup> Van Yuzuncu Yil University, Van, Turkey, emrebicek@yyu.edu.tr

<sup>2</sup> Van Yuzuncu Yil University, Van, Turkey, ecelik@yyu.edu.tr

<sup>3</sup> Van Yuzuncu Yil University, Van, Turkey, nacigenc@yyu.edu.tr

<sup>4</sup> Van Yuzuncu Yil University, Van, Turkey, serhatomer@gmail.com

<sup>5</sup> Van Yuzuncu Yil University, Van, Turkey, erolkina@yyu.edu.tr

<sup>6</sup> Van Yuzuncu Yil University, Van, Turkey, firatkapar@yyu.edu.tr

Since renewable energy production is more cost-effective than thermal energy, it has become the energy source that the whole world is concentrated on. In 2018, 26% of global energy production came from renewable energy. With a capacity of 100 Gigawatts (GW) (55%), solar energy became the most common renewable energy source installed in the world in 2018. Wind power (28%) and Hydroelectric (11%). Total installed solar power capacity in the world is 505 GW. Solar photovoltaic (PV) has become the worlds fastest- growing energy technology, with gigawatt-scale markets in an increasing number of countries. Optimal PV panel installation is very important in terms of cost and requirements. A number of estimation methods have recently been used to meet the long-term planning needs of solar panels for optimal use. In this study, the energy production estimation of a photovoltaic system with a capacity of 17.16 kW installed on the roof of Van Yuzuncu YI University Science Research and Application Center in Turkey was performed by using regression analysis. The meteorological data of the region were used in the analysis.

## References

- [1] Zheng, Z. W., Chen, Y. Y., Huo, M. M., and Zhao, B. (2011). An overview: the development of prediction technology of wind and photovoltaic power generation. *Energy Procedia*, 12, 601-608.
- [2] De Giorgi, M. G., Congedo, P. M., and Malvoni, M. (2014). Photovoltaic power forecasting using statistical methods: impact of weather data. *IET Science, Measurement & Technology*, 8(3), 90-97.
- [3] Ekstrom, J., Koivisto, M., Millar, J., Mellin, I., and Lehtonen, M. (2016). A statistical approach for hourly photovoltaic power generation modeling with generation locations without measured data. *Solar Energy*, 132, 173-187.
- [4] REN21, Renewables 2019 Global Status Report (Paris: June 2019)

# Uniformly Convergence of Singularly Perturbed Reaction-Diffusion Problems on Shishkin Mesh

K. Yamac<sup>1</sup>, F. Erdogan<sup>2</sup>,

<sup>1</sup> Van Yuzuncu Yil University, Van, Turkey, kyamac@yyu.edu.tr

<sup>2</sup> Van Yuzuncu Yil University, Van, Turkey, ferdogan@yyu.edu.tr

In this study, we investigated the uniform convergence of numerical solutions of singularly perturbed reaction-diffusion problems on the shishkin mesh. A Numerov-type exponentially fitted difference scheme was constructed and it was shown that the scheme was uniform convergence with respect to  $\epsilon$ . Convergence was supported by numerical example.

## References

- [1] J.J.H. Miller, E. O'Riordan, G.I. Shishkin, (1996). Fitted Numerical Methods for Singular Perturbation Problems. Error Estimates in the Maximum Norm for Linear Problems in One and Two Dimensions World Scientific, Singapore.
- [2] P.A. Farrell, A.F. Hegarty, J.J.H. Miller, E.O'Riordan and G.I.Shishkin, (2000). Robust Computational Techniques for Boundary Layers, Chapman-Hall/CRC, New York.
- [3] H. MacMullen, J.J.H. Miller, E. O'Riordan, G.I. Shishkin, (2001). A second order parameter-robust overlapping Schwarz method for reaction-diffusion problems with boundary layers. *J. Comput. Appl. Math.*, 130(12), pp.231-244.
- [4] K.C.Patidar,(2005).Highorderfittedoperatornumericalmethodforself-adjointsingular perturbation problems. *Applied Math.and Comp.* (Vol.171 pp.547-566).
- [5] G.M. Amiraliyev, (2005). The convergence of a finite difference method on layer-adapted mesh for a singularly perturbed system. *Applied Mathematics and Computation*, 162(3), pp. 1023-1034.
- [6] R.K. Bawa, (2007). A Paralel aproach for self-adjoint singular perturbation problems using Numerovs scheme. *International Journal of Computer Math.*, 84(3), pp.317- 323.
- [7] G.M. Amiraliyev, F. Erdogan, (2009). Difference schemes for a class of singularly perturbed initial value problems for delay differential equations. *Numer. Algorithms*, 52 (4), pp. 663-675.

