
JATIT

Welcome to Journal of Theoretical and Applied Information Technology

Home

Volumes

Submit Paper

Manuscript Status

Author Guidelines

Editorial Board

Indexing and Abstracting

Subscribe to JATIT

Contact Us

Frequency : Monthly

Submit Paper / Call for Papers

Journal receives papers in continuous flow and we will consider articles from a wide range of Information Technology disciplines encompassing the most basic research to the most innovative technologies. Please submit your papers electronically to our submission system at http://jatit.org/submit_paper.php in an MSWord, Pdf or compatible format so that they may be evaluated for publication in the upcoming issue. This journal uses a blinded review process; please remember to include all your personal identifiable information in the manuscript before submitting it for review, we will edit the necessary information at our side. Submissions to JATIT should be full research / review papers (properly indicated below main title in case of review paper).

Editorial Board

JATIT
Welcome To The Research Community

Welcome to Journal of Theoretical and Applied Information Technology

Journal of Theoretical and Applied Information Technology
March 2019 | Vol. 97 No.06

Frequency :MONTHLY

Title:	CORRELATES OF SUCCESS IN JAVA PROGRAMMING: PREDICTING THE PERFORMANCE OF STUDENTS IN A JAVA COURSE FROM THE RESULTS OF LOWER LEVEL COURSES
Author:	BASHIR S GALADANCI, MARYAM I MUKHTAR, SANAH A MUAZ
Abstract:	Over the years, a number of factors that could determine success in programming have been investigated; yet computer programming is still challenging to most students. This paper is aimed at determining the relationship among students' results of a third year Java Workshop Course with those of lower level courses taken in year one and year two and then developing predictive models based on this relationship. The examination results in 20 courses for a total of 109 students of Bayero University Kano that enrolled for BSc Computer Science course in the 2012/2013 session are analyzed using correlation and linear regression. The correlation results revealed that out of the 20 selected courses, the strongest correlation occurs in the course ((CSC2251: Analysis of Algorithms) ($r= 0.756$), $p<0.001$). Some programming courses, Computer Science theory courses (CSC) and Mathematics courses were also found to have strong correlations with the Java performance. The regression results revealed that a regression module, based upon the linear combination of (CSC 2203(VB), CSC2212 (C++), (CSC2251 (Analysis of Algorithms), CSC2202 (Data Structures) and CSC 1303(Introduction to Computer Science)) which is statistically significant at, $F_{(5, 103)} = 47.09$, $p = .000$ accounts for 70% variance in the Java Programming results.
Keywords:	Predicting Performance, Java Programming, Correlation, Regression, Students
Source:	Journal of Theoretical and Applied Information Technology 31st March 2019 -- Vol. 97. No. 06 -- 2019

Full Text

Title:	THE CITIZENS ACCEPTANCE FACTORS OF TRANSPORTATION APPLICATION ONLINE IN BATAM: AN ADAPTATION OF THE UTAUT2 MODEL AND INFORMATION SYSTEM SUCCESS MODEL
Author:	MANGAPUL SIAHAAN, NILO LEGOWO
Abstract:	The growth of the internet and the advance of information technology, made very significant changes to transportation in Indonesia especially in Batam City. The emergence of applications for online transportations services such as Gojek / Grab has made some people switch to using conventional transportation to online transportation. But with the development of online transportation, there are still people in Batam City taking action against online transportation. For this reason, researchers conducted research to look for factors that influence the acceptance of the people of Batam City on online transportation. The researcher used the UTAUT2 research model (Unified Theory of Acceptance and Use of Technology 2) and the IS Success Model DeLone & McLean. The construct of this research model uses constructs of performance expectancy, effort expectancy, facilitating conditions, price value, information quality, system quality, service quality, behavioral intention, user satisfaction, use behavior. The number of respondents in this study amounted to 400 respondents using SPSS Amos version 22. The results of the analysis showed that the factors that significantly influenced the acceptance of the people of Batam City on the application of online transportation services were seen from constructs of behavioral intention (user intention) 91.2%, performance expectancy is 51.7%, price value is 42.6%, while construct of user satisfaction is influenced by system quality by 68%, information quality is 28.2%, service quality is 12.6%.
Keywords:	Transportation application online, UTAUT2, Information System Success Model Five
Source:	Journal of Theoretical and Applied Information Technology 31st March 2019 -- Vol. 97. No. 06 -- 2019

Full Text

Title:	DEVELOPMENT AND ANALYSIS OF QUALITY COMPREHENSIVE MODEL WITH SCRUM TREE ALGORITHM FOR AUTOMATIC USER STORY PROCESSING
Author:	MRS. RUPALI A. MAHAJAN, DR. S. K. YADAV, DR. SURENDRA A. MAHAJAN
Abstract:	Scrum framework is needed to keep up openness of a project life cycle. If structure ceases to conserve the transparency subsequently it is vital to hunt down the reason behind same. This paper provides the new algorithm that helps you to examine the openness within tasks as per scrum user stories. The core assumption is that, user stories are back bone of work and this can be formulated at preliminary stage of project cycle and the actual performance of stories takes a long time to accomplish. If any of these tasks are time intensive, it is crucial to formulate sub-stories for major task and conduct whole Scrum cycle for same. To address this problem, proposed research focusing on decision goal modeling with scrum tree algorithm where project manager can get automatic goal scheduling effectively as per scrum stories. The evaluation of proposed system is done for real time project and also tested for Thurman dataset. Performance parameters like data centrality
Keywords:	Scrum, Agile, Kanban, Sprint, Sprint Retrospective
Source:	Journal of Theoretical and Applied Information Technology 31st March 2019 -- Vol. 97. No. 06 -- 2019

Full Text

Title:	EXPLORING THE GENERAL AWARENESS OF YOUNG USERS ACCORDING TO AIDA MODEL APPLIED TO SOCIAL NETWORKING ADS
Author:	OSAMA AHMED ABDELKADER, MOHAMED HASSAN RABIE
Abstract:	The information revolution represents an essential feature of the era that the world is currently living in. Social Networking is one of the important elements of this revolution. According to the literature review, previous studies focused on measuring of the model elements depending on specific perceptions toward determined ads. This study presents a

	contemporary application of AIDA model stages; Attention, Interest, Desire and Action on Social Networking Advertise (SNA), with the impact analysis of gender and SNA-image. This paper reviews the literature need of measuring the general awareness of SN users according to AIDA model applied to SNAs. Additionally, it explores the perceptions of youth who represent the largest sector among SN users, especially universities' students. A proposed questionnaire was contributed as a research instrument, the reliability and validity of the questionnaire were tested by a group of statistical measures included Cronbach's α and Exploratory Factor Analysis (EFA). Data were collected based on snowball sampling. There are 738 participants in this survey; 570 females and 168 males. After data collection, the goodness fit of the proposed model was examined and ensured by seven fit indicates. The main contributions of this study are exploring the general perception of young users about AIDA stages applying on SNA. Moreover, the influences of SNA-image and gender on AIDA stages were tested. These contributions enable marketers and advertisers to predict the effectiveness of their advertising campaigns through social networking sites (SNSs). Findings discussion, limitations of the study and suggested future researches were provided to extend the applicability of its contributions.
Keywords:	Marketing, Social Networking Advertise, AIDA, Forecasting, Ad Effectiveness.
Source:	Journal of Theoretical and Applied Information Technology 31st March 2019 -- Vol. 97. No. 06 -- 2019

Full Text

Title:	COMPREHENSIVE ANALYSIS OF SCOPE OF NEGATION FOR SENTIMENT ANALYSIS OVER SOCIAL MEDIA
Author:	NIKHIL KUMAR SINGH, DEEPAK SINGH TOMAR
Abstract:	Handle Negation is essential for effective sentiment analysis decision support system. Negation control comprised identification of negation cues, scope of negation and their influence within it. Negation can either invert, reduce or increase the polarity score. This paper present comprehensive assessment of recent research on negation control for sentiment analysis technique. Explore negation cue and scope detection techniques in collaboration with classification technique over social media data set. This assessment has included the evaluation of sentiment classification (Support vector machine, Navies Bayes, Linear Regression and Random Forest) and scope detection techniques (conjunction Analysis, Punctuation Mark and grammatical dependency tree) over presented prepossessing framework. This paper yield interesting result about collective response of negation scope detection and classification technique for sentiment analysis. Negative scope feature vector significantly increase the polarity classification accuracy of sentiment classification technique. Grammatical dependency tree in collaboration with SVM and Naves Bayes can detect negation with better accuracy as compere to conjunction and punctuation word scope detection technique.
Keywords:	Sentiment Analysis, Negation Cues, Scope Detection, Conjunction Analysis, Punctuation Mark, Grammatical Dependency Tree, Navies Bayes, Linear Regression, Random Forest, SVM.
Source:	Journal of Theoretical and Applied Information Technology 31st March 2019 -- Vol. 97. No. 06 -- 2019

