



**Karnataka Jnana Aayoga
(Karnataka Knowledge Commission)**

**Report of the Workshop on Fostering Innovation in
Data Science: Critical Areas and Critical Issues**

May 27, 2016

In Collaboration with Department of IT, BT and Science & Technology
Hosted by Department of Computer Science and Automation (CSA), IISc



(Report by KJA Task Group on Machine Intelligence)

**Karnataka Jnana Aayoga
Karnataka Knowledge Commission
(Government of Karnataka)**



Table of Content

Abstract	3
Acknowledgements	4
About Karnataka Jnana Aayoga	5-6
Keynote Address	7-8
Smart Cities	9-11
Technologies for Massive Online Education	12-13
Legal Issues in Data Science	14-15
Data Science in Action	16-18
Panel Discussion	19-20
Special Remarks by the Co-chair	21-22
Recommendations	23
Annexure-I-List of Participants	24-25



Abstract

This is a report on a workshop conducted by KJA Task Group on Machine Intelligence. The workshop titled ‘Fostering Innovation in Data Science: Critical Areas and Critical Issues’ was held on May 27, 2016 in Computer Science and Automation (CSA) department in Indian Institute of Science (IISc). The main agenda of this workshop was to explore the role of ML and Data Science in development of Smart Cities- an initiative of Government of India. Another agenda was to discuss legal issues regarding the availability and accessibility of administrative data collected by the government. The workshop brought together several leading data scientists from the industry to discuss the data-driven research being undertaken in their respective groups, which could be relevant for modern urban development, especially in Indian context.

Dr. K. Kasturirangan, Chairman, KJA and Ms. Manjula, V., Principal Secretary, Department of IT, BT and S&T, GoK delivered the keynote address. The workshop had four sessions-1) Smart Cities 2) Technologies for Massive Online Education 3) Science in Action and 4) Legal Issues in Data Science Data. First three sessions included talks by several representatives of industry and last session was a panel discussion involving government representatives and experts on law. Participants included TG-MI members, representatives from academia, government and industries besides research scholars. Mr. Venkatesh Valluri, KJA Member and TG Co-chair made a Special Remarks and Prof. Chiranjib Bhattacharyya, TG Co-chair designed and anchored the workshop.

Key outcomes of the workshop has been placed at the end as Recommendations for consideration by the TG-MI.



Acknowledgements

KJA Task Group on Machine Intelligence expresses its sincere thanks to Dr. K. Kasturirangan, world renowned space scientists and Chairman, KJA and to Ms. Manjula V., Principal Secretary, Department of IT, BT and S&T, GoK.

TG-MI is grateful to Department of Computer Science Automation (CSA), Indian Institute of Science for hosting the Workshop. TG extends its thanks to Prof. Y. Narahari, Chairman of Electrical Sciences Division, for inaugurating the workshop and for setting the agenda. TG is thankful to the staff and students of CSA Department for their help in organising the event. A special thanks to all the Speakers, GoK Officers, government and academia representatives for sharing their work and ideas. TG-MI is indeed thankful to Mrs. Tanusree Deb Barma, Director, ITBT, Gok for anchoring a Panel session.

TG-MI also expresses its thanks to Mr. Adway Mitra, Mr. Ankit, Mr. Srijith and Mr. S. Bohra of CSA Department for their support in capturing the workshop deliberation.



About Karnataka Jnana Aayoga (KJA)

Karnataka is emerging as a Knowledge State in the country and needs to prepare itself to be a knowledge society – through education, innovation, skill development, social development and economic growth. With a view to address these and develop a progressive society in Karnataka state, Government of Karnataka constituted Karnataka Jnana Aayoga. The present KJA has been formally reconstituted in December 2013 under the Chairmanship of Dr. K. Kasturirangan. KJA has 32 experts and department Secretaries as Members. The main aims and objectives of the present KJA are to recommend actions for institution building, policy innovation and excellence in the field of education, health, S&T, industry, entrepreneurship, research and innovation, traditional knowledge, agriculture, E-Governance, rural development and ANY other relevant areas.

Tasks of KJA are mainly “proof-of-concept” and get defined/formulated, either through internal discussions within KJA – mainly issues of public/societal/technological and knowledge relevance for the state OR are identified through interactions with GOK departments – mainly issues of governance and development in the state. KJA tasks are “anchored” with one or more departments of GOK – so that after proof-of-concept stage by KJA, any executive implementation can get effectively coordinated by relevant departments of GOK. As part of “proof-of-concept”, KJA takes up technical assessments, studies, small demonstrations, proto-typing analysis, policy analysis, social impact assessment etc – which result in relevant recommendations to GOK.

