

5th IIMA International Conference on
**Advanced Data Analysis,
Business Analytics and Intelligence**
April 8-9, 2017



Indian Institute of Management Ahmedabad, India

ABSTRACT BOOKLET

5th IIM A International Conference on
**Advanced Data Analysis, Business
Analytics and Intelligence**

April 08-09, 2017, Ahmedabad, India.

Indian Institute of Management Ahmedabad is happy to organize the 5th 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.

Dr Vikram Sarabhai and a few other public spirited industrialists founded the Indian Institute of Management, Ahmedabad in 1961 as an autonomous body with the active collaboration of the Government of India, Government of Gujarat, and industry. The Institute had initial collaboration with Harvard Business School which greatly influenced the Institute's approach to education. Gradually it emerged as a confluence of the best of eastern and western management approaches having strong ties with both industry and government.

The first IIMA International Conference on Advanced Data Analysis, Business Analytics and Intelligence (ICADABAI 2009), held on 6-7 June 2009 was attended by about 150 participants from academia and industry. A total of 116 research papers and case studies were presented in this conference.

The second conference in this series ICADABAI-2011 was held at IIM, Ahmedabad on 8-9 January, 2011. The two day conference had three key-note speeches delivered by eminent academicians and practitioners and two panel discussions on special topics aligned to the theme of the conference. This conference saw academicians and practitioners present a total of 100 research papers and case studies.

The third conference in this series ICADABAI 2013, was held on 13-14 April 2013. More than 150 people registered for this conference and the two day conference had six key note speeches delivered by eminent academicians and practitioners on special topics aligned to the theme of the conference.

The fourth conference in this series ICADABAI-2015 was held on 11-12, April 2015. The two day conference had four keynote speeches, two panel discussions, workshops on 4 different topics and more than 80 paper presentations.

Papers for ICADABAI-2017 were selected for regular presentation or short presentation based on the recommendation of the reviewers following a double blind peer-review process. We thank the reviewers for their generous support.

Conference Team of ICADABAI-2017

Arnab Kumar Laha (Conference Convener)

Indian Institute of Management Ahmedabad



Pravida Raja A.C. (Conference Manager)

Deepna. P (Conference Coordinator)

Indian Institute of Management Ahmedabad.



Mahesh K.C (Conference Member)

Institute of Management, Nirma University, Ahmedabad



Vaidyanathan Ramachandran (ICADABAI-HACKATHON 2017 Convener)

Nagravision India Pvt.Ltd, Bangalore

International Programme Committee members of ICADABAI-2017

1. Ananthakumar Usha – Indian Institute of Technology, Bombay.
2. Banerjee Buddhananda –Indian Institute of Technology Kharagpur.
3. Chowdhury Shovan- Indian Institute of Management, Kozhikode.
4. Ghosh Dilip Kumar- Saurashtra University, Rajkot.
5. Ghosh Pulak - Indian Institute of Management, Bangalore.
6. Ghosh Sujit Kumar - North Carolina State University, USA.
7. Guhathakurta Kousik- Indian Institute of Management, Indore.
8. Karandikar Rajeeva Laxman- Chennai Mathematical Institute.
9. Kumar Dinesh Unnikrishnan- Indian Institute of Management, Bangalore.
10. Kundu Debasis- Indian Institute of Technology, Kanpur.
11. Mahesh K.C - Nirma University, Ahmedabad.
12. Maitra Ranjan- Iowa State University, USA.
13. Mitra Murari – Indian Institute of Engineering Science and Technology, Shibpur.
14. Mukhopadhyay Chiranjit- Indian Institute of Science, Bangalore.
15. Nanath Krishnadas- Middlesex University, Dubai.
16. Pal Debdatta- Indian Institute of Management, Lucknow.
17. Ramachandran Vaidyanathan-Nagravision India Pvt. Ltd.
18. Ranjan Pritam- Indian Institute of Management, Indore.
19. Roy Vivekananda- Iowa State University, USA.
20. Sebastian M P -Indian Institute of Management, Kozhikode.
21. Sengupta Anindya - Fractal Analytics.
22. SenGupta Ashis - Indian Statistical Institute.
23. Singh Sanjeet- Indian Institute of Management, Calcutta.

Session Organizer of ICADABAI-2017

Atanu Biswas

Indian Statistical Institute, Kolkata



Workshop speakers of ICADABAI-2017

Krishnadas Nanath

Middlesex University, Dubai

Arnab Kumar Laha and Poonam Rathi

Indian Institute of Management Ahmedabad

We thank the following reviewers of ICADABAI 2017 for helping us in selecting the papers for presentation

- Abhiman Das - IIM Ahmedabad
- Abhishek - IIM Ahmedabad
- Anindya Chakrabarti - IIM Ahmedabad
- Apratim Guha - IIM Ahmedabad
- Biju Varkkey - IIM Ahmedabad
- Budhananda Banerjee - IIT, Kharagpur
- Chetan Soman - IIM Ahmedabad
- Debdatta Pal - IIM Lucknow
- Debjit Roy - IIM Ahmedabad
- Goutam Dutta - IIM Ahmedabad
- Jacob Joshy - IIM Ahmedabad
- Jahar Saha - IIM Ahmedabad (Retd.)
- Kanish Debnath - FLAME University, Pune
- Karthik Sriram - IIM Ahmedabad
- Kavitha Ranganathan - IIM Ahmedabad
- Krishnamoorthy Srikumar - IIM Ahmedabad
- Kriti Sharda - IIM Ahmedabad
- Mahesh K.C - Institute of Management , Nirma University, Ahmedabad
- Murari Mitra - Indian Institute of Engineering Science and Technology, Shibpur
- Nilesh Gupta - IIM Trichy
- P.K.Sinha - IIM Ahmedabad
- Prasun Das - ISI Kolkata
- Pritha Dev - IIM Ahmedabad
- Rajeev Sharma - IIM Ahmedabad
- Sachin Jayaswal - IIM Ahmedabad
- Sanjay Verma - IIM Ahmedabad
- Sanjeev Tripathi - IIM Ahmedabad
- Sharda Kirti - IIM Ahmedabad
- Sobhesh - IIM Ahmedabad
- Soumya Mukhopadhyay - IIM Ahmedabad
- Usha Ananthakumar - IIT Bombay
- Vishal Gupta - IIM Ahmedabad
- Vivekananada Roy - Iowa State University, USA

Contents

IC17/KN 1	The Science of Behavioral Nudges	1
IC17/KN 2	Key emerging trends in analytics and how this is helping organizations drive business outcomes	1
IC17/KN 3	Cognitive Systems to Drive Decision Making in Financial Services	1
IC17/KN4	Big Data, Advanced Analytics and IoT in industry	2
IC17/KN 5	Dimension Reduction- one way to big data analysis.	3
IC17/KN 6	Directional Statistics for High Volatility Big Data	4
IC17/KN 7	Next Generation Internet & Data Science.	4
IC17/KN 8	Prescriptive Analytics with Optimization: Challenges for Future	5
IC17/IT 1	Statistical Analysis of High-Velocity Data Streams.	6
IC17/IT 2	Riding the information wave : Driving Analytics Adoption in Life Insurance in India.	6
IC17/WS 1	Social Media Analytics using R Studio: Mixing the two domains:	7
IC17/WS 2	Functional Data Analysis	8
IC17/001	Predictors for top ranking Management institutions in India - A study using Discriminant analysis based on NIRF ranking data	8
IC17/002	Innovative Heuristics Modeling for Dynamic Project Cost Optimization.	9
IC17/003	The Impact of Green Washing on Green Confusion, Green Perceived Risk and Green Trust-An Indian Perspective	9
IC17/004	Impact of Stress on the Worklife Balance of Women Gaenacologists.	10
IC17/006	Click-Through Rate Estimation using CHAID Classification Tree Model	10
IC17/007	On optimum reliability acceptance sampling plans under generalized hybrid censoring	11
IC17/009	Devising Virtual Supply Chain Analytics for Agriculture Products during Festival season	11
IC17/014	Does Banking Competition Granger Cause Banking Stability: The Study of SEM Countries	12
IC17/015	Effect of Venture Capital Investment on Economic Growth: Evidence from GMM Estimates	13
IC17/ 017	Impact of Social Media Advertising on Consumer Buying Behavior	13
IC17/ 019	Determining Sportsperson's affinity for brand sponsorships: An Analytics approach	14
IC17/021	Sales Lobby – Account Insights Gateway	15
IC17/023	Group Messaging and its Impact on Quality of Life at Work Place – an exploratory study on WhatsApp's group communication platform	15

IC17/028	Optimization of ATM Network Using Geo-Spatial Data Analysis Techniques..	16
IC17/029	Assessing Student Employability to Help Recruiters Find the Right Candidates	16
IC17/033	An Inter Services Comparative Study on Job Satisfaction in Bangladesh: Application of Discriminant Function Analysis and Structural Equation Modeling	17
IC17/034	Applying Predictive Analytics in a Continuous Process Industry	18
IC17/035	Price transmission from crude oil to corn in the United States: New evidence from multivariate adaptive regression splines approach.....	18
IC17/037	A Stochastic Feedback Queuing Model with Encouraged Arrivals and Retention of Impatient Customers	19
IC17/039	A More Powerful test Identifying the Change In Mean of Functional Data	20
IC17/041	A Hollander-Proschan type test when Ageing is not Monotone	20
IC17/042	Determination of Optimal Inventory Policy – A Case Study.....	21
IC17/043	Estimation of Parameters of Misclassified Size Biased Borel Tanner Distribution	22
IC17/044	Predicting Success Probability in Professional Tennis Tournaments Using a Logistic Regression Model	23
IC17/049	Do bubbles and crashes in stock markets represent a seismic behaviour? An empirical investigation.....	23
IC17/051	A study of the Performance of ‘Make in India’ in Information Technology Hardware: An Early Inference from Perceptions of Senior Information Technology Hardware Professionals.....	24
IC17/053	Loan Loss Provisioning Practices in Indian Banks.....	25
IC17/055	Hausdorff Path Clustering and Hidden Markov Model Applied to Person Movement Prediction in Retail Spaces.....	25
IC17/059	“Estimation Of Fluid Flow Rate And Mixture Composition Using Low Cost Acoustic Sensors For Application In The Oil And Gas Industry”.....	26
IC17/061	From Surveys to Action: A Machine Learning Pathway to Realize Customer Satisfaction.....	26
IC17/064	Improving Email Marketing Campaign Success Rate Using Personalization ...	27
IC17/065	Automatic Detection of Tuberculosis using Deep Learning Methods.....	28
IC17/068	Ranking of Batsmen in ODI: A Factor Analysis Approach	28
IC17/076	Analytics-led Talent Acquisition for Improving Efficiency and Effectiveness...	29
IC17/078	Frame of Reference Biases: The Effect of Social Media on Behavioral Reinforcement	29
IC17/080	The relationship between insurance and Economic growth in India: A Cross- region study using an Econometric approach	30

IC17/082	In-database Analytics in the Age of Smart Meters	30
IC17/ 084	Role of Gender Bias & Household Amenities in Impeding Access to Financial Services	31
IC17/ 085	Exploratory and Predictive Data Analysis Using Machine Learning Tools and Techniques.	32
IC17/087	Feature Engineering Strategies: A Practitioner’s Guide.	32
IC17/ 090	Predicting customer churn for DTH: Building churn score card for DTH	33
IC17 / 091	Online Strategy to Influence Retail Customer Purchase Intention: Analysis through Structural Equation Modeling.	34
IC17/099	Taking The Bizarre Out of the Retail Bazaar In The Sky	34
IC17/101	Leveraging The Power of Emoticons to Enhance Polarity Classification of Text	35
IC17/102	Connected Cars and Driving Pattern: An Analytical Approach to Risk Based Insurance.	36
IC17/105	Hybrid Machine Learning and Cognitive approach using Telematics and IoT for tackling Downtime of Commercial Vehicles and Taxis.	36
IC17/111	Market Efficiency and Inflationary Impact of Agricultural Futures Trading in India.	37
IC17/112	Inverse Problem for Dynamic Computer Simulators	37
IC17/ 113	A Novel Approach for Cost-Sensitive Attribute Selection on Medical Datasets	38
IC17/114	Employee Branding: A Case Study	38
IC17/ 115	Incremental Learning for Location Prediction on Taxi-GPS Data Streams.	39
IC17/116	Robust Estimators on the Circle.	39
IC17/117	Agricultural Marketing Channel Choice and Its Determinants – A Farmers’ Perspective	40
IC17/ 118	New Approaches to Prediction using Functional Data Analysis	40
IC17/119	Response Adaptive Allocation with ordinal categorical responses	41
IC17/120	OLS: Is That So Useless for Regression with Categorical Data?	41
IC17/121	Robustness issues in circular-circular regression	42
IC17/ 122	Assessing Density Forecast for Energy Commodities: A Post- Financialization Study	42
IC17/123	Market Irrationality and Share Repurchases	43
IC17/124	Robustness of circular mean functional for families of asymmetric circular distribution	43
IC17/ 125	Idiosyncratic return characteristics and cross-section of expected stock returns.	44
IC17/126	Forefather distribution in Galton Watson Processes with age dependant structure in population.	44

IC17/ KN 1

The Science of Behavioral Nudges

Debashish Banerjee

Deloitte

Behavioral economists advocate the fact that humans are largely irrational and struggle to control their emotions. Irrationality in an individual's decision making is predictable and can be governed by behavioral nudges.