Full Text

Title:	OBJECT-BASED IMAGE ANALYSIS APPLIED FOR DIFFERENT STAGES OF RUBBER PLANTATIONS MAPPING USING THAICHOTE SATELLITE DATA
Author:	WASANA PUTKLANG, CHARAT MONGKOLSAWAT, RASAMEE SUWANWERAKAMTORN
Abstract:	During 2000 to 2011, rubber plantations rapidly expanded in northeast Thailand, which had not been historically planted. Information about planted areas and their distribution is a prerequisite for formulating land use planning and understanding its consequences on ecosystems. This study aimed to establish a model for digitally devising a synergistic approach to distinguishing the different stages of rubber plantations in the northeasternmost region of Thailand and a small portion of the Lao People's Democratic Republic (Lao PDR). The combination of Object-Based Image Analysis (OBIA), Vegetation Canopy Density (VCD), plant phenology and intensive ground observation was applied to THAICHOTE satellite data. Two levels of classification based on OBIA approach were performed. At the first level, multi-scale image segmentation of pansharpened imagery was performed to divide the image set into objects with different spectral and spatial characteristics. Incorporating the normalized difference vegetation index (NDVI) and brightness index (BI) into the objects, the image set was subdivided into four different subsets of VCD. Analyses were then performed at the next level classification on each of VCD subsets by using certain and a range of different approaches to discriminate stand age rubber tree plantations. Rubber tree phenology and OBIA feature optimization were used to differentiate the different stages of rubber plantations. The results indicated that the agreement between field-based classification and image-based classification was well correlated. The overall accuracy of 79.00 % and Cohen's kappa coefficient of 0.77 were achieved for the integrated models for the different stage of rubber plantations.
Keywords:	THAICHOTE satellite data; Different stages of rubber plantations; Object-based image analysis (OBIA); Vegetation canopy density (VCD); Plant phenology; Northeast Thailand
Source:	Journal of Theoretical and Applied Information Technology 31st March 2019 -- Vol. 97. No. 06 -- 2019

Full Text

Title:	EMPIRICAL STUDIES ON THE IMPACT OF SOFTWARE CUSTOMIZATION ON QUALITY ATTRIBUTES: A SYSTEMATIC REVIEW
Author:	ABDULRAZZAQ QASEM ALI , ABU BAKAR MD SULTAN , ABDUL AZIM ABD GHANI , HAZURA ZULZALIL
Abstract:	Software needs to deliver acceptable levels of quality, whilst meeting the functional demands of users. This need is gaining more consideration in Enterprise Resource Planning (ERP) and multi-tenant software. This study investigated the effect of software customization on external quality attributes via four pertinent primary studies, which were summarized, aggregated, discussed and presented through a systematic literature review. The primary studies in this review were chosen through the application of inclusion and exclusion criteria to applicable articles published during the period 2000-2016; they were analysed using research questions that focused on software customization, software quality attributes and measures, approaches, and impact results. The results revealed that each primary study had been applied in the context of ERP and demonstrated that different types of customizations can have various effects on particular quality attributes. Accordingly, further research is needed to determine the impact of customization on each external quality attribute.
Keywords:	Customization; ERP; Software Quality Attributes; Systematic Literature Review.

Source:	Journal of Theoretical and Applied Information Technology 31st March 2019 -- Vol. 97. No. 06 -- 2019

Full Text

Title:	ASSOCIATION RULES IMPLEMENTATION FOR AFFINITY ANALYSIS BETWEEN ELEMENTS COMPOSING MULTIMEDIA OBJECTS
Author:	MENDOZA-PALECHOR, FABIO E. , CARRASCAL OVIEDO, DE-LA-HOZ-FRANCO, EMIRO
Abstract:	The multimedia objects are a constantly growing resource in the world wide web, consequently it has generated as a necessity the design of methods and tools that allow to obtain new knowledge from the information analyzed. Association rules are a technique of Data Mining, whose purpose is to search for correlations between elements of a collection of data (data) as support for decision making from the identification and analysis of these correlations. Using algorithms such as: A priori, Frequent Parent Growth, QFP Algorithm, CBA, CMAR, CPAR, among others. On the other hand, multimedia applications today require the processing of unstructured data provided by multimedia objects, which are made up of text, images, audio and videos. For the storage, processing and management of multimedia objects, solutions have been generated that allow efficient search of data of interest to the end user, considering that the semantics of a multimedia object must be expressed by all the elements that composed of. In this article an analysis of the state of the art in relation to the implementation of the Association Rules in the processing of Multimedia objects is made, in addition the analysis of the consulted literature allows to generate questions about the possibility of generating a method of association rules for the analysis of these objects.
Keywords:	Association Rules, Multimedia Object, Data Mining, Data-Set, Correlations.
Source:	Journal of Theoretical and Applied Information Technology 31st March 2019 -- Vol. 97. No. 06 -- 2019

Full Text

Title:	ENHANCING ARABIC NAMED ENTITY RECOGNITION USING PARALLEL TECHNIQUES
Author:	ZIADOON A. OTAIWI, MOHAMMED A. OTAIR, KHALDOON A. AOTAIWE, AHMAD ODAT, FERAS ALMASHAKBAH
Abstract:	Named entities recognition systems (Proper Names) are used in the development of many natural language processing applications. There is a paucity of published research in the field of identifying the named entities from texts written in Arabic. This is due to the fact that the Arabic language has a specificity regarding the complexity of spelling and morphology, which is an obstacle to the development of a technique to identify the names of the Arabic entities or the so-called Arabic Named Entity Recognition system (ANER). This paper presented the experiments conducted to identify the appropriate technique to design a robust and reliable system for identifying Arabic entities. For this purpose, this study focuses on the most common state-of-art in the field of identification of Arabic named entities, then a comparison was made between five of the most famous tools that interested in identifying the Arab entities, after that, integrated each of two tools together to get 10 different parallel techniques. The results of the comparison between the tools showed that Rosette achieved the best results followed by Madamira, while it was the worst performance results in the gate tool and for hybrid systems, the R-F (combining Rosette and Farasa) achieved the best performance with better accuracy than individual tools.
Keywords:	Parallel Techniques, Arabic Named Entity, Named Entity Recognition, Tool
Source:	Journal of Theoretical and Applied Information Technology 31st March 2019 -- Vol. 97. No. 06 -- 2019

Full Text

Title:	ANALYSIS USABILITY AND CONTENT IN KNOWN SYSTEM IMPLEMENTATION
Author:	RITAWATI , AHMAD NURUL FAJAR
Abstract:	The purpose of this study is to analysis usability and content of KNOWN system implementation in Bank Syariah Mandiri (BSM). Based on the data, KNOWN system visitor is less than 10 percent on January to April 2017. The author will conduct research to find out whether the employee's user experience in using KNOWN has gone well or not. This study, will focus on 2 factors in user experience, namely usability and content. This study consists of five steps, such as: (1) Questionnaire Preparation, (2) Questionnaire Distribution, (3) Usability Testing, (4) Analyzing and Data Interpretation, and (5) Recommendation. The results of this study are how good the level of usability and content on the KNOWN site and KNOWN recommendations in the future for the company.
Keywords:	Usability, Usability Testing, Knowledge Management System
Source:	Journal of Theoretical and Applied Information Technology 31st March 2019 -- Vol. 97. No. 06 -- 2019

Full Text

Title:	INTERCONNECTED MODERN HIGHER EDUCATION IN INDONESIA BASED ON SOA
Author:	ALDIAN NURCAHYO, AHMAD NURUL FAJAR, SEPTIA REDISA SRIRATNASARI, MOH SUKRON MUFAQIH, SARIFUL BACHTIAR
Abstract:	The adoption of Service Oriented Architecture (SOA) in modern higher education will help to solve the problems of adaptability by reducing integration complexity of new systems, reusing services, interoperability and increasing the agility of learning system. Furthermore, share services and cloud computing solutions could helped modern higher education and universities to create low costs and efficient service management by collaboration and open environment. The creation off interconnected modern higher education ecosystem has become an important way for players in the field of education. This paper is to propose SOA adoption solutions for universities and education institutions by modelling and designing the interconnected modern higher education's system. The aim is to create a new educational environment where learners have digital course and information accessible easily to support their educational and career goals. On the other hand, the solution is to understand the way universities and education institution solved demands complexity of the digital learning system by reducing investment costs. Moreover, we are purposing new alternatives of the architecture systems that focused on interoperability and cross-platform functionality in order to deal with high level components diversity.
Keywords:	Service Oriented Architecture SOA, Interconnected Modern Higher Education, Universities, E-Learning.
Source:	Journal of Theoretical and Applied Information Technology 31st March 2019 -- Vol. 97. No. 06 -- 2019

Full Text

Title:	NOVEL ARCHITECTURE OF CLOUD COMPUTING FOR INTERNET OF THINGS
Author:	CHETNA DABAS
Abstract:	In the years to come, futuristic research and business opportunism will hugely rely on the fusion of two very popular technologies called Cloud Computing and Internet of Things. The victory of Internet of Things seeks for service accessories consisting up of crucial parameters like reliability, scalability, ubiquity, good performance to name a few. This further demands cloud computing to be in place for the concrete support of the Internet of Things paradigm. Cloud Computing is weaved with in-built intellect support for the integration with Internet of Things and on the other hand cloud environment aims at boosting up of the Internet of Things scalability and cost factors in great ways. This research paper presents review of literature, surveyed practices, points out research issues in the above context, and proposes a novel fused architecture of Cloud Computing for the Internet of Things technology, validates the proposed architecture while implementing an application using IOT devices and Google Cloud, compared the proposed work for MQTT protocol with the existing work and performs better than it, further, in another existing work the cloud component was completely missing whereas it is embedded in the proposed architecture for offering scalability in the Internet of Things scenario, at last this research paper concludes with the open research issues and assumptions along with limitations of the proposed work.
Keywords:	Internet of Things, Cloud Computing, Architecture, Applications, Games
Source:	Journal of Theoretical and Applied Information Technology 31st March 2019 -- Vol. 97. No. 06 -- 2019