Karnataka is one of the leading states in technology and machining with a wide range of government, private sector and academia involved in a range of technology – including, machines, information technology, aero-space technology, education technology and a host of industrial technologies. Given this status, Karnataka aspires to rise to greater heights so that its citizens can be assured of advanced technology development benefits and the state can position a futuristic technology-climate that will prepare for an advanced, mature and economically prosperous state and national development processes.

Recent developments in smart mobile systems and cloud computing are enabling the connection of millions of devices and that has amplification of any governance, technological, efficient, moral and ethical imperative through a network of machines. Smart Machines are now becoming a cognitive, contextually aware computing system capable of making decisions without human intervention – that is the way of future. Machine Intelligence technologies – 3d Printing, indigenous Cloud design, Internet-Of-things (IOT), robotics, Artificial Intelligence (AI), algorithmic societal tools, social analytics etc are getting integrated into education systems, high-end industrial capability, analytical governance and citizen-centric services of improved quality of life through efficient governance.

To develop a Machine Intelligence Plan for Karnataka, KJA decided to constitute a **Task Group on Machine Intelligence** composed of technology specialists/experts and government representatives. The Task Group was formally constituted in November 2015 under the Co-Chairmanship of Mr. Venkatesh Valluri and Prof. Chiranjib Bhattacharyya.



The main task of the KJA-Task Group on Machine Intelligence is to prepare a strategy for Karnataka to take up leadership in development of Machine Intelligence technologies through an integrated development of key technologies, applications development, education and research and industrialisation. The TG will also define a roadmap for development and industrialisation of technologies and applications of Machine Intelligence for the benefit of the society, state and nation-specifically in areas of healthcare, science & technology education, transportation, rural & urban development, governance, citizen services etc.,

About the Workshop

Data Science has created a huge buzz in the world of science and technology, and has tremendous potential for revolutionizing the traditional approaches to various real-life problems in every sector. While the initial impact has been made in the IT sector, it has obvious applications in the social and governance sectors also. Data Science is the key ingredient of Machine Learning (or Machine Intelligence), which is an elaborate mathematical framework for modelling data. MI enables computer-aided intelligent decision-making, in the context of known information.

To accomplish its mandate, the KJA Task Group on Machine Intelligence is exploring the scope and potential of using techniques of Machine Learning and Data Science for the purpose of governance. For this purpose, the TG is collaborating with Department of Information Technology, Biotechnology, Science and Technology (IT, BT and S&T) of Government of Karnataka, and Department of Computer Science and Automation (CSA), Indian Institute of Science (IISc). The plan is to focus on a few specific applications- healthcare, education, transportation, rural and urban development, governance, citizen services, consumer electronics, etc. The three bodies have agreed to hold a series of workshops that bring together representatives of the government, industry and academia to exchange ideas, perspectives and technical know-how on the matter.

The first workshop, held on January 18, 2016 focussed on capacity-building, by examining the undergraduate curriculum in various engineering colleges across Karnataka. The second workshop in this series was held on May 27, 2016 in CSA Department, IISc. The main agenda of this workshop was to explore the role of ML and Data Science in development of Smart Cities- an initiative of Government of India. Another agenda was to discuss legal issues regarding the availability and accessibility of administrative data collected by the government. The workshop brought together several leading data scientists from the industry to discuss the data-driven research being undertaken in their respective groups, which could be relevant for modern urban development, especially in Indian context.

The organizers of the workshop were Prof Chiranjib Bhattacharyya of CSA Department who specializes in Machine Learning, Dr.VenkatेशValluri – KJA member and chairperson on the Task Group, and Dr.MaulishreeAgrahari – ICT Skill Development Society, Government of Karnataka. The session was inaugurated by Prof. Y. Narahari of CSA Department, and started with two opening addresses – by Dr. K. Kasturirangan – eminent space scientist and current chairperson of the KJA, and Ms. V. Manjula – Principal Secretary, Dept. of IT, BT and S&T, Government of Karnataka. There were four sessions- 1) Smart Cities, 2) Technologies for Massive Online Education, 3) Legal Issues in Data Science and 4) Data Science in Action. Each session included talks by several representatives of industry, except Session 3 which was a panel discussion involving government representatives and experts on law. Finally, the participants put forward requests and questions to the Karnataka Government, which were responded to by Ms. Maulishree as a representative of the government.

Session 1: Keynote Address

The workshop started with introductory addresses by Dr. K. Kasturirangan and Ms. V. Manjula.

Dr. K. Kasturirangan, the current chairperson of KJA, started by thanking the IISc and Department of IT, BT and S&T of Government of Karnataka, for their initiative which he believes has tremendous potential. He personally thanked Dr. Venkatesh Valluri for his brilliant ideas about how Data Science and Machine Intelligence can be used for critical issues and Policy formulation. He recounted how computing technology has evolved since the 1960s, and has come to a stage where it can be used to interpret massive data and make judgements in the core areas of humanity. He identified three main areas of computing technology- development of parallel computing technology, techniques of acquisition and storage of big data, and algorithms for their analysis such as deep learning.

