IIMA

विद्याविनियोगाद्विकास:

Brij Disa Centre for Data Science and Artificial Intelligence

INDIAN INSTITUTE of MANAGEMENT AHMEDABAD

Annual Report

2022-2023

IIMA

Brij Disa Centre for Data Science and Artificial Intelligence

INDIAN INSTITUTE of MANAGEMENT AHMEDABAD

विद्याविनियोगाद्विकास:

CONTENTS

Brij Disa Centre for Data Science and Artificial Intelligence
IIMA INDIAN INSTITUTE of MANAGEMENT AHMEDABAD
विद्याविनियोगाद्विकास:

FOREWORD

As we stand at the cusp of the Artificial Intelligence (AI) era, it is worthwhile investigating what would be the key enablers in India's AI journey and why every sector shall need to navigate a unique path and rely on a different set of growth levers. Before addressing this, evaluating the nation's readiness for AI adoption is essential. Additionally, is readiness enough to spur growth and operationalize AI for wider adoption? Organizations around the world stand at Big Data 1.0, gathering, organizing, and analyzing large data sets at scale to augment their current operations. However, leading organizations that had invested in ramping up their data and digital capabilities over the last decade have now begun to reap benefits in the form of better understanding of their customers, suppliers, and businesses. The investment in AI, if done well, would possibly be one of very high returns.

In 2022, the Centre has taken up India's AI Readiness study in collaboration with Boston Consulting Group, India. A detailed, structured survey and discussions were conducted with CXO level leaders in Technology, Data analytics, Digital transformation, and Business Heads of 130 organizations spread across BFSI, Consumer Goods (CG) and Industrial Goods (IG). The study captures the journey of organizations from pre-covid years to the present as well as gauges their future aspirations. The findings of this study will be published as a separate report.

ChatGPT, though just a large language model, should be termed as an inflection point in AI research as it is going to have a significant impact on several businesses and would lead to applications that will bring multiple surprises to us within the next 5 years. The new AI technology will disrupt the labor market. For a very long time, we saw mechanization and automation of manual tasks. The current era of digitization has the potential of automating jobs which are of cognitive nature. The effect on the labor market will be large, especially in the context of the developing countries which often do not have economic safety nets. As with most disruptive technologies, certain job roles will be made redundant while a completely different set would come to the fore. This is skill polarization. While technology has been impacting white collar jobs for quite some time, blue collar jobs are being impacted considerably for the first time. What about the high skill jobs that involve complex analytical skills as well as soft skills. Many functions within such a role would be taken up by these tools. A recent paper says that 1% increase in the AI vacancy growth rate results in a 3.61% decrease [1]in establishment non- AI vacancy growth. This holds especially true when such technologies are deployed towardsinternal efficiency and automation.

Data, methods, and computation are three key ingredients to training an AI model. Most of the large language models are trained on publicly available massive text data using deep learning technologies that are common knowledge. However, the barrier to training such models is access to compute resources. Therefore, such high-performance large language models would remain proprietary with large corporations in the short run. These corporations are likely to charge users based on subscription or follow a freemium model. Instead of regulations, the need of the hour is to fuel research through public and private funding to democratize AI so that these models are available to the users across the world for free and in the language of their choice.

The Centre, with Prof Aditya Moses as Principal Investigator, is conducting in-depth research on the Future of Work. Through this initiative, the centre hopes to invite researchers to work under one umbrella and leverage on resources to grow this initiative as India navigates this era of dynamic jobs and skills.

So, where will jobs be created? There are 2 key areas where jobs are created when new technologies disrupt. One, in the disruption itself. Skilled professionals shall continue to build advanced versions and enhancements. Second, organizations would embrace this wave of technology and its variants and increasingly build products / services around these technologies. For instance, Large language models (LLM) such as ChatGPT have brought Prompt Engineering to the forefront.

[1] Stapleton, Katherine and Copestake, Alexander and Pople, Ashley, AI, firms and wages: Evidence from India (November 6, 2021). Available at SSRN: https://ssrn.com/abstract=3957858 or http://dx.doi.org/10.2139/ssrn.3957858

In 2022, the Large-Scale Optimization summer school was conducted at IIM Ahmedabad with Prof Sachin Jayaswal as the Coordinator. The Centre continues to support this initiative that brings together faculty from across top universities, researchers, and practitioners under one roof every year. (Please see Section: Large Scale Optimization summer school for more details).

Over the last year, the Centre has invited distinguished faculty and researchers from around the world and across management areas to showcase their latest research. Through this initiative, we hope to bring the latest research so that it would benefit the Indian research community and serves to encourage innovation and adoption of new data science use cases in the industry. (Please see Section: Seminars and Webinars for more details)

There is immense potential for exponential returns through the adoption of Data Science and AI in the social sector. The Centre in its pilot initiative, with Prof Ambrish Dongre as Principal Investigator, is working with Saajha to improve learning outcomes. We hope this initiative will allow us to contribute towards the National Education Policy (NEP) 2020 which has explicitly mentioned achieving Foundational Literacy and Numeracy (FLN) for all students at the end of Grade III by 2025. (Please see Section: Data driven research in the Social Sector).

We are aware how touch screen smartphones have impacted various businesses and our lives, and people call it disruptive. The question is, "Which launch was more exciting and surprising, that of a touch screen smartphone, or a conversational AI like ChatGPT?" If you think it is ChatGPT, then imagine what the upcoming plethora of use cases based on advanced data techniques have on offer for the future.

Building on our existing research projects, the Center will continue in its path towards a better understanding of the new emerging society and economy in the presence of an ever-changing business landscape with the advent of new disruptive technologies.

Brij Disa Centre for Data Science and Artificial Intelligence
INDIAN INSTITUTE of MANAGEMENT AHMEDABAD
विद्याविनियोगाद्विकास:

About Us

The Brij Disa Centre for Data Science and Artificial Intelligence (CDSA) at the Indian Institute of Management Ahmedabad (IIMA) provides a common platform to faculty, scholars, and practitioners for conducting and disseminating cutting-edge research on data analytics and artificial intelligence that offers solutions applicable to business, governance, and policy.

Besides generating action-oriented insights, CDSA is also responsible for dissemination of the knowledge generated to a wider audience both within and outside the realm of the Institute. Seminars, workshops, and conferences are regular activities at the Centre, which are conducted to reach out to and engage with stakeholders.

The Centre aims to forge synergistic and collaborative relationships between scholars and practitioners in data-intensive organizations, besides undertaking case-based research to understand the current industry practice and develop case studies for classroom teaching.

Furthermore, through its collaboration with the industry, CDSA takes up challenging consulting projects of considerable practical importance. These projects are targeted at providing an opportunity for students to participate in projects that aim at outcomes that can further benefit the organization and the business, at large.

A key offering from the Centre is the Annual Report, which would provide a holistic view of the Data Science and Artificial Intelligence industry, identify challenges and gaps, gauge scope of the industry and offer plausible solutions that can be utilized by the industry and policy makers.

Brij Disa Centre for Data Science and Artificial Intelligence
INDIAN INSTITUTE of MANAGEMENT AHMEDABAD
विद्याविनियोगाद्विकास:

What We Do

The activities of the centre are five-fold through which the members of the centre engage with businesses, policy makers, students, and other academic institutions.

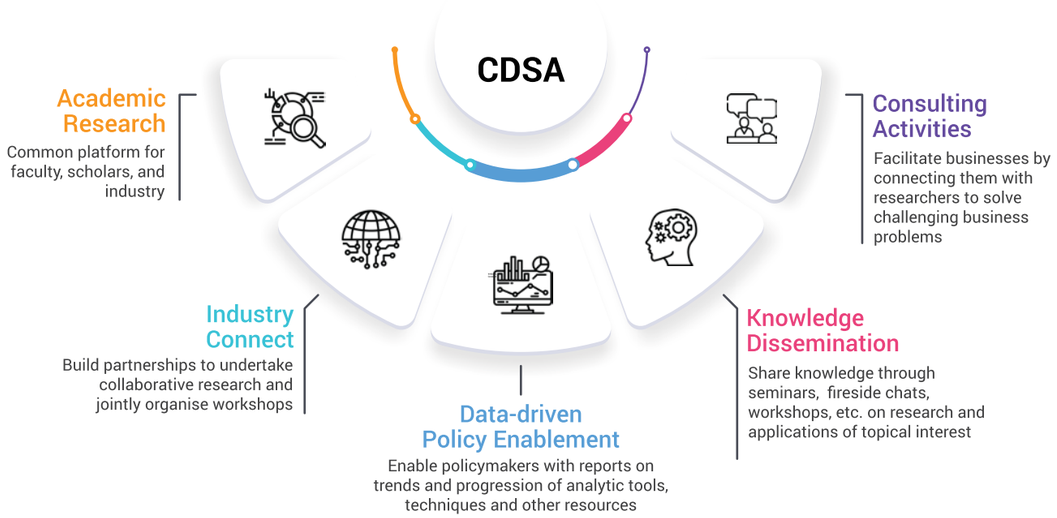
Academic Research
The centre supports and fosters academic research in the field of data science and AI.

Industry Connect
The centre connects with the industry by collaborating with them in conducting workshops and seminars for the audience that is keen to track the progress of data science and AI in practice.

Data-driven Policy Enablement
The centre enables the policy makes and the industries with reports on the trends and the progression of the analytics tools, techniques, market, opportunities, human capital, and other resources, within the country and worldwide.

Knowledge Dissemination
The centre disseminates research reports/findings to key stake holders including academics through a variety of channels viz., media outreach, participation in academic and professional conferences, authored articles, workshops, etc.

Consulting Activities
The centre facilitates businesses by connecting them with researchers who are working on cutting-edge problems and can help solve challenging business problems.



Brij Disa Centre for Data Science and Artificial Intelligence
INDIAN INSTITUTE of MANAGEMENT AHMEDABAD
विद्याविनियोगाद्विकासः

India AI Readiness Study

(In collaboration with Boston Consulting Group, India)

AI is altering the structure of jobs, replacing human performed tasks while also generating new tasks. However, evidence suggests that positive and complementarity effects of adoption of AI are still small relative to its displacement effects (Acemoglu et al., 2022).