Full Text

Title:	SEMANTIC INTEGRATION SHARING FOR E-GOVERNMENT DOMAINS ONTOLOGY: DESIGN AND IMPLEMENTATION USING OWL
Author:	ABDULLAH ALSHEHAB1, NASER N. ALAZEMI, HAWAF ABD ALHAKEM
Abstract:	Integrating information system in any domain for E-government is facing several problems, such as structural and semantic sharing in similar information for applications and services. In addition, representation across heterogeneous organizations. In this aspect, the semantic web played a dynamic role in sharing and managing knowledge sharing. Therefore, e-government services need to modeling, re-engineering governmental sharing, and processing to provide information delivery according to the citizen and locations. Semantic Web technology based on the ontology has brought promising solutions to the above engineering problems. This paper is ongoing for previous work presents a framework for generating semantic model ontologies in OWL syntax from a government service domain. However, current works employing OWL ontologies in e-government to the Semantic Web viewer, which implemented semantic sharing integration for e-government domains (Civil information, Health information, Education Information) ontology, which present the case of the State of Kuwait interoperability integration. The domain ontology is created using ontology knowledge-based editor Prot�g� and OWL-Viz. Thus facilitating to improve the knowledge sharing so that citizens can take more knowledge benefits from the e-government sharing. In addition, the ontology will be useful in the knowledge sharing of semantic e-government integration, and semantic Web applications. Thus facilitating the design of e-government systems that can be easily integrated and maintained. Web Prot�g� ontology can be easily sharing with a distributed group of users who can engage in collaborative authoring activities from wherever they happen to be logged in.
Keywords:	E-government; Semantic Web; Domain ontology; Services; Framework; owl; Integration; Citizens; Kuwait e-government; OWL-Viz; Prot�g�
Source:	Journal of Theoretical and Applied Information Technology 31st March 2019 -- Vol. 97. No. 06 -- 2019

Full Text

Title:	BOOTSRAPPING INSTANCE-BASED ONTOLOGY MATCHING VIA UNSUPERVISED GENERATION OF TRAINING SAMPLES
Author:	MANSIR ABUBAKAR, HAZLINA HAMDAN, NORWATI MUSTAPHA, TEH NORANIS MOHD ARIS
Abstract:	Training set is the key role player that can improve the performance of any classification task. Different techniques and methods are being applied to generate training set depending on its area of application. Researchers in data science and semantic web community use different kind of training sets generated to improve the performances of classifications and information retrieval capability. Operational Training Set Generator (TSG) should always solve a minimum of two issues; (1) it must address the computational cost in producing a reasonable outcome, thereby reducing the computational cost in the whole system. The run-time of TSG is near linear as in blocking approach and (2) it must produce the qualitative training sets. We use LogTfIdf as the cosine similarity function of two given vectors to produce Bag of Words (BoW); the tokenizer is developed to specially take care of delimiters that often come across URIs and other RDF essentials. We evaluated our UTSG on nine cross-domain benchmark ontologies publically available in OAEI website. The results obtained shows that our UTSG outperforms the two baseline TSGs previously developed to address similar problem.
Keywords:	Semantic Web, Link Open Data, Semantic Heterogeneity, Ontology Matching, Instance-based Matching, Training Set
Source:	Journal of Theoretical and Applied Information Technology 31st March 2019 -- Vol. 97. No. 06 -- 2019

Full Text

Title:	CLASSIFICATION OF PNEUMONIA PATIENTS RISK USING HYBRID GENETIC ALGORITHM-DISCRIMINANT ANALYSIS AND NA�VE BAYES
Author:	IRHAMAH, SITI MAR�ATUS RAHIMATIN, HERI KUSWANTO, LAKSMI WULANDARI
Abstract:	Pneumonia is the most common causes of death in developing countries, such as in Indonesia. Therefore, appropriate pneumonia classification is very important in determining the disease severity and to know the most appropriate treatment for the patient. In this study, Discriminant Analysis (DA), hybrid Genetic Algorithm- Discriminant Analysis (HGA-DA) and Na�ve Bayes (NB) are used to classify risk class of patient. GA is an artificial intelligent method that can avoid a trap in local optima and easy to implement in solving various objective functions and constraints, while NB is a simple but powerful method that returns not only prediction but also the degree of certainty. In this study, GA is used to improve multi-class classification performance of DA. Firstly, GA is used for variable selection in DA, and then a comparative study with other variable selection methods is performed. In addition, Genetic Algorithm is also implemented for parameter estimation.

	Analysis results show that there are differences in selected variables from four selection methods in classifying patient risk class. The use of hybrid methods of DA and GA in variable selection and parameter optimization stages gives better multi-class classification results than DA or NB, since it produces highest Geometric Mean (GM) and Area Under Curve (AUC) criterion.
Keywords:	Pneumonia, Multi-class Classification, Discriminant Analysis, Genetic Algorithm, Nave Bayes
Source:	Journal of Theoretical and Applied Information Technology 31st March 2019 -- Vol. 97. No. 06 -- 2019

Full Text

Title:	MODELING THE STRUCTURE OF SOCIAL NETWORKS BY USING THE PYTHAGOREAN SPIRAL
Author:	IZEM ACIA, WAKRIM MOHAMED, GHADI ABDERRAHIM
Abstract:	Since the good representation of the complex networks structure can effectively communicate more information and can help explore them and understand their behaviors, the purpose of this paper is to present and model the structure of social networks. Mathematical concepts can be a powerful tool for modeling. However, complex networks in the real world are far too complicated to model in their entirety. Therefore, it is important to identify the information that the generated model may help illustrate. In the present work, we use the spiral of Pythagoras and a combination of mathematical concepts in pre-topology, graph theory and fuzzy logic to generate two different models: the Rings Model and the Member-ship Matrix. The first proposed model has a geometric representation which makes all the nodes visible and every direct link between two nodes illustrates the number of, direct and indirect, paths that connect them and the cost of each one. The second proposed model is a matrix where every column depicts the relation between the corresponding node and the rest of the network. The most important benefit of the models presented in this paper is that they allow the human eye to get information about the strength of connection between the nodes, of complex networks, in an easy and fast way.
Keywords:	Spiral of Pythagoras, Pretopology, Fuzzy Logic, Social Networks, Complex Networks, Mathematical Modeling, Matrix, Rings Model
Source:	Journal of Theoretical and Applied Information Technology 31st March 2019 -- Vol. 97. No. 06 -- 2019

Full Text

Title:	RELIABLE LICENSE PLATE RECOGNITION SYSTEM BASED ON ADAPTIVE NON-PLATE FILTRATION METHODS
Author:	LAMA ALKHALED, MUHAMMAD EHSAN RANA
Abstract:	This paper proposes a technique to filter out the false negative detection of license plate caused by using edge algorithm to specifying the targeting areas. The proposed technique implements filters bases on the histogram value for the three main channels of the image (RGB). The paper consists of the following three main stages: (a) Detection based on edge detecting method, (b) calculating the histogram values of RGB color model, and (c) finally, verifying and detecting valid license plate regions by applying different histogram- based rules on features extracted out from the calculated histogram values. Experimental results show that the proposed method is very effective in assisting to neglect many noises caused by different conditions such as poor illumination and varied weather
Keywords:	license Plate Detection (LPD), Histogram Distribution, Detection Accuracy, False Rejection Rate (FRR)
Source:	Journal of Theoretical and Applied Information Technology 31st March 2019 -- Vol. 97. No. 06 -- 2019

Full Text

Title:	RESOURCE ALLOCATION OPTIMIZATION BASED ON CHANNEL QUALITY FOR LONG TERM EVOLUTION SYSTEMS (LTE)
Author:	SARA RIAHI, AZZEDDINE RIAHI
Abstract:	In this paper we focus on an important task of the eNodeB in the architecture of LTE networks, the RRM (Radio Resource Management) its goal is to accept or reject network connection requests, ensuring optimum distribution of radio resources between UEs (Users equipment). It consists mainly of two elements AC (Admission Control) and PS (Packet Scheduling). In this work we will focus on the PS, which achieves an efficient allocation of radio resources in both directions is to say Uplink (considered in our case) and downlink. Several approaches and algorithms have been proposed in the literature to address this need (allocate resources effectively) this diversity and variety of algorithms is related to the factors considered for the optimal management of radio resource, specifically the type of traffic and requested QoS by the EU. In this paper a study of several scheduling algorithms proposed for LTE (uplink and downlink) is made. Therefore, we offer our evaluation and reviews. In this paper we are interested in the allocation of radio resources in LTE uplink (uplink) in particular with a comparative study between the scheduling algorithms flows that are: Round Robin (RR) Max Min Fair (MMF), Maximum-Largest Weighted Delay First (M-LWDF) and Exponential Proportional Fair (PF EXP).We considered the real-time stream or RT (Video and VoIP), considering the QoS criteria: time, spectral efficiency and throughput. The results obtained show the advantages and disadvantages of using one algorithm over another.
Keywords:	Resource allocation, Uplink, Downlink, LTE, scheduling algorithms.
Source:	Journal of Theoretical and Applied Information Technology 31st March 2019 -- Vol. 97. No. 06 -- 2019

Full Text

Title:	POST-EDITING OF WORDS IN KAZAKH SENTENCES FOR INFORMATION RETRIEVAL
Author:	SHORMAKOVA A., ZHUMANOV ZH., RAKHIMOVA D.
Abstract:	In this paper, the proposed method for determining incorrect words for any context is described in more details. Also, the paper describes the completeness of the meaning of words in Kazakh sentences for incorrect words in post-editing. And the second main task is to find the most suitable variant for these words found by the full meaning of the word for any context. According to the results of the study, the relevance and weighty significance for the Kazakh language was shown to determine the full meaning of the word. After the work done, this methodology shows that it is very relevant and effective especially for the Turkic languages. Although the system is aimed to find the wrong words and analyze these words

	for post-editing, but these modules can be considered as separate modules for working with contexts. The work was also done for information retrieval and for the analysis of the initial input data. In this work, the essence of the work of a search system is indicated which uses the above-said algorithm for the algorithm for searching and processing data.
Keywords:	Completeness Of Word; Kazakh Language; Intelligent Analysis Of Sentences ; Machine Translation; Information Retrieval.
Source:	Journal of Theoretical and Applied Information Technology 31st March 2019 -- Vol. 97. No. 06 -- 2019