Dr. Kasturirangan went on to mention a few critical areas of social development where data science and machine intelligence can be a game-changer. The population has grown so fast in India and beyond, that conventional systems of human development are unable to keep up. He noted that in the field of healthcare, given the worldwide population explosion coupled with poverty and inequality, conventional techniques will take 300 years to development for a satisfactory universal healthcare system. However, the available clinical data can be used to model and simulate healthcare systems, which can lead to direct the efforts effectively and bring down the required time. He mentioned that the Indian judicial system suffers a massive backlog of cases due to inadequate number of legal professionals. However, Machine Intelligence may be capable of learning from past judgements on similar cases, and coming up with reasonable initial judgements which may be referred to human judges only if the concerned parties are interested, thus potentially reducing the workload on judges. He mentioned such a system being studied by University of Illinois in USA. He also mentioned that climate change is an issue affecting most people in India, especially farmers, for whom rainfall prediction is very important. In the presence of meteorological data obtained by remote sensing, the existing weather-forecasting techniques based on Global Circulation Models (GCM) can be augmented to make more accurate fine-grained weather forecasts.



Finally, he noted that in India, the government holds 80% of the data from critical fields of human development, and it needs to be sensitized about the potential of such data. Also, skill development and capacity building among government employees are required to make best use of such data.

Ms. V. Manjula, the principal secretary of Karnataka JnanaAyoga, in her address, acknowledged Dr. Kasturirangan's comment that the Karnataka government holds a lot of data, and often doesn't know how to use it. It has data related to Human Development Indices upto the Taluka level, but there are issues with the integrity with the data, and know-how on validation and analysis of such data is missing. Most governments in India are severely lagging behind the private sector in making

data-driven decision making. Such decision-making by the government can be very effective in keeping up with the ever-changing urban landscape, for example in prediction of urban floods given that the natural drainage systems in cities have changed due to urban development.

She mentioned the government's thoughts and plans about collection of data and opening it up to the general public for use. Government of India's policy of building optical fibre network can help in



reliable data acquisition and transmission from gram panchayats. Such data may be made available to individuals by the government, who can also tap social innovations by crowdsourcing 5 key problems every year. The Government of Karnataka may replicate Government of India's Open Data Policy in this regard. She mentioned some government initiatives such as BMTC's Intelligent Transport System, and the need to keep up with those, along with capacity-building and collaboration with institutions like IISc.

Prof Chiranjib Bhattacharyya, Professor of CSA department who specializes in Machine Learning and also is the Co-chair of Task Group on MI, introduced the program for the workshop. Citing recent references he remarked that opening up data can unlock 3 trillion US dollars per year globally in just 7 important sectors. In his opinion, this would be possible if the government plays the role of provider and regulator of data. There are 4 important characteristics of open data, namely: high degree of availability, machine readable, low cost, and legal rights to reuse and redistribute. If government can ensure availability of data following these characteristics it will have a very positive impact on our economy unlocking great value. This could be realised if government could set up an ecosystem where entrepreneurs and informed citizens can join hands to come up with innovative data-driven applications, such as mobile apps, in various sectors such as electricity, transportation, education, health, governance etc.

Session 2: Smart Cities

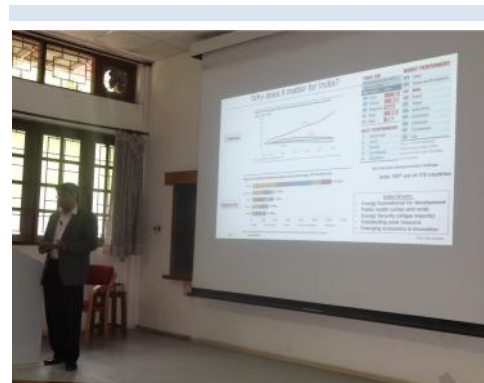
This session was chaired by **Prof. Amrutur Bharadwaj**, IISc. It included 4 talks- by AnandRangarajan from Google India, Dr. Shivkumar Kalyanraman from IBM India Research Lab, Dr. Koyel Mukherjee from Xerox Research Center India and Bipin Kumar, Gaia Smart Cities.

AnandRangarajan had done B.Tech from IIT Madras, followed by Masters' Degree from Stanford University. He currently works in Google India. The topic of his talk was “**What can Machine Intelligence do to make Cities Smarter?**” He started off by highlighting key technologies – small data analysis (personalization), big data storage and analysis, Machine Learning, Speech Processing, Language Processing, Image and Video Processing, Collaborative applications, Compute Engines and Cloud APIs, Knowledge graphs (to convert unstructured data to structured) and Mobile Sensors.

