The SatSure Newsletter

January - March 2021

www.satsure.co

The Indian AgTech Paradigm

Hemendra Mathur

The "India Agricultural Platform"

Nipun Mehrotra

Fair AI for Satellite Data Applications

Krishna G. Namboothiri

Open Data Cube in SatSure

Thaiseer Parammal

FOREWORD

It's been more than a year since the COVID-19 pandemic started, triggering and accelerating the digital transformation of businesses around the world. The last year has provided its fair share of challenges to companies by testing their character and business sustainability. We at SatSure had our fair share to overcome as well, but the team came together, and we ended up growing two times in size and generating five times more revenues than what we were doing just six months back.

We have had an exciting beginning to 2021, launching our open innovation platform SatSure Sparta, followed by the launch of SatSure Cygnus - an all-weather optical data virtual constellation service available on SatSure Sparta. SatSure Cygnus caters to the need for high-frequency vegetation monitoring during the rainy season when there can be long gaps in optical satellite data availability. And now, we are releasing the first edition of TSNL in 2021!

In this edition of the newsletter, we are bringing you the knowledge and opinions of experienced industry professionals on topics focussing on the government policies and their impact on AgTech startups; the transformation of agriculture in India using Artificial intelligence and large scale data platforms to connect farmers with digital services from the government, financial institutions, and AgTechs. We've also got two thought-provoking articles on the ethical dilemma of Artificial Intelligence, and the technological advances in harnessing the power of satellite data.

In the first article, Hemendra Mathur - Venture Capitalist and industry thought leader, in an interview with Prateep Basu, discusses the Indian Agtech paradigm. He assesses the current state of the Indian AgTech space for startups, the funding scene, and the impact of government policies on these startups' growth to drive the Indian AgriStack.

The second Article by Nipun Mehrotra, Founder of the Agri Collaboratory, talks about using Data and Artificial Intelligence to transform agriculture in India. The article focuses on Agri Collaboratory's work, together with other industry players, to enable this transformation.

The third article by Krishna G. Namboothiri, Data Science Product Manager at SatSure, talks about Fairness in Artificial Intelligence. Picking up from the article on the Explainability of AI, she discusses ethics in an AI model to solve business problems without creating an imbalance in the generated intelligence.

The last article is written by Thaiseer Parammal, Platform Manager at SatSure. He talks about the technological advances to harness Satellite Data's power using the Open Data Cube (ODC) framework. ODC is an open tool for the geospatial community to deliver decision intelligence on a real-time basis to solve satellite data's Big Data challenges.

As the pandemic continues to take a toll on most of the world, we at SatSure will always be thankful for the support shown by our TSNL readers to bring a thought leadership content platform.

We hope you enjoy reading this edition of TSNL as always!

Rashmit Singh Sukhmani

Rashmit Singh Sukhmani,
Co-Founder, CTO/CDO
SatSure

THE INDIAN AGTECH PARADIGM

A Conversation with

Hemendra Mathur

Venture Partner, Bharat Innovation Fund

1. If you had to draw comparisons between Indian AgTech in 2015 and now, what would be the most critical changes you think will go on to define the new decade?