Full Text

Title:	NEW ALGORITHM IN IMAGE COMPRESSION BASED ON DISCRETE COSINE TRANSFORM
Author:	ZYAD THALJI
Abstract:	Digital image compression is the most important field in digital image processing. Communication technologies and storage are using compression to compressed digital images. Digital image compression provides storage space when storing images on storage media and saving time when transferring data over communication networks. This paper aimed to compress an image by resizing the image dimension, the resizing image based on pixel location. The study used many methods to reach its goals. The method of separating the colour image is used to split the image into three intensity images (red intensity images, blue intensity images and green intensity images). The Discrete Cosine Transform (DCT) is used after the splitting the resizing image, resizing the image dimension reduced the dimension of the image to quarter. The new algorithm was applied for ten selected images from the internet, the experiment results show that the new algorithm has a high efficiency in image compression.
Keywords:	Image Compression, Image Resize, Discrete Cosine Transform, Inverse Discrete Cosine Transform, Colour Image.
Source:	Journal of Theoretical and Applied Information Technology 31st March 2019 -- Vol. 97. No. 06 -- 2019

Full Text

Title:	AN ANALYTICAL APPROACH FOR BANKRUPTCY PREDICTION USING BIG DATA AND MACHINE LEARNING TECHNIQUE
Author:	S.SAROJINI DEVI , Dr Y.RADHIKA
Abstract:	Bankruptcy is defined as a legal procedure used to claim the identity of an organization or a person on the basis of their creditworthiness and debtor. There is a huge requirement for effective prediction models to evaluate the risks arising in bankruptcy at early stages and to overcome financial losses. Another critical risk considered in bankruptcy is periodic increase in financial data and updates, which increase the size of the system. The existing tools and techniques are limited to less number of inputs and big data cannot be processed directly with these techniques. Big data analytic technique, such as Hadoop, provides an opportunity to process large financial data, which is retrieved from data source such as DataStream, FAME, Company, and House. In this research, bankruptcy prediction model has been developed by using a combined approach of Big Data and Naive Bayes machine learning algorithm. A huge amount of financial data is extracted and stored in Hadoop database through key pair conversion technique and big data analysis and prediction is carried out using naive Bayes
Keywords:	Big Data Analytics; Bankruptcy; Naive Bayes Algorithm; Financial Models; Datastream; FAME (Forecasting Analysis And Modelling Environment); Industrial Risk; Management Risk; Financial Flexibility; Credibility; Competitiveness; Operational Risk.
Source:	Journal of Theoretical and Applied Information Technology 31st March 2019 -- Vol. 97. No. 06 -- 2019

Full Text

Title:	DEVELOPING BIG DATA FRAME WORKS FOR EFFICIENT INTELLIGENT TRANSPORT SYSTEM FOR SAFETY AND SECURITY
Author:	P. KIRAN KUMAR , A. VIJAY GOPAL
Abstract:	Big data is an advanced technology where we can use its services for our application. Big data handles large volumes of data upon a sudden rise and request of the users. Transportation system is one where we have to handle big data. Based on the gathered data the decision has to be taken for safer measures and efficient, profitable and secured. In this paper we conducted a survey on big data in ITS (Intelligent Transportation System). Data collection, analytics and frame works can be explained. We concluded from our survey by using big data we will resolve our traditional traffic problems. We attempted to provide architecture for ITS using big data thencalculation of bayonet traffic flow, calculation ofaverage speed of a road, querying the travel path of a vehicle, checking and controlling the fake vehicles.
Keywords:	ITS, big data, vehicle tracking, Large Data Storage, Cloud Computing.
Source:	Journal of Theoretical and Applied Information Technology 31st March 2019 -- Vol. 97. No. 06 -- 2019

Full Text

Journal of Theoretical and Applied Information Technology
31st March 2019. Vol.97. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

CLASSIFICATION OF PNEUMONIA PATIENTS RISK USING HYBRID GENETIC ALGORITHM-DISCRIMINANT ANALYSIS AND NAÏVE BAYES

[1]IRHAMAH, [2]SITI MAR'ATUS RAHIMATIN, [3]HERI KUSWANTO, [4]LAKSMI WULANDARI

[1,2,3]Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

[4]Faculty of Medicine, Universitas Airlangga, Surabaya, Indonesia

E-mail: [1]irhamah@statistika.its.ac.id

ABSTRACT

Pneumonia is the most common causes of death in developing countries, such as in Indonesia. Therefore, appropriate pneumonia classification is very important in determining the disease severity and to know the most appropriate treatment for the patient. In this study, Discriminant Analysis (DA), hybrid Genetic Algorithm- Discriminant Analysis (HGA-DA) and Naïve Bayes (NB) are used to classify risk class of patient. GA is an artificial intelligent method that can avoid a trap in local optima and easy to implement in solving various objective functions and constraints, while NB is a simple but powerful method that returns not only prediction but also the degree of certainty. In this study, GA is used to improve multi-class classification performance of DA. Firstly, GA is used for variable selection in DA, and then a comparative study with other variable selection methods is performed. In addition, Genetic Algorithm is also implemented for parameter estimation. Analysis results show that there are differences in selected variables from four selection methods in classifying patient risk class. The use of hybrid methods of DA and GA in variable selection and parameter optimization stages gives better multi-class classification results than DA or NB, since it produces highest Geometric Mean (GM) and Area Under Curve (AUC) criterion.

Keywords: *Pneumonia, Multi-class Classification, Discriminant Analysis, Genetic Algorithm, Naïve Bayes*

1. INTRODUCTION

Pneumonia is a type of disease that causes serious problems that can lead to consolidation of lung tissue and disturbances of local gas exchange, and also cause high morbidity [1]. According to the Basic Health Research (RISKESDAS) [2], the prevalence of pneumonia in Indonesia in 2013 is 4.5%. Compared to other provinces in Java, East Java has the highest prevalence of Acute Respiratory Infection (ARI) that is 4.2%. Pneumonia is typically grouped into community acquired (Community Acquired Pneumonia) and hospital admissions (Hospital Acquired Pneumonia) [3]. Community pneumonia (CAP) is ranked on fourth of the ten most-treated diseases per year [4].

Early examination is necessary in the prevention of pneumonia. Knowing the level of classification of the disease is very important in order to accelerate the determination of the most appropriate treatment for the patient. Assessment on the severity of pneumonia is an important component in the management of community pneumonia. This led to the emergence of various scoring systems such as PSI, CURB-65, modified ATS (m-ATS) and so forth [3]. Some commonly used scoring systems are the PSI system (Pneumonia Severity Index) developed by the PORT (Pneumonia Patient Outcome Research Team) recommended by the American Thoracic Society Guidelines (ATS) and the CURB-65 system which is a recommendation from the BTS (British Thoracic Society).

In this study, the Genetic Algorithm is used in variable selection and optimization of the estimated parameters of the discriminant function. Researches related to variable selection methods and parameter optimization in the Discriminant Analysis have been done by [5] and [6] resulted in the conclusion that the classification using hybrid analysis of discriminant-genetic algorithm has a higher accuracy than using only discriminant. Thus, in this study, Genetic Algorithm is used for variable

selection and parameter estimation of discriminant function in classifying pneumonia patients' risk class. Furthermore, the results will be compared to Naïve Bayes classification method result. Although included in a simple classification method, a study conducted by [7] concludes that Naïve Bayes produced a higher classification accuracy than the Artificial Neural Network method. Another study conducted by [8] gives the conclusion that the Naïve Bayes method has the highest accuracy value compared to the Support Vector Machine (SVM) and Random Forest methods. Hence, in this research, a comparative study of the performance of Discriminat Analysis, Hybrid Genetic Algorithm-Discriminat Analysis and Naïve Bayes to classify pneumonia patients risk are conducted.

2. DISCRIMINANT ANALYSIS

Discriminant Analysis is a multivariate technique concerned with separating distinct sets of objects and to allocate new objects into previously defined groups. Discriminant analysis is one of the dependency techniques, finding the effect of the dependent variable based on several independentvariables [9]. Discriminant analysis works when the measurements made on categorical dependent variable and continuous independent variables [10]. Discriminant analysis is one of statistical methods that can be used for classification. The linear discriminant function takes form of a linear combination of coefficients of p variables and their respective variables in the study as equation(1).

A variate of the independent variables selected for their discriminatory power used in the prediction of group membership. The predicted value of the discriminant is the discriminant Z score, which is calculated for each object in the analysis. It takes the form of the linear equation (1).

$$Z_{ji} = a + w_1 x_{1i} + w_2 x_{2i} + \ldots + w_p x_{pi} \quad (1)$$

where

Z_{ji} : discriminant Z score of discriminant function j for object i

a : constant/ intercept

w_p : discriminant weight for the independent variable p

x_{pi} : independent variable p and object i [10]

In discriminant analysis, there are assumptions that must be fulfilled prior to performing the analysis:

(1) The data follows a multivariate normal distribution

The Mardia test can be used to test the null hypothesis H_0 of multivariate normality. The Mardia's skewness statistic for this hypothesis is

$$\text{MS} = \frac{n}{6}\hat{\gamma}_{1,p} \quad (2)$$

and Mardia's kurtosis statistic

$$\text{MK} = \hat{\gamma}_{2,p} \quad (3)$$

where

$$\hat{\gamma}_{1,p} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n} m_{it}^3 \quad (4)$$

$$\hat{\gamma}_{2,p} = \frac{1}{n}\sum_{i=1}^{n} m_{ii}^2 \quad (5)$$

with $m_{it} = (x_i - \bar{x})'^{S^{-1}}(x_t - \bar{x})$ is mahalanobis distance. [11]

(2) Homogeneity of variance-covariance matrices. Box's M tests the null hypothesis of homogeneity of covariance matrices.