Next, he mentioned several current applications of Data Science being used by Google for urban development at various places in the world. These applications include adaptive transport system based on real-time traffic information, healthcare, energy, smart management of resources, like lighting and water distribution, monitoring of air quality and pollution, waste management etc. He observed that many of the solutions which have been developed in other cities of the world, can be imported directly to Indian cities, such as public health and epidemic detection system, and providing personalized education on the lines of Khan Academy. However, he also identified issues specific to Indian cities- crisis response, planning and coordination of rescue efforts. He mentioned recent natural disasters like 2013 Uttarakhand floods and 2015 Chennai floods, and pointed out that crowd-sourced information collection could have significantly improved decision-making. He pointed out the importance of modelling and simulation of natural disasters to formulate response plans.

Finally, he made requests to the government for an open data policy, which would create an environment for sharing of data between agencies through common repositories, and such sharing would be incentivized. According to him, the biggest promise lies in tying big data to small data for offering personalized solutions.

The second speaker **Shivkumar Kalyanraman** is also an alumnus of IIT Madras. He obtained his PhD from Ohio State University and has had long stints in academia as faculty member. He is currently a researcher in IBM India Research Lab, Bangalore. He is the program director for special initiatives. The title of his talk was “**The Energy-Transportation Nexus**”. He started by pointing how India has a very poor record in the power and energy sector, where coal-powered



energy is causing carbon dioxide emission to continuously increase, resulting in toxic air. He stressed on the need of a network of renewable energy sources, and pointed out that energy and transport are interrelated. He pointed out that India has the potential to generate tremendous amount of clean solar energy, on a small area of land, through the use of rooftop devices. But there are challenges posed by dust.

He mentioned several initiatives of his group, such as solar-enabled air conditioning system, a system to balance demand and supply of energy and reducing power-cuts by shifting loads from peak time, using e-waste to create low-cost reusable batteries (Project UrJar). He mentioned initiatives of collecting pollution data from sensors covering the basin of river Ganga (Project GangaWatch), and similar initiatives for river Kaveri. Like the previous speaker, he also requested the government to promote open data and data sharing across agencies.

The third talk was by **Dr.Koyel Mukherjee** from Xerox Research Center, who talked about ride-sharing for urban mobility. She pointed out that the huge passenger volume challenge existing cab services. Manual scheduling becomes inefficient in the presence of last minute cancellations and change in pick-up and drop-points, and vehicle utilization is poor, resulting in added pollution and traffic congestion. Thus, the need of the hour for cab services is to develop intelligent and efficient scheduling algorithms. She discussed **Xshare**: a project by XRCI to build such a system where a shared taxi can re-route itself in the event of cancellation by passengers. The system includes a routing algorithm based on passengers' pick-up or drop-down points by which a taxi can work out the shortest path that it needs to take, while minimizing the waiting times of each passenger. Another part of the system is a billing system, where each passenger is provided a discount for every detour taken by the taxi to accommodate other passengers, however the total revenue of the taxi does not suffer. She discussed how passengers can be further encouraged to avail the ride-sharing schemes, such as by respecting their preferences in deciding sharing partners.



She also talked about a multi-modal transport plan which allows each passenger to plan a trip making most efficient and comfortable use of resources- shared taxi, public transport and walking-to provide last-mile connectivity. This project, in turn, can identify the demands of various routes at various times of the day, and help the government in planning and scheduling public transport system.

She requested from the government very detailed and accurate information about public transport services, such as timings, routes, delays and passenger volumes. She also suggested that the government can provide green points to commuters who use public transport, which would carry suitable incentives.

The fourth speaker of the session, **Bipin Kumar**, is the founder of Gaia Smart Cities. His group has come up with a data-driven plan for Raipur city, which has resulted in the city getting selected by the Government of India to be developed as a "Smart City". He started his talk by giving examples of cities in ancient civilizations, and how modern cities can borrow ideas from them. He noted that we have crossed the Information Age (1990-2010), and the current era (2010-onwards) is the Age of

multi-disciplinary integration, where all kinds of data should be analysed for improved delivery of services. He noted that in India, the various services and facilities in each city, such as security, transportation, water distribution, healthcare, energy etc are designed to work independent of each other. However, not only are the various services inside a city interdependent, even different cities are interdependent. He mentioned the need to come up with new designs for various public services taking such dependences into account.



Regarding the availability of data, he noted that while the Government of India has made huge volumes of citizen data available in the social and administrative sector through Data Portal India (data.gov.in), various departments provide the data in completely different formats, and there is a need to standardize the formats

to improve machine-readability, which in turn will facilitate app-development. In a question regarding which sectors should be prioritized for data-driven research, he said that while his personal opinion is water resources, people's opinions seem to indicate great concern for the transport systems.

