विद्याविनियोगाद्विकासः
IIM
AHMEDABAD

4th IIMA International Conference on

Advanced Data Analysis, Business Analytics and Intelligence

April 11 - 12, 2015, Ahmedabad, India

ABSTRACT BOOKLET

4th IIMA International Conference on Advanced Data Analysis, Business Analytics and Intelligence

April 11-12, 2015

Indian Institute of Management Ahmedabad, India

4th IIMA International Conference on

Advanced Data Analysis, Business Analytics and Intelligence

April 11-12, 2015, Ahmedabad, India.

Indian Institute of Management Ahmedabad is happy to organize the 4th international conference dedicated to advanced data analysis, business analytics and business intelligence and the objectives of the conference are to facilitate sharing of:

- Research based knowledge related to advanced data analysis, business analytics and business intelligence among academicians and practitioners
- Case studies and novel business applications of tools and techniques of advanced data analysis, business analytics and business intelligence among academicians and practitioners.

Papers were selected for regular presentation or short presentation based on the recommendation of the reviewers following a double blind peer-review process. The entire ICADABAI-2015 conference will be video recorded and turned into an interactive DVD.

Conference Team of ICADABAI-2015

Arnab Kumar Laha (Conference Convener)
Indian Institute of Management, Ahmedabad

Pravida Raja A.C. (Conference Manager)
Indian Institute of Management,Ahmedabad.

Mahesh K. C. (Conference Member)
Institute of Management, Nirma University, Ahmedabad

Members of International Programme Committee of ICADABAI-2015

1. Bag Dinabandhu- National Institute of Technology, Rourkela
2. Basu Devraj- University of Strathclyde, UK
3. Basu Sumanta - Indian Institute of Management, Calcutta
4. Biswas Atanu- Indian Statistical Institute
5. Carol Anne Hargreaves- National University of Singapore
6. Chakraborty Arnab - Accenture Digital
7. Chaterjee Snigdhanshu- University of Minnesota, USA
8. Das Samarjit - Indian Statistical Institute
9. Doctor Gayatri- CEPT University, Ahmedabad.
10. Gattu Nandu- GVK Biosciences Private Limited, Hyderabad.
11. Ghosh Dilip Kumar- Saurashtra University, Rajkot
12. Ghosh Pulak - Indian Institute of Management, Bangalore
13. Ghosh Sujit K. - North Carolina State University, USA
14. Kaneko Shinji - A Hiroshima University
15. Karandikar Rajeeva Laxman- Chennai Mathematical Institute
16. Kumar Dinesh Unikrishnan- Indian Institute of Management, Bangalore
17. Laha Arijit - Infosys
18. Maitra Ranjan- Iowa State University, USA
19. Mitra Murari - Indian Institute of Engineering Science and Technology, Shibpur.
20. Mohanty Bhabha Krishna- Indian Institute of Management, Lucknow
21. Mukhopadhyay Chiranjit- Indian Institute of Science, Bangalore
22. Pal Sandip K- Unisys Global Service India
23. Patil Hemant - Dhirubhai Ambani Institute of Information and Communication technology, Gandhinagar
24. Paudyal Krishna- University of Strathclyde, UK.
25. Pochiraju Bhimasankaram- Indian School of Business, Hyderabad
26. Prabhu Jayant K- Wipro Technologies
27. Sapra Amar- Indian Institute of Management, Bangalore
28. Sarkar Abhinanda- Myra School of Business, Mysore
29. Sarkar Nityananda - Indian Statistical Institute
30. Sen Sudipta K- SAS
31. SenGupta Ashis - Indian Statistical Institute
32. Sengupta Raghu Nandan- Indian Institute of Technology Kanpur
33. Singh Sanjeet- Indian Institute of Management, Calcutta
34. Velamakanni Srikanth- Fractal Analytics

Session Organizer of ICADABAI-2015

Prasun Das, Indian Statistical Institute, Kolkata

Sponsors of ICADABAI-2015

GOLD SPONSORS

Fractal Analytics Pvt. Ltd. General Mills India Pvt. Ltd.

SILVER SPONSOR

SAS Institute (India) Pvt. Ltd

Bronze Sponsors

Cubic Computing (P)Ltd. Emerald Group Publishing (India) Pvt.Ltd

Acknowledgement

Indian Institute of Management, Ahmedabad.,

Progressive Innovations, New Delhi.

Referees of ICADABAI 2015:

- **Biju Varkkey**
  IIM Ahmedabad
- **Debjit Roy**
  IIM Ahmedabad
- **Sanjeev Tripathi**
  IIM Ahmedabad
- **Ramanathan Subramaniam**
  IIM Ahmedabad
- **Karthik Sriram**
  IIM Ahmedabad
- **Viswanath Pingali**
  IIM Ahmedabad
- **Chiranjit Mukhopadhyay**
  IISc- Bangalore
- **Mahesh K.C**
  Institute of Managemet Nirma University, Ahmedabad
- **Dinesh Kumar Unnikrishnan**
  IIM Bangalore
- **Kirti Sharda**
  IIM Ahmedabad
- **Abhishek**
  IIM Ahmedabad
- **Ranjan Maitra**
  ISU-USA
- **Sachin Jayaswal**
  IIM Ahmedabad
- **Krishnamoorthy Srikumar**
  IIM Ahmedabad
- **Apratim Guha**
  IIM Ahmedabad
- **Debdatta Pal**
  IIM Raipur
- **Kavitha Ranganathan**
  IIM Ahmedabad
- **Samarjit Das**
  ISI Kolkata
- **Arijit Laha**
  Infosys Ltd., Hyderabad
- **Nandu Gattu**
  GVK Biosciences Private Limited, Hyderabad
- **Jayant Varma**
  IIM Ahmedabad
- **Mohanty Bhabha Krishna**
  IIM Lucknow
- **Agarwalla Sobhesh Kumar**
  IIM Ahmedabad
- **Sumanta Basu**
  IIM Kolkata
- **Saral Mukherjee**
  IIM Ahmedabad
- **Sanjay Verma**
  IIM Ahmedabad
- **Rajeev Sharma**
  IIM Ahmedabad
- **Pulak Ghosh**
  IIM Bangalore
- **Goutam Dutta**
  IIM Ahmedabad
- **Amit Garg**
  IIM Ahmedabad
- **D. K Ghosh**
  Saurashtra University, Rajkot.
- **Sanjeet Singh**
  IIM Kolkata
- **Nilesh Gupta**
  IIM Trichy
- **Ankur Sarin**
  IIM Ahmedabad
- **Anand Jaiswal**
  IIM Ahmedabad
- **Abhinanda Sarkar**
  Myra School of Business, Mysore
- **Snigdhanshu Chatterjee**
  University of Minnesota, USA

Contents

IC 15/ KN-1

Multi Armed Bandit Sampling in Nested Simulation for Financial Portfolio Risk Measurement

Sandeep Juneja
Tata Institute of Fundamental Research, Mumbai

Consider a financial portfolio comprising many and diverse financial securities including interest rate, equity, credit, and commodity financial derivatives as well as CDO tranches. Risk measurement of such portfolios is an important problem faced by large banks. The value of the portfolio at any state-time can be seen to be a conditional expectation that is typically not amenable to analysis and needs nested simulation estimation. However, naïve implementations can be computationally prohibitive.

Multi-armed-bandit techniques are common in online machine learning literature. In this talk we adapt these powerful computational techniques to portfolio risk measurement and develop a multi-armed-bandit based sampling method to determine whether at any time portfolio loss exceeds a specified threshold. We show that the proposed method is computationally `optimal' amongst a large class of sampling methods.

IC 15/KN-2

Drug Repurposing Analytics: Identification of New Uses of Old/ Failed Drugs

Nandu Gattu
GVK Lifesciences, Hyderabad

Drug discovery and development process requires an average of $1.5 Billion dollars and takes 14 years. During this process, significant number of compounds fail before it reaches to the pharmacy shelf and its cost millions of dollars to pharmaceutical companies. Drug Repurposing is a strategy in which we take failed or shelved compounds from late stage clinical development and identify new indications to rescue the investment. This approach requires integration and analysis of data from diverse areas of drug discovery and development including but not limited to chemistry, biology, genetics, genomics, proteomics, metabolomics, pharmacology, adverse events and clinical trials. A comprehensive and multi-pronged Drug Repurposing Data Analytics Platform has developed using number of proprietary and public biological/chemical/clinical databases and algorithms. The multi-pronged drug repurposing approach explores and exploits the relationship between drug–target–disease at various levels using eight different strategies such as structural similarity, adverse events, pathways, gene expression, genome-wide association studies, interactome, literature mining and clinical trials. The current presentation will elaborate more about the power of big data analytics in pharma world along with couple of case studies

IC 15/KN-3

Role of Statistics in the Big Data Era

Rajeeva Karandikar
Chennai Mathematical Institute, Chennai, India.

One often hears that with tons of data available, who needs statistics? We will examine this issue and argue that for getting deeper insights into the phenomenon that is generating the data, statistical thought, models and ideas of statistical inference are important ingredients. Availability of large chunks of data does not replace these ideas. Of course statistics has to evolve and we can no longer live in a world where n=30 was considered large.

IC 15/KN-3

Statistical Perspective of Clinical Trials

Atanu Biswas
Indian Statistical Institute, Kolkata, India.

The logistic and issues in clinical trials are discussed from a statistical point of view. We start with the four phases of experimentation in a clinical trial. We then discuss the important issues of objectives, ethics, randomization, sample size determination, power in a clinical trial. We then discuss the logistics of response-adaptive clinical trials which are used to have some ethical gain by allocating a larger proportion of patients to the eventual better treatment. Different response-adaptive designs are discussed in the context.

IC-15/040

Pre-Opening Call Auction of Stock Price Discovery Using Multivariate Lognormal Montecarlo Simulation

Saurabh Goel
Delhi Technological University, New Delhi,

Dinabandhu Bag
National Institute of Technology Rourkela

Market microstructure design and testing of efficient market hypotheses have been of long interest in security research. Market microstructure discussions also include the impact of call auctions in price discovery and bringing out efficiency in the functioning of markets. Call auctions form a method of trading, where orders over a small time period are collected and the market opening price are derived based on aggregated supply and demand for the underlying security. Empirically the effect of introduction of opening call has not been found to be unambiguous. This paper uses Monte Carlo simulation of co-dependent prices to explain the equilibrium price formation process. It also simulates the equilibrium prices using multivariate lognormal distribution and derives the price discovery impact of the pre-opening call auctions on normal markets. Our simulation results can generate volatility estimates close to the actual historical volatility of pre-opening equilibrium price and discovery efficiency. We provide an application of Multivariate Monte Carlo simulation to improve upon the volatility estimates by using daily pre-opening auction data for over 50 stocks listed in NSE.

IC 15/104

Impact of Stock Market Liberalization on Returns Volatility and Economic Growth: an Evidence from Asian Economies

Rahul Roy, Kavya T B
Department of Commerce, School of Management,
Pondicherry University, Pondicherry, India-605 014

Shijin Santhakumar
Department of Commerce, School of Management,
Pondicherry University, Pondicherry, India-605 014

The present study tries to quantify the dynamics persistence among and impact of, stock market liberalization on returns volatility in 14 Asian emerging economies through modelling the GARCH and TGARCH model of volatility. The impact of stock market financial liberalization on returns volatility in Asian economies is witnessed subsequently by categorizing the study period into three phases, namely pre-, during, and post-liberalization. The dynamicity and behaviour in returns volatility further entangled through employing the daily, weekly and monthly frequency data set to witness the impact of financial liberalization on volatility returns in Asian economies including global financial crunch of 2007-08. Successively, the second phase of the study tries to assess the impact of stock market liberalization on industrial growth in the Indian economy. The literature insists and accounts for the direct relationship among financial liberalization and economic growth. Thus, the third phase of the present study tries to examine the impact of financial stock market liberalization on the economic growth in emerging Asian economies owing to different course of liberalization, after controlling the macro-economic factors.

IC 15/ 110

Bond Market Development and Economic Growth: Is There a Causal Effect

Rudra P. Pradhan, Danish Zaki, Ranapratap Maradana, Debaleena Chatterjee
Indian Institute of Technology Kharagpur (India)

Uday Kiran Marepalli
Birla Institute of Technology, Mesra (India)

The paper examines the long-run relationship between bond market development and economic growth in 34 selected countries for the period 1993-2011. The statistical methods used in this study are Principal Component Analysis (PCA) and panel vector auto-regressive (VAR) model. The PCA is used to construct the composite index of bond market development, which can project the overall position of bond market in the 34 selected countries. On the other hand, panel VAR model is used to know the casual-nexus between bond market development and economic growth. The panel VAR technique is the application of cointegration and Granger causality analysis on a panel of cross sectional units. The empirical investigations starts with unit root and cointegration check. Using panel unit root test and panel cointegration test, the study finds that bond market development, and economic growth are integrated of order one and they are cointegrated, indicating the existence of long run association between the two. The empirical findings of panel VAR model suggest the existence of bidirectional Granger causality between bond market development economic growth. The policy implication of this study is that the economic policies should recognize the differences in the bond market-growth nexus in order to maintain sustainable financial development and economic growth in the selected 34 countries.

IC15/024

Decisive Lending Rate Analytics for Rural Clients' Clusters

Baba Gnanakumar P,
Sri Krishna Arts & Science College, Coimbatore

This analytics creates clusters for banking clients to determine the lending rate without affecting the prime lending rate of RBI. This enables to create a sustainable lending policy for the economically backward people of India.

Indian banks are evolving with innovative methods in loan approval process, in order to retain their customers and to get rid of competition. Retail customers having good credit score are ready to bargain for less lending rate. In this research, we construct the analytics with the aim of fixing the transparent credit approval system by the banks . This analytics enables the bankers/micro-finance institutions to revise the lending rates of loans based on the creditworthiness of the customers. The model has been evolved based on the data collected from 328 retail loan borrowers. We create clustering procedure after exploring the relationship between lending rate and credit scoring based on vector error correction modelling. The study revealed that, a hundred-point swell in credit score decreases the lending rate by 40 basic points. The analytics portray that the impressive clients will be encourages to have minimum prime lending rate; whereas , modest customers will be charged with higher lending rate. We justify the floating lending rate analytics based on the credit scores as classified based on the clients' clustering process.

IC 15/025

Price Discovery and Impact on Volatility: A Study of Indian Derivatives Market

Divya Verma Gakhar
Gurugobind Sing Indraprastha University, New Delhi

Derivatives were introduced in year 2000, with an objective that they will help in price discovery in the spot market and also volatility of the market will reduce. The present study tries to evaluate the impact of introduction on derivatives spot market volatility and whether spot market leads the futures market or vice versa. The data was collected for a period of 18 years from 1996-2014 of Nifty and futures Index. The results of AR (1) - GARCH (1,1) model shows that volatility has reduced after introduction of derivatives. The results of Cointegration test and Vector Error Correction Model also show that long run relationship exists in both the market and Nifty leads the Nifty futures market.

IC 15/ 072

An Alternative Preference Elicitation Procedure in Predicting Choice

Ram Kumar Dhurkari
T.A. Pai Management Institute, Manipal,

Anjan Kumar Swain
Indian Institute of Management Kozhikode

This research proposes the development and implementation of a better method (MAGL: Multi Attribute Gain Loss) for multi-attribute decision making (MADM) under certainty. MAGL is based upon the tenets of norm theory, prospect theory, Kauffman's complexity theory and context dependent choice theories. Since the proposed method is a compositional approach using the frameworks of Multi-Criteria Decision Analysis (MCDA), its effectiveness is tested against Analytic Hierarchy Process (AHP) on two MADM problems for predicting the choice behavior of the DM. With less number of preferences taken from each individual in MAGL (order of n) in comparison to AHP (order of n^2), the solution prescribed by the former resembles more closely with the actual choice behavior of the individuals. The analytical model provided by MAGL is consistent and strengthening the arguments of academicians and practitioners that the overall satisfaction function might not be linear and/or symmetric. The results of two studies conducted during the course of this research conclude that the alternative valuations on different attributes is a nonlinear- reference point dependent function of the associated objective value of the alternative on the same attribute. The integration rule which best describes how evaluations are integrated into overall valuations follow a non-linear, non-compensatory context dependent relation, overweighing of negative information.

