



Creating Jobs. Changing Lives

Introduction to Data Driven Decision Making & AI for Good Governance

**as part of
Foundation Course
for ISS & IES at**

**Dr MCR HRD Institute,
Hyderabad**

Wadhvani Institute of Technology and Policy
Wadhvani Foundation



WITP Team & Faculty



Prakash Kumar,
CEO, WITP



Kamal Das,
Dean, WITP



Sudhir Aggarwal
Director & Faculty
Digital Transformation, WITP

- Former IAS officer of 1985 batch.
- Worked with central and State Governments as well as PSUs.
- Worked with Cisco, Microsoft, and Deloitte.
- First CEO of Goods and Service Tax Network (GSTN), Govt. Of India
- B-Tech, IIT Kanpur; MPA, Public Finance, Saitama University, Japan

- Rated the best senior faculty in Data Science 2021
- Kaggle Grandmaster, Ranked in the top 0.1% Data Scientists (out of over 1.8 lakh) in the world
- At Manipal-Jigsaw, Program Director for four courses ranked in the top 5 Executive Analytics Courses with IIM Indore, University of Chicago, SDA Bocconi Asia.
- Over 20 years of experience; Worked with Moody's, Barclays, Lehman, S&P, JPMorgan
- BE, BIT Mesra; MBA, IIM Ahmedabad

- Over 35 years of Industry experience, working with government ecosystem for 2 decades
- Adjunct Faculty with RGIPT
- 3 Copyrights in credit and 3 papers published
- Named 'IT Industry Leader' in 2010 by CSDMS
- Worked with Deloitte, Microsoft, Thomson Reuters, Sify Tech, Oracle, IBM, and HCL
- Created Government business for Sify Tech
- Graduate from DU and Alumni ISB, Hyderabad

Agenda

February 22nd (Wednesday) :
“Data-Driven Decision Making“

Group A: 9:30 AM to 11:20 AM
Group B: 11:40 AM to 1:30 PM

February 23rd (Thursday):
"Applications of Artificial Intelligence for Good Governance"

Group B: 9:30 AM to 11:20 AM
Group A: 11:40 AM to 1:30 PM.

Learning Outcomes

After the training participants will be able to:

After the training participants will be able to:

- Understand the applications of different emerging technologies such as Analytics/Data Science, AI-ML-DL, CV and NLP.
- Identify work-problems where use of selected emerging technologies can provide a solution
- Understand the nuances of data collection and feature engineering to generate better insights
- Conduct detailed visual analysis on data to find patterns and make data driven decisions
- Appreciate the limitation of emerging technologies and in which areas it may not be prudent to use these technologies

Digitisation

Digitization is conversion something that's done in an analogue or manual way into digital one

Vs



Digitalisation

Digitalization is what happens when one applies technology to improve a process that was otherwise done manually

Vs



Digital Transformation

Digital transformation is a business model change that occurs because of the rise of technology.

Emerging Technologies for Digital Transformation

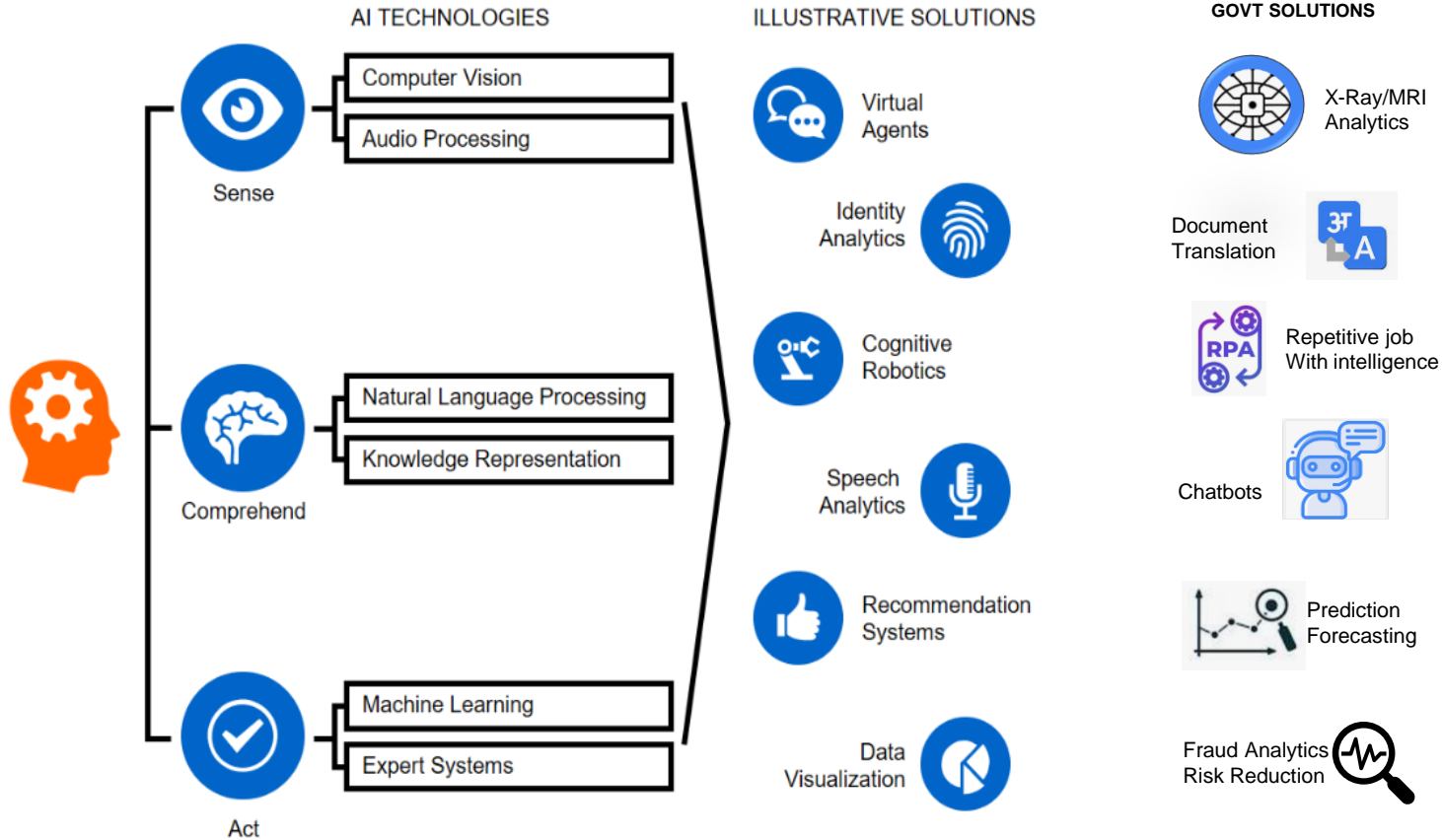
Departmental Applications

Infrastructure/Common Platform

- **Transactional Data**
- **Ready to use Emerging Technologies on large scale**

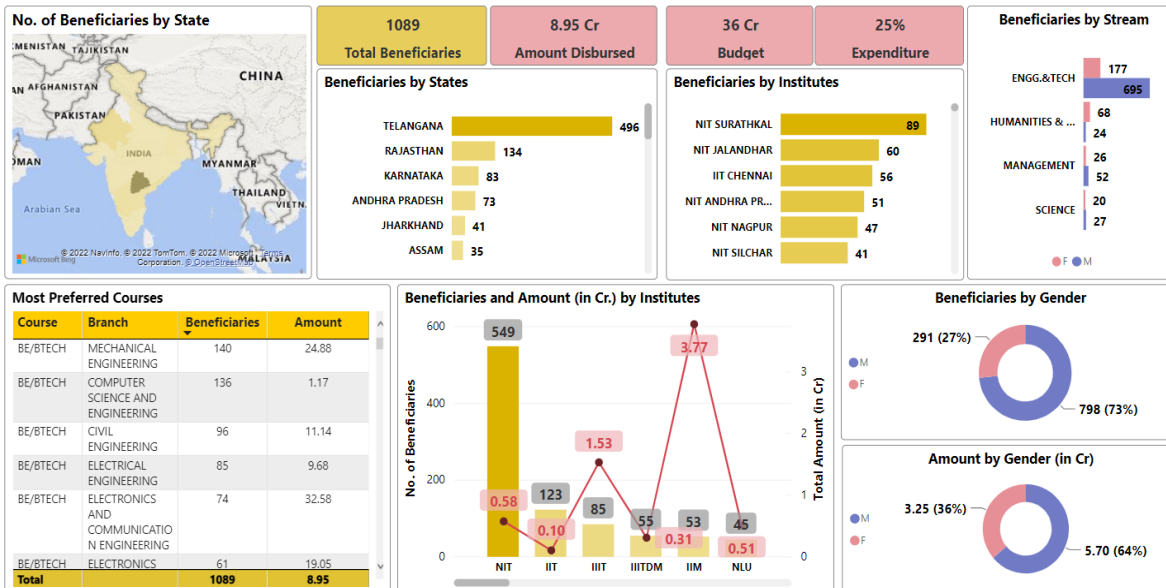
Major Emerging Technologies

✓ Analytics/Data Science	✓ AI, ML	✓ Computer Vision (CV)
✓ Natural Language Programming (NLP)	Big data	Cloud
Blockchain	5G/6G	AR/VR/ MR/Metaverse
IOT	3D Printing	Drones
Quantum Computing	Nano Technology	And many many more...



Major areas of Application of ET

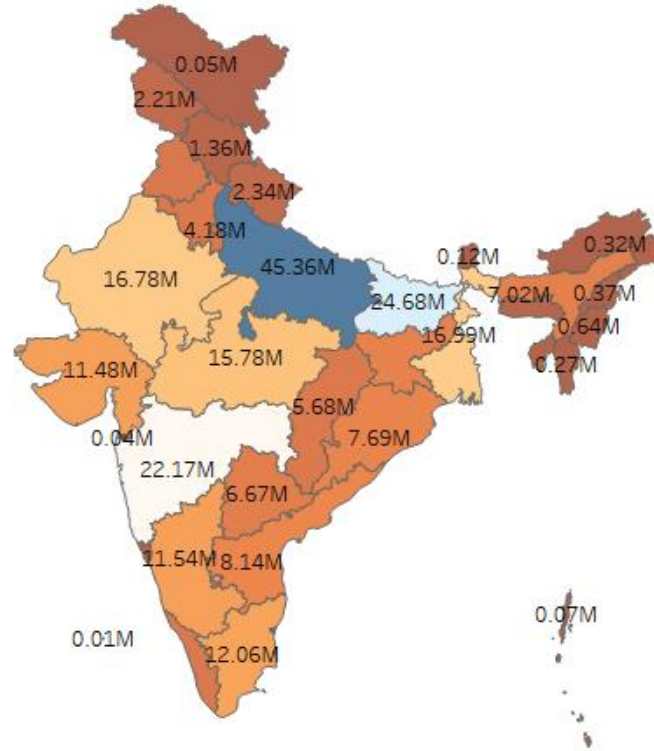
- For fraud analytics and risk reduction
- To predict/forecast
- Improve efficiency and optimisation of resources (logistics, traffic planning etc.)
- Measurement of impact of policies and programs
- Ability to do repetitive job with intelligence
- Where devices can see and interpret vision like a human being
- Where devices can understand oral instruction or converse with humans
- Where faster pattern analysis is required
- Where record needs to be made immutable etc.



- Dashboard is the first step
- This should lead to evidence based decision making, after slicing and dicing the data.
- Correlating data from other sources
- Measuring impact of policies in terms of policy objectives
 - Reduction of pollution
 - Reduction of malnutrition
 - Reduction of stunting etc.