Session 3: Technologies for Massive Online Education

This session was chaired by Dr. Ravi Kannan from Microsoft Research, India. It included two talks- by B. Ashok from Microsoft Research India, and Dr. Manish Gupta from Xerox Research Center, India.

The first talk was by **B. Ashok**, who described **Microsoft's Open Education Platform**. He started by



expressing concern about the quality of higher education of India, and its availability to students. Although there are over 33 thousand colleges and over 8 lakh faculty members available, only 18% students are able to avail higher education. Even otherwise, the examination systems in most colleges emphasize rote learning, faculty members are overworked, and most students are not motivated to learn any more than the bare minimum needed to pass exams and secure jobs. Although Massive Open Online Courses (MOOC) have made world-class study material

openly accessible, it was found that they are not well-received by college students in India, especially in second and third tier colleges. The reasons were found to be that students often find the instructors' accent hard to understand, and the content of such courses are often very different from the college syllabi in India.

To address these issues, Microsoft has come up with MEC: Massively Empowered Classroom, where teachers can edit course content from other sources, including MOOCs, and create personalized classroom lectures. They are found to solve the issues which make MOOCs ineffective in India. The system can also be used to find fine details about the students' learning pattern. Similarly, it is possible to gain detailed insights about teachers, and the system itself. Apart from standard college courses, this framework has also been used to provide vocational training to electricians and plumbers, e-courses to farmers and rural health workers, and also for raising awareness campaigns about sanitation and other issues.

Mr. Ashok mentioned another project performed by Microsoft: identifying school students who are likely to drop out. This was carried out in a particular district of Andhra Pradesh, and factors like the socio-economic backgrounds of students, and the infrastructure of the schools were taken into account for prediction. The idea was that, if such students can be identified, then special corrective measures can be taken. The exercise turned out to be particularly successful, and there are plans to replicate it for the entire state.

Dr. Manish Gupta from Xerox talked about the need to impart personalized education to the students. Indian students face problems with the existing frameworks based on massive open online courses (MOOC) due to their one fits all strategy. They find it difficult to follow the accent of the professors teaching the courses or even follow the languages in which these courses are taught. Dr. Gupta pointed out that learning from videos is taking the role the textbooks have done previously.





Therefore, it is important to provide features present in the text books such as table of contents, index of key phrases, note making, bookmarking etc. Xerox is contributing here by building technologies with these features so that students can learn from the videos as efficiently as from the textbooks. Their system automatically detects shifts in topic within a video lecture and can use those to create a table of content. The systems rely on speech recognition technologies and it is hard due to domain specific vocabularies involved in the videos of a particular subject. The technology they had developed also allows teachers to compile teaching material from multiple online course contents. Dr. Gupta highlighted the usefulness of learning from the videos as they help one to automatically assess the engagement level of students, identify difficult topics and revise the syllabus accordingly. An open challenge in education is in developing methodologies which will impart creativity among the students and videos personalized for students based on assessments.

Session 4: Legal Issues in Data Science

Unlike the other session which dealt with the technical aspects, this session aimed to look at legal issues regarding open data, especially in the administrative sector. This session was chaired by Tanusree Deb Barma, who is the Director of Information Technology and Biotechnology in the Government of Karnataka. She coordinated a discussion which involved Dr. Sony Pellisery from National Law School of India (NLSI), Dr. Pranesh Prakash from Center of Internet and Society (CIS) and Ms. Jamuna from BESCO, Government of Karnataka.



Dr. Pranesh Prakash started by remarking that in India, the laws regarding access to data are quite outdated. Even though the government has an open data policy and laws like RTI which encourage it to actively disseminate data, they are rarely followed. He observed that though the laws are outdated and need to be changed, it is possible to overcome legal hurdles through policies. Most data related to the social sector that have been hosted on Data Portal India need to be in universally accessible formats. He specifically mentioned two very important ethical issues regarding data that need to be taken care of through laws: privacy and discrimination. He said that being a democracy, India should not turn into a “surveillance society”, and should not allow arbitrary profiling of people by anyone, not even by the government. Not all kinds of data about individuals should be mineable. Research efforts should be made at anonymizing data, and privacy-preserving data analysis. Also, since data is not available from all people with equal ease, this may result in discrimination- people who can afford to publish their data quickly through devices like smartphones, may grab the attention of policymakers more than those who do not have such facilities, but are likely to be in more urgent need.

Dr. Sony Pellisery talked about Open Data from the perspective of social sciences. He asked and examined a question: “Do we have a good governance system for big data?” He admitted that social science has till now considered data from the social sector as “administrative data”, but in view of the recent abundance of such data, it must understand and study “big data”. This is different from



traditional administrative data because of its massiveness and unstructured form. He pointed out that various interest groups, with potentially conflicting interests, are vying for access to big data from social sector- citizen's group who want completely open data, business groups which want controlled data for their business interest, academic groups which are often neutral to the question of openness, and the government. In the atmosphere of such conflicts of interests, a financial architecture for data is needed.

