ISSN 2394 - 9554

# International Journal of Research in Science and Technology

Volume 4, Issue 2: April - June 2017



Indian Academicians and Researchers Association www.iaraedu.com

Volume 4, Issue 2 : April – June 2017

Editor

# Dr. Nurul Fadly Habidin

Faculty of Management and Economics Universiti Pendidikan Sultan Idris Malaysia

# Members of Editorial Advisory Board

**Mr. Nakibur Rahman** Ex. General Manager ( Project ) Bongaigoan Refinery, IOC Ltd, Assam

**Dr. Alka Agarwal** Director, Mewar Institute of Management, Ghaziabad

**Prof. (Dr.) Sudhansu Ranjan Mohapatra** Dean, Faculty of Law, Sambalpur University, Sambalpur

**Dr. P. Malyadri** Principal, Government Degree College, Hyderabad

**Prof.(Dr.) Shareef Hoque** Professor, North South University, Bangladesh

**Prof.(Dr.) Michael J. Riordan** Professor, Sanda University, Jiashan, China

**Prof.(Dr.) James Steve** Professor, Fresno Pacific University, California, USA

**Prof.(Dr.) Chris Wilson** Professor, Curtin University, Singapore

**Prof. (Dr.) Amer A. Taqa** Professor, DBS Department, University of Mosul, Iraq

**Dr. Pranjal Sharma** Professor, Himalaya University, Arunachal Pradesh

**Dr. Neetu Singh** HOD, Department of Biotechnology, Mewar Institute, Vasundhara, Ghaziabad **Dr. Mukesh Saxena** Pro Vice Chancellor, University of Technology and Management, Shillong

**Dr. Archana A. Ghatule** Director, SKN Sinhgad Business School, Pandharpur

**Prof. (Dr.) Monoj Kumar Chowdhury** Professor, Department of Business Administration, Guahati University, Guwahati

**Prof. (Dr.) Baljeet Singh Hothi** Professor, Gitarattan International Business School, Delhi

**Prof. (Dr.) Badiuddin Ahmed** Professor & Head, Department of Commerce, Maulana Azad Nationl Urdu University, Hyderabad

**Dr. Anindita Sharma** Dean & Associate Professor, Jaipuria School of Business, Indirapuram, Ghaziabad

**Prof. (Dr.) Jose Vargas Hernandez** Research Professor, University of Guadalajara,Jalisco, México

**Prof. (Dr.) P. Madhu Sudana Rao** Professor, Mekelle University, Mekelle, Ethiopia

**Prof. (Dr.) Himanshu Pandey** Professor, Department of Mathematics and Statistics Gorakhpur University, Gorakhpur

**Prof. (Dr.) Agbo Johnson Madaki** Faculty, Faculty of Law, Catholic University of Eastern Africa, Nairobi, Kenya

**Prof. (Dr.) D. Durga Bhavani** Professor, CVR College of Engineering, Hyderabad, Telangana **Prof. (Dr.) Shashi Singhal** Professor, Amity University, Jaipur

**Prof. (Dr.) Alireza Heidari** Professor, Faculty of Chemistry, California South University, California, USA

Prof. (Dr.) A. MahadevanProfessorS. G. School of Business Management, Salem

**Prof. (Dr.) Hemant Sharma** Professor, Amity University, Haryana

**Dr. C. Shalini Kumar** Principal, Vidhya Sagar Women's College, Chengalpet

**Prof. (Dr.) Badar Alam Iqbal** Adjunct Professor, Monarch University, Switzerland

**Prof.(Dr.) D. Madan Mohan** Professor, Indur PG College of MBA, Bodhan, Nizamabad

**Dr. Sandeep Kumar Sahratia** Professor Sreyas Institute of Engineering & Technology

**Dr. S. Balamurugan** Director - Research & Development, Mindnotix Technologies, Coimbatore

Dr. Dhananjay Prabhakar Awasarikar

Associate Professor, Suryadutta Institute, Pune

**Dr. Mohammad Younis** Associate Professor, King Abdullah University, Saudi Arabia

**Dr. Kavita Gidwani** Associate Professor, Chanakya Technical Campus, Jaipur

**Dr. Vijit Chaturvedi** Associate Professor, Amity University, Noida

**Dr. Marwan Mustafa Shammot** Associate Professor, King Saud University, Saudi Arabia **Prof. (Dr.) Aradhna Yadav** Professor, Krupanidhi School of Management, Bengaluru