1. Data Driven Decision Making

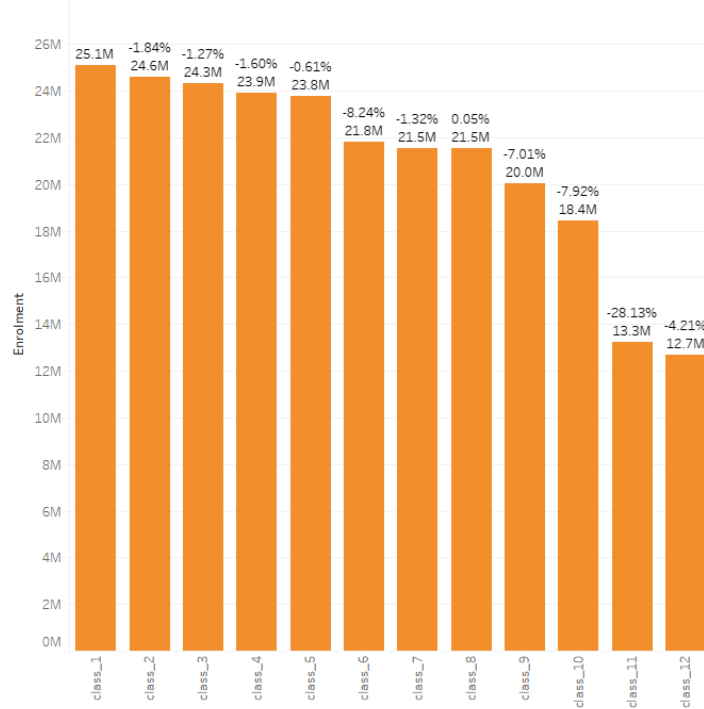
India School Enrollment Data



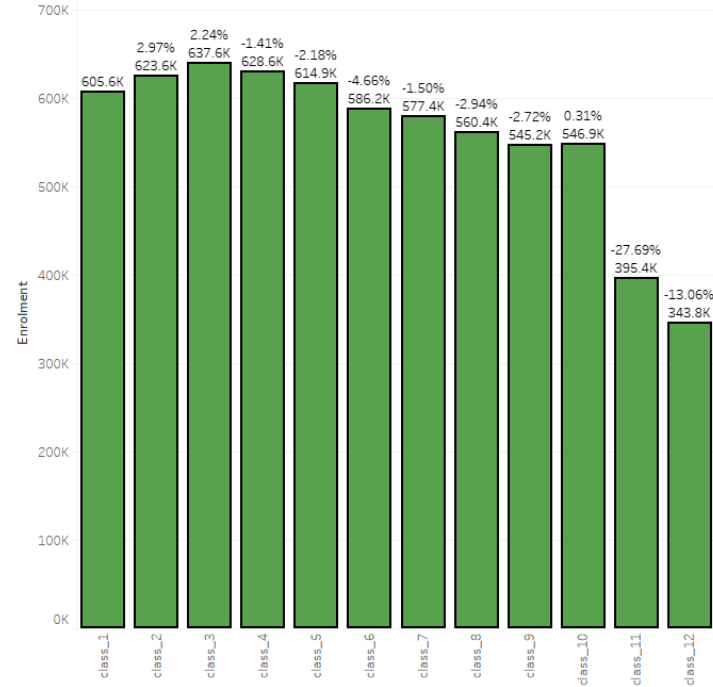
Is School Drop out higher in India or Telangana?

Telangana : Lower Dropout till Class 11, higher in Class 12

India



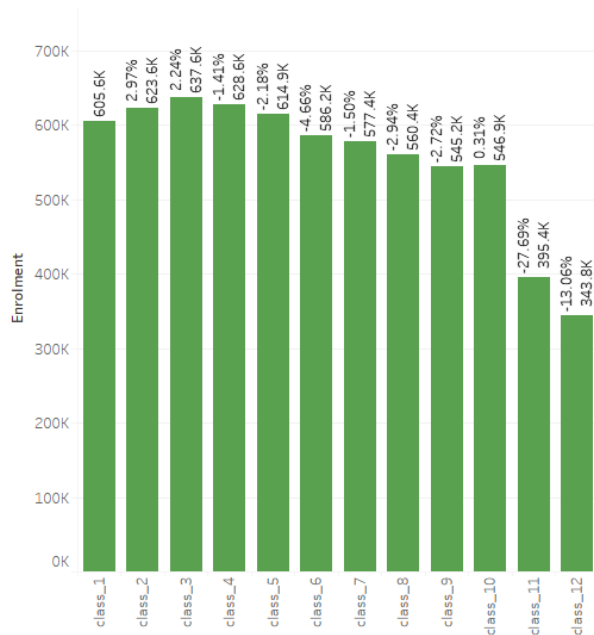
Telangana



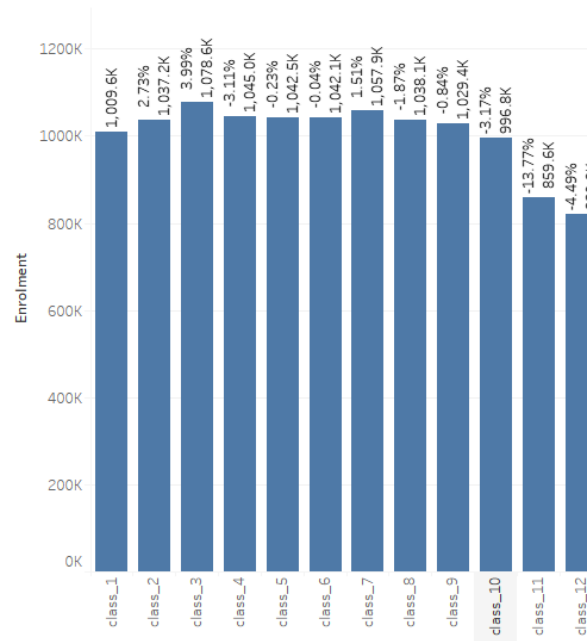
Is School Drop out: higher or lower in Telangana ,Kerala or Tamil Nadu?

Kerala & TN have lower dropout than Telangana

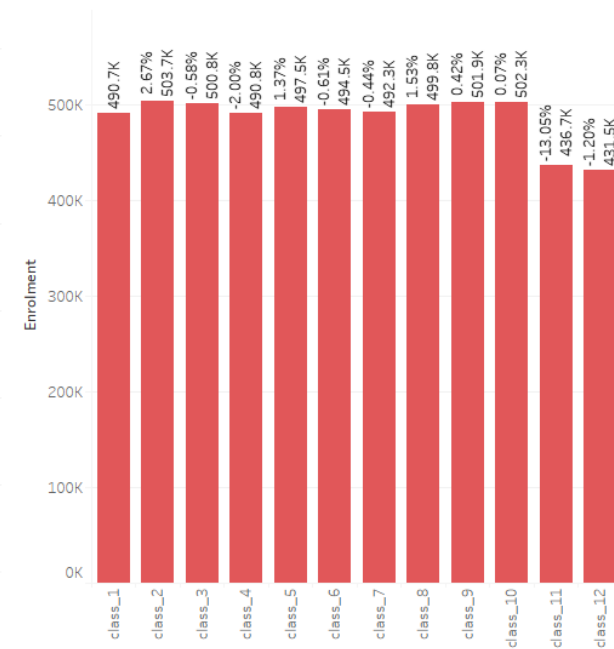
Telangana



Tamil Nadu



Kerala



Defining Analytics



Raw data



Cleaned data

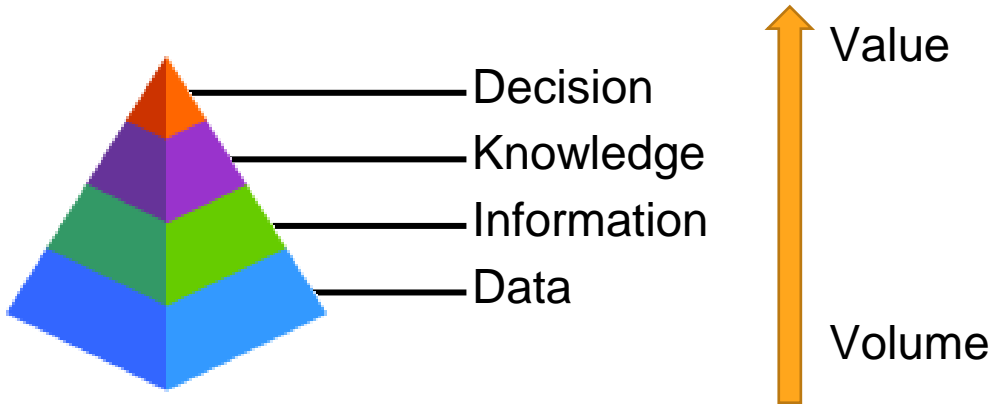


Information

Data Science or **analytics** is the science and art of **analyzing raw data**, converting it into **information** to solve a **problem**.

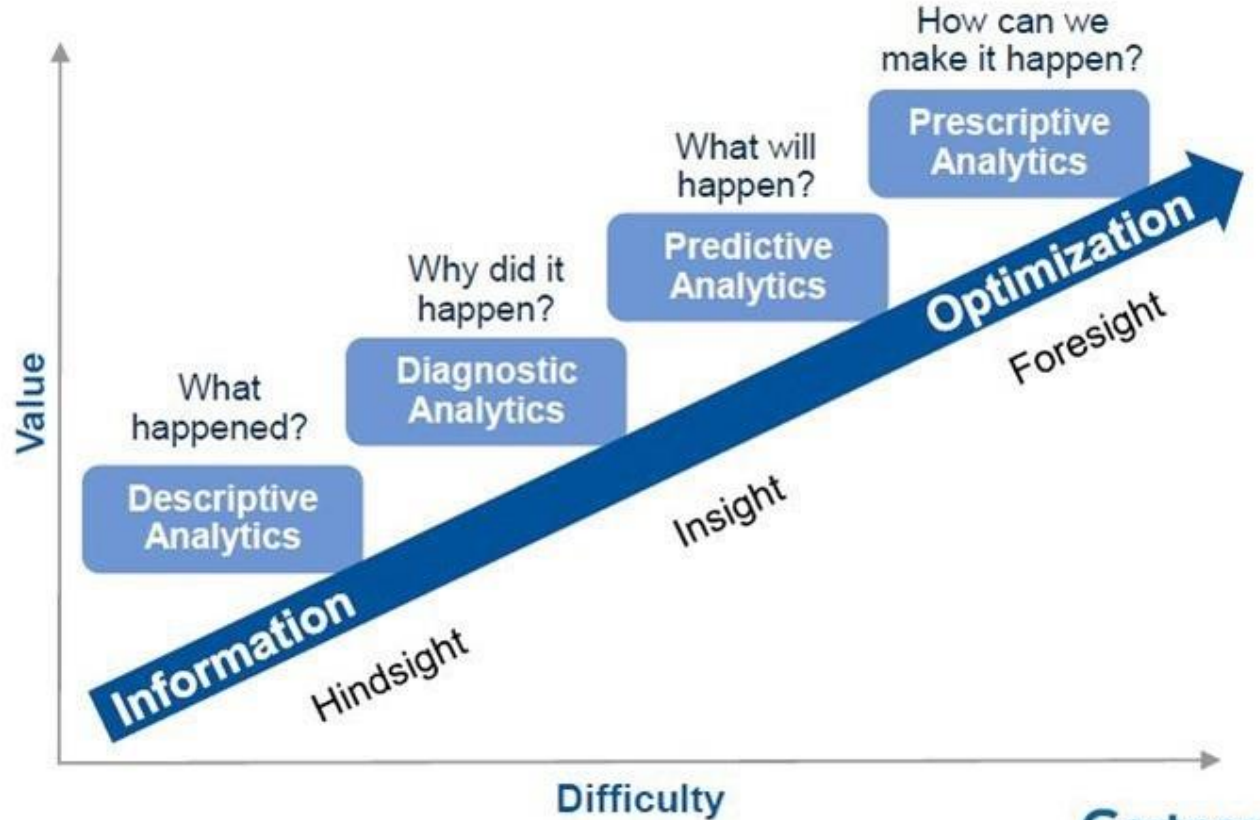
Purpose of Data Science

- The purpose of Data Science to convert the **volume** of data into useful **value** through analytical process.