MAGL can help marketing/product managers analyze the dominance of attribute levels in the selection/rejection of an alternative. Since the context or the choice set plays an active role in the process of choice and MAGL is able to model the context dependent choice behavior of the consumers, the marketing/product managers can design new products and analyze which combination of these attribute levels will perform better with the choice set already available in the market. With the preferences on product attributes from a sample population, marketing/product managers can also analyze the changes in the market share with the introduction (recall) of products in (from) the market of the similar consumer population.

IC 15/147

Analyzing the Impact of Global Financial Crisis on the Interconnectedness of Asian Stock Market Using Network Science

Jitendra Aswani
Indira Gandhi Institute of Development Research (IGIDR)
Mumbai

In the first section of this study, impact of Global Financial Crisis (GFC) on the synchronization of fourteen Asian stock Markets of these countries: Hong Kong, India, Thailand, Singapore, Taiwan, Pakistan, Bangladesh, South Korea, Malaysia, Indonesia, Japan, China, Philippines and Sri Lanka, has been analyzed using the network science and its metrics like degree of node, clustering coefficient and network density. Then in the second section of this study by introducing the US stock market in existing network and developing a Minimum Spanning Tree (MST) spread of crisis from the US stock market to Asian Stock Markets (ASM) has been explained. Data used for this study is adjusted closing price of these indices from 6th January, 2000 to 15th September, 2013 which further divided into three sub periods: pre , during and post crisis. Using network analysis, it is found that Asian stock markets become more interdependent during the crisis than pre and post crisis, and also Hong Kong, India, South Korea and Japan are systemic important stock markets in the Asian region. Therefore, failure or shock to any of these systemic important stock markets can cause contagion to other stock market of this region. This study is useful for global investors' in portfolio management especially during the crisis period and also for policy makers in formulating the financial regulation norms by knowing the connections between the stock markets and how the system of these stock markets changes in crisis period and after that.

IC 15/010

Effects of Energy Price Rise on Investment: Firm Level Evidence from Indian Manufacturing Sector

Anver Sadath C
Department of Economics, Central University of Kerala

Rajesh Acharya H,
Department of Humanities, Social Sciences and Management, National Institute of Technology Karnataka

This paper analyses the effects of the rising prices of energy products on the investment of a large panel of manufacturing firms in India during 1993- 2013. The prime motivation behind this study is the absence of an empirical study into this research issue exclusively on Indian economy. The empirical results obtained by estimating an Error Correction Model (ECM) using Generalized Method of Moments (GMM) show that energy price rise has negative effect on the investment of firms in the manufacturing sector. The negative effect is transmitted to the firm's investment through both demand-side and supply-side factors. The transmission also depends upon factors such as the energy intensity of production. The results also show that the sales-growth-investment relationship becomes weak in the face of the rising prices of the energy which could be due to the cautious approach to investment adopted by the firms, as suggested by theories of irreversible investment. Therefore, it calls for the attention of the policy makers to evolve a comprehensive energy-policy to ensure continuous supply of energy at affordable prices to the manufacturers.

IC 15/ 050

Algorithm for Modular-Capacitated Multi-Period Plant Location Problem with capacity closure constraint

Vikram Batra, Y K Agarwal
Indian Institute of Management, Lucknow

Selection of location for manufacturing plants is a strategic decision for an organization. Shifts in customer demand during the plant's lifespan can alter the attractiveness of a particular location, turning an optimal location of one period into a strategic blunder for the future. Closure or relocation of plants may be unviable, due to external factors and these inefficient locations would result in excess transportation costs, which cannot be offset, no matter how well the production plans or inventory are optimized in the operational level plans.

The complexity of modeling such problems has limited much of the traditional facility location research to simplified static (single-period) models.

This paper presents an algorithm to generate the optimal sequence for opening plants and installing modular capacity units across locations during a multi-period planning horizon. The objective is to achieve the lowest cumulative cost of transportation and capital investment. The algorithm was applied to a randomly generated set of locations (50 customers and 20 candidate plants) over a 10 year demand horizon. The multi-period model achieved a capacity sequence with a cumulative cost 3.2% lower than the year-on-year planned sequence.

To demonstrate the algorithm on an industry application, it was applied for the Indian automobile industry. This industry is a good candidate for the model as it has high transportation costs and capital is built in modular assembly lines. Also, there is a large ecosystem of vendors, workmen and their families which makes closure or relocation of capacity unviable.

IC 15/098

Performance Analysis of Warehouse Systems with Dynamic Batching

Vibhuti Dhingra, Debjit Roy
Indian Institute of Management, Ahmedabad

Jennifer Pazour
University of Central Florida

Order picking process is one of the largest components of the total operational costs in a warehouse. A robust layout design and efficient utilization of resources are necessary to optimize the performance of order picking systems and to ensure high levels of customer satisfaction. This paper introduces a parts-to-picker order fulfillment strategy called "dynamic batching" and studies its impact on the warehouse performance. In a dynamic batching policy, multiple orders are processed by the picker in parallel but an individual order is released as soon as it is filled instead of on a first-come, first-served basis. A constant work-in-process is maintained at the pick station as a new order arrives as soon as a completed order departs. This facilitates advantages of both sequential processing (by meeting individual order deadlines) and static order batching (by fulfilling several orders simultaneously thereby reducing the completion time per order). The implementation of a dynamic batching policy requires investment in sophisticated and expensive material handling equipment. In this paper, we develop an analytical model to estimate the throughput of the pick station under a dynamic

batching policy for a fixed number of order positions at the pick station. We study the trade-offs between the throughputs for a dynamic and static order batching policy. In addition, we explore various environmental factors (such as item commonality between orders and batch size) and determine the conditions under which the policy is most beneficial.

IC 15/ 111

Implementation and Validation of Multi Attribute Decision Making Technique for Supplier Selection in Supply Chain Management

Prof. Nital P. Nirmal
Department of Production Engineering, Shantilal Shah Engineering College, Sidsar Campus, Bhavnagar, Gujarat- 364002, India

Mr. Rakesh P. Nirmal
Cementing Section, O.N.G.C. Mehsana Asset, Mehsana- 384003, India

Dr. Mangal G. Bhatt
Shantilal Shah Engineering College
Sidsar Campus, Bhavnagar – 364002, Gujarat, India

In Today's Competitive Environment it is impossible to successfully produce low cost & high quality product, without satisfactory supplier. Thus one of the important purchasing decisions in smoothly running Supply Chain Management (SCM) is Supplier Selection. In this paper, the research paper pattern with the importance of SCM and its different Modelling areas, Supplier Selection Methodology (SSM) to learn the different research work carried out for Supplier Selection criteria evaluation and show the various methods for Supplier Criteria Evaluation and Ranking in SCM field have been discussed. In the current research paper authors aims to implement and validate the Complex Proportional Assessment (COPRAS) method in supplier selection and in conclusion comparing with ELECTRI-III ((Elimination Et Choix Traduisant la Realite elimination and choice translation reality) the multi attribute decision making method with the same input data. After analyzing and comparing it authors explore that the COPRAS method is more sensitive and simpler then ELECTRI-III which is also one of the Multi Attribute Decision Making (MADM) techniques. while in the given model provide relatively more sensitive than other supplier selection methods, same methodology can applied for selecting supplier in the Manufacturing, Pharmaceutical, Automobile, Ship Building, Agricultural Industries, FMCG (Fast Moving Consumer Goods), etc. where supplier plays vital role for the industries. The method will also use to selection phase of different areas of SCM Like, Logistics and transportation decision making, Reverse Supply Chain Management, related issues, Vendor Managed Inventory (VMI), Strategic decision making and uses of Information technology in SCM.

IC 15/ 016

Analytics for Improving Talent Acquisition Processes

Rajiv Srivastava, Girish Keshav Palshikar, Sachin Pawar
Tata Consultancy Services Limited, Pune, India.

Talent Acquisition (TA) is an important function within HR, responsible for recruiting high quality people for given job positions through various sources under stringent deadlines and cost constraints. Given the importance of TA in the overall successful operations and growth of any organization, in this paper we identify specific "business questions" focused on

analyzing various aspects of the TA processes, analyze past TA data using statistical analysis techniques and to discover novel patterns/insights and actionable knowledge which can help in improving the cost, efficiency and quality of recruitments. Our predictive analytics is mainly related to various durations and delays in TA, candidate selection or rejection, offer acceptance by selected candidates, root cause analysis for offer decline. We also use the data-mining technique of subgroup discovery to identify interesting patterns (e.g., candidate subgroups having unusually high decline ratios). We illustrate the approaches through a real-life dataset.

IC 15/061

Optimizing Merchant Discount Rate (MDR) for One of the Largest Merchant Acquirers in South America

Mathur Gautam, Goyal Tarang, Mudgil Vinay
Fiserv, Bangalore

Merchant Acquirers ('MA') play the essential role of accepting/ capturing, processing and settling the payments made through cards (credit, debit, etc.) for purchases made at merchants/ retailers. This MA had a client portfolio of around ~one million merchants with the mass market segment consisting of SMB merchants forming the major portion. Due to highly competitive market, the MA was interested to identify the right set of merchants who were less elastic and then adjust their Merchant Discount Rate (MDR) by few basis points to improve revenue & profitability without losing market or wallet share.

As a first step in this study, a robust statistical segmentation was developed using unsupervised machine learning technique to reveal the pattern of merchant's spend, product usage, growth trend(QoQ)and customer risk. A non-linear differential price elasticity function was used to analyze the impact of price and the influence of a number of qualitative factors on revenue. The results of modeling suggested that the most significant factors impacting price besides merchant elasticity were - rental price of POS, regional effect, number of excess POS or PDV machines at the merchant location. For each individual segment, price elasticity along with an optimization algorithm with complex constraints (maximum % change in price, maximum loss of market share etc.) was developed to optimize the MDR price and achieve optimal revenue. An in-market testing of this solution was done for a sample of merchants. A factorial design of experiment was used to design the campaign and build a test & control group. Based on the price change suggested by the pricing model, more than 150% incremental revenue lift was observed for test over control.

IC 15/ 088

An Application of Demographic Data as a Surrogate for Epidemiological Profiling in India

K Vinay Kumar
DATAWISE, Hyderabad

K Vijay Kumar
Decision Sciences, Citibank, India.

The absence of nation-wide profiling of the health condition of people in India has lead to the Government of India tending to take a broad brush approach with very little reference to the specific health condition of people in a region. This causes inefficient health management

mechanism with the government erring on the side of caution in order to ensure coverage. The key challenges arising out of inadequate and poor quality of data are uneven health care spending, and mismatch between disease profiles and provision of care.

As a result, a significant gap exists in the ability of policy makers to understand the implications for public health in those geographies which are not extensively covered. A resultant approach is a broad-brush health intervention which is inefficient while achieving its objective of universal health coverage.

Many of the data collected through the Census have a strong correlation to the health status of the individual. A question therefore arises as to whether it is possible to use demographic data available through the Census, for the purpose of deriving healthcare indicators, and thus using some of the data available through census as a means of deriving epidemiological data.

The results of the mapping of each of the demographic data collected from Census, initially within the NFHS data set, and subsequently, mapped to the Census dataset provides an insight into the potential disease that individuals in various geographies are likely to suffer. This data, once extrapolated to the ward level in a city provides a level of granularity which helps in a better understanding of the health parameters that are actionable.

IC 15 / 089

Visualization & Analysis of Employee Interactions A study of Inter-Departmental Communication Effectiveness at Hansa Cequity

Bharath V., Hemanth Kattamuri, Vaibhav Verdhan
Hansa Customer Equity Pvt. Ltd., Mumbai, Maharashtra

In any Knowledge Organization, measuring the collaboration and communication effectiveness between and within different departments is extremely important. Hansa Cequity is a Marketing Analytics Consulting firm having more than 200 employees across 6 key departments. Almost every project that is taken up requires collaboration from almost all the divisions. Hence identifying the bottlenecks in information flow is a need of the hour. In this paper, we have attempted to analyze the interactions between employees across different departments at Hansa Cequity. The connections between the employees were visualized in the form of Ego-Networks. Network Centrality measures like Weighted Degree, Betweenness, Closeness and Modularity were used to understand key influencers, employees who enable communication flow across teams and employees who are isolated from the rest of the network. The teams which had high bonding relative to others were found out using Community detection technique. Based on the centrality measures, employees were classified into similar groups. The demographic variables of these groups were then overlapped with their centrality measures to identify the key factors that drive their observed behaviour, performance at work and group formation. Centrality measures were also computed at a department level to understand whether all the departments have good interaction. This work has helped in identifying the key communication gaps and cross-departmental information flow barriers within Hansa Cequity. Based on the findings, several initiatives have been taken up by the Human Resources team to improve cross-functional collaboration.

IC 15/026

Recognizing trust in natural language in Amazon's online reviews

Maitreyee Tewari
School of Computing, University of Dundee DD1 4HN, Scotland UK

The goal of this project is to build a system which could extract and classify ethotic statements (ethos relates to trustworthiness, credibility and reliability of seller) about Amazon's service from a corpus of Amazon's reviews.

Until now processing and extracting ethos was done manually. With this project, we take the first step in automating the process of trust extraction from product reviews.

The paper includes discussion on ethos extraction and has used natural language processing, machine learning and python to achieve the goals. Specifically we have used sentiment analysis, argument mining, python, supervised and semi-supervised machine learning algorithms such as Naive Bayes and Maxent Classifiers.

The contribution of this project is the development of two classifiers. One classifier that classifies sentences into ethos support and ethos attack and the other classifier that extracts ethotic statements from a corpus of Amazon reviews. These classifiers provide an initial solution to automatic ethos extraction.

IC 15/-039

PROCESS MINING OF SUTURE MANUFACTURING FIRMS WITH DEFINE, MEASURE, ANALYZE, IMPROVE AND CONTROL METHODOLOGY

Dhanya. J. S
CET School of Management, Kerala

Variation is the impending problem of any Quality Control activity. This extent of variation may be from inestimably small to very large which are unacceptable to conformation. The variations beyond the specification limits as per the product requirements results in rejection and rework, thereby adding the cost to the company. These defects are quantified in terms of Defects per Million Opportunities. This paper delineates the empirical application of Six Sigma and Define, Measure, Analyze, Improve and Control methodology to reduce product defects within a suture manufacturing organisation. Currently the suture market currently exceeds $1.3 billion annually. Sutures, as medical devices, are subject to strict quality control. The study has been conducted to examine the different aspects to be considered for improving the quality of suture manufacturing process and thereby reducing the wastages. The study involves a Six Sigma Define, Measure, Analyze, Improve and Control approach in analyzing the manufacturing performance of the organization. Since the study analyses the rejection rate of the products; different tools like Pareto Chart, Cause and Effect diagram is used for analysis. The analysis from employing Six Sigma and Define, Measure, Analyze, Improve and Control methodology indicated that high moisture content influenced the amount of defective sutures produced. This paper aims to rank the emphasis placed on critical factors and quality management principles that determine the success of Six Sigma as it applies to improvement of quality in suture manufacturing process.

IC 15/ 051

Effect of Macro-Economic Factors on Demand for Two-Wheelers in India

Vikram Batra
IIM Lucknow

The traditional approach for forecasting sales demand in the two-wheeler industry in India primarily relies on naïve models with some overlay of judgmental forecasting methods. The empirical data on sales for the last 30 years shows a high level of variability on a monthly, quarterly and annual basis and the naïve traditional forecasting approaches are unable to provide reasonable levels of forecast accuracy.

This paper aims to determine the coincident and leading macroeconomic and other external factors influencing annual two-wheeler demand. This can help manufacturers better anticipate demand trends over a 1-3 year horizon and install adequate capacities and plan production levels.