**Prof.(Dr.) Robert Allen** Professor Carnegie Mellon University, Australia

**Prof. (Dr.) S. Nallusamy** Professor & Dean, Dr. M.G.R. Educational & Research Institute,Chennai

**Prof. (Dr.) Ravi Kumar Bommisetti** Professor, Amrita Sai Institute of Science & Technology, Paritala

**Dr. Syed Mehartaj Begum** Professor, Hamdard University, New Delhi

**Dr. Darshana Narayanan** Head of Research, Pymetrics, New York, USA

**Dr. Rosemary Ekechukwu** Associate Dean, University of Port Harcourt, Nigeria

**Dr. P.V. Praveen Sundar** Director, Shanmuga Industries Arts and Science College

**Dr. Manoj P. K.** Associate Professor, Cochin University of Science and Technology

**Dr. Indu Santosh** Associate Professor, Dr. C. V.Raman University, Chhattisgath

**Dr. Pranjal Sharma** Associate Professor, Department of Management Mile Stone Institute of Higher Management, Ghaziabad

**Dr. Lalata K Pani** Reader, Bhadrak Autonomous College, Bhadrak, Odisha

**Dr. Pradeepta Kishore Sahoo** Associate Professor, B.S.A, Institute of Law, Faridabad

**Dr. R. Navaneeth Krishnan** Associate Professor, Bharathiyan College of Engg & Tech, Puducherry **Dr. Mahendra Daiya** Associate Professor, JIET Group of Institutions, Jodhpur

**Dr. Parbin Sultana** Associate Professor, University of Science & Technology Meghalaya

**Dr. Kalpesh T. Patel** Principal (In-charge) Shree G. N. Patel Commerce College, Nanikadi

**Dr. Juhab Hussain** Assistant Professor, King Abdulaziz University, Saudi Arabia

**Dr. V. Tulasi Das** Assistant Professor, Acharya Nagarjuna University, Guntur, A.P.

**Dr. Urmila Yadav** Assistant Professor, Sharda University, Greater Noida

**Dr. M. Kanagarathinam** Head, Department of Commerce Nehru Arts and Science College, Coimbatore

**Dr. V. Ananthaswamy** Assistant Professor The Madura College (Autonomous), Madurai

**Dr. S. R. Boselin Prabhu** Assistant Professor, SVS College of Engineering, Coimbatore

**Dr. A. Anbu** Assistant Professor, Achariya College of Education, Puducherry **Dr. G. Valarmathi** Associate Professor, Vidhya Sagar Women's College, Chengalpet

**Dr. M. I. Qadir** Assistant Professor, Bahauddin Zakariya University, Pakistan

**Dr. Brijesh H. Joshi** Principal (In-charge) B. L. Parikh College of BBA, Palanpur

**Dr. Namita Dixit** Associate Professor, ITS Institute of Management, Ghaziabad

**Dr. Nidhi Agrawal** Assistant Professor, Institute of Technology & Science, Ghaziabad

**Dr. Ashutosh Pandey** Assistant Professor, Lovely Professional University, Punjab

**Dr. Subha Ganguly** Scientist (Food Microbiology) West Bengal University of A. & F Sciences, Kolkata

**Dr. R. Suresh** Assistant Professor, Department of Management Mahatma Gandhi University

**Dr. V. Subba Reddy** Assistant Professor, RGM Group of Institutions, Kadapa

**Dr. R. Jayanthi** Assistant Professor, Vidhya Sagar Women's College, Chengalpattu

Copyright @ 2017 Indian Academicians and Researchers Association, Guwahati All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, or stored in any retrieval system of any nature without prior written permission. Application for permission for other use of copyright material including permission to reproduce extracts in other published works shall be made to the publishers. Full acknowledgment of author, publishers and source must be given.

The views expressed in the articles are those of the contributors and not necessarily of the Editorial Board or the IARA. Although every care has been taken to avoid errors or omissions, this publication is being published on the condition and understanding that information given in this journal is merely for reference and must not be taken as having authority of or binding in any way on the authors, editors and publishers, who do not owe any responsibility for any damage or loss to any person, for the result of any action taken on the basis of this work. All disputes are subject to Guwahati jurisdiction only.

Volume 4, Issue 2 : April – June 2017

# CONTENTS

# **Research Papers**

QSAR: MODERN TOOL OF DRUG DESIGN	1 - 6
S. S. Chourasia	
INFORMATION TECHNOLOGY IN DATA SCIENCE	7 – 10

Sonia S. Joshi

### **QSAR: MODERN TOOL OF DRUG DESIGN**

S. S. Chourasia

Department of Chemistry, M. B. Patel College, Deori

### ABSTRACT

The application of quantitative structure–activity relationships (QSARs) has significantly revolutionised the whole process of drug discovery. QSAR models eases out the calculation of physicochemical properties (e.g., lipophobicity), the prediction of biological activity (or toxicity) of the potential drug molecule. It also predicts very well the absorption, distribution, metabolism, and excretion (ADME). QSAR has the most significant application in the preclinical stages of the drug discovery as this stage involves tedious work with high cost of experimentation. QSAR narrow downs and filter large chemical databases, and select suitable drug candidates. This article presents some current challenges and applications for the discovery and optimization of drug candidates based on QSAR methodology.