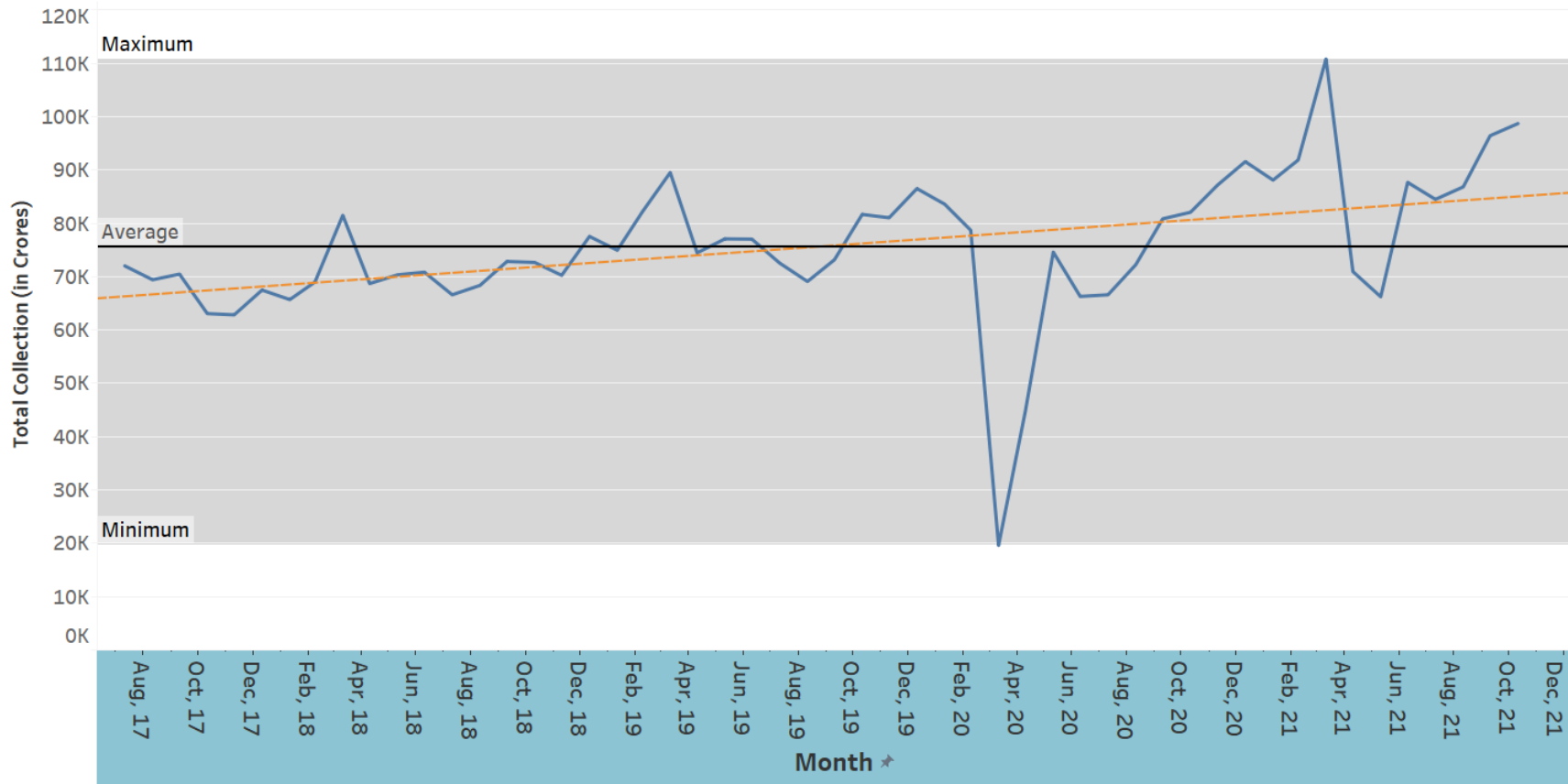


Analytics Maturity Model

Analytics maturity model



Monthly GST Collection



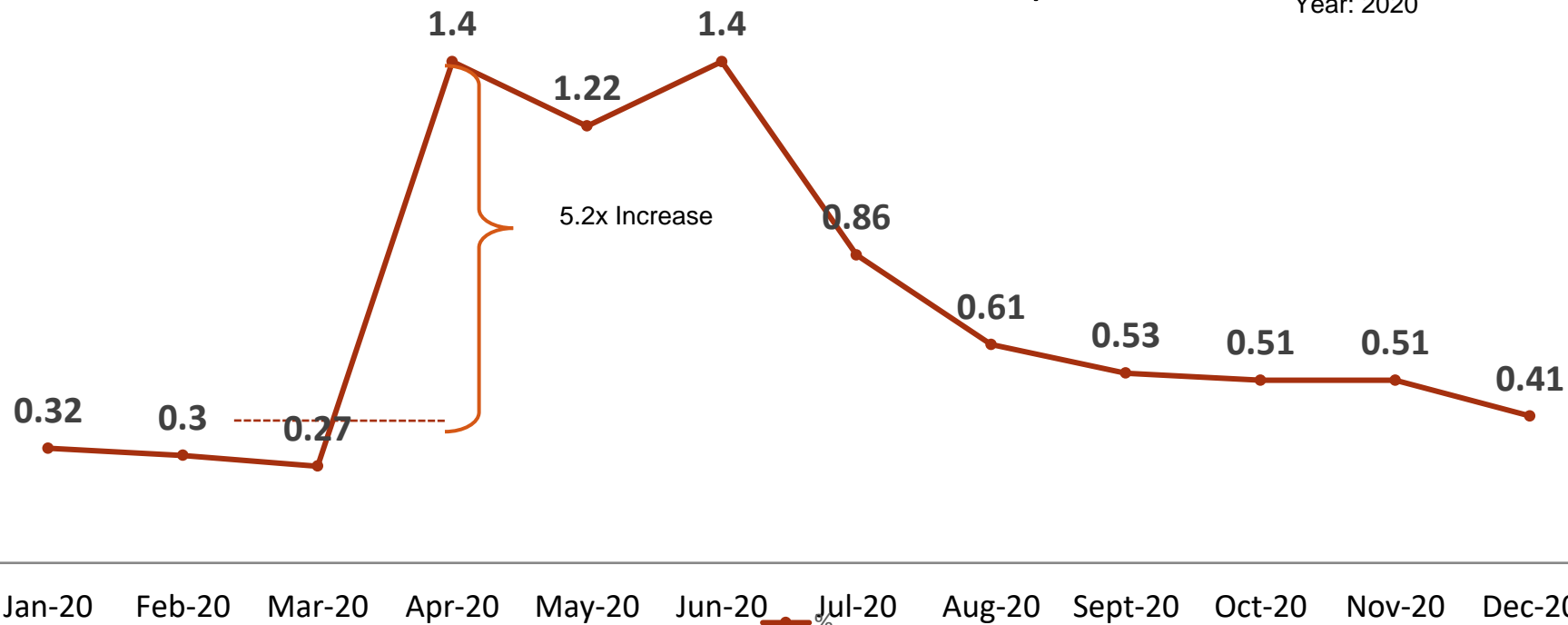
Analytics maturity model

	Descriptive	Diagnostic	Predictive	Prescriptive
	What happened?	Why did it happen?	What will happen?	How can we make it happen?
Monthly GST collection	What was the tax collected during different months	Was it higher/lower than expected/history? Why?	With three years data can we predict for next 3 months or year?	What should we do to increase taxes or reduce under-reporting?

Temporal pattern of malnutrition

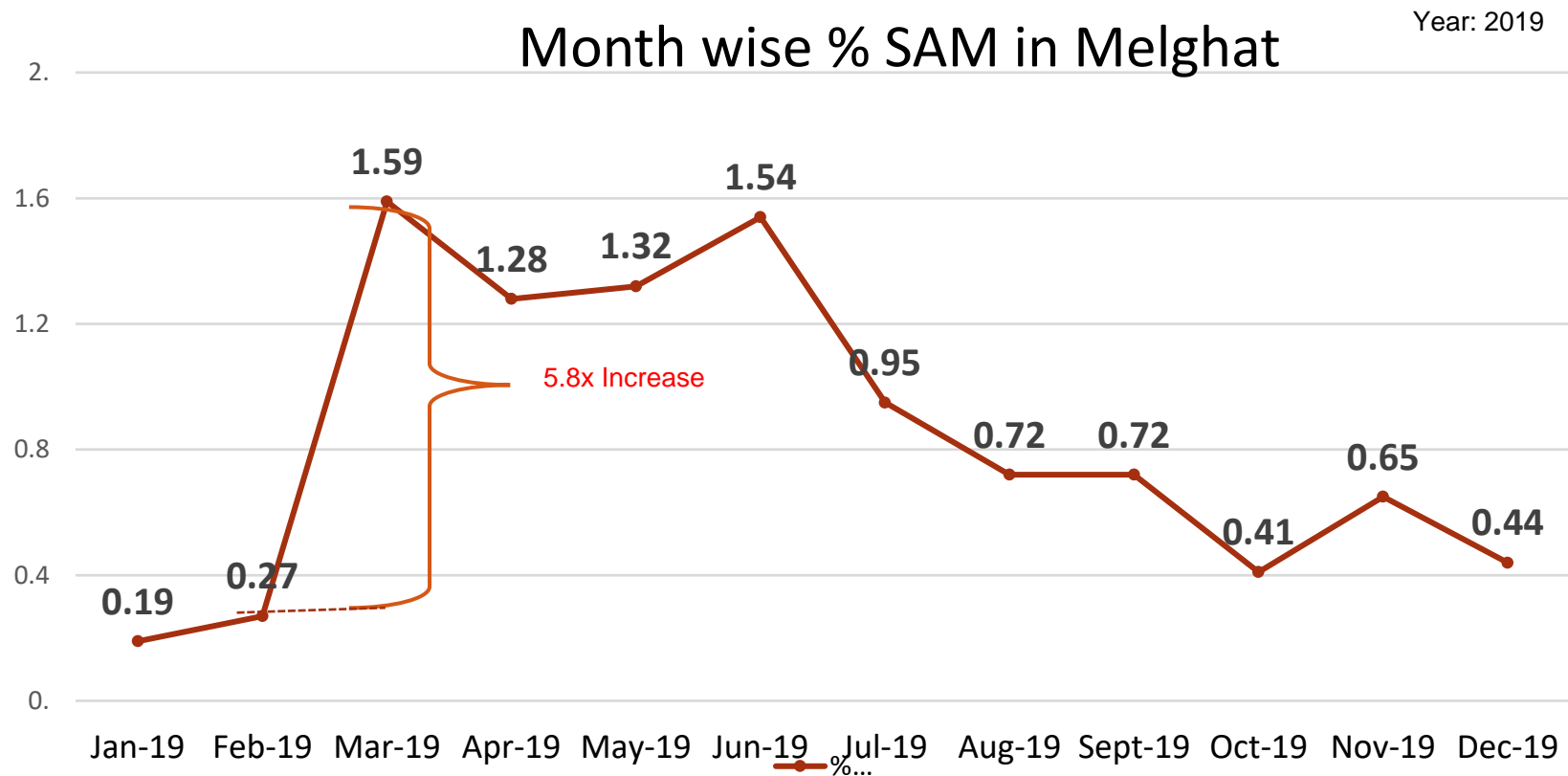
Month wise % SAM in Melghat (SAM : Severe Acute Malnutrition)

Year: 2020



Source of Data: **District MPR**

Temporal pattern of malnutrition (Contd.)



Source of Data: **District MPR**

Learnings from Field Visits



Our questions/thoughts were the following:

- Do we know where children are going exactly every year?
 - No institutionalised mechanism to capture migration, map corridors and therefore target prevention of distress driven migration
- Can we map these places and think about portability?
- Expecting people to register at institutions is a tall order. Demand side needs to be matched with supply side.
- Grievance redressal mechanisms, communication systems

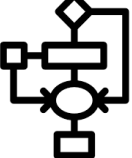
Genesis of Maharashtra MTS

The Maharashtra Migration Tracking System (Maha MTS) was piloted in 2021, in reciprocation to instructions of then Chief Secretary Shri. Ajoy Mehta to Principal Secretary DWCD Mrs. Kundan, which aims **to ensure the portability of Women & Child Development-related schemes and services to the internal migrants.**

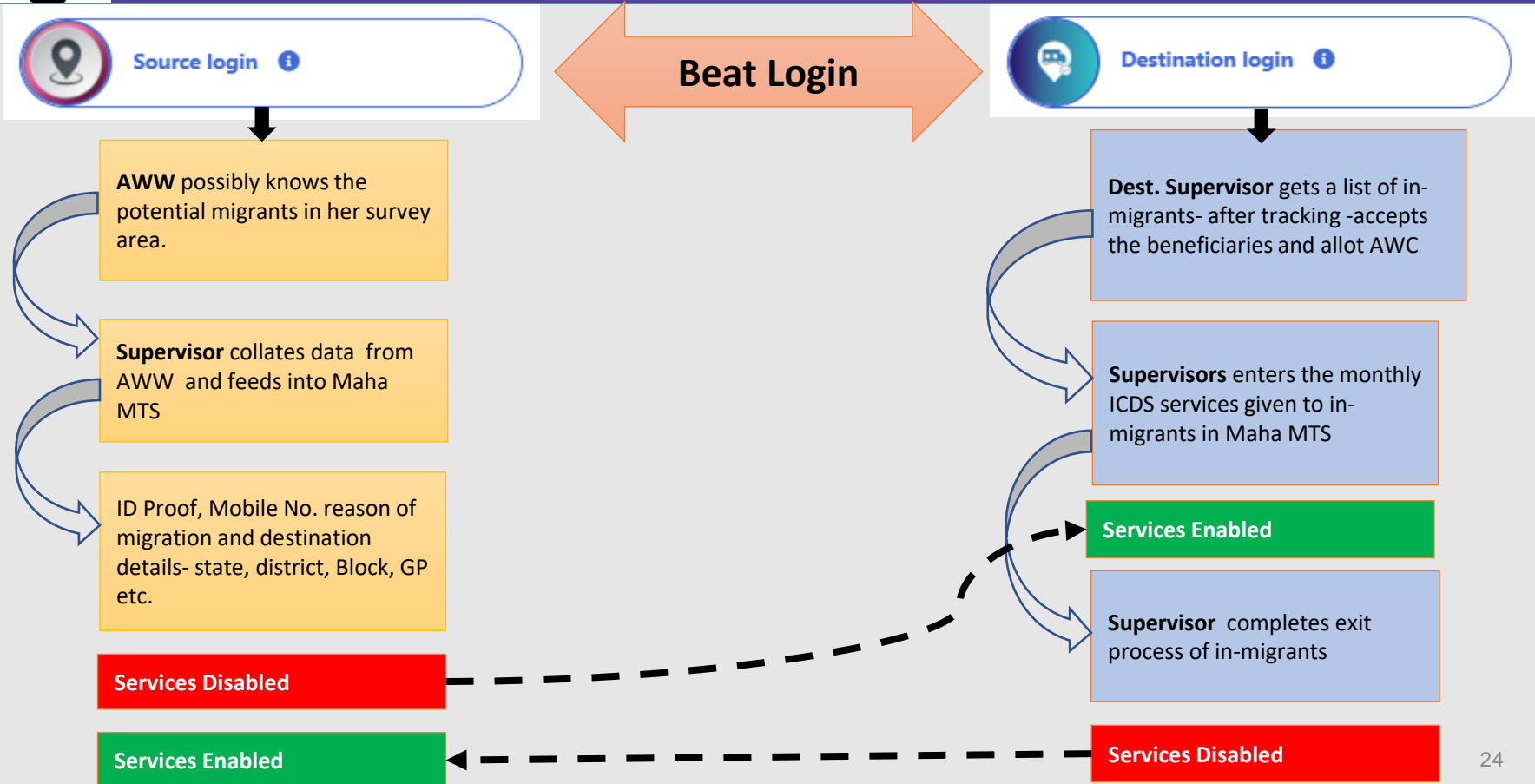
In order to fast-track the work of MTS, Principal Secretary DWCD constituted a state-level committee with ICDS Commissioner and Sr Consultant RJMCHN mission as chairperson and Member Secretary, respectively. The multifarious inputs were sought from various field officers of ICDS, DWCD, as well as UNICEF, IIT-Mumbai, and IIPS-Mumbai.