The research design involved collection of longitudinal data for different economic and demographic indicators. These covered indicators affecting (i) aggregate demand (population, urbanization level, per capita income), (ii) affordability (petrol price, interest rates, household wealth, inflation) and (iii) impulse purchase (rainfall level, festival season).

Two regression models were derived using statistical analysis of the collected data to forecast sales. These models showed that sales of two-wheelers increase with (a) higher per capita income, (b) increase in pump petrol price, (c) increase in price of bullion gold and (d) lower consumer finance rates.

The difference in sales during high and low monsoon years was found to be not statistically significant (at $\alpha = 0.05$) and festival season sales were found to be higher than rest of year sales with a factor of between 3.31% and 10.26% (confidence interval 95%).

IC 15/ 083

How Is Whatsapp Driving Businesses To Enhance Conversion Rates? – An Empirical Approach

Madhavi R, Aditya Rao, Srisudhan R
Jain University, Banglore

In our research paper, the researchers have witnessed a steady acceptance by business enterprises to embrace WhatsApp as a major platform of promotion. The methodology the researchers have adopted for this research paper consisted of a questionnaire, which contains the profile of the individual filling in the paper, as well as a set of common questions regarding the use of WhatsApp in the respondents' business endeavours. There are real world examples of successful social media advertising campaigns in the research paper; ones that adapted itself to a new medium. This proves that big businesses are not against or immune to adapting to changing trends. Since WhatsApp is free of cost and there is ease of usage, it has become the go to promotions outlet for up & coming as well as semi formal businesses. That, along with various other facts like speed, time conservation and reach is the major benefits of using WhatsApp for a business enterprise. A few examples of the effective use of WhatsApp by professional enterprises are listed in order to show the prior real world application of the

App. The reasons for WhatsApp's tremendous popularity are also mentioned, with experts weighing in their esteemed opinions. We then have the research data listed out, which includes various facts about the business as well as its proprietor, the volume of increase in sales, profit and reach that WhatsApp has brought to their table, and how important WhatsApp is to their enterprise's growth in the long run. This is followed by the conclusion of the paper and the references for the paper.

IC 15/100

Exploring Health Related Interactions on Social Media: An Analysis Of Customer Perceptions And Managerial Implications

Shahazadi Shaik, M.V.S. Kameshwar Rao
GITAM School of International Business, GITAM University, Visakhapatnam, Andhra Pradesh

T. Vasudha
Freelance – Teaching, Andhra Pradesh

Sandeep Gupta
Eli Lilly and Company, Haryana

Globally, and particularly in the developed countries, social media platforms are fast emerging as cost-effective and accessible channels for both healthcare providers and online patients, offering strategic advantages to the overall health industry . However in the Indian context, their role is limited to dissemination of information rather than for sophisticated health care integration.

This paper attempts to explore the perceptions and readiness of the Indian customers towards health care interactions on social media. The survey results reveal a very nascent scenario. The sample is divided between the socially aware and unaware of such interactions and the ensuing issues of privacy and misuse of information. Majority of the sample prefers anonymity in discussing health in public. They would rather prefer a specialist opinion on health issues than discussing in the public domain. Revelations point towards a significant use of social media in gaining awareness about ailments and their preventive cure. Other emerging uses that appear to gain acceptance are for allied medical services, managing appointments, writing reviews, and sharing experiences. Weight and wellness top the topics for discussion on social media, followed by matters related to heart ailments, child care, and psychological disorders.

Educating and offering personalized value-added services to the customers through innovative engagement are the immediate implications for the relevant stakeholders of the health industry. Social Media Analytics is found to be a strategic tool towards this end in gaining insights into customer needs and behaviour and responding in an effective and socially responsible manner.

IC 15/ 120

TA Study on Consumer Demography and Preference towards Organic Food Products – Application of Discriminant Analysis

Vedha Balaji
Christ University

Organic food products are popular across Europe and United States of America. Asia is not far behind with India being a prominent player. The concept of organic food products is not new to Indian farmers. However there is not much of a consumption taking place

domestically despite the fact that India is one of the top ten players in the world when it comes to the number of farmers engaged in organic cultivation. This study was conducted to understand the discrepancy between preference and consumption amongst urban Indians in the city of Bangalore.

The study covers both primary investigation and secondary literature review. Data was collected with the help of structured questionnaires and analyzed using descriptive statistics and discriminant analysis to arrive at a decision rule to help retailers classify customers.

Majority of the respondents constituted male customers in the sample data when compared to females. Most of the respondents were less than 30 years of age followed by an older population above 50 years of age group. It was also observed that majority of the respondents in the sample data belonged to a monthly household income of 30,000 - 50,000 Rs. while the average spend per month belonged to the category of 500 - 1,000 Rs.

To conclude, it is possible to arrive at a decision rule by using the above mentioned demographic characteristics of customers into either regular or non-regular users of organic food products.

IC 15/129

An Application of Queuing Theory to Analyze Non-Clinical process in Hospital Operations

Deepak Yaduvanshi
Manipal Hospitals, Jaipur

Ashu Sharma
Narsee Monjee Institute of Management Studies, Mumbai

Waiting is inherent to the services as services are perishable and intangible. The perceived waiting is much more annoying for the patients than the actual waiting time for the patients. The studies had revealed that the subjective waiting time have higher co-relation with the frustration of the patients than the objective waiting time. This parameter was taken care of while giving the recommendations for the improvement of the process efficiency.

This study comprises of the in depth analysis of OPD from different dimensions. Basically the calculation of time taken by the outdoor patients at each visit for the Registration, for the new patients and billing for the other patients. The time and delay points in consultation are also calculated.

Also, to know about the patients perception on the delay and the probable reason and required solutions for the same. Based on all the observations and data collection performed during the study,a SWOT analysis was also done for the OPD of Fortis escorts Hospital Jaipur(FEHJ). SWOT analysis gives an opportunity to the organization to dissect the complete problem statement and come out with the solutions knowing where the organization can excel and where there is a scope of improvement to make the working and processes better.

Finally, after examining the problem analytically using the queuing model, measures were suggested to improve the delay points and make the service system more efficient and to gain a high patient satisfaction rating on which a hospital thrives upon.

IC 15/155

A study on Consumer Decision making towards organised and unorganised retailing in Mumbai City

Indu Mehta
Prin.L.N. Welingkar Institute of Management Development & Research Mumbai

Deepak Phophase
Infosys, US

In urban areas a marketplace achieves differentiation from different market formats – unorganized shops (shopping malls) and street markets (Fashion Street) through the pursuit of singular orientations following the hedonic and utilitarian dimensions of shopping. This article discusses how consumer decision-making styles shift towards shopping at malls as well as street markets in Mumbai City. The study attempts to analyze the impact of the socio demographic locations of street markets on the convenience and costs borne by consumers.

This study discusses the impact of growing congestion in urban marketplace due to overlapping of market territories of shopping malls and street markets in urban areas of Mumbai. This study also entails the important factors that can help symbiotic existence of both shopping malls and street markets in urban areas of Mumbai. The analysis has been carried out in reference to the shopping conveniences and shopping behavior. The results of the study evidence the complex consumer decision-making styles towards shopping malls and street markets based on competitive advantages. Street markets are largely preferred by the shoppers as they exhibit ethnic and cultural attributes while cross-cultural attributes of stores and shopping ambience is found to be one of the major determinants of shopping behavior.

IC 15/023

Forecasting Techniques For Air Cargo Forecasting And Air Traffic Forecasting In Changing Economic Environment

A.S. Pandey
Gitarattan International Business School, Rohini, New Delhi

Research on forecasting techniques for air cargo forecasting and air traffic forecasting is extensive and includes many studies that have tested alternative methods in order to determine which ones are most effective. In this paper, I had reviewed this evidence in order to provide guidelines for forecasting techniques for air cargo forecasting and air traffic forecasting. The coverage includes methodologies and assumptions involved in the current Air Traffic and Air Cargo Forecast, Air Cargo demand, Market Demand, Facility and Infrastructure Demand, Air Cargo Services and Other Supply Factors, Airport Traffic and Other Activity, forecasting techniques and models including Simple Growth Rate Model , Time series analysis, Econometric analysis, estimation of income and price elasticity's of passenger demand, Statistical Estimation of Model through example. Finally, Air Cargo Tonnage Forecast Summary and a case study of Indian Airports are also covered under Air traffic forecasting. Further, this paper also describes trend analysis, industry forecasts, and an econometric regression approach through example. This example illustrates the need for professional judgment in the forecasting process. In general, there is a need for statistical methods that incorporate the manager's domain knowledge. This includes Simple Growth Rate Model, Time series analysis and Econometric methods.

The paper concludes that Cargo forecasts are important in understanding an airport's demand growth, assessing market risk, and predicting financial gains/losses to develop management strategy. Airports that accurately forecast their future traffic will better anticipate the needs of their customers, and thus will be in a better position to develop to their full potential.

IC 15/029

Corporate bond yield approximation using plece wise cubic spline and Residuals Estimation with Liquidity factors

Dinabandhu Bag
School of Management, National Institute of Technology, Rourkela

Corporate bond yields are not easy to approximate on account of their limited trading frequency and irregularity in issuance. The corporate bond market in India demonstrates poor investor participation. The purpose of this paper is to present an account of piece wise cubic spline methodology of yield estimation that can be feasibly implemented by practioners. This paper attempts to approximate the corporate bond yield using residual maturity for listed and traded bonds using traded data from the period January 2001 to October 2014. We use a two stage process, where in the first stage, we model the yield behavior of corporate bonds against their residual maturity nodes of 3 years, 5 years and 20 years, respectively. In the second stage, we identify the significant factors of bond liquidity that impacts the cubic spline residual to improve accuracy in estimation. The first step of piece wise cubic spline estimates the 3 node parameters to predict the logarithmic yield to maturities against their remaining maturity in years. The second step intends to analyze the prediction errors and provides for correction in accuracy incorporating factors of bond liquidity such as issuer, issue and liquidity factors, respectively. The dual approach explains the variation in yields and could be used for identifying quality papers and the reasons for lower investor participation in bond markets and for arriving at term structure in related markets. This does not include pure central govt loans, state loans, treasury bills, dated securities, bond of departmental undertaking or india development bonds.

IC 15 /071

Vendor Managed Inventory (VMI) Practice: Some Insight Through Fresh Pilot Exploration From Indian Industries

Vinay A. Parikh, Mangal G. Bhatt
S.S.Engg. College, Bhavnagar, Gujarat

India is now-a-days considered as a fast developing country and earning the fruits of liberalization. Vendor-Managed Inventory (VMI) is a family of business models in which buyer of a product provides certain information to a vendor (Supply Chain) supplier of that product and the supplier takes full responsibility for maintaining an agreed inventory of the material, usually at the buyer's consumption location. Here an effort has been to identify various research constructs like supplier selection for VMI parts, Information, stages where supply chain inventory held, were identified among several Indian Industries in order to verify the VMI concept as a fresh pilot exploration. Various benefits of VMI models were derived as well as several organizational objectives achieved through VMI like customer service

improvement, improvement in profit margin, improvement in rate of return etc. Strategic drivers like competition, shorter product life cycle, corporate restructuring were viewed with obstacles like ineffective organizational structure, lack of suitable IT infrastructure. Result shows that 25% Industries adopted VMI at Full-fledged and 75 % adopted it partially. VMI model is on an average 63% beneficial to Reduction in transaction and 50 % each for Leveling in production process and inbound quality performance. It is concluded that lack of mutual understanding amongst partners and lack of decision making tools are the major obstacles for adoption.

IC 15/087

Impact of the monthly, day of the week and trading month effects on Bombay Stock Exchange Indices Returns for the period 2000-2014.

Surya Dev
IMIS Bhubaneswar

The paper investigates monthly, day of the week impact and trading month effect on Indian stock market, i.e. the Bombay Stock Exchange. The study is based on daily data across 19 indices of the Bombay Stock Exchange so that the effects can be judged for the entire market and across sectors. The analysis is carried out for the entire period. The period is again divided into pre and post subprime crisis and the analysis is repeated. The subprime crisis period in this study is taken to be 1st Jan 2008 to 31st Dec 2009. In this study, first an OLS is carried out followed by an ARCH model as the residual analysis shows an ARCH effect. The closing prices of daily data are used to calculate returns of the indices. The BSE Sensex, BSE 100, BSE 200, BSE 500, BSE Capital Goods BSE Consumer durables, BSE FMCG, BSE Health Care, BSEIT, BSE PSU, BSE Tech, BSE Metal, BSE Oil Gas are studied for the period 1st March 2000 to 28th October 2014. The other indices BSE Auto, BSE Bankex, BSE Midcap, BSE Small Cap and BSE Realty are studied from their inception till 28th Oct 2014. In the monthly effect it is seen that September and December appears to be months which have higher daily returns in the post subprime crisis period while November and December are the two months that have shown higher returns in pre subprime crisis period. If the overall data is taken into account, the month of December displays higher daily return. Similarly, Wednesday displays higher return in comparison to other days in the pre subprime crisis period and the overall data. But in the post subprime crisis period Monday displays higher returns in comparison to the other days of the week. It is also observed that returns are normally higher in the first fortnight of the month.

IC 15/ 094

Corporate Social Responsibility Disclosure and Corporate Financial performance: A Study of Indian Software and IT sector

Ruchi Gupta, Gaurav Agrawal
ABV - Indian Institute of Information Technology and Management,
Gwalior

Till date organizations used corporate social responsibility (CSR) as competitive advantage to show case their concern and behave in responsible manner by voluntary contribution for social development. But compulsory contribution of 2% profits for CSR activities make it matter of importance given in strategies formulation these days, due to implied accountability on corporate to justify their operation with respect to society and environment. In this regards sector specific activities differ on the basis of impact they made on society and environment. The aim of this paper is to identify reporting pattern of IT specific companies and find out the impact of size and profitability on disclosure norms. Content analysis and regression was used to fulfil the objectives. Annual reports of 17 IT companies which come under top 500 companies were selected for the study on the basis of income was collected from the official website of the companies. The results indicates that social issues are more important in comparison to environmental issues for CSR contribution and large companies contribute and disclose more but most of the companies contribute in the growth of society for long term sustainability. So CSR and financial performance of a company are interlinked and same goes with corporate social disclosure.

IC 15/101

Exploring the Impact of Integration in Corporate Bond Markets Using Multiple Regression with Error Correction

Ruchira Panda, Dinabandhu Bag
National Institute of Technology,
Rourkela, Odisha

This paper relates to study of bond market integration by using yield based measures of integration. It also provides insight into the impact of variation in yield which caused by developments in related markets and indicators of financial development multiple regression. Although the broader issue of financial markets integration have been examined in the literature with greater attention to stock markets and more in the global context. There are few empirical works on bond markets integration and fewer works in India. The measures proposed here are also indicative of the money and credit markets. The objective is to assess the current level of integration using multiple regressions and error correction. For fixed-income securities, yield-based measures are more suitable to check the law of one price. Such measures could be based on either nominal or real yields and nominal yields may be preferred since purchasing power parity (PPP) may not hold in the event of capital flow restrictions, trade barriers and transaction costs. The data includes daily traded bonds reported in FIMMDA and NSE archives during ten year period from January 1999 to December 2010. We find significant of bond yield with instruments of other markets such as domestic credit markets, money markets, and external credit markets. We also find relatively lower degree with forex markets and no relationship with any of the commodity or economic indicators.

IC 15/ 136:

Deriving Constant UGDP Using Econometric Models and Calculating Real GDP Using Valuation of Key Natural Resources

Ravikumar Gajbiye, Kavita Laghate
JBIMS, Mumbai University, Mumbai

Wealth Creation is the primary motive for Business, but Unequal Wealth Distribution causes unbalanced economic growth. Economic disequilibrium is a global phenomenon arising out of cyclic nature of Business at macroeconomic level. Shifting dynamics and dormant statics need to be stabilized to a state of macro-constancy to ensure uniform growth.

Macro-Constancy of economy, based on the Law of Conservation of Energy, realigns business processes to Energy dimensions. The authors propose the concept of Universal Gross Domestic Product (UGDP) as a measure to counter business cycles for a stable world economy. The robustness of an economy through interdependence of its various productive sectors is demonstrated by an example using Input-Output analysis and Leontifs's Principle.