Keywords: QSAR, Drug design, New chemical entities (NCEs), ADME.

### **INTRODUCTION**

Discovery and development of a new drug is an expensive and time consuming process. The therapeutic effects and hazards of the discovered drugs to health are assessed by *in-vivo* tests and by various experimental techniques. However, involvement of animal models raises ethical questions. Therefore, alternative methods are being developed to for drug testing. *In-silico* methods offers alternative path due to their lower cost. They offer significant contribution to the identification and development of effective drugs from new chemical entities (NCEs). The computational tools are principally used for

- Conformational analysis of molecular structure (e.g., molecular dynamics)
- Characterization of drug-target interactions (e.g., molecular docking); and
- Assessment and optimization of drug activity using quantitative structure-activity relationships (QSARs).

QSAR methods are useful evaluating physicochemical features, physico-chemical properties and biological effects which govern a biological response in drug design. As a result, QSAR has emerged as low-cost tool filtering out novel "hits" and for "lead" optimization which is important process of drug discovery and development. The foundation of modern QSAR was laid by Prof. Corwin Hansch and co-workers during the early 1960s<sup>[1]</sup>. Since then, various QSAR models have come into existence for understanding of biological and physicochemical properties of NCEs and their evaluation. This emergence of QSAR for design and development of new drugs is evident by the cascade of publications and QSAR-based software. The aim of this article is to signify the application of QSAR in drug design.

### PRECLINICAL PHASES: FROM NEW CHEMICAL ENTITIES (NCES) TO DRUG CANDIDATES

To ensure the therapeutic effect and the safety of the new chemical entities (NCEs), the benefits (therapeutic effects) and the risk (toxic effects) of the NCEs are evaluated, respectively, during the preclinical and the clinical phases of development<sup>[2,3]</sup>. For a particular disease state once the target has been validated, relevant data is gathered in the preclinical drug development in order to propose a drug candidate for clinical test. Using *insilico, in-vitro methods* and animal models, the pharmacological profile and the acute toxicity of the drug candidate are assessed during the preclinical stage. Figure 1 present the three-stage procedure allowing the selection of the most effective NCEs (the drug candidates).

### **Stage 1: Hit idetification**

This stage aims to identify "hit" compounds from diverse libraries (corporate, commercial, etc.) and/or by medical observations for a given target (receptor, enzyme, etc.). Various techniques like high-throughput screening (HTS) and *in-silico* evaluations are used to screen NCEs with suitable pharmacodynamic (PD) activity. The PD properties of a given molecular entity are defined as the physiological and biochemical effects of the entity on the body.

### **Stage 2: Lead identification**

This stage is a key milestone of the drug candidate discovery process. The pharmacokinetic (PK) properties govern the bioavailability of the NCEs and, therefore, the correct delivery of the drug to its target site. The PK properties are represented by the processes of absorption, distribution, metabolism, and elimination (ADME) under-gone by the NCEs in the organism. During this stage, "hit" molecules presenting good ADME and physicochemical properties are identified and taken further as lead compounds.

ISSN 2394 - 9954



Figure 1: Strategy for a validated pharmaceutical target. Drug candidates are selected following various stages of development: (1) the selection of "hit" compounds, (2) the optimization of "lead" compounds, and (3-5) the choice of the drug candidates

### **Stage 3: Lead Optimization**

This stage involves the evaluation of various properties of lead analogs in order to propose the drug candidates. Accordingly, lead's molecular scaffold is modified to generate lead analogs. Then the drug candidates are selected based on the chemical structures with the optimal potency, solubility, and ADME profile. QSAR and molecular docking have a major role to play in selecting strategies for lead finding and optimization. Once the preclinical phase is complete, the selected drug candidates can be subjected to phases I, II, and III of clinical development.

### DRUG DISCOVERY PROCESS: A TIME CONSUMING PROCESS

Simple rules of thumb, such as the "rule of 5" (also known as Lipinski's rules can be used to filter molecules which are likely to be only weakly bio available. However, experimental evaluation can be time consuming, expensive, or be subject to ethical barriers. In this context, QSAR methods analyses allow the identification of the structural and physicochemical features modulating the activities of compounds.

Thus QSAR method by passes the cumbersome process of laboratory work of optimizing the drug candidate directly gives the potential drug molecules for the preclinical analysis. Figure 2 represents the time involved in the drug discovery process and impact of QSAR in drug design.



Fig2: Impact of QSAR methodology in drug design and discovery process.