The development of software was done by RJMCHN Mission with support from IT Vendor, which is funded by UNICEF- Maharashtra.

The total cost of the project is Rs. 7 lakh (half for the pilot phase and half for expansion to 36 districts) and the ownership of the software source code rests with DWCD.

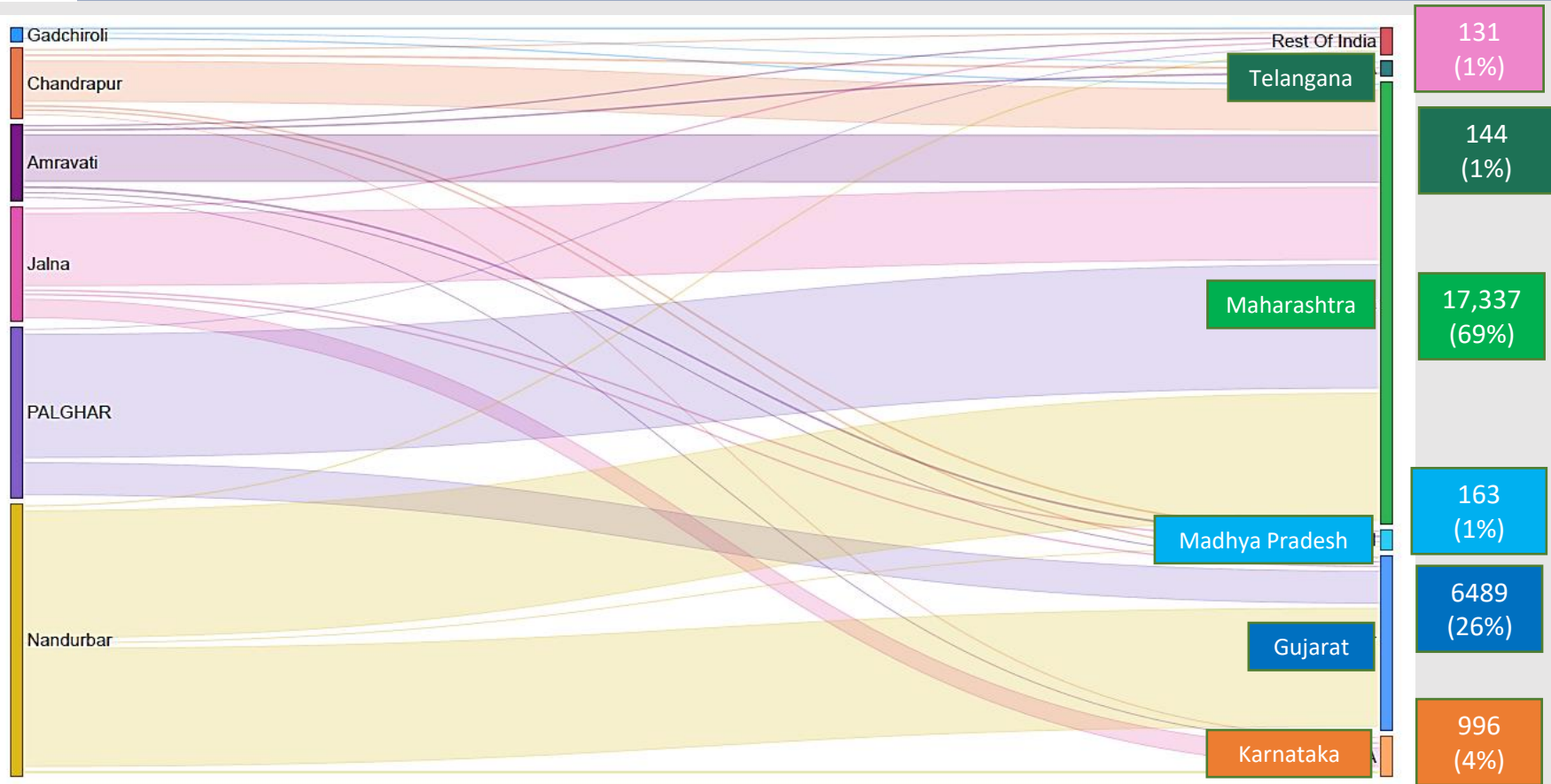


Workflow of Maha MTS





Learnings from Pilot in Source Districts: Migration Corridor



Source: Maha MTS Data, 10 June, 2022

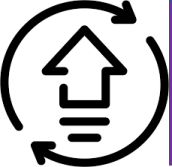


Merits of Maha MTS

1. Mapping Migration Corridor: Mapping of high out-migrant and in-migrant AWC, GP, ICDS Beat, ICDS Projects, and Districts - wise

1. Causality: Reason-wise seasonality
2. Scope of ensuring 100% portability of ICDS and ICPS services- **Pro-active Approach**
3. Enlisted potential remedial measures to halt migration
4. Interdepartmental Convergence (Indicators captured)

Sr. No	Department	Target Population	Services
1	Education	6-14 migrating children	School enrollment
2	Public Health	0-6 years children, Pregnant women, and Lactating mothers	Immunization
3	MGNREGA	Migrants	Focus on livelihood opportunities
4	Labor	Migrants	Shelter, Drinking water, and sanitation
5	Food and Civil Supply	Migrants	Ration
6	MSRLM	Migrants	Focus on livelihood opportunities



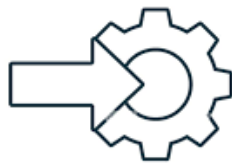
Maha MTS version 2.0

Maha MTS Version 1.0	Maha MTS version 2.0
6 Pilot districts	Statewide expansion of MTS for all 36 districts
End-to-end rural migration tracking only	End-to-end rural and urban migration tracking
Only Potential ICDS and ICPS migrating beneficiaries captured	Household-wise enumeration of potential ICDS and ICPS migrating beneficiaries (Duplication logic)
Data entry on the web-based portal only.	Mobile-App-based data entry (both offline data entry option available)
Enumeration of ICPS migrant beneficiaries only	End-to-end tracking, accepting and ensuring portability of services to staying back 0 to 18 years
No return mechanism	Enabled return mechanism
	Registration of In-Migrants by Destination



Way Forward

Integration of Maha MTS into Poshan Tracker



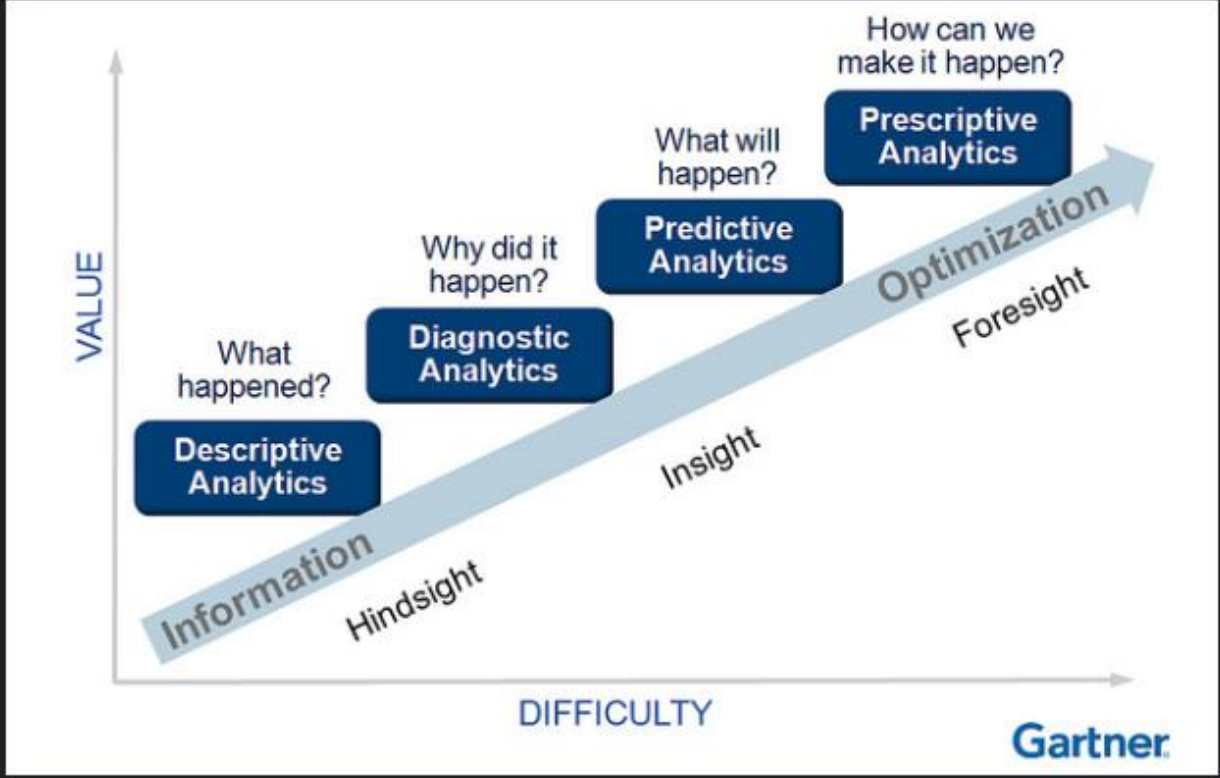
INTEGRATION



Reactive
Approach

Proactive
Approach

Analytics maturity model



Test Your Learning

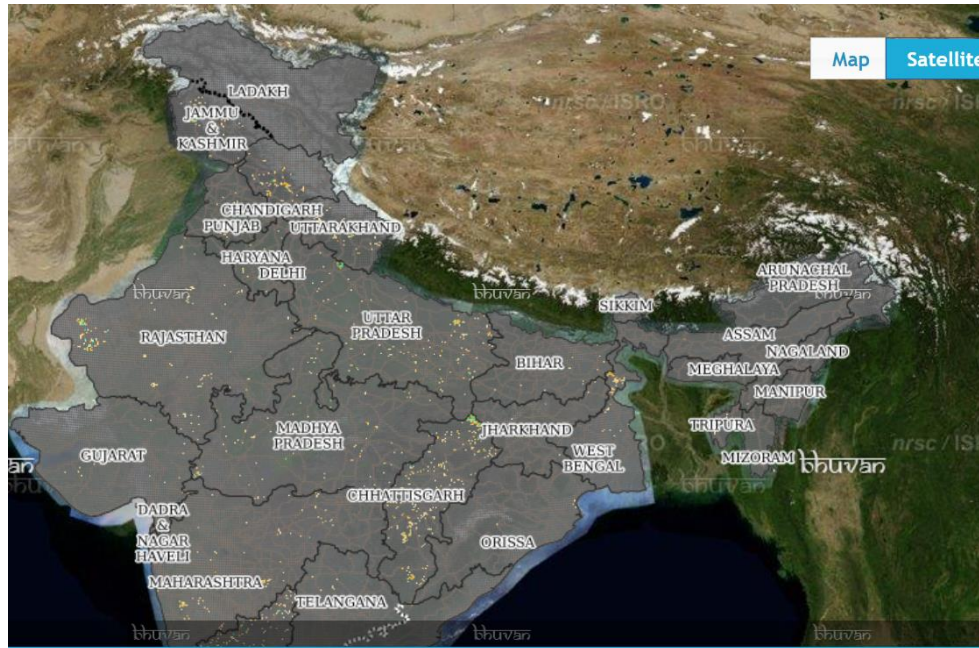
- A. Descriptive
 - B. Diagnostic
 - C. Predictive
 - D. Prescriptive
1. What percentage of PMJJBY insurance holders will renew the insurance next year?
 2. Percentage of PMJJBY insurance holders last year who did not renew the insurance this year
 3. How can we increase renewal rate of PMJJBY?
 4. Why is the Percentage of PMJJBY insurance holders last year who did not renew the insurance this year higher or lower than industry average/ last few years? What are the reasons for this?