Natural Resources are the *raison d'etre* for business; the real wealth of nations. Yet, nations underplay their significance in reporting economic growth. The authors propose to engineer optimization strategies aimed at intelligent utilization of natural resources using regression analysis techniques.

India should lead the economic development of nations by adopting an active role as universal ombudsman under the Triple Helix concept, in controlling, monitoring and regulating the fast depleting stock of the natural resources of the world.

IC 15/163

Occupational Stress among custodians of civic law: An exploratory study

Malini Nandi Majumdar
Indian Institute of Social Welfare and Business Management, Kolkata, India

Avijan Dutta
National Institute of Technology, Durgapur, West Bengal, India

Kalyan Sengupta
Dept of Computer Application, Indian Institute of Social Welfare and Business Management, Kolkata, India

Purpose of the study: The purpose of this paper is to identify the factors responsible for creating occupational stress among West Bengal Police Officials.

Design/methodology/approach – The authors employ an existing standardized scale on Occupational Stressors with some modifications resulting from pilot survey as well as in depth and focus group interviews. The researchers draw on a data set of 310 structured questionnaires covering four hierarchies from Sub Inspectors to ADC/ ASP and above from 5 districts and one commisionerate under West Bengal Police jurisdiction to understand their perceptions about the stressors that police officials often encounter while discharging their duties.

Findings: A confirmatory factor analysis (CFA) has confirmed the five constructs in the data set namely Organizational Stressors (OS), Hierarchical Stressors (HS), Situational Stressors (SS) and Environmental Stressors (ES) and Personal Stressors (PS). CFA suggested a good model fit with GFI>0.9, NFI>0.9, IFI>0.9 and CFI>0. Goodness of fit is initially analyzed through Chi-square value of our proposed model (2.81, df 73, p=.000).

Research Implications: Police are supposed to maintain law and order of the society. On many occasions owing to some internal and external influence they are unable to deal with the situation those results into helplessness at their professional as well as personal life. The paper reveals the major occupational stressors and suggests various measures to deal with these effectively and reduce the negative impact of them.

IC 15/013

Application of Analytical Hierarchy Process for Establishing Customer Relationship Complexity

Harsha Kodandaram
Business Analytics & Research, Fidelity Investments, Bangalore

Bidhan Roy
Business Analytics & Research,
Fidelity Investments, Bangalore

Financial services entities and other service organizations have a large number of customer relationships of varying scope and complexity. Customer service managers (CSMs) at the financial services entity are each tasked with managing a set of customer relationships and providing service to the set of customers. Traditionally, the allocation of customers to CSMs has been done using a subjective approach (i.e., CSM 1 seems capable of managing the customer relationship with Customer X or dividing the number of customers by the number of CSMs) without regard to the specific needs and complex nature of either the incoming customer relationship or the CSM's current customer relationships.

In this paper, we describe the application of Analytical Hierarchy Process (AHP) Optimization Technique in the area of Customer Relationship Capacity model. We tried to define the relationship complexity Score associated with the customer based upon financial plan attributes and personality characteristics. The method described herein also provides the advantage of dynamically evaluating and adjusting the current workload for a CSM to ensure an appropriate level of relationship complexity is assigned to each CSM and also to identify top performers for incoming customer service relationships.

IC 15/ 046

Statistical Modeling Approach Using Unstructured Data

Dipanmoy Roy, Daphne D Costa, Praveen Krishna
HP, Bangalore

This paper aims at finding the key factors that drive the average star ratings at e-commerce site for any product. The text reviews provided by customers on the websites are analyzed. The customer sentiments scores are calculated from the text reviews using text mining and are placed into various categories. A model is built with sentiment scores of the categories as the independent variable and average star rating as the dependent. The method also finds which categories drive customers from lower star ratings to higher star ratings. Shapley Value regression is used to find major drivers driving average star rating while ordinal logistic regression is used to identify drivers that drive customers from low star ratings to higher star ratings. We have used this methodology on customer reviews about HPs Inkjet printers on Amazon.com website and the recommendations have helped HP to improve its star ratings on the website for its Inkjet products.

IC 15/ 053

Measuring Relative Impact of KPIs – A Game Theory Approach

Adinarayana Murty S M, Suresh Venkata Medabala
HP, Bangalore

Customer satisfaction (CSAT) and Net Promoters Score (NPS) are related to key measures of future financial performance for firms. The ability to find key drivers for CSAT & NPS is an important step in any organizations' strategy that leads to high quality of service and long-term relationship with customers. One of the traditional techniques to assess the relative importance of attributes is using standardized regression coefficients. But the presence of high degree of multicollinearity among the attributes leads to imprecise and unstable coefficients and thus assessing relative importance among predictors is a challenging task. To handle this type of problem, Shapley value regression (which was developed by Nobel Prize winner Lloyd S. Shapley), which is a cooperative game theory concept using R-square decomposition is applied. In this method, the R-squared is decomposed into contributions from the different predictors in the model and these contributions are referred as importance measures. Shapley's solution allocates the importance of each predictor fairly even in the presence of high multicollinearity in the data.

IC 15/062

Forecasting Daily Revenue For Financial Planning And Budgeting

Mathur Gautam, Goyal Tarang, Muramalla Neelima
Fiserv, Banglore

Appropriate transactions volume prediction is very crucial for making different Strategic and Tactical decisions at various management levels in any field. This paper analyzes the possible reasons behind daily transaction volume fluctuations for Bill Payment (BPay) and different factors influencing it.

A forecasting solution around the Decomposition Concept of observed/explainable factors was developed, and a statistical Unobserved Components Model (UCM) was fit on the decomposed transactions volume line to forecast daily volumes for BPay in future time periods. Models were built for three main channels – Electronic, Paper Managed and Paper Unmanaged. The results show that error percentage (when forecasted volume was compared to actual volume) was less than 1% for Electronic and Paper Managed Channels whereas for Paper Unmanaged, the error was 3.6%.

IC 15/ 082

Using Neural Networks and Machine Learning to explain Impact of Predictors in a Classification Scenario

Pradeepta Mishra, Tuhin Das
Ma Foi Analytics and Business Services, Bangalore

This paper as an attempt in predicting the winning probability of a deal using two different methods multinomial logistics model and neural networks (NN) model. The objective

of the study is to find out the relevant predictors with their important score for each of the outcomes of the dependent variable. Using Machine learning method and the traditional statistical modeling method we compared the results and found the machine learning method more appropriate. With the advances in emerging fields, it is plausible that newer techniques like NN and Machine Learning (ML) may finally be able to not just provide predictions that are more robust than conventional methods, but also explain how each predictor variable impacts the prediction. The data were analyzed using open source software tool R (R-Studio). One of the key findings of the study was not only to arrive at a better overall model accuracy, but also it looked at the outcome value taking into consideration the business interest. This exercise is an attempt to arrive at a methodology using such a combination to not only predict which deals being pursued by the sales team of a healthcare intermediary are likely to result in a desired outcome (Win), but also show how each factor will impact the outcome, and to what extent. This would not only enable the users to determine which deals to prioritize but also determine what specific actions they need to take to achieve this. This could open new doors for B2B sales analytics in general and the Healthcare sector in particular.

IC 15/ 139

Fraud Analytics….
Like finding a needle in a haystack…

Anupriya Verma
Deloitte Consulting, LLP

An insurance claim prepared with the intention to deceive, conceal or distort relevant information that eventually accounts for health care benefits for an individual or a particular group is defined as fraudulent health insurance claim. It can be committed by anybody – either by a policyholder, a healthcare provider or even its employees.The essential components of fraud include intention to deceive, derive benefits from Insurance industry, preparation of exaggerated or inflated claims or medical bills and malafide intention to induce the firm to pay more than it otherwise would. Devising innovative methods and tactics including pressure tactics, favoritism, nepotism etc. form a part of fraud which is a hazard growing by leaps and bounds since the last decade.

According to a recent survey it is estimated that the number of false claims in the industry is approximately 15 per cent of total claims. The report suggests that the healthcare industry in India is losing approximately Rs. 600-Rs 800 crores incurred on fraudulent claims annually. Health insurance is a bleeding sector with very high claims ratio. Hence, in order to make health insurance a viable sector, it is essential to concentrate on elimination or minimization of fake claims. Predicting such rare fraud events from a sequence of events with categorical features is very difficult and considered as a real world problem. It is a typical segmentation problem to identify high risk claims.

There is a significant opportunity for us to pursue opportunities to continue to increase adoption of predictive analytics in various aspects of healthcare analytics including health care claims fraud detection.

In the study of such rare events data with binary dependent variables, where the success rate is much lesser than the failure rate, we often find it difficult to explain and predict the probability of success or alternatively, the probability of failure. Popular statistical procedures, such as logistic regression, can sharply underestimate the probability of rare events unless

proper attention is paid to take care of the rare event data. There can be many reasons for the relatively low number of successes in the dataset. The dataset itself might be in such a way that the number of successes is much lesser compared to the number of failures or the method used to collect data might have induced some biasness in the data. Example: A sample of 1000 individuals where 66% are male (Data collected at a football match). Several regression methods like Logistic and Probit can be employed to predict the probability of success/ failure. Several modifications like Over Sampling and Under Sampling can be made to the existing dataset to remove the sampling bias induced by the sampling scheme.

In the current assignment, we are only interested in ranking the observations based on the probability of success. We will compare different methods and weigh them against each other.

Two different approaches have been considered to deal with the dataset. Firstly, by using the entire dataset to construct a model. Secondly, select a sample of observations from the original dataset on which the models would be constructed. These samples have been selected using different sampling schemes namely Simple Random Sampling and Stratified Sampling. For each of these data extracts, two different models based on Logistic regression and Probit regression have been built and have been compared against each other. All the methods are then compared against each other using lift curves. Precision and recall values for optimal cut-off have also been determined and have been compared.

Oversampling has been performed by firstly, selecting all the fraud cases in the modeling dataset into the sample and then randomly selecting thrice as many non-fraud cases from the modeling dataset as there are fraud cases in the sample and obtaining the resampled dataset.

Stratification is a process of dividing members of the population into homogeneous subgroups before sampling. Stratification should observe rules such as the strata should be mutually exclusive, every element in the population must be assigned to only one stratum, the strata should also be collectively exhaustive and no population element can be excluded. Simple random sampling or systematic sampling is then performed on each stratum

In the current study, Industry group has been identified as the variable to divide the entire population into strata. The proportion of non-fraud cases in each stratum in the original population has been noted. Non-fraud cases are selected such that the relative proportion of non-fraud cases in the sample remains the same as in the entire population. Three different variations of Stratified sampling have been tried.

Variation 1: Non-fraud cases are drawn in such a way that the ratio of the total number of non-fraud cases to the total number of fraud cases (fraud ratio) in the sample is 3.

Variation 2: Non-fraud cases are drawn in such a way that the ratio of the total number of non-fraud cases to the total number of fraud cases (fraud ratio) in the sample is 1.

Variation 3: A variation of stratified sampling where the fraud ratio is approximately 3, but the number of observations sampled from each industry code depends on the fraud ratio of each industry code.

The average fraud ratio in the sample in variation 2 would be around 50% whereas in variation 1 and variation 3, this would be around 25%.

All the sampling schemes suggest that logistic and probit produce similar segmentation results. When we look at the rank correlation coefficients between the estimated probabilities for all the methods, we observe that all correlations are very high. This suggests that the ranking of observations based on different methods remains pretty much the same irrespective of the method we employ.

For every method, the cut-off value beyond which applicants are flagged as fraud is determined and precision recall values have been calculated at these cut-offs. By comparing the precision and recall values obtained in each of the methods show that segmentation is not affected by the sampling scheme or the regression technique and that inferences can be made using the entire population directly.

IC 15/141

Location Analytics – Store Location Analysis

Kalyani Dacha and Kranthi Ram Nekkalapu
Deloitte Consulting LLP

Retail chains compete against one another in terms of reaching the customer. While companies often try to attract customers by targeting them with the right set of offers and promotions, it is imperative that they are located in a locality that would best cater to the needs of the customers. Market domination comes not just by how they reach the service is provided but also by how reachable they geographically are to start with. Choosing the right location to open a new branch plays a huge role in the success/failure of the branch.

The demographic characteristics of a location play a huge role in forecasting the revenue that can be generated from a location. This total revenue can be understood to be split between different competitors that share the market space. In addition to competition from other retailers, there will be competition from the retail chain's own branches in case there are multiple branches within the same locality. Choosing among different locations to open a new branch is a matter of comparing the expected incremental revenues at different locations given a new unit comes in that location.

In the current study for a large retail chain in the US specialized in renting out electronics, furniture and other items to people, using predictive modeling, we identified factors that would contribute to the sales of the store. In addition, we also quantified cannibalization caused to an existing store by the opening of a new store in close proximity. Combining the two estimates, we calculated the total incremental sales for the retail chain. Different sites that have the potential to host the new store have been scored using the current model and the top sites have been identified. Several statistical techniques like multivariate regression modeling, Random Forests, Lasso Regression, Clustering and Association rules have been used in the process.

In this extensive data analysis, more than 12000 variables for nearly 3000 stores have been analyzed to determine the factors that best describe a successful store environment. Identified the store characteristics consistent with success and ascertained myths to dispel. It is expected that approximately $20M will be generated in annual incremental revenue by year 5 from opening up to 40 stores in more attractive locations. There will be an additional $5.5M contribution in annual revenue by year 5 from retaining more promising closure candidate stores. The effort also paved way to provide foundation for efficient processes to increase store openings, leading to additional revenue gains.

IC 15/018

Study of Dimensions of Consumer Characteristics of Online tourism companies in India

Nidhi Phutela
Lovely Professional University, Jalandhar

Hirak Dasgupta
Balaji Institute of Modern Management, Pune, Maharashtra

Analysing customer perceptions is an important requirement not only for offline stores, but for online stores as well. Study of customer perceptions helps the company in understanding the requirements and expectations of the customers in a better way. This study is an extension of the researcher's previous study of identifying antecedents of consumer trust for e-tourism companies in India. This study tries to explore the relationship between factors that build trust of the consumers on the online companies and their satisfaction level. Apart from studying the preferences and expectations of the consumers from e-tourism companies in India, the researcher has also tried to study the customer perceptions in context of online shopping in general. Data was collected by the researcher from 367 tourists (customers) who plan their tours online in Delhi NCR. Hypothesis testing has been done by using ANOVA and t-test. Findings of the study revealed that that amongst seven Consumer Trust dimensions, only "Promotional deals" and "Price Competitiveness" have a significant positive impact on "Customer Satisfaction", whereas the impact of other five variables are not significant.

IC 15/ 003

What's a Cricketer's Worth? : Predicting Bid Prices for Indian Premier League Auctions

Siddhartha K. Rastogi
Indian Institute of Management Indore India

Indian Premier League is a tournament of twenty-over format cricket, played between teams representing different cities of India. It started in the year 2008 and has established itself over past six editions as a grand annual sports marketing affair. The team franchises as well as cricketers are auctioned every season under certain conditions. Despite such wealth of information, the studies on IPL auctions are rare barring the three cited models. The present paper studies the results of the year 2011 English-style auction of cricketers and recalibrates the old yet most accurate model so far presented by Rastogi and Deodhar (2009). Both the models use ordinary least square method of regression albeit with different variable. The old model lacks predictive power, whereas the recalibrated model presented in this model displays better predictive capability. It also succeeds in reducing overall predictability gap as well as stands significantly parsimonious *vis-à-vis* all the previously proposed models. Further, the final model presented is applied on 2013 and 2014 auction data to show superior results.