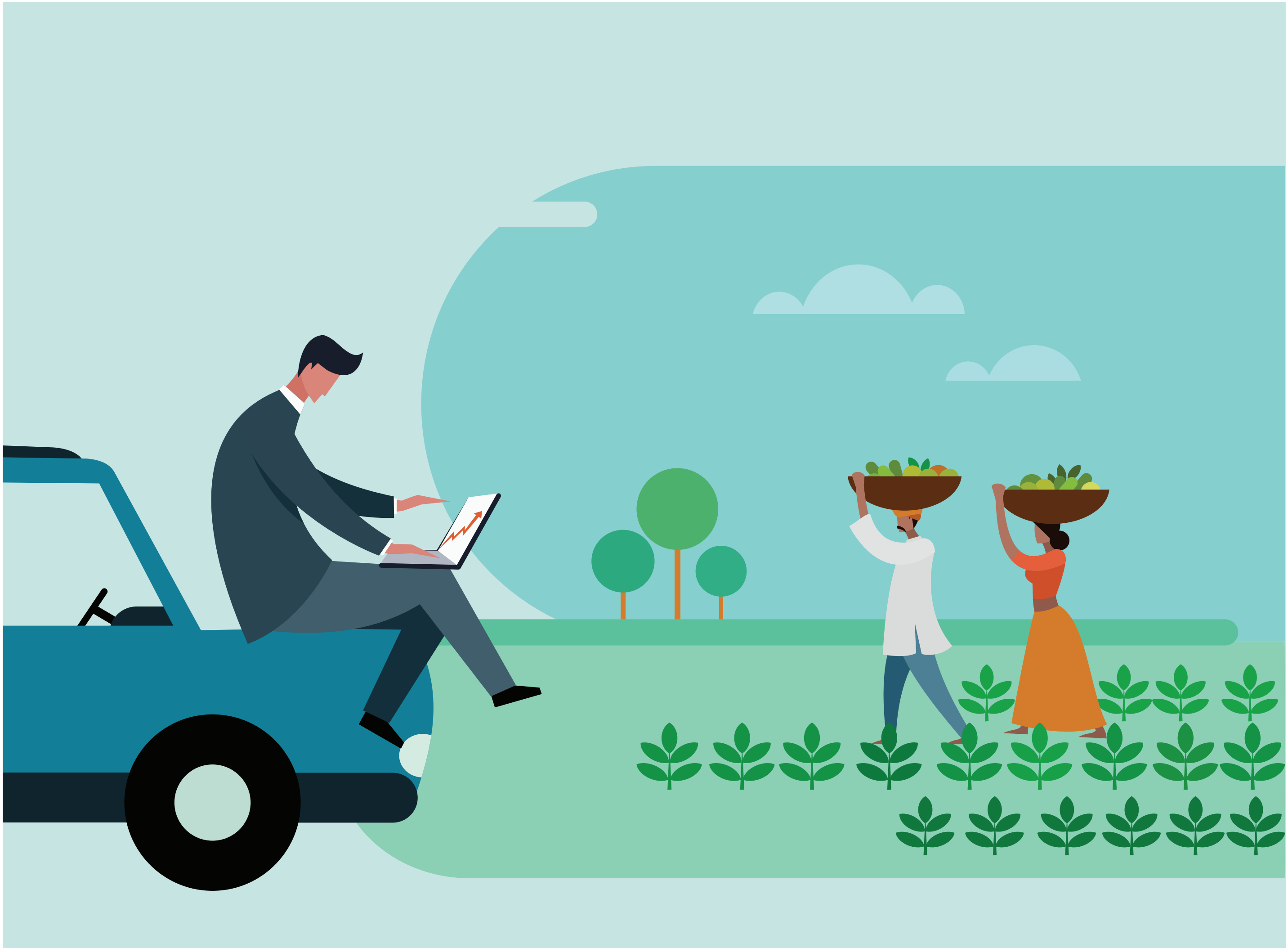
A lot has happened in the last five years, with the inflection point for Indian AgTech being the year 2017, when many smart, well-qualified entrepreneurs entered the fray with strong tech backgrounds and brought in equally strong teams, with years of corporate and tech experience behind them. This brought a sea change in how tech was envisioned to solve the agriculture sector's challenges. For the first time, we saw these ideas in action to improve farm economics and drive efficiency in a complex food supply chain.

The faster-than-expected adoption of data and technology interventions by the farmers and the value chain players in the seeds, fertilisers, food processing, insurance, and banking sectors over the last three years is a defining point in making business models for AgTech work. In comparison to the industry in 2015, two established things are the path to scale for AgTechs and growing investor interest in this sector. In 2020 alone, 53 deals closed in agritech, which is astounding, despite it being a pandemic year.

However, I feel that two things are still missing in this equation - a path to sustainability (across all AgTech segments) and penetration of real deep tech solutions in AgTech. Most of the AgTech models currently are supply-chain oriented, which is the low hanging fruit for the entrepreneurs and the investors. The AgTech sector has a lot of potential to develop and absorb deep tech, including digital tech (sensors, IoT, blockchain, machine learning, image processing applications) and biotech (biostimulants, bio fertilisers and bio pesticides; gene editing, microbial applications, food nutrition etc.) However, it is still untapped due to the lack of enough innovations and early-stage adoption challenges.

It is also interesting to note that the scope of AgTech is expanding; for example, we now have AgriFinTechs, Farm robotics, Vertical/indoor farming verticals propagating with unique and differentiated business models.