# Factorial Moment Generating Function Of Sample Minimum Of Order Statistics From Geometric Distribution

A. Demir<sup>1</sup>, A.M. Karakas<sup>2</sup>, S. Calik<sup>3</sup>

<sup>1</sup> Firat University, Elazig, Turkey, a.binici128@gmail.com

<sup>2</sup> Bitlis Eren University, Bitlis, Turkey, aysekarakas5767@gmail.com

<sup>3</sup> Firat University, Elazig, Turkey, scalik@firat.edu.tr

In this study, factorial moment generating function of sample minimum from a geometric distribution of statistics based on order statistics are considered. The moments of the sample minimum of order statistics from a geometric distribution are obtained with the help of factorial moment generating function. Using these moments, the expected value and variance were obtained as algebraic and numerical.

## References

- [1] Abdel-ATY, S.H., (1954) . Ordered variables in discontinuous distributions. *Stat. Neerlandica*, 8, 61-82.
- [2] Ahsanullah, M. and Nevzorov, V. B., (2001). *Ordered Random Variables*, Nova Science Publishers, Inc., New York.
- [3] Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N., (1992). *A First Course in Order Statistics*, John Wiley and Sons, New York.
- [4] Balakrishnan, N., (1986). Order Statistics from Discrete Distributions. *Commun. Statist Theory Meth.*, 15(3), 657- 675.
- [5] Bugatekin, A., (2015). Moment Generating Functions of Sample Minimum of Order Statistics from Geometric Distributions. NWSA.2015.10.3.3A0073.
- [6] Calik, S., Gngr, M. and Colak, C., (2010). On the Moments of Order Statistics From Discrete Distribution. *Pak. J. Statist.*, 26(2), 417-426.
- [7] David, H. A. and Nagaraja, H. N., (2003). *Order Statistics*. John Wiley and Sons, New York.
- [8] Khatri, C. G., 1962. Distribution of Order Statistics for Discrete Case, *Ann. Inst. Statist. Math.*, 14, 167-171.
- [9] Nagaraja, H. N., 1992. Order Statistics from Discrete Distribution (with discussion). *Statistics*, 23, 189-216.

# Comparisons of Methods of Estimation Generalized Exponential Distribution

A. Duva Yilmaz<sup>1</sup>, A.K. Demirel<sup>2</sup>

<sup>1</sup> Van Yuzuncu Yil University, Van, Turkey, [asumanduva@yyu.edu.tr](mailto:asumanduva@yyu.edu.tr)

<sup>2</sup> Ordu University, Ordu, Turkey, [aysekubrademirel@gmail.com](mailto:aysekubrademirel@gmail.com)

The generalized exponential distribution plays an important role in modeling various data set in many areas such as, biological, medical and reliability function. Also, the distribution is used an alternative to Gamma and Weibull distributions in many situations. Thus, it is very important to determine the best estimation method for distribution parameters. The main objective in this study is to determine the best estimators of the unknown parameters generalized exponential distribution. Hence, we briefly discussed different methods for estimation of the unknown parameters of generalized exponential distribution. Furthermore, the performances of the estimators are compared with respect to their biases and mean square errors through a simulation study. Finally, a real data set is analyzed for better understanding of methods presented in this study.