(3) Equality of Group Means
The tests of equality of group means measure each independent variable's potential before the model is created

Variable selection methods used in this study are stepwise method, forward selection, backward elimination and genetic algorithm. The details of the method are given as follows:

1. *Stepwise Method*
It involves entering the independent variables into the discriminant function one at a time on the basis of their discriminating power. The stepwise approach follows sequential process of adding or deleting variables in the following manner:
   - Choose the single best discriminating variable
   - Pair the initial variable that is best able to improve the discriminating power of the function in combination with the first variable.
   - Select additional variables. Note that as additional variables are included, some

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

previously selected variables may be removed if the information they contain about group differences is available in some combination of the other variables included at later stages.

- Consider the process completed when either all independent variables are included in the function or the excluded variables are judged as not contributing significantly to further discrimination. [10]

2. *Backward Elimination*
   Discriminant analysis is performed by using model containing all independent variables. Then the independent variable having the smallest partial F value or largest *Wilks' Λ* value is chosen. If p_{value} indicates that this independent variable is significant at the αlevel then the procedure terminates by choosing the model containing all independent variables. If this independent variable is not significant at the α level or $p_{value} \geq \alpha$, it is removed from the model and the discriminant analysis is performed by using all the remaining independent variables. At each step an independent variable is removed from the model if it has the smallest partial F value and it is not significant at the α level. The procedure terminates when no independent variable in the model can be removed.
3. *Forward Selection*
   Variable selection is done by entering one by one independent variable that has the largest partial F value and $p_{value} < \alpha$.
4. Genetic Algorithm

   This method selects the predictor by selecting combination of variables that produces the best fitness value.

The coefficients are estimated such that the function maximizes the distance between the two centroids. That happens when a ratio (λ)-between group sum of squares to within group sum of squares is maximized [12]. Then the vector of coefficients a that maximizes the ratio given as follows.

$$\frac{\boldsymbol{\alpha}'\boldsymbol{B}\boldsymbol{\alpha}}{\boldsymbol{\alpha}'\boldsymbol{W}\boldsymbol{\alpha}} = \frac{\boldsymbol{\alpha}'(\sum_{j=1}^{k}(\overrightarrow{\bar{x}_j}-\bar{x})(\bar{x}_j-\bar{x})')\boldsymbol{\alpha}}{\boldsymbol{\alpha}'(\sum_{j=1}^{k}\sum_{k=1}^{n_l}(x_{jk}-\bar{x}_j)(x_{jk}-\bar{x}_j)')\boldsymbol{\alpha}} \quad (6)$$

where:

$\boldsymbol{\alpha}$ = coefficient vector of discriminant function

$\bar{\boldsymbol{x}}$ = mean vector

$\boldsymbol{x}_{jk}$ = measurement on independent variable j and category k

Then we calculate the value of discriminant function for the classification and then allocate x to π_l if

$$\sum_{j=1}^{r}(\hat{y}_j - \bar{y}_{lj})^2 = \sum_{j=1}^{r}[\hat{a}_j'(x-\bar{x}_l)]^2 \leq \sum_{j=1}^{r}[\hat{a}_j'(x-\bar{x}_m)]^2, l \neq m \quad (7)$$

where $\hat{a}'_j$ is parameter coefficient vector, $y_{lj} = \hat{a}'_j \bar{x}_l$ and$r \leq s$. x will be allocated to π_l when the value of $\sum_{j=1}^{r}(\hat{y}_j - \bar{y}_{lj})^2$ is minimum. The performance of classification function can be evaluated by calculating the geometric mean (GM) and Area Under Curve (AUC) using formula below.

$$G\_mean = \left(\prod_{j=1}^{k} R_j\right)^{\frac{1}{k}} \quad (8)$$

$$AUC = \frac{1}{k}\sum_{j=1}^{k} R_j \quad (9)$$

Where

$$R_j = \frac{n_{ii}}{\sum_{j=1}^{k} n_{ij}}, j = 1,2,..,k \quad (10)$$

3. GENETIC ALGORITHM

The Genetic Algorithm is used to solve both constrained and unconstrained optimization problems based on natural selection, the process that drives biological evolution. Genetic algorithm can used to solve a variety of optimization problems that are not well suited for standard optimization algorithms, including problems in which the objective function is discontinuous, non-differentiable, stochastic, or highly non-linear [13]. Each iteration consists of the following steps [14]:

(i) Initialization: The initialization value used comes from the encoding of a gene in chromosome which represents the individual genes.

(ii) Evaluate the fitness of each chromosome in the population.

(iii) Selection: Apply Roulette Wheel Selection where each chromosome occupies a circular piece of roulette wheel proportionately according to the fitness value.

(iv) Crossover: a process that occurs on the chromosome aimed to increase the diversity of chromosome in a population.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

(v) Mutation: used to prevent the algorithm from being trapped in the optimum local solution [15].
(vi) Evaluation. Then the fitness of the new chromosomes is evaluated.
(vii) Replacement. During the last step, individuals from the old population are replaced by the new ones.
(viii) Elitism. Save and copy several best chromosomes into next iteration.

4. NAÏVE BAYES CLASSIFIER

Naive Bayes is one of the simplest probabilistic classifiers based on the Bayes theorem. Naïve Bayes Classification is statistical classification method can be used for predicting membership of class [16]. Generally, Bayes theorem can be formulated as

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \quad (11)$$

If there is a quantitative or continuous attribute, then $P(X_l|Y)$ will be calculated using normal distribution approach

$$P(X_l|Y) = \frac{1}{\sigma_{lj}\sqrt{2\pi}} exp\left(-\frac{(\bar{X}_l - \mu_{lj})^2}{2\sigma_{lj}^2}\right) \quad (12)$$

Probability estimation $P(X_l|Y)$ can be calculated for each attribute X_l and class Y so that new data can be classified into certain Y based on largest probability value among it.

5. PNEUMONIA

Pneumonia is an inflammation that affects the lung parenchyma that leads to consolidation of lung tissue and local gas exchange disturbances [17]. The first management of pneumonia patients after being diagnosed is the determination of the treatment site based on the severity of pneumonia. Assessment of the severity of pneumonia is an important component in the management of community pneumonia. Some of the commonly used scoring systems are the PSI system (Pneumonia Severity Index) developed by the American Thoracic Society Guidelines (ATS) and the CURB-65 system which is a recommendation from the BTS (British Thoracic Society) [3]. These scores are also used as a guide for selection of antibiotic therapy and morning treatment of patients with pneumonia. PSI Scoring System is presented in Table 1 and 2.

Table 1: Scoring System on Community Pneumonia PSI System

Characteristics of Patients	Score
Demographic Factors	
Age: Male	age (year)
Female	age (year)-10
Homa care	+10
Complicated Desease	
Malignancy	+30
Liver disease	+20
Cognestive heart failure	+10
Cerebrovascular disease	+10
Kidney illness	+10
Physical Examination	
Changes in mental status	+20
Respiratory > 30 times/minute	+20
Systolic ≥ 90 mmHg	+20
Temperature <35°C or >40°C	+15
Pulse ≥125 times/minute	+10
Laboratory or Radiology Result	
Arterial Blood Gas Analysis : pH < 7,35	+30
BUN (*Blood urea nitrogen*)>30 mg/dL	+20
Sodium <130 mEq/liter	+20
Glucose >250 mg/dL	+10
Hematocrit <30%	+10
PO_2≤60 mmHg	+10
Pleural Effusion	+10

Then the points of the PORT result are summed up. The summations are then categorized according to the risk class, so that appropriate treatment can be determined.

Table 2: Risk Score with PSI System

Risk	Class	Total Score	Recommendation
Low	I	0	Outpatient
	II	<70	Outpatient
	III	71-90	Inpatient/Outpatient
Intermediate	IV	91-130	Inpatient
High	V	>130	Inpatient

In addition, the CURB-65 system is a very practical, memorable and valuable score model. The advantages of this score are its easy use and are designed to better assess disease severity rather than assessing pneumonia patients with mortality risk. The CURB-65 system is given in Table 3.

Table 3: Prediction Score of CURB-65

Characteristics	Score
Loss of consciousness	1
Blood Urea Nitrogren> 20 mg/dL	1
Respiratory Frequency ≥ 30 per minute	1
Blood preassure (systolic < 90 mmHg or diastolic ≤ 60 mmHg)	1
Age > 65 years old	1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Then the CURB-65 score is summed and categorized based on its severity as shown in Table 4.

Table 4: Risk Classes with CURB-65 System

Total Score	Risk Classes	Recommended Treatment
0	Low	Outpatient
1	Low	Outpatient
2	Intermediates	Inpatient / Outpatient
3	Intermediates to High	Inpatient / Outpatient
4 or 5	High	Inpatient/ ICU

6. PROCEDURE

The data used in this research is a secondary data obtained from medical record data of pneumonia patients at hospital '*T*' in Surabaya. The data consists of samples of pneumonia patients in 2015. The Pneumonia patient data is classified into 4 categories of dependent variables, 5 independent variables and 196 observations. The variables used in this study are shown in Table 5.

Table 5: Research Variables

Symbol	Variable	Scale
Y	Risk Class: I/II/III/IV/V	Ordinal
X_1	Age (years)	Rasio
X_2	Systolic (mmHg)	Rasio
X_3	Diastolic (mmHg)	Rasio
X_4	Respiratory Frequency per minute	Rasio
X_5	Blood Urea Nitrogen (mg/dL)	Rasio

The steps of the analysis are described below.