Volume 4, Issue 2 : April - June, 2017

### REASONS FOR FAILURE IN DRUG CANDIDATE DEVELOPMENT

The development of a typical drug may cost up to one billion US dollars and 10-15 years of time. Even though after applying stringent procedure and substantial financial investment of drug development tested in the preclinical phase, out of 5000 molecules, only one reaches the market<sup>[4]</sup>. Kennedy et al.<sup>[5]</sup> have studied the factors which cause failures during clinical assessments (Figure 3). Insufficient pharmacokinetic properties (39%), lack of efficacy (30%), toxicity in animals (11%), and adverse effects to man (10%) are the major reasons to explain the exit of molecules in pharmaceutical research. Poor ADME properties are major cause of failure during drug development<sup>[6,7]</sup>.



Figure 3: Reasons for the failure of the development of NCEs in clinical development<sup>[5]</sup>

To address ADME and potency during the early phase of drug discovery and development, numerous computational tools, have been  $proposed^{[8-10]}$ . These tools vary from very trivial "rules of thumb," e.g., Lipinski's rule of  $5^{[11]}$  to more complex and multivariate QSAR models, which includes the use of models based on neural networks<sup>[12,13]</sup>.

### QSAR PROCESS

### DEVELOPMENT OF TRAINING AND TEST SET

The first step in deriving a QSAR model is to gather and select the molecules with activity data to include in the training set and test set. The resulting information can be downloaded in different formats (e.g., SMILES, sdf file, txt file). Using the QSAR methods, Chourasia et al<sup>[14]</sup> designed and evaluated a series of 4-methyl-2-(p-substitutedphenyl) quinoline derivatives as potential antifungal agents. The general structure of the synthesized quinoline derivatives is shown in figure 4.



R= -H,-CH<sub>3</sub>,-OCH<sub>3</sub>,-OH,-NO<sub>2</sub>,-Cl,-Br,-COOH

### Fig 4: Structure of the synthesized test set quinoline derivatives

Various physicochemical descriptors like constitutional, molecular, steric, and electronic were calculated by the Datawarrier<sup>[15]</sup> Software .The descriptors used to develop the model are listed and described in the Table 2.

Table 2. Descriptors used for the training and test set						
S. No.	0. Descriptor used					
1	Total Molweight in g/mol; natural abundance	TMW				
2	cLogP; P; conc(octanol) /conc(water)	cLogP				
3	Total surface area(from polar and non-polar SAS Approximation)	TSA				
4	Druglikeness	Drug				
5	Lipophilic Ligand Efficiancy (LLE)	LLE				
6	Ligand Efficiency Lipophilic Price(LELP)	LELP				
7	Molecular shape Index	MSI				
8	Rotatable Bond Count	RBC				

### **REGRESSION ANALYSIS**

Using the calculated descriptors multiple linear regressions were performed using the software  $SPSS^{[16]}$  by stepwise method. The best model derived from the regression analysis was used to predict the biological activity of the synthesised compounds. No outliers have been determined and the equations were derived using the entire training data set (n=30). The summary of the model generated is given in the Table 3.

Table 3. Summary of the QSAR model generated by the regression analysis									
Model Summary									
Model	R	R	Adjusted R	Std. Error of	Durbin-Watson				
		Square	Square	the Estimate					
1	0.76 2	0.580	0.420	0.9830729	2.296				
Predictors: (Constant), Rotatable_Bonds, Total_Surface_Area, Druglikeness,									
Shape_Index, Total_Molweight, LLE, cLogP, LELP									

Dependent Variable: Reported activity

### VALIDATION TEST

The equations 1 and 2 are generated by the model. The equations, then, were used to calculate the unknown activities of the test set compounds.

### QSAR model for A. Niger

Biological activity (ZOI) = (18.354)+ (-0.007\* **TMW**)+ (4.359\***cLogP**)+(0.132\* **TSA**)+ (0.315\*Drug)+ (2.963\***MSI**)+ (-0.419\***RBC**) ------ eq. (1)

### **QSAR** model for C. Albicans

Biological activity (ZOI) = (18.354)+ (-0.007\* **TMW**)+ (4.359\***cLogP**)+( 0.132\***TSA**)+ (0.315\*Drug)+ (2.963\* **MSI**)+ (-0.419\***RBC**) ------ eq. (2)

Using the equations, the predicted antifungal activities of the synthesized test set quinoline compounds were calculated. This gave the predicted antifungal activity of the test set compound with unknown activity. Had conventional method of synthesis and determination of the antifungal activity being applied, it would cost money and time for the same pocess.

Volume 4, Issue 2 : April - June, 2017

### IN-VITRO ANTIMICROBIAL ACTIVITY

Once the QSAR predicted activity of the test set quinoline compounds are determined, the model has to be validated by comparing the predicted activities with the experimental ones. For this standard method were applied and the *in-vitro* antifungal activities of the synthesized test set quinoline compounds were determined. Figure 5 gives a comparative account of the predicted and experimental activities.