IC 15/009

Review of efficiency of k-means algorithm on studies related to cardio vascular diseases

Avijit Kumar Chaudhuri
Technoindia – Banipur, West Bengal

D. Sinha
Indian Institute of Foreign Trade, Kolkata – 700091

K. S. Thyagaraj
Department of Management Studies, Indian School of Mines

Medical science industry has huge amount of data, but unfortunately most of this data is not mined to find out hidden information present inside the data. Advanced data mining techniques can be used to discover those entire hidden pattern inside the data. Different models can develop from these techniques which will be useful for doctors to take effective decision. The k-means is the simplest, most commonly used and good behavior clustering algorithm used in many applications. It has been observed that conventional k-means algorithms are sensitive to the initial cluster centers, and tend to be trapped by local optima. This may result in TYPE I error causing potential patients, say, of Cardio vascular Diseases (CVD) unattended. The study aims at using step-wise clustering approach to identify different age groups prone to CVDs for different combinations of variables with k-means algorithm.

IC 15/065

Efficiency and Performance of Manufacturing Sector in India: An Application of Data Envelopment Analysis

Shaik Saleem, G. Somasekhar, M. Srinivasa Reddy
S. V. University, Tirupati.

To compete in this globalized world, productivity of manufacturing sector is crucial factor. The significance of productivity in increasing national welfare is now universally recognized. In every county, industry or organization, the main source of economic growth is a result of an increase in productivity. Nowadays, it is widely accepted that productivity is a key performance benchmark for firm involved in the manufacturing sector. This is because improvement in productivity is related to increased profitability, lower costs and sustainable competitiveness. In this paper an attempt is made to assess the performance and efficiency of the manufacturing sector of India based on financial variables for the period from 2008-09 to 2012-13. Data Envelopment Analysis method is employed to assess the efficiency of sub-sectors by considering operating expenses and Net Fixed Assets as inputs, and sales and profit after tax as outputs. The study highlighted Consumer goods sector, Construction materials sector and Machinery sector are consistently efficient and their performance in terms of return of investment is comparatively better than the rest of the sub-sectors. Further super efficiency model highlighted construction material sector as the most efficient sector. Spearman rank correlation and concordance analysis showed stability in efficiency rankings of subsectors over the period of time.

IC 15/ 079

Identification of Critical Risk Factors for Nuclear Power Projects in India using Risk FMEA

Rohan Dudam, Balkrishna Narkhede, Yogesh Ghadage
VJTI,Mumbai

Rakesh Raut
NITIE,Mumbai

Nuclear power is a reliable, do not have carbon emissions and is a most concentrated form of energy. The major drawback is a large amount of nuclear wastes are generated and disposal of these wastes is a major challenge. There is huge demand for electrical energy to accelerate the growth in the field of industrialization, agriculturization and urbanization.To overcome the crisis of electrical energy, large numbers of nuclear power projects are being under construction. Uncertainty is one of the major inherent difficulties in Nuclear Power Plant (NPP) projects. NPP projects are more risky.

This paper identifies critical risk factors in NPP projects in India. It is observed that identified risk factors are mainly related to construction, organization, design, government bodies, natural disasters and social issues, etc. It is observed that these risks are spread through the whole project life cycle and many risks occur at more than one phase. With identified factors a survey by experts has carried to collect responses related to probability of occurrence, impact on project & detection difficulties of the risks. This analysis facilitated to scrutinize the high probability factors. Identifying and mitigating project risks are crucial steps in managing successful projects. This article proposes the Failure Mode and Effects Analysis (FMEA) format to quantify and analyze project risks. The new technique is labelled the project risk FMEA (RFMEA).

IC 15/014

Exploring the Big Data Analytics imperative: Value creation versus Value capture.

Surabhi Verma, Som Sekhar Bhattacharyya
NITIE, Mumbai

Big data analytics has emerging as an excellent exemplar (for both as a tool and a technique) for managing processes and as a governance mechanisms which could harness insights from such voluminous data. This data when analyzed could create value. Nonetheless, the very concept of big data analytics has represented a context that required the managers of firms to rethink their strategy. The purpose of this research paper was to identify how and why big data analytics helps firms creating value or capturing value. Exploratory qualitative content analysis was undertaken for this study. The study instrument was in-depth semi-structured open ended interviews with big data analytics experts. The findings justify big data analytics could create and capture value in organizations. Organizations could gain competitive advantage by using big data analytics to improve management process, operational process and strategic decisions. The study theoretically suggested the elements that determined the value creation and value capture by BDA. The paper is one of the first known empirical study of value created and captured by BDA in organizations across sectors.

IC 15/ 043

Behavioural Analytics in Telecom

Vani Haridasan
SSN School of Management, Chennai

The telecom industry is witnessing an increase in adoption of customer experience management solutions, social media tools and analytics solutions to understand the customers and provide them with improved and customized offerings. Focusing on customer experience would enable telecom service providers to respond more effectively to customer requirements, build customer loyalty, and create a stronger value perception in the minds of customers. Behaviour is increasingly recognized as a key component in business intelligence and problem-solving. Human behaviour has been increasingly highlighted for pattern analysis and business intelligence in many areas such as customer relationship management, social computing intrusion detection, fraud detection, outlier detection and group decision making. This paper aims at understanding the behavioural aspects that play a greater role in understanding customer loyalty, improving service quality and thereby enhancing customer experience. The study focuses on improving the customer experience management practices of Indian Mobile Service Providers and identifies the latent factors which influence the customer loyalty in terms of advocacy, repeated purchase and defection. The findings of this study emphasizes on focusing on the underlying behavioural aspects that would further strengthen the current knowledge in determining the relationship among perceived service quality, service loyalty and customer loyalty. The path models highlight the fact that loyalty is essential for service business survival and this study is an early attempt to improve them.

IC 15/164

The fundamental plot for health microinsurance success

Kanish Debnath
Indian Institute of Management Ahmedabad

The core products for the financial inclusion of poor households are savings, credit and insurance. While savings and credit, commonly aggregated under microfinance, have been successful, insurance products have struggled to be viable. Health insurance, in particular, face high incurred claim ratios1. This happens mainly on the account of insurance appearing resourceful only at the time of need. Many micro-insurers have tried to counter this problem of adverse selection by pre-screening clients, reducing coverage to only hospital stays, and increasing the premium amount but with not much success. In the light of the above, many health microinsurance schemes need to be subsidized either by the government or through donor support. In order to find how adverse selection can be countered properly, this study explores the work of a community health microinsurance organization working out of Pune and Mumbai, who have become self-sustainable. At the same time, they do not exclude any individual based on age, nor do they have differential premiums according to age. A review of 5,255 health insurance claims for the recent three years of its operation finds that there is no systematic exclusion of any age group. Further there is a healthy participation across all age groups. It appears that their linkage of microinsurance with microfinance, where it is compulsory for the borrower to insure self and is also motivated to insure her family members, has reduced adverse selection.

[1] Incurred Claims Ratio = Incurred Claims / Earned Premium

IC 15/ 052

Understanding Big Data Analytics and Its Application for Brands – An Exploratory Study

Rasananda Panda, Diwakar Sharma
MICA, Ahmedabad, Gujarat

This exploratory study aims to understand the Big Data Analytics to improve the products or services delivered to the customers. The main objective is to see the applications of big data analytics in helping the brands deal with the huge customer base and their ever-changing demands.This study includes the discussion on platforms used to generate insights from the data and their relevance to the brands. It will act as a guide to the marketers and business analysts who need to assess the big datasets and use analytics tools to take business decisions. The brands have realized the need to monitor and use the huge data being generated by the consumers to have the competitive advantage in the industry. Hence the study will benefit the mid-level managers to understand the importance of the Big Data analytics and the technological aspects of the platforms used. The study includes three cases of the implementation of Big Data in large organizations to understand the application of the technology in the industry. Paper also discusses the applications of Big Data in different domains such as marketing management, risk management, supply chain management and operations management.

IC 15/ 058

Socioeconomic Implications of Wind Energy – A Perspective from Gujarat

Amit Purohit, Naran M. Pindoriya
IIT Gandhinagar, Gujarat

India possesses an abundant reserve of renewable sources of energy. These renewable energy sources may not limit to wind, solar, hydro and bio-mass. In order to harness energy from such renewable sources, a lot of policies and infrastructure has been planned and placed by the Government. However, implementation and optimal utilization of the renewable sources of energy has not been substantial as such sources are highly intermittent in nature. In this paper, we have selected wind energy, one of the most intermittent sources of energy, its trends, technical and operational factors that play a key role in generation of electricity. The study is limited to policies and wind trends of Gujarat, one of the richest sources of wind energy resources in the country. The study aims at creating a model to forecast wind energy generation to counter the variability of wind energy and minimize financial as well as operational hassles for the wind energy generator as well as the system operator. The wind profiles, wind energy produced and various institutional factors are considered. Linear regression is employed to predict the wind energy generation. Indian Electricity Grid codes, working dependencies, errors encountered during statistical modeling and respective losses to the wind energy generators are taken in consideration for a coherent understanding and analysis of implications of wind energy.

IC 15/074

Assessment and Prospects of Big Data Analytics in Public Policy in India

Sachin Kumar
Delhi University, Delhi

Rampal Singh
Deen Dayal Upadhaya College, Delhi Universty, Delhi

Saibal K Pal
Defense Research Development Organization, Delhi

Indian is emerging economy with potential to grow at double digit GDP growth rate. Sustainable growth and sustainable development require investment in technological innovations in governance product as well as governance processes. India is investing immensely to harness the full potential of resources. Incorporation of information and communication technology into governance process provides mainly two benefits. One is delivery of governance products and services in time bound, transparent manner with accountability. Second is collection of information from the actual consumer about products and services provided by the government. Second function may play a better role in improving effectiveness and efficiency of policy. India is country of great diversity with many traditions, cultures, food habits, environmental changes, flora and fauna, people perceptions and their employment habits. This poses challenges in front of the policy makers as to deal with vast amount of drivers information to analysis and formulate policy which is effectives and efficient in governance products delivery. This research paper discusses failures of policy making processes, reasons, problems with exiting systems, challenges posed by diversity and population in Indian context. This study first analysis the problems and challenges in policy domains at present and prospects of solution for solving them using machine learning and big data analysis by a proposing framework. This framework tries to improve the policy making cycle problems and provides better prospect for the future policy making by analyzing scenarios of the coming times.

IC 15/116

Context Development For Internet Security In The Big Data Analytics World

Nitin Varma, Pradip Kumar Bala
IIM, Ranchi

For the first time computer launched foreign assaults on U.S. infrastructure were ranked higher in the U.S. intelligence community's annual review of worldwide threats than worries about terrorism (Dilanian, 2013). That alone points to the importance and criticality of internet security for national as well as international security.

As the world heads towards massive penetration of mobile devices, as the era of IOT (Internet of Things) dawns and as the world ultimately transitions from early-Pervasive to Ubiquitous computing, the internet and therefore internet security are bound to be massively and inextricably enmeshed with human existence, while multiplying by many orders on complexity, sensitivity and criticality.

This paper traverses the vast landscape of internet security quickly to construct and provide a first ever big picture of the internet security context. It finds that most existing paradigms,

already insufficient to provide sustainable internet security, are going to fall woefully short either on protection or coverage, or even be rendered obsolete. The paper then discusses the necessity for Big Data analytics and machine learning. Taking cues from the shortcomings pointed out by early Big Data practitioners, the paper recommends metadata and context driven Big Data practice for security analytics. The importance of machine learning for delivering context and context-awareness, is emphasized, to emerge with a set of potential winning approaches centering on malbehavior prevention and resilience. A contextual framework for malbehavior that may serve as the core of context-aware security systems, and a wider contextual framework for sustainable internet security are also proposed.

IC 15/ 119

Data Analysis Using Probabilistic Graphical Models

Heena Timani
Ahmedabad University
Ahmedabad, Gujarat, India

Mayuri Pandya
Bhavnagar University
Bhavnagar, Gujarat

Data mining is a multidisciplinary field, drawn from varying areas as artificial intelligence, database technology, data visualization and machine learning. Using a combination of machine learning, statistical analysis, modeling techniques and database technology, data mining finds patterns and subtle relationships in data and infers rules that allow the prediction of future results. Data mining offers tools for discovery of relationship, patterns and knowledge from a massive database in order to guide decision about future activity. Probabilistic Graphical Models also known as Bayesian networks are popular and powerful tool in data mining. They have many applications in commercial decision support. Typical applications include market segmentation, customer profiling, fraud detection, evaluation of retail promotions, credit risk analysis and banking sector. In this paper the knowledge discovery from various databases using Bayesian network and Bayesian classification techniques are discussed. Practical machine learning data mining open source software are used for knowledge discovery and data analysis.

IC 15/W-1

Bayesian Data Analysis

Karthik Sriram
Indian Institute of Management Ahmedabad

The work shop will introduce participants to fundamental ideas in Bayesian data analysis. A basic understanding of probability is assumed although required key ideas will be briefly discussed. Starting with conditional probability and Bayes theorem, we will discuss the importance of prior information in some data analysis problems and how such prior can be combined with data to obtain a posterior distribution. We will discuss different types of priors. We will work through some examples using excel and/or R software.

IC 15/W-2

Text Mining: Theory, Methods and Applications

Srikumar Krishnamoorthy
Indian Institute of Management Ahmedabad

It is estimated that more than eighty percent of the data generated by organizations are in unstructured textual form. The advent of Web 2.0 has further fueled the proliferation of user generated textual content about organizations products, services and policies. However, most of the analytical models used by organizations today for decision making largely rely on structured data sources. Text mining is an emerging and exciting research area that aims to help organizations leverage the rich goldmine of unstructured textual content. It is an interdisciplinary field that uses fundamental ideas from statistics, information retrieval, data mining, machine learning, and natural language processing. This tutorial primarily aims to introduce the participants to the fundamentals of text mining, describe the state-of-the-art methods for key text extraction, mining and sentiment analysis tasks, and discuss several real world business and scientific applications. The tutorial will also present recent research trends and future research directions in the area of text mining and analytics.

IC 15/ 031

Predicting risk of Rejection in Non-Submitted Claims

**Ayush Dubey, Amar Jain, Venkatesh Pagidimarri, Shivang Singhal,
Vamsi Chandra Kasivajjala, Suvomoy Sarkar**
Enlightiks USA

A 2011 study by the U.S. Government Accountability Office found that claim denial rates vary significantly among states and health insurers. Of the small number of states tracking such information, denials ranged between 11 percent and 24 percent of claims

The following are results from the National Health Insurer Report Card (NHIRC) years 2008-2013 that address denials.

Percentages of claim lines denied: What percentage of claim lines submitted are denied by the payer for reasons other than a claim edit? A denial is defined as: allowed amount equal to the billed charge and the payment equals $0.

Hence if a system is in place to predict the risk of rejection much before the claim is actually submitted, the percentage of rejection of claims for a provider can be reduced and its revenue improved.

The problem is modeled as supervised classification problem, training data consisted of all the labelled claims and model accuracy is checked on test data set. Overall data set was spread among multiple tables, proper understanding of features and data preprocessing lead us to combine the tables in sensible manner over which model can be built. Multiple models are built with proper tuning, and their performance are compared. We found Decision Tree and Logistic Regression as top models which were giving around 90% accuracy.

IC 15/055

Retail Credit Risk Model Validation: Performance and Stability Aspects

Varun Aggarwal, Sanjukta Rej
Gurgaon Haryana

Like a model without sufficient validation may only be a hypothesis (Stein 2002), a hypothesis without reason may just be an intuition. With the growing competition for acquisition of new and retention of existing retail banking customers, all stakeholders (financial regulators, auditors and business leads) have shifted their focus on accurate and timely validation of risk models. Intuitions and hypotheses are formulated into models at the time of development using statistical techniques, which get implemented into production system only after rigorous out-of-sample and out-of-time validation. Real test of the model, however, begins after implementation. In this paper we attempt to cover various aspects of validation of credit risk models after they get implemented.

This paper is not an attempt to provide a laundry list of metrics but a conscious effort to add value to existing literature by focusing on application of appropriate methodologies and computation of relevant metrics for reviewers to comment on aspects of model performance and stability. Our approach looks at different dimensions of a credit risk model for detailed assessment. For such comprehensive validation we consider a) Power Testing b) Calibration and c) Sampling strategies

All along the paper, we provide the readers with key validation aspects and corresponding metrics which serve as critical inputs for a model evaluator to approve or decline usage of implemented scores. Towards the end, the paper presents few illustrative snapshots of a recommended tool useful for analyzing model performance and stability results on regular basis.