## References

- [1] Chen, D. G., and Lio, Y. L. (2010). Parameter estimations for generalized exponential distribution under progressive type-I interval censoring. *Computational Statistics & Data Analysis*, 54(6), 1581-1591.
- [2] Casella G and Berger R.L. (2002). *Statistical inference* (Vol.5). Duxbury, USA.
- [3] Kundu D. and Pradhan B. (2009). On progressively censored generalized exponential distribution test, 18(3), 497.
- [4] Kara M. and Koak V. (2016). *Statistical inference with generalized exponential distribution*, Yuzuncu Yil University.

# Interdependence Of Bitcoin And Other Crypto Money Indicators: Cd Vine Copula Approach

A.M. Karatas<sup>1</sup>, A. Demir<sup>2</sup>, S. Calik<sup>3</sup>

<sup>1</sup> *University of Bitlis Eren, Bitlis, Turkey, aysekarakas5767@gmail.com*

<sup>2</sup> *Firat University, Elazig, Turkey, a.binici128@gmail.com*

<sup>3</sup> *Firat University, Elazig, Turkey, scalik@firat.edu.tr*

This paper aims to examine the relationship between bitcoin and other crypto money indicators with the CD Vine Copula Approach method. In the study, we use closing prices of Bitcoin, Bitcoin Cash, Ethereum, Litecoin and XRP. The results show that there is a weak dependence between bitcoin and prominent financial indicators. These findings indicate the necessity of more detailed studies.

## References

- [1] Brechmann, E., Schepsmeier, U. (2013). Cdvine: Modeling dependence with c-and d-vine copulas in r. *Journal of Statistical Software*, 52(3), 1-27.
- [2] Kim, D., Kim, J. M., Liao, S. M., Jung, Y. S. (2013). Mixture of D-vine copulas for modeling dependence. *Computational Statistics and Data Analysis*, 64, 1-19.
- [3] Czado, C., Brechmann, E. C., Gruber, L. (2013). Selection of vine copulas. In *Copulae in Mathematical and Quantitative Finance*. Springer, Berlin, Heidelberg 17-37.

# New Complex Hyperbolic Mixed Dark Soliton Solutions for Some Nonlinear Partial Differential Equations

W. Gao<sup>1</sup>, H.M. Baskonus<sup>2</sup>, J. L. G. Guirao<sup>3</sup>

<sup>1</sup> *Yunnan Normal University, Yunnan, China, gaowei@ynnu.edu.cn*

<sup>2</sup> *Harran University, Sanliurfa, Turkey, hmbaskonus@gmail.com*

<sup>3</sup> *Technical University of Cartagena, Cartagena, Spain, juan.garcia@upct.es*

This work focuses on obtaining new complex hyperbolic and mixed dark solutions for some nonlinear partial differential equations, namely, (2+1)-dimensional asymmetrical Nizhnik-Novikov-Veselov and SawadaKotera (SK) equations via sine-Gordon expansion method. This powerful method is based on two important properties of sine-Gordon equation. We generate new solitary wave solutions to the governing models. With the help of symbolic computation package programmes, we plot some grafical surfaces of them including high and lower points in a large range of independant variables. The results for the governing models are graphically introduced.

## References

- [1] Liu, J.G., (2019). Lump-type solutions and interaction solutions for the (2 + 1)-dimensional asymmetrical Nizhnik-Novikov-Veselov equation. *European Physical Journal Plus*, 134(56), 1-6.
- [2] Zhao, Z.L. , Chen, Y., Han, B., (2017). Lump soliton, mixed lump stripe and periodic lump solutions of a (2+1)-dimensional asymmetrical Nizhnik Novikov Veselov equation. *Modern Physics Letters B*, 31, 1750157.
- [3] Osman, M.S., (2017). Multiwave solutions of time-fractional (2+1)-dimensional Nizhnik-Novikov-Veselov equations. *Pramana*, 88(67), 1-10.