While predictive analytics identifies "which" individuals or customers to target, behavioral analytics attempts to design environments that can persuade them to make the right decisions. When the ultimate goal is behavioral change, predictive analytics and the science of behavioral nudges can serve as two parts of a greater, more effective whole. In this session on the science of behavioral nudges, the audience will be exposed to the importance of behavioral nudges and how behavioral science in association with data science can be used effectively to solve business problems.

IC 17/ KN 2

Key emerging trends in analytics and how this is helping organizations drive business outcomes

Discussion will focus on key emerging trends in analytics and how this is helping organizations drive business outcomes

- Analytics as a Service
- Power of predictive analytics to drive intelligent operations
- Artificial intelligence for unstructured data and judgement based work
- Data centric businesses

IC 17/KN 3

Cognitive Systems to Drive Decision Making in Financial Services

Sriram Raghavan

IBM Research-India and CTO, IBM India/South Asia

Data, in all forms, is expanding as a resource to be utilized. Yet in many industries and professions, the data explosion is outstripping the human capacity to understand the meaning hidden within that data. Cognitive computing is able to unlock the potential in all data - internal, external, structured, unstructured, voice, and visual. Enterprises can make better operational decisions, understand customer wants and needs, communicate in real time, and optimize business processes, leveraging the power of cognitive computing to understand, reason, and learn.

In this talk, I will begin with an overview of IBM's worldwide cognitive computing research agenda and lay out the key pillars of our innovation agenda in this space. I will then present two case studies, from the financial services domain, that illustrate the transformative power of cognitive computing technology.

The first case study looks at the power of deep natural language understanding and knowledge representation to support rich analysis and understanding of corporations. We show that, using unstructured data primarily extracted from corporate annual reports, expert opinions, press coverage, etc., it is possible to define several qualitative attributes based on the corpus of textual data, which are summaries of firm characteristics that are not directly comprehended as a numeric value.

In the second case study, I will focus on the problem of regulatory compliance. Monitoring regulations, tracking changes and ensuring compliance has been a highly manual process because of the complexity involved in understanding regulations and control policies that are written in complex natural language. I will describe how cognitive technologies can dramatically transform this process and drive significant efficiencies and cost reduction.

IC17/KN4

Big Data, Advanced Analytics and IoT in industry

Sudipta Sen

McKinsey and Co.

The mega trend of Big Data (BD), Advanced Analytics (AA) and Internet of Things (IoT) is shaping the future of industry in truly disruptive ways. These technologies enable innovation, help create value, facilitate decision making, understand the customer behaviour by granular segmentation of customers, enhance customer satisfaction and enhance productivity and efficiency.

Big Data is everywhere. The volume, ubiquity and potential of Big Data presents a huge opportunity for companies to create value. In the auto industry, Big Data is helping save costs by reducing product complexity. In the IT industry, it drives excellence in R&D. Declining prices of memory and computing power have boosted the use of Big Data by in memory analytics. So massive is the impact of Big Data that it is being hailed as the "new oil".

Advanced Analytics is helping the world's leading companies create multibillion-dollar products using machine learning algorithms in medicine, retail, banking, transport and supply chain organizations.

IoT is a new engine for product innovation and investor value creation. This is seen across wearable devices, industrial equipments like pumps and turbines, oil and gas supply, logistics chain initiatives, reliability of heavy vehicles, self-driving cars and aviation. Companies are using the connective power of IoT to generate impactful solutions and create value.

Harnessing the benefits of these technologies requires mastering a new landscape of foundational practices across Data, Analytics, IT, Employees and Processes. To maximize impact, it is essential for companies to have a clear vision, employ use cases for leveraging Big Data and lay the necessary foundations by identifying data sources required and capabilities and tools needed.

Benefits of Big Data, Advanced Analytics (BDAA) and IoT are already being reaped across sectors. At a petrochemical plant, BDAA in operations helped maximize yield and profit. BDAA used in predictive maintenance accurately predicted transmission failure events across a truck fleet more than 80 percent of the time. In the pharma business, BDAA has helped reduce costs and shorten time-to-market for drugs.

Companies can capture Big Data, IoT and Advanced Analytics opportunities through four guiding principles: (1) develop a “business first” BDAA strategy, working with functional and business leaders to identify opportunities and develop priorities for enhanced insights, data and capabilities, (2) invest in data infrastructure and tools and identify application areas for BD, AA and IOT, (3) build cross-functional teams supplemented with experts and develop capabilities to capture automation and optimization opportunities, and (4) define processes to translate insights into business value.

IC 17/ KN 5

Dimension Reduction- one way to big data analysis

Asis Kumar Chattopadhyay

Department of Statistics, Calcutta University

For multivariate analysis with $p(>1)$ variables the problem that often arises is the ambiguous nature of the correlation or covariance matrix. When p is moderately or very large it is generally difficult to identify the true nature of relationship among the variables as well as observations from the covariance or correlation matrix. Under such situations a very common way to simplify the matter is to reduce the dimension by considering only those variables (actual or derived) which are truly responsible for the overall variation in order to analyze the data.

Principal component analysis (PCA) is a very common dimension reduction procedure. PCA was invented in 1901 by Karl Pearson, as an analogue of the principal axis theorem in mechanics. It was later independently developed by Harold Hotelling. It was also discussed by several authors in different forms. But PCA has several limitations. Here the components are uncorrelated and not independent and only under Gaussianity, the components become independent. As a result PCA works better for Gaussian data. Further the components are difficult to interpret physically being mixture of several random variables.

More recently, Independent Component Analysis (ICA) has emerged as a strong competitor to PCA and factor analysis. ICA finds a set of source data that are mutually independent and not only uncorrelated like PCA. ICA was primarily developed for non-Gaussian data in order to find independent components responsible for a larger

part of the variation. ICA separates statistically independent original source data from an observed set of data mixtures.

ICA has been used for data analysis in different areas like signal processing, pattern recognition, econometrics, Astrophysics etc. Andrew Back (RIKEN, Japan) and Andreas Weigend (L N Stern School of Business, New York) used ICA to explore whether it can give some indication of the underlying structure of the stock market. The target was to find interpretable factors of instantaneous stock returns. Such factors could include news, response to very large trades and unexplained noise.

IC 17/ KN 6

Directional Statistics for High Volatility Big Data

Ashis SenGupta

Indian Statistical Institute, Kolkata, West Bengal, INDIA and Augusta University, Augusta, Georgia, USA

In this era of emerging complex problems, both small and big data – linear and non-linear, exhibit challenging characteristics which need to be carefully modelled. Marked presence of asymmetry, multimodality, high volatility, long and fat tails, non-linear dependency, etc. are common features of contemporary data, including important applications in Reliability analysis. Notwithstanding pitfalls, ideas from several disciplines do enrich the contribution of the research work. Directional statistics is one such scientific “key technology” as can be exploited to address these problems elegantly. In this talk, we consider the problem of obtaining probability distributions for modelling high volatility. The work of Mandelbrot has shown the appropriateness of the stable families of distributions for high volatility. However, in general, these families do not possess any analytical closed form for their probability density functions. This leads to the complexity of inference involving the parameters of such distributions. We overcome this problem of modelling high volatility data by appealing to the area of probability distributions for directional data. A new family of possibly multimodal, asymmetric and heavy-tail distribution is presented. The usual fat-tail, Cauchy and t, distributions are encompassed by this family and it has even tails comparable to that of the stable family. The problem of estimation of the parameters for such distributions as mentioned above is taken up. We apply our results to several real-life examples including one from bankruptcy.

IC 17/ KN 7

Next Generation Internet & Data Science

Ashish Ghosh

Indian Statistical Institute Kolkata

The Internet used today can be described as a network of computers, which connects one user to others around the globe. Most of the usage and application of the “Internet of Computers” involves human intervention. The future of the Internet would be a world where manual intervention for the objects on the network could

be minimized and its functionalities would be automatic and smart. This internet would not only connect computers and smart phones; it would be a network of smart objects, the “Internet of Things”. These “things” would be smart enough to sense, process and decide a corresponding action, Examples include smart appliances (refrigerator, lights, air conditioners), traffic signals, smart body monitors, etc. The individual objects along with the network would collect process and exchange data strategically. This interconnected network along with all the smart objects working together in correspondence with each other form a larger “Cyber Physical System” (like smart cities, smart hospital, etc). A working CPS would generate tons of data, hence efficient processing and effective use of this data is very crucial. There will be data from everywhere like climate data, social network data, video data, medical data, scientific data, etc. Storing these data for analytics may not always be feasible and analyzing them in real time will also be too difficult. Traditional analysis tools are not well suited to capture the complete essence of this massive data. The volume, velocity and variety is too large for comprehensive analysis, and the range of potential correlations and relationships between disparate data sources are too great for any analyst to test all hypotheses and derive all the value buried in the data. Some algorithms already have good capability of letting computers do the heavy thinking for us in case of smaller data. But, we are striving for more to deal with large volumes of such data in a short time. Therefore, we need to revisit old algorithms from statistics, machine learning, data mining and big data analytics and improvise them to tame such big data. Major innovations in big data analytics are still to take place; but, it is believed that emergence of such novel analytics is to come in near future from various domains.

IC 17/ KN 8

Prescriptive Analytics with Optimization: Challenges for Future

Goutam Dutta

Production and Quantitative Methods Area
IIM Ahmedabad

Analytics is the science of interpretation, and communication of meaningful patterns in data. It is very important to identify the pattern of data in areas rich with recorded information; Analytics relies on the simultaneous application of statistics, computer science and operations research to quantify performance. Business Application of Analytics started in 1950s and continues today in several fields of management like pricing and revenue management, manufacturing planning, supply chain, project management, healthcare, military and many other systems in public and private systems.

Optimization of business processes demonstrates a distinctive advantages as it only demonstrates the pattern in the data and shows the relationship (descriptive), but also shows what needs to done (prescriptive). A firm needs to be bold to experiment and implement that change suggested by the model. Most of the research in optimization has avoided the most difficult part of the modelling process: solving real world problem

and implementation of the model in the real world. While the research in algorithms have taken about ninety five percent of the efforts , only five percent of the people are possibly solving or trying to solve real practical problems. Future challenges lie with solving such practical problems and develop meaningful methodology out of it.

The talk will summarize the history of optimization for business problem solving from 1950s and discuss the challenges of analytics professional in a business setting.

IC 17 IT1

Statistical Analysis of High-Velocity Data Streams

Saumyadipta Pyne

Indian institute of Public Health, Hyderabad.

In many scientific and healthcare applications, high-velocity stream data are generated routinely and relentlessly. Stream data present unique challenges for statistical analysis. For instance, the data cannot be saved indefinitely, which could make it difficult to fit certain models optimally. Also, the models might require systematic updating as new data arrive. The discussion will focus on specific applications and algorithms for stream data analysis.

IC 17/ IT 2

Riding the information wave : Driving Analytics Adoption in Life Insurance in India

Sayan Sen

HDFC Life

As per the India Brand Equity Foundation of the Ministry of Commerce and Industry, Govt. of India, the life insurance market has grown from US\$10.5 billion in FY 02 to US\$27.5 billion in FY16, growing at a CAGR of 7.5 per cent. Although LIC accounts for 70% of the total market size, there is intense competition among 23 private players as well to increase market share. There is an increasing pressure to innovate on products and services. Bancassurance remains the dominant channel but the insurers are increasingly using various modes of digital outreach for sales and services. The market is undergoing structural changes such as open architecture in bancassurance, liberalisation of foreign investment ceiling from 26% to 49% and section 45 of Insurance Amendment Act, 2015 on claims repudiation. Greater competition, increased distributor payouts and regulatory changes will shrink profits. Hence there is a greater need to stay updated on market trend, respond to customer needs, improvise operating margin and manage risk.