Solution: Test Your Learning

- A. Descriptive
 - 2. Percentage of PMJJBY insurance holders last year who did not renew the insurance this year
- B. Diagnostic
 - 4. Why is the Percentage of PMJJBY insurance holders last year who did not renew the insurance this year higher or lower than industry average/ last few years? What are the reasons for this?
- C. Predictive
 - 1. What percentage of PMJJBY insurance holders will renew the insurance next year?
- D. Prescriptive
 - 3. How can we increase renewal rate of PMJJBY?

Leveraging work done by NRSC of ISRO

[ISRO's Geoportal](#) | [Gateway to Indian Earth Observation](#) | [Mobile Postal \(nrsc.gov.in\)](#)



Discussion

Question:

What analytics can we use to make the work of MGNREGA officials more efficient?

Discussion

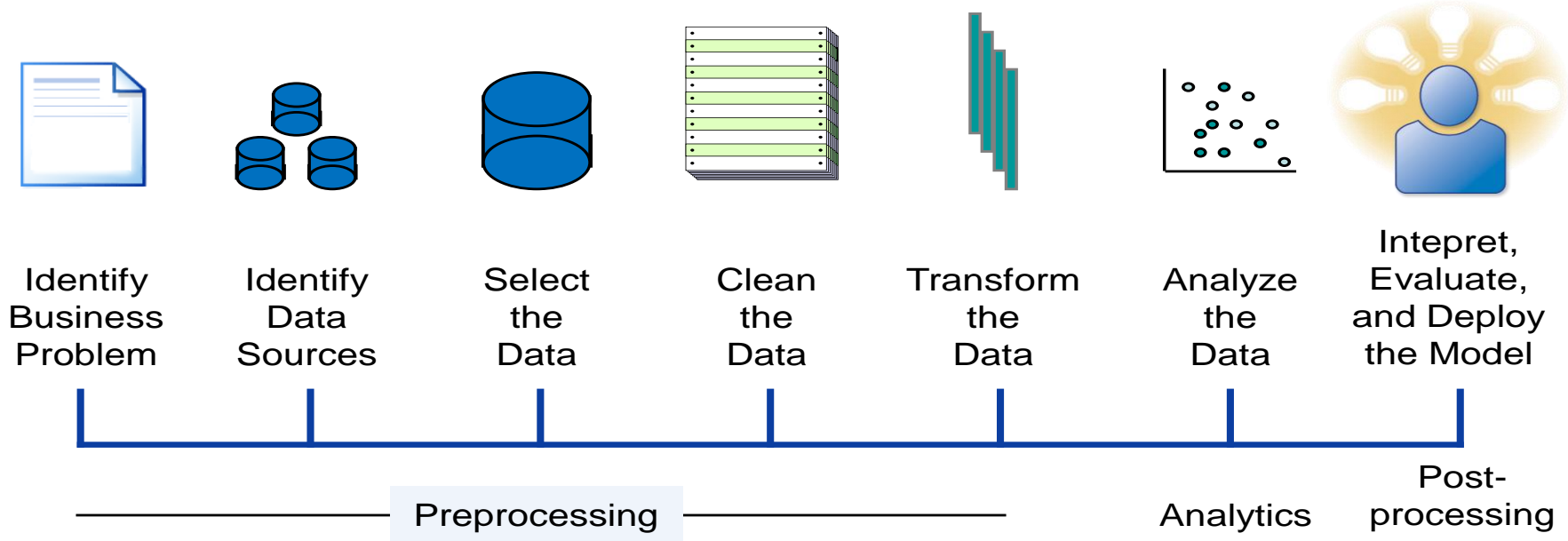
Problem definition

- Alert system for fresh work application

Approach

- Set pre-defined rules for the desired number of each work item and if a certain number of similar work have already been done in the village/area, raise an alert.

Process Model



WITP How could Data Analytics help Government Departments?



Keeping track of program implementations



Making data driven policy interventions and decisions



Analyzing root cause of exceptional/under performance



Fraud/Crime Detection with data such as Income Tax returns, GPS location, CCTV footage etc.

Keeping track of program:

All departments: for tracking implementation; target achievement

Education: Sudden dip in school attendance/ pass %

Health: Sudden increase in number of communicable disease in a particular area

Making data driven policy intervention:

Tax Dept.: Handling refund

Social welfare: Refinement of eligibility condition

Root cause analysis

All depts.: Grievances at a particular point in processing of applications.

To measure Performance of officers

Fraud Detection

GST: Refund

Welfare Departments: Ineligible people

Scholarship: Ineligible student

Pension: Ineligible or dead person

Important Points:

- Data source
- Data quality

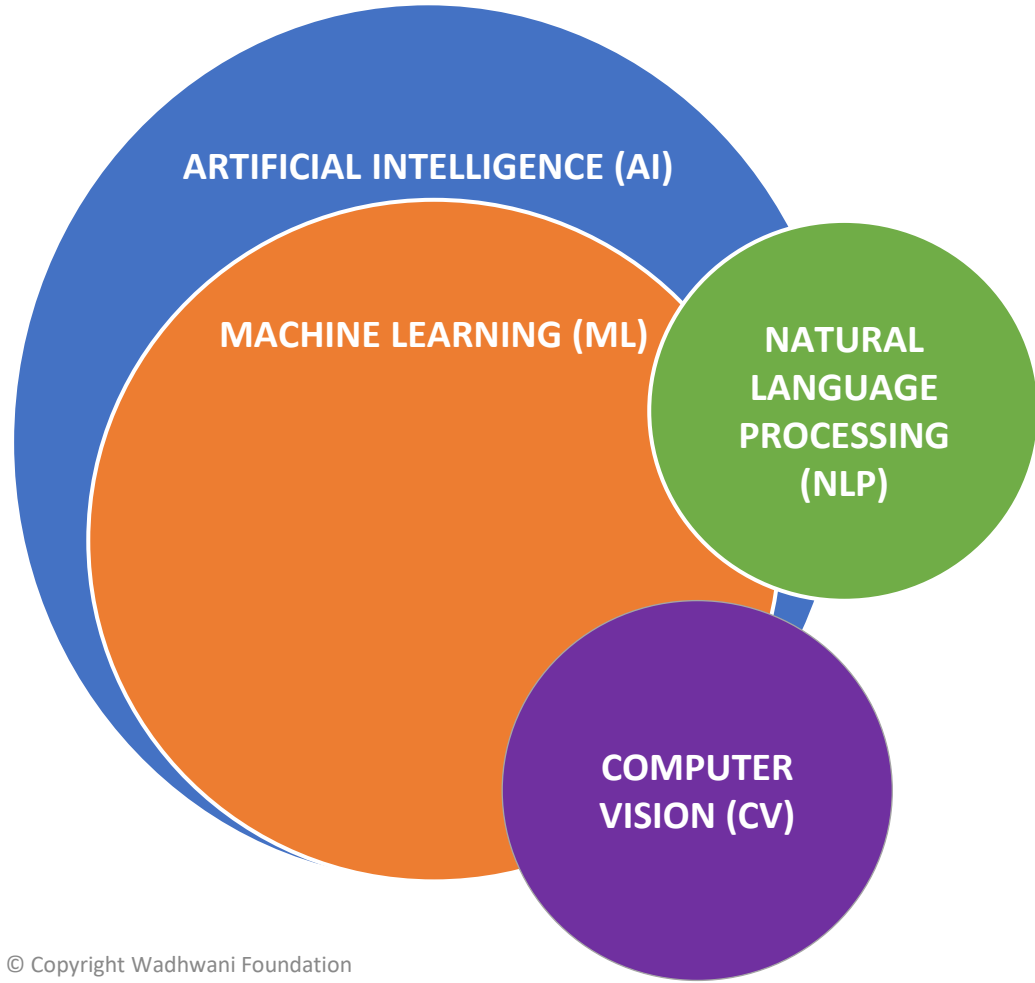
Take home Assignment

- We discussed Data Science/Analytics today
- Read about AI-ML, NLP and CV
- Suggest some ideas where some of these technologies may be used in governance!

"Applications of Artificial Intelligence for Good Governance"

February 23rd (Thursday)

AI & ML

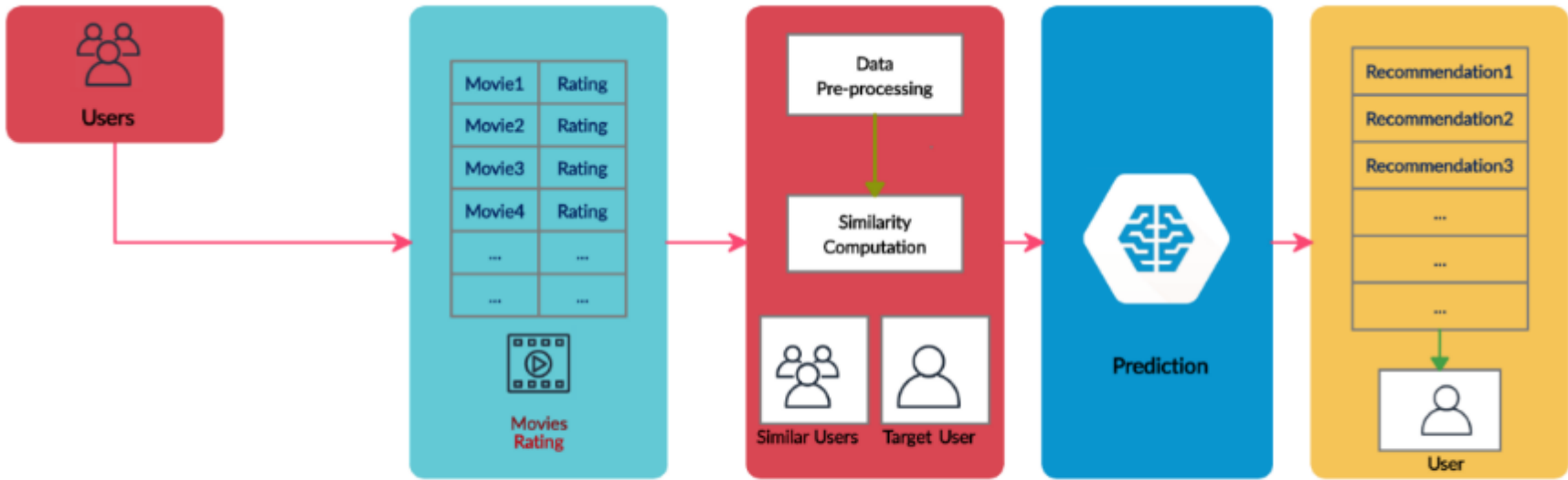




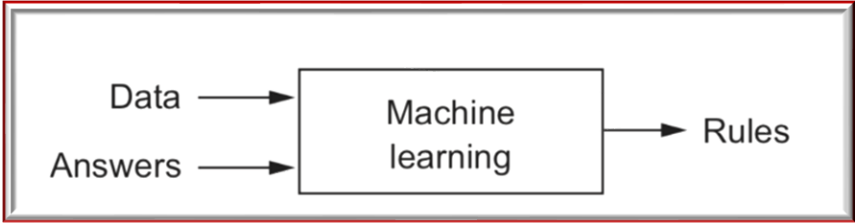
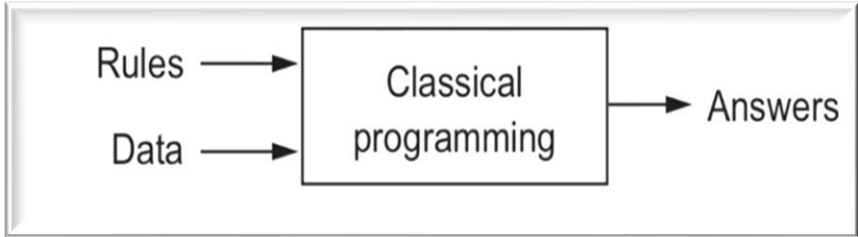
Extra topping: AI/ML