As representative of the government, **Ms. Tanusree** pointed out based on her experience of administrative work in Tripura, that the government has access to all details of 37 crores of people through MGNREGA job cards. So much data has a lot of power, with potential for both use and misuse. She mentioned Mother and Child Scheme, where mothers' and babies' health are kept track of continuously, and child mortality immunization rate has improved from 70% to 99%. She also mentioned government initiatives in which the government has organized competitions and hackathons with social sector data.

Ms. Jamuna from BESCO pointed out that the electricity-supplying body has recently come up with smart initiatives like Restricted Accelerated Power Reform Program. She agreed that government has a lot of data in the social sector, and it should continuously innovate to improve usage of such data. The government should always make grounded or reasonable promises, and it must remain answerable to people.

Session 5: Data Science in Action

This session was chaired by Prof Y. Narahari from CSA Department, IISc. The aim of this session was to demonstrate, through real-world examples, how far the industry has utilized Data Science and Machine Intelligence to revolutionize its performance. The first talk was by Dr. Swami Manohar from Microsoft, who was one of the founder-members of the Simputer Project, India. He was followed by Dr. Krishnendu Chaudhuri of Flipkart Imaging Sciences, Dr. Babu Narayanan of General Electricity, and Dr. Ramesh Hariharan from Strand Life Sciences.



The first talk of the session was by **Dr. Swami Manohar**, who commenced by relating his experience of recording *harvest data* for village accountants at Panchayat level using the *Simputer*, an indigenous computing device developed by a group of four faculty members of the Department of Computer Science and Automation. He pointed out that nowadays, with the proliferation of smart devices which are equipped with various kinds of sensors, data acquisition from people is easier than

ever. He referred to various projects by the the Central government and Karnataka government, which aim at collecting and using data on a large scale to provide services to citizens. He highlighted the Aadhar project of the Government of India, and also various initiatives by Karnataka government bodies, such as BMTCL, BESCOM, KEB, etc. He mentioned initiatives by private industries- banks, hospitals, financial corporations etc, as also in the software sector, such as IndiaStack and Open API initiatives. He had a word of praise for Bangalore's startup culture, where several people with knowledge of Data Science and Machine Learning have started to work on impactful problems.

Among challenges, he emphasized that both the private sector and the public sector have a tendency of holding on to their own data, and to overcome this, he mentioned government policies such as NDSAP. While open data culture should be encouraged, well-defined laws are also needed on the issue, and illegal acquisition of data ought to attract severe penalties. He suggested data collaborations involving the private and public sectors along with academia, where they can exchange data and technical know-how to solve problems of public importance.

The second talk of the session was by **Dr. Krishnendu Chaudhuri** from Flipkart Imaging Sciences. He talked about several state-of-the-art applications of Computer Vision using Deep Learning techniques that have been successful in recent times. The examples included a system where a computer provides near-perfect textual description of any image

provided as input; predicts tags for any given image; achieves very high accuracy in face recognition from images where the persons in question appear in a wide range of poses, etc. He remarked that the accuracy numbers of these systems are so high that they are hardly inferior to human beings. Regarding Flipkart's services, he talked about automatic product recommendation for users based on visual similarities – if any user views a particular product, other products of the same kind which look similar to the original product are also suggested. Similarly, if a user provides an image as input to search a particular item that appears in the image, the system is automatically able to identify it from its catalogue.



Talking about the power of deep learning, Dr. Chaudhury pointed out that using a large combination of non-linear functions, it aims to follow the paradigm of the human visual system. He listed some new potential applications for governance- in security and healthcare domains. Regarding security, he talked about how face recognition can be used real-time in CCTV-based video surveillance, so that if a wanted criminal comes into camera range, the police can be alerted immediately.

Also, any suspicious person's image can be quickly compared to archived images to check if that person had ever been profiled by the police anywhere in the country. In the healthcare sector, he suggested setting up semi-automated kiosks in villages where there is a shortage of medical professionals, for machine-based diagnosis. He also suggested that data science can help city doctors to serve remote villages without physically going there. Also, city medicine shops would be able to collaborate effectively with local delivery facilities, to allow quick transfer of medicines to remote locations.

The third talk of this session was by **Dr. Babu Narayanan** from General Electric (GE). His talk focused on Industrial Internet of Things, also called as Industrial IoT. He talked about ongoing work in GE where they build large machines. They have a system to collect data from their machines using sensors and store this data in the cloud, on which analytics are carried out. By analysing this data, their goal is to create a digital twin, i.e. digital copies of industrial machines, which can exactly simulate the operations of the actual machines. Such simulation comes handy in studying the performances of these large systems.