Brij Disa Centre in association with BCG have explored Data and AI readiness of organization over the last year. The data collection involved in-depth interviews with leading Chief Data and Analytics Officers across industries as well as a survey specifically targeting data and data-associated functions in the organization.
The goal is to measure the AI readiness of key Indian Companies: Companies from Banking, Financial Services, and Insurance (BFSI), Consumer Goods & Retail and Industrial. Approximately 60% of AI-led value-added will be from these industries.

Prof. Anindya S Chakrabarti presented the initial findings of this study as part of discussion on 'AI, Policy Issues & Technological Revolutions' at University of Sussex Business School.

What does the management think? Some initial insights.

BFSI
• "The focus is on whether you can get a customer without a salesperson being present."
• "The cost of discovery is quite high to find and engage startups with the right talent to work and co- innovate."
• "Focus is on providing a seamless digital experience across products and minimize disruptions."

Consumer Goods & Retail
• "In India, the challenge is with respect to the cleanliness of the data. Many government organizations do not have a unified database."
• "Recruitment to build the right team size is a challenge. Finding the right talent is difficult." -

Expectations for the future of India's skilled labor population:
• Expect an increase in hiring because of 5G expansion, automation, and digitization.
• India needs 30 million digitally skilled professionals by 2026 and 50% of the current workforce will have to re-skill in emerging technologies.
• 25-30% increase in demand for tech-talent in non-tech industries
• Easier to absorb engineering (AI/ML/Full stack Engineers) than non-engineering talent (Corporate and Operational roles)

Detailed findings will be published in the form of a report.

Brij Disa Centre for Data Science and Artificial Intelligence
INDIAN INSTITUTE of MANAGEMENT AHMEDABAD
विद्याविनियोगाद्विकासः

Large Scale Optimization Summer School

Prof. Sachin Jayaswal
Coordinator LSO

The Brij Disa Centre for Data Science and Artificial Intelligence has successfully organized the second edition of the Large-Scale Optimization (LSO) Summer School at the Indian Institute of Management Ahmedabad, which commenced from 06 May 2022 to 13th May 2022. The previous edition of the summer school, conducted at the Indian Institute of Management Indore in 2019, received an encouraging response from the academic and business communities.

The colloquium serves as an informal forum for researchers and practitioners to share their ideas on optimization, aiming to establish new collaborations or strengthen existing ones and provide Ph.D. students with exposure to the international community. The primary objective is to emphasize recent developments in theory, computation, and applications in these fields. The workshop incorporates a tutorial on each LSO technique, along with research presentations on problems that successfully apply these methods.

Target Audience
• Masters/PhD students in Operations Research/ Management Science/ Industrial engineering
• Faculty Members working with Integer Programmes
• Industry Professionals in Optimization/Logistics/Supply Chain Domain

Faculty:
· Prof. Amit Kumar Vatsa, Indian Institute of Management Indore
· Prof. Ankur Sinha, Indian Institute of Management Ahmedabad
· Prof. Ashutosh Mahajan, Indian Institute of Technology Bombay
· Prof. Faiz Hamid, Indian Institute of Technology Kanpur
· Prof. Goutam Dutta, Indian Institute of Management Ahmedabad
· Prof. Jyotirmoy Dalal, Indian Institute of Management Lucknow
· Prof. Sachin Jayaswal, Indian Institute of Management Ahmedabad (Co-ordinator)
· Prof. Saurabh Chandra, Indian Institute of Management Indore
· Prof. Yogesh Kumar Agarwal, Jaipuria Institute of Management (Retd., IIM Lucknow)

Here are some details about the speakers and the concepts that were covered during the summer school:

Prof. Yogesh Agarwal

Prof. Yogesh Agarwal is a Distinguished Professor at Jaipuria Institute of Management with expertise in Operations Research and Quantitative Techniques. His research interests include integer programming, combinatorial optimization, and telecommunications network optimization.

During summer school, Professor Yogesh Agarwal gave a lecture on integer programming, where he delved into the differences between IP and LP problems. He explained how cutting planes and valid inequalities can be used to solve IP problems in branch and bound settings, covering the Gomory cutting planes, Knapsack, and Extended cover inequalities. Additionally, he highlighted the importance of obtaining the most violated cover inequalities by solving the separation problems.

Prof. Sachin Jayaswal

Prof. Sachin Jayaswal is a Professor of Operations and Decision Sciences at the Indian Institute of Management Ahmedabad. His research interests primarily lie in areas such as facility location, stochastic optimization, and supply chain management.

During summer school, Professor Jayaswal presented a detailed discussion on modeling Mixed Integer Programming (MIP) problems. He explained the application of Bender's decomposition in detail with examples. The discussion further elaborated on the Lagrangian relaxation technique for solving integer and mixed-integer problems. He also covered Dantzig-Wolfe decomposition in detail, a popular technique for solving large linear programs.

Prof. Ankur Sinha

Prof. Ankur Sinha is an Associate Professor of Operations and Decision Sciences at the Indian Institute of Management Ahmedabad. His research interests primarily lie in business intelligence, optimization, and big data analytics.

Prof. Sinha took a lecture on non-linear optimization, where he covered various methods to deal with non-linear problems. He covered Newton and Quasi-Newton methods for unconstrained optimization problems. Further, he elaborated on constrained optimization methods such Frank-Wolfe algorithm and Sequential quadratic programming.

Prof. Sinha also gave a lecture on multi-level optimization, focusing on bilevel optimization problems. He covered various solution approaches for multilevel optimization problems.

Brij Disa Centre for Data Science and Artificial Intelligence
INDIAN INSTITUTE of MANAGEMENT AHMEDABAD
विद्याविनियोगाद्विकास:

Prof. Ashutosh Mahajan

Prof Ashutosh Mahajan is an Associate Professor in the Department of Industrial Engineering and Operations Research at IIT Bombay. His research interests primarily lie in Operations Research, focusing on developing algorithms, software, and theory for solving optimization problems involving integer variables and discrete choices.

During summer school, Prof Mahajan delivered a lecture on mixed integer nonlinear programs (MINLP). He elaborated on the methods to deal with non-linearity in the problems and presented a functional approach to solving convex MINLPs using linearization techniques and spatial branching.

Prof. Sourabh Chandra

Prof. Saurabh Chandra is an Associate Professor at IIM Indore in the Department of Operations Management and Quantitative Techniques. His current research interests include maritime logistics, mathematical optimization, supply chain management, and automotive logistics in India.

Professor Chandra conducted a lecture on the use of the Branch and Cut method for solving the Inventory Routing Problem (IRP). The talk aimed to explain the mathematical formulation of the IRP problem using DFJ Constraints, sub-tour elimination (STE) constraints, and other relevant constraints. He also covered metaheuristics, strategies that guide the search process of an optimization problem.

Prof. Amit Kumar Vatsa

Prof. Amit Kumar Vatsa is an Assistant Professor at IIM Indore in the Department of Operations Management and Quantitative Techniques. His research interests include Operations research, Large-scale Optimization, Facility Location, and Logistics.

Prof. Amit took sessions on modeling integer and mixed-integer problems and Benders partitioning along with Prof. Sachin Jayaswal. Also, Prof Amit covered the facility location problem in order to explain generalized benders decomposition and outer approximation.

LSO CLASS OF 2022

Data-driven Research in Social Sector

Brij Disa Centre for Data Science and AI is excited to collaborate with Saajha on enhancing parental engagement in student learning. Saajha is catalyzing India's public education reforms through parent and community leadership. Saajha, since its inception in 2014, has worked across 2000+ schools in Delhi, Maharashtra, Jharkhand and Karnataka. Training, personalized support, policy reforms, adaptive technology have been the strategic drivers for Saajha to enable parental participation.

The centre, collaborating with Professor Ambrish Dongre, will support this initiative by guiding Saajha through data-led insights. Through this initiative, the centre hopes to bring more awareness in India's education sector on "Data Science use-cases" in maximizing outcomes.

Prof. Ambrish Dongre

Low learning levels is an important challenge for educators and policymakers in India. The National Education Policy (NEP) 2020 has explicitly mentioned achieving Foundational Literacy and Numeracy (FLN) for all students at the end of Grade III by 2025 as an important objective. Parental role in helping their child improve learning has received relatively less attention in the Indian context. Our collaboration with Saajha could throw some light on ways in which parents could be helped better so that their engagement with child's learning and child's school improves.

Saajha, through a well-orchestrated digital push, has modernized its systems in 2022 and brought over 80,000 parents to their platform. With their app, they are now able to manage their learning programs and interventions more effectively.

Saajha

Data Science Collaboration in the Power Sector

On 7th July 2022, DVC stepped into its 75th year of existence. As one of the oldest diversified power organizations of the country, DVC has drawn up an expansion plan till 2030. In line with the plans of the Government of India, DVC has taken up an ambitious target of achieving 15000 MW of installed capacity by 2030 where about 50% will come from renewable energy sources. The expansion program includes solar plants, pumped hydro projects and brownfield thermal units. Diversification into new businesses such as tourism, EV infrastructure and green hydrogen are being planned. As a fully integrated and diversified power organization, DVC will continue to fulfil its mandate in the years to come.

Over the last year, the Centre has worked with the DVC management to understand their challenges across the value chain. The centre has embarked on a 1-year collaboration plan to understand and make recommendation for data science use cases in the Power sector. During a joint meeting on 15th February at DVC HQ at Kolkata, Shri Ram Naresh Singh, Chairman DVC shared insights into the Power sector, operational challenges and expressed confidence in data science driven initiatives to improve operational performance. Prof Ankur Sinha shared insights on data science adoption across industries and sectors. Working together, the goal is to maximize operational efficiency to deliver value to end consumers.

