Indian Institute of Management Ahmedabad is happy to organize the 6th 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 ICABADAI-2015 was held on 11-12 April 2015. The two-day conference had four key-note speeches, two panel discussions, workshops on 4 different topics and more than 80 paper presentations.

The fifth conference in this series ICADABAI-2017 was held on 8-9, April 2017. The two-day conference had eight key note speeches and two workshops delivered by illustrious academicians and practitioners. This conference saw more than 60 paper presentations. Selected papers of ICADABAI-2017 were published as an edited book volume:


Papers for ICADABAI-2019 were selected for regular presentation and 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-2019

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

Pravida Raja A.C. (Conference Manager)
Pina Dudhnath Kori (Conference Coordinator)
Indian Institute of Management Ahmedabad.

Conference Team
Pina Dudhnath Kori
Pratishtha Batra
Pravida Raja A.C.
Rohit Bhardwaj
Shylaja Deepak

Vaidyanathan Ramachandran APAC (India), Kaltura Inc
(ICADABAI-HACKATHON 2019 Convener)
International Programme Committee (IPC) members of ICADABAI-2019

1. Abhishek – Institute of Management Technology, Ghaziabad
4. Banerjee Gautam – Business Brio
5. Basu Srabashi – Great Learning
7. Das Kiranmoy – Indian Statistical Institute
8. Das Paulami – Principal, Data Scientist, Brillio
9. Dutta Somak – Iowa State University
10. Dutta Subhajit – Indian Institute of Technology Kanpur
11. Ghosh Ashish – Indian Statistical Institute
12. Ghosh Sujit – North Carolina State University
15. Banerjee Buddhananda – Indian Institute of Technology Kharagpur
16. Pant Siddharth – SG Analytics
17. Pal Debdatta – Indian Institute of Management Lucknow
18. Pal Sandip – IBM
19. Paul Pritam Kanti- Bridgei2i Analytics Solutions
22. Sen Sayan - HDFC Life Data Labs
23. Singh Sanjeet- Indian Institute of Management, Calcutta
24. Singh Sudhanshu – Genpact
25. Ranjan Pritam – Indian Institute of Management Indore
26. Roy Vivekananda – Iowa State University
27. R P Suresh – Accenture
Session Organizers of ICADABAI-2019

R P Suresh
Accenture

Vivekananda Roy
IOWA State University

Workshop speakers of ICADABAI-2019

Gayatri Pal
WalmartLabs, Bangalore

Murari Mitra
IIEST Shibpur

Sponsor of ICADABAI-HACKATHON 2019

IBM®
We thank the following reviewers of ICADABAI 2019 for helping us in selecting the papers for presentation

- Abhishek - IMT Ghaziabad
- Aruna Divya – IIM Ahmedabad
- Aashish Argade – IRMA Gujarat
- Aditya Christopher Moses - IIM Ahmedabad
- Budhananda Banerjee - IIT Kharagpur
- Debdatta Pal - IIM Lucknow
- D. Karthik - IIM Trichy
- Goutam Dutta - IIM Ahmedabad
- Harshal Lowalekar - IIM Indore
- Joy Chowdhury – Goa Institute of Management
- Kanish Debnath - FLAME University, Pune
- Kousik Guhathakurta - IIM Indore
- Krishnamoorthy Srikumar - IIM Ahmedabad
- Mahesh K.C - Institute of Management, Nirma University, Ahmedabad
- Manish Aggarwal - IIM Ahmedabad
- Naman Desai - IIM Ahmedabad
- Premilla D’ Cruz - IIM Ahmedabad
- Pritam Ranjan - IIM Indore
- Rudra Prakash Pradhan - IIT Kharagpur
- Sandip Chakrabarti –IIM Ahmedabad
- Santosh Kr. Prusty - IIM Shillong
- Sanjeev Tripathi - IIM Indore
- Saral Mukherjee – IIM Ahmedabad
- Sayan Putatunda - VMWARE Software India Pvt.Ltd.
- Sharad Nath Bhattacharya –IIM Shillong
- Suman Mitra – DAIICT, Gandhinagar
- Vijaya Sherry Chand - IIM Ahmedabad
- Rama Mohana Turaga –IIM Ahmedabad
- Sobhesh Aggarwalla –IIM Ahmedabad
- Ankur Sinha –IIM Ahmedabad
Contents

IC19/KN 1  Man, Machine and Métier : Trends in adoption of Artificial Intelligence for Business. 1
IC 19/IN1  Global-fitting of the response surface of an expensive computer simulator .......... 1
IC 19/IN2  A new perspective from a Dirichlet model for forecasting outstanding liabilities of non-life insurers ................................................................. 2
IC 19/IN3  Probability Models and Statistical Inference for High Volatility Big Data .......... 2
IC 19/IN4  Big Data, Bigger Aspiration, Too Little Expertise .................................................. 3
IC 19/IN5  Analysis of Credit Rating Model for SME Borrowers .............................................. 3
IC 19/IN6  From Pillar to Posts-Conceiving and Delivering Trusted AI .................................. 4
IC 19/IN7  Convolutional neural networks based image recognition ....................................... 4
IC 19/WS 1  Using R for Production .......................................................................................... 5
IC 19/WS 2  Reliability Shock Models and Applications .......................................................... 5
IC 19/001  Factors Affecting Work And Family Life Imbalance Among Women Gynaecologists In Government And Private Hospitals in Kerala ........................................... 6
IC 19/002  Descriptive Analytics during Disruptive Periods of Investments – A case study on Gold & Gold ETFs .................................................................................. 6
IC 19/003  A Geometric Approach to Estimate the Mean Shift in Change-Point Analysis ......... 7
IC 19/006  Artificial Intelligence based cost reduction for Customer Retention Management in the Indian Life Insurance Industry ..................................................... 8
IC 19/007  Implementing Learning Analytic Tools in Predicting Students’ Performance in a Business School ............................................................................................. 8
IC 19/008  Solving the CECL Riddle through Risk Analytics ..................................................... 9
IC 19/010  Comparative Study of Different Benchmarking Techniques ................................. 10
IC 19/011  Optimization of Initial Credit Limit Using Comprehensive Customer Features .......... 11
IC 19/012  Improved Forecasting Techniques for New Product Introduction using Ensemble Models ........................................................................................................... 12
IC 19/013  Mitigating Agricultural Lending Risk: An Advanced Analytical Approach ............... 12
IC 19/015  Supply Chain Contract for Short Life-Cycle Products with Production Capacity Constraint .................................................................................................................. 13
IC 19/018  Applying Analytics Techniques to effectively understand match phases in major European soccer tournaments (2013/14-2017/18 seasons) ........................................... 13
IC 19/019  Unraveling Fake News “Segmentation & Classification based Detection Approach”. 14
IC 19/021  Vehicle Routing Case Study .................................................................................... 14
IC 19/022  Food Index Forecasting ............................................................................................ 15
IC 19/023  Cognitive Customer Value Management: A systematic approach for customer churn analytics in a non-contractual context ......................................................... 15
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayesian Optimization for Multiple Boosting</td>
<td>16</td>
</tr>
<tr>
<td>Explainable Artificial Intelligence Model: Analysis of Neural Network Parameters</td>
<td>17</td>
</tr>
<tr>
<td>Assessing Service Quality using Sentiment Analysis – Evidence from Wellness Resorts in India</td>
<td>17</td>
</tr>
<tr>
<td>Credit Scoring Using Machine Learning Algorithms</td>
<td>18</td>
</tr>
<tr>
<td>An ARDL &amp; Cointegration Approach for Analyzing Determinants of Foreign Portfolio Investors' in India</td>
<td>19</td>
</tr>
<tr>
<td>Public Sector Banks Lending for Residential and Commercial Real Estate in India</td>
<td>19</td>
</tr>
<tr>
<td>India-Bimstec Bilateral Trade Activities: A Gravity Model Approach</td>
<td>20</td>
</tr>
<tr>
<td>Performance Analysis of Optical Character Recognition Tools in Natural Scene Text Detection using OCR Post-Processing</td>
<td>20</td>
</tr>
<tr>
<td>Project Delays and Business Analytics – Innovation, Risk Mitigation and Emerging Horizons</td>
<td>21</td>
</tr>
<tr>
<td>Real Earnings Management Practices for Meeting Earnings Benchmarks: Indian Evidence</td>
<td>21</td>
</tr>
<tr>
<td>Adoption and Implementation of HR Analytics in Damodar Valley Corporation (DVC)</td>
<td>22</td>
</tr>
<tr>
<td>Managing Class Imbalance in Big Data - A Distributed Approach to Synthetic Minority Oversampling Technique</td>
<td>23</td>
</tr>
<tr>
<td>Integrating TODIM into Data Mining for the Evaluation of Customer Lifetime Value</td>
<td>24</td>
</tr>
<tr>
<td>Privacy and Social Media: Do Users Really Care?</td>
<td>24</td>
</tr>
<tr>
<td>The Effect of Sustainability on Foreign Direct Investment Inflows: Evidence from developing countries</td>
<td>25</td>
</tr>
<tr>
<td>Analysing Regression and Decision Tree Models for Credit Risk</td>
<td>26</td>
</tr>
<tr>
<td>Stacking With Dynamic Weights on Base Models</td>
<td>26</td>
</tr>
<tr>
<td>Improving Blast Furnace Operations through Advanced Analytics</td>
<td>27</td>
</tr>
<tr>
<td>Decriminalisation of Algorithmic Trading: A Perspective From Indian Stock Market</td>
<td>27</td>
</tr>
<tr>
<td>Banks to become future Technology Companies</td>
<td>28</td>
</tr>
<tr>
<td>Stock Selection Based on Cluster Analysis: Supplementing The Stochastic Models For Decision Making</td>
<td>29</td>
</tr>
<tr>
<td>Machine Learning Approach for Formulating and Improvising Existing Local and Federal Government Policy</td>
<td>29</td>
</tr>
<tr>
<td>Side Effects of Communication - Intermediating Effect of Feedback Loop on NPS® Score</td>
<td>30</td>
</tr>
<tr>
<td>Time series analysis of the US Airline Industry: What financial statements do not reveal</td>
<td>31</td>
</tr>
<tr>
<td>Response Prediction and Ranking Models for Large Scale Ecommerce Search</td>
<td>31</td>
</tr>
<tr>
<td>IC 19/073</td>
<td>Knowledge discovery based decision support system for application of systems thinking concepts in maternal and child health system strengthening</td>
</tr>
<tr>
<td>IC 19/075</td>
<td>Characterization of the Skew Raised-Cosine Distribution</td>
</tr>
<tr>
<td>IC 19/076</td>
<td>Application of Association Rule Mining in a Clothing Retail Store</td>
</tr>
<tr>
<td>IC 19/077</td>
<td>Implementing and democratizing personalized campaigns at scale and speed</td>
</tr>
<tr>
<td>IC 19/078</td>
<td>Style Scanner- Personalized Visual Search &amp; Recommendations</td>
</tr>
<tr>
<td>IC 19/082</td>
<td>Frequency-dependent connectedness of geopolitical risk across BRICS nations</td>
</tr>
<tr>
<td>IC 19/086</td>
<td>Connectedness of Markets with Heterogeneous Agents and the Information Cascades</td>
</tr>
<tr>
<td>IC 19/089</td>
<td>Dynamic Regression Approach to Identifying the Determinants of Inward Foreign Direct Investment for India</td>
</tr>
<tr>
<td>IC 19/092</td>
<td>Best approach direction for spherical random variables</td>
</tr>
<tr>
<td>IC 19/093</td>
<td>Role of Infrastructure and Taxation on Economic Growth in Middle Income Countries</td>
</tr>
<tr>
<td>IC 19/095</td>
<td>Drivers of Governance: An Econometric Analysis</td>
</tr>
<tr>
<td>IC 19/096</td>
<td>Stochastic comparisons of systems with heterogeneous log-logistic components</td>
</tr>
<tr>
<td>IC 19/097</td>
<td>Care2Vec: A Deep learning approach for the classification of self-care problems in physically disabled children</td>
</tr>
<tr>
<td>IC 19/098</td>
<td>Predictive likelihood for coherent forecasting of count time series</td>
</tr>
<tr>
<td>IC 19/099</td>
<td>On Modeling the High-Frequency Time Series Data</td>
</tr>
<tr>
<td>IC 19/100</td>
<td>Coalescence in Branching Processes with age dependent structure in population</td>
</tr>
<tr>
<td>IC 19/101</td>
<td>Gender equality at work: Are we there yet?</td>
</tr>
<tr>
<td>IC 19/103</td>
<td>Predicting Failures Beyond Warranty in Automotive industry</td>
</tr>
<tr>
<td>IC 19/104</td>
<td>Machine learning in Retail</td>
</tr>
<tr>
<td>IC 19/105</td>
<td>How Machines Embed Business Intelligence</td>
</tr>
<tr>
<td>IC 19/106</td>
<td>Why prices of medical procedures vary across providers</td>
</tr>
<tr>
<td>IC 19/108</td>
<td>Land transitions and climate change impacts on urban flooding in 42 cities across India</td>
</tr>
<tr>
<td>IC 19/109</td>
<td>A Perfect Mixture Model for Credit Risk and Actuarial Applications</td>
</tr>
<tr>
<td>IC 19/111</td>
<td>SB-robustness of Estimators of Directional Correlation Coefficient</td>
</tr>
<tr>
<td>IC 19/112</td>
<td>A Robust Test for the Spherical Mean Direction</td>
</tr>
<tr>
<td>IC 19/113</td>
<td>Impact of Artificial Intelligence on Indian IT / BPM Employment</td>
</tr>
<tr>
<td>IC 19/114</td>
<td>Violence, Contagion and Informal Sector</td>
</tr>
</tbody>
</table>
Man, Machine and Métier: Trends in adoption of Artificial Intelligence for Business

Kingshuk Banerjee
Cognitive Computing and Analytics
IBM Global Business Services.