Cheese: Data

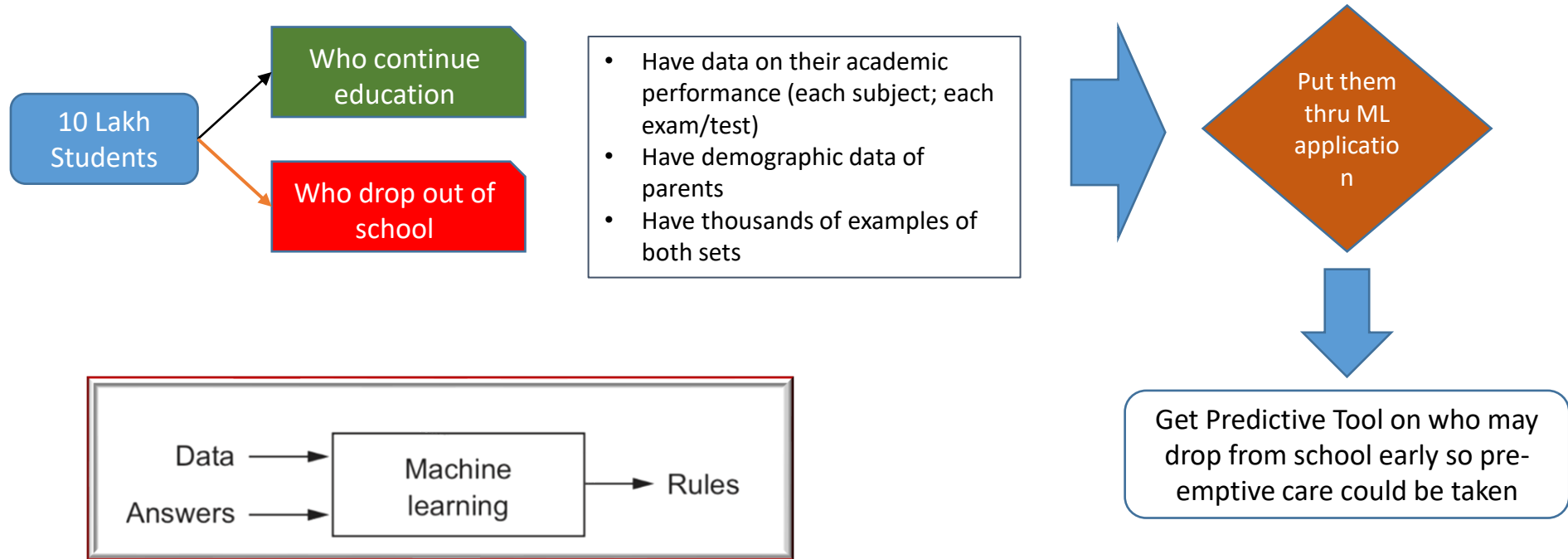
Pizza Base: Applications



NETFLIX



Prediction of who may drop from school: One example of Machine Learning Application



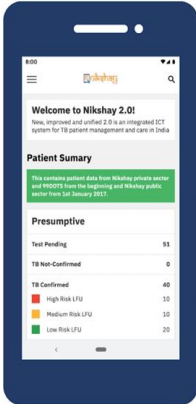
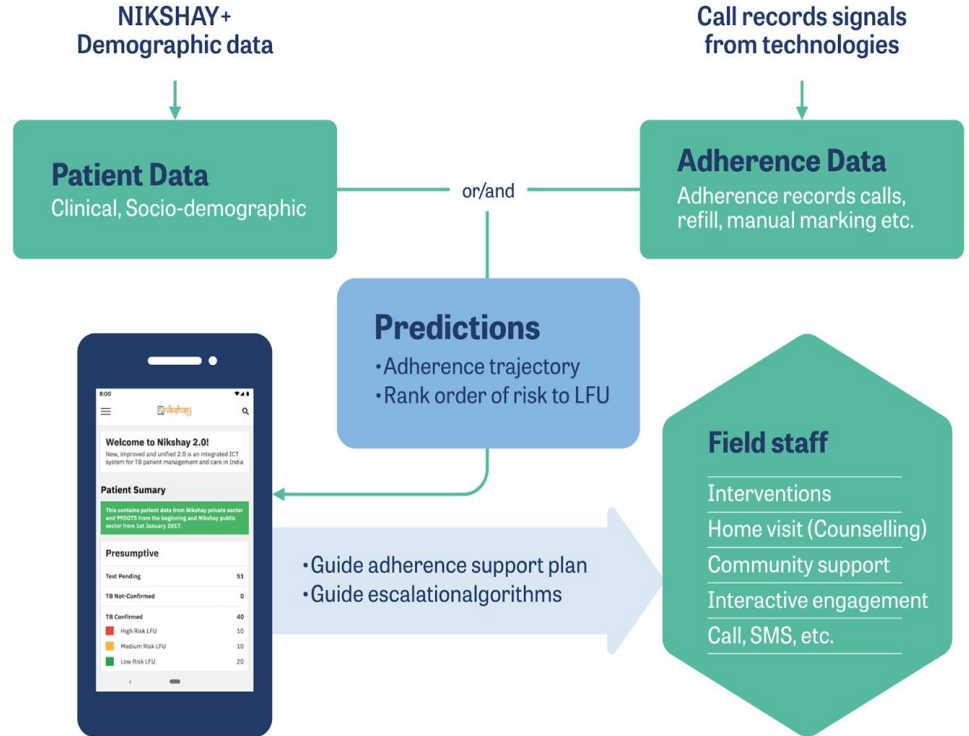
We get rules based on patterns seen by ML algorithm in the dataset

Prediction of risk for Loss to Follow Up (LFU) among TB patients using AI

Treatment success of TB patients highly depends on the adherence to the treatment regimen. An extreme form of non-adherence is Lost to follow up (LFU), defined as a TB patient whose treatment was interrupted for one or more consecutive month(s).

Early prediction is critical since these:

- Are silent transmitters of Tuberculosis
- Have higher risk of development and amplification of drug resistant TB
- Cause increase in overall deaths
- Cause increase in risk of morbidity
- Cause increase in cost burden



AGRICULTURE

Pest Management

Pink bollworm is the most economically damaging pest for cotton.



20-30%

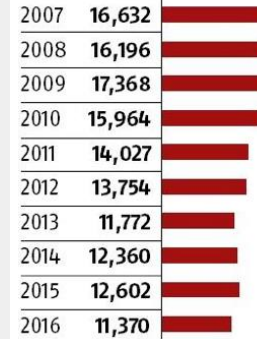
Of avoidable crop damage occurs due to pests such as the Pink Bollworm



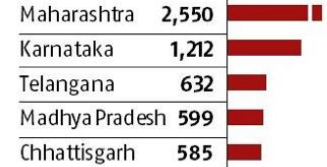
50%

Of all pesticides are used on cotton crops in India.

SUICIDE IN AGRICULTURAL SECTOR



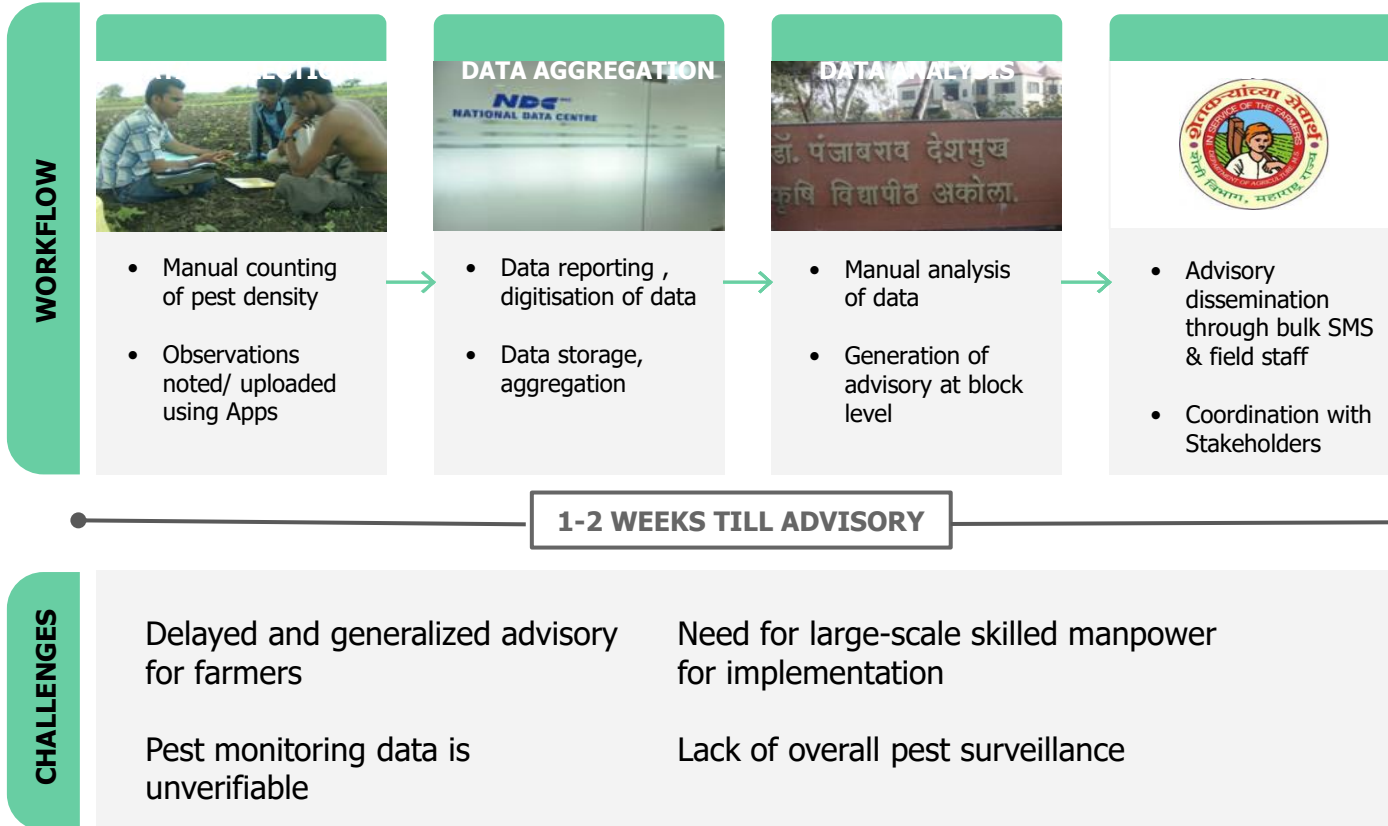
STATES REPORTING MOST FARMER SUICIDES IN 2016



BUSINESS STANDARD

Loss of income and indebtedness have been the cause of an increase in farmer suicides in key agricultural states.

Existing agricultural program workflow



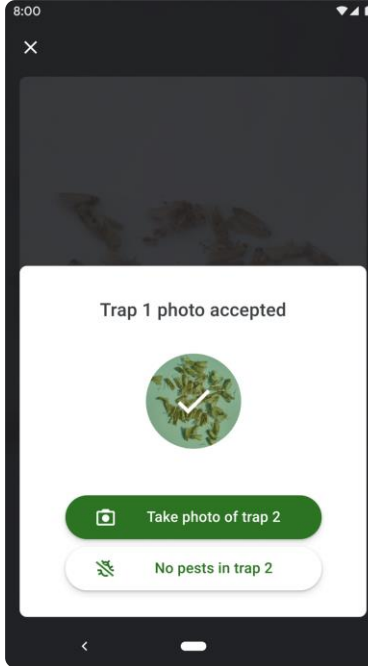
A farmer or extension worker empties the moths from the pheromone trap.



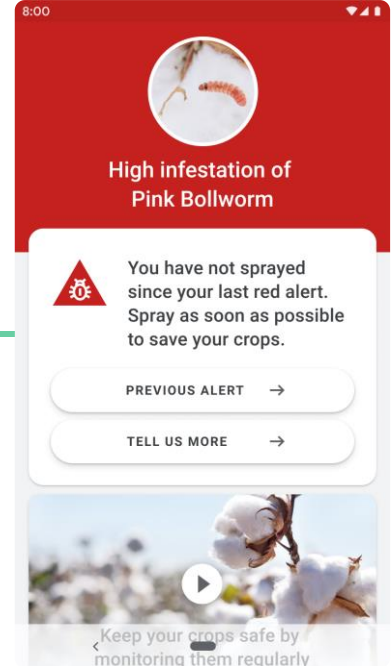
He then opens the app and captures 2 trap images.



The pests are detected and counted from the image uploaded.