Fig. 5: Comparison of predicted and experimental antifungal activities of the synthesized test set quinoline compounds

A set of 8 quinoline compounds were developed and a comparative study of the predicted and experimental activities was made. It is clear from the graph that the *in-vitro* antifungal activities are close the predicted activities. This itself validates the generated QSAR model

### **RESULT AND DISCUSSION**

The predicted and experimental antifungal activity for the synthesized test set quinoline compounds have been found in good agreement. A low value of  $R^2$  suggest that the descriptor used for the model are significant and show little co-relation with one another. This validated the generated QSAR model and hence it can be used to predict the antifungal activities of the other quinoline derivatives with unknown biological activity.

### CONCLUSION

The generated QSAR model in this case offers an alternative quick and economic path for the pre-clinical phase i.e. target validation, lead identification and lead optimization in the drug discovery process.

### REFERENCES

- 1. Selassie C D (2003) 'History of quantitative structure-activity relationships. In: Abraham D J (ed) Burger's medicinal chemistry and drug discovery, 6th edn. Wiley, New York.
- 2. Drayer J I, Burns J P (1995), 'From discovery to market: the development of pharmaceuticals'. In:Wolff ME (ed) Burger's medicinal chemistry and drug discovery, 5th edn. Willey, New York.
- Lesko L J, Rowland M, Peck C C et al. (2000), 'Optimizing the science of drug development: Opportunities for better candidate selection and accelerated evaluation in humans'. Eur J PharmSci 10:9– 14.
- 4. DiMasi J A (1995), 'Success rates for new drugs entering clinical testing in the United States'. Clin Pharmacol Ther 58:1–14.
- 5. Kennedy T (1997), 'Managing the drug discovery/development interface. Drug Disc Today' 2: 436–444.
- 6. Wenlock M C, Austin R P, Barton P et al. (2003), 'A comparison of physiochemical property profiles of development and marketed oral drugs'. J Med Chem 46:1250–1256.
- 7. Lin J H, Lu A Y H (1997), 'Role of pharmacokinetics and metabolism in drug discovery and development'. Pharmacol Rev 49:403–449.
- 8. Cruciani G, Crivori P, Carrupt P A et al. (2000), 'Molecular fields in quantitative structure permeation relationships: the VolSurf approach'. J Mol Struct: THEOCHEM 503:17–30.
- 9. Lombardo F, Gifford E, Shalaeva M Y (2003), 'In silico ADME prediction: Data, models, facts and myths'. Mini Rev Med Chem 8:861–875.

Volume 4, Issue 2 : April - June, 2017

- 10. Hansch C, Leo A, Mekapati S B et al. (2004), 'QSAR and ADME'. BioorgMed Chem 12:3391–3400.
- 11. Lipinski C A, Lombardo F, Dominy B W et al. (1997), 'Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings'. Adv Drug Del Rev 23:3–25.
- 12. Yamashita F, Hashida M (2003), 'Mechanistic and empirical modeling of skin permeation of drugs'. Adv Drug Del Rev 55:1185–1199.
- 13. Wang Z, Yan A, Yuan Q et al. (2008), 'Explorations into modeling human oral bioavailability'. Eur J Med Chem 43:2442–2452.
- 14. Chourasia S. S, Rahangdale P. K., Inam F. Inam (2017), Quantitative Structure Activity Relationship studies of 4-methyl-2-(p-substitutedphenyl)quinoline derivatives as potential antifungal agents'. IJAEMS,Vol-3, Issue-5,607-612.
- 15. www.openmolecules.org
- 16. http://bioinformatikz.googlepages.com

Volume 4, Issue 2 : April - June, 2017

### INFORMATION TECHNOLOGY IN DATA SCIENCE

Sonia S. Joshi

Department of Statistics, P.E.Society's Modern College of Arts, Science and Commerce, Pune

### ABSTRACT

The most successful person is the one who has the best information and the ability to apply it effectively. This remark summarizes the business of information technologies in the context of generation, storage, processing and use of information. Today, cutting-edge technologies such as computers, software and artificial intelligence, fiber optics and networking have an immense impact on information technology. Among the many applications of information technology, three of particular importances are traditional telephony, mobile cellular telephony, and data processing and communication. Advancement in information technology, affect many industries and society as a whole.

This research article discusses the effect of advanced computer science and information technology in the context of data science.

Keywords: E-Commerce, Data Science, Machine Learning, Data Analytics, SAS, SQL, Python, Exploratory Data Analysis, Confirmatory Data Analysis.