This discourse deals in complex interplay of people, process and technology for developing next-generation business systems. The focus is on Analytics, Artificial Intelligence (AI) and Automation. The scope of observation on emerging trends of interaction between man and machine is industry-agnostic, and specific to enterprise domains like Marketing, Procurement, Finance and Human Resource Management. This treatment details the disruptive trends like Conversational Systems, Digital workforce and Cognitive Digitization that are currently transforming global enterprises. It touches upon the underlying implements - the tools and technology - for laying the foundation of AI systems for smart business. It also draws upon the outcomes of operational systems for feedback learning on the efficiency and efficacy of AI-based business.

Global-fitting of the response surface of an expensive computer simulator

Pritam Ranjan
Indian Institute of Management Indore

Feng Yang and C. Devon Lin
Queens University, Canada

In many application areas such as aerospace, renewable energy, cosmology, climate modelling, and manufacturing, computer simulators are widely used to mimic complex physical systems that are either too expensive or infeasible to observe. Experimentation with such computer simulators (or computer models) is a substitute for the underlying physical experiment. One fundamental issue in these computer experiments is to carefully select the input settings where the computer simulator is run, particularly, when the simulator is expensive to run. This is very common, as realistic simulators of complex phenomena can take from seconds to hours for even one run.

There are several popular objectives of computer experiments, for instance, estimation of global optimum, pre-specified threshold, upper/lower quantile region of the response surface, the overall response surface, identification of important input variables, uncertainty quantification, or the calibration of the computer model itself. In this talk, I will focus on the design issue (i.e., how to carefully select the set of input points) for the goal of producing more accurate prediction which is important for risk assessment and decision making.
Ranjan, Bingham and Michailidis (2008, Technometrics, 50, 527-541) proposed an expected improvement criterion under a sequential design framework for estimating a pre-specified contour of the simulator output surface. In this paper, we generalize the expected improvement criteria for contour estimation, and propose novel usage of simultaneous and sequential contour estimations for global fitting of the response surface. Several numerical examples are used to demonstrate the effectiveness of the proposed approaches as compared to the state-of-the-art methods like one-shot approach via Latin hypercube design with space-filling property, IMSE optimal designs, and sequential techniques such as D-optimal designs.

IC 19/IN2

A new perspective from a Dirichlet model for forecasting outstanding liabilities of non-life insurers

Karthik Sriram
Indian Institute of Management Ahmedabad

Forecasting the outstanding claim liabilities to set adequate reserves is critical for a nonlife insurer’s solvency. Chain-Ladder and Bornhuetter-Ferguson are two prominent actuarial approaches used for this task. The selection between the two approaches is often ad hoc due to different underlying assumptions. We introduce a Dirichlet model that provides a common statistical framework for the two approaches, with some appealing properties. Depending on the type of information available, the model inference naturally leads to either Chain-Ladder or Bornhuetter-Ferguson prediction. Using claims data on Worker’s compensation insurance from several US insurers, we discuss both frequentist and Bayesian inference.

IC 19/IN3

Probability Models and Statistical Inference for High Volatility Big Data

Ashis SenGupta
Indian Statistical Institute, India; Augusta University, USA and Middle East Technical University, Turkey

Over the last few years, strong asymmetry and high volatility have inevitably characterized modern financial Big data. However, probability distributions exhibiting both these properties seem not to have been used in their full flexibilities for modelling such data. The non-existence of tractable analytic forms of the density functions for appreciably long-tailed, skewed distributions as required for emerging high-frequency Big data render these non-amenable to the commonly used likelihood-based methods for statistical inference. In this work, several flexible distributions are introduced. Motivated by the resulting elegant forms of the related characteristic function that are derived, and the domain of attraction provided by the Khinchin metric, we enhance our approach for obtaining strongly consistent estimators of the
parameters. A novel numerical procedure is employed. An algorithm for simulating high volatile random variates is presented and used for studying the finite sample performances of the proposed estimators. A real-life financial dataset from Blockchain / Bitcoin is analyzed to illustrate our proposed methodology.

**IC 19/IN4**

**Big Data, Bigger Aspiration, Too Little Expertise**

**Atanu Biswas**  
Indian Statistical Institute, Kolkata

We discuss how data is becoming bigger day by day, and how we aspire to use the loads of data in politics, sports, healthcare, national policy making, and everything. But, is our current expertise adequate for this purpose? An assessment is attempted.

**IC 19/IN5**

**Analysis of Credit Rating Model for SME Borrowers**

**D. K. Ghosh**  
UGC BSR Faculty Fellow, Saurashtra University Rajkot

Credit scoring model include both quantitative & qualitative aspect. As it accounts for both aspect to reach to a score which gives a holistic view about the creditworthiness of the borrower. Quantitative aspect is taken care of as discussed above, including some more aspects & properly defining the cut-off for each parameter & sub parameter.

Qualitative variables has been accounted by defining dummy variable & giving them a standard definition in order to maintain consistency in the subjective parameters of the model across different users.

Parameters were decided keeping in view some essential aspects which are must for the existence, sustenance & growth of the borrowers business. Observations based on past performance of the accounts (under defined sample) over a period of time have been considered to take a view on qualitative parameters.

In addition to using Z score as deduced above from which turn around time (TAT) of taking decision for approving loans can be significantly reduced, it can be also used to compare two borrowers based on their credit rating or fine tuning the credit rating on a scale of one to ten. For this more parameters are required for analysis. Others parameters of Quantitative and Qualitative nature has been formulated based on the past experience and Z-score deducted in this study. The financial Risk has been calculated through Z score and other Risk namely, Business Risk, Industry Risk, management Risk and Operational Risk are deducted through the points are carried out here which has been described from past experience of the bankers.
From Pillar to Posts-Conceiving and Delivering Trusted AI

Kalapriya Kannan
IBM Research Labs, India

Talk will cover advances in Artificial intelligence in recent year - some of them are topics revisited due to the changing requirements and some of them are new due to the new requirements. The talk will spread across some of the recent works in Trusted AI, delivery of AI solution through a trusted platform and continuous improvement of the metrics using goal based active learning techniques.

Trusted AI has four pillars - Fair AI, Explainability, Robustness and Lineage. Fair AI attempts to study the impact of the `unwanted bias` in the data, models and propose mitigation algorithms. Explainability answers the question `Is it easy to understand` the outcome of AI models. Different stakeholders require explanations for different purposes and with different objectives, and explanations will have to be tailored to their needs. `Did anyone tamper with it?` is the question that Robustness tries to address and is about identifying and protecting the models against attacks. Finally the trust is enhanced when one can say `Is it accountable` for the outcomes provided by the AI Systems. I will very briefly go over all these areas in this topic.

While AI advances are focussing on increasing various aspects, another important aspect is the delivery of the AI solutions in a trusted platform. I will walk through some of the advances in AI marketplaces and how adding `blockchain` can add the spice to enhance trust -especially in cross organisation setting. We will also see how AI systems can increase reliability in a typically information asymmetric world of AI marketplaces.

Finally, continuous learning is an important aspect of the improvements and adaptability of the AI systems to new and untested data sets. I will walk through some of the advances that are moving in the direction of the continuous learning for AI systems.

Convolutional neural networks based image recognition

Ashish Ghosh
Indian Statistical Institute

Image recognition is very interesting and challenging field of study. While it is very easy for human and animal brains to recognize objects, computers have difficulty with the same task. Computers perceive images as mere pixel values and cannot capture the semantics of the image. Neural networks are, therefore, an immediate
choice to achieve human like recognition. The general applicability of neural networks is one of their advantages, but this advantage turns into a liability when dealing with images. The convolutional neural networks make a conscious trade-off. If we consider an image, proximity has a strong relation with similarity in it and convolutional neural networks specifically take advantage of this fact. In this lecture, we would delve deeper into the workings of a convolutional neural network and how it is used for image recognition tasks.

IC 19/WS1
Using R for Production
Gayatri Pal
Walmart Labs, India

Abstract: Objective of the workshop is to introduce participants using R in production. Recently, as more enterprises started using R for their analytic solutioning, there is an increasing need to understand if the same codebase can be used for the production. The workshop will guide the participants to run through a R deployment hands-on exercises, such as R Suite applications for large data. Additionally, the session will start with a brief introduction to large scale parallel computing like Sparkly R and use the same for operationalization in production.

IC 19/WS2
Reliability Shock Models and Applications
Murari Mitra
IIEST Shibpur

A unit may fail because of changes within it or due to changes in the environment. Failure often occurs as a result of accumulated shocks received over time. Various shock models have been studied extensively and find widespread applications in diverse areas of reliability theory and applied probability. Shocks arriving in the system can be interpreted as the claims in risk analysis, demand in inventory control, health hazards in biometry. One is interested in studying the ageing properties of the survival function for components or systems or biological entities under some information on the magnitude and occurrence time of the shocks. The life distribution functions $H(t)$ of a device subjected to a sequence of shocks occurring randomly in time according to a counting process is considered as a function of the probabilities $P^k$ of surviving the first k shocks. Various properties of the discrete failure distributions $P_k$ are shown to be reflected in corresponding properties of the continuous life distribution $H(t)$. The development of various natural ageing classes via shock model theory is explored.
Factors Affecting Work And Family Life Imbalance Among Women Gynaecologists In Government And Private Hospitals in Kerala

P.N. Harikumar
PG Department of Commerce & Tourism, Catholicate College, Pathanamthitta, Kerala
Susha D
Research & PG Department of Mathematics, Catholicate College, Pathanamthitta, Kerala

Doctors’ career discomfort, job stress and job burnout have recently received much attention, and considerable evidence suggests that dissatisfaction with medical practice is increasing. Physicians’ career dissatisfaction and burnout have been linked with high level of work-family imbalance. Working women gynaecologists are experiencing difficulty in balancing their work and family due to the excessive work pressure, too little time for themselves, and the need to fulfil others’ expectations of them. The majority of the lady gynaecologists experience job spill over into the home as they have to put in longer hours of work. Major consequences of poor work and family-life balance are high levels of stress and anxiety, disharmony at home, causing job burnout at hospitals and inability to realize full potential. They feel irritable and resentful often due to their inability to balance work and family life. Despite the concerns outlined above, the relationship between the physicians’ career dissatisfaction, work-life imbalance, and job burnout is under-explored; therefore, we set out to characterize these relationships. This study aims at identifying the factors affecting the work-family-life imbalance of the lady gynaecologists working in the government and private hospitals of Kerala. How such imbalances affect their job satisfaction as well as institutional, professional and family commitments is also examined in detail. The mental and physical exhaustion of the lady gynaecologists of private hospitals are very high when compared to the doctors in government hospitals, so that they face high level of job stress and job burnout. Their work-life imbalances will result in their family-life imbalance and consequently, the situation creates high levels of work-life conflict among them. Work-family-life conflict is very high among those who are living within the nuclear family structure and are also willing or ready to work more beyond normal working hours in hospitals. Work-family-life imbalances or conflicts affect job satisfaction, as well as the professional, institutional and family commitments of such doctors negatively.

Descriptive Analytics during Disruptive Periods of Investments – A case study on Gold & Gold ETFs

P.Baba Gnanakumar
School of Management, Kristu Jayanti College, Bangalore

Gold ETF, which has been introduced in 2007, in the Indian market as an alternative to investment in physical gold witnessed a heavy outflow of investment during
the period 2017-18; whereas, investment in Gold has increased. This research aims to find out the reason for this phenomenon and to create investment analytics between Gold and Gold ETF. We apply K-means of clustering for identifying the bullish/bearish trend in returns and ROC analysis to diagnose the goodness of predictability. The investment analytics is based on short-term gains during the sporadic trends. We found that the decrease in Gold ETF investments is due to less intra-day returns in Gold ETF as compared with Gold. We conclude that the returns from Gold ETF and physical Gold will have an equilibrium effect during the bullish period only. The bearish trend in Gold ETF may be hedged through Gold but not vice-versa. The reason for the negative effect has been portrayed in the ROC curve. During bearish trend, the mutual fund organisations of Gold ETFs are unable to market the product; whereas in case of physical Gold, investors are not having negative perception. However, during bullish trend, the investment in both physical Gold and Gold ETFs are yielding same returns. This research enables the mutual fund managers to decide the investment analytics among Gold ETFs.