1. Perform descriptive statistics and data exploration.
2. Partitioning data into training data and testing data with a ratio of 90:10 and 80:20
3. Selecting variables in discriminant analysis using forward selection, backward elimination, stepwise method, and genetic algorithm. The steps in variable selection using the genetic algorithm as follows:
   a. Representing the independent variable into the chromosome and determines the initialization value.
   b. Evaluating chromosomes based on fitness values, namely the value of misclassification.
   c. Making a selection process with roulette wheel selection (RWS).
   d. Crossovers with probability of crossing (*Pc)* is 0.8
   e. Perform mutation processes with mutation probabilities (*Pm)* is 0.1 and
   f. Carry out the elitism process.
   g. Change the old population with a new generation.
   h. Repeating the process from step d until convergent fitness value is get
4. Perform discriminant analysis :
   a. Test the assumption of multivariate normal distribution, the homogeneity of variance-covariance matrix and the groups mean difference for the entire independent variable.
   b. Detect multicollinearity among independent variables.
   c. Estimating the coefficients of the discriminant function parameters.
   d. Calculate the GM and AUC from testing dataset
5. Estimate parameter using genetic algorithms with similar steps such as in variable selection, the difference is in the chromosome representation.
6. Analyzing data using Naïve Bayes
   a. Calculate the probability value of each parameter in each category.
   b. Determine the final probability of all parameters for each category.
   c. Determine the category group based on the highest probability value.
   d. Calculate the GM and AUC for testing dataset
   e. Compare the value of classification performance criterions from the results of analysis using discriminant analysis methods, hybrid discriminant analysis-genetic algorithms and Naive Bayes.
7. Draw a conclusion

7. RESULTS AND DISCUSSION

7.1 Characteristics of Pneumonia Patients at Hospital 'T' in Surabaya

Pneumonia patient's medical record data consist of 5 categories on dependent variable, but there is no patient included in class 1 category so the number of categories is 4 with total 96 observations. Based on Figure 1, it shows that the proportion of each category is imbalanced.

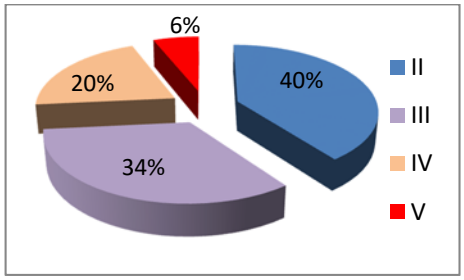


Figure 1: Proportion of Each Class of Pneumonia Risk

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

7.2 Discriminant Analysis Results

Before performing discriminant analysis, the data is firstly partitioned into training data and testing data with proportions of 90:10 and 80: 20. Furthermore, the discriminant analysis assumptions testing and multicollinearity detection are conducted. The results are as shown in Tabel 6.

Table 6: The Result of Discriminant Assumption Testing (90:10 split)

Assumption	Test Statistics		*pvalue*
Multivariate Normal	*Mardia Skweness*	246.509	0.000
	Mardia Kurtosis	4.621	0.000
Homogeneity of Variance and Covariance Matrices	*Box's M*	41.978	0.003

The table showed that both assumptions of normal multivariate distribution and homogeneity variance-covariance matrices are violated. To overcome this, the data need to be transformed using Box Cox transformation as in Table 7.

Table 7: Lambda Value of Box-Cox Transformation (90:10 split)

Variable	Lambda value
X_1	1.00
X_2	-0.50
X_3	0.50
X_4	-0.50
X_5	0.00

Furthermore, multivariate normality is tested for transformed data, and also the homogeneity variance-covariance assumption as shown in Table 8. After suitable transformation, both assumptions of normal multivariate distribution and homogeneity variance-covariance are satisfied.

Table 8: The Result of Assumption Testing of Transformed Data (90:10 split)

Assumption	Test Statistics		*pvalue*
Multivariate Normal	*Mardia Skweness*	46.959	0.085
	Mardia Kurtosis	0.5809	0.561
Homogeneity of Variance and Covariance Matrices	*Box's M*	60.319	0.176

Furthermore, the Equality of Group Means test is performed for each independent variable as summarized in Table 9. Table 9 shows that the hypothesis of equality of group means is rejected for X_2 and X_4.

Table 9: Equality of Group Means Test Result from Transformed Data (90:10 split)

Variable	Wilks' Lambda	*pvalue*
X_1	0.623	0.000
X_2	0.979	0.309
X_3	0.982	0.380
X_4	0.927	0.004
X_5	0.736	0.000

Next, multicollinearity detection is performed.

Table 10: Correlation among independent variables (90:10 split)

	X_1	X_2	X_3	X_4
X_2	0.259 0.001*			
X_3	0.064 0.403*	0.710 0.000*		
X_4	-0.121 0.112*	-0.161 0.033*	-0.045 0.558*	
X_5	0.307 0.000*	-0.049 0.524*	-0.031 0.679*	0.079 0.300*

* p_{value}

Based on Table 10, it can be seen that the correlation among several independent variables are significant. A high correlation value indicates a case of multicollinearity. Furthermore, variable selection is carried out using forward selection, backward elimination, stepwise method and genetic algorithm. Variable selection using Genetic Algorithm is initialized by coding the independent variable with 1 for variable included in the model and 0 for variables not included in the model. Figure 2 illustrates the variable selection process using genetic algorithm, where $X_1, X_4$and X_5 are used in the modeling.

1	0	0	1	1
X_1	X_2	X_3	X_4	X_5

Figure 2: Example of Chromosome Representation in Variable Selection

After representation step, then generate 100 chromosomes (solution) for the initial population and calculate the fitness value, namely the value of classification accuracy as shown in Table 11.

Table 11: Ilustration of Initial Chromosome and its Fitness Value

Chromosome #	Chromosome					*Fitness Value*
	X_1	X_2	X_3	X_4	X_5	
1	1	1	1	1	1	0.412
2	1	0	1	1	1	0.263
⋮	⋮	⋮	⋮	⋮	⋮	⋮
100	1	1	1	1	1	0.022

Journal of Theoretical and Applied Information Technology
31st March 2019. Vol.97. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The next stage is selecting the parent chromosome by using the Roulette Wheel method so that 100 parent chromosomes are obtained. After that, Crossover is performed with the probability of crossover is 0.8. An illustration of a single point crossover is given in Figure 3.

Parent	X_1	X_2	X_3	X_4	X_5	*rand*
5	1	1	1	1	1	0.56
6	1	0	0	1	0	0.72
Offspring	Crossover ($rn = 2$)					
5	1	1	0	1	0	
6	1	0	1	1	1	

Figure 3: Illustration of Single Point Crossover

Figure 3 shows the crossover process occurring on parent chromosomes no 5 and 6 with random numbers 2. Then the mutation process is carried out using uniform mutations with the mutation probability of 0.1. Illustration of the mutation process using the uniform mutation method is shown in Figure 4.

Chromosome	X_1	X_2	X_3	X_4	X_5
	0.01	0.65	0.07	0.82	0.03
1	1	1	1	0	0
Mutation Process					
1	0	1	0	0	1

Figure 4: Illustration of Uniform Mutation in Variable Selection

Figure 4 illustrates the process of mutation in chromosome 1. Genes mutated are gene 1, gene 3 and gene 5 where 1 is flipped into 0 and vice versa.The next step is elitism to maintain the best chromosomes in the population. The genetic algorithm variable selection process is continued until it gets a convergent fitness value. After the genetic algorithm process is complete, the results of the selected variables are obtained as shown in Table 12.

Table 12: Final Population from Genetic Algorithm for Variable Selection

Chromosome #	Chromosome					*Fitness Value*
	X_1	X_2	X_3	X_4	X_5	
1	1	0	0	1	1	0.3499
2	1	0	0	1	1	0.3499
⋮	⋮	⋮	⋮	⋮	⋮	⋮
100	1	0	1	1	1	0.6135

The comparison results of the four selection methods are presented in Table 13. The classification performances for testing dataset also summarized.

Table 13. The Performance Comparison of Variable Selection Methods in Discriminat Analysis (90:10 split)

Selection Method	Data	Selected Variables	GM	AUC
All variables used	Data	X_1, X_2, X_3, X_4, X_5	0.4478	0.4821
	Transformed data	X_1, X_2, X_3, X_4, X_5	0.4597	0.4866
Forward Selection	Data	X_1, X_4, X_5	0.5325	0.5444
	Transformed data	X_1, X_2, X_4, X_5	0.4811	0.5178
Backward Elimination	Data	X_1, X_4, X_5	0.5325	0.5444
	Transformed data	X_1, X_2, X_4, X_5	0.4811	0.5178
Stepwise Method	Data	X_1, X_4, X_5	0.5325	0.5444
	Transformed data	$\boldsymbol{X_1, X_4, X_5}$	**0.6501**	**0.6741**
Genetic Algorithm	Data	X_1, X_4, X_5	0.5325	0.5444
	Transformed data	$\boldsymbol{X_1, X_4, X_5}$	**0.6501**	**0.6741**

From Table 13, we can see that variables selected from forward selection method, backward elimination, stepwise method and genetic algorithms are different, both from data that has met the assumptions and those that violate the assumptions. In addition, it was found that the highest AUC (0.6741) and GM (0.6501) are obtained from Stepwise method and Genetic Algorithm, with the selected variables are X_1, X_4 and X_5.

After obtaining the selected variables, then genetic algorithm is used to estimate the parameters of the discriminant function that can minimize misclassification. The steps taken to optimize parameter estimation are the same as genetic algorithms for variable selection, the difference is in the chromosome representation in initialization step.