Future of Work

(In collaboration with Right Management India)

Prof. Aditya Moses

The world of work and workplaces has undergone immense change in the last few years. The COVID-19 pandemic made work from home the new normal. However, the work from home benefit is not uniformly distributed across countries and occupations. On an average, one in five jobs across the globe can be performed from home; however, this number masks enormous heterogeneity across countries because the ability to telework is correlated with income. In high-income countries one in every three jobs is suitable for home-based work, while in low-income countries only one in every 26 jobs can be done at home . Furthermore, technology has disrupted the way we work . A large amount of work is now done digitally or through algorithmic control . This has put pressure on employees to reskill themselves. Similar to the work from home concept the most vulnerable jobs to technology are the low income jobs . However, what organizations are doing or not doing to build capacity, reskill employees and alternative work arrangements has not been explored in detail. Our central question is to understand how work is changing and what organizations are doing. In order to explore we focus on traditional industries, since these industries have had to play catch-up with the new disruptive start-ups. Second, we focus on only the sales and marketing function. Sales and marketing was considered to be a field job. This meant the individuals in these roles had to meet potential customers; however, technology such as AI and data analytics have changes the role across industries .

The Centre with Prof Aditya Moses as Principal Investigator is exploring organizational readiness in the context of Future of Work. The outcome is a framework that would help organizations become future-ready. Our framework suggests that in order to become future ready organizations need to focus on skill-based hiring, re-skilling employees systematically and providing employees with a career crafting opportunity. Furthermore, we argue that organizational culture will influence the effectiveness of these practices on future readiness.

Data was gathered for this study through in-depth interviews, surveys and from public sources.

Interviews with 17 Chief Human Resource Officer's (CHRO's)	Survey of Professionals in Customer facing roles

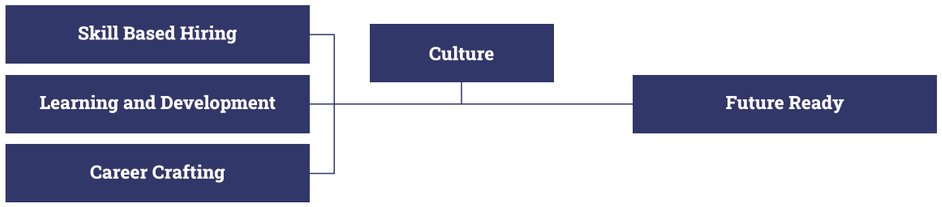


Figure : Framework for Organizations to become Future-ready.

[2] Press Release, Titan Company. https://www.titancompany.in/media/press-release/store-opening-closing

[3] Titan Annual Report 2020-21, https://www.titancompany.in/sites/default/files/Annual%20Report%202020-21_0.pdf

Brij Disa Centre for Data Science and Artificial Intelligence
INDIAN INSTITUTE of MANAGEMENT AHMEDABAD
विद्याविनियोगाद्विकासः

Case Study: Titan's Analytics Journey

Arindam Banerjee

Sourav Borah

The Centre is happy to share the completion of case study capturing the analytics journey of Titan. This initiative was led by Prof Arindam Banerjee and Prof Sourav Borah.

As of November 2022, Titan is among the few companies that had generated considerable wealth for shareholders. A company whose corporate history could be traced back to the Tata Group and the Tamil Nadu Industrial Development Corporation (TIDCO) had become one of India's largest lifestyle companies. Its presence spanned multiple categories such as jewellery, watches, fragrances, eyewear and precision engineering. With a market capitalisation of more than USD 20 billion, Titan featured in the NIFTY 50 index, making it one of the most valuable Indian companies and the second most valuable company within the Tata Group of companies. Titan has more than 2,000 stores, more than 2.5 million square feet of retail space and 7,000 multi-brand watch outlets; its retail footprint spanned 26 Indian states and 3 union territories,. Titan owns some of India's most admired brands—Titan, Tanishq, Fastrack, Titan EyePlus, Taneira, Raga and Edge—and was a leader (with regard to market share) in the Watches, Jewellery and Eyewear segments.

Ongoing Research Projects

An iterative gradient-based bilevel approach for hyperparameter tuning in machine learning

Prof. Ankur Sinha

Hyperparameter tuning in the area of machine learning is often achieved using naïve techniques, such as random search and grid search that only lead to an approximate set of hyperparameters. Although techniques such as Bayesian optimization perform an intelligent search on the domain of hyperparameters, it does not guarantee an optimal solution. A major drawback of most of these approaches is that as the number of hyperparameters increases, the search domain increases exponentially, thereby increasing the computational cost and making the approaches slow. The hyperparameter optimization problem is inherently a bilevel optimization task, and there exist studies that have attempted bilevel solution methodologies for solving this problem. These techniques often assume a unique set of weights that minimizes the loss on the training set. Such an assumption is violated by deep learning architectures. Our study is on gradient-based bilevel optimization method for solving the hyperparameter optimization problem. The method is general and can be easily applied to any class of machine learning algorithms that involve continuous hyperparameters.

Hiring for the Future – A People Analytics Approach

Prof. Aditya Moses

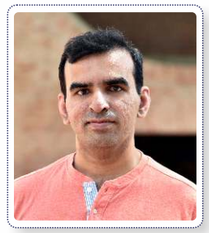
The future of work is a critical aspect for many organizations. A 2020 report by the World Economic Forum suggests that among the various challenges faced by organization one of the most critical areas is skill gaps. The top skills and skill groups which employers see as rising in prominence in the lead up to 2025 include groups such as critical thinking and analysis as well as problem-solving, and skills in self-management such as active learning, resilience, stress tolerance and flexibility. On average, companies estimate that around 40% of workers will require reskilling of six months or less and 94% of business leaders report that they expect employees to pick up new skills on the job, a sharp uptake from 65% in 2018. The changing nature of work and the exponential technology development imply that employees need to constantly re-skill and up-skill. In the current environment, while knowledge can be accessed via multiple sources the behaviours to develop oneself become more important. What behaviours will organizations require for ensuring they have a workforce that can reskill and upskill exponentially? This will be the primary area of research for this study.

Using a data-driven approach, this study uses surveys and NLP to understand which behavioural traits enable re-skilling at pace. We will employ text-mining methods and techniques to identify behavioural traits that help in re-skilling. The insights from this will be further validated and tested using a survey instrument administered to a large sample of individuals.

Brij Disa Centre for Data Science and Artificial Intelligence
INDIAN INSTITUTE of MANAGEMENT AHMEDABAD
विद्याविनियोगाद्विकास:

Can an AI Coach Help You Lose More Weight Than a Human Coach: Empirical Evidence From a Mobile Fitness Tracking App

Prof. Anuj Kapoor



Artificial intelligence(AI) assisted tools are increasingly being used in health care contexts to provide advice and motivation. But whether AI can be a good or even better substitute for human involvement in these contexts is an open question. We provide empirical evidence to answer this question specifically in the context of fitness tracking mobile applications (apps). In addition to facilitating the tracking of activity and food intake, such apps provide advice and motivation in the form of targeted messages to their consumers, and this can be done through human coaches or an AI coach. An AI coach allows these apps to scale their offerings to a larger number of consumers, available on demand to consumers, and potentially more finely targeted by leveraging vast amounts of data. On the other hand, human coaches might be better placed to show empathy, and consumers might also feel more accountable to humans. We compare human and AI coaches on their effectiveness in helping consumers achieve their weight-loss goals. Our empirical analysis is in the context of a large-scale mobile app that offers consumers different levels of subscription plans with human and AI coaches respectively, and specifically compares adopters of the two kinds of plans on their weight loss and goal achievement. We address the potential self-selection in plans by employing a matching-based approach. We find, for our sample of almost 65000 consumers that human-based plans do better than those in AI-based plans in helping them achieve their goals, but that this differs by consumer characteristics including age, gender and body mass index (BMI).

High-frequency trading: Measuring latency from big data

Prof. Anirban Banerjee

Over the last decade, the Indian market has seen significant growth in algorithmic trading and more specifically, high-frequency trading (HFT) activity. During this period, we have witnessed a significant change in the trading landscape as presently close to half of the trading volume in the stock exchanges is contributed by algorithms. This rise has not always been smooth as there have been calls for regulations to restrict algorithmic trading activity due to the fear of probable market manipulation. Latency is considered one of the most important market parameters for HFTs. Using a large novel dataset of order and trade level data from the NSE, we would like to inspect how the latency in the Indian market has changed and if that has caused any shift in the way HFTs operate. We would also like to observe how the different market quality parameters have evolved over this time.

Causes, Symptoms and Consequences of Sociocultural polarization

Prof. Samrat Gupta

The Information and Communication Technology (ICT) provides users unparalleled access to information from around the globe. In spite of demographic differences, people can communicate, express and evolve their opinions on topics ranging from politics to culture. The wide-ranging information exchange on digital media can lead to two scenarios viz. formation of public sphere or formation of echo chambers. While the public sphere, which promotes greater diversity, is a well-researched domain, substantially less research has been conducted on echo chambers in relation to socio-cultural events. Despite the huge affirmative impact of socio-cultural events on society, the proliferation of controversies around them and reinforcement through echo chambers is collectively having malefic effects on societies. Recent controversies around socio-cultural products such as movies, painting, books, cartoons, etc. resulted in serious outcomes. For example, Indian movie Padmavat brought polarization of public perception which further reinforced through echo chambers and escalated into widespread agitations. It led to mass destruction of property and human suffering during agitation. We believe this represents a mounting problem for society, one that is likely to intensify in the era of social media. Thus, understanding the causes, symptoms and consequences of socio-cultural polarization is critical and would be valuable for developing interventions to reduce unhealthy societal and organizational polarisations.

Employee Reviews - A Text Mining Perspective

Prof. Adrija Majumdar

With the emergence of web 2.0, there is a deluge of online text. Technologies like online communities, social media, crowd funding platforms have further contributed to the volume of content. From the firm's perspective, understanding consumers' sentiment from the text is of supreme importance. The literature on online reviews has predominantly focused on ascertaining consumer sentiment of a firm's products and services. We extend this stream of research and focus on analyzing reviews that employees post regarding their organizations. The study will seek to identify different dimensions that employees highlight in their reviews and their association with overall job satisfaction. We further wish to understand if employees' perception of the firms also impacts the firm's performance. The unstructured and noisy nature of the text data often poses significant challenges for organizations in leveraging them for decision making. We will employ text mining methods and techniques to quantitatively analyse the large dataset of employee reviews. The research will have implications for both theory and practice.