IC 19/003

A Geometric Approach to Estimate the Mean Shift in Change-Point Analysis

Buddhananda Banerjee
Department of Mathematics
Indian Institute of Technology Kharagpur, India

Arnab Kumar Laha
Production and Quantitative Methods
Indian Institute of Management Ahmedabad, India

Estimation of the mean shift in a temporally ordered sequence of random variables with a possible existence of change-point is an important problem in many disciplines. In the available literature of more than fifty years the estimation methods of the mean shift is usually dealt as a two-step problem. A test for the existence of a change-point is followed by an estimation process of the mean shift, which is known as testimator. We notice that the existing formulation of the problem is suffering from over parametrisation. We undertake it as a purely estimation problem. We establish that the maximum likelihood estimator (MLE) always gives a false alarm indicating an existence of a change-point in the given sequence even though there is no change-point at all. After modelling the parameter space as a modified horn torus, the newly introduced estimation method of the mean shift is assessed with a proper Riemannian metric on that conic manifold and established its superiority over the performance of the MLE.
Artificial Intelligence based cost reduction for Customer Retention Management in the Indian Life Insurance Industry

Sanjay Thawakar  
Analytics Centre of Excellence, Max Life Insurance

Vibhu Srivastava  
Analytics Centre of Excellence

Customer retention, measured as percentage policies renewed every year (persistency ratio), is one of the most important metrics for any life insurer. Due to several factors, including complexity of life insurance products, gap in understanding the importance of policy renewals and lack of appropriate engagement with the customers, higher lapsation rates for life insurance policies have been observed globally, specifically in India. Typically, for a life insurance company, policy renewal premiums drive close to 60-70% revenue and retaining customers for a period of more than 6-7 years is critical to business profitability.

Customer Retention Operations primarily include engaging with customers through telephonic renewal calls or other mediums to pay renewal premiums on time. With close to 70% of total policies present in the premium renewal base, tracking, scheduling, execution of customer retention calls and campaigns contribute to a major cost head for life insurers.

In this paper, the authors present an advanced analytic solution to effectively manage customer retention costs and improve overall persistency. The paper demonstrates the use of several Machine learning and Deep Learning Neural Network based models to classify the customers based on propensity of not paying renewal premiums on time. The study includes a comparative analysis of model performance with the Deep Learning Neural Network model showing the highest performance. The propensity scores were used to create a solution driving differentiated retention strategy, matching customer segment with appropriate renewal efforts to reduce customer retention cost and improve persistency.

Implementing Learning Analytic Tools in Predicting Students’ Performance in a Business School

R.Sujatha and B.Uma Maheswari  
PSG Institute of Management, Coimbatore, Tamil Nadu, India

In recent times, information technology and big data are two buzz words that have impacted all sectors including education. Research in the field of educational data mining and learning analytics is in its nascent stage. Applying analytics in education...
is the need of the hour especially in the context of a developing economy like India. It is time for educational institutions to use machine learning tools to enhance teaching learning experience. This study deploys learning analytics technique using the data of students undergoing a post graduate management program and attempts to create a system of preventive feedback mechanism for faculty and students. In the first part, logistic regression was used to identify the academic status of foundation courses in the first semester. Six models were developed and ‘specificity’ scores were used to test the validity of the models. The specificity scores of the models ranged between sixty percentage and ninety three percentage thus validating the predictive capability of the model. In the second part of the study, the stepwise regression model was used to predict the marks of the student in the capstone course. The results showed that as the student progresses into second semester courses, the tenth and higher secondary board becomes irrelevant. Performance in the first semester courses greatly influences the results of the second semester. Deployment of the models developed in this study would go a long way in not only enhancing students’ performance but also more fruitful faculty engagement.

IC 19/008
Solving the CECL Riddle through Risk Analytics
Varun Aggarwal and Jyoti Thakral
EXL Analytics, India

Issued in June 2016, and set to take effect in 2020 for SEC filers and in subsequent years for other financial institutions that follow U.S. GAAP, the Current Expected Credit Loss (CECL) accounting standard reflects a sea change from the 40 years old traditional loss accounting standards of Financial Accounting Standards Board (FASB). Introspection of 2007-2009 financial crisis identified “delays in recognition of credit losses” by lenders as one of the primary areas of improvement for U.S. banking and financial services industry. The financial regulators since then have been attempting to bring meaningful transformations to guard financial institutions against systemic risk. CECL is a new accounting standard that will change how the U.S. financial institutions account for expected credit losses.

While CECL accounting standard is intended to make financial institutions prudent at provisioning for losses, its implementation poses significant challenges. One of the biggest changes is moving away from loss forecast horizon of one year to life of asset. Accounting for lifetime losses for short-term closed-end installment loans such as auto loans and personal loans will be different from approaches for revolving loans such as credit cards and the long term closed-end loans such as mortgages, where estimation of life of loan itself requires a robust framework.

This paper aims at highlighting a) key challenges for CECL implementation for large and small financial institutions, b) life of loan estimation frameworks for different types of financial assets; and c) CECL measurement approaches, varying from portfolio level to account level methodologies.
Comparative Study of Different Benchmarking Techniques

Stuti Mohan and Swati Verma
Experian Credit Information Company of India

"Obviously, the highest type of efficiency is that which can utilize existing material to the best advantage." — Jawaharlal Nehru

This is so true for the banking industry, which currently is undergoing tremendous upheavals. With the NBFCs and the Fintechs challenging the incumbents, the banks need to be efficient, now more than ever.

The paper aims at estimating and comparing banking efficiency scores for Indian scheduled commercial banks using two approaches :-(1) 2 stage DEA Analysis (2) Stochastic Frontier Analysis

Stage 1 DEA results are estimated using operating profits and advances as output variables. The efficiency scores, obtained are then regressed on the financial ratios(CAMEL), macroeconomic variables and dummy of the political variables as the part of the 2nd stage of the analysis.

As a part of the third stage of our analysis, we estimated the same data with the same input-output variables using the stochastic frontier analysis.

HDFC, ICICI, AXIS bank are coming out to be the top performers and Indian overseas bank, Nainital bank are coming out to be worst performers under DEA. The private banks are outperforming the nationalized banks in terms of the efficiency under the DEA approach. The SFA analysis gives an efficiency score of more than 0.90 to most of the banks which is an overestimation.

Across the approaches, liquidity, earning ability and the asset quality ratios are coming to be significant determinants of efficiency.

Our study captures the difference in efficiency scores obtained from both approaches. The results obtained from SFA are in a more restricted range, and not providing better differentiation between private and nationalized banks. DEA provides score in a more wider range because of the linear programming structure of the method and hence have more reliable estimates.
Optimization of Initial Credit Limit Using Comprehensive Customer Features

Shreya Piplani and Geetika Bansal
Experian Credit Information Company of India

Credit card is one of the most commonly used financial instrument within banking space across the world. While credit card started as a replacement for cash transaction, the product has evolved multi-fold to become the first source of funds for large majority of the earning population in India. Approximately 40mn adult population have credit card and we estimate almost 32Mn of these using the plastic for their spending needs. Credit Card is the most convenient and flexible form of credit with an interest free period during which there is no interest incurred on the outstanding balance. Credit card issuers determine the appropriate limit based on various factors such as income, limit on other cards, employment type, credit history, relationship with the issuer etc.

While credit card is an attractive product for banks, it is often associated with high delinquencies, inactivation and dormancy. In the current scenario, approximately 18% of the population does not activate their credit card post issuance and 45% utilizes less than 25% of the credit assigned to them. For the credit issuers, this eventually leads to capital blockage and operational costs for inactive customers. As a consequence, there is an opportunity lost in terms of procuring prospective customers and generating more revenue with minimum possible risk.

This paper resolves the above defined conundrum by arriving at the optimal initial credit limit of a customer that maximizes the revenue and minimizes the risk associated in the same environment subject to constraints on essential factors such as income, credit worthiness, other limits etc. Data collected as a part of application information and historical credit behaviour on other credit lines is used for the analysis. Key factors that ascertain the revenue component are activation of the card, retail and cash spend on cards which defines the utilisation of the limit, payment behaviour which defines the revolving pattern and fees being charged by the issuer such as annual fees, late payment charges, etc. However, with increased expected revenue there is a higher risk which is accounted for by considering the probability to default and the exposure at default on the given limit.

Separate models were developed to arrive at the expected values of the each of constituents mentioned above using multiple techniques like Logistic Regression, Gradient Boosting, ARIMA forecasting etc. Each component model captured various attributes such as credit utilization, number of distinct credit lines, payment behaviour on other cards, primary and secondary card utilization etc. Minimum and maximum accuracy of each component model was 56% and 72% respectively. Post arriving at the optimal limit, 26% of the customers were assigned a higher limit, keeping the risk factor constant leading to an expected growth in revenue by 16%. On the other hand, 10% of the customers were identified with lower expected utilization or with higher probability of default. Hence, this segment can be assigned a comparatively lower limit without any change in the overall revenue.
IC 19/012

Improved Forecasting Techniques for New Product Introduction using Ensemble Models

Sam Albert P, Ananda Swarup Das, Asad Ahamad and Anil Kumar Goyal
Deloitte Consulting, India

New Product Introduction (NPI) is a challenging problem as there are no historical data available to accurately forecast the demand of the consumer goods. Overestimation leads to warehousing issues as well as loss of revenue in logistics while underestimation leads to lost sales. In this work, we show how ensemble techniques like Random Forest in conjugate with Dynamic Time Warping based K-NN method can help to achieve forecasting accuracy in the range of 75% to 80%. Our method has been tested in a consumer goods industry and have been found to be accurate in the aforesaid range.

IC 19/013

Mitigating Agricultural Lending Risk: An Advanced Analytical Approach

Aditi Singh and Nishtha Jain
Experian Credit Information Company of India

As per the Situation Assessment Survey (SAS) for Agricultural Households by NSSO 70th round, in 2012-13, almost 40% of the agricultural households still relied on non-institutional sources for their credit needs, an increase of almost 11% over 1990-91. Moneylenders still form a major part, around 26%, of that non-institutional credit. Even with the rising credit disbursements and loan waivers, we haven't been able to improve the situation of our farmers. In FY 2018, banks disbursed only an additional 6.37% to this sector which is the lowest in a decade. Lack of sufficient information about the agricultural finance landscape and mounting NPAs have contributed to banks being reluctant to lend in this sector. In this study, we aim to build a credit risk assessment model for farmers to bridge the gap between them and the formal credit sector. We obtained a robust model that uses NSS data features and gives a better prediction of a farmer going to default than the one built on features currently captured by banks. These additional variables give necessary insights into farmer characteristics that could 1) help banks identify low risk farmers and expand their lending and 2) help the government identify problem areas where their intervention is necessary to uplift farmers and bring them under the purview of formal credit.
Supply Chain Contract for Short Life-Cycle Products with Production Capacity Constraint

Raunaq Srivastava  
Indian Institute of Management Ranchi, Jharkhand

Pritee Ray  
Indian Institute of Management Ranchi, Jharkhand

This paper considers a two-stage supply chain with one manufacturer and one retailer for short life cycle products, in a risk neutral setting. The retailer’s demand is dependent on the initial stock levels and price of the product displayed at the store. The retailer’s order quantity is limited by the capacity of the manufacturer. This paper proposes a centralized setting and three decentralized settings with revenue sharing contract, buyback contract and a combined contract. Through numerical analysis, it is observed that the retailer’s expected profit is better in buyback contract compared with the revenue sharing contract and vice versa for the manufacturer. For retailers who prefer the revenue sharing mechanism due to lower initial cost of procurement, the combined contract improves the expected profits than only a revenue sharing contract. We construct a game where both the manufacturer and the retailer take chances leading the contract negotiation. We find that in a manufacturer led game, the combined contract is preferred while improving the retailer’s profits. Sensitivity analysis is carried out to determine the effect of factors like stock coefficient, price sensitivity and capacity of plant on the overall supply chain performance. While stock dependency and demand uncertainty are inversely proportional to the supply-chain performance, increase in price sensitivity and capacity of the plant, improve the overall supply-chain performance.

Applying Analytics Techniques to effectively understand match phases in major European soccer tournaments (2013/14-2017/18 seasons)

Seth Riddheish Anilkumar  
Indian Institute of Management Calcutta, West Bengal

This paper tries to understand match outcomes and trends shown by football teams over extended time spans. It tries to show why certain styles of play are prevalent (or not), and why certain results tend to form a larger trend by themselves. Things that have been mapped include the importance of scoring/conceding the first goal, home vs away, expected outcomes from statistical models and how often they pan out, statistical anomalies that are simply not explained by current and previous works in the domain (a club’s current season, impact injuries in English grounds) and also attempts to find player specific insights besides testing over the ongoing season for maintaining validity.
It must be borne in mind that timelines in sport are hurried, and that actual on ground events often do not pan out as per simulation or models- some phenomenon in current and previous seasons still continue to confound logic. Another honest attempt in this paper is detailing why certain things are not analysed, and what information asymmetry exists as we try to predict sporting outcomes. The output is also eliminating some parts of sample size from consideration (domestic cups, for instance)- justification for this is also given, as well as why building a macro model may not work.