IC 15/103

Identification of Best Practice HRMs in NTPC With the help of Statistical Tools

E.Nanda Kishore
NTPC Ltd. Ramagundam

Organisation is collection of individuals joined together to meet the purpose given by society that could not have been done by one individual. Many researchers tried to understand the relationship between the Human Resource Management practices and performance of the individuals and organisational performance. Mark A Huselid and Brain E Becker in their paper "The strategic Impact of High Performance Work systems (1995)" viewed "High Performance Work Systems" as a key strategic lever both as means to develop and sustain core competencies and as a necessary condition for strategy implementation.

There is limited work that is exploring direct relationship between HR Management systems and organisational performance (Either intermediate or strategic levels) is available. In recent years, there has been much interest in the notion of 'best practice' human resource management (HRM). Sometimes this is referred to as 'high performance work systems' (Berg 1999; Appelbaum *et al* 2000), 'high commitment' HRM (Walton 1985; Guest 2001a, 2001b) or

'high involvement' HRM (Wood 1999a). Whatever the terminology, the idea is that a particular set (or number) of HR practices has the potential to bring about improved organisational performance for all organisations. This paper tries to explore the set of best practices in the area of HR management those lead to intermediate levels of organisational performance such as Low attrition, Peaceful Industrial Relations climate etc. This study was done in NTPC by defining ***"Best Practice HRM" as the HR Practice that is significantly impacting positively more than 5 HR Results*** with the premise that HR results influence other outcomes of organisation such as Economical and societal results also. For this purpose 26 Practices/Policies of HR of NTPC have been considered and their influence on each of the HR results is checked. Following HR Practices are found as "Best Practice HRM" in NTPC.

- Pay Structure.
- Job Security.
- Township Facilities.
- Safety.
- Recreation Facilities.

IC 15/140

How Predictive Analytics can help E-Retailers?

Abhishek Kuppili and Kalyani Dacha
Deloitte Consulting, LLP.

The rapid emergence of e-commerce sites has made the web-space an exciting and interactive business platform for producers, marketers and consumers. In their quest for increasing their share of customer's wallet online retailers are battling aggressively by offering better customer experience through personalized recommendations, customized promotional offers, etc. And to enhance their customer experience and stay ahead of competition online retailers are adopting advanced technologies and capturing information on their customer's online activity on their site by various means resulting in an explosion in the volume of data being collected. They are eager to understand how this big-data can be used to increase customer's click rate and to ensure customers aren't attiring.

The analysis of this huge volume of data presents a lot of opportunities as well as challenges. In this paper we discuss about our experiences in working with this data for a Korean online retailer. We present our approaches/solutions on following three topics which have helped the client in enhancing their customer experience.

1. Improving the performance of their recommendation engine which resulted in 76% increase in click-through-rate.
2. Develop customer shopping pattern engine to score customers in each of the 34 client identified shopping patterns. These scores will in turn be used by client for selective promotions through E-mail/SMS.
3. Identifying if context like time, weekday/weekend, weather, location, etc., affect the browsing behaviors of customers.

We are going to propose how E-Retailers can make use of the big-data they are collecting in this paper.

IC 15/ 128

To extract and relate Human Sentiments using Text Mining through Bayesian Learning

Chandramouli Banerjee
Skybits Technologies, Kolkata

A huge amount of text data is available through different sources across the internet where the users write reviews relating to the product and service related features. The text representation is a function of the inherent topics which reflects the reviewers perception.The aim is therefore to address the challenge to discover the hidden sentiments from these reviews and classify them as well predict the performance of the future reviews through rank ordering between and within the positive and negative sentiments.Addition to these latent structures the data also depends on a variety of nuisance parameters that are irrelevant to the task.

The methodology adopted is Bayesian Aspect Mining of Text data using the concept of mixture of Stick Breaking Processes Representation thereby leading to Hierarchical Aspect Sentiment Modeling through the technique of Dirichlet Processes embedded with a recursive Chinese Restaurant Process. The approach uses nested stick-breaking processes to allow for trees of unbounded width and depth, where data can live at any node and are infinitely exchangeable. One can view the model as providing infinite mixtures where the components have a dependency structure corresponding to an evolutionary diffusion down a tree. By using a stick-breaking approach, Markov chain Monte Carlo methods can be applied based on slice sampling to perform Bayesian inference and simulate from the posterior distribution on trees.This can be fairly extended to infinitely exchangeable mixture processes.

IC 15/142

Making the Data Complete: An approach to Deal with Missing Data Problems

Divya Sharma
Deloitte Consulting LLP, Hyderabad

How the missing values should be treated is a universal dilemma in the world of statistics. The subject has seen a lot of research in the past mainly on the methods of handling missing values.

While in theory there are many methods of dealing with missing values, the actual method to be applied depends on the type of variables in particular and type of data in general. This paper is an attempt to bring together all the methods available for treating missing values in machine learning and statistical data analysis, assessing the situations where these should be applied (or not applied) with some examples and case studies. This is also an effort to explore some of the relatively newer techniques of dealing with the missing values (like multiple imputation and Maximum likelihood recently developed by SAS and R), the advantages they have to offer over other techniques as also the situations where they should not be used. This will be done through comparing and presenting six different approaches to treat missing values on a sample dataset and show how drastically the conclusions of predictive models can differ based on the method of treating missing values applied.

IC 15/ 146

Market Intelligence & Analytics 2.0

Gaurav Kumar
Fidelity Investments, Bangalore

Analytics and market intelligence maybe the buzz words of today, but these concepts aren't new. These can, in some shape and form, be traced back to the evolution of mankind right from the agrarian to industrial to the modern day information economy. Yes admittedly the nature & quantum of data, the sources to trace and track it, and its applications have undergone a huge transformations over the years – none more so than in the last decade. There are five key drivers that necessitate formation of evolved market intelligence & analytics function of the future – digitization & data explosion, changing competitive dynamics, social media & rising interconnectedness, the mobile data & device revolution, and the importance of doing it right. Market Intelligence 2.0 will require a metamorphosed approach to corporate decision making, and will be characterized by eight key attributes. The millions of dollar being poured into it together with its cognizance as a competitive differentiator will see the analytics function getting a seat the decision table at one end and being measured against hard metrics at the other. Shortage of talent together with the need to look at data through innovative lens will necessitate hiring for aptitude and domain knowledge, and mandate agility to scale-up on new tools and techniques. Technology is likely to emerge as a key enabler to make the function more efficient and cutting-edge, as well as drive creation of self-help tools. Digital analytics will become a key pillar embedding itself in all forms of analysis.

IC 15/148

Field Services efficiency within Telecom Network enabled by Analytics

Kaushik Sanyal, Rajan Shingari, Saswata Sarkar
Accenture Services Private Limited, Bangalore

As the gamut of activities within the area of Field Services increases, it is increasingly becoming an important aspect of differentiation within Telecom operators and Network Providers. Today organizations have come to realize that customer satisfaction and enhanced subscriber experience is driven by an intelligent and optimized Field Service operation, rather than it being designated as a cost bucket with a traditional focus on increasing productivity. Thus, the focus on field service has extended to newer aspects of improved forecasting, identifying the root causes of network failure and proper scheduling and dispatching of personnel to get maximum efficiency. In this regard, analytics, specifically aspects of advanced analytics have become an important tool within the Field Services organization to help becoming more effective in addressing present-day challenges.

Any impactful analytical solution would not be effective without a strong work around Root Cause Analysis, Forecasting of Trouble Tickets, and Capacity Planning with the help of workforce management of field force personnel. Advanced analytics is increasingly processing a wide range of data both internal and external that helps in applying these techniques to get meaningful insights. Our talk today will cover how these techniques will help organizations in identifying the value drivers impacting every issue.

In conclusion an end-to-end solution around Field Services will be presented which will highlight the various ways advanced analytics can help operators make smarter interventions within their overall processes to enhance customer experience while retaining an optimized offering.

IC 15/W-3

An Introduction to Directional Data Analysis

Arnab Kumar Laha
Indian Institute of Management Ahmedabad

Mahesh K.C
Institute of Management, Nirma University, Ahmedabad.

In many scientific fields, the measurements are directions. Analysis of such data is substantially different from the standard (univariate / multivariate) statistical analysis. In this workshop some of the fundamental concepts of Directional Data analysis will be discussed. Some commonly used parametric models and methods to analyse circular and spherical data will be introduced. Some real-life examples from different application fields will be given.

IC 15/W-4

An Introduction to Splines with Statistical Applications

Dhiman Bhadra
Indian Institute of Management Ahmedabad.

This talk will outline the basic concepts and applications of splines as a flexible statistical tool for modeling noisy and non-linear patterns, something that is so ubiquitous in data sets collected in real-life.Starting with some motivating examples from real data sets, the talk will explore the underlying concept of splines and why they work. Some important characteristics and properties of splines will be taken up next which will be followed by some applications of spline modeling in real data sets. A brief preview of some statistical softwares used for spline-modeling will also be given. The overall intent of the talk is to make the audience aware of the wide-ranging applicability of this statistical technique in data analysis, specifically when linear regression modeling fails.

IC 15/042

Determinants of India's Food Grain Production: Evidence from Quantile Autoregressive Distributed Lag Model

Debdatta Pal, Subrata Kumar Mitra
Indian Institute of Management, Raipur

This study intends to identify the determinants of food grain production in India by employing Quantile Autoregressive Distributed Lag model of Cho *et al.* (2014). QARDL modelling approach simultaneously captures both the long-run relationship and the associated short-run dynamics across a range of quantiles of the conditional distribution of the dependent variable in a fully parametric setting. The strength of the QARDL model has been shown in the empirical assessment of food grain production using the time series data of rainfall, fertilizer use, and pesticide consumption. It is found that rainfall has an asymmetric impact on food grain production.

IC 15/ 152

Emotional labor Strategies and burnout: Mediating role of asakti-anasakti for ASHA workers

Jatin Pandey
IIM Ahmedabad

Manjari Singh
IIM Ahmedabad

The mediating relationship of Asakti (i.e. attachment) and Anasakti (i.e. non-attachment) between emotional labor strategies (surface and deep acting) and burnout has not been explored in studies before. We bridge this gap through a quantitative model using Structural equation modeling (SEM). Emotional labor represents the effort to bring forth or suppress certain emotions required to perform a job role. Emotional labor is performed through surface acting where inner emotions are not changed and deep acting where inner emotions are changed. Burnout is a psychological state associated with long term exhaustion and low interest in job. Asakti-Anasakti are Indian psychological constructs which are regarded as bi-polar emotional states wherein an individual high in asakti forms emotional attachments more quickly as compared to an individual who is high in anasakti.

This study is particularly critical for jobs where emotional labor plays an important role in effectively and efficiently carrying out job responsibilities. One such role is that of accredited social health activists (ASHAs). We measured emotional labor strategies for surface acting and deep acting through five and three items of (Grandey, 2003). The measure of burnout was contextualized version of the Copenhagen Burnout Inventory (Kristensen, Borritz, Villadsen, & Christensen, 2005) and Asakti anasakti through 40 item instrument of Test of Asakti-Anasakti Anasakti (Bhushan & Jha, 2005). Asakti-Anasakti scale was reduced to five factors using factor analysis. Anderson and Gerbing's (1988) two-stage approach was followed, the first stage warrants examination of the convergent as well as discriminant validity of the measurement model by means of confirmatory factor analysis and the second stage involves investigating the model that have been hypothesized. Results from 116 ASHAs bring forth that TAA mediates the relation between emotional labor strategies and burnout.

IC15/157

The Impact of Paid and Earned Media on Consumer Choice: A Real-Time Experience Tracking Approach

Bhuvanesh Pareek
Indian Institute of Management, Indore

Rajiv Sinha
Arizona State University

Pulak Ghosh
Indian Institute of Management, Bnaglaore

Hugh Wilson, Emma K. Macdonals
School of Management, Cranfield University

Although both paid and earned media influence choice decisions, practitioners have little guidance on their relative impact, as paid and earned media have largely been studied separately. Furthermore, most research examining multiple media types ignores the individual's perceptual response to exposures, despite its known importance. To address these weaknesses, the authors propose a model that jointly considers individual consumers' exposure and valenced response to media over time, the evolution of brand choice intentions, and actual brand choice or purchase. The model is applied to two categories, choice of political

parties and alcoholic drinks purchases. Data collection uses an emerging method based on text messaging called real-time experience tracking (RET), which allows respondents to report both paid and earned media exposures, and their response to them, as they occur. Results reveal the differential impact over time of paid and earned media, and show the importance of exposure valence and not just frequency, contrary to the predominant practitioner emphasis on share-of-voice metrics.

IC 15/161

When 'Refurbished' offers makes 'New' offers look unfair

Pinaki Roy, Arnab K Laha, Neharika Vohra, Sanjeev Tripathi
Indian Institute of Management, Ahmedabad

Selling refurbished goods at discounted rates has become popular as an option for managing returned and damaged goods. In this article we focus on the effect of discounted refurbished product offers on the price fairness perceptions of the corresponding new unit offers. Building and comparing findings from previous studies on price unfairness, we determine whether the effect of refurbished product discounts is different from regular product discounts. Further, we explore the effect of selling strategies such as seller identity and warranty strength. Using a survey based between subjects experiment design and regression based path analysis, we explore the above phenomenon through two 2X2X2 experiments using a student based sample. We find that seller identity, warranty strength, and trust in the seller of the refurbished unit offer have significant effect on the price unfairness perceptions formed with regards to the corresponding full priced new unit offer. Prior research has found a reduction in willingness to pay for the more expensive new unit offers when customers are exposed to discounted refurbished unit offers. The present study finds price unfairness as a mediator for this behaviour. In doing so it connects a predominantly analytical and mostly operations management based literature to marketing based concepts. The findings hold insights into the effect of pricing and warranty decisions of original goods vis-à-vis refurbished goods.

IC 15/166

Real estate investment selection and empirical analysis of property prices: Study of select residential projects for Gurgaon, India

Mridul Upreti, Aakriti Bhatia
Jones Lang LaSalle Investment Advisors Pvt. Ltd.

Sanjay Sehgal, Piyush Pandey
Department of Financial Studies, University of Delhi

The paper studies the residential micro market of the Gurgaon region of the Delhi NCR, India to identify key determinants of real estate investment selection and perform empirical analysis of property prices. Primary survey suggested that developer's goodwill is the most important factor for investors in case of under construction residential properties (forward projects). Other factors include location, amenities, project density and construction quality. These factors enjoy almost equal importance for selecting completed (spot projects). The factor information can be used to construct property quality rating classes. High risk adjusted returns were provided by high quality spot projects and low quality forward projects. Long run equilibrium relationship was observed between spot and forward prices with the former playing the lead role. GDP and non-food bank credit are the macroeconomic variables that

can predict property prices. Highest pre-tax Internal Rate of Return was observed for forward projects in first quarter holding itself while for spot projects it was around 8th quarter. The research has implication for property developers, real estate investors and market regulators. The study contributed to real estate investment literature for emerging markets.

IC 15/033

Predicting Risk of Diabetes in Non-Diabetic Population

Shivang Singhal, Amar Jain, Venkatesh Pagidimarri, Ayush Dubey, Vamsichandra Kasivajjala, Suvomoy Sarkar
Enlightiks USA

Diabetes Mellitus is a chronic debilitating disease affecting a major population of the developed and developing countries. The occurrence of type 2 diabetes mellitus (T2DM) is rising rapidly among middle-aged American adults. It has been estimated that the prevalence of diabetes in the United States increased from 7.3% in 1993 to 7.9% by the year 2000, and greater frequencies are forecast for the future Prediction of chronic conditions like DM that have a definable onset can help to guide interventions and health policy development. Prediction of future incidence of this disease will enable adequate fund allocation for delivery of care to be planned. This white paper discusses the approach and statistical models used to predict diabetes mellitus in a population with unknown status for diabetes. The prediction is for at present and at three months' time frame allowing a practitioner to pick up patients at risk of acquiring diabetes. The problem is modeled as supervised classification problem, training data consisted of all the labelled patients and model accuracy is validated on test data set. Multiple models are built with proper tuning, and their performances are compared. Support Vector Machine and Random Forest have better accuracy compared to other models.