### I. INTRODUCTION

In the past few decades there has been great revolution in computing and communication technology. One can book a movie ticket, shop online, transfer money via different commercial websites on a single click. Such a progressed transformation is due to advances in handy devices, digital media and geo-based technology. Such multi-channel shopping revolution has put the customer in control, convincing the retailers to look for a single, seamless approach that lets them interact with their customers anytime and anywhere, across any and all channels. E-commerce companies have now turned to Data Science for focused customer group targeting and evaluating campaign strategies., E-commerce companies are greatly using data science to push the customers to the wide range of products & sell it as per customer's purchasing power and concern. Use of Data Science plays an important role in tracking the entire journey of a customer, from entry to exit on a particular commercial website. An average online shopper may not realise that every click is being monitored and that all purchases being made are captured from beginning to end. Due to this the customers get easily divided into different sectors based on the combination of their purchase patterns and demographic details . This not only provides an interactive and richer experience to customers, but also improves the retailer's ability to increase sales through cross-sell and up-sell.

### **II. WHAT IS DATA SCIENCE?**

Data science, also known as data-driven science, is an interdisciplinary field about scientific methods, processes, and systems to extract knowledge or insights from data in various forms, either structured or unstructured. Data science is a "concept to unify statistics, data analysis and their related methods" in order to "understand and analyze actual phenomena with data". It employs techniques and theories drawn from many fields within the broad areas of Mathematics, Statistics, Information Science and Computer Science.



Fig.1: Data science process flowchart -"Doing Data Science", Cathy O'Neil and Rachel Schutt, 2013

Volume 4, Issue 2 : April - June, 2017



Fig. 2 : Data Science as in Venn Diagram

The above venn diagram elaborates Data Science as the combination of fields such as Computer Science, Mathematics and Statistics .The domain knowledge about the particular business also plays vital role in Data Science. Machine learning and Data Analytics are the blends of aforesaid fields.

### Machine Learning

The term machine learning means the automatic detection of significant patterns in data. Since last couple of years it has become a common implement in almost any task that requires essence of information gathered from large data sets. The search engines like Google, Yahoo are nothing but machine learning based technologies which are used to bring us the best results by placing worthwhile advertisements. The anti-virus software are also machine learning based technologies which filter our email messages, credit card transactions are secured by a software that learns how to detect frauds. Digital cameras learn to detect faces and voice commands given by us are recognized by intelligent personal assistance applications on smart-phones . The accident prevention systems in cars are built using machine learning algorithms. Machine learning is also widely used in scientific applications such as bioinformatics, medicine, and astronomy. Thus Machine Learning tools give us programs with the ability to "learn" and adapt.

### Data Analytics

Data Analytics is the science of inspecting raw data with the purpose of locating patterns and drawing inferences about the information by applying an algorithmic or mechanical process to derive insights. Data analytics methodologies consists of two types of techniques :

- i) Exploratory Data Analysis (EDA) It is applied to find patterns and relationships in data.
- ii) Confirmatory Data Analysis (CDA)- It involves use of various statistical techniques to determine whether the claims stated about a data set are proper or not.

An eminent statistician John W. Tukey says- "EDA is often compared to detective work, while CDA is akin to the work of a judge or jury during a court trial"

### **III. APPLICATIONS OF DATA SCIENCE**

### • Web browsing

The web search engines like Google, Yahoo, AOL, Ask, Bing etc. deliver the best results for our online search queries in fraction of seconds using the data science algorithms.

### • Targeted Digital Advertisements

Traditional advertising is now well complimented by the Digital advertising / marketing field due to effective use of Data Science. Display banners & advertisements on various websites and social media sites, digital bill boards at airports, are decided by use of data science. This has significantly helped to increase Click Through Rate (CTR) of digital advertisements. This is the reason why one sees ads of job/ immigration opportunities while his/her friend sees ad of kitchenware products in the same place at the same time.

### Recommender Systems

A lot of companies like Amazon, Twitter, Google Play, Netflix, LinkedIn have aggressively used this system to promote their products / suggestions in accordance with user's interest and relevance of information. The recommendations are made based on previous search results of a user.

Volume 4, Issue 2 : April - June, 2017

### • Image Recognition

Social media websites automatically helps to tag your friends when you upload your photos there. Such tagging is done using Image Recognition .It also gives you a suggestion to connect & find new friends through mutual networks. This automatic tag suggestion feature uses face recognition algorithm. Similarly, while paying or transferring money, one can scan a barcode or QR code in web browser using mobile phone. It uses image recognition.

### Airline Route Planning

Airline Industry across the world is known to have heavy losses which are attributed to various reasons. Maintaining occupancy ratio and operating profits. High fluctuation in air fuel prices and need to offer heavy discounts to customers have further made the situation worse. Use of data science has tactically provided the solution and areas of improvements. Now using data science, the airline companies can predict flight delay, decide which class of airplanes to buy, whether to directly land at the destination, or take a halt in between, decide the air fares based on passenger density on a given route and effectively drive customer loyalty programs etc.