IC 19/019

Unraveling Fake News “Segmentation & Classification based Detection Approach”

Abhimanyu Tyagi, Akash Chopra, Akshayjeet Singh, Bhuven Khurana, Pranil Ranjan Ghosh, Sameer Uttam, Pankaj Mittal, Sandeep Kumar Sharma
Deloitte Consulting, India

Fake news is one of the most ardently contested socio-political influencer in recent times. Many countries around the globe are facing flak for non-containment of impact originating from a fake news on any platform. With people around the globe defining “media” as both content and social media platforms, the trust in journalism is being perceived as trust in all the information on these platforms.

As per a recent survey almost 89% of the population have a strong opinion that the platforms must take responsibility for spread of fake news. This calls for a need of a sturdy framework or methodology to flag and contain the fake news. In our paper, we focus on an intelligent framework, which can flag any content on social media platforms into four buckets: “Trusted,” “Non-Trusted,” “Hoax,” and “Opinion.” Our focus would be to extract the right information even if it is embedded in a misrepresented fact and compare it against the trusted sources to arrive at a conclusion and flag it appropriately.

IC 19/021

Vehicle Routing Case Study

Asad Ahamad, Deepu Unnikrishnan, Rohini Prasad and Sam Albert P
Deloitte Consulting, India

Vehicle routing problems (hereafter referred to as VRPs) are common in all transportation systems from courier services, retail distribution networks, public/private transport etc. However, since VRPs are combinatorial problems, and are NP-hard problems there is no singular approach that caters to all requirements. Therefore, VRPs are still widely researched and multiple heuristics are still being developed to find near optimal solutions in a reasonable amount of time, because brute force approaches (such as branch and bound) can never be applied to big datasets. All the current best approaches iteratively improve feasible solutions using Meta-heuristic
approaches. Meta-heuristics algorithms (predominantly evolutionary algorithms) are currently the best at solving combinatorial problems, and will continue to be so for VRPs, at least till the onset of Quantum Computing.

The case study will discuss a quick, flexible and modularized meta-heuristic approach that provides near optimal solutions for VRPs in real life scenarios, that are often more complex as compared to standard theoretical problems. The approaches used to tackle real constraints on the delivery nodes, drivers, trucks and routes will be addressed in the case study. This will be followed by a discussion about the operational benefits and cost savings that were observed while applying this approach to an actual service center.

IC 19/022

Food Index Forecasting

Kalyani Dacha, Abir Sinha and Ramya Cherukupalli
Deloitte Consulting, India

Designing efficient and robust algorithms for accurate forecast of price index is one of the most prevalent challenges in the food market business. With the exponential rate of development, evolution of sophisticated algorithms and the availability of fast computing platforms, it has now become possible to effectively and efficiently extract, store, process and analyze food price index data with diversity in its contents. Consumer products companies in US approached us to use the price index data for over six food categories and forecast for the upcoming 18 months so that they can get an idea about the upcoming price trends, which will help in making more informed decisions in their business.

In this paper, all the data used for analyzing purpose was external (freely available) data. We used the monthly price index data for US for raw/processed food categories, published by United States Department of Labor for the period January 2010 till May 2018. Different univariate and multivariate time series modeling approaches were used to model the price data and best fitted model was chosen for each of the individual categories.

IC 19/023

Cognitive Customer Value Management: A systematic approach for customer churn analytics in a non-contractual context

Kalyani Dacha, Vamsee Krishna Mamanduru, Abir Sinha, Sujit Acharya
Deloitte Consulting, India

In the current transformational digital world, markets have become extremely competitive and businesses are reaching the saturation point in no time with
In our paper, we illustrate in greater details how Bayesian Optimization technique can be utilized to solve real world combinatorial optimization problems while simultaneously refining on the hyperparameter tuning issue that has been a popular challenge for machine learning modelers across industries. The paper provides a formal mathematical structure to the Multiple Boosting methodology which forms the backbone of our improvised algorithm wherein varied and multiple objective functions are combined to form one joint boosting function to be chased which needs to be methodically accurate and is business-intuitive at the same time.

Interestingly, mathematical structure of BOMB relies on a two-part process. In part one, an aggregated factor is made by normalizing the individual variables in a unique mathematical procedure. Factors are selected based on the requirements of the global solution that the modeler wants to fulfill. Factor selection methodology is advised to be business driven, but mathematical modelling can accomplish the task too. In part two, the aggregated factor is normalized again in the same procedure (as in part one) to calculate weights and the specific weight entailing objective function is constructed as the single objective function which is used for hyperparameter optimization in the Bayesian system.
IC 19/025

Explainable Artificial Intelligence Model: Analysis of Neural Network Parameters

Sandip Kumar Pal and Amol A Bhave
Cognitive Business and Decision Support, IBM India

Kingshuk Banerjee
Cognitive Computing and Analytics, IBM Global Business Services

In recent years, Artificial Neural Network (ANN) is becoming a popular technology to extract the extremely complex pattern in the data across different segments of research areas and industrial applications. Most of the Artificial Intelligence researchers are now focused to build smart and user-friendly applications which can assist humans to make the appropriate decision in the business. The aim to build these applications is mainly to reduce the human errors and minimize influence of individual perceptions in the decision-making process. There is no doubt that this technology will be able to lead to a world where we can enjoy AI driven applications for our day to day life and making some important decisions more accurately. But what if, we want to know the explanation and reason behind the decision of AI system. What if, we want to understand the most important factors of the decision-making processes of such applications. Due to the intense complexity of inherent structure of AI algorithm, usually researchers define the Artificial Neural Network as "Black-Box" whereas the traditional Statistical Learning models are more transparent, interpretable and explainable with respect to data and underlying business hypothesis. In this article we will present Transparent Neural Network (TraNN) by examining and explaining the Network Structure (Model Size) using statistical methods. Our aim is to create a framework to derive the right size and relevant connections of network which can explain the data and address the business quarries. In this paper we will be restricting us to analyze the Feed-Forward Neural Network model through Non-Linear Regression Model and analyze the parameter properties guided by Statistical Distribution, Machine Learning Algorithm and Simulation technique.

IC 19/028

Assessing Service Quality using Sentiment Analysis – Evidence from Wellness Resorts in India

Dibya Nandan Mishra and Rajeev Kumar Panda
School of Management, National Institute of Technology Rourkela, India

The purpose of this study is to identify dimensions of service quality that are important in creating value for a customer in wellness resorts in India. This value indeed impacts the customers’ satisfaction level. The paper may help to identify the required area of improvements in the field of service, relationship and hospitality management. Sentiment Analysis approach is used to process public reviews shared on TripAdvisor and similar platform to identify customers’ opinion on the services
provided by wellness resorts in India. A machine learning based Sentiment Analysis approach is used. The research tries to identify factors of service quality and rank by their impact on building customer satisfaction level. This study may support Managers and Policy Makers to identify areas of improvement to help develop the wellness resorts in India. Future researchers may be reinforced with understanding customer experience through social media rather than traditional questionnaire-based surveys.

IC 19/029

Credit Scoring Using Machine Learning Algorithms

Sandhya Kuruganti
Ex-SVP Analytics, Citibank

Gunnvant Singh
Jigsaw Academy

There is enough evidence to support the fact that machine learning has the potential to outperform traditional solutions in real-world settings. However, the application of these advanced algorithms for credit lending decisions in the financial services industry has been a reluctant effort despite rigorous academic pursuits with several publications in peer reviewed journals. This lag could be attributed to a multitude of factors – complex black box algorithms, relative performance vis-à-vis traditional approaches and adherence to regulatory requirements, among others.

The case study presents a practitioners’ approach for building credit scorecards to predict customer default behavior for a loan portfolio. Using a real-world lending data set, traditional and machine learning algorithms are deployed. A scorecard is developed using logistic regression based on weight of evidence. Machine learning individual classifiers of support vector machine and artificial neural networks are applied on the same data. Ensemble techniques of bagging and boosting are also deployed. Assessment of relative performance of the resulting scorecards based on a comprehensive set of metrics as per industry practice is showcased. Concluding results as well as summary findings of interim simulations are discussed. Further, insights and learnings from hyper-parameter tuning and “white boxing” of the algorithms is attempted. Lastly, benefits and challenges of leveraging machine learning algorithms for credit scoring model development and its implementation are highlighted.
An ARDL & Cointegration Approach for Analyzing Determinants of Foreign Portfolio Investors’ in India

Parul Kumar
Maharaja Agrasen Institute of Management Studies

Sunil Kumar
Department of Commerce, IGNOU

R K Sharma
BVIMR

This paper examines the relationship of FPI Net flows domestic with international financial & macroeconomic indicators. The time frame covered by the study is from January 2000 till December 2017. Auto Regressive Distributed Lag method along with the Co-integration Analysis is used. Results highlighted that the major determinants of FPI in India are, Nifty returns, Wholesale Price Index, Index of Industrial Production, Rupee Dollar Exchange Rate, NSE Market Capitalization, Foreign Exchange Reserves and in terms of international factors are S&P 500 returns and MSCI Emerging Market returns. LIBOR, CMR, Broad money & MSCI World Index returns are not significant in explaining the variations in FPI to India.

Public Sector Banks Lending for Residential and Commercial Real Estate in India

Tanu Aggarwal and Priya Solomon
Amity University, Noida, India

India is considered as the attractive business source in the service sector because of favorable demographics and economic growth which lead to attract property investors for Real Estate Development. Real Estate Sector is like a treasure for the Indian economy.

The Real Estate Sector is considered as the major part of our economy and has the multiplier effect which helps in increasing the economic growth of our country. It is contributing about 6% to country’s Gross Domestic Product. The Residential and Commercial Real estate has been increasing at a continuous rate because of increased population, migration of people to urban areas, increasing the income level of the people and rise of nuclear families in India. Real Estate Sector Development is lucrative for business in India. The demand for the residential loans is more as compared to commercial loans by public sector banks which have been shown by using Trend Analysis. The Partial Least square structured equation modeling (PL-SEM) shows that there is the impact of residential lending, commercial lending on total real estate sector development in India. The results also depict that Public sector banks are contributing more for the Residential Real Estate as compared to Commercial Real Estate. This reflects that Residential Real Estate Sector has more demand in the market as compared to commercial real estate sector development.
India-Bimstec Bilateral Trade Activities: A Gravity Model Approach

Gurpreet Kaur and Akriti Gupta
Mittal School of Business, Lovely Professional University, Phagwara

The Bay of Bengal Initiative for Multi-Sectoral Technical and Economic Cooperation convergence of economic interests coming out of India’s Look East Policy and Thailand’s Look West Policy. Its objective was to integrate the regions on both sides of the Bay of Bengal. The development of BIMSTEC countries is indispensable for the forward march of Asia as a whole. The present attempt to analyze the India-BIMSTEC trade activities after the formation of BIMSTEC bloc from 1997 to 2017. The study contains two parts- first part examine the trade creation and trade diversion effect of trade India and BIMSTEC countries and second part focus on forecasting of trade. The model estimated with panel data and analyze the trade flows between India and BIMSTEC countries. To fulfill the objective of study econometrics models i.e. gravity model and Auto Regressive Integrated Moving Average (ARIMA) has been used. Model estimates sets of regression equations to measure the effects of Regional Trade Agreements (RTAs) using Ordinary Least Squares (OLS) with nation dummies to capture country specific fixed effects. The Study revealed that all coefficients of regional dummy variables are mostly positive and significant, indicating agreements tend to enhance more trade than bilateral trade agreements and study also stated that based on India’s trade with BIMSTEC region, there exists a scope for intraregional trade in future. Therefore, efforts at the international level are required to be made to increase intraregional trade to earn a fair name for BIMSTEC in the world trade.

Performance Analysis of Optical Character Recognition Tools in Natural Scene Text Detection using OCR Post-Processing

Sneha Banerjee, Parteek Bhatia, Sawinder Kaur,
Department Of Computer Science & Engineering, TIET, Patiala, Punjab, India

In recent times, there has been a rapid advancement in digital data mainly in visual formats, such as images from the web, mobiles, digital cameras, screenshots, etc. Thus, there is a need to retrieve the information within such textual images. Optical Character Recognition (OCR) is a method used for converting textual images into readable text format. There are various Optical Character Recognition tools available which help in converting visual data into editable textual documents. None of the available Optical Character Recognition tools are perfect in extracting information from the images accurately, therefore a method of post-processing on the retrieved text to improve the accuracy of the detected text from images has been proposed. In this paper, a performance analysis between various Optical Character Recognition
tools like Tesseract-OCR, Google Cloud Vision and Amazon rekognition is presented on natural scene images. It has also been observed that after removing spelling errors from the identified text in images resulted in a significant improvement in the accuracy of the output text. There has been an improvement of around 2% in the case of natural scene images and approximately 8% in the case of text obtained from handwritten images.