Predictive analytics is used to develop bespoke solutions in driving strategic response to address these business challenges. Although the life insurance industry has long been using actuarial insights to manage mortality and investment risks, recent developments in technology to store and manage information and apply algorithm

based inference has expanded the opportunities to leverage data driven decisions. Some of the successful use cases have been in predictive underwriting, policy renewals, cross-selling and loyalty marketing, claims fraud identification, online product recommendation engine and customer sentiment analysis.

Insurers in emerging markets are also taking note of the impact delivered by analytics based business models in other industries including ecommerce, credit cards, banking, retail and telecom. The domestic companies are using various models of Analytics integration including establishing in house Centre of Excellence as well as engaging third party consulting firms. Analytics is well-poised to become an integral part of the insurance organisation alongside other established departments like sales & marketing, finance, operations, technology and HR to help create sustainable value and stay ahead in competition.

IC 17/WS 1

Social Media Analytics using R Studio: Mixing the two domains:

Krishnadas Nanath

Middlesex University, Dubai

Data Science and Social Media With the growth of Social Media, the amount of data generated by these platforms is increasing exponentially. This becomes a riveting sight for Data Scientists and Analysts and presents an opportunity to extract, process and visualize this data for added business value. This workshop introduces the participants to the world of Social Media Analytics and the use of Developer platforms (Twitter Dev, LinkedIn Developer and Facebook Developer) to begin with. Participants will explore various methods of data extraction from social media and how to apply statistical methods on social media datasets. Real world case studies on Hashtag monitoring will be discussed from Social Media Analytics Consulting perspective.

Live demonstrations and tutorials will be given on using tools like R to connect with Twitter Database using Twitter Developer account. Live tweets will be extracted based on a given Hashtag and then Sentimental Analysis of these tweets would be performed in real time. Use of several tools for conducting prediction and classification on social media data will also be covered in the workshop. This workshop will also include the Analytics on Facebook Fan pages which includes extraction of Fan page data and Statistical analysis on extracted data. Information on how analytics is used in Sponsored Ads on Facebook will also be given.

Functional Data Analysis

Arnab Kumar Laha, Poonam Rathi

Indian Institute of Management Ahmedabad

In many real life situations the data are obtained as curves or functions. Mathematically, functional data can be thought of as realisations of sample paths of an underlying stochastic process. The topic of Functional Data Analysis (FDA) deals with analysis of such data. With the advancement of technology, collection of data continuously during a time interval has become much easier than before. However, in some cases it may be relevant or possible only to collect data intermittently at several time points. Functional data arises in various scientific and business applications such usage history of telecom customers, sensor information coming from equipments like engines of vehicles, data on the electrical activity of the heart (ECG) or the scalp (EEG), growth curves for children, medical imaging, booking curves for hotel rooms, flights etc. FDA provides the statistical methodology for analysing such type of data and it differs from conventional statistics where the observations considered are either scalars or finite-dimensional vectors. In this workshop, we will provide an introduction to functional data analysis. Beginning with a brief discussion on the theoretical foundations, we discuss the descriptive statistical measures, some simple inferential procedures, functional regression ideas, classification and clustering with functional data. Several examples and real life business applications will be discussed.

IC 17/001

Predictors for top ranking Management institutions in India - A study using Discriminant analysis based on NIRF ranking data

Avadhanam Ramesh

Vignana Jyothi Institute of Management, Bachupally,
Hyderabad, Telangana

R.Srinivasa Rao

Vignana Jyothi Institute of Management, Bachupally,
Hyderabad, Telangana

Degree in Management domain is most vaunted post-graduation degree for many Indian students. At one time Economist Intelligence Unit found the Indian institutes of Management was the toughest business school in the world to get into based on number of aspirants and availability of seats. However in the recent decade, proliferation of Business schools in India and associated rankings by different magazines caused confusion among the stakeholders and sometimes seriously raised the issues of trust and ethics. In the light of the above precarious situation, National Institutional Ranking Framework initiative by Ministry of Human Resources Development comes at opportune time for evaluating several business schools in India and has serious implications for the stakeholders. The study uses National Institutional Ranking Framework ranking data for Business schools and uses Discriminant analysis for classification and prediction for top fifty business schools in India. The study also uses Analysis of Variance One way and post hoc to

assess the significant difference between different business schools in India. The study found that important predictors for top ranking Business schools (top 25 and below 25) in the order of importance include Teaching, Learning & Resources , Research, Professional Practice & Collaborative Performance, Perception, Graduation outcomes and Outreach. There is no significant difference between Indian Institutes of Management and Business schools set up by Indian Institutes of Technology in India. However significant differences exist between Indian Institutes of Management and private business schools in India.

IC 17/ 002

Innovative Heuristics Modeling for Dynamic Project Cost Optimization

Perumalsamy Radhakrishnan

Emirates College for Management and Information Technology, Dubai, United Arab Emirates

Jeyanthi Natarajan

Sri GVG Visalakshi College ,Udumalpet, TN, India

Efficient and effective project management significantly improves the bottom line of an organization as well as enhances service level provided to customers. The dynamic nature of cost escalations of the various tasks of the project is a serious issue during project execution stage. Efficient project management is a complex process in an effort to execute the project within the budget provisions. The complexity of the problem increases when more number of tasks and longer duration of the project are involved. In this paper, a Particle Swarm optimization methodology is proposed and implemented to generate essential predictive analytics to overcome the impasse in maintaining the optimal project cost performance.

IC 17/003

The Impact of Green Washing on Green Confusion, Green Perceived Risk and Green Trust-An Indian Perspective

Sangeeta Trott

SIA Business School, Mumbai

Purpose :The main purpose of the study is to examine the impact of green washing on green confusion, green perceived risk and green trust in the Indian scenario. Where the whole world is talking about green, companies in a haste to increase consumer preference tend to mislead consumers about environmental concerns which is termed as "Green Washing".*Design/methodology/approach*:The study makes use of written questionnaire method to collect data from approx 500 consumers all over India. The study makes use of SEM to study the hypothesized relationship between various constructs.*Findings* :Findings of the study show that awareness among green washing activities is less and green washing activities have a positive impact

on green consumer confusion and green perceived risk which in turn negatively affects green trust. *Practical implications:* The study is of great use to the consumers in understanding that green washing activities create confusion and increase perceived risk which undermines green brand trust. It will also help the government to design some regulatory measures to curb the green washing practices in the country like India. *Originality:* The paper makes a unique contribution to the study in India as it is the first study in India which intends to find out the consumer perception of green washing activities in India and secondly intends to find out the impact of green washing on green consumer confusion, green perceived risk and green brand trust.

IC 17/004

Impact of Stress on the Worklife Balance of Women Gynaecologists

P N Harikumar

PG Department of Commerce & Tourism, Catholicate College, Pathanamthitta, Kerala-689645

Susha D

Research & PG Department of Mathematics, Catholicate College, Pathanamthitta, Kerala-689645

Stress is a major problem faced by an individual in the workplace. Occupational stress has so many implications for an individual's mental and physical health. In the medical field, during the last few decades, there was a drastic change in the nature of work and work situations. These changes have imposed increased demands on the part of lady doctors such as technological upgradation, over-time work and achievement of success. Gynaecologists are expected to perform more than one task at a time. As part of their work, they are asked to work for more than the normal working hours. These conditions lead to job stress and its related problems.

IC 17/006

Click-Through Rate Estimation using CHAID Classification Tree Model

Rajan Gupta

University of Delhi

Saibal K Pal

Defence R&D Organization

Click-Through Rate (CTR) is referred to as the number of clicks on a particular advertisement as compared to the number of impressions on it. It is an important measure to find the effectiveness of any online ad campaign. The effectiveness of online ads through calculations of ROI can be done through measurement of CTR. There are multiple ways of detecting CTR in past, however this study focuses on machine learning based classification model. Important parameters are judged on the basis of user behavior towards online ads and CHAID tree model is used to classify the pattern for successful and unsuccessful clicks. The model is implemented using SPSS version 21.0. The dataset used for the testing has been taken from Kaggle website as the data is from an anonymous company's ad campaign given to Kaggle for research purpose. Various rules are generated and advertisement's click/non-click pattern is determined

through the model. A total of 83.8% accuracy is reported for the classification model used. This implies that CHAID can be used for less critical problems where very high stakes are not involved. This study is useful for online marketers and analytics professionals for assessing the CHAID model's performance in online advertising world. Researchers can be benefited through the subject understanding and testing their algorithmic results on the given dataset and comparing the accuracy of their models, respectively.

IC 17/ 007

On optimum reliability acceptance sampling plans under generalized hybrid censoring

Tanmay Sen

Indian Institute of Technology Patna

Ritwik Bhattacharya

Centro de Investigacion en Matematicas (CIMAT),
Monterrey, Mexico

Yogesh Mani Tripathi

Indian Institute of Technology Patna

Biswabrata Pradhan

SQC and OR Unit, Indian Statistical Institute, Kolkata

Acceptance sampling plan plays an important role for taking decision on acceptability of a lot or batch of products. A variable sampling plan to determine the acceptability of lot based on lifetime of the product under consideration is called reliability acceptance sampling plan (RASP). In a RASP, we need to determine the sample size and an acceptability constant. Decision on lot acceptability of a lot is taken based on the observed lifetime from a life test. In practice, censored life tests are conducted because of time and cost constrains. Type-I and Type-II censoring schemes are the common censoring schemes applied in life testing experiments. Hybrid censoring scheme, a mixture of Type-I and Type-II censoring schemes, are also used in many instances. This work consider a generalized hybrid censoring (GHC) scheme. We consider Type-I GHC scheme. This work considers determination of optimum RASP for the Weibull lifetime distribution for given producer's and consumer's risks. We have considered a variance minimization criterion under a cost constraint. A simulation study is undertaken to assess the performance of the optimum RASP.

IC 17 /009

Devising Virtual Supply Chain Analytics for Agriculture Products during Festival season

Baba gnanakumar.P

Sri Krishna Arts and Science College, Coimbatore

During the festival season, the demand will be high for agricultural products, but matching the supply and demand is a difficult process. In this context, this research has been conducted during "Onam" festival season in the state of Kerala, with the aim of identifying the supply chain analytics. After conducting a survey among

521 farmers, the supply chain process flow has been identified and implemented. The benefits of the analytics are measured after implementation. This research enables the farmers to create the direct marketing activities during the festival time. The research conclude that virtual supply chain network will be effectively articulated in a situation of heavy demand, multiple sourcing, and logistics pattern which is controlled by the supplier.

IC 17/014

Does Banking Competition Granger Cause Banking Stability: The Study of SEM Countries

Manju Jayakumar, Rudra P Pradhan, Debaleena Chatterjee, Ajoy K. Sarangi
Indian Institute of Technology Kharagpur (India)

Using a panel vector auto-regressive model, we study the interactions between banking competition and banking stability in Single European Market (SEM) countries between 1996 and 2014. The countries include in this analysis are Albania, Andorra, Armenia, Austria, Belarus, Belgium, Bosnia, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, the Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, and the United Kingdom.

We attempt to assess whether banking competition has influenced the banking stability in these selected European countries, and whether a policy focus on banking competition is appropriate as an approach to boost banking stability. We use six measures of banking competition and five measures of banking stability. The six banking competition indicators are Lerner index, Boone indicator, H-statistic, CR-5 firm concentration ratio, CR-3 firm concentration ratio, and foreign ownership. The cointegration and Granger causality techniques in the dynamic panels are used to establish the possible long-run and short-run relationship between banking stability and banking competition.

Our results point to a myriad of short-run results, pointing to the existence of unidirectional or bidirectional causal links between the variables in some cases. However, our remarkable uniform result is that banking competition is the key driver of banking stability in European countries in the long run. The study finally suggests that banking competition matter- positively and negatively- for the efficient functioning of a financial market and their regulation has become one of the key objectives of financial policy to achieve the banking stability.

Effect of Venture Capital Investment on Economic Growth: Evidence from GMM Estimates

Rana P. Maradana, Rudra P Pradhan, Saurav Dash, Danish B. Zaki, Kunal Gaurav

Indian Institute of Technology Kharagpur (India)

Using the System Generalized Method of Moments (SGMM) estimate, we trace the effect of venture capital investment on economic growth in European countries between 1989 and 2014. The countries include in this analysis are Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Norway, Poland, Portugal, Romania, Spain, Sweden, the Netherlands, and the United Kingdom.

We attempt to assess whether venture capital investment has influenced the economic growth in these selected European countries, and whether a policy focus on venture capital investment is appropriate as an approach to boost economic growth. We use three measures of venture capital investment and selected macroeconomic determinants. The three venture capital investment indicators are venture capital at early stage investment, venture capital later stage investment and total venture capital investment. The system GMM estimates confirm that the effect of venture capital investment on per capita economic growth is positive and significant across these three selected venture capital indicators. Additionally, the other factors that affect this process are innovation, income tax rate, interest rate, inflation rate, unemployment rate, research and development expenditure, and stock market capitalization. The findings of this study can give some clarity in existing debate on the causality between venture capital investments and economic growth.