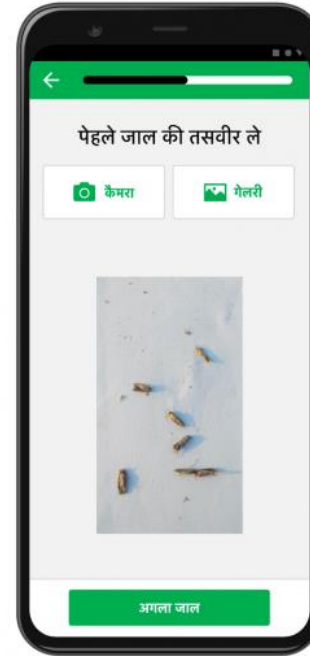


An advisory is then generated on the app.



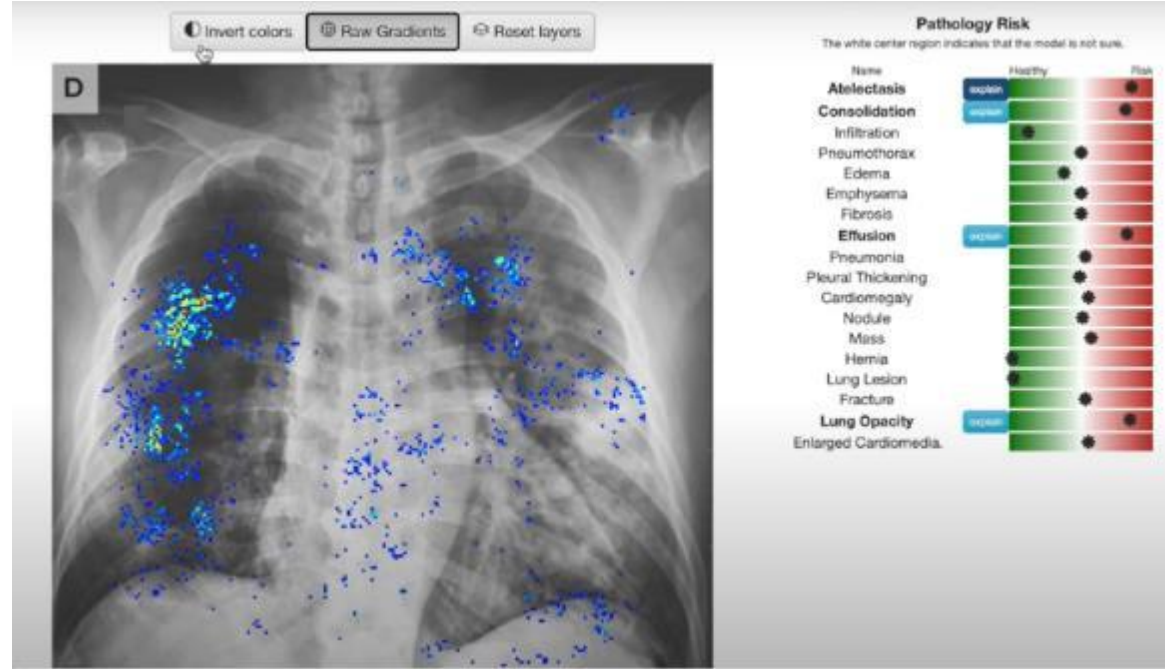
25

Regional language app
Where pictures are uploaded



AI algorithms can read chest X-rays at resident radiologist levels

3. Computer vision

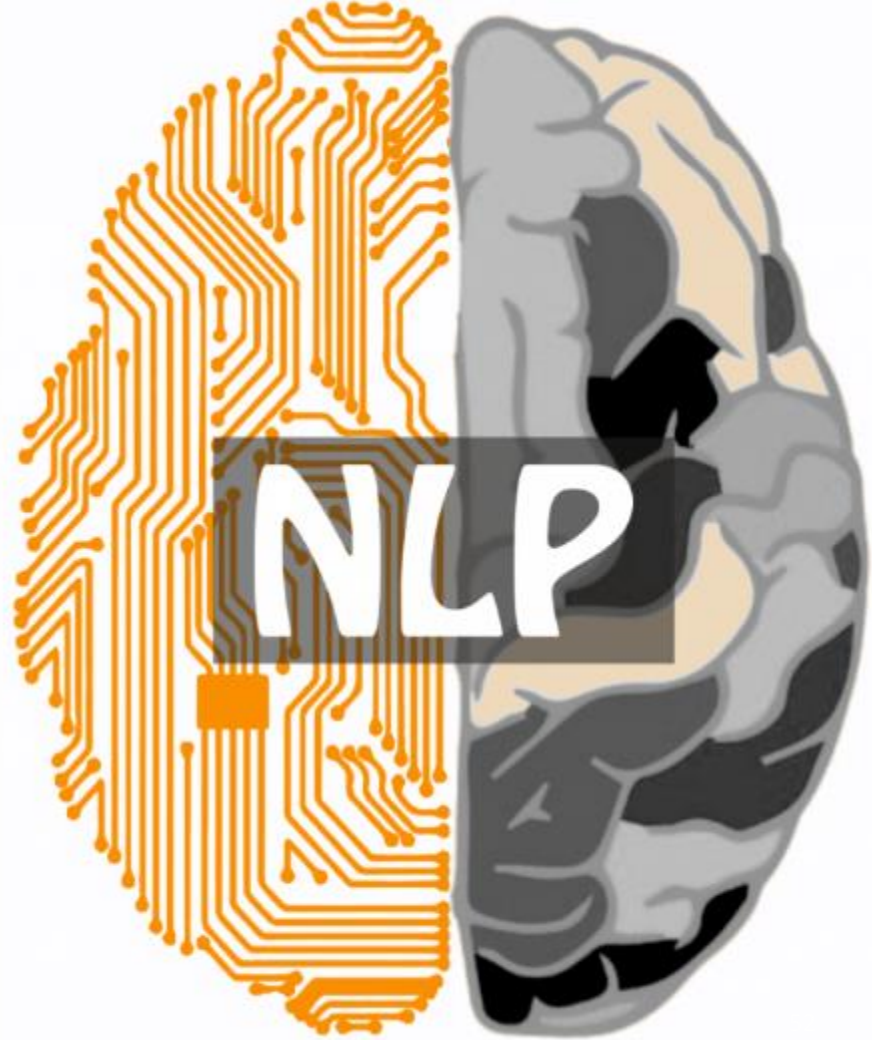


Examples of Use of ML for Prediction

Solution	Dataset	Delivery Mechanism	End User
Fraud Detection: ML powered application being used in GST for early identification of fraudsters or identify risky refund applications or risky new registrants	Registration data, Refund data, past fraud data etc.	GST Departments	Tax officers in GST
Pregnancy Risk Stratification: ML Powered predictor of maternal and mortality at various stages of pregnancy to reduce negative outcomes by triaging and access to differentiated care	Digitised antenatal care data	Through existing frontline health workers smartphone applications and platforms	Frontline health workers
Risk scoring in case of taxpayers under GST or loan applicants or applicants for any welfare program etc.	Part labelled data	Concerned department	Officers of the department

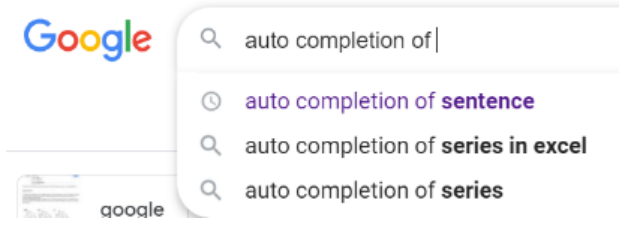
Solution	Dataset	Delivery Mechanism	End User
Face recognition: non-touch attendance; confirmation of examinee in examination hall	Face scan of people	Through an App	Employees, Examinee
Healthcare: Medical image processing (X-Ray or MRI); Early detection of eye diseases; Cancer detection, Tumour detection;	X-ray or MRI	Through an application	Patients, Radiologists Doctors
Agriculture: crop wellbeing and quality; bug invasion; pinpointing weeds; flowering detection; yield assessment; Animal monitoring; irrigation management etc.	Photo or video of crop, bugs, animals etc.	Through an application	Farmers
Traffic System: Number plate reading for generation of challan for traffic violation; lane tracking; automated vehicle classification at tolls; etc.	Photo/video	Through an application	Traffic police
Transportation: Autonomous vehicles; moving violations detection; parking occupancy detection; road condition monitoring; civil infrastructure inspection and monitoring ; driver attentiveness detection etc.	Photo/video	Through an application	Traffic police
Security: Surveillance; Face recognition; Crowd detection; Human abnormal behavior detection; Illegal parking detection; Speeding vehicle detection	Photo/video feed	Through an application	Security agencies
Matching signature/photo While making application online, live video is taken using the mobile phone based app to ensure liveness and match it with photo in supported identity documents	Photos and video	Through an application	Any department which asks for photo during filing of application

Application area	Organisation	Description of the application	Technology Used
Remote monitoring	IRCTC	Monitoring Railway kitchens to detect rodent activities and deficiencies in hygiene	Artificial Intelligence Machine Learning Computer Vision
Water Supply	Bangalore Water Supply Company	Water distribution improvement program	IoT, ML, Analytics
Cataract screening	Govt of Telangana	Cataract screening app and whether a LASIK, cataract or other surgery will be successful or not	Machine Learning Computer Vision
TB Project	Wadhvani Institute of Artificial Intelligence	To predict patients who will stop medication midway	ML