# Complex Solitons in the Conformable (2+1)-dimensional Ablowitz-Kaup-Newell-Segur Equation

W. Gao<sup>1</sup>, G. Yel<sup>2</sup>, H.M. Baskonus<sup>3</sup>, C. Cattani<sup>4,5</sup>

<sup>1</sup> Yunnan Normal University, Yunnan, China, gaowei@ynnu.edu.cn

<sup>2</sup> Final International University, Kyrenia Mersin 10, Turkey, gulnur.yel@final.edu.tr

<sup>3</sup> Harran University, Sanliurfa, Turkey, hmbaskonus@gmail.com

<sup>4</sup> Tuscia University, Viterbo, Italy, cattani@unitus.it

<sup>5</sup> Ton Duc Thang University, Ho Chi Minh City, Vietnam

In this study, we study the Conformable (2+1)-dimensional Ablowitz-Kaup-Newell-Segur equation in order to show the existence of complex combined dark-bright soliton solutions. To this purpose an effective method which is the Sine-Gordon expansion method is used. The 2D and 3D surfaces under some suitable values of parameters are also plotted.

## References

- [1] He, J. H., (2005). Application of Homotopy Perturbation Method to Nonlinear Wave Equations. *Chaos, Solitons and Fractals*, 26, 695700.
- [2] He, J. H., (2005). Homotopy perturbation method for bifurcation of nonlinear problems, *Int. J. Nonlinear Sci. Numer. Simul.*, 6, 207208.
- [3] Liao, S. J., (2003). Beyond Perturbation: Introduction to the Homotopy Analysis Method, Chapman & Hall/CRC Press.
- [4] Kocak, Z. F., Bulut, H., Yel, G., (2014). The solution of fractional wave equation by using modified trial equation method and homotopy analysis method, AIP Conference Proceedings,

# Fitting Irregular Migration Data in Turkey to Itô Stochastic Differential Equation

A. Shamilov<sup>1</sup>, F. Erdogan<sup>2</sup>, S. Ozdemir<sup>3</sup>

<sup>1</sup> *Eskisehir Technical University, Eskisehir, Turkey, asamilov@eskisehir.edu.tr*

<sup>2</sup> *Van Yuzuncu Yil University, Van, Turkey, ferdogan@yyu.edu.tr*

<sup>3</sup> *Van Yuzuncu Yil University, Van, Turkey, sevdaoazdemir@yyu.edu.tr*

In this study Irregular Migration Data is investigated by Ito Stochastic Differential Equation Modeling (SDEM). A typical one-dimensional stochastic differential equation has the form

$$dX(t) = f(t, X)dt + g(t, X)dW(t) \quad (1)$$

for  $0 \leq t \leq T$  where  $X(0, \omega) \in H_RV$ ,  $H_RV$  is a Hilbert space of random variable,  $X(t, \omega)$  is a stochastic process not a deterministic function.  $W(t, \omega) = W(t)$  is a Wiener process or Brownian motion and since it is nowhere differentiable [1].

Firstly, parameters of SDE (see equation (1)) which occurs in biological problems are estimated by using the maximum likelihood procedure. Moreover, by applying Euler-Maruyama Approximation Method are obtained approximate trajectory of stochastic process which is solution of mentioned SDE.

This method allows to obtain approximate trajectory according to observation  $x_i$  on interval  $[t_i, T]$  by formula

$$X_i = X(t_i, \omega) = X(t_{i-1}, \omega) + f(t_{i-1}, X(t_{i-1}, \omega))\Delta t + g(t_{i-1}, X(t_{i-1}, \omega))\sqrt{\Delta t}\delta(0, 1) \quad (2)$$

where  $t_0 = 0; t \in [0, T]; t_i = i\Delta t/K; i = 1, \dots, K, \Delta t = T/K$ ,  $K$  is number of steps using Euler-Maruyama method.

The performances of SDE are established by Chi-Square criteria, Root Mean Square Error (RMSE) criteria. It should be noted that in this research, the data of the number of irregular migrants those who have been captured in Turkey between 2005 and 2019 (until 29.05.2019) is examined. The results are acquired by using statistical software R-Studio and MATLAB R2013a. These results are also corroborated by graphical representation.