Brij Disa Centre for Data Science and Artificial Intelligence
IIMA INDIAN INSTITUTE of MANAGEMENT AHMEDABAD
विद्याविनियोगाद्विकास:

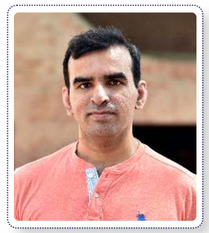
Financial networks from big data: A multivariate time series based approach

Prof. Anindya S. Chakrabarti

Financial markets exhibit non-trivial comovement and dependency structure. The standard approach in the finance literature is to consider the market in Its aggregate form. A more recent 'data'-oriented approach emphasizes a more granular decomposition of the market so that the aggregate dynamics can be broken down into contributions arising from individual assets. This leads to two analytical problems. First, one has to necessarily deal with a large amount of data such that the process scales with the volume of data (large N and large T where T>>N). Two, analyzing such a large volume of data requires toolkits which are at the intersection of econometrics and machine learning. In this project, the goal is to construct large scale financial networks based on multivariate time series data to capture the dynamics of the system. The main idea is to provide an algorithmic approach to convert time series into networks such that the properties of time series are also inherited by the resulting network. The spectral structure of the comovement network is known to capture, at least partially, the booms and busts in the markets. Here, we take up two specific problems. One, how reliably does the spectral structure reflect the system for the case where T~N. Two, a large chunk of the literature on networks construction depends on bivariate modelling which is subject to failure due to multiple hypothesis testing. Therefore, an imminent question is how to construct a network with a direct multivariate model.

Voice AI Effectiveness for Debt Collection

Prof. Anuj Kapoor



Conversational AI Chatbots are rapidly evolving as new technologies with both business potential and customer reactance. This study exploits large-scale field experiment data on thousands of customers who are randomized to receive highly structured outbound sales calls from chatbots. We vary features like the gender of the bot along with the formal or informal tone of the bot. In this paper, we propose a dynamic outbound call experimentation policy. The proposed approach extends multi-armed bandit (MAB) algorithms, from statistical machine learning, to include microeconomic choice theory. Our automated outbound call policy solves this MAB problem using a scalable distribution-free algorithm. Beyond the actual experiment, we plan to counterfactual simulations to evaluate a range of alternative model specifications and allocation rules in MAB policies.

Purchase/Biding behaviour of new and used anthropomorphized and non-anthropomorphized toaster products on eBay and classifying the toasters using ML techniques

Prof. Hyokjin Kwak

Major empirical methods: web scrapping, data pre-processing, independent t-test, machine learning classification method (CNN and ResNet-32). This project aims to study how anthropomorphized 'brand new' and 'used' toaster products affect consumer purchase or bidding behaviour. To do this, I scrapped all the toaster data information like product name, number of consumers watching the product, bidding details etc., from eBay website, toaster products were then manually labelled as Anthropomorphized "AB" or Non- Anthropomorphized "Non-AB." Exploratory data analysis (EDA) was used to look at the attributes of the data. IBM SPSS software is also used to analyse the independent sample t-test. This test compares the means of two independent groups, AB and Non-AB, to see if there is statistical evidence that the relevant attribute means are significantly different. Also, multiple deep learning approaches were used to classify the AB and NAB toaster images.

Effective Amul Brand Positioning Through Topical Ads and Brand Mascot: A Sentimental Analysis on Twitter Data

Prof. Hyokjin Kwak

Major empirical methods: web scrapping, data analysis, sentimental analysis. The objective of this study is to check the purchasing behaviour of consumers in response to creative advertisements that have been posted on Amul's twitter handles and to also apply sentimental classification techniques to the comments that have been posted on Amul's twitter handles. Initially, I compiled all available data on Amul's most successful advertisements from 2019 to 2021, which can be found on the company's website and neatly labeled according to year. Information about advertisements was then culled from the Amul Twitter account and matched with captions taken directly from the company's website. The properties of the data were investigated using exploratory data analysis (EDA) methods. IBM SPSS is also used to analyse the results of the independent sample t-test, which compares the means of two groups to determine whether or not there is statistical evidence of the relevant attributes are significantly different. In addition to that, a sentimental analysis was carried out on the user comments left on the advertisement that was uploaded to Amul's Twitter handle (https://twitter.com/Amul_Coop)

Sentimental Analysis on Amazon Book Reviews in India (vs US)

Prof. Hyokjin Kwak

Major empirical methods: web scrapping, data pre-processing, sentimental and emotions analysis. This research will compare the emotional and sentimental analysis of book reviews that were posted on Amazon.in (India) and Amazon.com (US) by taking into consideration of other characteristics such as book ratings, book cost, discounts available on the book, etc. This study will focus mostly on the sentiments and emotions expressed in Amazon USA and India book reviews. I explored various algorithms like VADER (Valence Aware Dictionary and Sentiment Reasoner), Textblob, SentiBERT etc. to detect the sentiments present in the product reviews also explored emotions like happy, fear, disgust, anticipation, joy, sadness, surprise, and trust.

Brij Disa Centre for Data Science and Artificial Intelligence
INDIAN INSTITUTE of MANAGEMENT AHMEDABAD
विद्याविनियोगाद्विकास:

Production Recommendation using Product Reviews from Amazon India (vs US)

Prof. Hyokjin Kwak

Major empirical methods: web scrapping, data pre-processing, text analysis which includes identifying bi-gram and trigram words In this study, I will be performing an analysis of the textual contents, beginning with the reviews of the product and continuing all the way through to the purchase recommendation. In addition, a semi-automatic method will be utilised to extract terms from the text in reviews, and a knowledge graph will be utilised. The extracted phrases were connected to the various technical aspects of the items. After that, vector representations of the graph elements will be trained, which will result in a significant improvement in the overall quality of the recommendations. I intend to look into Adaptive Text Rank, which is based on a set of technical characteristics and a collection of sentiment words; the SOTA BERT model, which matches terms with the technical features of the products; the TransE method, which trains vector representations of graph elements; and the ABAE method, which highlights important characteristics for products based on a collection of reviews.

Building Fake Review Detection Model for Amazon India (vs US)

Prof. Hyokjin Kwak

Major empirical methods: web scrapping, data pre-processing, text analysis which includes identifying bi-gram and trigram words, collection of large labelled dataset of fake and non- fake reviews from official sources. The most recent AI text generation models, such as GPT-2 and GPT-3, along with a variety of transformer models, can be utilised to generate fake computer-generated reviews. Researchers in the field of data science have demonstrated how artificial intelligence reviews may be generated and identified with the use of machine learning. Additionally, fake reviews can be purchased in bulk from various sources. Researchers have uncovered a wide range of potential traits that can help humans and models determine the difference between a fake review and a genuine one. These features were discovered after studying fake reviews in great detail. Because fake reviews frequently utilise the same language, the review content itself is generally considered to be the most important aspect. This is especially true in cases where multiple reviews were written by the same individual, firm, or other sources. Review length, Sentiment, Helpfulness, Reviews per user, and Verified review are some of the often-used features that can be noticed in studies on fake review detection. In order to create a model that is able to differentiate between the fake reviews that are posted on Amazon.com and Amazon.in, I will be conducting research into text-classification algorithms and NLP pre-processing techniques, such as count vectorization and TF-IDF, as well as machine and deep learning techniques.

Modelling on Online Product Reviews from Amazon India (vs USA)

Prof. Hyokjin Kwak

Major empirical methods: web scrapping, data pre-processing, text analysis, Latent Dirichlet Allocation (LDA). The technique of automatically identifying topics that are present in a text item and deriving hidden patterns that are represented by a text corpus is referred to as topic modelling. The usage of topic models is beneficial for a variety of tasks, including the clustering of texts, the organisation of huge blocks of textual data, the recovery of information from unstructured text, and the selection of features. My goal here is to extract a set number of relevant word groups from the reviews based on sentiment, i.e., positive, negative, or neutral. These word groups are essentially the issues that will aid in determining what the customers are actually talking about in the reviews. This will inform us which subjects are frequently addressed by Indian and American reviews when they favour or dislike the product. This can be expanded upon in terms of emotions as well.

Data-driven auction design: A computational approach

Prof. Jeevant Rampal

Auctions are often used to sell property rights for liquor licenses, spectrum licenses, land and mineral rights, and construction projects etc. This project investigates potential improvements in these auctions using a computational data-driven approach. The first part of this project will be to collect primary data of the participants and their choices in auctions. Subsequently, using the game-theoretic properties of the chosen auction design, we will computationally estimate the true (unobservable) value distribution across players of the object(s) being auctioned (e.g., liquor licenses). The estimation method used will be non-parametric "distance minimization" between the observed out-of-sample distribution of bids, and the predicted out-of-sample distribution of bids using optimally calibrated parameter values. E.g., Athey, Levin, and Seira (QJE 2011) use their estimated model to make comparative static predictions and test that for fit against data from timber auctions. Finally, to analyse which auction design would have best met the various aims of the auction designer, we will use the calibrated model, parameters, and the estimated valuations of the bidders. In particular, using these we will simulate the revenue, efficiency, and other metrics of importance for different auction designs. In addition to the use of simulation described above, to analyse alternate auction designs, we will use simulations of variations of the estimated model, parameters (like risk aversion, budgets etc.), and value distributions to analyse the different rates with which different auction designs can meet the various possible aims of the auction designer.

Multi-period Facility Interdiction Problem

Prof. Sachin Jayaswal

We propose to study a multiperiod interdiction problem, in which the leader (attacker) with a limited interdiction budget decides the sequence of facilities to interdict (destroy) over time so as to inflict the maximum cumulative damage to the follower. The follower's objective is to serve a given set of demand points from the surviving subset of facilities the minimum cumulative cost across all periods. for this, his decisions include the assignments of demand nodes to the surviving facilities and the allocation of his limited budget to the revival of interdicted facilities and the protection of the surviving facilities against their interdiction in the future periods. The multi-period version of the problem, which is the focus of the proposed study, presents additional complexity due to the leader's interdiction decisions constrained by the follower's protection decisions. The objective of the proposed study is to design efficient exact solution methods for this challenging bilevel integer program.