NLP

Auto completion



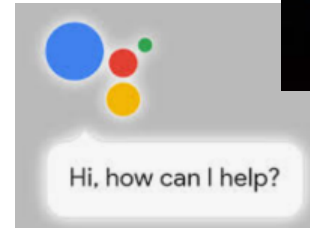
AI based Chatbot



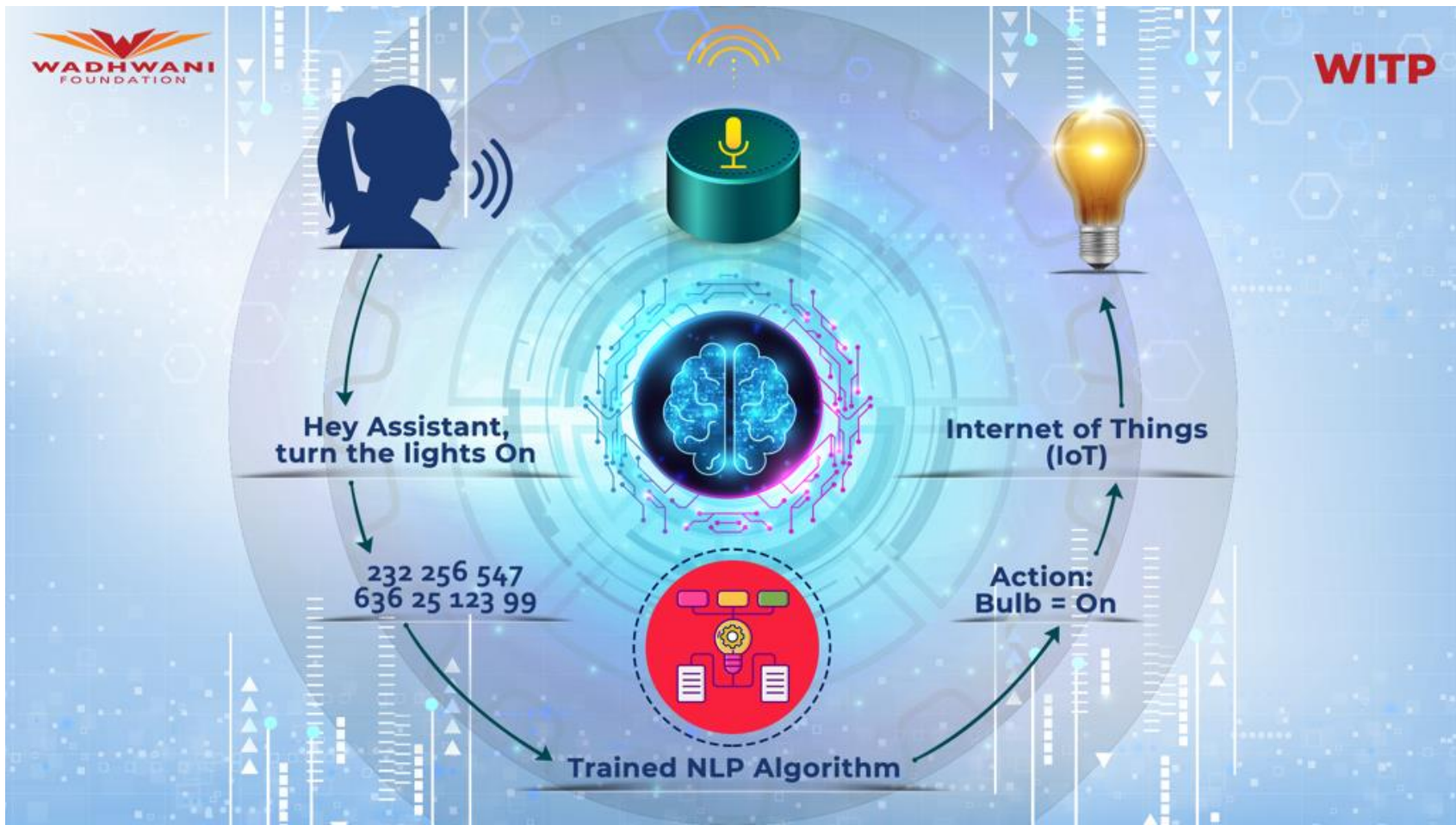
Writing Books/Articles

**Best AI Story & Novel
Writing Tools**

Voice Assistants



Natural Language Processing: Voice Assistants

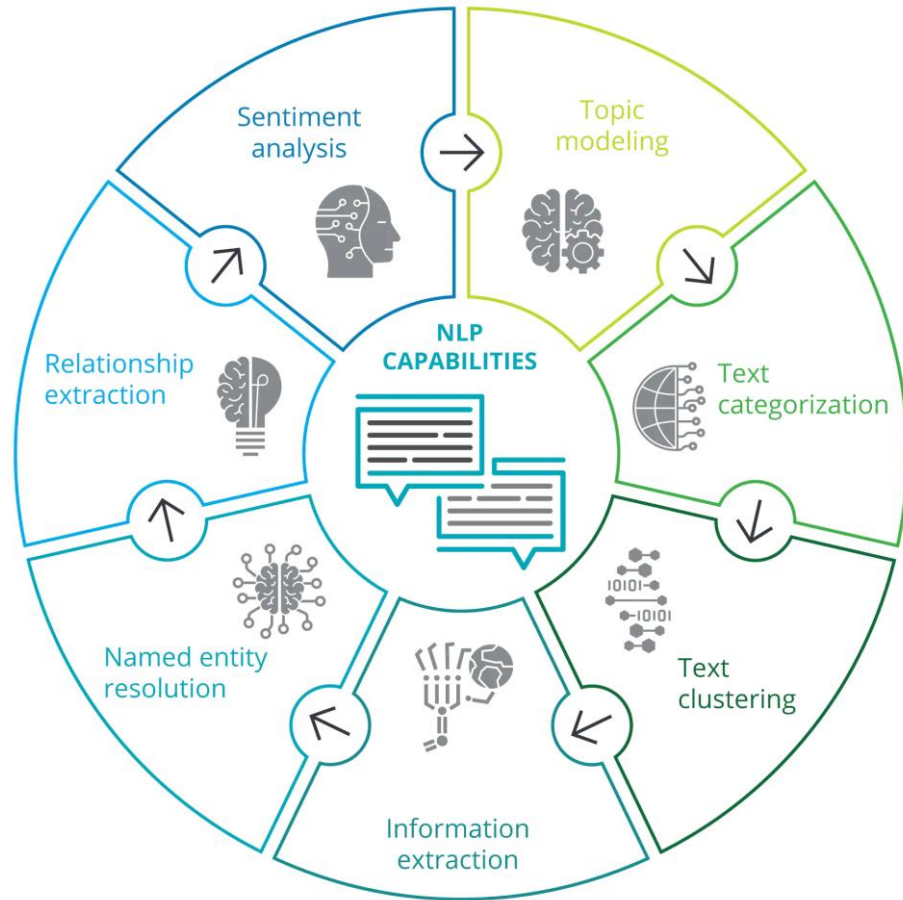


Key NLP Capabilities

- Voice assistants
- AI Chatbots
- Writing Books

FIGURE 2

Key NLP capabilities



राम राम, मुझे
कुछ मदद
चाहिए



मेरी पेंशन
नहीं आई दो
महिने से

हाँ

जी हुकुम

क्या तुम्हारा पेंशन इसी
मोबाइल से जुड़ा है?

कृपया एक मिनट
प्रतीक्षा करें।

अधिकारी आज आपसे
बात करेंगे.
मैं कल तुम्हें फोन करूंगा।



Chatbots

Imagine a villager could call up and talk to Chatbot in Telugu and gets the reply in Telugu and matters resolved. The voice /video chatbot calls next day and takes the feedback from the complainant.



నాకు కొంత సహాయం కావాలి

గత 2 నెలలుగా నాకు
పింఛను అందలేదు

అవును

ధన్యవాదాలు

అభివందనాలు

మీ పెన్షన్ ఖాతా ఈ మొబైల్
నెంబరుకు లింక్ చేయబడింది

దయచేసి 1 నిమిషం వేచి
ఉండండి

సంబంధిత అధికారి ఈ రోజు మీకు
కాల్ చేస్తారు. పరిష్కారంపై ఫీడ్
బ్యాక్ పొందడం కొరకు నేను రేపు
తిరిగి కాల్ చేస్తాను.



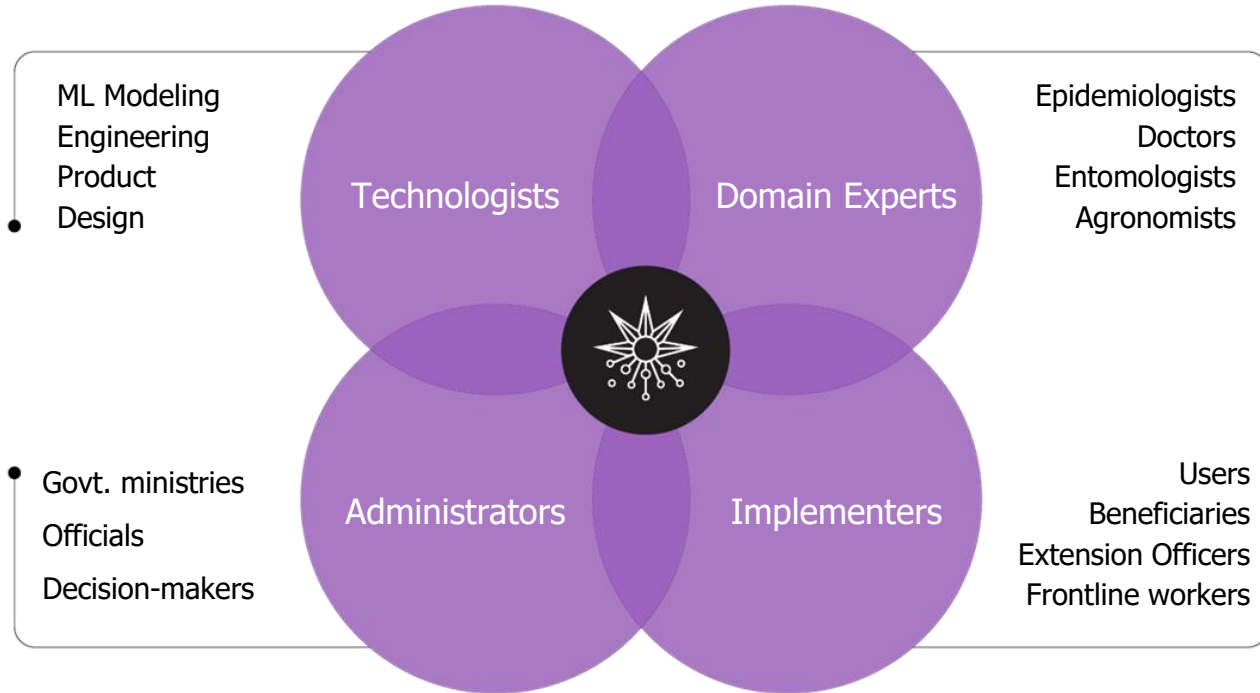
Examples of Use of NLP in the Government

	Solution	Dataset	Delivery Mechanism	End User
Natural Language Processing (NLP)	Chatbots to assist people <ul style="list-style-type: none"> in getting information from websites, checking status of application, Lodging complaints etc. 	Information available on website; FAQs; Access to software for accessing status	Through website Already being used by few departments of Delhi Govt.	Citizen and businesses
	Translating documents	Training datasets on translations done to train the model	As an App	Government officials
	Summarization of long reports	Long reports and summaries created in past.	As an App	Government officials
	Analysis of email and other forms of communication to identify insider trading etc.	Email or Chats	As an additional application	Intelligence agencies
	One Indian language to another language translation in real time	This is being handled by MEITY under National Language Translation Mission (NLTM)		

Few Important Points

- AI/ML is not panacea for all ills. It can't be force-fitted everywhere.
- In many cases ethical issues come up (like our bias gets reflected in the model).
- AI/ML projects are different from regular IT or eGov Projects where we have clarity on deliverables, timelines. Unlike IT Applications, ET based models require long time for development as they involve iterative process. Also, involvement of subject matter expert from the user department is for longer period.

A radically collaborative approach is needed to develop AI-based solutions that make a difference



You can reach us at:



Contact us at:

Prakash Kumar
CEO, WITP

prakash.kumar@wfglobal.org

Kamal Das
Dean, WITP

kamal.das@wfglobal.org

Helpdesk: helpdesk.witp@wfglobal.org

Few Examples of Other Emerging Technologies

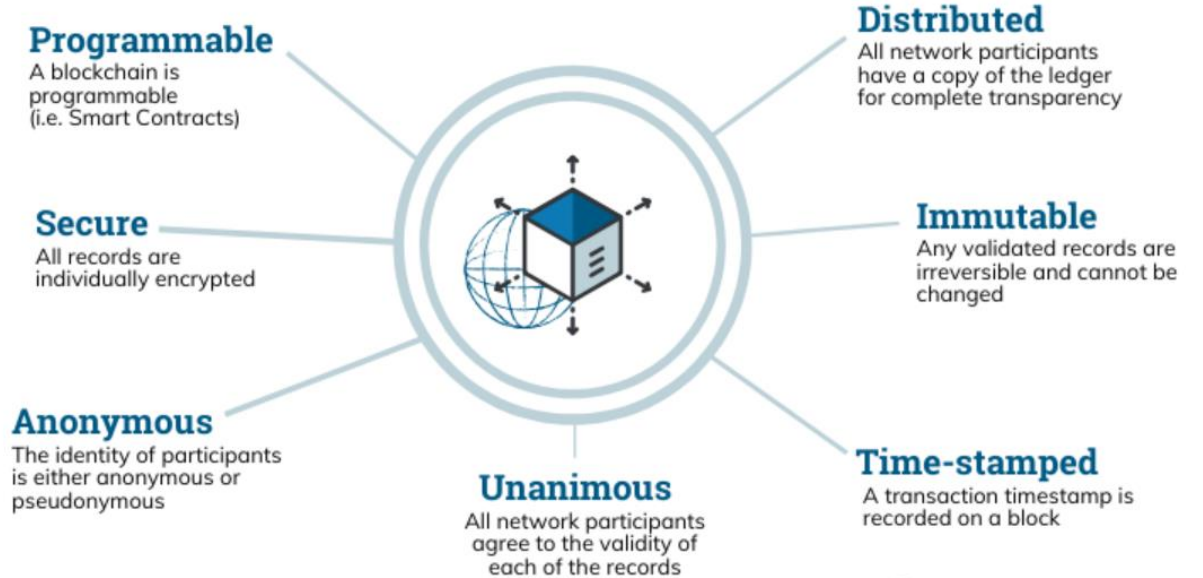
#	Technology	Application of technology (few examples)
1	IoT	Street Light, Precision Agriculture, Logistics, Smart Parking, Remote Health Monitoring, Smart Homes, Air Pollution Monitoring System etc.
2	Drones	Agriculture, Disaster Management, Construction, Military, transportation, survey, forestry etc.
3	AR/VR	Immersive Learning, Disaster Management, Tourism, Défense Forces training etc.
4	RPA	Data migration (of data among government legacy systems and organization platforms at certain times or at triggering events); Public sentiment analysis; Data entry from scanned forms; Converting pdf doc to text for processing; copy/paste from one system to another; Application processing (by validating ID from other database, processing citizen queries, validating data etc.);



Few Ideas for Govt

- Land registration system (Started by Karnataka/UP)
- Making mark sheet fool proof (CBSC)
- Marriage certificate and caste certificate issued by District offices
- Management of sensitive data like forensic reports
- Smart Contract

The Properties of Distributed Ledger Technology (DLT)



© Euromoney Learning 2020

A blockchain is essentially a digital ledger of transactions that is duplicated and distributed across the entire network of computer systems on the blockchain.

Each block in the chain contains a number of transactions, and every time a new transaction occurs on the blockchain, a record of that transaction is added to every participant's ledger. The decentralized database managed by multiple participants is known as Distributed Ledger Technology (DLT).

Blockchain is a type of DLT in which transactions are recorded with an immutable cryptographic signature called a [hash](#).