The first step of genetic algorithm for parameter estimation is to initiate chromosomes with the total of 100 where each chromosome consists of 15 genes. One of the chromosomes is a solution (parameter estimate) obtained from discriminant function analysis. Figure 5 shows one example of chromosome representation for linear discriminant function analysis.

-11.3	0.7	⋯	0.9	..	..	16.6	-0.01	⋯	0.64
α	W_1	⋯	W_p			α	W_1	⋯	W_p
$\hat{y}_1$						$\hat{y}_{100}$			

Figure 5: Illustration of Chromosome Representation for Parameter Estimation

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The classification results for data partition 90:10 using four selection methods followed by genetic algorithm for parameter estimation are summarized in Table 14, both for data where the assumptions are not met and the second for transformed data where the assumptions are met.

Table 14: Classification Performance of Hybrid DA-GA for Data Partition 90:10

Selection Method	Data		Transformed Data	
	GM	AUC	GM	AUC
No Selection	0.7071	0.7316	0.7194	0.7321
Forward Selection	0.6049	0.6161	0.6874	0.7098
Backward Elimination	0.6049	0.6161	0.6874	0.7098
Stepwise Method	0.6049	0.6161	**0.7530**	**0.7321**
Genetic Algorithm	0.6049	0.6161	**0.7530**	**0.7321**

It was shown that the best GM and AUC are obtained from Stepwise Method and Genetic Algorithm as selection methods followed by genetic algorithm for parameter estimation. The selected variables are X_1, X_4 and X_5 so the best Hybrid Genetic Algorithm- Discriminant Analysis model is as follows.

$$Z_1 = -15.08 + 0.0721X_1 + 2.6153X_4 + 0.8699X_5$$

$$Z_2 = -2.741 - 0.056X_1 + 0.514X_4 + 0.9877X_5$$

$$Z_3 = 0.733 + 0.0389X_1 + 4.0537X_4 + 0.7328X_5$$

The values of the discriminant function are the contribution values of each variable to classification. The coefficient with a positive value means the chance of an observation entering a class is increasing, while the coefficient with a negative value means the chance for an observation to enter a class is decreasing.

The same analysis is performed on the Data Partition 80:20. The assumption testing indicates that multivariate normal assumptions and homogeneous variance-covariance are violated so that transformation needs to be done. The Box-cox transformation value for data partition 80:20 are similar to transformation in data partition 90:10. After the assumptions are met then variable selection is done. The results of the variable selection from the four methods are summarized in Table 15.

Table 15: The Performance Comparison of Variable Selection Methods in Discriminant Analysis (80:20 split)

Selection Method	Data	Selected Variables	GM	AUC
All variables used	Data	X_1, X_2, X_3, X_4, X_5	0.4331	0.5468
	Transformed data	X_1, X_2, X_3, X_4, X_5	0.5150	0.5781
Forward Selection	Data	X_1, X_2, X_4, X_5	0.4331	0.5468
	Transformed data	X_1, X_2, X_4, X_5	0.5150	0.5781
Backward Elimination	Data	X_1, X_2, X_4, X_5	0.4331	0.5468
	Transformed data	X_1, X_2, X_4, X_5	0.5150	0.5781
Stepwise Method	Data	X_1, X_4, X_5	0.4746	0.6093
	Transformed data	X_1, X_4, X_5	0.5837	**0.6584**
Genetic Algorithm	Data	X_1, X_2, X_5	**0.6050**	0.6383
	Transformed data	X_1, X_4, X_5	0.5837	**0.6584**

Table 15 shows that different selection methods can yield on different variables. Genetic Algorithm performs best in classifying patients risk based on GM. Based on AUC, Stepwise and Genetic Algorithm methods perform best. Table 16 presents the results of the classification performance from the four methods for Data Partition 80:20 followed by the step of optimizing the parameter estimation using Genetic Algorithm.

Table 16: Classification Performance of Hybrid DA-GA for Data Partition 80:20

Selection Method	Data		Transformed Data	
	GM	AUC	GM	AUC
No Selection	0.7339	0.7500	0.6804	0.6853
Forward Selection	**0.7477**	**0.7679**	**0.7477**	0.7210
Backward Elimination	**0.7477**	**0.7679**	**0.7477**	0.7210
Stepwise Method	0.7339	0.7522	0.7031	**0.7321**
Genetic Algorithm	0.6839	0.7187	0.7031	**0.7321**

It is shown in Table 16 that for data partition 80:20, forward selection and backward elimination hybridized with GA in parameter optimization can produce highest GM and AUC that are 74.77% and 76.79% respectively. The selected variables are X_1, X_2, X_4 and X_5. The best hybrid Discriminant Analysis–Genetic Algorithm model is

$$Z_1 = 0.14 + 0.658X_1 + 0.099X_2 + 0.059X_4 + 0.049X_5$$

$$Z_2 = 0.348 + 0.861X_1 + 0.135X_2 + 0.172X_4 + 0.379X_5$$

$$Z_3 = 0.156 + 0.681X_1 + 0.026X_2 + 0.645X_4 + 0.018X_5$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

7.3 Naïve Bayes Analysis

The initial step in Naïve Bayes analysis is to calculate the prior probability value. In the partition data 90:10, the training data consists of 70 observations in class II, 59 observations in class III, 36 observations in class IV and 10 observations in class V. Prior probability values for Data Partition 90:10 are presented in Table 17.

Table 17: Prior Probability for Data Partition 90:10

Class	Probability
II	0.4000
III	0.3371
IV	0.2058
V	0.0571

Next, calculate the mean and standard deviation of each variable in each category. After that, the partial probability values are calculated using the Gauss density function as in equation (12) and then calculate the posterior probability values, which are then used to determine the prediction results on the **testing data**. Predictions results are then compared to the actual data.

Table 18: Classification Performance of Naïve Bayes Classifier (90:10 split)

		Predicted				
		π_1	π_2	π_3	π_4	n_l
Actual	π_1	6	1	1	0	8
	π_2	2	4	0	1	7
	π_3	1	2	1	0	4
	π_4	0	1	0	1	2

From the results in Table 18, then calculate the miss classification value using expanding geometric-mean (GM) and Area Under Curve (AUC) for both data and transformed data. Similar steps are applied for transformation data and Data partition 80:20. Based on Table 19, it can be seen that by using GM, it is possible to obtain 0% classification accuracy of testing data. These occurs because there are one or more classes in predicted data that have no members.

In addition, the highest GM values for Data Partition 90:10 and 80:20 are obtained from data where the assumptions are met and the selection methods used are Stepwise methods and Genetic algorithm, which are 50.87% and 64.6 % respectively. The same conclusion is obtained from AUC comparison as presented in Table 20.

Table 19: Classification Performance of Naïve Bayes based on Geometric Mean (GM)

Selection Method	90:10		80:20	
	Data	Tranformed Data	Data	Transformed Data
No Variable Selection	0.4811	0	0.5087	0
Forward Selection	0	0	0.5287	0
Backward Elimination	0	0	0.5287	0
Stepsiwe Method	0	**0.5087**	0.5087	**0.6460**
Genetic Algorithm	0	**0.5087**	0.4370	**0.6460**

Table 20: Classification Performance of Naïve Bayes based on Area Under Curve (AUC)

Selection Method	90:10		80:20	
	Data	Tranformed Data	Data	Transformed Data
No Selection	0.5178	0.0952	0.5339	0.0975
Forward Selection	0.3404	0.0952	0.5284	0.0975
Backward Elimination	0.3404	0.0952	0.5284	0.0975
Stepsiwe Method	0.3404	**0.5339**	0.5339	**0.6069**
Genetic Algorithm	0.3404	**0.5339**	0.4841	**0.6069**

7.4 Comparison between Hybrid Genetic Algorithm-Discriminant Analysis and Naïve Bayes

After obtaining the value of GM and AUC from every method, a comparison is carried out.

Table 21: Overall Comparison

Method	Partition	Data	Selected Variables	Classification GM	Classification AUC
Discriminant Analysis	90:10	Data	X_1, X_4, X_5	0.5325	0.5444
		Transf	X_1, X_4, X_5	0.6501	0.6741
	80:20	Data	X_1, X_2, X_5	0.6050	0.6383
		Transf	X_1, X_4, X_5	0.6584	0.6584
Hybrid DA-GA	90:10	Data	X_1, X_2, X_3, X_4, X_5	0.7071	0.7316
		Transf	X_1, X_4, X_5	**0.7530**	0.7321
	80:20	Data	X_1, X_2, X_4, X_5	0.7477	**0.7679**
		Transf	X_1, X_2, X_4, X_5	0.7477	0.7210
Naïve Bayes	90:10	Data	$X_1, X_2, X_3 X_4, X_5$	0.4811	0.5178
		Transf	X_1, X_4, X_5	0.5087	0.5339
	80:20	Data	X_1, X_4, X_5	0.5287	0.5339
		Transf	X_1, X_4, X_5	0.6460	0.6069

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 20 shows that highest values of GM and AUC are 74.77% and 76.79% respectively, obtained from hybrid genetic algorithm-discriminant analysis. The selected variables are X_1, X_2, X_4 and X_5.