He pointed out some applications where his team is using data analytics to solve real-life problems. First of them is Faster Medical Diagnosis where they applied Deep Learning algorithms to detect breast cancer from ultra-sound images. They



achieved reasonable accuracy in their model. They also use analytics for power plant fuel usage in their company, which has led to 1-2% savings. One more application which he talked about was a movement planner which resulted in 10% velocity improvement during train scheduling. Their analytics also come in handy in achieving nearly 3% output increase in wind power. But the challenges which are being faced in all these applications are volume and variety of data, cybersecurity, missing information, and lack of self-learning systems which necessitates human intervention. His long term vision is to build software defined machines.

The final talk of this session was delivered by **Dr. Ramesh Hariharan**— an adjunct faculty of CSA, IISc, and CTO of Strand Life Sciences, and another member of the Simputer Project. During his talk, he mainly focused on the problems of data analytics in the healthcare domain. One of the key problems which he pointed out was the availability of data. The collection of healthcare data at a large scale is very expensive. In the human body, there



are tens of trillions of cells and out of them, any one of them may turn out cancerous and to detect this cell is like finding a needle in a haystack. The reason is there can be a large number of combinations of these genomic patterns and to traverse over all of this combinatorial space is computationally expensive. Apart from being expensive, the availability of prototypes of each of these combinations is hard to collect, in the first place. He proposed India to be the ideal place for collection of such data, courtesy large number of patients here, but for that he proposed that the government should come forward and join hands with industry and academia to facilitate this. With the data available right now, his team has developed analytics models to detect early cancer and their model turned out to be successful in 50% of the cases.

Session 6: Panel Discussion

At the end of all the talks, there was a panel discussion where members of the audience had several questions to ask, or requests to make, to the Karnataka government. **Ms. MaulishreeAgrahari** responded to these questions, as the government representative. Some of the main points raised are mentioned below:

Openness of government in participating in this endeavour:

Government departments are not very keen to interact with the private sector. There is a need to change the mindset of government to see private sector as partners. Such change of mindset is required in all aspects, and then only something fruitful can be reached.

Open data is important, but there is a need for an open set of problems faced by the government also

This opinion got a great support from everyone. There are a lot of problems in the social sector but we should identify and prioritize them carefully and make a set of them so that



leaders of technology can take up these problems and solve them. It may help if we can concretely define pilot projects.

There are constraints enforced by NIAC on the education system.

What type of graduates should we

produce and how the curriculum needs to be changed to carry out these goals?

This was mostly the content of the first edition of this workshop series. IISc is taking a leading role in restructuring the course curriculums in undergraduate education in Karnataka. There are nearly 300 engineering institutions in Karnataka and they are actively engaged in research activities to produce well-trained graduates.

Data Science in Karnataka

Karnataka is a leader in data science education especially in Bangalore because of the presence of a number of MNCs here. The focus of their education is to look into the



problems of tomorrow.

Sharing of data from the point of citizens is very sensitive. If government promises confidentiality to citizen's data, should they trust and share the data?

Yes, this is a very critical issue and the government is sensitive towards the confidentiality of public data and data sharing. But of course, there are no disclaimers, government can put on this.

Special Remarks by the Co-chair

The Co-chair of the Machine Intelligence Task Group, **Mr. VenkateshValluri**, Member, KJA lauded the efforts carried out by government, academia and industry in conducting this workshop. The main motive of this task force is to use the knowledge of Data Science to enhance the living standards of the society including all sectors like water management, electricity, waste management and healthcare. The massive challenge which comes with this is the scale, given the dense and diverse population in the country. It is pretty much clear that technology helps, but questions remain about the best possible way to use it. Can we have data for this? How will technology help in using this data to carry out the goals? The major fields of study are Data Science, Data Analytics and Data Acquisition. Regarding the first two, he pointed out that in India, IISc is the best place to develop the necessary technology. Regarding the third, he endorsed mass deployment of sensors in public places to collect data and using technologies like IoT for rendering such the data so collected. He urged the government, academia and industry to come forward and collaborate together to make this possible.



Commitments by the Industry

The corporate bodies present expressed their desire to conduct pilot projects with the government, in different areas. Microsoft representatives, who had presented their work on open education platforms, expressed willingness to make their platform available to schools in Karnataka. Xerox offered to do pilot projects in transport systems and education sector, and IBM in energy analytics. Google offered to share their data and computing platforms for these projects.

Closing Remarks

Dr. Maulishree Agrahari extended her gratitude towards all the participants of this workshop. As a concluding remark, she focused on the formation of a concrete group which aims at sharing of data between government and academia/industry. She also urged the



other two stakeholders to help government in resolving the challenges faced. She encouraged the idea of pilot projects with private partners.