Brij Disa Centre for Data Science and Artificial Intelligence
INDIAN INSTITUTE of MANAGEMENT AHMEDABAD
विद्याविनियोगाद्विकास:

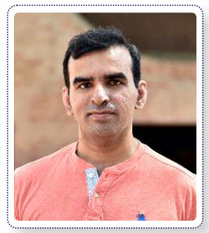
Financial networks from big data: A multivariate time series based approach

Prof. Anindya S. Chakrabarti

Financial markets exhibit non-trivial comovement and dependency structure. The standard approach in the finance literature is to consider the market in Its aggregate form. A more recent 'data'-oriented approach emphasizes a more granular decomposition of the market so that the aggregate dynamics can be broken down into contributions arising from individual assets. This leads to two analytical problems. First, one has to necessarily deal with a large amount of data such that the process scales with the volume of data (large N and large T where T>>N). Two, analyzing such a large volume of data requires toolkits which are at the intersection of econometrics and machine learning. In this project, the goal is to construct large scale financial networks based on multivariate time series data to capture the dynamics of the system. The main idea is to provide an algorithmic approach to convert time series into networks such that the properties of time series are also inherited by the resulting network. The spectral structure of the comovement network is known to capture, at least partially, the booms and busts in the markets. Here, we take up two specific problems. One, how reliably does the spectral structure reflect the system for the case where T~N. Two, a large chunk of the literature on networks construction depends on bivariate modelling which is subject to failure due to multiple hypothesis testing. Therefore, an imminent question is how to construct a network with a direct multivariate model.

Voice AI Effectiveness for Debt Collection

Prof. Anuj Kapoor

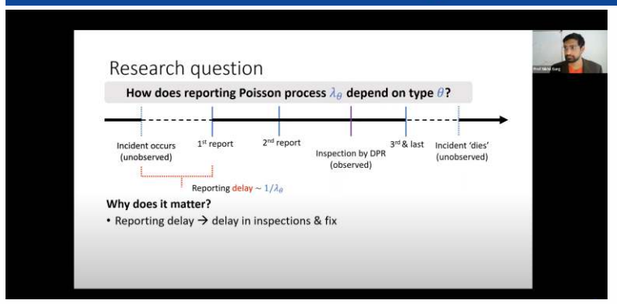


Conversational AI Chatbots are rapidly evolving as new technologies with both business potential and customer reactance. This study exploits large-scale field experiment data on thousands of customers who are randomized to receive highly structured outbound sales calls from chatbots. We vary features like the gender of the bot along with the formal or informal tone of the bot. In this paper, we propose a dynamic outbound call experimentation policy. The proposed approach extends multi-armed bandit (MAB) algorithms, from statistical machine learning, to include microeconomic choice theory. Our automated outbound call policy solves this MAB problem using a scalable distribution-free algorithm. Beyond the actual experiment, we plan to counterfactual simulations to evaluate a range of alternative model specifications and allocation rules in MAB policies.

Seminars and Webinars

Title: Auditing and Designing for Equity in Resident Crowdsourcing

Abstract: Modern city governance relies heavily on crowdsourcing (or "coproduction") to identify problems such as downed trees and power-lines. A major concern in these systems is that residents do not report problems at the same rates, leading to an inefficient and inequitable allocation of government resources. However, measuring such under-reporting is a difficult statistical task, as, almost by definition, we do not observe incidents that are not reported. Thus, distinguishing between low reporting rates and low ground-truth incident rates is challenging. First, joint with Zhi Liu, we develop a method to identify (heterogeneous) reporting rates, without using external (proxy) ground truth data. Our insight is that rates on duplicate reports about the same incident can be leveraged, to turn the question into a standard Poisson rate estimation task—even though the full incident reporting interval is also unobserved. We apply our method to over 100,000 resident reports made to the New York City Department of Parks and Recreation, finding that there are substantial spatial and socio-economic disparities in reporting rates, even after controlling for incident characteristics. Second, I'll briefly overview our work in redesigning inspection decisions to improve system efficiency and equity.

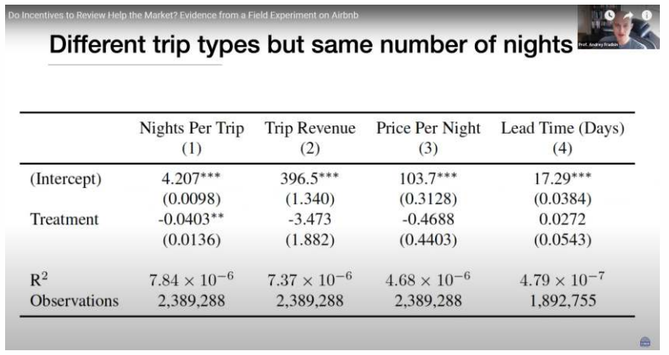


Link to webinar:

Brij Disa Centre for Data Science and Artificial Intelligence
INDIAN INSTITUTE of MANAGEMENT AHMEDABAD
विद्याविनियोगाद्विकास:

Title: Do Incentives to Review Help the Market? Evidence from a Field Experiment on Airbnb

Abstract: Online reviews are typically written by volunteers and, consequently, accurate information about seller quality may be underprovided. We study the extent of this under-provision in a randomized experiment conducted by Airbnb. In the treatment, buyers are offered a coupon to review listings that have no prior reviews. The treatment induces additional reviews, which are more negative on average. Induced reviews do not change nights sold, although they affect the types of transactions that occur. Measures of transaction quality for treated sellers do not improve. We show how market conditions and the design of the reputation system can explain our findings.


Do Incentives to Review Help the Market? Evidence from a Field Experiment on Airbnb

Different trip types but same number of nights

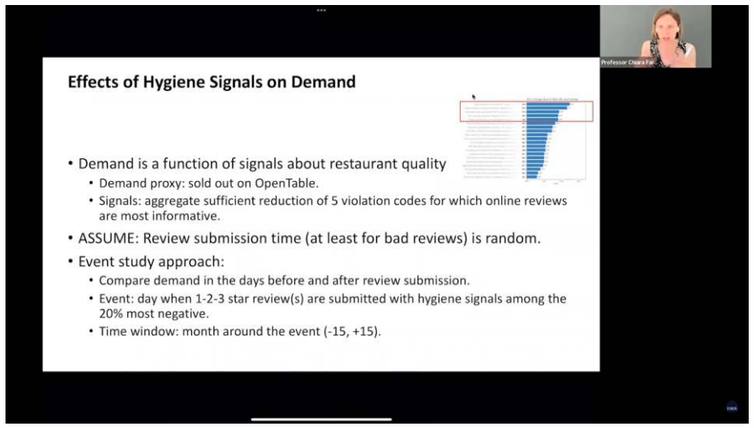
	Nights Per Trip (1)	Trip Revenue (2)	Price Per Night (3)	Lead Time (Days) (4)
(Intercept)	4.207***	396.5***	103.7***	17.29***
	(0.0098)	(1.340)	(0.3128)	(0.0384)
Treatment	-0.0403**	-3.473	-0.4688	0.0272
	(0.0136)	(1.882)	(0.4403)	(0.0543)
R^2	7.84×10^{-6}	7.37×10^{-6}	4.68×10^{-6}	4.79×10^{-7}
Observations	2,389,288	2,389,288	2,389,288	1,892,755

Link to webinar:

Brij Disa Centre for Data Science and Artificial Intelligence
INDIAN INSTITUTE of MANAGEMENT AHMEDABAD
विद्याविनियोगाद्विकास:

Title: Consumer Reviews and Regulation: Evidence from NYC Restaurants (joint work with Georgios Zervas)

Abstract: We investigate the informativeness of hygiene signals in online reviews, and their effect on consumer choice and restaurant hygiene. We first extract signals of hygiene from Yelp. We find that reviews are more informative about hygiene dimensions that consumers directly experience -- food temperature and pests -- than other dimensions. Next, we find causal evidence that consumer demand is sensitive to these hygiene signals. We also find suggestive evidence that restaurants that are more exposed to Yelp are cleaner along dimensions for which online reviews are more informative. Our results have implications for government inspections when consumers rate providers online.

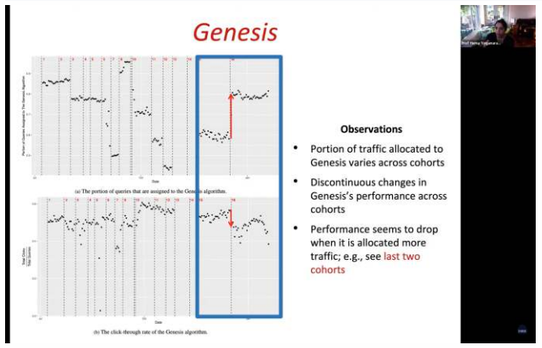


Link to webinar:

Brij Disa Centre for Data Science and Artificial Intelligence
INDIAN INSTITUTE of MANAGEMENT AHMEDABAD
विद्याविनियोगाद्विकास:

Title: A Bias Correction Approach for Interference in Ranking Experoments

Abstract: Online marketplaces use ranking algorithms to determine the rank-ordering of items sold on their websites. The standard practice is to determine the optimal algorithm using A/B tests. We present a theoretical framework to characterize the Total Average Treatment Effect (TATE) of a ranking algorithm in an A/B test and show that naive TATE estimates can be biased due to interference. We propose a bias-correction approach that can recover the TATE of a ranking algorithm based on past A/B tests, even if those tests suffer from a combination of interference issues. Our solution leverages data across multiple experiments and identifies observations in partial equilibrium in each experiment, i.e., items close to their positions under the true counterfactual equilibrium of interest. We apply our framework to data from a travel website and present comprehensive evidence for interference bias in this setting. Next, we use our solution concept to build a customized deep learning model to predict the true TATE of the main algorithm of interest in our data. Counterfactual estimates from our model show that naive TATE estimates of clicks and bookings can be biased by as much as 15% and 28%, respectively.