Impact of Social Media Advertising on Consumer Buying Behavior

Ayush Tiwari, Ayushi Tewari, Ranjana Vyas

Department of Management Studies, IIT Allahabad, U.P.

Consumer behavior refers to the study of people, groups and organizations where process involves in selection, purchase and consumption of various services and goods in order to cater their requirements and demands. There are cultural, social, psychological and personal factors which effect the buying patterns of consumers. Social media encompassing various online communication channels forms an important part of social factors which have an outside influence on purchasing decisions directly or indirectly. Social media advertisements contribute a great deal in marketing messages of various business organizations. It creates a two way interaction between businesses and targeted customers and subsequently effects the consumer buying behavior.

The purpose of this research paper is to analyze the impact of social media advertising on consumer buying behavior. The main focus of our research is to incorporate the survey of different individuals comprising students and working professionals from age group 15-50 years. The results revealed the significant impact of social media advertisements on consumer buying behavior. Further, it was concluded that social media advertisements' content, credibility and their location do have a positive impact on buying behavior of consumers.

The findings of the research were impactful as it helps to reveal that across gender, there is no significant difference in buying behavior due to advertisements shown on social media websites. Further, it was revealed that there exists no dependency or relationship between the gender and the information they prefer to look on social media websites

IC 17/ 019

Determining Sportsperson's affinity for brand sponsorships: An Analytics approach

Anindo Bhattacharjee, Amit Kumar, Rimi Moitra, Tanmay Trivedi

ASMSOC, NMIMS (Deemed to be University), Mumbai

Across the world, we're seeing ever closer convergence between the sport and entertainment industries, as both sectors continue to rise to the challenges of digital technologies which shape the way we spend our leisure time. At the same time sponsorships and media rights emerge as the main engines of growth, putting the traditional dominance of gate revenues under pressure.

Sports has always been a unifying element in a country with wide diversity of population, breaking barriers and cultural differences. India has been a nation where cricket has dominated all other forms of sports and athletics. Thus sports sponsorship has been the topic of most marketers. Cricket is the prime sports that attracts the highest share of sports sponsorship in India

Sports analytics is an emerging field which has traditionally existed mostly in the form of studies by statisticians. Sports Analytics can be defined as "The process of using sports related data anything from player statistics to game day weather to find meaningful patterns and communicate those patterns to help make decisions" (Stolbunov, 2014).

This paper attempts to propose a quantitative model for brand sponsorship of sportsperson. The model that we propose has two prominent multidimensional vector spaces which can be represented as two matrices viz. Performance Matrix (A) and Personality Matrix (B). The Sponsorship Matrix (S) is scalar product of the two vector spaces and represents the affinity of a player to attract brand sponsorships.

The future scope of this research can be generalized to other sports apart from cricket.

IC 17/021

Sales Lobby – Account Insights Gateway

Shripal Gandhi, Nameeta Naik, Ed Garcia

HPE Customer and Market Insights

The goal is to turn data into visualizations, and visualizations into actionable insights by using Sales lobby visualization tool.

Sales lobby provides analytic insights to the Pan HPE's Sales & Marketing stakeholders and Executives across the globe (all 3 regions) with the objective of finding and creating new sales opportunities and to provide Business Insights. These insights are leveraged by sales and marketing team in converting it to business to bring in more revenue to HPE. Sales lobby is a self-service visual data mining and business intelligence solution providing account level insights into HPE Customer install base, HPE Competitor install base, HPE Win loss deals, Account news, Account financial performance etc. In short it is providing 360 Degree view of HPE customer. The complete customer and competitor data is stored in Vertica database and dashboards are developed in Tableau. Sales lobby tool is live from February 2016 and more than 3000 users have registered with it. Sales lobby is currently covering 200K+ accounts and the number is increasing with every data refresh.

IC 17/023

Group Messaging and its Impact on Quality of Life at Work Place – an exploratory study on WhatsApp's group communication platform

Aditya B. Rao

Institute of Management, Christ University,
Bangalore, Karnataka, India

Meghana A. Singh

Institute of Management, Christ University,
Bangalore, Karnataka, India

Srisudhan R

Deluxe DigiCaptions Pvt Ltd, Bangalore, Karnataka,
India

With increasing workloads and change being the only constant, life at workplace is governed by dynamics of interpersonal and group communication needs. While shorter goals have compelled time constraints, the key to any piece of work turning out successful can be attributed to quicker and easier communication. And that is what WhatsApp has come to offer! This instant messenger has provided a commendable platform for groups of many sizes and varying duration to connect, interact and execute work which otherwise can be delayed adding woes to the existing set of challenges. At the operational level, exchange of information and input on a timely note can redefine the role of technology by decoding its simplicity in complex situations. WhatsApp has been instrumental in enabling easy flow of work both vertically and horizontally at a work place and has been a go-to platform for keeping in touch. However, it has not quite been a platform where official communication is delegated on a regular basis.

IC 17/028

Optimization of ATM Network Using Geo-Spatial Data Analysis Techniques

Lalit Barhate, Rashmi Hegde

Analytics, DATAWISE, a business analytics company based out of Hyderabad, India.

Banking sector in India is witnessing an era of reforms with moves such as demonetization which have pushed digitalization in banking sector. With consistent thrust on making the economy cashless, digital modes of banking such as ATMs, Internet Banking, and Unified Payment Interfaces are gaining significance rapidly.

With increasing card usage by all sections of society, banks have to spend significant amount of resources on setting up ATM networks across geographies. It is very crucial to understand the existing network of ATMs as well as the expected ATM utilization and impact on network efficiency before considering adding new ATMs to the existing network.

This paper provides a methodology to optimize ATM network using geo-spatial analysis, assisting in installing new ATMs based on the efficiencies in the current network as well as predicting the expected usability of proposed ATMs. As an initial step, Geo spatial data layer superimposed on top of the existing networks of ATMs mapped to population levels increases the usability and effectiveness of ATM network establishment. The model can be further refined for expected usage patterns, customer footfall, and propensity to purchase, or usage.

IC 17/029

Assessing Student Employability to Help Recruiters Find the Right Candidates

Saksham Agrawal

IIM Ahmedabad PGP 2010

India produces lakhs of engineers every year, but employability has been a concern for educators and recruiters alike. Cost of technical recruitment has been driven extremely high due to inconsistent quality of talent available. Objective assessment of students has helped, but recruiters still use simple cutoffs followed by subjective methods like group discussions and personal interviews which are inherently unscalable and expensive.

In present work, we looked at most recent employment data for almost 50000 fresh engineering graduates and established precise objective relationships between students' scores across several dimensions, and the students' employability as measured by actual job offers made to them by employers. Through regression modeling, we were able to develop a composite score that identified employable candidates with 5X precision than any individual score.

We then went a step further to cluster employers on the basis of their ability to choose the stars from amongst a large pool of candidates, to identify employers who can be advised to further optimize their recruitment spend.

IC 17/033

An Inter Services Comparative Study on Job Satisfaction in Bangladesh: Application of Discriminant Function Analysis and Structural Equation Modeling

Nigar Sultana, Md. Mahedi Hasan, Mohammad Tauhidul Islam

Bangladesh University of Professionals (BUP)

Despite a significant number of research for finding the factors of job satisfaction and its impact on the performance of the employees in the organizations, a comprehensive identification of the factors amongst different types of organizations yet to be undertaken in Bangladesh. The findings of those researches on the factors of job satisfaction rarely explicate their specification among different types of organizations, which is assumed to be different. Thus, we argue that a comprehensive study to find the factors of job satisfaction of the employees in different types of organizations is relevant and necessary. Such a study would allow to reach more explicit and comprehensive understanding on the specification of the factors of job satisfaction amongst different types of organizations, and to enhance the performance of the employees. We would like to address the research by means of a systematic and conceptual analysis of the responses on job satisfaction from the employees working in different types of organization such as “Education”, “Manufacturing”, “Bank”, “Government Services”, “Private / Multinational Organizations”. In that regards, data has been collected from 288 employees, at least 30 from each type of organization by stratified random sampling technique with proportional allocation. As research method we have used Discriminant Function Analysis (DFA) and Structural Equation Modeling (SEM) to meet the objective of the study. We have found that the factors of job satisfaction are notably different in organization of varying nature. Those specific factors of job satisfaction for each type of organization can be addressed in order to put into effect, so that it can get efficacious outcomes. The contribution of this paper will also help the policy makers to devise an effective specific plan for the organization to ensure job satisfaction of the employees that can lead to the better performance of the employees and profitability of the organization.

Applying Predictive Analytics in a Continuous Process Industry

Nitin Merh

Symbiosis University of Applied Sciences, Indore, Madhya Pradesh

In this paper an attempt is made to develop data driven models on pilot data set for predicting fault in machines of continuous process industry on various selected attributes using techniques of Multiple Linear Regression Model (MLR), Regression Tree (RT) and Artificial Neural Networks (ANN). Association rules are also derived from the available data set. Efforts are also made to predict total shutdown time of machines. These machines are used for manufacturing components machined for Heavy Commercial Vehicles (HCV), Light Commercial Vehicles (LCV), Multi Axle Vehicle (MA) and Tractors. To check the robustness of models a comparison is made between the results derived from various techniques discussed above. Performance evaluation is done on the basis of the errors calculated between the actual and predicted values of down time. Based on actual and predicted results various error scores are calculated to evaluate best model and check robustness of the models under study. Training and validation of the model is done using datasets collected from Indore CNC Pvt. Limited a manufacturing unit located at Pithampur industrial area near Indore, Madhya Pradesh, India. In the current paper an association is also developed between the attributes and occurrence of the fault. The developed model will be used on the bigger data set which will help the stakeholders of the organization for smooth functioning of the unit and for better governance in the organization. XLMiner is used for model development and simulations.

Price transmission from crude oil to corn in the United States: New evidence from multivariate adaptive regression splines approach

Debdatta Pal

Indian Institute of Management Lucknow

Subrata K. Mitra

Indian Institute of Management Raipur

This article explored price transmission from crude oil to corn in the United States based on monthly price data from January, 1996 to March, 2015. There were strong economic reasons to believe that crude prices had strong influence on corn price changes as 37 percent of corn produced was used as feedstock of ethanol. By employing a non-parametric multivariate adaptive regression splines (MARS) approach of Friedman (1991), we found that the importance of crude oil in corn price change was relatively weak and prime determinants of corn prices were price variations of other major agricultural commodities, namely wheat and soybean. This was contrary to the long standing claim of oil-food nexus by a wide array of research. We also found corroborative evidence of asymmetric price transmission from both

wheat and soybean to corn that lead to an uneven distribution of consumer and producer surplus. This called for right policy intervention for mitigating adverse impact of price acceleration.

IC 17 037

A Stochastic Feedback Queuing Model with Encouraged Arrivals and Retention of Impatient Customers

Bhupender Kumar Som

JIMS Rohini, New Delhi

Globalization has introduced never ending competition in business. This competition empowers customers' and somewhat ensures quality at reduced cost. High competition along with uncertain customer behavior complicates the situation further for organizations. In order to stay ahead in the competition, organizations introduce various discounts and offer to attract customers. These discounts and offers encourage customers to visit the particular firm (online or offline). Encouraged arrivals result in heavy rush at times. Due to this customers have to wait longer in queues before they can be serviced. Long waiting times, results in customer impatient and a customer may decide to abandon the facility without completion of service, known as reneging. Reneging results in loss of goodwill and revenue both. Further, heavy rush and critical occupation of service counters, may lead to unsatisfactory service and some customers may remain unsatisfied with the service. These customers (known as feedback customers) may retire to the system rather than leaving the facility satisfactorily. Unsatisfactory service in these situations may cause harm to the brand image and business of the firm. To counter the situation, an effective management policy shall be designed. If the performance of the system undergoing such pattern, can be measured in advance with some probability, an effective management policy can be designed and implemented. A concrete platform for measuring performance of the system can be produced by developing a stochastic mathematical model. Hence, in this paper a stochastic model addressing all practically valid and contemporary challenges mentioned above is developed by classical queuing theory model development approach. The model is solved for steady-state solution iteratively. Economic analysis of the model is also performed by introduction of cost-model. The necessary measures of performance are derived and numerical illustrations are presented. MATLAB is used for analysis as and when needed.