IC 19 /039
Project Delays and Business Analytics – Innovation, Risk Mitigation and Emerging Horizons

A.A.Attarwala and C.S.Balasubramaniam
Kohinoor Business School, Vidyavihar (west), Mumbai

Project delays coupled with cost overruns have been causing heavy concern to the Governments both at the Union and State level in the past. The factors responsible have been identified as the delays in clearances, lack of supporting infrastructure, law and order problems, and delay in shifting of utilities, pre-project activities, equipment supply, fund constraints and geological surprises. As per Government’s online monitoring system, of the 351 infrastructure projects worth over Rs 1,000 Crore, 127 were delayed and 115 faced cost overruns as of February 2018. As many as 51 faced both time and cost overruns. Financing the infrastructure and mobilising required investment by State Government agencies would be a challenging task outside the Government budgetary support. Traditional techniques of project assessment which have been used in the past have neither helped the Governments nor the beneficiaries, contractors and other stakeholders. Big Data / Business analytics tools which can be powerful impact making in the implementation and financial terms, risk mitigation measures are presented as exploratory models. Parameters such as cost, resources, size of the project, duration are considered. Innovative Options and alternatives to reduce the cost and risk would be enabling the study to arrive at optimization approach. Emerging Horizons indicate the novelty, technology advancements and growing perspectives.

IC 19/041
Real Earnings Management Practices for Meeting Earnings Benchmarks: Indian Evidence

Karan Gandhi
University Business School, Panjab University, Chandigarh

Literature posits that earnings benchmarks pressurize and incentivize managers to take abnormal operating decisions. This study examines whether managers overproduce and/or reduce research and development expenses and/or reduce selling, general, and administrative expenses to meet two important earnings
benchmarks, i.e., avoiding losses and sustaining last year's earnings. It also categorizes selling, general, and administrative expenses into three categories, viz., marketing expenses, welfare and training expenses, and the remaining as other general and administrative expenses, so as to examine spending on which of these categories is reduced for meeting the aforesaid earnings benchmarks.

The results, based on large sample of Indian public companies during the period 2000 to 2016, suggest that, on average, managers overproduce and reduce selling, general, and administrative expenses to avoid losses and to sustain last year's earnings. However, this study does not find evidence on reduction made in research and development expenses for meeting both the earnings benchmarks.

Individual examination of three categories of selling, general, and administrative expenses reveals that, on average, managers reduce marketing expenses, welfare and training expenses, and other general and administrative expenses to avoid losses. However, to sustain last year's earnings, managers reduce marketing expenses and welfare and training expenses. The evidence on management of selling, general, and administrative expenses and its three categories for meeting earnings benchmarks are robust on at least two of the four alternate cross-sectional estimation models except for the evidence on management of marketing expenses for sustaining last year’s earnings. Lastly, findings suggest that avoiding losses benchmark is more important vis-à-vis benchmark of sustaining last year’s earnings.

IC  19/043

Adoption and Implementation of HR Analytics in Damodar Valley Corporation (DVC)

Binay Tudu
Damodar Valley Corporation, Kolkata, Part-time Ph.D. Research Scholar, Indian Institute of Technology, (ISM), Dhanbad, Jharkhand
Saumya Singh
Department of Management Studies, Indian institute of Technology, (ISM), Dhanbad, Jharkhand

This paper is mainly a survey research on adoption and implementation of HR analytics in one of the Indian Public Sector Undertakings primarily engaged in thermal power generation, transmission and distribution. The key objective of the study is to assess the knowledge of HR analytics amongst HR officers in DVC and to measure the level of preparedness regarding adoption of HR analytics in the day-to-day functions for cutting down on administration and establishment cost. The survey was carried out of total 50 HR executives in three managerial levels, i.e. Junior, Middle and Senior Management.

Seven numbers of questionnaire statements, based on different parameters were distributed both offline and online for conducting surveys. The survey response was measured and scaled on a Likert Scale to determine the predictive analytics. In operational analytics, total count of each response at each level is shown in the pie chart represents the percentage of outcome of the survey for each managerial level.
The response “agree” is less than 60% at all hierarchy levels in operational analytics. The coefficient of correlation ranges from 0.89 to 0.98 in case of all the three identified predictive analytics. It shows that there is a strong correlation between variables. In all three predictive analytics, regression model explains linear relationship between dependent and independent variables. The Coefficient of Determination ($R^2$) is from 0.79 to 0.96 in predictive analytics, it means regression does explain a high fitted linear model which exactly ensures good approximation of the actual relationship between dependent and independent variables.

Accordingly, HR Strategy Map has been designed with reference to the identified measurable and quantifiable HR objectives of DVC for preparation of the future roadmap for adoption and implementation of HR analytics.

IC 19/046
Managing Class Imbalance in Big Data- A Distributed Approach to Synthetic Minority Oversampling Technique

Avnish Kumar Rastogi and Nitin Narang
HCL Technologies, India

Most real-world prediction problems like cancer prediction, etc. are problems involving imbalanced data. Here an imbalance in the dataset is represented by a mismatch in class representation. An application of classification algorithms on imbalanced data is biased in favor of the majority class and gets further biased in case of high-dimensional data. Synthetic Minority Oversampling Technique (SMOTE) is one technique proposed by Chawla (2002) that vastly improves prediction in imbalance datasets. Unfortunately, existing SMOTE (Chawla 2002) implementations are inept to handle large data sizes, demanding need for distributed computing solution. In this paper we investigate challenges for implementing SMOTE in distributed environment and present algorithm and results of our distributed SMOTE implementation. Our implementation is done in Apache Spark using variations of high and low dimensional data as well as large and small datasets. The results are compared to existing monolithic implementation in python SMOTE. We were able to successfully demonstrate a distributed version of Spark SMOTE which generated quality artificial samples better than the existing python implementation while preserving spatial distribution of the data.
Integrating TODIM into Data Mining for the Evaluation of Customer Lifetime Value

M. Ray  
Birla School of Management, Birla Global University (BGU), Bhubaneswar  
B. K. Mangaraj  
XLRI-Xavier School of Management, Jamshedpur

TODIM is a Multi Criteria Decision Making technique used to order the alternatives based on pair comparisons where the relative measure of dominance of one alternative over another is used to get the complete ranking of the alternatives. Here, we integrate this technique with data mining in a fuzzy decision environment for the evaluation of Customer Lifetime Value (CLV). This construct measures loyalty and can be used to segment customers to design appropriate marketing strategies. This work considers a problem of ordering customer segments based on four criteria for evaluation of the CLV, viz. length (L), recency (R), frequency (F) and monetary value (M). Further, Fuzzy Analytic Hierarchy Process (FAHP) is used as a tool for group decision-making to determine the weights of LRFM criteria. We considered nine decision-makers from different management layer of the sales department for making pairwise comparison of the LRFM variables in Saaty’s 9-point scale. The consistent judgment matrices obtained from AHP analysis was then used to construct FAHP matrix for its further analysis.

To implement the model to a real life case of Indian retail sector, we prepared customers’ CLV file for generating L, R, F and M data of customers by accessing store’s data-base for a particular period. Then fuzzy c-means clustering (FCM) was carried out by using MATLAB 12.0 for identifying clusters of customers. The inputs from FAHP and FCM were then used in TODIM model to get weights of these clusters to help managers in policy making and designing suitable marketing strategies.

Privacy and Social Media: Do Users Really Care?

Swati Bhargava  
IILM Academy of Higher Learning, Jaipur

Social media provide platforms for us to share personal information as part of our social interaction with other people. Most of these online services and platforms are also freely available. However, our personal information tends to be collected and used commercially (not always in the most transparent way), and this is the price we pay for free access.

In this research paper, we are focusing on the modern use of social media. Our aim is to investigate the degree social media users care about their privacy issues when sharing personal information via social media platforms. In the means of investigating
to what extent users share information with other users, and to what extent they are aware of the information social media owners collect and (also) sell to third-party companies.

In addition, our intention is to gain insight into social media users’ awareness of their personal information being used for commercial use. To what degree would such information affect their willingness to share personal information? Lastly, our aim is to investigate if a person’s opinion of privacy affects user behaviour. Do the users really read and consider the content of privacy policies and do they actively take control of the privacy settings?

Through conducting a user survey, it gained insight into social media users’ habits and their knowledge concerning the discussed topics. An evaluation of our findings ultimately leads to the conclusion of the privacy paradox holding true for the users involved in our study. We found that users claim to care about privacy online and that many have knowledge concerning the aspects analysed. However, they do not read the documentation and still utilise the services provided without having a clear understanding of how the technologies work on the Internet.

IC 19/050

The Effect of Sustainability on Foreign Direct Investment Inflows: Evidence from developing countries

Neha Singh and Cheshta Kapuria
Faculty of Management Studies, University of Delhi, New Delhi, India

The growth of foreign direct investments (FDI) in the world has been significant since 1990s. The transnational flow of FDI has increased manifold over the years. However, the strengthened process of globalization poses the issue concerning the quality of FDI. Particularly, the dilemma on balance between growth and sustainability which led to the need for enhancing the quality of FDI i.e. Sustainable FDI. With this backdrop, the paper aims to examine the impact of economic, environmental, governance and socials dimensions on Sustainable FDI for select 22 Asian developing countries. Dynamic panel (two-step Difference GMM) estimation method has been applied for the time period 2000 – 2016. Our econometric results indicate that, there exists a positive significant relationship between control of corruption, political stability and electricity consumption in relation to sustainable FDI. On the contrast, CO2 emissions have a negative relationship with sustainable FDI. The result underlines deficiencies in the information technology variable, which has a non-significant yet positive relationship with sustainable FDI. Based on the results, the paper concludes from a broad policy perspective.
Analyzing Regression and Decision Tree Models for Credit Risk

Tarun Jain, Kapilanjan Bhardwaj
KPMG Advisory Services, India

This paper is related to the development and validation of credit risk models using predictive analytics. Our work is divided into two parts, first, we are using logistic regression and decision tree approaches for model development. Second, we are validating models developed by both these approaches to check the model accuracy.

Both the logistic regression and decision tree approaches are useful for classification problem and have their own significance. It is up to the data modeller to acknowledge the best models that fit in the data and business problem. The result shows that Logistic Regression is more suitable to forecast losses and also useful to find probability of default for each customer as every obligor has its own default probability based on its characteristics whereas decision tree approach is more suitable for creating homogeneous and heterogeneous pools. Hence, only limited sets of pools are created which is not sufficient to provide a granular distinction between the obligors in terms of default risk. Hence, Logistic Regression is more useful in predicting Probability of Default, whereas decision tree approach is more useful to take strategic business decision for specific segments of customers.

Stacking With Dynamic Weights on Base Models

Biswaroop Mookherjee and Abhishek Halder
Tata Consultancy Services (TCS), Kolkata, India

Stacking is used to combine models based on different techniques using a second level model to come up with higher accuracy. The second level model essentially uses the values predicted by different base level models as independent variables, while the dependent variable remains the observed one. Though fit of the base level models differ at various parts of the data, the second level model uses same set of weights on base level models on the whole data. We have derived two methods where we replace the second level model by a linear combination of base model outputs where the weights vary. In our methods, we select a part of the data based on some predefined condition of proximity for classification of a new observation. Then weights are assigned on different base models considering their accuracy in that part of the data. In one method all points in the neighbourhood gets equal importance while in the other method points get importance based on proximity. The algorithms apply same principle on each of the new observations which get their neighbourhoods in different parts of the data; thus, weights vary. The new ensemble methods are tried on different datasets from different fields and found to give better results than conventional stacking.
IC 19/058

Improving Blast Furnace Operations through Advanced Analytics

Rishabh Agrawal and Suresh R.P.
Supply Chain & Operational Analytics, Accenture Applied Intelligence

Hot blast is an input to the blast furnace and is instrumental in blast furnace efficiency. The current state of stove operations is not standardized. Many times, operators based on their experience take critical decisions. To standardize the decision-making process in the most optimum way, an analytics research project with an iron and steel manufacturing industry was kick started. The paper describes the present control system of the hot blast heating process and describes a model complementing the control system. The model is built by using K-means clustering algorithm and Principal Component Analysis to recommend the operators the critical variables set-point. The process variables in the plant are continuously changing and thus makes impossible for operators to take the most optimum decision accounting all the variables. The model covers all the critical variables whether controllable or non-controllable and is aimed at increasing the heat recovery in stoves and increase the temperature of the Hot blast. The end-result of the research would be to reduce the variability and increase the median temperature; eventually reducing the cost of Hot metal. The dashboard displaying model generated recommendations to the operators and also monitors the compliance by operators on week by week basis. Based on different business rules, the model is scheduled to be re-tuned in specified intervals which takes care of the change in efficiencies of process equipment.