## References

- [1] E.J. Allen, (2007). Modeling with Ito Stochastic Differential Equations, Springer, USA.
- [2] J. Bak, A. Nielsen, H. Madsen, (1999). Goodness of fit of stochastic differential equations. 21. Symposium i Anvendt Statistik, Copenhagen Business School, Copenhagen, Denmark.
- [3] Shamilov A., (2007). Im Teorisi, Olasilik ve Lebesgue Integrali, Anadolu Universitesi Yayinlari, Eskisehir.
- [4] Shamilov A., Bozdag B., (2011). Hisse Senetlerinin Fiyatlandirilmasi icin Yeni Bir Stokastik Model Onerisi, Istatistik Arastirma Dergisi, 8(2), pp.21–26.

# Attitudes of Students Towards Coding Learning Supported With Educational Computer Games: A Case Study for Van Province

A. Yaman<sup>1</sup>, C. Guler<sup>2</sup>

<sup>1</sup> Van Yuzuncu Yil University, Van, Turkey, kalime1986@hotmail.com.tr

<sup>2</sup> Van Yuzuncu Yil University, Van, Turkey, cetinguler@yyu.edu.tr

Computer training and coding education has become an indispensable part of educational environments as a consequence of rapid developments in information technologies and accessibility of technology. The coding education, which takes place in the curriculum from the beginning of primary school, brings success in various scopes such as analytical thinking, creativity and problem solving. It is well known that children having a close interest in information technologies gain a strong capability in solving encountered problems and analysing obtained results.

The structuring of coding learning is undoubtedly one of the most important elements of the educational process defined in information technologies. Accordingly, the most efficient method in structuring of coding learning is the coding education supported with educational computer games. In this study, an attitude scale for the attitude of students towards coding education supported with educational computer games was utilized with a view to determine the attitude of students towards coding education supported with educational computer games. A total number of 173 students were selected from 5th, 6th and 7th grades of four schools, including two private and two public schools, affiliated to the National Educational Directorate of Van Province of Turkey, in spring term of 2018-2019 academic year. It is expected that obtained results from this study can contribute to future studies related to attitudes of students towards coding education supported with educational computer games in terms of school, grade, age and gender characteristics.

## References

- [1] Voogt, J., Fisser, P., Pareja Roblin, N., Tondeur, J., and van Braak, J. (2013). Technological pedagogical content knowledge: a review of the literature. *Journal of Computer Assisted Learning*, 29(2), 109-121.
- [2] Drayton, B., Falk, J. K., Stroud, R., Hobbs, K., and Hammerman, J. (2010). After installation: Ubiquitous computing and high school science in three experienced, high-technology schools. *Journal of Technology, Learning, and Assessment*, 9(3), n3.
- [3] Cope, B., and Kalantzis, M. (2009). Multiliteracies: New literacies, new learning. *Pedagogies: An International Journal*, 4(3), 164-195.

# Design and Optimization of Graphene Quantum Dot-based Luminescent Solar Concentrator Using Monte-Carlo Simulation

A. Rostami<sup>1</sup>, M. Rastkar Mirzaei<sup>1</sup>, S. Matloub<sup>1</sup>, H. Mirtagioglu<sup>2</sup>

<sup>1</sup> *University of Tabriz, Tabriz, Iran, rostami@tabrizu.ac.ir*

<sup>2</sup> *Bitlis Eren University, Bitlis, Turkey, hamitsa86@gmail.com*

Monte Carlo (MC) methods are a subset of computational algorithms that use the process of repeated random sampling to make numerical estimations of unknown parameters. It is suitable for the case of luminescent solar concentrators (LSCs) as there is no deterministic analysis for an LSC. Furthermore, the physical processes responsible for its performance has many coupled degree of freedom. We have used Monte-Carlo ray tracing simulation to design and optimize an efficient LSC. We have used graphene quantum dots as luminescent material because of its unique properties compared to inorganic quantum dots like low toxicity, photostability, and tunable photoluminescence. The focus of this study is the optimizing the graphene quantum dot concentration and the waveguide size and scale. We have discussed the choice of efficient LSC scale to make a balance between maximum obtainable optical power and energy flux gain (e.g., cost of solar cells). Moreover, we have optimized the parameters and discussed the results for different graphene quantum dots with different quantum yields. These results optimization method is general and can be applied to all other LSCs.