A More Powerful test Identifying the Change In Mean of Functional Data

Buddhananda Banerjee

Indian Institute of Technology Kharagpur

Satyaki MazumderIndian Institute of Science Education and Research,
Kolkata

An existence of change point in a sequence of temporally ordered functional data demands more attention in its statistical analysis to make a better use of it. Introducing a dynamic estimator of covariance kernel we propose a new methodology for testing an existence of change in the mean of temporally ordered functional data. Though a similar estimator is used for the covariance in finite dimension, we introduce it for the functional data in this context for the first time. From this viewpoint, the proposed estimator of covariance kernel is the most natural one when the sequence of functional data may possess a change point. We prove that the proposed test statistic is asymptotically pivotal under the null hypothesis and is consistent under the alternative. It is also shown that our testing procedure outperforms the existing one(s) in terms of power and provides satisfactory results when applied on real data.

A Hollander-Proschan type test when Ageing is not Monotone

Shyamal Ghos, Murari Mitra

Department of Mathematics, Indian Institute of Engineering Science and Technology, Shibpur, Howrah, west Bengal

In this paper we present a class of tests for testing exponentiality against New Worse than Better than Used in Expectation (NWBUE) alternatives which constitute a substantially large class of life distributions defined on the basis of a trend change in the Mean Residual Life (MRL) function. It is shown that the test statistics reduce to linear functions of order statistics, the so-called L-estimators. Both the small sample and the asymptotic distributions of the family of test statistics have been derived. The test is shown to be consistent against all alternatives belonging to the NWBUE class. Our family of test statistics includes the test proposed by Klefsjo (1983) as a special case in spite of the fact that our test is based on a totally different approach. The test also includes the test of Anis and Mitra (2011) and that of Hollander and Proschan (1975). We also rectify the main asymptotic result in Mitra and Anis (2005) which was subsequently utilized in Anis and Ghosh (2015). The performance of the test against various alternatives has been assessed by means of a simulation study. Finally, we apply our test to a real life data set.

Determination of Optimal Inventory Policy – A Case Study

M. Z. Anis

Indian Statistical Institute, Calcutta

Proper inventory planning is of key importance for the success of almost every industry in general and of engineering industry in particular. Large inventory essentially means blocking of precious (and perhaps, scarce) capital. Whitin [1] reported that inventories are referred to as the “graveyard” of American business, as surplus stocks have been a principal cause of business failure. In contrast, the Japanese industries have reduced inventories as much as possible to increase productivity by the Just-in-time Management (JITM) control system (see, for example, Hall [2] and O’Connor [3]). On the other hand, inadequate inventory can lead to shortages. These shortages can be very critical leading to huge tangible losses (e.g. monetary losses) and intangible losses (e.g. loss of goodwill).

Between these two extremes, there are situations where inventory can be maintained at a balanced level. It is the aim of the science of inventory theory to manage the inventory position at the “optimal” level. The simplest model is the deterministic model where the nature of demand is known with certainty. However, in real life situations, such a phenomenon is rarely met with in practice. Demand is usually probabilistic in nature. Many models have been proposed for such a situation, see for example Sharma [4], Axsäter [5] and Silver et al. [6]. These models get more involved when lead-time is present, which may not be constant. Magson [7] describes such a situation. The models get even more complicated when reviewing period is allowed to be another parameter. Donaldson [8] discusses such a situation. Cheng and Sethi [9] consider a periodic review inventory model where demand is influenced by promotion decision.

This case study is based on a probabilistic reorder point lot size system with constant lead-time. It describes the application of scientific inventory management at the Stores Department of an engineering unit, which had hitherto **not** used any OR principles. This unit manufactures ferro-alloys used in the production of alloy steel. It had a large number of items in the stores. Careful scrutiny of records showed that there were items that had been lying unused even for very long periods of time. In fact some 5% of the items had not been used even for the past seven years.

No shortage of any technically important item had ever been reported. Rather, there were problems of storage and warehousing. Thus, a large volume of capital is locked up by inventory. Since no scientific inventory management procedure was in place, it was felt that inventory management based on scientific principles need to be introduced systematically, **but slowly**. It is with this background that this study was taken up. It was decided to develop a scientific inventory policy for one high value critical item as a confidence-building exercise. Initially, an important item, **welding electrode – 4mm**, was chosen for the study. This welding electrode is a critical item used for welding purposes in the blast furnace. If any problem in the blast furnace is reported, it must be attended to immediately, and hence no shortages can be permitted for this

item. During the eight month data collection period it was observed that the lead time for this item is only 8 days; however, a high inventory of 248 packets was maintained even though the maximum daily demand (during this data collection period of eight months) was only 13 packets, and more importantly in about 20% of the time there was no demand. Hence it is worthwhile to model for this inventory problem and implement the same at the shop floor.

Based on discussion with the stores and purchase personnel the characteristics governing this item were listed. This exercise revealed that what is required is a model, which can incorporate the probabilistic nature of demand with constant lead time, and yet be simple enough for implementation at the shop floor. It was felt that a *Probabilistic Reorder Point Lot Size System with Constant Lead Time* might be the most appropriate model. Since the plant is primitive in the usage of scientific inventory management principles, it was decided to review the stock daily. Thus the reviewing period of one day was chosen. Based on the data collected and the subsequent analysis, the following inventory policy was prescribed:

Review the stock position daily; if the amount in inventory at any point becomes 44 packets or less, order 114 packets.

It should be noted that it is not difficult to review the stock position on a daily basis as their materials handling system is computerised. Each transaction triggers an immediate updating of the status. Thus transaction reporting ensures that the stock status is always known. The expected reduction in costs in using the recommended scientific inventory policy is over 23%. When these results were presented to the management, they were impressed and decided to accept this scientific policy on an experimental basis.

The company has adopted this policy of daily review, and placing an order only when dictated by the inventory policy. It has been in operation and no stock out has occurred. Encouraged by these findings, the company decided to implement scientific inventory policies for other high value items.

IC 17/043

Estimation of Parameters of Misclassified Size Biased Borel Tanner Distribution

B. S. Trivedi

Amrut Mody School of Management (AMSoM),
Ahmedabad University, Navrangpura, Ahmedabad,
Gujarat, India.

M. N. Patel

Department of Statistics, School of Sciences, Gujarat
University, Ahmedabad, Gujarat, India.

Now a day statistical data analysis is a great interest in each and every field of management, business, engineering, medical etc. Different types of tools and techniques of statistical methodology are used for data analysis. In this paper we have considered a statistical data analysis and estimation for the data by using size biased Borel – Tanner distribution. At the time of classification and analysis there

may arise error, like a classification of an observation in the other class instead of the actual class. Such type of data is known as misclassified data. Also, when samples are drawn without a proper frame, which means sampling mechanisms are not given an equal chance to select units but they select units with a probability proportional to some measure of the unit size, the resultant distribution is known as a size biased distribution or weighted distributions. In this paper, we have studied Misclassified size biased Borel – Tanner distribution and estimated its parameters by applying method of maximum likelihood, method of moment and Bayes' estimation method. Simulation study has been carried out for the purpose of comparison among all the three methods of estimation.

IC 17/044

Predicting Success Probability in Professional Tennis Tournaments Using a Logistic Regression Model

Saurabh Srivastava

EXL Service, Gurgaon

With a global audience of over 1 billion, professional tennis is the most widely followed individual sports in the world. The present study attempts to model the probability of success for a tennis player in a men's singles tournament of a given type (ATP 250, ATP 500, ATP Masters and Grand Slams) so as to enable his management team to take better decisions with respect to his calendar planning. The model in this study tries to arrive at the probability of success in a given category of tournament by modelling a dependant variable x to capture the success of an athlete in that tournament (measured by his ability to reach the quarterfinals), using the logistic regression method. The scorecard that is built uses six variables to arrive at the probability of success, which can be used to rank order the tournaments in a given category for a player, and can be subsequently augmented by a decision tree or linear programming method to help a player make the most optimum selection of tournaments.

IC 17/049

Do bubbles and crashes in stock markets represent a seismic behaviour? An empirical investigation

Kousik Guhathakurta, Brijesh K Mishra

Indian Institute of Management Indore

Various financial crises have brought about significant suffering to the human beings in the past century. Despite regular occurrence of crashes, and despite the advancements in the econometric techniques, there has not been a model in classical financial economics that attempts to predict a crash with some definitive assurance. Off late, however, there have been attempts to develop models for understanding

and predicting crashes by taking cues from build-up of seismic activity leading to an earthquake, which signifies a phase transition, or a critical point. A crash is considered a critical point accordingly, when pent-up overheating of stock market gets released in one instance. Following the literature that draws a parallel between seismic events around an earthquake and gradual build-up in stock market before a crash, this paper empirically tests nine major crashes and finds that crashes can indeed be characterised using log-periodic power laws as in case of earthquakes. This finding, however, should only be considered in view of limitations inherent in using log-periodic power laws for stock markets.

IC17/051

A study of the Performance of 'Make in India' in Information Technology Hardware: An Early Inference from Perceptions of Senior Information Technology Hardware Professionals.

Ashutosh D. Gaur

School of Doctoral Research and Innovation, GLS University, Ahmedabad

Jasmin Padiya

Faculty of Management, GLS University, Ahmedabad

The purpose of this study is to identify the perceptions of senior 'Information Technology Managers' about Make in India program of government of India on job creation, manufacturing and policy reforms. Data was collected through a self-administered questionnaire with 5 point Likert scale using non-probability convenient sampling method. A sample of 42 responded was collected from mail and social networking site (Facebook, LinkedIn) from Sr. IT professionals. Total 33 responded were selected for weighted average and independent t-test, rest 9 responses were not included due to incomplete response. Qualitative study is based on online journals, reports and newspaper articles.

The study identifies the hypothesis which exhibits the perceptions of senior information technology professionals about make in India, job creation, infrastructure and policy reforms.

The prime factor for IT hardware manufacture is improvement in R&D emerge from data analysis. Start-up will create more jobs is second out come and exports of IT hardware from India will significantly change the manufacturing in IT hardware sector is third major output of this study. Further expansion into international market requires improvement in research & development in the product design and manufacturing .Process should be standardized for best quality products. 'Make in India' has potential to generate employment and eradicate poverty from the nation.

Loan Loss Provisioning Practices in Indian Banks

Divya GuptaInstitute of Management & Information Science,
Vivekanand Marg, Bhubaneswar, Odisha**Sunita Mall**MICA, The School of Ideas, Shela, Ahmedabad,
Gujarat

RBI has started getting stringent on loan loss provisioning in the banks. Today banks are not only maintaining provisions for substandard assets but also they have to keep aside surplus for standard assets. The objective of our study is to find out the factors determining loan loss provisioning in banks in India. Variables considered for our research are loan loss provisions to total assets, loans to total assets, capital adequacy ratio of the banks, bank holding of securities to total assets, earning before tax to total asset ratio, non performing assets to total loans, credit growth rate of the bank and GDP annual growth rate. Our dependent variable is loan loss provisions to total assets. We have taken the data for the last 11 years from 2004-05 to 2014-15. The scope of our study is 46 commercial banks in India where we have checked the determinants of loan loss provisioning by analysing whether the loan loss provision is procyclical or countercyclical. The impact of various variables on loan loss provision has been tested and the result indicates that asset size, credit growth, NPA level, earning of the banks and macroeconomic variable (GDP annual growth rate) have a greater impact on the provisioning of the banks. The analysis has been done by using OLS regression and dynamic GMM approach recommended by Arellano and Bond (1991). The findings of the study also reveal that Indian banks have shown countercyclical provisioning and the banks tend to increase the provisions with increase in banks earnings and GDP annual growth rate.

Hausdorff Path Clustering and Hidden Markov Model Applied to Person Movement Prediction in Retail Spaces

Francisco Romaldo Mendes

Deloitte Consulting LLP

Current advances in technology allow for the efficient capturing and storage of high resolution and high frequency person movement data. The advent of wi-fi position triangulation has allowed us to capture human movement with a great deal of accuracy inside a closed urban structure, e.g., a University or a shopping mall and at a very high frequency (an observation every second). While there have been significant advances in our ability to capture this data, advances in robust modelling techniques for movement prediction have been largely absent. In this paper we aim to present our theoretical insights based on person movement data collected from Deloitte University, West Lake Texas. We outline two independent approaches - the first aims to model the problem of person movement prediction while the second aims to classify people based on their movement history. This paper presents a technique

to model person movement data using a Hidden Markov Model on a large data set with multiple rooms and locations of importance as well as a technique to classify users based on their path data (movement traces). We achieve an accuracy of 65% in prediction accuracy of the Hidden Markov Model.