IC 19/059

Decriminalisation of Algorithmic Trading: A Perspective From Indian Stock Market

Ritesh Kumar Dubey and Rajneesh Ranjan Jha
ICFAI Business School (IBS) Gurgaon, India

Often algorithmic trading (AT) have been seen with the skepticism of having ill effects on both markets and market participants. And, regulators have always been critical about AT, if AT can be allowed in long run. AT has been the scapegoat for any mishaps happening in the markets and AT and HFT have been taking the blame for any of the market crashes across the globe. It has been almost a decade since the introduction of AT in the Indian market and its necessary to figure out if we really need to criminalize or decriminalize the AT. There is limited research which examines the impact of algorithmic trading on market quality. Whatever existing literature we have, it mostly probes algorithmic trading with respect to the quote driven markets. But, as an emerging market India and the other emerging order driven economies pose various differences in the market structure than that of the developed quote driven markets. In this study, we have direct identification of AT unlike most of the existing studies where proxies have been used and we examine the relationship of
With e-commerce giant like Amazon moving to Amazon cash in US and lending cash to SME’s and Rakuten who operates Japan’s largest Online credit card company by transaction value will only appear Banks to lose vulnerable special status in coming years.

With the benefits of Big Data, Advanced Analytics, Internet of Things, emergence of Smart Phones and Wearables, Biometric, Robo-advisors and Voice assistants like Alexa and Siri, has helped in reducing human interactions and increased ROI on technology and customers retention for Big Technology firms and Fintech companies.

Banks will have a stiff challenge from big IT firms instead of Fintech companies in meeting customer expectations due to dearth of technical skills and knowledge.

This paper conducts a systematic and extensive review of recent trends, reports and articles on what all factors that will drive Banks to transition and become a future technology company.

This study is conducted based on a descriptive research design. The results in this study indicate that focus on 1. Developing a “consumer first” approach, 2. Investment in Banking Corporate Asset and Infrastructure 3. Changes to Workforce – From Hiring Data Scientist instead of Accountants and 4. Data insights into business value will help Banks to slowly transition and become the future technology company.

Possible directions for future research on Banking technologies are also discussed.
Stock Selection Based on Cluster Analysis: Supplementing The Stochastic Models For Decision Making

Rajneesh Ranjan Jha, Ritesh Kumar Dubey and Tavleen Kaur
ICFAI Business School (IBS) Gurgaon, India

Clustering technique is used across the various disciplines; especially in behavioral and life sciences (Jensen, 1971) and recently in the field of management it finds a lot of emphasis in marketing discipline. Portfolio management by managers requires them to analyze the different levels of risk and returns that they are exposing themselves to. Intuition behind diversifying the portfolios is to check the complementing nature of the exposures. Stochastic modeling has been the trend for the fund managers for creating the portfolios. Use of cluster analysis has not been much utilized by the finance manager. Cluster analysis attempts to classify items into clusters so that there is more homogeneity in the cluster than between the cluster. In this study we examine how cluster analysis can reveal important insights for portfolio management. We use cluster analysis to observe that stocks of selected companies of the National Stock exchange (NSE) can be categorized according to their degree of correlation. Stock returns are likely to be similar for an industry due to restrained macroeconomic environmental variables. In this study we use cluster analysis as our main tool to classify the stocks into a number of categories and examine how those categories change during different time periods. After clustering the stocks as per return or risk criterion, an informed investor can make use of the available analysis to trade according to his or her investment needs. With this paper we propose that cluster analysis can supplement the classic tools of stochastic modeling.

Machine Learning Approach for Formulating and Improvising Existing Local and Federal Government Policy

Siddarth Pai Manur and P C Siddhalingaswamy
Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal, Karnataka

Vasanth Kamath
T. A. Pai Management Institute, Manipal, Karnataka

Governments have a responsibility to deliver transparency, efficiency and accountability. In this regard, they have to make the best use of data that they have collected from their operations. Departments like Health, Education, Urban Infrastructure, Surface Transport, Shipping, Power, Law and Order do collect an enormous amount of data which is being generated over a continuous time frame. The effective use of data can really create a lot of value to citizens by improving their
experience with the Government and can greatly build the trust. Machine learning is a branch of data science that provides framework and ability to help the system to learn from the data with the mathematical modelling without explicitly being programmed to do so. We can use the Machine learning techniques for solving problems in Social science and generally in framing public policy so as to improve the service delivery in Government and experience for all its stakeholders.

We take a given problem in the public policy space and preprocess the data that is available and make suitable assumptions so as to adapt the suitable machine learning algorithm. We consider overall well-being of the ordinary citizen while framing the solution so that development is sustainable. We also study the correlation of improvement of one metric and its consequence over the other. In this study, we discuss various schemes of state and central Governments, to study and to suggest how machine learning algorithms can be used by them to improve the quality of living of its citizens.

IC 19/068
Side Effects of Communication - Intermediating Effect of Feedback Loop on NPS® Score

Lalit Barhate and Vinay Kumar
DATAWISE, Hyderabad, India.

It has long been established that there is a very strong correlation between Customer Satisfaction and Company Performance. It is also well established that Employee Satisfaction is an antecedent to Customer Satisfaction. Ample research also exists to demonstrate that influencing Employee Performance could lead to better Customer Satisfaction.

Net Promoter Score (NPS®) has been proven as an effective means of measurement and improvement of Service Quality, which demonstrates Customer Satisfaction and enhances loyalty. Enhanced loyalty is expected to assist higher growth in revenue, which depends on superior service quality and further helps to improve employee performance.

This study investigates the relationship between the scores given by customers on the NPS® surveys, the subsequent interventions by the Client team, and its impact on the subsequent NPS® scores over a period of time.

NPS® scores for the Client showed continuous improvements over the time for which the NPS® scoring was adopted. As a part of the process, feedback was sought to be captured for respondents giving low scores, with a possibility of corrective action being taken by the operations team of the Company. Subsequent measures of NPS® scores for such customers showed that these respondents improved the NPS® scores, clearly demonstrating that interventions through feedback surveys helped to increase the NPS® score for the organization as a whole.
IC 19/069

Time series analysis of the US Airline Industry: What financial statements do not reveal

Mohammed Alzeer and Ritvik Kuila
Datawise, Hyderabad, India

In the Airline Industry, financial performance has not always been the best indicator of corporate performance, (Richard A. Riley, 2003). Airlines all over the world are known for using Financial Engineering to improve presentation of their financial position. Some well-known methods of Financial Engineering include Sale & Leaseback, Manpower Outsourcing, Mergers & Acquisitions, etc. (E. Han Kim, 1993), (Kyle S. Wells, 2012). Investment Analysts have often relied on operational metrics such as Load Factor, Revenue Passenger Mile, Available Seats Mile, Number of Passengers etc. as a useful means of determining the health of the industry. With airline ticket prices remaining largely steady over the last decade (Bureau of Transportation Statistics, 2018), it is expected that the Financial Performance should reflect Operating Performance.

This study has examined the performance of Airline industry over a period of 10 years from 2008 to 2017. In specific, the study has examined the performance of large listed US based airlines - Delta Airlines, American Airlines, Southwest Airlines, and United Airlines. This study documents evidence indicating that the Operational Metrics of an airline are a better indicator of the share market performance of the airline as compared to the Financial Metrics.

IC 19/071

Response Prediction and Ranking Models for Large Scale Ecommerce Search

Seinjuti Chatterjee, Ravi Shankar Mishra, Sagar Raichandani and Prasad Joshi
Unbxd, Bangalore

The search engine funnel works in a progression series of impressions to clicks to carts to orders and the first optimization that naturally gets employed in the funnel is maximizing the probability of a click. If we can imagine a search session being the joint likelihood of the site features, user features, query features, product features, time series features and the outcome being the possibility of a click followed by cart or order, then it becomes the perfect situation of applying supervised machine learning models. The naive model at Unbxd calculates a composite score of click-ability and buy-ability of a product learnt over historical clicks, carts and orders as follows:

\[
\text{Score} = \sum_{n=0}^{N} \left( \frac{\text{clicks}}{(n + \alpha \cdot e^{0.5})} \right) \times \left( 1 + \frac{\text{carts}}{\text{clicks}} \right) \times \left( 1 + \frac{\text{orders}}{\text{carts}} \right) \times e^{-n/3}
\]
Where \( r \) = rank of the product and \( a \) and \( b \) are constants and \( m \) is lookback days.

To dynamically rank products we exploit various attributes extracted from impression and served search results such as site, query, device, user, time of day, day of week, query category, product category, query_dwelltime, time_to_first_click, time_to_first_cart, time_to_first_order, was_autosuggest_used, wifi_connection_type_of_user depending on the data collection granularity exposed through the search API. With these features and the event outcome we can fit a **Logistic Regression** model as a binomial classification task which emits the probability of click given a new impression based on its features. The beta coefficients of the model and ROC curve helps us understand the discriminating ability of the model to explain CTR through features. The business impact of the model is measured using inhouse A/B experimentation framework which tracks revenue metrics like CTR, CVR and STR apart from precision and recall per search and their statistical significance. Important business insights over multiple ecommerce domains like fashion, electronics, grocery etc in terms of CTR heatmaps helps achieve goals of search engine revenue optimization while personalizing the search experience for the user side by side.

**IC 19/073**

**Knowledge discovery based decision support system for application of systems thinking concepts in maternal and child health system strengthening**

**Partha Saha**  
BITS-Pilani, K. K. Birla Goa Campus

Maternal and Child Healthcare (MCH) has always been regarded as one of the important issues globally. Healthcare outcome of a region is very much dependent on factors like coverage of MCH interventions or availability of healthcare services and those factors are also interlinked with each other. Previously Health System (HS) was usually conceptualized as monolithic system. But it has been found out that HSs are also complex and dynamic systems build up with several building blocks. Paradigm shift is required from linear and reductionist research approaches to dynamic and holistic approaches for disclosing the complexity within HS components. Researchers have tried to find out influencing factors by analysing the complex HS with the help of Causal Loop Diagrams (CLD) and system dynamics modelling. Main purpose of the research work was to propose an alternative analytical framework for reduction of inequity among regions by generating knowledge about HS building blocks applying systems thinking approach and also to develop an interactive knowledge discovery based decision support system on web platform which would facilitate healthcare policy makers to integrate the analytical framework in the process of decision making.

For achieving aim of the research work, three interim objectives have been prepared. First objective was identifying all relevant variables under three major categories, i.e MCH interventions; Educational, Social, and Economic (ESE) parameters; and HS building blocks category. Second objective was development of the analytical framework and third objective was development of the DSS for incorporating the
proposed framework in the process of healthcare policy making. Results have confirmed that the proposed analytical framework would help healthcare policy makers to generate holistic understanding about MCH related elements separately for good, moderate, and poor performing regions before preparing strategic action plans for reduction of inequity among regions. Major advantage of the developed DSS is that users need not to be expert in statistical or data mining techniques for using the system.

IC 19/075
Characterization of the Skew Raised-Cosine Distribution

M. Ahsanullah
Rider University, Lawrenceville, NJ, USA

M. Z. Anis
SQC & OR Unit, Indian Statistical Institute, Calcutta

The raised cosine distribution has been used in communication systems. This is a symmetric distribution. However, many times data shows asymmetric distribution. In this work we consider skew raised cosine distribution. Several distributional properties and some characterizations based on the truncated first moment are presented.

IC 19/076
Application of Association Rule Mining in a Clothing Retail Store

Akshay Jain, Shrey Jain and Nitin Merh
SVKM’s NMIMS Deemed to be University, Indore Campus, Indore, Madhya Pradesh

In this paper an attempt has been made to understand the buying pattern of customer with the help of market basket analysis, which is an important tool in modern retailing industry. Retailing defined as the timely delivery of goods and services demanded by end customers at prices that are competitive and affordable. Through association rule in data mining we have tried to understand consumer behavior, brand importance, seasonality effect, buying pattern, product basket from the data. Data Mining is the practice of analyzing database to gather and generate new information. Association Rule Mining is a process of finding rules that govern relation between set of items. Market Basket Analysis is a modelling technique based upon the theory that explains the buying relation between a certain groups of items. The transactional data is collected from the retail clothing store “Try Us” located in Indore, Madhya Pradesh from the period November 26, 2017 to September 19, 2018 with help of Point of Sale (POS) and bar code scanner. “Try Us” is a small retail store with multiple brands and is planning to upgrade store to a multilevel store. Data Mining would benefit the overall
store performance. Frontline Solver® Analytic Solver Data Mining (XLMiner) is used for simulations.