IC 15/ 076

Analytics Driven Cross Sell: How Analytics boosted cross sell ratios for one of India's largest DTH companies

Ajay Kelkar, Bharath V., Akash Agrawal
Hansa Customer Equity Pvt. Ltd., Mumbai, Maharashtra

Girish Vardarajan
Tata Sky Pvt. Ltd, Mumbai, Maharashtra

In India, content distribution in television industry is a high volume and low ARPU (Average Revenue Per User) industry. Even though Direct to Home (DTH) service providers enjoy a higher ARPU owing to their higher penetration in the Urban markets, they still report ARPU to be less than 5-6$ per month. With such low revenue per user, profitability is the biggest challenge faced by such service providers. Every DTH provider usually has more than 20 value added offerings including but not limited to interactive channels, movies on demand, mobile applications, HD services etc. India being so diverse in its culture and the television consumption patterns and due to wide variance in the economic strata of the subscribers, it becomes very important to identify the right product for the right customers. We built next product purchase algorithms called Subscriber Preference Modeler (SPM) for one of India's largest DTH Company. These algorithms aim to cross sell bundles as well as a number of unique services such as gaming applications. We have described how analytics & smart use

of Big Data played a leading role in improving cross sell and reducing churn rates to industry best levels. We have articulated the techniques used on imperfect data and how they impacted business. In this paper, we have described how unique attributes of ethnicity and innovative analytics techniques were used to build the SPM engine which has now become the principal cross sell engine for the DTH Company.

IC 15/149

Customer Experience Management Powered by Advanced Analytics

Vivek Saha, Sachin Sehgal
Accenture, Gurgaon, Haryana

Customer experience management becomes a C-level priority and coordination effort as CSPs are changing business practices and business models that require alignment of technical KPIs with customer experience and perception KPIs. This has a trickledown effect on front- and back-office requirements.

Most CSPs have launched first- or second-generation customer improvement initiatives related to operations support system (OSS) and business support system (BSS) to better leverage the wealth of data available and gain improved insights into customer behavior. They historically focused on improvement in network infrastructure to deliver optimal network performance for their wireless, wi-fi and wireline businesses. Over the time, they have realized that this approach is not fool-proof since there is a gap between the customer view of service quality and traditional network operations view. Hence, CSPs are looking to understand performance from a customer's perspective and optimize their network performance accordingly. This will help them in differentiating themselves from their competitors in the market, enhance customer retention and drive revenue growth.

Advanced analytics is playing a defining role in real time measurement, optimization and improvement of network customer experience. This poses significant challenges and provides opportunities since it involves processing huge amount of streaming data generated from different network machines (like Nodes, Access Points, Towers etc.) and client devices. Additionally, advanced machine learning algorithms can be applied on this data for focused results. Through an integrated approach of application of analytics on data generated across end-to-end components of network, end user experience can be measured, evaluated and optimized. This can help in better decision making.

In this talk, we will share Accenture Customer Experience Suite powered with advanced analytics intellectual capital that are focused towards optimization of end- to-end network performance for CSP's wireless, wireline and Wi-FI businesses. These assets/IPs are used to measure real time KPIs of network performance to enable insight generation, and identification of bottlenecks. We will also discuss application of machine learning techniques on data generated from machines to solve host of problems related to network operations like – prediction of network issues before failure, optimization of customer experience and forecasting demand for long term Network infrastructure planning. We will also share case studies that will help the attendees understand the Network KPIs, implement strategies and machine learning techniques in context of practical examples.

IC 15/154

Transforming Big Data Into Actionable Insights

Rohit Chauhan
Mastercard Advisor, U.S.A.

In order for big data to be refined and made actionable, it will require the pressure of advanced analytics and then marketing adoption of the learnings found within the unrefined "big" source. This paper will show that four emerging data trends will enable the necessary evolution from big data to advanced customer segment analytics. This new approach will be enhanced by transactional data and secured by new technology and compliance regulations. Data must transform the customer experience, create operational efficiency and define new business models.

Four emerging trends will drive the growth of big data and its transition to advanced analytics: mobile, social, cloud-based data and the Internet of Things. Companies must apply the right analytics pressure to create diamonds of data rather than unrefined blocks. In order to make data actionable it will need to be secured, subjected to proper compliance and enhance the consumer experience. Data and technology are allowing us to answer very difficult questions that have not yet been asked e.g., purchase sequence.

The complete data set that emerges from the four sources mentioned as well as Internet traffic analytics has come very close to producing a 360 degree picture of the customer. In order to see that picture with clarity, transactional data is critical. Transactional data will produce new levels of actionable customer segments that define the future of advanced analytics. Prescriptive analytics represent the nirvana state for the discipline. Companies should strive to approach that ideal rather than become consumed with admiring their past descriptive analytics successes. Integrating analytics at key points of intersection in real-time will allows maximization of the benefits of insights to the bottom line.

IC 15/162

Applications of Statistical Tools and Techniques for Advance Data Analysis

CUBIC QUALITY , Bangalore

Quality products and services are fundamental to customer satisfaction. Industry will distinguish their products and services as exceptional in ways that the customer values. To achieve business excellence and remain leaders in industry, continuous improve of products, services, and processes at a rate exceeding that of competitors is required. Minitab serves everyone who practices or teaches data analysis. Our mission is to help them achieve their goals by providing high quality statistical software and related products and services that are accurate, reliable, and easy to use. In between our session we will present different business scenarios of how Minitab 17 has been playing vital role globally across the industry to achieve business excellence and also how Minitab 17 has been changing the way people analyze problems for more than 40 years.

IC 15/153

Design of Near-optimal Irregular Fractional Plans Satisfying Multi-optimality Criteria

Prasun Das
Indian Statistical Institute

In an industrial set-up, conditions of orthogonality and optimality of a statistical experimental design often get violated. This paper attempts to find an optimal design which performs uniformly better with respect to multi-design-optimality criteria. Initially, using non-dominated sorting genetic algorithm (NSGA-II) with single D-optimality criterion, a near-optimal design is searched for linear, interaction, quadratic, pure quadratic and some pre-specified models. The solutions obtained are verified with the existing upper bounds. In the second phase, Pareto-optimal solutions are obtained with multi-design-optimality criteria. The method of comparison is illustrated with an example along with its demerits in terms of design efficiency.

IC 15/108

A New Method of Construction of E-optimal Generalized Group Divisible Designs With Two Groups

D.K.Ghosh
Saurashtra University, Rajkot, Gujarat

Alex Thannippara
M.G.University, Kottayam, Kerala

Sreejith V.
University of Kerala, Thiruvananthapuram, Kerala

S.C.Bagui
University of West Florida, Pensacola, USA

In the class of equireplicate, proper and binary designs, the balanced incomplete block designs are the only variance-balanced designs. These designs have very strong optimality properties, whenever they exist. If we restrict our attention to only equireplicate, proper and binary designs, then it will be seen that BIBDs do not exist for every combination of the parameters, satisfying the necessary conditions. Even if a BIBD exists for a given value of v, the number of treatments and k, the block size, it might sometimes require too many replications, resulting in the increased size of the experiment. In situations where the size of the experiment is limited due to cost and other considerations, one might have to sacrifice the property of variance-balance and look for designs that are available with reasonable reasonable number of experimental units. Quite naturally, such designs might be called unbalanced or partially balanced. A variety of partially balanced designs are now available, among which the most important ones are the Partially Balanced Incomplete Block (PBIB) Designs. Generalized Group Divisible Designs are a particular class of partially balanced incomplete block designs. Construction of Generalized Group Divisible Designs and their optimality have been discussed by some authors like Jacroux(1980), Srivastav and Morgan (1998), Thannippara et. al (2009), Ghosh et. al (2012). In this article we present a new method of construction of generalized group divisible design with two groups from balanced incomplete block designs. E-optimality of generalized group divisible designs constructed in this investigation is also discussed.

IC 15/113

A Family of Tests for NBUE Alternatives using Linear Function of Order Statistics

M. Z. ANIS
Indian Statistical Institute

Murari Mitra
Indian Institute of Science Engineering and Technology, Shibpur,

In this note we develop a family of test statistics for testing exponentiality against NBUE alternatives. The exact and asymptotic distributions of the test statistics are derived. The test statistics are shown to be asymptotically normal and consistent. This family of test statistics includes the test proposed by Hollander and Proschan (1975) as a special case. Efficiency studies have also been done.

IC 15/160

On Sampling and Measurement Guidelines for Statistical Experiments Influenced by Measurement Errors

Tapan Bagchi

With today's push for Six Sigma DMAIC, the issue of measurement errors in quality engineering is now widely debated. As a result naive to theoretically sophisticated suggestions have emerged. This has led to formulating Gauge R&R guidelines by AIAG and others for the recommended choice and use of gauges to measure product characteristics. This study extends that effort. It links the overall *power* of ANOVA tests in DOE to the *number of treatments employed*, *parts produced in each treatment*, and *measurements taken on each part*. It invokes Patnaik's work (1949) to explore how these quantities may be optimized while conducting DOE to achieve a pre-stated detection power

IC 15/067

A Case Study on EM Algorithm for Bivariate Pareto

Arabin Kumar Dey, Rajat Kamra, Bhukya Ganesh
IIT, Guwahati

The main focus of this paper is to estimate a specific Marshal-Olkin bivariate pareto distribution taking the location and scale parameters as zero and one respectively. The distribution has three parameters. EM algorithm is used to estimate the parameters and some simulation is made to see the performance of the EM algorithm. A real-life data analysis is shown for illustrative purpose.

IC 15/133

Zero-inflated Poisson integer-valued AR processes

Atanu Biswas, Raju Maiti, Samarjit Das
Indian Statistical Institute

Seng Huat Ong
University of Malaya

Apratim Guha
Indian Institute of Management, Ahmedabad

Poisson integer-valued auto-regressive of order 1 [or PINAR(1)] process due to Al-Osh and Alzaid (1987) and McKenzie (1988) has been received significant attention towards modeling low count time series during the last two decades because of its simplicity. But in many cases, the process seems to be inadequate, in mostly when data appears to be over-dispersed. This over-dispersion occurs mainly due to presence of some extreme values or large number of zeros or due to presence of both.

In this presentation, we intend to present two models to accommodate such over-dispersion. We first develop a zero-inflated Poisson INAR(1) process as an alternate of the PINAR(1) process when the number of zeros in data is larger than the expected number of zeros by Poisson process. We investigate some properties it carries like stationarity, auto-correlation structure, conditional distribution with a detailed study on h-step ahead coherent forecasting. A comparison study among different methods of parameter estimation, namely Yule-Walker, conditional least squares and maximum likelihood, are carried out through some simulation study.

Then, as an alternative, based on Pegram's mixing operator, another kind of stationary zero-inflated integer-valued time series process of order p with Poisson marginal or ZIPPAR(p) process is constructed for modeling count time series consisting a large number of zeros compared to standard Poisson time series processes. Several properties like stationarity, ergodicity of this model are theoretically examined. Estimates of the model parameters are studied using the above-mentioned three methods of estimation. Also h-step ahead coherent forecasting distributions of the proposed process for $p = 1, 2$ are derived. Detailed simulation studies are also carried out.

The proposed procedures are illustrated by using two real life over-dispersed data sets – one data set giving the monthly number of submissions with skin-lesions from a region in New Zealand and the other data on daily counts of terrorist attacks in India.

IC 15/168

Survival Analysis of Breast Cancer data using Kaplan-Meier and Cox Regression Model

D. K. Ghosh
Department of Statistics, Saurashtra University, Rajkot

In this paper we have used Kaplan-Meier and Cox Regression Model to obtain the Survival rate, Survival time, Survival curve and Hazard function for the patient suffering with Breast cancer in Saurashtra-Kutch region. The main objective of this investigation is to find percentage of breast Cancer Patients in respect of Treatment wise, stage wise and year wise.

Here two study groups are considered as breast cancer women with surgery and without surgery. Moreover, the survival time of the two studied groups are compared by using Log-Rank test. In addition, Cox regression model is used to find the relationship between Survival time and treatment, Surgery, Chemotherapy, Radio therapy etc.

Here it is concluded that, if the patient are not operated then their survival time is reducing drastically. However if they were operated then survival time is reducing slowly. Moreover, after five year their survival is continuing. That is, Survival time is not decreasing up to 78 months. In contradiction the Hazard function is increasing in both the cases. In case of no surgery hazard function increases rapidly but in case of surgery it increases slowly.

Words and Phrases: Chemotherapy, Cox Regression, Kaplan – Meier model.

Survival analysis is a one of the best tools to examine and model the time it takes for event to occur. A typical such event is death. Survival analysis is a terminology comes out from this event. In this investigation we have collected data of breast cancer patients from Rajkot cancer hospital. We have used logistic regression model to measure the relationship between the categorical dependent variable and one or more independent variables which are usually continuous but not necessary always. Here we have used probability scores as the predicted values of the dependent variables. We have also compared two groups by using Cox regression model. For this investigation, we have taken four treatments namely, surgery along with other treatments, radiotherapy along with other treatments, chemotherapy along with other treatments and hormonal therapy along with other treatments to find the survival of the breast cancer patients after giving this four treatments. We have also obtained relative hazard or hazard ratio for the patients receiving the treatment and not receiving the treatment. It is shown that there is better risk for patients receiving treatment as compared to not receiving treatments. Finally it has been concluded that the risk of the patients receiving the treatment, say, hormonal therapy along with other treatments is less than as compared to the receiving the remaining treatments surgery along with other treatments, radiotherapy along with other treatments and chemotherapy along with other treatments of the breast cancer patients respectively. This result has been also supported by the survival graphs.

IC 15/156

Robustness of Sharpe Ratio

Mahesh K.C
Institute of Management, Nirma University Ahmedabad

Arnab. K. Laha
Indian Institute of Management, Ahmedabad

The Sharpe ratio is the most commonly used tool for risk adjusted return of a financial asset. The computation of the Sharpe Ratio involves estimation of the expected returns and volatilities of the various assets in a portfolio. Since the statistical properties of an estimator are dependent not only on the data quality but also onan equally important base that is formed by prior assumptions (explicit or implicit) about randomness, independence, distributional models etc. It is well known that most of the common statistical procedures are very sensitive to even minor deviations from the assumed model or in the presence of gross errors in the data. For example, both the arithmetic mean and the standard deviation are extremely sensitive to such deviations. In this paper we propose a robust alternative of the Sharpe ratio estimator and study its robustness properties for some well known return distributions such as Normal, Log-normal, Cauchy and Double exponential distributions.

IC 15/169

Is Global Warming affecting Indian cities? - Change point Analysis with Functional Data

Poonam Rathi, Arnab Kumar Laha
Indian Institute of Management Ahmedabad;

In recent years there has been considerable concern expressed worldwide regarding increase in temperature popularly called the global warming problem. However, not much work has been done on this in the Indian context. The existence of this phenomenon for six Indian cities- Bangalore, Delhi, Guwahati, Jodhpur, Mumbai, and Madurai for the period 1960 to 2000 is investigated using the techniques of functional data analysis. Anova of the functional data indicates that the six Indian cities have significantly different weather patterns that led us to examine the data for each of these cities separately. Change point analysis with functional data as described in Aue. et.al (2009) is carried out for the monthly average temperature data for the six cities. It is found that there has been a rise in the average temperature for Bangalore and Mumbai with the change point being in the middle of 1970's whereas for Guwahati, there is an increase in temperature with the change point being in the middle of 1980's. For Mumbai, presence of a change point is detected at the year 1975 but the nature of the change is not very clear. The temperature data of Delhi and Jodhpur does not show any change point.