# **A Review: Big Data Technologies with Hadoop Distributed Filesystem**

R. R. Asaad<sup>1</sup>, A.Y. Hussien<sup>2</sup>, S. M. Almufti<sup>3</sup>

<sup>1</sup> *Nawroz University, Duhok, Iraq, renas\_rekany@yahoo.com*

<sup>2</sup> *Van Yuzuncu Yil University, Van, Turkey, amar.yehya85@yahoo.com*

<sup>3</sup> *Nawroz University, Duhok, Iraq, saman.almofty@gmail.com*

Today Big Data, is any set of data that is larger than the capacity to be processed using traditional database tools to capture, share, transfer, store, manage and analyze within an acceptable time frame; from the point of view of service providers, Organizations need to deal with a large amount of data for the purpose of analysis. And IT department are facing tremendous challenge in protecting and analyzing these increased volumes of information. The reason organizations are collecting and storing more data than ever before is because their business depends on it. The type of information being created is no more traditional database-driven data referred to as structured data rather it is data that include documents, images, audio, video, and social media contents known as unstructured data or Big Data. Big Data Analytics is a way of extracting value from these huge volumes of information, and it drives new market opportunities and maximizes customer retention. Moreover, this paper focuses on discussing and understanding Big Data technologies and Analytics system with Hadoop distributed filesystem (HDFS). This can help predict future, obtain information, take proactive actions and make way for better strategic decision making.

# **Guler and Linaro et al Model in an Investigation of the Neuronal Dynamics using noise Comparative Study**

R. R. Asaad<sup>1</sup>, A.Y. Hussien<sup>2</sup>, S. M. Almufti<sup>3</sup>

<sup>1</sup> *Nawroz University, Duhok, Iraq, renas\_rekany@yahoo.com*

<sup>2</sup> *Van Yuzuncu Yil University, Van, Turkey, amar.yehya85@yahoo.com*

<sup>3</sup> *Nawroz University, Duhok, Iraq, saman.almofty@gmail.com*

Recently, theoretical arguments, numerical simulation and experiments shown that ion channel noise in neurons can have deep impact on the behavior of the neuron's dynamical when there is a limited size for the membrane space. It can be create different models of Linaro al equations by using stochastic differential equations to find the impacts of ion channel noise, and it has been analytically put forward the Guler model. More recently, Guler has discussed that in small neurons the rate functions for the closing and opening of gates are under the effect of the noise. In this research, the investigation of dynamics neurons are determined with noise rate functions. The exact Markov simulations will be employ during the investigation with above analytical models. Comparatively, the results will be presented from these models. The research aims to show more details on the phenomenon recently outlined by Guler.

# A Robust Confidence Interval and Ratios of Coverage to Width for the Population Coefficient of Variation: A Comparative Monte Carlo Study

H.E. Akyuz

*Bitlis Eren University, Bitlis, Turkey, heakyuz@beu.edu.tr*

In this study, it is proposed a robust confidence interval for the population coefficient of variation. It is compared with some existing confidence intervals, which are proposed by McKay [1], Hendricks and Robey [2], and Donner and Zou [3], in terms of coverage probability, average width and ratio of coverage to width for normal and some skewed distributions. A simulation study is conducted to evaluate performances of the proposed confidence interval in MATLAB. Simulation study shows that the coverage probabilities of the proposed confidence interval are very close to nominal confidence level for  $\alpha=0.05$ . It is seen that the its average widths are narrower than the existing confidence intervals. Also, ratios of coverage to width of proposed confidence interval are higher than the others. Thus, it is recommended to use the robust confidence interval for the coefficient of variation.

## References

- [1] McKay, A.T. (1932). Distribution of the coefficient of variation and the extended t distribution. *Journal of the Royal Statistical Society*, 95 (4), 695-698.
- [2] Hendricks, W.A., and Robey, W.K. (1936). The sampling distribution of the coefficient of variation. *The Annals of Mathematical Statistics*, 7 (3), 29-132.
- [3] Donner, A., and Zou, G.Y. (2010). Closed-form confidence intervals for function of the normal mean and standard deviation. *Statistical Methods in Medical Research*, 21 (4), 347-359.