**IC 19/077**

Implementing and democratizing personalized campaigns at scale and speed

_Shirvam Lakshminarayan and Saurav Chakravorty_
Brillio

In this paper, we present a solution that enables B2C and B2B companies to use cutting-edge analytical models to achieve hyper-personalized recommendations for their customers. While the need for personalization as a strategy has been largely established, the challenges faced by companies in implementing this vary from case-to-case. This publication covers the broad gamut of challenges faced by, mostly, non-digital native businesses to implement a personalization strategy and how our solution mitigates some of those challenges by adopting a ground-up approach through the concept of offer banks and a highly optimized assembly process flow. The paper will also explain our on-going efforts to create a self-service platform that will enable businesses to truly achieve personalization at scale. In this regard, we will share briefly some of the success stories that we have seen while delivering this solution for one of our customers.

**IC 19/078**

Style Scanner- Personalized Visual Search & Recommendations

_Abhishek Kushwaha Saurav Chakravorty and Paulami Das_
Brillio

In this paper, we present a visual search and recommendation system that supports typical shopping behaviour. We present a unified Convolutional Neural Network architecture, to learn embeddings, which is a way to capture notion of similarity. We will introduce the concept of embeddings with respect to similarity and show how we try to achieve required embeddings with various loss functions. We demonstrate various model architectures based on availability of data. We also demonstrate a semi-automatic way of creating labelled dataset for training. We will talk about the concept of accuracy with respect to similarity, which is complicated as similarity is subjective. Finally, we present an end to end system for deployment.
IC 19/082

Frequency-dependent connectedness of geopolitical risk across BRICS nations

Debdatta Pal
Indian Institute of Management Lucknow

Subrata K. Mitra
Indian Institute of Management Raipur

We explore connectedness of geopolitical risk across BRICS economies in time-frequency sphere. Using frequency-dependent connectedness measure of Barunik and Krehlik (2018) on geopolitical risk indices, we find, in short-term, China is the key transmitter of risk to Brazil, Russia, and India. However, at long-term horizon, geopolitical risk returns are not connected across BRICS nations.

IC 19/086

Connectedness of Markets with Heterogeneous Agents and the Information Cascades

Avijit Ghosh¹, Aditya Chourasiya², Lakshay Bansal³, Abhijeet Chandra⁴
¹²³ Financial Engineering Program, ¹²³⁴ Vinod Gupta School of Management
Indian Institute of Technology Kharagpur, Kharagpur 721 302, India

Macroeconomic integration of global financial markets is often characterized as complex systems where ever-increasing interactions among a vast number of agents make it difficult for the traditional economic theory to provide a realistic approximation of market dynamics. Economic systems are increasingly interdependent through cross-country networks of credit and investment, trade relations, or supply chains, and highlight the need for an integration of network theory and economic models to reduce the risk of global failure of financial systems. Our aim is to study the cross holdings of entities in terms of input-output and look at a time varying feature to examine the changes in the network. We also study the ripple effects caused due to the failure of entities inside the model. Using the WIOD (World Input Output Database) dataset covering 28 countries from the European Union and 15 other major countries (total 43 countries across 56 industries) in the world for the period from 2000 to 2014, we present the evidence on the nature of interconnectedness that global markets exhibit in terms of their input-output representing the cross-holdings. The interdependence of some markets in a global network is strongly correlated with not only the size of the markets, but also the direction of trades/cross-holdings, and the type of industries that dominate in their input-output data. With growth model estimation, we are able to project the cascades of failures in the network significantly. Our findings employ innovative approaches such as network formation approach and graph theory to explain the interconnectedness of markets across the world, and contribute significantly to the theoretical issues related to market integration and risk spill-over.
Dynamic Regression Approach to Identifying the Determinants of Inward Foreign Direct Investment for India

Sharmiladevi J.C
Symbiosis Centre for Management Studies, Symbiosis International (Deemed University), Pune, Maharashtra

One of the remarkable features of globalization in the 1990s for India was the flow of private capital in the form of foreign direct investment (FDI), which is an important source of development finance that contributes to productivity gains by providing new investments, better technology, management expertise, and export markets to Indian economy. The study identifies some of the major macroeconomic factors which act as potential determinants of inward FDI. Variables analyzed includes inward FDI, gross domestic product, exchange rate, exports, gross enrolment ratio, urban population, inflation and one period lagged value of inward FDI. Period of study ranges from 1990-91 to 2013-14. Dynamic Regression Modelling is used in this research to understand the self-reinforcing effects of past values of FDI. Results of the model indicate that urbanization and lagged values of inward FDI are statistically significant in determining inward FDI. The results are statistically significant. It indicates that towns and cities are creating a landmark not only for overall development but also acts as a tool in bringing in more capital from abroad through inward FDI. Also it is proved that, the mechanism of inward FDI is taking into account not only the current values but also the past inflows of FDI which together influences future potential investments.

Best approach direction for spherical random variables

Jayant Jha
NOVARTIS Healthcare, Hyderabad

Spherical random variables have support in subsets of $S^{d-1}$. They have wide application in geology, astronomy, shape analysis and crystallography, among others. In two dimensions, the support of the random variable is on the circumference of a unit circle and it is called a circular or directional random variable. Wind direction, direction of migration of birds and date or time of occurrence of an event are the most common examples.

We attempt the estimation of the directional quantiles for spherical random variables. We first form a notion of directional quantiles in this regard. The consideration of directional quantiles for spherical random variables is very natural as no transformation needs to be done on the raw dataset and hence, it induces important insights in data analysis examples and helps in deriving some properties unique to this particular case.
Based on the directional quantiles, we define the concept of best approach direction. We provide exact polynomial time algorithms for the computation of its estimator on circle and hyperspheres. Naturally, the algorithms employed are also important in fast computation of connected highest sample density region when the observations are circular or spherical in nature. After that, we discuss some properties of the population best approach direction for different family of distributions and discuss the inferential procedures related to its sample variant. Then, we study the asymptotic distribution of the sample best approach direction using the known asymptotic distribution of a directional estimator. Simulation and real data analyses are done to illustrate the results.

IC 19/093

Role of Infrastructure and Taxation on Economic Growth in Middle Income Countries

Rudra P. Pradhan
Indian Institute of Technology Kharagpur, India

“It is shortage of resources, and not inadequate incentives, which limits the pace of economic development. Indeed, the importance of public revenue from the point of view of accelerated economic development could hardly be exaggerated (Kaldor, 1963).”

This study examines the role of infrastructure and taxation on economic growth in middle income countries (MICs). We deploy various cases, depending upon the types of infrastructure and tax intensity to examine this nexus. It addresses two standard questions, “how does infrastructure affect economic growth?”, and “how does tax propensity affect economic growth?”. Using generalized methods of moments (GMM), our answer to these questions are very much certain in the case of infrastructure, as it promotes economic growth positively, and quite uncertain in the case of taxation, as it shows both positive and negative impact on economic growth. Additionally, we also find the impact of other macroeconomic variables, such as population, inflation, human capital, etc, on economic growth while investigating the relationship among infrastructure, taxation and economic growth in MICs.

IC 19/095

Drivers of Governance: An Econometric Analysis

Sakshi Malik and Cheshta Kapuria
Faculty of Management Studies, University of Delhi

Quality of governance has a fundamental bearing on the economic as well as social growth of an economy. Sustainable development is achieved not merely by launching policies, but by ensuring that these policies yield results at grassroots level. Effective governance by institutions and policymakers can lead to efficiencies in implementation of policies at both micro and macro levels. Given that governance
is indispensable to the process of sustainable development, it is crucial to study its determinants. The present paper uses two step system Generalized Method of Moments (GMM) technique on a panel dataset of select developing countries from 2005-2017 and analyses the impact of macroeconomic indicators on the level of governance. The results indicate that infrastructure and urbanization favourably impact the level of governance, while per capita income level has shown to have no significant impact on the country’s level of governance. A pertinent finding of this paper is that globalisation has a positive and significant relationship with the quality of governance in a country. Based on the econometric results, the paper concludes from a broad policy perspective.

IC 19/096

Stochastic comparisons of systems with heterogeneous log-logistic components

Shyamal Ghosh, Priyanka Majumder and Murari Mitra
Department of Mathematics, Indian Institute of Engineering Science and Technology, Shibpur, India

Order statistics occupy a place of remarkable importance in both theory and practice. They play a vital role in statistical inference, reliability theory, life testing, operations research, actuarial science, auction theory, hydrology and many other related areas and have received a lot of attention in the literature during the last few decades. It is a well-known that the k-th order statistic \( X_{k:n} \) corresponds to the lifetime of a \((n-k+1)\)-out-of-\(n\) system, a very popular structure of redundancy in fault-tolerant systems, which have been studied extensively. A series system can be regarded as a \(n\)-out-of-\(n\) system while a parallel system is a \(1\)-out-of-\(n\) system. In this paper we study stochastic comparisons for both parallel and series systems having heterogeneous log-logistic distributed components. The log-logistic distribution is a very flexible family of life distributions which has been applied in a vast number of disciplines. The comparisons are carried out in the sense of stochastic, reversed hazard rate, hazard rate and likelihood ratio orderings which are widely used in the literature for fair and reasonable comparison. The effects of the changes in the scale parameters or the shape parameters on the magnitude of smallest and largest order statistics according to above mentioned orderings are also investigated. We also study the comparison in the case of multiple outlier log-logistic distribution model.
Care2Vec: A Deep learning approach for the classification of self-care problems in physically disabled children

Sayan Putatunda
VMware Software, India

Accurate classification of self-care problems in children who suffer from physical and motor affliction is an important problem in the healthcare industry. This is a difficult and a time consuming process and it needs the expertise of occupational therapists. In recent years, healthcare professionals have opened up to the idea of using expert systems and artificial intelligence in the diagnosis and classification of self-care problems. In this study, we propose a new deep learning based approach named Care2Vec for solving the self-care classification problem. We use a real world self care activities dataset that is based on a conceptual framework designed by the World Health Organization (WHO). The conceptual framework is known as the International Classification of Functioning, Disability, and Health for Children and Youth (ICF-CY), which is a widely used standard framework for analyzing self-care activity records.

Deep learning is a form of representation learning and in recent years, it has been very successful in various fields such as computer vision, speech processing and more. Care2Vec is a mix of unsupervised and supervised learning where we use Autoencoders and Deep neural networks as a two step modeling process. We compare Care2Vec with traditional methods reported in the literature for solving the self-care classification problem (such as, Artificial Neural Networks and Decision trees) in both multi-class classification and binary classification settings. We use k-fold cross validation while applying the different methodologies. The evaluation metrics used were the Mean Cross Validation (CV) score and the Mean Area under the curve (AUC). We found that the Care2Vec method has better prediction accuracy than the prevalent methods and so a recommended approach for the self care classification problem. The adoption of Care2Vec can help expert therapists in making better diagnostic decisions and will thus lead to better treatment.

Predictive likelihood for coherent forecasting of count time series

S. Mukhopadhyay
Department of Mathematics, IIT Bombay

In contrast to time series for Gaussian responses, where numerous forecasting methods are available, literature on forecasting for count type time series is still very sparse. However, time series data in the form of counts is frequently measured in
various fields such as finance, insurance, biomedical, and public health. As an example, consider a disease surveillance study, where health officials record the number of disease cases over a certain time period to understand the disease trajectory. The main interest in such surveillance studies is to forecast disease counts in the future. However, forecasting disease counts in these situations is complex, due to the fact that the required forecasts have to be consistent with the nonnegative and integer-valued sample space of such count time series. Usage of the estimated mean at a future time point which is a noninteger, as a suitable point estimate for the future count as practiced in usual ARIMA forecasting techniques gives rise to an incoherent forecast. In this talk, a new forecasting method based on the concept of the profile predictive likelihood function is proposed for discrete-valued processes. In particular, generalized autoregressive moving average (GARMA) models for Poisson distributed data are explored in detail. Highest density regions are used to construct forecasting regions. The proposed forecast estimates and regions are coherent. Large-sample results are derived for the forecasting distribution. Numerical studies are used to establish the performance of the proposed forecasting method. Robustness of the proposed method to possible misspecifications in the model is also studied.

On Modeling the High-Frequency Time Series Data

Sourish Das
Chennai Mathematical Institute, India

In the era of data-driven science, high-frequency time series data become ubiquitous from the stock market to the power plant. However, no single magic model can be used for all kind of high-frequency time series data. In this talk, we will discuss that the domain demands a mathematical model that fits its set of features and hence we should develop the statistical analysis along the mathematical models that satisfy the domain requirement.

In this presentation, we will present the study of the consumer’s electric power consumption data, where the data observed in the one-minute interval level and objective is to predict the demand at one-minute interval levels for 24 hours. We use Fourier series approach to develop the predictive model. The successful implementation of the Fourier series method depends on the correct selection of the number of harmonics and estimation of corresponding coefficients of each harmonic. In this work, we resort to statistical machine learning technique like the lasso and the elastic net to select the right harmonic for the fitted model. The advantage of the elastic net is it can make the harmonic selection like the lasso; also it can take care of the multicollinearity in the feature space like the ridge regression. We compare the prediction result with some ‘of the self’ machine learning algorithms.