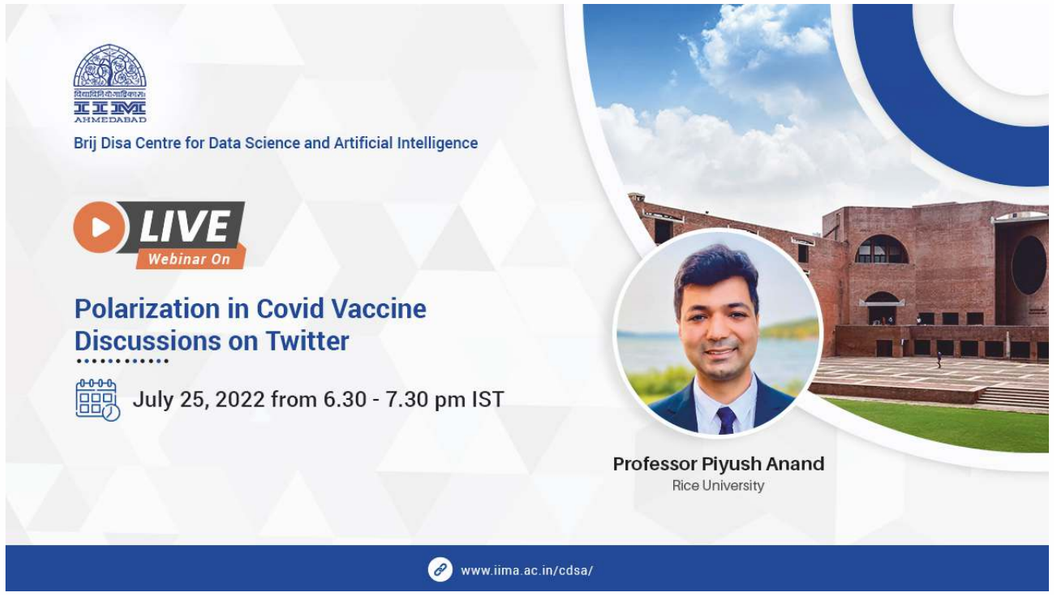


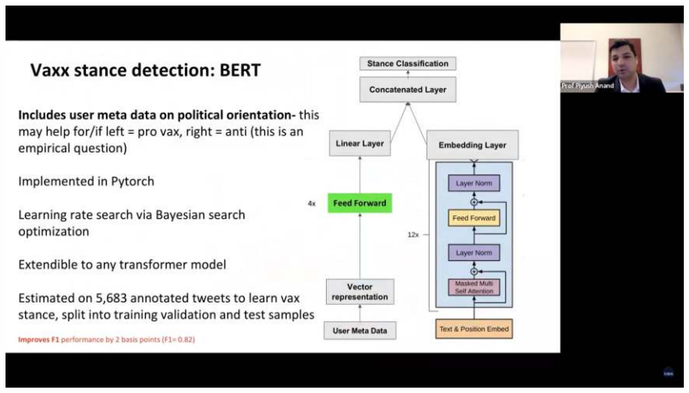
Link to webinar:

Brij Disa Centre for Data Science and Artificial Intelligence
INDIAN INSTITUTE of MANAGEMENT AHMEDABAD
विद्याविनियोगाद्विकास:

Title: Polarization in Covid Vaccine Discussions on Twitter

Abstract: We explore the nature of vaccine conversations on Twitter from January 2019-June 2021, i.e. just before Covid to after major vaccine announcements and rollout. We answer the following questions: (1) Can Twitter be a good info-veillance tool for measuring vaccine sentiments? (2) Do vaccine conversations get more politically polarized during Covid? (3) Do vaccine conversations comprise more opinions over verifiable facts? (4) Do vaccine conversations become more toxic during Covid? We provide time series trends as well as causal estimates around two major vaccine announcements. We discuss implications for managers, health policy regulators, and platform managers.





Link to webinar:

Brij Disa Centre for Data Science and Artificial Intelligence
INDIAN INSTITUTE of MANAGEMENT AHMEDABAD
विद्याविनियोगाद्विकास:

Title: You Cannot Optimize What You Do Not Know: Evidence From A Reform

Abstract: Firms may try to maximize profits but fail due to partial knowledge about the value of possible actions. I use a reform to argue that the only model of the firm consistent with the data is one with mis-optimization due to insufficient knowledge of the world. A reform in Israel limited prices to end with X0 as the cents digits (e.g. 2.90 but not 2.99), thus limiting the set of actions available to the firm. Since consumers are left-digit biased, specifically with demand discretely falling by 5%-9% at round prices, optimal pricing implies bunching at just-below prices (e.g. 2.99 or 2.90) and avoiding round prices (e.g. 3.00). Supermarkets act as if they know this discontinuous demand structure but underestimate its magnitude before the reform, setting just-below prices for 45% of prices and rarely using round prices. If price-setting before the reform was driven by the correct model of demand, firms' response to the reform would have been to update immediately according to their beliefs and avoid round prices. Yet, pricing after the reform is inconsistent with their pre-reform revealed beliefs, setting 20% of clearly dominated prices for almost a year. Whether firms were optimizing a wrong model, or making decisions in a model-free way, their knowledge had to be partial. Partial knowledge and incomplete learning can explain how firms behave sub-optimally in a persistent way and challenges counterfactual exercises that rely on the assumption of model-based optimization. I suggest an approach of almost-optimization to calculate counterfactuals that are more likely to contain the truth.

Brij Disa Centre for Data Science and Artificial Intelligence
INDIAN INSTITUTE of MANAGEMENT AHMEDABAD
विद्याविनियोगाद्विकास:

Title: Misinformation in the context of Covid-19 Pandemic: an Investigation of health harm characteristics and related social media conversations

Abstract: In the environment of health crises, misinformation dissemination on social media has been shown to result in serious harms. In investigating misinformation harm, this research-in-progress focuses on the following research questions in the context of safety: (1) How can the perceptions of harm from misinformation be characterized in health crises, particularly for the COVID-19 pandemic misinformation? (2) What are the novel characteristics of harm perceptions from health crisis misinformation during COVID-19 global pandemic? (3) What are the types of social media conversations that are associated with certain characteristics of COVID-19 health crisis misinformation harm perceptions? Focusing on the COVID-19 context, we address these research questions through two studies: In Study 1, we adapt the psychometric paradigm as used in risk perception literature to misinformation harm perception and examine several COVID-19 misinformation scenarios. We discover that health misinformation harm can be characterized as three factors, i.e., unknown and dread (both of which are well known in the risk literature) and manageability (a concept that has hitherto not been discussed in the literature and is discovered in this research). In Study 2, we follow a data-driven approach that utilizes text mining to examine the association between harm perceptions and nature of communication in Twitterverse. We suggest propositions for future research. This research is one of the early attempts at quantitatively measuring health misinformation harms in the pandemic context. Our findings draw insights and suggestions for future related studies on misinformation harm.

Brij Disa Centre for Data Science and Artificial Intelligence
INDIAN INSTITUTE of MANAGEMENT AHMEDABAD
विद्याविनियोगाद्विकास:

Title: Scalable bundling via dense product embeddings

Abstract: Bundling, the practice of jointly selling two or more products at a discount, is a widely used strategy in industry and a well- examined concept in academia. Scholars have largely focused on theoretical studies in the context of monopolistic firms and assumed product relationships (e.g., complementarity in usage). There is, however, little empirical guidance on how to actually create bundles, especially at the scale of thousands of products, which is common today. Here, we use a machine learning-driven approach for designing bundles in a large-scale, cross-category retail setting. We leverage historical purchases and consideration sets determined from clickstream data to generate dense representations (embeddings) of products. We impose minimal structure on these embeddings and develop heuristics for complementarity and substitutability among products. Subsequently, we use the heuristics to create multiple bundles for each of 4,500 focal products and test their performance using a field experiment with a large retailer. We use the experimental data to optimize the bundle design policy with offline policy learning. Our optimized policy is robust across product categories, generalizes well to the retailer's entire assortment, and provides an expected improvement of 35% ($5 per 100 visits) in revenue from bundles over a baseline policy using product co-purchase rates.

Potential complements for **Organic Russet Potatoes**

Product	Score
Organic Celery Hearts, 16 Oz	0.75
Organic Grape Tomatoes, 1 Pint	0.74
Organic Green Bell Peppers, 2 Ct	0.74
Organic Carrots, 2 Lb	0.73
Organic Cauliflower, 1 Ct	0.73
Organic Garlic, 8 Oz	0.72
The Farmers Hen Large Organic Eggs, 1 Dozen	0.70
Organic Bananas, Minimum 5 Ct	0.69
Organic Broccoli Crowns, 2 Ct	0.69
Organic Romaine Hearts, 3 Ct	0.69

Potential substitutes for **Organic Russet Potatoes**

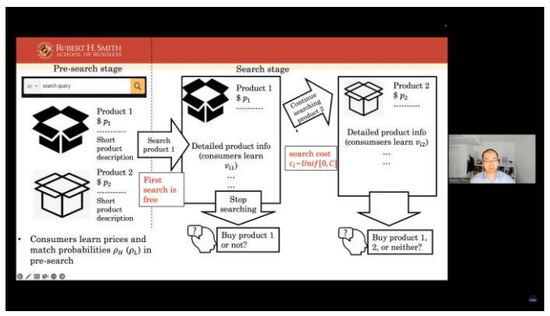
Product	Score
Green Giant Organic Golden Potatoes, 3 Lb	0.95
Green Giant Organic Red Potatoes, 3 Lb	0.95
Organic Russet Potatoes, 3 Lb	0.95
Green Giant Klondike Gourmet Petite Purple-Purple Fleshed Potatoes, 24 Oz	0.93
Green Giant Klondike Fingerling Potatoes, 24 Oz	0.93
Green Giant Golden Potatoes, 5 Lb	0.92
Organic Sweet Potatoes, 3 Lb	0.92
Green Giant Klondike Petite Red-White Fleshed Potatoes, 24 Oz	0.92
The Little Potato Garlic Herb Potato Microwave Kit, 16 Oz	0.92
The Little Potato Company Garlic Herb Oven Griller Kit, 16 Oz	0.91