As far as agfintechs space is concerned, we have a huge opportunity of $200 billion the priority sector lending in agriculture, with another $ 100 billon gaps each for post harvest value chain financing and ancillary industries such as dairy, aquaculture, livestock, etc. This opportunity to me is bigger than the mortgage or urban-centric personal lending market. As the organisation of the agriculture supply chain keeps improving, it will offer more opportunities for digitisation and more so for AgFinTechs to enable deeper penetration of institutional lending, digital payments, savings, parametric insurance to the sector.



2. Given the maximum quantum of funding that has gone towards e-commerce and retailing startups, what do you think about the 'tech' part in Indian AgriTech?

By the end of this decade, I am bullish that the investment rate in AgTech will be at least $ 1 billion annually, and we will see a cumulative investment of over $ 10 billion in the next ten years. The multiplier effect of this scale of investment in the rural economy is beyond imagination. Agtech is one of the sectors which can bridge the gap between India and Bharat, drive climate resilience, reduce food loss and unlock huge value for the farmers.

A significant part of the investment will still go to supply chain startups because the problem's scale is vast and is a low hanging opportunity. While deep tech startups have higher defensibility, investors' ability to evaluate them and entrepreneurs ability to go to market is still work-in-progress. Deep tech investors have to have patience for market development and product development of Deep tech applications in agriculture. Their growth trajectories will be different from that of supply chain focused startups.

A lot of capital for such startups will come from outside of India because of proven Deep tech applications, especially in places like Israel, Germany, U.S., etc. So it is only a matter of time when Indian AgTechs also get the same level of interest. Business models have to be proven as well for these, alongside the technology component.

3. To what extent would incubators have played a role in this spurt of AgriTech startups' growth in the last five years? And how do you see their role evolving in the next five years?

Incubators have played a vital role in building the AgTech ecosystem in India. Organisations such as Villgro, Indigram Labs, ICRISAT, CIIE.CO, A-Idea Naarm and Pusa Krishi had supported startups when AgTech was not even a term here in India. Building enabling ecosystems is what role incubators should continue to play.

We also need incubators that are rural-centric and not just in the premium institutions of the country. I believe this is a missing piece today amidst the success stories of AgTech incubators in India. Current incubation hubs are very city-centric, and we are missing out on some of the essential components that make AgTechs work, i.e. the connection and trust with farmers where the local ecosystem can play a vital role.

In my opinion, incubators should also take up the challenges of taking their incubatees' solutions to farmers, helping them partner with FPOs / MSMEs, and promoting rural entrepreneurship specifically in digital-assist solutions microprocessing, farm-level value addition.

4. What impact have you seen on this evolving AgriTech ecosystem by startup-friendly government policies?

The government's recognition of AgTech as an industry in the last three years is one of the ecosystem's key achievements. This has helped agtechs engage with policymakers at both central and state levels. I feel that the government can support agtechs in the next few years by building necessary physical (farm-level storage, processing) and digital (Agristack) infrastructure and making regulations aligned with evolving agtech business models.

We also need a high-velocity catalytic fund from the government for giving the first cheque to seed-stage startups for rapid testing of innovations. Government platforms like Startup India and Agnii have done a great job. It will be good if they can build sector-specific offerings with dedicated resources for sectors like agriculture.

5. With large technology companies like Google, Microsoft, Cisco getting into the agriculture sector, how do you see their entry affecting the VC interest and scaling for AgriTech startups?

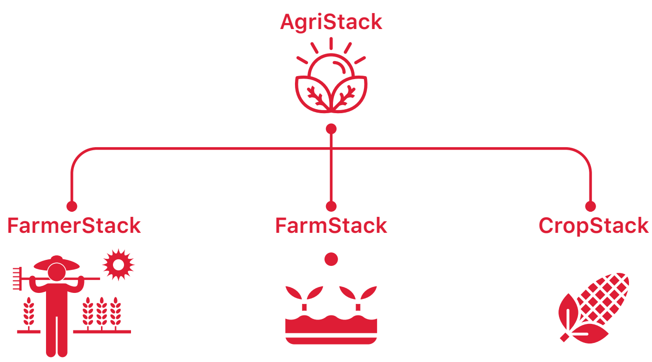
Firstly, it is interesting to see these companies enter the agriculture space. I see them acting more like System Integrators than pure solution providers. These large technology companies collaborate with startups, which bodes well for everyone in the ecosystem (entrepreneurs, investors) as M&As are fueled by such close working relationships. They can build and enable public platforms offering various services that might not be possible for one startup to pull together. The interest in the agriculture sector by the likes of Google, Microsoft, AWS, and Cisco signals is a positive sign and also an indicator of the AgTech ecosystem maturing.