Recommendations

1. It was felt that Government needs to embrace the Open Data culture. Open Data Culture means data is readily available for use by third parties. Government owns many useful data sources. Govt. should endeavour to make data available for use by citizens and business within reasonable restrictions.
2. Data Science has an important role in Smart cities. Any Smart city should have ability to collect and make available the data which could be used by 3rd parties or city administration. Government role should be to create appropriate policy frameworks and Data Infrastructure, possibly leveraging NIC, to facilitate smart city.
3. There are several legal issues related to Data Sharing, Exploitation and Privacy. Government should engage with legal experts to come up with a suitable policy.
4. One could emulate initiatives like Open Data SF, or Open Data London for Bengaluru to start a culture of Open Data.





Annexure-I-List of Participants

Dignitaries

1. Dr. K. Kasturirangan, Chairman, Karnataka Janan Aayoga
2. Ms. Manjula V, IAS, Principal Secretary, Dept. of IT, BT & ST, GoK

TG MI Members, Speakers, Representatives and Participants

3. Mr. Venkatesh Valluri, Co-Chair, TG-MI
4. Prof. Chiranjib Bhattacharya, Co-Chair, TG-MI
5. Dr. Mukund Rao, Member, TG-MI
6. Prof. Ramesh Hariharan, Member, TG-MI
7. Prof. Ravi Kannan, Member, TG-MI
8. Mrs. Tanushree Deb Barma, IAS, Director, ITBT, Gok
9. Prof. Amrutur, Professor, IISc
10. Prof. Chandrashekar, Associate Professor & Associate Dean (Academics), IIIT-B
11. Dr. A. Maulishree, Member Secretary, TG-MI
12. Mr. R. R. Ranganath, PCW, DG & Inspector General of Police office, GoK, Bengaluru
13. Mr. Kumar. S, Head, KSP Data Centre, GoK
14. Ms. Jamuna, AGM, BESCO, Bengaluru
15. Mr. Shivkumar, DGM, BESCO, Bengaluru
16. Ms. Koyel Mukherjee, Senior Research Scientist, Xerox Research Center
17. Mr. B. Ashok, Microsoft
18. Mr. Manish Gupta, Vice President, Xerox Research Center
19. Dr. Pranesh Prakash, Policy Director, Center for Internet and Society (CIS)
20. Mr. Krishnendu Chaudhury, Principal Scientist, Flipkart
21. Mr. Babu Narayan, GE, Industrial Internet
22. Dr. Swami Manohar, Microsoft
23. Dr. Shivkumar Kalyanaraman, Research, IBM, Bengaluru
24. Mr. Bipin P. Kumar, Gaia Smart Cities
25. Mr. Sathish, Faculty, CSA, IISc
26. Mr. K. Gopinath, Faculty, CSA, IISc
27. Dr. P. Ramakanth Kumar, Dean Academics, R.V.College, Bengaluru
28. Dr. H.N. Narasimha Murthy, Prof. & Head, Dept of Mechanical Engg, R.V. College, Bengaluru
29. Dr. N.K. Cauvery, Prof. & Head, Dept of Information Science & Engg, R.V. College, Bengaluru
30. Dr. G. Shobha, Prof & Head, Dept. of Computer Science and Engg, R.V. College, Bengaluru
31. Dr. M. Uttarakumari, Prof. Dept of Electronics & Communication Engg, R.V. College, Bengaluru
32. Dr. Viraj Kumar, Professor, P.E.S. Institute of Technology, Bengaluru
33. Mr. Subramaniam K.V, Professor, P.E.S. Institute of Technology, Bengaluru
34. Prof. T. N. Nagabhushan, Professor/Chief, Sri Jayachamarajendra College of Engineering, Mysuru.



35. Mr. Suman Dev .H.A, Assistant Professor, Malnad College of Engineering, Hassan
36. Mr. H.C. Vijaylaksmi, Associate Professor , Sri Jayachamarajendra College of Engineering, Mysuru
37. Dr. Padma Srinivas, Professor, Sri Jayachamarajendra College of Engineering, Mysuru.
38. Mr. Ankit Jauhari, Research Student, IISc
39. Mr. Supratip Ghose, Research Fellow, CSA, IISc
40. Mr. Srijith. P.K, RA, IISc
41. Ms. Aarthi, Senior Manager, Educational Initiatives
42. Mr. Vanya Rakesh, Programme Officer, Center for Internet and Society
43. Ms. Shreyas.G.S, Student, IISc
44. Mr. Adway, Research Scholar, IISc
45. Mr. Aashish.V, Student, IISc
46. Mr. Satya, Student, IISc
47. Ms. Divya, IISc
48. Mr. Himanshu Sinha, IISc
49. Ms. Sweta Sharma, Student, IISc
50. Ms. Sakina Bohra, IISc
51. Dr. B.S. Padmavathi, Convenor, KJA TG-MI
52. Mr. Deepak, SRA, KJA