Link to webinar:

Brij Disa Centre for Data Science and Artificial Intelligence
INDIAN INSTITUTE of MANAGEMENT AHMEDABAD
विद्याविनियोगाद्विकासः

Title: Search Neutrality and Competition between First-party and Third-party Sellers

Abstract: Retail platforms often serve both as a bridge connecting buyers and third-party sellers and as a competing (first-party) seller. Many dominant platforms self-preferentially prioritize first-party products, despite their potential lower relevance, over third-party products in consumers' personalized search rankings. Antitrust agencies propose search-neutrality regulations that forbid such self-preference practices to protect consumers and third-party sellers. This paper shows that search neutrality, despite its good intention, may unintendedly harm consumers and third-party sellers after accounting for the strategic decisions of the platform and third-party sellers. In the short term, search neutrality can weaken the price competition between the platform and third-party sellers, which will harm consumers if most consumers ex ante prefer the third-party product. By contrast, the platform's profit may increase due to the alleviated competition. In the long term, search neutrality can incentivize the platform to raise its commission rate to preempt third-party sellers' entry if their entry cost is intermediate, which further harms both consumers and third-party sellers. These unintended consequences stem from two unique features of online retailing platforms: platforms personalize consumers' search rankings, and consumers observe product prices before searching a product.

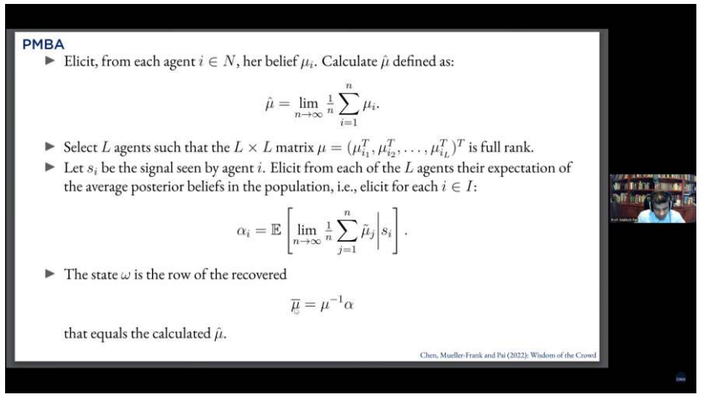


Link to webinar:

Brij Disa Centre for Data Science and Artificial Intelligence
INDIAN INSTITUTE of MANAGEMENT AHMEDABAD
विद्याविनियोगाद्विकासः

Title: Mechanisms to Infer the Wisdom of the Crowd. (joint work with Yi-Chun Chen and Manuel Mueller-Frank).

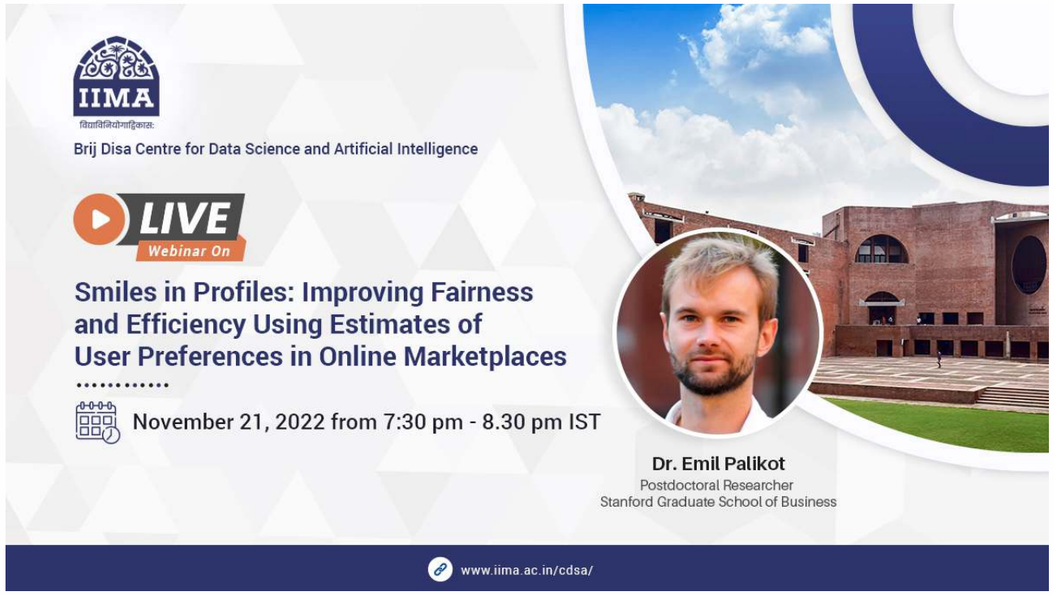
Abstract: The classic wisdom-of-the-crowd problem asks how a principal can "aggregate" information about the unknown state of the world from agents without understanding the information structure among them. We revisit the "Surprisingly Popular" (SP) mechanism of Prelec, Seung and McCoy (2017) and provide a general theoretical foundation for when it can/cannot work. We propose a "population mean based aggregation" mechanism that generalizes the SP. It only requires eliciting agents' beliefs about the state, and also eliciting some agents' expectations of the average belief in the population. We show that this procedure fully aggregates information: in an infinite population, it always infers the true state of the world. The procedure can accommodate correlation in agents' information, misspecified beliefs, any finite number of possible states of the world, and only requires very weak assumptions on the information structure.

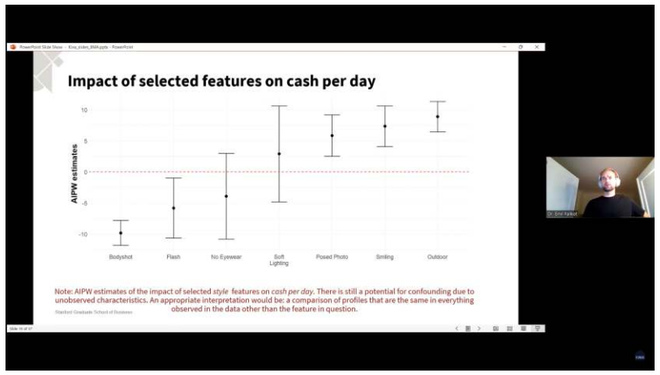


Link to webinar:

Title: Smiles in Profiles: Improving Fairness and Efficiency Using Estimates of User Preferences in Online Marketplaces

Abstract: Online platforms often face challenges being both fair (i.e., non-discriminatory) and efficient (i.e., maximizing revenue). Using computer vision algorithms and observational data from a micro-lending marketplace, we find that choices made by borrowers creating online profiles impact both of these objectives. We further support this conclusion with a web-based randomized survey experiment. In the experiment, we create profile images using Generative Adversarial Networks that differ in a specific feature and estimate it's impact on lender demand. We then counterfactually evaluate alternative platform policies and identify particular approaches to influencing the changeable profile photo features that can ameliorate the fairness-efficiency tension.

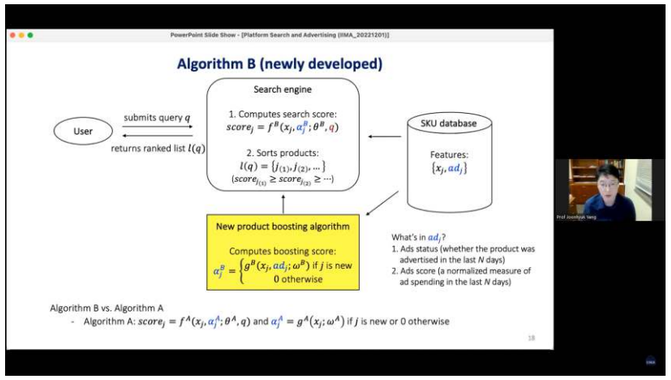




Link to webinar:

Title: Advertising as Information for Ranking E-Commerce Search Listings

Abstract: Search engines and e-commerce platforms have substantial difficulty exposing new products to their users on account of an information problem: new products typically do not have enough sales or other user-engagement that enables platforms to reliably assess product quality. This paper evaluates the role of advertising in providing information to the platform regarding new product quality so as to solve this "cold-start" problem and to engineer higher quality organic listings. Using a large-scale experiment implemented at JD.com---a large e-commerce platform in China---we show that using ad propensity information for ranking new products benefits both the platform and consumers. Our findings showcase a new channel by which advertising can potentially improve outcomes for consumers and platforms in e-commerce, through its ability to reveal information that can be used by platforms to improve search ranking algorithms.

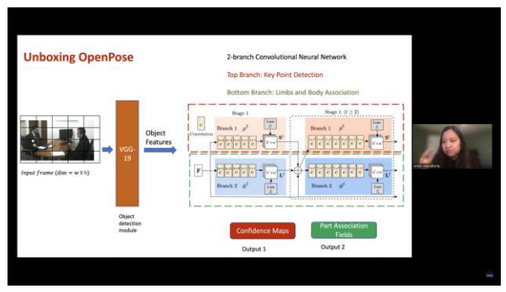


Link to webinar:

Brij Disa Centre for Data Science and Artificial Intelligence
INDIAN INSTITUTE of MANAGEMENT AHMEDABAD
विद्याविनियोगाद्विकास:

Title: AI and AI-Human Based Salesforce Hiring Using Interview Videos (joint with K. Chiong, H.Dover and K. Sudhir)

Abstract: We study the problem of AI and AI-human based screening (eliminating bottom candidates) and selection (hiring top candidates) in salesforce hiring. Using videos of structured interviews and ratings from multiple recruiting experts on standard performance criteria, we develop an AI model of salesforce "skill" prediction by extracting theory-relevant objective measures of interviewee performance from videos. Using the model, we address two issues: First, to aid interpretability, we assess what mode of unstructured data from the interview (text, audio and video) and what specific behaviors (e.g., body language, conversation style) drive AI predictions of salesperson's ability. We find that while screening decisions are heavily impacted by audio-visual features like "energetic voice" and "open posture"; selection relies more heavily on textual content and linguistic features such as being precise and quantitative. We also find that there is an optimal amount of hand gesturing, body posture movement and conversational interactivity for success in selling. Second, we consider a hybrid AI-human model where we augment AI predictions with human interventions in a Bayesian framework. The hybrid model improves accuracy but increases cost (of human labor). Hence, we propose a cost-effective way of deploying AI in salesforce hiring with humans in the loop — use AI for screening and augment it with human judgments based on early stages of interview for selection.