### • Fraud and Risk Detection

Data Science is also applied in Finance sector. Finance companies were losing their profits because of amount overdue by customers and losses every year. However, they had a lot of data which were collected during the initial paper work while sanctioning loans. They decided to practise data science in order to liberate them out of losses. Hence banking companies learned how to collect the data via customer profiling, their past expenditures and other essential variables to analyze the probabilities of risk and default. Moreover, it also helped them to push their banking products based on customer's purchasing power.

Apart from the applications mentioned above, data science is also used in Marketing, Finance, Human Resources, Health Care, Government Policies and every possible industry where data gets generated. Using data science, the marketing departments of companies decide which products are best for Up selling and Cross selling, based on the data related to customer's choice. In addition, predicting the wallet share of a customer, which customer is likely to churn, which customer should be kicked for high value product and many other questions can be easily answered by data science. Finance (Credit Risk, Fraud), Human Resources (which employees are most likely to leave, employees performance, decide employees bonus) and many other tasks are easily accomplished using data science in these disciplines.

### IV. SKILL SET AND SCOPE FOR BEING A DATA SCIENTIST

Data Scientist have a huge scope in India, for experienced as well as fresher who want to make a career in data science. These are some of the top Indian companies that hire data scientists –

### Service-based company -

Fractal Analytics, Mu Sigma, Citi, HCL, Uber, Goldman Sachs, IBM, CapGemini , Accenture.

Product-based company- Amazon, Flipkart, Paytm, Haptik etc.

The skills needed to be data scientists:

- In-depth knowledge of SAS and/or R. For Data Science, R is generally preferred.
- Python coding: Python is the most common coding language that is used in data science along with Java, Perl, C/C++.
- SQL database/coding: Preferred candidates can write and execute complex queries in SQL.
- Working with unstructured data: It is extremely important that a Data Scientist is able to work with unstructured data—whether from social media, video feeds, audio or other sources.

### **V. CONCLUSION**

Today, developments in Information Technology have changed the socio-economic status of the society in terms of flexibility in computing, communication and business. It has generated n number of job opportunities in the fields of computer science as well as E-Commerce. People with exact domain knowledge and proper skill set regarding a particular job profile gets absorbed immediately with impressive salary structure. This indicates a positive sign to the growth of Indian Economy.

### VI. REFERENCES

1. Tukey John W. "Exploratory Data Analysis", Pearson's Publication, 1977

Volume 4, Issue 2 : April - June, 2017

- 2. Shai Shalev-Shwartz and Shai Ben-David , "Understanding Machine Learning: From Theory to Algorithms", Published 2014 by Cambridge University Press.
- 3. Ellioty, Jane, Catherine Marsh, "Exploring Data-An Introduction To Data Analysis for Social Scientists", Polity Press, 2008.
- 4. https://www.analyticsvidhya.com/blog/2015/09/applications\_data\_science/
- 5. https://www.yourstory.com/2016/11/big-data-impacting-e-commerce-industry/
- 6. https://www.searchdatamanagement.techtarget.com/
- 7. https://www.simplylearn.com/data-science/

# MANUSCRIPT SUBMISSION

# **GUIDELINES FOR CONTRIBUTORS**

- 1. Manuscripts should be submitted preferably through email and the research article / paper should preferably not exceed 8 10 pages in all.
- 2. Book review must contain the name of the author and the book reviewed, the place of publication and publisher, date of publication, number of pages and price.
- 3. Manuscripts should be typed in 12 font-size, Times New Roman, single spaced with 1" margin on a standard A4 size paper. Manuscripts should be organized in the following order: title, name(s) of author(s) and his/her (their) complete affiliation(s) including zip code(s), Abstract (not exceeding 350 words), Introduction, Main body of paper, Conclusion and References.
- 4. The title of the paper should be in capital letters, bold, size 16" and centered at the top of the first page. The author(s) and affiliations(s) should be centered, bold, size 14" and single-spaced, beginning from the second line below the title.

# First Author Name1, Second Author Name2, Third Author Name3

1Author Designation, Department, Organization, City, email id

2Author Designation, Department, Organization, City, email id

3Author Designation, Department, Organization, City, email id

- 5. The abstract should summarize the context, content and conclusions of the paper in less than 350 words in 12 points italic Times New Roman. The abstract should have about five key words in alphabetical order separated by comma of 12 points italic Times New Roman.
- 6. Figures and tables should be centered, separately numbered, self explained. Please note that table titles must be above the table and sources of data should be mentioned below the table. The authors should ensure that tables and figures are referred to from the main text.

# **EXAMPLES OF REFERENCES**

All references must be arranged first alphabetically and then it may be further sorted chronologically also.

# • Single author journal article:

Fox, S. (1984). Empowerment as a catalyst for change: an example for the food industry. *Supply Chain Management*, 2(3), 29–33.