IC17/059

“Estimation Of Fluid Flow Rate And Mixture Composition Using Low Cost Acoustic Sensors For Application In The Oil And Gas Industry”

Pradyumn Singh, Karthikeyan G, Mark Shapiro, Shiyuan Gu, Bill Roberts
Deloitte Consulting LLP

One of the key problems in oil production is to determine flow rate of oil/gas mixture in the pipeline. Currently available methods rely on flow meters that are expensive, require installation of dedicated tank separators, and allow to measure the flow rate from a single well at a time for a short period of time. We provided a proof-of-concept using accelerometers with acoustic range bandwidth to estimate flow rates. The key idea is that flow of mixture inside the pipe generates acoustic range vibrations that can be sensed by an accelerometer magnetically attached to the pipe, and analytic methods can be used to estimate the flow rate from the accelerometer signal. Our approach uses inexpensive sensors that can be installed on pipes coming from all wells and provide continuous estimate of flow rates for each well.

To simulate oil production conditions, we designed an experiment to collect data at two university laboratories. A closed-circuit pipeline was assembled, the flow rate and oil/air mixture composition was controlled using an oil pump and air compressor. First, we estimated autoregressive models for each condition. For test data, we calculated a normalized log-likelihood score using the AR coefficients. For 63 test conditions, a multiclass AUC was 0.9676. Second, we used a sliding window approach to convert the sensor time series to a sequence of cepstral coefficient vectors for each condition. These cepstral sequences were used as predictors in a Gaussian Mixture Model and Hidden Markov Model with a diagonal accuracy of 92.5%.

IC 17/061

From Surveys to Action: A Machine Learning Pathway to Realize Customer Satisfaction

Neogy, Arkosnato, Thapliyal, Piyush
Deloitte Consulting LLP

Survey data forms one of the key structured datasets that capture feedback on customer experience around a product or service offering. While survey data has been frequently used to measure performance within firms and across competitors

in industries, a clear roadmap charting a path from survey feedback to actionable insights that address operational shortcomings remains obscure. In this paper we outline a data-driven approach to quantifying areas of opportunity using rated survey data (ratings being integers in a given range). The algorithm allows us to identify factors where a given firm has a statistically significant disadvantage compared to competitors in an industry or market. Following this, we demonstrate a quantitative method to simulate the potential impact (on a metric of interest based on the ratings) of improving on the identified factors. Methods to calculate the impact both for individual improvements in factors as well as the best set and sequence of a group of factors are explored. Additionally, natural language processing techniques are used on free-form customer comments to precisely identify dissatisfaction points within each category of service. To make this actionable we propose a reinforcement learning framework that experiments with live-market response by tweaking the said factors. This statistically concrete end-to-end battery of techniques allows leveraging customer feedback to drive surgical improvements in customer experience, aiming to replace adhoc manual and case-specific interventions with uncertain returns.

IC 17/ 064

Improving Email Marketing Campaign Success Rate Using Personalization

Gyanendra Singh, Himanshu Singh, Sonika Shrivastav

Experian

Email marketing provides one of the best methods for direct communication with consumers. However, the success rate of an E-mail marketing Campaign is often low because of its generic content and inadequate segmentation of customers. This paper aims to showcase the application of a two-step personalization process to improve effective open and click rates for email marketing campaigns. Consumer behaviour is monitored over a period of time in terms of email opens and click pattern. This behaviour is stream-lined into keywords sorted in order of user preference. The keywords are updated at regular intervals to account for behavioural changes in user preference. While sending the email, keywords relevant to the campaign are picked individually for each user. These keywords are used to form attractive subject lines using probabilistic language models such as noisy channel model and Hidden Markov model.

Automatic Detection of Tuberculosis using Deep Learning Methods

Manoj Raju, Arun Aswath, Amrit Kadam, Venkatesh Pagidimarri

Enlightiks Business Solutions, Bangalore

In this paper, we present a Deep Learning based approach for automatically detecting tuberculosis manifestation from chest X-ray images. India is the country with the highest burden of tuberculosis. A chest radiograph in symptomatic patients is used to diagnose active tuberculosis. This screening method is ideally done at the primary health care centers where a clinician is available and sometimes through mobile X-ray unit. The major challenge for this method of screening is timely reporting and further follow up of patient for initiation of treatment. We built multiple convolutional neural networks, the state of the art deep learning algorithm, to build the model for automatic tuberculosis diagnosis. We classified the chest X-rays into 2 categories, namely, tuberculosis presence and tuberculosis absence. The dataset used to train the model contained 678 images, having 340 normal chest X-rays and 338 chest X-rays with tuberculosis manifestation. The validation dataset contained 235 images, which observed a sensitivity of 84.91% and a specificity of 93.02%. This demonstrates the potential of convolutional neural networks to automatically classify chest X-rays in real time.

Ranking of Batsmen in ODI: A Factor Analysis Approach

Jimut Bahan Chakrabarty, Prashant Nair

Indian Institute of Management Kozhikode

Ranking of players is an important aspect of any sport. Player rankings are of concern to sports authorities, the players and the enthusiasts and with commercialization of sports it is even more important to the investors. This paper explores the factor analysis approach to rank players in One Day International (ODI) cricket. The paper ranks batsmen who have played during the calendar year 2015. It uses a dynamic approach of generating factor scores on a match by match basis which may be used for further analysis such as valuation of players, as the ranks can be considered as a good representation of a player's form and performance. The model uses appropriate variables affecting the performance of a batsman, many of which have been ignored by the earlier ranking systems including the most widely used ICC ranking system. The factor analysis approach can also be extended to the other formats of the game. It can also be used to rank bowlers, all-rounders and wicket-keepers using apposite variables.

IC17/076

Analytics-led Talent Acquisition for Improving Efficiency and Effectiveness

Girish Keshav Palshikar, Swapnil Hingmire, Rajiv Srivastava, Ankita Jain, Mahek Shah, Sachin Pawar, Saheb Chourasia

TCS

Large IT organizations every year hire tens of thousands of employees through multiple sourcing channels for their growth and talent replenishment. Assuming that for each hire at-least ten potential profiles are scrutinized and evaluated, the Talent Acquisition (TA) personnel end up processing half a million candidate profiles having multiple technical and domain skills. The scale and tight timelines of operations lead to possibility of suboptimal talent selection due to misinterpretation or inadequate technical evaluation of candidate profiles. Such recruitment process implementation due to manual, biased and subjective evaluation may result in a lower job and organizational fit leading to poor talent quality. With the increased adoption of data and text mining technologies the recruitment processes are also being reimagined to become effective and efficient. The major information sources viz. candidate profiles, the job descriptions (JDs) and TA process task outcomes are captured in the eHRM systems. The authors present a set of critical functional components built for improving efficiency and effectiveness in recruitment process. Through multiple real-life case studies conducted in a large multi-national IT company these components have been verified for effectiveness. Some of the important components elaborated in this paper are a resume information extraction tool, a job matching engine, a method for skill similarity computation and a JD completion module for verifying and completing a JD for quality job specification. The tests performed using large datasets of the text extraction modules for resume and JD as well as job search engine show high performance.

IC 17/078

Frame of Reference Biases: The Effect of Social Media on Behavioral Reinforcement

Keerthi Paturi, K. Vinay Kumar

Datawise Management Services India Private Limited, Hyderabad, India.

Since its evolution, the use and utility of Social Media has been a matter of debate in many forums across the world. There have been many debates regarding the merits, demerits, implications, and long term impact of Social Media. While on the one hand, it has democratized the media, making it much less unidirectional, on the other hand, the empowerment of a possibly less credible source of information to create conversations whose ability to influence an individual's Frame of Reference and outcome calls for greater research.

The source of information is usually validated based on a number of factors such as

presence of implicit bias, consideration of Frame of Reference, and point of view of primary source creator. Personal feelings influence the way individuals and groups choose sides and view the facts. Social Media posts ability to influence the Frame of Reference of an individual are different from that of a Main-Stream Media (MSM) post. A person is likely to actively seek to link himself/herself to posts that exhibit an emotion that is to resonate with his/her own. Such behavior could potentially increase the virality of a post irrespective of its credibility. Social media posts from persons within ones network are likely to be taken with a greater degree of credibility, thereby reinforcing the impact on Frame of Reference of a post. This paper presents an analysis of the impact of Social Networking Sites (SNS) on an individual's Frame of Reference and its implications for his expression and behavior.

IC 17/080

The relationship between insurance and Economic growth in India: A Cross-region study using an Econometric approach

Sunita Mall

MICA, Ahmedabad

Tushar Ranjan Panigrahi

St.Francis Institute of Management and Research ,
Mumbai

Insurance is an important part in the financial sector that contributes significantly to the economy of a country. Insurance market contributes to the economic growth as a financial intermediary and also helps in managing risk more effectively. This piece of research work made an attempt to examine the relationship between insurance and economic growth in India considering the state level data and contributing to the existing literature. The data is collected for Fifteen states of India and covers the time period for 2000 to 2015. Endogenous growth model with a modified Cobb-Douglass production function is used. This result implies that the insurance policies which can improve the insurance penetration in different states of India should be promoted. The relationship between physical capital and economic growth indicates that more investments should be made on the policies of infrastructure like health facilities, road etc. This research work could be useful for the state Governments to improve the economic growth and also is useful for the development of the insurance sector in India.

IC 17/082

In-database Analytics in the Age of Smart Meters

Ankit Mahato

Fuzzy Logix, Inc., NC, USA

The rapid depletion of energy resources has made energy security and management a major area of interest for the utility industry. Smart utility service is a building block of the IoT framework, and the growing amount of data captured by smart

meters provide an opportunity to gather various analytical insights which help in better understanding the energy consumption patterns, energy leak, and peak load management. This article demonstrates how analytics are traditionally performed outside the database using specialized analytical servers and toolkits, but with the advent of big data, performing these computations in-memory is no longer feasible. In-database analytics helps in performing faster analytics by pushing analytics inside the database. Through in-database analytics, the utility industry can now gather insights which were not available previously. This article presents a real-life use case of an energy utility firm whose objective is to identify the households which have similar energy usage patterns. A hierarchical clustering based approach is implemented with additional cluster analysis feature (Duda Index) to optimize the stopping criteria for cluster division. The final clusters obtained demonstrate the effectiveness of the developed clustering algorithm for this use case. The identified consumer clusters enable the utility service provider in preparing segment-specific campaigns and tariffs to distribute the peak energy loads.

IC 17/ 084

Role of Gender Bias & Household Amenities in Impeding Access to Financial Services

K Vijay Kumar, Marthal T Packiam
State Bank of India

Ranjitha Korrapati
SRM University

While much research has focused on the benefits accruing from access to financial services among the less affluent, and among poor women, and some research is available on measuring such access, most research has been silent on the reasons for the underlying disparity. Our research explores the causes and effects of the exclusion of female-led households from the formal financial services sector.

This paper uses data from the population and agricultural censuses of India and overlays that with data from State Bank of India to understand the underlying causes behind women's exclusion from the formal financial sector.

While 55% of non-female led households avail banking services in Andhra Pradesh, less than 41% of female headed households do so. While overall participation of households in the formal banking sector is higher among urban households than in rural households, female-headed households are disadvantaged in both rural and urban areas.

Availing Banking services often places a significant demand on the time of the consumer. Access to drinking water, toilets, bathrooms, kindling wood, electricity, implements improving agricultural productivity and efficiency, and better health care are all factors which affect the time available for women to visit banks. This is exacerbated by the relative lack of bank offices and outlets in rural areas, further raising the barriers for financial services access.

Since improving access to and usage of banking services among women requires

provision of amenities that increase the “free” time available to womenfolk, programs and policies should be drafted appropriately.

IC 17/ 085

Exploratory and Predictive Data Analysis Using Machine Learning Tools and Techniques

Heena Timani

Ahmedabad University, Ahmedabad , Gujarat

Shail Jayesh Deliwala

UBS (Pune center of Swiss bank headquartered in Zurich, Switzerland)

Exploratory analysis is the starting point of any data analysis project. It enables us to get a feel of the data. It lets us study summary statistics, explore data visually, check distribution, spread, modality, and so on. Many hypotheses are also tested at this stage. Some actionable insights may come to light as well. Signals and patterns discovered at this stage help with predictive modeling. Exploratory analysis often goes hand in hand with cleaning and preparing the data. In this paper we have applied XGBoost famous gradient boosting algorithm. It has both linear model solver and tree learning algorithms. It is extremely fast as it has capacity to do parallel computations on a single machine and distributed version for scenarios with big data. This model is often described as a *blackbox*, which means it works well but it is not trivial to understand how. Indeed, the model is potentially made of hundreds or thousands of decision trees. It is known for its speed and accurate predictive power. We have compared the results of applying XGBoost and logistic regression to the data. Among both, XGBoost performs better on classification and prediction tasks. Python libraries NumPy, pandas, scikit-learn, and SciPy have been used in this paper.