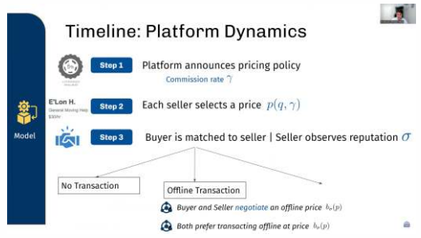


Link to webinar:

Brij Disa Centre for Data Science and Artificial Intelligence
INDIAN INSTITUTE of MANAGEMENT AHMEDABAD
विद्याविनियोगाद्विकास:

Title: Disintermediation in Online Platforms: The Role of Information Quality and Pricing

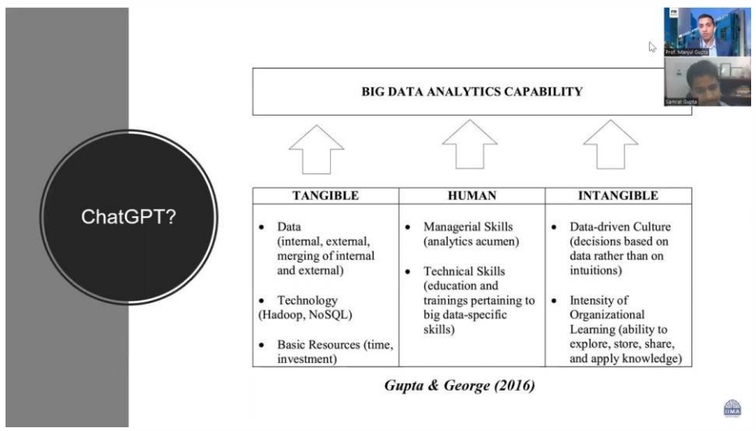
Abstract: Online platforms like AirBnB and Upwork generate revenue by matching prospective buyers and sellers and extracting commissions from completed transactions. Disintermediation, where sellers transact offline with buyers to bypass commission fees, can lead to significant revenue losses for these platforms. For example:
a) the talent outsourcing platform ZBJ estimates that up to 90% of their service providers' transactions may occur off-platform; b) a recent study estimates that online labour markets (e.g., cleaning platforms) lose out on 25% of potential transactions due to disintermediation; c) in the extreme case, disintermediation can threaten the viability of the platform itself as in the case of the home-cleaning platform Homejoy, which was shut down in 2015. In this talk, we will use a modelling and analytical framework to examine how disintermediation impacts the platform's incentives, commission rate, and revenue. First, we argue that in the presence of disintermediation, platforms have strong incentives to keep its review systems noisy as providing more information about risky buyers can amplify disintermediation and hurt platform revenue. This goes against conventional wisdom that lowering information asymmetry can benefit platforms by enabling more efficient matches. Second, and somewhat surprisingly, we show that the risk of disintermediation may actually force platforms to increase the commission rate, i.e., the platform may not lower the commission rate to encourage users to stay online. Instead, they may look to make up for lost revenue by overcharging the remaining users. Finally, motivated by platforms like Thumbtack which have migrated from commissions to charging sellers for "leads", we examine whether platform-access fees or seller-side subscriptions can hedge against losses from disintermediation. While access fees do indeed reduce the incentives for disintermediating, they may drive away sellers who face too much uncertainty, and as a result, fall short of recovering the full revenue lost due to disintermediation.



Link to webinar:

Title: In the world of ChatGPTs, not every organization is equipped to take advantage of the next generation Artificial Intelligence artifacts

Abstract: The era of ChatGPTs has begun such that organizations in all industries would eventually be investing in these artificial intelligence (AI) artifacts. However, we know from prior literature that investments alone do not generate competitive advantage; instead, firms need to create capabilities that rival firms find hard to match. When all organizations of similar sizes have access to tools like ChatGPT, simply owning a technological artifact will not be sufficient to generate a competitive advantage. The firm would need to develop a unique combination of a variety of firm-specific resources that would result in the creation of an AI-oriented capability that would outperform the capabilities of the rival firms.



Link to webinar:

Brij Disa Centre for Data Science and Artificial Intelligence
INDIAN INSTITUTE of MANAGEMENT AHMEDABAD
विद्याविनियोगाद्विकासः

IKIGAI workshop: Democratizing AI through Lo-No code platforms

It was a pleasure hosting Prof Devavrat Shah from Massachusetts Institute of Technology over a Seminar and Ikigai to conduct hands-on training workshop on Creating AI data Apps using Ikigai platform at #iimahmedabad. Ikigai CTO & Co-founder Devavrat Shah with their team Vinay Gahlot, Chanikya Kesarapu guided the participants to build an AI/ML data app. The workshop covered the end-to-end process of data extraction, cleaning, stitching, training the data on an ML model, visualization, and deploying the model in an AI app to recommend values of target columns for new data. By the end of the session, participants created their own functional AI data app for housing prices recommendation. Ikigai is a platform that offers a range of data management and analysis tools, including data visualization, reporting, and dashboards. One of the key features of Ikigai is its ability to be customized to suit the needs of different industries, making it a horizontal platform that can be used across various sectors. Founded out of MIT, the powerful Ikigai tool is mainly used by operations teams that use data for mission-critical operations. The no-code, cloud based end-to-end platform works alongside or on top of existing data/application stacks seamlessly enabling it to achieve time-to-value within days.

Brij Disa Centre for Data Science and Artificial Intelligence
INDIAN INSTITUTE of MANAGEMENT AHMEDABAD
विद्याविनियोगाद्विकास:

CDSA Budget Annual Accounts for the FY 2022-23

No	Particulars	Utilized amount for 2022-23
1	Total revenue at beginning of Year	1,98,52,554
2	Donor fund and top ups - revenue	1,50,00,000
3	Total revenue	3,48,52,554
4	Interest	13,89,679
5	Lab infrastructure (Capital purchase)	2,29,472
6	Salary	70,11,761
7	Travel	29,551
8	Activities	4,19,660
9	Other expenses	2,11,499
10	Institute overhead	7,50,795
11	Total expenditure	86,52,738
	Balance for 2023-24	2,75,89,493

Brij Disa Centre for Data Science and Artificial Intelligence
INDIAN INSTITUTE of MANAGEMENT AHMEDABAD
विद्याविनियोगाद्विकास:

Executive Committee - Current Members

Anindya S. Chakrabarti
Co-Chairperson

Ankur Sinha
Co-Chairperson

Adrija Majumdar

Anuj Kapoor

Chetan Soman

Karthik Sriram

The Team

Debjit Ghatak
Centre Head

Neaketa Chawla
Post-doctoral
Research Associate

Satender
Pre-doctoral
Research Associate

Nebu Varghese
Pre-doctoral
Research Associate

Vaishnav Garg
Pre-doctoral
Research Associate

Rahul Sharma
Pre-doctoral
Research Associate

Shaswot Nayak
Pre-doctoral
Research Associate

Sayantan Pramanick
Pre-doctoral
Research Associate

Jaydeep Gohel
Centre-secretary

Advisory Board

Anand Deshpande

Tiziana Di Matteo

Sanjiv Das

Sunil Gupta

Brij Disa Centre for Data Science and Artificial Intelligence
INDIAN INSTITUTE of MANAGEMENT AHMEDABAD
विद्याविनियोगाद्विकास:

Current Centre Members


Anindya S. Chakrabarti


Anuj Kapoor


Adrija Majumdar


Ankur Sinha


Chetan Soman


Anirban Banerjee


Dhiman Bhadra


Indranil Bose


Swanand Deodhar


Anil Deolalikar


Andrea Lodi


Thomas Lux


Tanmoy Majilla


Neelkant Rawal


Pekka Malo


Aditya Christopher Moses


Sundaravalli Narayanaswami


Sriram Sankaranarayanan


Pankaj Setia


Avinash Sharma


Karthik Sriram


Sobhesh Kumar Agarwalla


Cheong Siew Ann


Arindam Banerjee


Samrat Gupta


Sachin Jayaswal


Sudha Ram


Hyokjin Kwak


Sheri Markose


M P Ram Mohan

Brij Disa Centre for Data Science and Artificial Intelligence
INDIAN INSTITUTE of MANAGEMENT AHMEDABAD
विद्याविनियोगाद्विकास:


Peng Shi


Hemant Kumar Singh


Pranav Singh


Sitabhra Sinha


Soumya Mukhopadhyay


Suprateek Sarker


Abhishek Tripathi


Arvind Tripathi


Ellapulli Vasudevan


Sanjay Verma


Pritha Dev


Ambrish Dongre


Jeevant Rampal


Viswanath Pingali


Anish Sugathan


Mohsen Mohaghegh

Annual Report | 2022-2023

Articles published by Centre in Quarterly Communique

Brij Disa Centre for Data Science and Artificial Intelligence
INDIAN INSTITUTE of MANAGEMENT AHMEDABAD
विद्याविनियोगाद्विकास:

IIMA

विद्याविनियोगाद्विकासः

INDIAN INSTITUTE *of* MANAGEMENT AHMEDABAD

भारतीय प्रबंध संस्थान अहमदाबाद

Admissions Office (PGP & PGP-FABM)

वस्त्रापुर, अहमदाबाद 380015, भारत | Vastrapur, Ahmedabad 380015, India.

Phone: +91-79-7152 7742 | **Email:** cdsa@iima.ac.in | **Website:** www.iima.ac.in