Bateson, C. D.,(2006), 'Doing Business after the Fall: The Virtue of Moral Hypocrisy', Journal of Business Ethics, 66: 321 – 335

# • Multiple author journal article:

Khan, M. R., Islam, A. F. M. M., & Das, D. (1886). A Factor Analytic Study on the Validity of a Union Commitment Scale. *Journal of Applied Psychology*, *12*(1), 129-136.

Liu, W.B, Wongcha A, & Peng, K.C. (2012), "Adopting Super-Efficiency And Tobit Model On Analyzing the Efficiency of Teacher's Colleges In Thailand", International Journal on New Trends In Education and Their Implications, Vol.3.3, 108 – 114.

# • Text Book:

Simchi-Levi, D., Kaminsky, P., & Simchi-Levi, E. (2007). *Designing and Managing the Supply Chain: Concepts, Strategies and Case Studies* (3rd ed.). New York: McGraw-Hill.

S. Neelamegham," Marketing in India, Cases and Reading, Vikas Publishing House Pvt. Ltd, III Edition, 2000.

# • Edited book having one editor:

Raine, A. (Ed.). (2006). Crime and schizophrenia: Causes and cures. New York: Nova Science.

# • Edited book having more than one editor:

Greenspan, E. L., & Rosenberg, M. (Eds.). (2009). *Martin's annual criminal code:Student edition 2010*. Aurora, ON: Canada Law Book.

# • Chapter in edited book having one editor:

Bessley, M., & Wilson, P. (1984). Public policy and small firms in Britain. In Levicki, C. (Ed.), *Small Business Theory and Policy* (pp. 111–126). London: Croom Helm.

# • Chapter in edited book having more than one editor:

Young, M. E., & Wasserman, E. A. (2005). Theories of learning. In K. Lamberts, & R. L. Goldstone (Eds.), *Handbook of cognition* (pp. 161-182). Thousand Oaks, CA: Sage.

• Electronic sources should include the URL of the website at which they may be found, as shown:

Sillick, T. J., & Schutte, N. S. (2006). Emotional intelligence and self-esteem mediate between perceived early parental love and adult happiness. *E-Journal of Applied Psychology*, 2(2), 38-48. Retrieved from http://ojs.lib.swin.edu.au/index.php/ejap

# • Unpublished dissertation/ paper:

Uddin, K. (2000). A Study of Corporate Governance in a Developing Country: A Case of Bangladesh (Unpublished Dissertation). Lingnan University, Hong Kong.

# • Article in newspaper:

Yunus, M. (2005, March 23). Micro Credit and Poverty Alleviation in Bangladesh. *The Bangladesh Observer*, p. 9.

# • Article in magazine:

Holloway, M. (2005, August 6). When extinct isn't. Scientific American, 293, 22-23.

# • Website of any institution:

Central Bank of India (2005). *Income Recognition Norms Definition of NPA*. Retrieved August 10, 2005, from http://www.centralbankofindia.co.in/ home/index1.htm, viewed on

- 7. The submission implies that the work has not been published earlier elsewhere and is not under consideration to be published anywhere else if selected for publication in the journal of Indian Academicians and Researchers Association.
- 8. Decision of the Editorial Board regarding selection/rejection of the articles will be final.



# INDIAN ACADEMICIANS & RESEARCHERS ASSOCIATION

# **Major Objectives**

- To encourage scholarly work in research
- To provide a forum for discussion of problems related to educational research
- To conduct workshops, seminars, conferences etc. on educational research
- To provide financial assistance to the research scholars
- To encourage Researcher to become involved in systematic research activities
- To foster the exchange of ideas and knowledge across the globe

# **Services Offered**

- Free Membership with certificate
- Publication of Conference Proceeding
- Organize Joint Conference / FDP
- Outsource Survey for Research Project
- Outsource Journal Publication for Institute
- Information on job vacancies

Indian Academicians and Researchers Association Shanti Path ,Opp. Darwin Campus II, Zoo Road Tiniali, Guwahati, Assam Mobile : +919999817591, email : info@iaraedu.com www.iaraedu.com

# EF EMPYREAL PUBLISHING HOUSE

- Assistant in Synopsis & Thesis writing
- Assistant in Research paper writing
- Publish Thesis into Book with ISBN
- Publish Edited Book with ISBN
- Outsource Journal Publication with ISSN for Institute and private universities.
- Publish Conference Proceeding with ISBN
- Booking of ISBN
- Outsource Survey for Research Project

# Publish Your Thesis into Book with ISBN "Become An Author"

# **EMPYREAL PUBLISHING HOUSE**

Zoo Road Tiniali, Guwahati, Assam Mobile : +919999817591, email : info@editedbook.in, www.editedbook.in