IC 15/ 159

Using Analytics to pitch the right customer with right product at the right time through right channel

Salil Kumar, Subhankar Mukherjee, Mufazzal Arif
HDFC Bank, Mumbai

Outbound calling is the most commonly used channel for Service organizations in order to engage customers and cross-sell their products. However it's also one of the most intrusive ways of customer engagement. Most of the instances the customer is not in the right frame of mind to receive the communication that service provider wants to send to him / her. With more and more customers registering for DNC and NDNC, the accessibility of such customers is also reducing.

In such a challenging environment, it is most appropriate to engage with the customer which is preferred by the customer. As in the recent times there are multiple touch points (Branch, Inbound/Outbound call, Social Media etc.) for a customer to connect with the bank.

Given the value of these customer-led interactions, it's equally important to put forward a relevant message to the customer at each of such interactions. There was a need to develop a solution that helps HDFC Bank identify the most relevant product / service at customer level based on all the information available at customer level.

IC 15/138

Discontinuation Risk Models

Ramya Cherukupalli and Syed Abrar Hussaini
Deloitte Consulting, LLP, Hyderabad

Discontinuation or Medication non-adherence is a condition when people who stop taking their medicine resulting in worsening health and potentially unnecessary hospitalization. Medication non-adherence doesn't just impact the individual who is not taking his or her medicine; it impacts national economy, with experts estimating the cost of unnecessary medical treatment resulting from non-adherence at nearly $289bn per year. Given the scale of the problem, it is imperative that the pharmaceutical companies take all measures to curb the rates of dis-continuations.

A large specialty pharmaceutical in US approached us to use advanced analytics techniques to anticipate patient behavior based on the patient characteristics, recent and historical behavior. Data provided was vast, which included therapy related, physician and health system, patient attitude, patient condition, frequency of visits, patient participation / engagement in support activities by the pharma manufacturer and patient demographics details.

Using the wealth of information, Discontinuation Risk Models (DRM) was built to determine overall discontinuation risk and core characteristics of patients likely to discontinue. The target variable chosen mainly addresses four main adherence issues namely -

- Persistence,
- Compliance,
- Switching and
- Related reasons

All variables were developed using statistical software (SAS) by aggregating 300 files and 19 million records to create a single patient-level file. After data integration and aggregation, the team performed a series of exploratory data analyses (EDA), including:

- Univariate (single-variable) to identify candidate variables for multivariate analysis
- Correlation analysis to determine collinearity between potential multivariate predictor variables
- Principal component analysis as a means of dimension reduction of collinear, themed variables
- Logistic regression, assigning a predicted probability of an patients' propensity to leave therapy

Patient segments are developed using a decision tree model. Based on the risk profiles obtained for individual patients, targeted interventions that are responsive to his or her health outcomes and needs are recommended.

Retention is a problem faced by many industries – retailers and manufacturers want to maintain customers for their products/brands, banks and insurers want to maximize the "share of wallet". The methodology created for the DRM is its cross-functional and can be applied flexibly across industries. DRM have multiple advantages as listed -

1) Accurate: segment patients to their likely adherence behavior
2) Usable: nodes which can be acted on
3) Measurable: results can be monitored against for future sustainability

IC 15/167

Witness the Future of Analytics

Harpreet Singh Paintal
SAS Institute (India), Mumbai

In this era of Big Data, regardless of size or industry sector, organizations collect all types and amounts of data. Unfortunately, traditional architectures and existing infrastructures aren't designed to deliver fast analytical processing needed for rapid insights. The need of the hour is to be proactive and predictive, instead of being reactive.

Democratization of Data

Let's face it, many business users are uncomfortable working with data, let alone big data. They see it as something reserved for Analysts or IT Specialists. It is important to remove that mind block and put information into their hands. This interaction between right people and right information is where real value lies. By allowing users to visually explore big data, visual analytics spurs an environment providing actionable insights to complex business questions at the speed of your thought. This is bringing big data to life.

Analytics for everyone

Analytics should be tailored for ease of use so everyone can create analytic visualizations without learning new skills or engaging IT. Powerful, yet easy-to-use analytics include,

- Automated forecasting selects most appropriate forecasting method for data chosen.
- Decision trees graphically depict the most likely outcomes.
- Network diagrams show how complex data is interconnected.
- Text analytics applies sentiment analysis to twitter streams or customer comments providing quick insight into the hot topics being discussed.
- Path analysis gives insight into the dynamics of flow of how customers navigate through your website, or how shipments flow through your distribution centers

In a nut-shell:

Visual Analytics empowers organizations to convert the so called 'big data challenge' into a new world of Opportunities. **And this is where the Future of Analytics lies ….**

IC 15/ 032

Predicting Policy Renewal Probability in Auto Insurance Sector

Venkatesh Pagidimarri, Vamsi Chandra Kasivajjala, Deepti Miyan
Enlightiks Business Solutions, Banglore

We are entering a new phase of the information era, in which organizations query and analyze huge volumes of diverse data in real-time to improve outcomes for their most critical business processes. In today's competitive edge, the concept of Predictive Analytics is used in every business sector. This paper discusses about how **Predictive Analytics** helps in increasing the **policy renewal rate** for Auto Insurance sector. Customer demographics data, Type of vehicle, historical claims data is used to understand the behaviour of customers in renewing the policies.

The model identifies the parameters which impacts the **policy renewal rates** in customers. Customers were grouped in to High, Medium and Low segments based on their policy renewal probabilities. This will help the insurers to identify their **potential customers** that are more likely to go for the policy renewal. The last section of the paper outlines the financial benefits in increasing the policy renewal rate using the predictive model.

IC 15/ 158

Using Analytics to set lending product Cross-sell targets at bank level

Mufazzal Arif
HDFC Bank, Mumbai

Branch is the most commonly used channel for Service organizations in order to engage customers and cross-sell their products. However it's also one of the most intrusive ways of customer engagement.

Given the value of these customer-led interactions, it's equally important for the branch to understand its own customer base, branch demographics, location demographics, peer performance and competition to increase the market share of the organization. In such a challenging and competitive environment, it is most important to have a right target set for a branch against the best performing branch from its peers.

In order to set asset cross-sell targets at branch level, we have used frontier analysis method with particular emphasis given to Data Envelopment Analysis (DEA) method. Three asset lending utility was designed for setting branch level target and this has resulted into an improved product holding at customer level and increase the market share.

IC 15/170

A New Method for Estimating the Beta measure of a Stock using Circular Statistics

Arnab Kumar Laha
Indian Institute of Management Ahmedabad

Kalyani Shelat
National Institute of Technology, Surat

The Beta which is a widely used measure of risk in stock markets is calculated by using the ordinary least squares (OLS) regression method. While the OLS method performs well in estimating the beta coefficient when the returns of the stock and that of the market is highly correlated it doesn't perform that well when the correlation is low. In this paper we propose a new approach to calculate the beta coefficient using ideas from circular statistics and examine its performance through extensive simulations. We find that when the correlation is low and/ or the sample size is small this new estimator performs better than the OLS estimator and the Theil-Sen estimator of beta in terms of having a lower mean squared error (MSE) and mean absolute bias (MAB). We apply this new method to some stocks whose returns have low correlation with the market returns and find that the value of beta given by our new method is quite different from that given by the OLS method. These findings are likely to have substantial importance in understanding risk related to such stocks.

IC 15/132

Portfolio Optimization Using Krylov Subspace method MINRES-QLP

Arnab Kumar Laha
Indian Institute of Management Ahmedabad

Tanmay Sen and Sriparna Saha
Indian Institute of Technology Patna

In order to solve the problems of Global Minimum Variance (GMV) portfolio and Mean-Variance (MV) portfolio the inversion of the sample covariance matrix of the security returns is required. This matrix may become ill-condition in some situations. For the case when the covariance matrix is symmetric, ill-conditioned and near singular, a Krylov subspace based method MINRES-QLP that uses QLP decomposition instead of QR decomposition is used to derive the GMV and MV portfolios. The portfolio obtained using MINRES-QLP is compared with those obtained using the LASSO and Ridge portfolio. In case of portfolio variance MINRES-QLP shows better performance than the other two. It is further seen that our portfolio tracks the market indices NASDAQ 100 and S&P 500 but it fails to track the FTSE 100 index in the recession period 2008-2009.

IC 15/ 118

Kolmogorov-Smirnov Test for Hybrid Censored Data

Buddhananda Banerjee
Indian Institute of Science Education and Research, Kolkata

Biswabrata Pradhan
SQC & OR Unit, Indian Statistical Institute, Kolkata

In the literature of reliability theory and survival analysis several censoring schemes are introduced and modified over the time to deal with the life-testing problems. But no such asymptotic result or methodology is available to perform a goodness-of-fit test for the hybrid censored data (see Balakrishnan and Kundu, 2013). We give an alternative representation of the Kolmogorov-Smirnov (KS) statistics under the Type-I censoring. The new representation leads us to approximate the limiting distributions of the KS statistic as a function of the Brownian bridge for the Type-II, Type-I hybrid and Type-II hybrid censored data. The approximated distributions are used to obtain the critical values of the tests in this context. It is shown that the proposed KS test procedure for the Type-II censoring has more power than that of a modified KS test available in the literature for moderate and large sample size.

IC 15/165

Estimating Export Demand: An Empirical Analysis

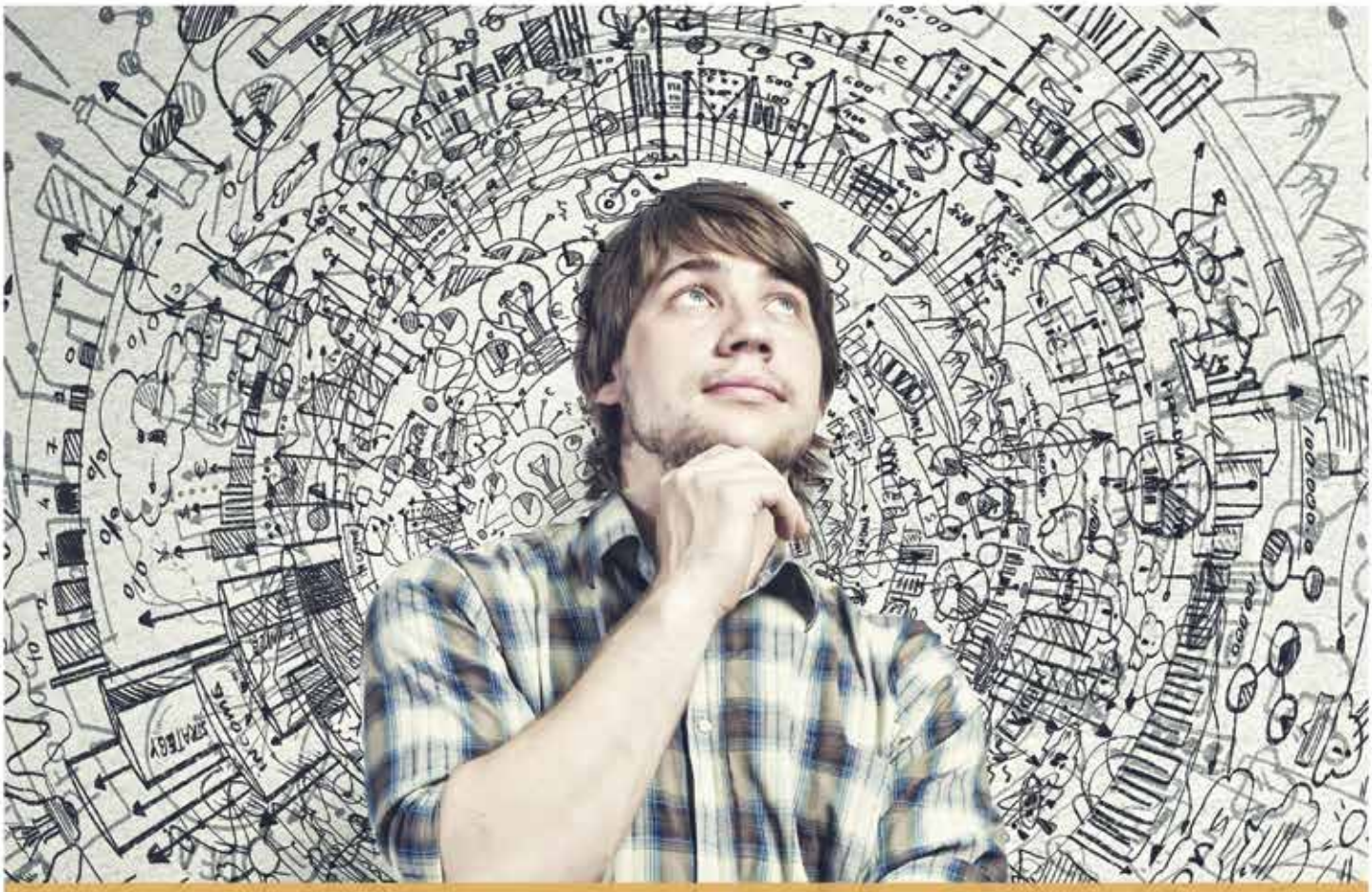
Joy Chowdhury
Goa Institute of Management, Goa

Arnab Kumar Laha
Indian Institute of Management, Ahmedabad

Aman Mahawar
IIT Kharagpur

Export of a country plays major role since it affects many economic indicators/parameters/variables like current account deficit, foreign exchange reserves etc. Generally all the economies try to boost up the export sector and formulate strategy for the higher growth in export. Thus it is very important to find out the factors which determine the export for an economy. In this paper we have tried to find out the determinants of bilateral export which

will be helpful for the policy makers in formulating policy. We have considered 15 countries which mean 105 country pairs. We have estimated the bilateral export demand function by using gravity model. We found that distance between two countries, product of GDP between two countries, openness, and real effective exchange rate are significantly affecting the bilateral export. We also found that exports are higher for countries sharing the common border. Moreover, common language has also positive impact on the bilateral export.



Ignite your future in analytics

Collaborate with the smartest people to create insight, impact and innovation.

Industry influence:

Srikanth Velamakanni, Co- founder and CEO, ranked as India's most influential analytics leader for two consecutive years by Analytics Magazine

Featured in "The Best Companies To Work For" issue by Business Today

"Vendor to watch" in managed analytics services by Gartner

"Cool Vendors" in Analytics by Gartner

Are you a future Fractalite?

✉ Email us at careers@fractalanalytics.com

Fractal®

General Mills

- World's sixth largest food company
- Produce and market more than 100 consumer brands in more than 100 countries
- 41,000 global employees
- $17.9 billion in fiscal 2014 net sales*

More than 91% of our employees strongly recommend General Mills as a great place to work.

Häagen-Dazs
Betty Crocker
Totino's
Cascadian Farm ORGANIC
PROGRESSO
Cheerios
Cinnamon Toast Crunch
FRUIT ROLL-UPS
Yoplait
NATURE VALLEY
Pillsbury
Green Giant
MUIR GLEN Organic

World's Most Reputable Companies
Reputation Institute

Best Companies for Multicultural Women
Working Mother magazine

Top Companies for Executive Women
National Association for Female Executives

Best Employers for Healthy Lifestyles
National Business Group on Health

Linda Baptista - Manager
Solutions Development
5+ years with General Mills

General Mills continuously strives to improve business processes and remove waste out of the system. There are opportunities to work cross functionally which helps in overall development. If you are looking to be part of a global team working in best-in-class business process and technology, General Mills is the place to be.

Learn more about job opportunities at
General Mills India Private Limited:
902 Ventura, Hiranandani Business Park,
Powai, Mumbai, India 400076
Email: gbs.recruiting@genmills.com
Website: www.generalmills.co.in
Board No – 91-22- 40430430 / 42430430

We live our championship values every day.

Do the **right thing**, all the time
Play to **win**
Grow and **inspire**
Win as a **team**
Act boldly, move quickly

GENERAL MILLS

With Best Compliments from

Our Sponsors - ICADABAI - 2015

GOLD SPONSORS

Fractal®

GENERAL MILLS

SILVER SPONSOR

sas®

BRONZE SPONSORS

Emerald

Minitab®