IC 17/087

Feature Engineering Strategies: A Practitioner’s Guide

Varun Aggarwal, Tuhin Chatterjee

EXL Analytics

In the field of predictive modeling, variable jugglery has been long used to boost performance of statistical models. For developing high performing models, statisticians may resort from business intuitions to much more sophisticated approaches of inventing variables. Nevertheless, the idea of feature engineering has always remained ingrained in a modeler’s toolkit. Moreover, the governing thought of variable transformation is not always the marginal gains in model performance but may also be driven by the requirement to make adequate business sense from the raw attributes.

Feature engineering is one of the most crucial and value adding steps during model development process. In competitive scenarios such as development of challenger models and the data mining contests, it demands unparalleled creative and imaginative powers from the modeler. To aid the modeling fraternity, this paper enlists a variety of

empirically validated approaches and groundbreaking ideas based on our experience in data mining, modeling and analytics. These methodologies range from business sense based variable binning and variable interactions to more complex approaches of decision tree application and choice of mathematical transformations.

Apparently, there is no universal rule to achieve improvement in model performance due to existence of real-life intricacies. The paper examines various scenarios and provides guiding principles of variable creation and transformations to equip the statistical modelers with powerful tools for model performance enhancements. This paper is an attempt to help our contemporary modelers and researchers to make appropriate choices about variable transformation techniques.

IC 17/ 090

Predicting customer churn for DTH: Building churn score card for DTH

Gyanendra Singh, Ankit Goenka, Chandan Chintu

Experian Credit Information Company of India Private Limited, Maharashtra

Customer churn is when the customers either switch from their incumbent service provider to its competitor or stop using the service all together. According to the existing research, it costs five to six times more to acquire customers than to retain existing customers, thus emphasizing the importance of managing churn by organization. In this paper, we build Churn prediction model for one of the Direct to Home (DTH) operator of India for its customer base. We use data provided by DTH operator to build model. Considering the large base of DTH customers, we conduct segment analysis to determine segments where the factors that affect churn are different across segments, indicating the need to have different models. Analysis showed that age on the network was the key segmentation driver and hence separate models were built for each segment. We further apply various data mining techniques such as – Logistic regression, Random forest and Gradient boosting method to build the model for each segment. Gradient boosting method outperforms both Logistic regression and Random forest on measures of AUC and Gini. The proposed model correctly classifies churn customers between 76 % and 78% depending on segment. The primary driver of churn across all the segments for DTH are age of the customer, type of package subscribed, longest delinquency, maximum amount recharged and maximum valid days for which customer has recharged their set up box. The paper also shows that customer experience while interacting with operator and quality of device is equally important attribute.

Online Strategy to Influence Retail Customer Purchase Intention: Analysis through Structural Equation Modeling

Shantanu Prasad, Subhasis Sen

Symbiosis Centre for Management and Human Resource Development (SCMHRD), Symbiosis International University, Pune, India

The purpose of this study is to come up with online strategy in terms of blogging that influence retail customer purchase intention. Review of literature led to identification of five constructs: blogging, electronic word of mouth, purchase intention, brand reputation and conviction. Customers of a big e-retailer were surveyed at five different locations as customers' experiences. The data was collected by administering questionnaire and was analyzed using Amos 18 with the application of structural equation modeling. It was found that electronic word of mouth mediates the impact of blogging on customers purchase intention. Conviction and brand reputation are identified to interact with electronic word of mouth and moderate the same and customer purchase intention relationship. When conviction and brand reputation are high, experiencing electronic word of mouth from blogging has a high positive impact on customers purchase intention. In contrast, when conviction and brand reputation are low, experiencing electronic word of mouth from blogging does not have a significant impact on customer purchase intention.

Taking The Bizarre Out of the Retail Bazaar In The Sky

Nitin Varma

IIM Ranchi

Pradip Kumar Bala

IIT Kharagpur

Every year millions of flights cumulatively carry billions of passengers. With changing times, a flight ticket is no longer the final sale. Airlines and associated businesses have realized flyers seek and indulge in much more during travel. This understanding has lent cognition to the opportunity of air travel retail. Exploitation of this opportunity requires better understanding of flyer needs.

Extant research is deficient in literature that identifies or understands flyer travel experiences and needs better. Flyers are happy to share their air travel experiences via reviews on various Social Media sites. However, this Social Big Data is not easy to process without special software and skills, and attributes that lend a certain 'black box' aspect to it can cause loss of context - making drawing context-aware insights a challenge.

This research demonstrates the use of commonly available internet technologies, that are available on just any Personal Computer (PC), by carrying out Social Big Data Analysis on flyer reviews data. Through the demonstration of context-aware

components AskEasy and LearnContext developed as part of this work, this research shows insights can be drawn from Social Big Data to better understand air travel experiences of flyers. It demonstrates a methodology that brings Big Data analytics closer to just any analytics professional. This research demonstrates how it may be possible to process Big Data real-time in an interactive manner even on a mobile interface, using natural English language for querying without having to learn highly specialized software, databases or SQL like query languages.

IC 17/101

Leveraging The Power of Emoticons to Enhance Polarity Classification of Text

Chandrasekhar Rangu

HCL Technologies

Social media has become an integral part of modern-day living. Emoticons are widely used in everyone's conversation and are very much useful to convey the desired intent in a crisp and precised manner. Emoticons will play a crucial role in analysing social media information especially in case of identifying the polarity of review comments.

Polarity Classification also referred to as *opinion mining* is the process of extracting the subjective information from the source with the help of Natural Language Processing (NLP) and text analysis techniques. Polarity of a user text can generally be classified as Positive, Negative and Neutral. It is widely used in Market analysing, Product Intelligence and Customer relationship management. There are different ways to calculate the polarity of a review. The basic approach of identifying polarity of a review text is with the help of Lexica dictionaries by matching the dictionary of words. It basically ignores the emoticons by default, as they take different form altogether once extracted from source, typically encoding form of the emoticon and cleaned during text processing.

In this study we are presenting an approach to include the power of emoticons through text classification approach by converting the emoticons into its encoding form initially and then replacing them with the text description available in the pre-prepared emoticons dictionary. In this approach Polarity classification accuracy on the data set with Emoticons has achieved 8% more compared to the accuracy on same data set without including emoticons.

Connected Cars and Driving Pattern: An Analytical Approach to Risk Based Insurance

SrinivasaRao Valluru

HCL Technologies, Hyderabad.

Insurance is a means of protection from financial loss. Due to heavy competition in Insurance industry, insurance companies are announcing new policies with additional benefits to increase their customer base. In this process, profit margins are reducing day by day especially in the segment of vehicle insurance. Customers are expecting personalized insurance plans specific to their needs and insurance companies also want to charge high premium for the customers with risky driving behaviour and less for safe driving. Insurance companies are trying to reduce costs by leveraging historical risk data and advanced analytics. With the help of emerging technologies like IoT, companies are capturing real time vehicle movement data to monitor driving behaviour of their customers. By applying advanced analytics techniques on this data, insurance companies can derive better insights of their customer's driving patterns to assess the risk involved.

The objective of the study is to provide an analytical approach to classify driving patterns which will lead to risk based insurance premiums. Capturing of the driving behaviour also helps insurance companies to quickly determine the fault when accidents took place which is a key factor in claims settlement.

Data comprises of vehicle movement information captured through multiple sensors connected to the cars. With the help of advanced machine learning techniques driving patterns are categorized as Risky, Potential Risky and Safe zones. By quantifying the percentage of time vehicle was driven in Risky, Potential Risky and Safe zone, insurance companies can fix the premium based on the risk assessed with each category.

Hybrid Machine Learning and Cognitive approach using Telematics and IoT for tackling Downtime of Commercial Vehicles and Taxis

Chiranjiv Roy, Lakshminarayanan S

Mercedes Benz R&D India

Defect Diagnostics and research has been key topic for many researchers for decades now. Exponentially increasing constant research in this field has evolved to develop devices which can transmit performance of each component of a vehicle through logs. Pattern recognition and Machine Learning is a famous field of application to find failure symptoms in this field. These techniques are quite powerful in learning when plethora of historical information is readily available to mine, however with new generation of devices using Internet of Things (IoT) to communicate in real time lacks

much history and heavily impacts learning. This research proposes to understand the complex and mammoth Telematics data through a hybrid mixture of subject models and then in-stream classify them by proximity towards learned features and validate them through human intervention of engineers. We use an ensemble of best-in-class tools and techniques like NLP, Machine Learning, Streaming Big Data prediction on Telematics and IoT generated data for Commercial Vehicles. The Service Engineers are provided with chat-bots powered by the machine learning models to quickly becoming a suite of application for Human Machine Interaction.

IC 17/111

Market Efficiency and Inflationary Impact of Agricultural Futures Trading in India

Archana Singh

DSM, Delhi Technology University, Delhi

Narinder Pal Singh

Jagan Institute of Management Studies, Rohini, Delhi

In an efficient futures market, the producers, traders and consumers can hedge their price risk. But, increased trading in commodities may also result in higher inflation. Also, there have been allegations on commodity futures trading that it has resulted in higher inflation especially in agricultural commodities. Thus in this study, we analyse the market efficiency of agricultural futures market and the effect of futures trading on inflation with special reference to chana (chickpea) market in India. This study is for a time frame of ten years from 2005-2014. The data on closing prices of chana in futures and spot markets, and futures trading volume has been collected from National Commodity and Derivatives Exchange, and WPI monthly data from Office of the Economic Advisor, Government of India. The collected data is analysed for efficiency using Johansen cointegration approach and inflationary effect using Toda Yamamoto (TY) version of Granger causality test. From the results, we find that the spot and futures prices for chana are cointegrated and unbiased i.e. the chana (chickpea) futures market is efficient. But, the futures trading in chana has inflationary impact. This research has got direct implications for government and market participants.

IC 17/112

Inverse Problem for Dynamic Computer Simulators

Pritam Ranjan,

Indian Institute of Management Indore, MP, India

The recent accelerated growth in the computing power has generated popularization of experimentation with computer models (or simulators) in various physical and engineering applications. Realistic computer models of complex phenomena can be very expensive and statistical surrogates are used for detailed investigation. Inverse problem refers to finding a set of input parameters of a computer model that generates a pre-specified output. For computationally expensive simulators, it is a challenging problem, and has been thoroughly investigated for a scalar-valued simulator. In

our context, however, we consider a computer simulator that returns time-series whenever it is run with an input setting. In this talk, I will present several innovative techniques for solving this problem. I will illustrate the proposed methodology using several test functions and real-life applications.

IC 17/ 113

A Novel Approach for Cost-Sensitive Attribute Selection on Medical Datasets

Kaushar Ghanchi

Navgujarat College of Computer Applications

Nileshkumar Modi

Baba Saheb Ambedakar University, Ahmedabad
Gujarat India

Ratvinder Singh Grewal

Laurentian University, Greater Sudbury, Ontario,
Canada.

Health Analytics is significant for improved patient care and outcomes. Health records need important analysis to obtain vital information. There are many ways to gain knowledge from health records. All the attributes of the health records may or may not be useful in all types of knowledge extraction. So it is required to identify the most required and meaningful attribute from the dataset. Attribute selection methods are used to identify the best relevant feature from the dataset. Out of all attribute selection methods only cost sensitive learning considers different cost while analysing these data. Health managers and clinicians require some model to minimise several cost associated with health care. We can consider these several types of cost in cost sensitive feature selection and classification process. Here we present a novel approach for the cost sensitive feature selection of most important and meaningful feature from the hepatitis dataset of UCI repository. We consider misclassification cost, test cost and risk cost. The main purpose of the research is to minimise the total average cost of each feature. We found that risk cost of each feature plays vital role in the selection of the concern feature in further classification process

IC 17/114

Employee Branding: A Case Study

D. K. Ghosh

Department of Statistics, Saurashtra University Rajkot, Gujarat.

In this investigation our main effort is to study how the employee branding function satisfactorily in various Indian companies. For an example for various daily required items like clothes, shoes, electrical goods, mobiles etc. employee also needs the branding name of the company. In facts the company needs to maintain the employee brand and due to this the degree of satisfactory level of the employee is maintained. Moreover the effectiveness of the organization can be improved with such practices. Considering this facts we collected data of employee branding from ten companies

through questionnaires. These data were further analyzed using Discriminant analysis for company and employee branding impacting the organizational success through SPSS software. Through this analysis in this investigation, we observed that if the companies are able to maintain employee branding then they will definitely earn more profit through the most important factors. That is, satisfaction of employee depends upon employee branding which is further converted in to consumer's satisfaction.