6. We have discussed several aspects of the AgTech ecosystem, but it would be incomplete without mentioning the 'India Agristack'. How do you think it can address some of the key problems in Indian agriculture, and how can the AgriStack be operationalised?

The Agristack is conceptualised as a public digital platform for ready and almost instant access to farmers. This platform can be the hotbed for driving disruptive innovations in the agricultural sector. I have proposed an architecture involving three essential building blocks as a starting point for building India AgriStack, including - FarmerStack, FarmStack and CropStack.

FarmerStack is essentially determining who the farmer is, whom we want to reach out to. FarmStack includes the location and dimensions of farm size. This is important to estimate farming needs and the income potential from the farm. Farmer and Farm stacks can be combined to establish who the farmer is, where he is located, his input needs, and how much he can earn from his farm. The CropStack includes data on the number of crops and types of crops a farmer is growing, and it is integral to potential interventions needed to improve farm economics.

There are several possible application areas of Agristack. Some of them are:



Credit to Farmers: Agristack can be used by banks and MFIs to assess the creditworthiness of farmers and then tailor-make products for each one of them. As per the priority sector norms, 18 per cent of ANBC (Average Net Bank Credit) is for lending to Agriculture. FY21-22's Priority Sector Lending (PSL) target for agriculture stands at Rs. 16.5 lakh crores. Banks' ability to reach out to farmers, judge their credit-worthiness, and monitor crops for recovery can be facilitated by Agristack.

Crop Insurance: The number of farmers (or farm holdings) covered under the Prime Minister's Fasal Bima Yojna (PMFBY) is about six crores. The majority of the farmers who availed crop insurance are the ones who have taken loan from banks. Only about 30% of the gross cropped area is covered under insurance. Agristack can improve the access of crop insurance to wider farmer bases (non-loanee farmers specifically) and help bring more crop area under coverage. It can also help Insurance companies understand the farmer and farm risk profile needed to decide the premium and settle the claims.

Direct Benefit Transfer (DBT) The government can use Agristack to design and implement schemes for farmer welfare and implement DBT. Given the government's intent to move farmer welfare schemes to DBT, Agristack can play a catalytic role in this process.

Market Linkages: Corporates selling to or buying from farmers can also benefit immensely with direct and targeted access to farmers. The digitisation will improve farmers's access to quality inputs and connect them directly with multiple buyers, improving their price realisation.

Building AgriStack has to be a collaborative effort, including governments (both federal and state) and innovators. It is a one-time investment in building such infrastructure, and then there will be a systematic effort to maintain and continuously update it. Unlike IndiaStack, AgriStack will have dynamicity given multiple crop cycles, farmer migration (especially tenant farmers and sharecroppers) and changes in farm boundaries over some time.

Farmers can also be incentivised to share the data in building the AgriStack. For example, the direct benefit transfer scheme can make it mandatory for farmers to share certain data points (for their benefit) with the government agencies. They can share it directly or through local government bodies or any government-designated person in the village. The data can be selectively audited, or compliance tools can be built to check and maintain data quality.

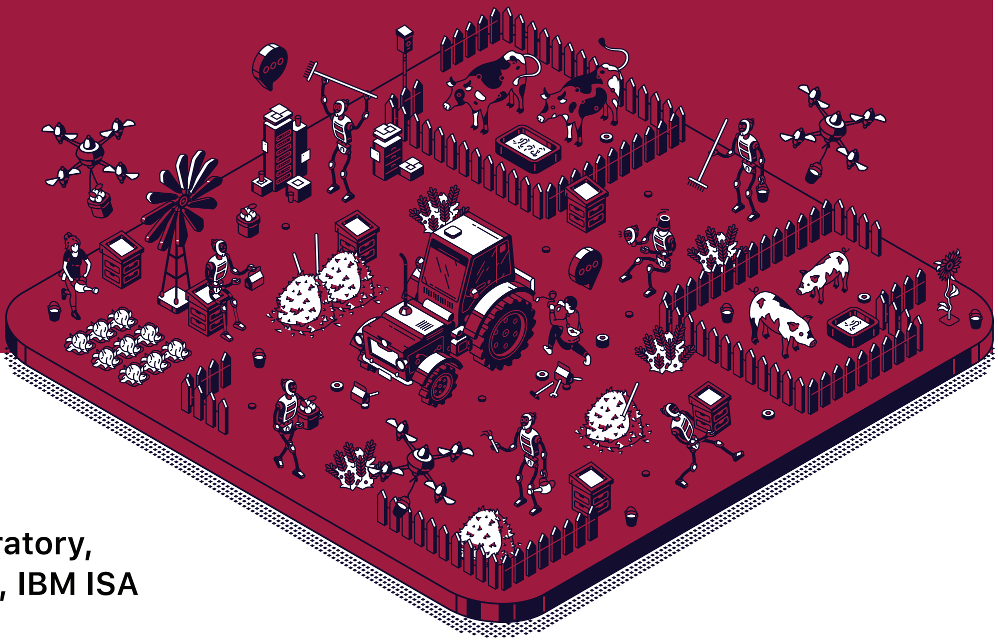
These are some of the ways in which AgriStack can be operationalised. I do not doubt that the effort in building AgriStack is worth the effort, and it could it could transform how product and services are delivered to farmers. It can also give a solid boost to the AgTechs who today spend most of their time collecting and cleaning data instead of building applications around it.