8. CONCLUSION

It can be concluded that:

1. From the data, there were no patients that classified into class I, 40% patients are in class II, 34% in class III, 20% in class IV and 6% are in class V.
2. The variable that can be used to discriminate patients risk in Data Partition 90:10 and 80:20 are different. In data partition 90:10 and 80:20 when the assumptions are met, the variables that selected from *forward selection* and *backward elimination* are X_1, X_2, X_4, and X_5, while *stepwise method* and Genetic Algorithm select X_1, X_4 and X_5. The same selected variables are produced from *forward selection, backward elimination, stepwise method* and Genetic Algorithm for Data Partition 90:10 when the assumptions are violated. For Data Partition 80:20 when the assumptions are violated, forward selection and backward elimination gives the same result that is X_1, X_2, $X_4$and X_5 are selected; while stepwise method selects X_1, X_4, and X_5 and Genetic Algorithm selects X_1, X_2, and X_5.
3. Optimization of parameter estimation using genetic algorithm in discriminant analysis gives better classification result than discriminant analysis, both for data when assumption is met or not.
4. The highest GM and AUC are 74,77% and 76,79%, obtained from Hybrid Genetic Algorithm-Discriminant Analysis with the selected variables are $X_1, X_2, X_4$and X_5.
5. Multi-class Classification performance using GM and AUC give the same result that is the Hybrid Genetic Algorithm-Discriminant Analysis produces better classification performance than Naïve Bayes.

Further study can be carried out using simulation data with various characteristics of the data, so that it can be known the performance of the genetic algorithm for variable selection and parameter estimation based on simulated data characteristics.

ACKNOWLEDGEMENT

The authors gratefully thank the Ministry of Research, Technology and Higher Education, Republic of Indonesia and Institut Teknologi Sepuluh Nopember Surabaya Indonesia for the financial support under “Penelitian Fundamental”.

REFERENCES

[1] Depkes R.I. (2002). *Pedoman Pemberantasan Penyakit Infeksi Saluran Pernafasan Akut untuk Penanggulangan Pneumonia pada Balita dalam Pelita VI.* Jakarta: Dirjen PPM & PLP.

[2] Badan Penelitian dan Pengembangan Kesehatan. (2013). *Riset Kesehatan Dasar : RISKESDAS 2013.* Jakarta: Kementrian Kesehatan Republik Indonesia.

[3] Tierney, L. M., McPhee, S. J., & Papadakis, M. A. (2002). *Diagnosis dan Terapi Kedokteran (Penyakit Dalam).* (G. Abdul, Y. T. Sofyatul, Erlina, & Isnatin, Trans.) Jakarta: Salemba Medika.

[4] Perhimpunan Dokter Paru Indonesia. (2003). *Pneumonia Kominiti: Pedoman Diagnosis & Penatalaksanaan di Indonesia.* Jakarta: Perhimpunan Dokter Paru Indonesia.

[5] Gradianta, R. D., & Irhamah. (2014). *Klasifikasi Pasien Penderita Diabetes Melitus Tipe Dua Menggunakan Metode Analisis Diskriminan Hybrid Algoritma Genetika.* Jurnal Sains dan Seni ITS, Vol 3 no 2. Surabaya: Institut Teknologi Sepuluh Nopember.

[6] Kurnianto, I. P., & Irhamah. (2016). Forest Type Classification Based Spectral Characteristics using Hybrid Discriminant Analysis and Genetic Algorthm. *Proceedings of the 6th Annual Basic Science International Conference* (pp. 401-405). Malang: Brawijaya University.

[7] Islam, S., & Islam, R. (2011). Modeling Spammer Behavior: Artificial Neural Network vs. Naïve Bayesian Classifier. *Artificial Neural Networks*, Chi Leung Patrick Hui. IntechOpen. DOI: 10.5772/16002.

[8] Nayak, A., & Natarajan, D. S. (2016). Comparative Study of Naive Bayes, Support Vector Machine and Random Forest Classifiers in Sentiment Analysis of Twitter Feeds. *International Journal of Advance Studies in Computer Science and Engineering (IJASCSE), 5*(1), 16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[9] Johnson, R. A., & Winchern, D. W. (2007). *Applied Multivariate Statistical Analysis (6th ed.).* New Jersey: Pearson Prentice Hall

[10] Hair, J., Black, W., Barbin, B. J., & Anderson, R. (2010). *Multivariate Data Analysis 7th Edition.* New Jersey: Pearson Prentice Hall.

[11] Mardia, K.V. (1970) *Measure of Multivariate Skewness and Kurtosis with Applications.* Biometrika, 57(3):519-530

[12] Dillon, W. R., & Metthew, G. (1984). *Multivariate Analysis Methods and Application.* United States of America: John Wiley & Sons, Inc.

[13] Gen, M., & Cheng, R. (1997). *Genetic Algorithms and Engineerung Design.* New York: John Wiley & Sons, Inc.

[14] Irhamah & Ismail, Z. (2009). A Breeder Genetic Algorithm for Vehicle Routing Problem with Stochastic Demands. *Journal of Applied Sciences Research*, 5(11): 1998-2005.

[15] Sivanandam, S. N., & Deepa, S. N. (2008). *Introduction to Genetic Algorithms.* New York: Springer-Verlag Berlin Heidelberg.

[16] Wu, X., & Kumar, V. (2009). *The Top Ten Algorthms in Data Mining.* New York: CRC Press.

[17] Sudoyo, A. W., Setiyohadi, B., Alwi, I., Simadibrata, M., &Setiati , S. (2009). *Buku Ajar Imu Penyakit Dalam Jilid II Edisi V.* Jakarta: Interna Publishing.

Source details

Journal of Theoretical and Applied Information Technology

Scopus coverage years: 2005, from 2008 to Present
Publisher: Little Lion Scientific
ISSN: 1992-8645 E-ISSN: 1817-3195
Subject area: Computer Science: General Computer Science | Mathematics: Theoretical Computer Science
Source type: Journal

CiteScore 2020
1.3

SJR 2020
0.153

SNIP 2020
0.410

View all documents > | Set document alert | Save to source list | Source Homepage

CiteScore | CiteScore rank & trend | Scopus content coverage

Improved CiteScore methodology

CiteScore 2020 counts the citations received in 2017-2020 to articles, reviews, conference papers, book chapters and data papers published in 2017-2020, and divides this by the number of publications published in 2017-2020. Learn more >

CiteScore 2020

$$1.3 = \frac{3.003 \text{ Citations 2017 - 2020}}{2.227 \text{ Documents 2017 - 2020}}$$

Calculated on 05 May, 2021

CiteScoreTracker 2021

$$1.2 = \frac{2.590 \text{ Citations to date}}{2.102 \text{ Documents to date}}$$

Last updated on 06 March, 2022 • Updated monthly

CiteScore rank 2020

Category	Rank	Percentile
Computer Science General Computer Science	#144/226	36th
Mathematics Theoretical Computer Science	#101/120	16th

View CiteScore methodology > | CiteScore FAQ > | Add CiteScore to your site

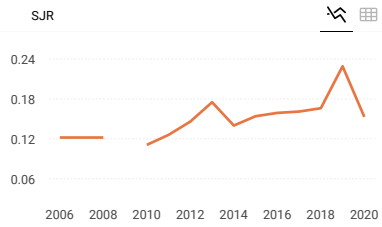
Journal of Theoretical and Applied Information Technology

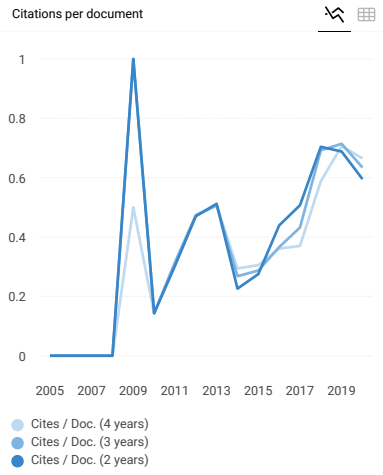
COUNTRY	SUBJECT AREA AND CATEGORY	PUBLISHER	H-INDEX
Pakistan Universities and research institutions in Pakistan	Computer Science Computer Science (miscellaneous) Mathematics Theoretical Computer Science	Little Lion Scientific	29

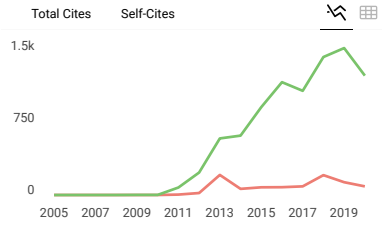
PUBLICATION TYPE	ISSN	COVERAGE
Journals	18173195, 19928645	2005, 2008-2020

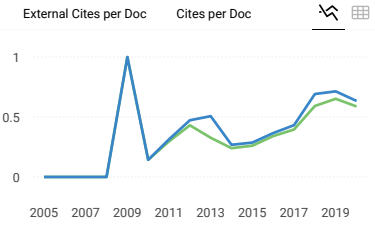
SCOPE

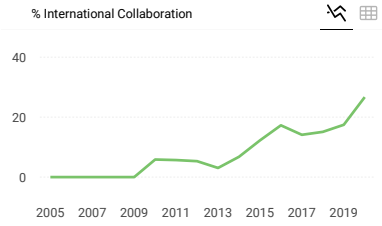
Information not localized

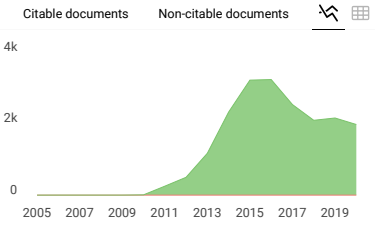


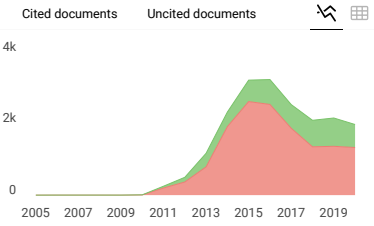












← Show this widget in your own website

Just copy the code below and paste within your html code:

```
<a href="https://www.scimag
```

SCImago Graphica

Explore, visually communicate and make sense of data with our **new free tool**.

Get it

Metrics based on Scopus® data as of April 2021

N **nashat** 4 weeks ago

it is a very bad journal
There is no DOI to the paper
the paper does not appear in google scholar
you cannot make citations to the paper
it is predatory