IC 17/ 115

Incremental Learning for Location Prediction on Taxi-GPS Data Streams

Arnab Kumar Laha, Sayan Putatunda
Indian Institute of Management Ahmedabad

The prediction of the destination location at the time of pickup is an important problem with potential for substantial impact on the efficiency of a GPS enabled taxi service. While this problem has been explored earlier in the batch data set-up, but in this paper we propose new solutions in the streaming data set-up. The analysis of data streams offers a great opportunity for development of new methodologies and applications in the area of Intelligent Transportation Systems. In this paper, we examine four incremental learning methods using a damped window model namely, Multivariate multiple regression, Spherical-spherical regression, Randomized spherical K-NN regression and an Ensemble of these methods for their effectiveness in solving the destination prediction problem. The performances of these methods on several large datasets are evaluated using suitably chosen metrics and they were also compared with some other existing methods. The Multivariate multiple regression method and the Ensemble of the three methods are found to be the two best performers. The next pickup location problem is also considered and the aforementioned methods are examined for their suitability using real world datasets. As in the case of destination prediction problem, here also we find that the Multivariate multiple regression method and the Ensemble of the three methods give better performance than the rest.

IC 17/116

Robust Estimators on the Circle

Mahesh K C
Institute of Management, Nirma University

In many natural and physical sciences such as biology, geology, ecology, geography, earth science, medicine etc. measurements are directions. These directions are conveniently represented as angles on the circle. Most of the statistical analysis is based on a number of explicit or implicit assumptions and in reality it may often happens one or more of these assumptions fail to hold. One common phenomenon in analyzing real-life data sets is the presence of one or few outliers-observations

which are very different from the rest. The presence of such observations completely distorts the statistical analysis and inference made will not be reliable. An alternative way to deal with such problems is the use of robust procedures. In this paper we study some robust estimators of parameters of some well-known distributions on the circle.

IC 17/117

Agricultural Marketing Channel Choice and Its Determinants – A Farmers’ Perspective

Aashish Argade, Arnab K. Laha

Indian Institute of Management, Ahmedabad

Agricultural marketing reforms in India were initiated in 2003 with the Central government proposing the Model APMC Act to provide guidelines for the states to amend their respective APMC Acts. Some of the important issues that were given prominence in the amendments were contract farming, direct marketing, private markets and electronic trading. This study aims to understand how agricultural marketing takes place after the implementation of reforms. The first part of the study attempts to know the different marketing channels available to farmers and how the farmer makes a choice out of the available options. APMC and farm gate sale emerged as the most used options and hence binary logistic regression was employed to identify the significant predictors that impacted farmers’ choice. The type of crop (horticultural), facilities available at the APMC and credit offered by the buyer were found to be the significant predictors for APMC as the channel choice. The second part of the study applied conjoint analysis by providing different parameters of a market place and options for each parameter. Farmers were asked to rank four choice sets. The responses were analyzed using Kemeny rule or Kemeny-Young method to select the median choice. Farmers preferred a market place within the village, with warehousing facility, followed by APMC with all facilities including warehouse receipt financing, scientific assaying and online sale and payment. The study has important policy implications for furthering agricultural marketing reforms.

IC 17/ 118

New Approaches to Prediction using Functional Data Analysis

Poonam Rathi, Arnab Kumar Laha

Indian Institute of Management, Ahmedabad

In this paper we address the problem of prediction with functional data. We discuss several new methods for predicting the future values of a partially observed curve when it can be assumed that the data is coming from an underlying Gaussian Process. When the underlying process can be assumed to be stationary with powered exponential covariance function we suggest two new predictors and compare their performance. In some real life situations the data may come from a mixture of two

stationary Gaussian Processes, there we introduce two new methods of prediction and compare their performance. In case the data comes from a non-stationary process we propose a modification of the powered exponential covariance function and study the performance of the predictors mentioned above using three real-life data sets. The results indicate that the KM-Predictor in which the training data is clustered using the K-Means algorithm before prediction can be used in several real life situations.

IC 17/119

Response Adaptive Allocation with ordinal categorical responses

Rahul Bhattacharya and Soumyadeep Das
Calcutta University.

Atanu Biswas
Indian Statistical Institute, Kolkata

A response adaptive allocation is developed for two treatment clinical trials when the responses to treatments are ordinal categorical in nature. Several exact and asymptotic properties of the procedure are studied and acceptability is further investigated through a real data set with categorical responses.

IC17/120

OLS: Is That So Useless for Regression with Categorical Data?

Soumyadeep Das
University of Calcutta.

Atanu Biswas, Samarjit Das
Indian Statistical Institute, Kolkata.

Binary/categorical response data abound in many application areas and poses a unique problem; OLS based model may lead to negative estimate for probability of a particular category; and does not provide coherent forecast for the response variable. This unique and undesirable property of linear regression with categorical data impedes the use of OLS which otherwise is the simplest and distributionally robust method. The logit or probit kind of solution is heavily distribution dependent or link-function dependent. Failure of such distributional assumption of the underlying latent variable model may cost the estimators heavily; may lead to biased and inconsistent estimates, in general. In this paper, we attempt to fix the inherent problem of linear regression by suggesting a simple manipulation; which, in turn, leads to consistent estimates of probability of a category; and results in coherent forecasts for the response variable. We show that the proposed solution provides comparable estimates; and sometimes, with respect to some criterion, the proposed method is even slightly better than the logit kind of models. Here we consider different underlying error distributions and compare the performances of the two models (in terms of their respective residual sum of squares and also in term of mutual information) based on simulated data. It is evidenced that the OLS performs better for many distributions viz., Gamma, Laplace and Uniform error distributions.

Robustness issues in circular-circular regression

Jayant Jha, Atanu Biswas

Indian Statistical Institute, Kolkatta

Tsung-Chi Cheng

National Chengchi University, Taipei, Taiwan.

In this paper, we attempt to address the robustness issues in circular-circular regression. We consider the Mobius transformation based circular-circular regression model of Kato et al. (2008). Then, we discuss the robustness issue of the estimators in this model. We propose maximum trimmed cosine estimator in this context and discuss the algorithm for its computation. We also discuss some properties including the breakdown point of the estimators. Simulation studies and a real data analysis are used to illustrate the proposed methodology.

Assessing Density Forecast for Energy Commodities: A Post-Financialization Study

Deepak Bisht, Arnab K. Laha

Indian Institute of Management Ahmedabad

A density forecast of a random variable is an estimation of the probability distribution of the possible future values of that variable. Futures and options together provide such a density forecast for the underlying asset which is often referred as Risk Neutral Density (RND). Derivative pricing literature suggests that RND differs from real world density (RWD) due to the risk aversion of market participants. As risk premia (if exists) contaminates RND forecasts, care must be taken while making risk management decisions. Most of the literature is focused on detecting the risk premium in commodity futures, i.e., locational bias. Results based on pre-financialization futures data were able to detect risk premium in energy commodities in accordance to the Keynesian-Hicks hypothesis. Unlike pre-financialization era where producer hedging was dominant, now long-only commodity index funds dominate the commodity futures markets. In recent times, there is a substantial increase in the liquidity of commodity options. These developments inspired us to investigate the nature of risk premia in the current setting. In this article, we evaluate the reliability of RND forecasts for two most widely traded energy commodities, WTI Crude Oil and Natural Gas. Our data set consists of 8 years of monthly data on NYMEX traded futures and OTM options. We find that implied RND forecasts of one-month, two-month, and three-month don't show any statistically significant systematic risk premia. With no evidence of systematic bias in location and dispersion, we feel RND implied by short dated derivatives can be used for risk management and regulatory reporting purposes.

Market Irrationality and Share Repurchases

Joshy Jacob, Suman Saurabh

Indian Institute of Management, Ahmedabad

There is significant variation in the number of repurchase announcements and the targeted amount over time. It could be attributed to several factors, such as the varying tax advantage and changing stock valuations. It is well documented that the tax-attractiveness trigger more repurchase announcements. More repurchases are announced in markets characterized by low stock valuation. It is widely known that irrationality influences stock valuation and firms often attempt to favorably exploit the market irrationality. We examine the possible influence of market irrationality on repurchase announcement outcomes and motives. The paper investigates how repurchases announced in markets characterized by high-irrational sentiment vary from those announced in low-sentiment markets. Our findings indicate that while announcements in low-sentiment phase earn lower abnormal returns in the short-run, in the long-run, they outperform announcers in the high-sentiment phase. Also, the long-run operating performance of announcers in the low-sentiment markets is superior. The long-run stock and operating performance suggest that announcers in the low-sentiment markets are more likely to be motivated by the need for signaling. We further examine how announcers across the two phases differ in their motives. We find that announcers in the low-sentiment markets are larger, more profitable and liquid, and have higher leverage relative to announcers in high-sentiment markets. A logit regression indicates that announcers in the low-sentiment market are more likely to signal undervaluation, distribute excess cash, and increase leverage through repurchases. Announcers in the high-sentiment phase, on the other hand, are more likely to be motivated by the price-support and anti-dilution role of repurchases. Overall, our results suggest that the market irrationality influences both the firms' decision to repurchase and the market reaction to the announcement.

Robustness of circular mean functional for families of asymmetric circular distribution

Arnab Kumar Laha, Pravida Raja A.C.

Indian Institute of Management Ahmedabad

Mahesh K.C.

Institute of Management, Nirma University, Ahmedabad.

Most of the tractable distributions currently available for modeling circular data are symmetric around a modal direction. Only in recent years, some asymmetric extensions of well-known circular models have been proposed in the literature. One such extension is given by Kato and Jones (2010), who proposed a three parameter asymmetric family for modeling an asymmetrically distributed datasets on the circle. Robustness and SB-robustness of estimators for the asymmetric families of distributions on the circle has not been studied much in the literature. In this paper we derive the influence function of an asymmetric unimodal circular distribution and study the SB-robustness of the Circular

Mean Functional (CMF) for asymmetric families of distribution. It is shown that CMF is SB-robust at a sub-family of three parameter asymmetric Kato-Jones distributions.

IC 17/ 125

Idiosyncratic return characteristics and cross-section of expected stock returns

Nilesh Kumar Gupta

Indian Institute of Management Tiruchirappalli

Joshy Jacob

Indian Institute of Management Ahmedabad

The cross-section of expected stock returns is unaffected by idiosyncratic skewness when investors are fully diversified. However, in a world with un-diversified investors the idiosyncratic characteristics of stock returns could be priced in the market. Recent research has brought out that investors with lottery preferences may prefer to concentrate their portfolios in stocks with significant positive skewness. Investment stocks with high idiosyncratic skewness and volatility could be driven either by the exaggeration of small probabilities or by overconfidence. We investigate the extent to which proxies of idiosyncratic skewness influences cross-sectional expected returns in India. Given the high volatility and information asymmetry prevalent in the emerging markets, it is possible the speculative features of stock returns might have a greater explanatory role in the cross-sectional returns in emerging markets.

We sort all the stocks listed at the Bombay Stock Exchange in India into portfolios using a proxy for idiosyncratic skewness, measured as the maximum single day returns earned by a stock during a certain period in the past. We examine the performance of the portfolios over a future period. The results primarily indicate that stock portfolios with greater idiosyncratic skewness earn relatively lower alphas compared to their less skewed counterparts, even after controlling for many widely used cross-sectional risk-factors, such as size, value, and momentum. We further examine how the overall market irrationality captured by a sentiment index influences the speculative trading in the market. The results suggest that the speculative idiosyncratic features of past returns significantly account for cross-sectional returns in the Indian market.

IC 17/126

Forefather distribution in Galton Watson Processes with age dependant structure in population

Arnab Kumar Laha , Sumit Kumar Yadav

Indian Institute of Management Ahmedabad

In this paper we examine the structure of a variant of the Galton-Watson branching process with particular reference to the distribution of the number of forefathers of the individuals in the current generation. Starting with a simplified model with the offspring distribution being Poisson but allowing the number of time periods that an individual can survive to vary within a limited range we find the age distribution of the number of forefathers using extensive simulation. For the cases, where the number

of time periods an individual survives is equal to 2, exact expression for expected number of individuals having “k” forefathers has been derived. A similar simulation analysis is carried out assuming the offspring distribution to be binomial and also negative binomial. Some interesting insights and possible applications are discussed.



Announcement

Selected papers of ICADABAI-2017 will be published by Springer as an edited volume after peer review.



ICADABAI HACKATHON-2017

in collaboration with





“Gold sponsor of ICADABAI-2017”

**iNautix Technologies India Pvt. Ltd,
a BNY Mellon Company**