If time permits, we will present another case study of the stock market data observed at the one-minute interval level, we discuss why the above approach of Fourier series will not work, and we implemented risk analysis with fractional Brownian motion with Dirichlet Process prior and we compared the result with other popular methods.
IC 19/100

Coalescence in Branching Processes with age dependent structure in population

Sumit Kumar Yadav and Arnab Kumar Laha
Indian Institute of Management Ahmedabad

Branching process and their variants are a widely used mathematical model in the study of population dynamics, in which all individuals in a given generation produces some random number of individuals for the next generation. In the recent past, branching process has also found applications in areas like operations research, marketing, finance, genetics etc. A problem that has caught attention in the context of coalescence in branching process is as follows: Assume that the branching process is started by one individual in 0-th generation and the population size of the tree obtained by branching process in generation n is greater than 1. Next, pick two individuals from n-th generation at random and trace their lines of descent back till they meet. Call that random generation by $X(n)$. The objective is to study the properties of $X(n)$. While this problem has been studied by many authors for simple and multitype discrete time branching processes, not much attention has been given for the realistic extension when one individual is allowed to survive for more than one generation and can also give birth more than once. We study this problem for some deterministic and random cases. Explicit expressions about some mathematical properties of $X(n)$ have been derived for broad classes of deterministic trees. For random trees, we provide explicit expression for some special cases. We also derive properties of $X(n)$ as n goes to infinity. Additionally, simulation analysis has also been performed and some interesting insights are discussed.

IC 19/101

Gender equality at work: Are we there yet?

Kanish Debnath
FLAME University, Pune

Multiple studies demonstrate that Female Labor Participation Rate (FLPR thereafter) follows a U-shaped relationship with the log of real GDP per capita in PPP terms, and an M-shaped distribution with married women’s age. These distributions are not independent as fertility decisions not only reduces the labor supply for middle-aged married women but also reinforces the U-shaped relationship. If both distributions are considered together, a three-dimensional surface of FLPR can be described, with FLPR falling for middle-income countries and for middle-aged married women. However, with the changing nature of work and improving legal support systems in most countries, FLPR is expected to catch-up to Male Labor Participation Rates (MLPR thereafter). This paper therefore further investigates into the decadal improvements in FLPR surfaces vis-à-vis MLPR surfaces through a non-parametric estimation and comparison between them. The comparisons across six decades, that is to say from 1960 – 2010 show that FLPR has been continually improving against the MLPR, however significant differences still remain.
Predicting Failures Beyond Warranty in Automotive Industry

Suresh R. P.
Supply Chain and Operations Analytics, Accenture Digital, Bangalore

In Automotive industry, it is of interest to predict failures of products and components thereof, so as to help engineers, financial managers and supply chain managers in taking appropriate Warranty decisions (e.g., extension of warranty), in determining warranty reserve fund, in determining stock levels of spare parts, in providing alerts to customers, and also in arriving at bulk replacement or recall decisions. For most components, the manufacturers have the failures data only until the warranty period, and the data that exist beyond warranty period are not very reliable or largely incomplete. Hence the prediction of failures beyond warranty period needs to be based primarily on failure data during warranty period. Some parametric based models (Nordman and Meeker - Technometrics 2002, Ebrahami E – Reliability Engineering and System Safety 2009) exist in the literature, and by estimating the parameters based on the data till the warranty period. However, since typically, 3-4% of items / parts fail (see e.g., Lawless et al, Technometrics 2009) in the warranty period, such parameter estimates may not be reliable.

In this paper, we discuss some approaches to estimate the failures beyond the warranty period using the behavior / properties of the normalized spacings of order statistics observed till the warranty period. We validate the method with specific data sets.

Machine Learning in Retail

Raj Vaidyanathan
Data Science & Data Engineering, Wal-mart Labs, India

Machine learning and AI has made considerable progress in the areas of consumer applications. It has taken more time for it to make an impact in enterprise applications. Reasons have been many, access to clean and large amount of data being a challenge, lack of data governance, lack of academic and industry partnerships and volatility of the industry itself. Having said that, over the last few years, considerable progress has been made and there are various attempts at establishing AI for enterprises. At Walmart, we realize the importance of data and digitization. We have adopted the latest of technologies and innovated to accommodate the explosion of data and its usability. Concepts of data lake along with state of art machine learning techniques have enabled a data-driven business transformation. With the right combination of proprietary tools and open source technology, we at Walmart have developed platforms that has helped democratize the use of data. This has helped in development of various analytical and machine learning applications across different retail functions like sourcing, merchandising, supply chain, etc. These applications lead to faster business decisions which in turn results in better customer experience. This talk would give a sneak peek into some of the work that is going on within Walmart that leverages Machine Learning algorithms to drive business decisions.
How Machines Embed Business Intelligence

Sai Sundarakrishna  
Caterpillar

Learning is characterized by how machines store, shape and process the underlying knowledge representation. A brief introduction to how a machine uses calculus, linear algebra and computational strategies for their learning process in relevance to the business context would be provided. Starting from regularization techniques to tree-based models, how different kind of business data and applications are modeled would be the initial focus. Slowly, as the parameter space explodes, how kernel tricks and other sophisticated deep learning models structure the knowledge and address the domain and problem complexity would be discussed. Finally, how this knowledge can be converted to actionable intelligence in different practical business settings would be the take away conclusion.

Why prices of medical procedures vary across providers

Shreya Surendhran and Bhavana Kamaraj  
National Institute of Technology Tiruchirappalli

Arnab Kumar Laha  
Indian Institute of Management Ahmedabad

Healthcare in India is very dynamic and operates in an environment with rapid social, economic and technical changes. Unlike many developed countries, most of the expenses incurred by the patients are out-of-their own pockets. Lack of price transparency in Indian healthcare sector is one of the biggest difficulties faced by the patients today and is a stumbling block for new welfare schemes such as Ayushman Bharat. Price transparency coupled with accurate effectiveness indicators of providers and availability of reasonably priced health insurance options could lead to a greater patient satisfaction and improve overall efficiency in the healthcare sector. Working with data on prices of 14 common medical procedures announced by the healthcare providers located across 10 cities of India on their websites, healthcare provider specific attributes and city characteristics, this paper examines the variation in prices of these procedures. A multiple linear regression analysis is carried out to identify the key variables that explain the variation in prices of the procedures. It is found that the category of hospital (i.e. public or private), cost of living index of a city and the accreditation status of the healthcare provider (i.e. whether the hospital is accredited by an agency such as NABH) explains the variation in the price of medical procedures substantially.
Land transitions and climate change impacts on urban flooding in 42 cities across India

Vidhee Avashia and Amit Garg
Indian Institute of Management Ahmedabad

Urban development induces land transitions impact urban hydrology resulting in increased flooding risks. Climate change related precipitation changes are an added complexity to the flood risks that the cities face. This study looks into role of land use changes in urban flooding across 42 Indian cities. Landsat images for 1990, 2000, 2010 and 2017 have been processed using hybrid classification technique to arrive at the land use shares for all cities. A typical event count study using newspaper archives has been conducted to create a flooding event database. A logistic model approach considering time lag been used here to establish the relation between rainfall amount, share of built-up spaces and occurrence of flooding event. Future projections of occurrence of flooding events under three climate change scenarios- RCPs 2.6, 4.5 and 8.5, and three urban development scenarios have been carried out. The results suggest that cities need to preserve the land uses that act as a sponge – the green and blue spaces. As these spaces decrease, the projected flooding events increase. Under RCP 2.6 scenario, number of flooding events are significantly lower (95% confidence) than under RCPs 4.5 and 8.5. The expected flooding occurrences between RCP4.5 and RCP8.5 are not significantly different (95% confidence) for many scenarios suggesting that cities need to aim for below 2°C world or would have to face devastating consequences in future. This study highlights the need for Indian cities to undertake integrated spatial planning measures for resilient and sustainable urban future.

A Perfect Mixture Model for Credit Risk and Actuarial Applications

Fareed Thayil (TPM)
WMO Arts & Science College Muttil, Wayanad, Kerala

Credit risk models are used by financial companies to evaluate in advance the insolvency risk caused by credits that enter into default. Many models for credit risk have been developed over the past few decades. In this paper, we focus on those models that can be formulated in terms of the probability of default and survival of an account by using survival analysis techniques. In the conventional survival analysis, we have one event of interest and we study the life time in the presence of censoring. In the case of credit risk, we consider two events, default and payoff. An account originated in a date, may default or payoff in a future date. In other words, an account will close either as a defaulted account or closed good in a later date. We introduce a perfect mixture model, a general model, which can be used for estimating probability of payment / default, to estimate conditional default and conditional payoff survival
functions. We discuss the application of the model in different areas of credit risk assessment and also possible actuarial applications.

IC 19/111

SB-robustness of Estimators of Directional Correlation Coefficient

Arnab Kumar Laha and Pravida Raja A.C.
Indian Institute of Management Ahmedabad

Bivariate circular data arises in many real-life situations and it is often of interest to understand the interrelationship between the two directional variates. The joint distribution of two circular random variables results in a distribution defined on the torus where T is the unit circle. Unlike in the linear case there is no well-accepted definition of correlation among two circular variates. In this paper we discuss the robustness and SB-robustness of estimates of the bivariate circular-circular correlation functional introduced by Jammalamadaka and Sarma (1988) (to be called JS-correlation). We work with the bivariate distribution on the torus introduced by Shieh and Johnson (2005) which has von-Mises marginal as the underlying distribution. We derive the influence function of the JS-correlation functional and study the Bias Robustness (B-robustness) and Standardized Bias Robustness (SB-robustness) through the Influence Function approach with the dispersion measure being the Clifford Torus distance. It is found that the JS-correlation is B-Robust but is not SB-Robust at the entire Shieh-Johnson family of distributions with respect to Clifford Torus distance dispersion measure. We then identify a subfamily of Shieh-Johnson distributions such that the JS-correlation is SB-robust at this subfamily.

References:


IC 19/112

A Robust Test for the Spherical Mean Direction

Mahesh K.C
Institute of Management, Nirma University

Arnab Kumar Laha
Indian Institute of Management Ahmedabad

The spherical data arise in many areas of scientific experimentation in which observations are made on directions in three dimensions. For example, the arrival directions of showers of cosmic rays, measurements of magnetic remanence in rocks, observed wind directions at a given place, measurements of ocean current directions etc. The von-Mises-Fisher distribution is the most widely used distribution for modelling spherical data. Several tests for the mean direction of the von-Mises-Fisher distribution have been developed in the literature but the robustness of these tests has not been explored in detail. We study the robustness of three single sample tests using different test functionals namely- the likelihood ratio test functional, the spherical mean test functional and the $\gamma$-spherical trimmed mean test functional for the spherical mean direction of von-Mises-Fisher distribution based on the approach of He, Simpson and Portnoy.

IC 19/113

Impact of Artificial Intelligence on Indian IT / BPM Employment

Ronit Bhattacharya
Indian Institute of Management Ahmedabad

Purpose: This paper explores the risks and benefits of Artificial Intelligence (AI) from an Indian perspective, focusing on its effects on employment in the Indian IT/BPM industry. We sketch out the scale, scope and timing of the likely impact over the next 5 years, and recommend needed actions by key stakeholders.

Methods: A review of existing literature, expert views, available primary and secondary data provides a range of estimates for the speed to maturity of AI, the likely demand/supply of talent for AI jobs in India, and the likely job losses from AI-based automation. This data is used to develop 3 core uncertainties - AI speed to maturity, AI job creation, and AI job destruction. For job creation, there is uncertainty in terms of both demand for AI talent as well as supply of AI talent. For job destruction, there is uncertainty on the % of jobs in the IT-BPM sector that are susceptible to automation, and the degree to which these jobs are likely to be automated. Combining these uncertainties, 4 “most plausible” scenarios are constructed, ranging from “AI Skeptic” to “Worst Case”.

Results: In the most likely scenarios, AI is estimated to cause net direct job losses ranging from 140k to 700k jobs.

Conclusions: Policy responses to AI must adopt a dual pronged approach in order to both encourage the growth of AI talent as well as mitigate the harm to laid off employees. Aggressive re-skilling will be key to achieving both objectives.
We attempt to theoretically model the interplay between economic backwardness and the quality of democracy in a poor country. Large informal sector and high political violence are two defining features of less developed economies like India. We focus on two possible channels through which economic backwardness affects the quality of democracy - political violence and the informal sector. We start with the research question - is there any causal link? Our present model follows that framework of clientelist politics using global game. We illustrate how contagious voting behaviour of the informal sector affects political violence. Our study attempts to explain how political violence within a democratic framework may persist in a developing economy. We find that a growing informal sector may be detrimental for development in a poorly-institutionalized society. We also explain the puzzle that why well-performing incumbents engage in violence even in a democracy.
Announcement

Selected papers of ICADABAI-2019 will be published by Springer as an edited volume after peer review.
ICADABAI HACKATHON-2019

in collaboration with

IBM®