About the Interviewee

Hemendra Mathur is the Venture Partner at Bharat Innovation Fund, a deep tech sector agnostic fund and Co-Founder of ThinkAg, a not-for-profit platform for accelerating the adoption of innovation in agriculture. He has over 12 years of experience in early-stage and growth capital investing as well as over 12 years of management consulting and investment banking experience in the food, agribusiness, retail and consumer goods industry for corporates as well as the Government of India and state governments. He also serves as the Chairman of the Task Force on Agri Startups at FICCI, India.

The "India Agricultural Platform"

DATA Flows & AI to Assist in the Sectoral Transformation



Nipun Mehrotra
Founder - The Agri Collaboratory,
Former Chief Digital Officer, IBM ISA

Agriculture utilises around 50% of India's workforce, 90% of freshwater and 46% of the land. Yet it barely generates 14% of GDP and 10% of exports. Most farms are small-sized, with half of them being rain-fed. The majority of farmers (~100 Million) lack low cost, institutional credit. All this, combined with rapid soil degradation and significant crop wastage (Rs 90,000 Crore), makes the path forward look arduous. However, the situation is not entirely bleak. Production has outpaced population growth for decades, catapulting India into one of the largest global producers of wheat, rice, sugarcane, cotton, milk, pulses, fruits, vegetables and more. But despite volumetric self-sufficiency, India struggles with 33% malnutrition - ironically amongst farmers themselves.

Keeping these gaps in mind, **'The Agri Collaboratory'** was formed as a not-for-profit, specialised Food and Agriculture "Think and Do Tank", working with a long-term mission to reposition India as a leading global agricultural Innovator by 2030. The guiding philosophy of 'The Agri Collaboratory' is to realise the national agricultural vision into reality is anchored in solving several intractable issues and transforming how we collaborate and collectively drive execution. We agreed that working on complex, national scale challenges requires taking the long view (15 - 30 years) with well-calibrated milestones. We focus on outcomes in four discrete yet interconnected areas, led by subsets of alliance members:

[**Economic** - e.g., Credit access | **Indian Agri Platform** - e.g., decision making | **Sustainability** - e.g., Water usage | **Innovation Quotient -** farmer to researcher.]

Among these subsets, the Indian **Agricultural Platform (IAP)** is catering to the need for an open, scalable, integrating platform that democratises access to Agri information, credit, insurance and markets; incubates innovative business models; and enables better decision making. The IAP shall be created by the ecosystem, regulated by the Government, and is envisioned as an "enabling framework of Data and Services (applications) around a data exchange".

Half our farmers cannot get credit easily because of incorrect or lack of land records, and financiers have limited credit-risk assessment information. As in Selvam's example, the IAP can facilitate this by triangulating data real-time from several entities. These are a farm management startup revealing cropping history, satellite data for estimated yield & water source, geolocation coordinates with Aadhaar helps a lender assess credit-risk. The lender readily pays a 0.5% assessment fee to the

Use Case 1: Agri Fintech (Farm Credit)

Consider this scenario, leveraging the IAP for a credit use-case in the year 2024:

Selvam, a small and marginal rice farmer with a two-acre farm near Madurai, logs into the IAP using retina scan, fills in a loan request in Tamil using iconography, attaches photos of himself and the farm. He still doesn't have formal land records in his name, since his father passed away, but accords consent for his farm data to be accessed from different entities (Govt, Start-ups, FPOs etc.): Aadhaar, geo-location, three year's crop type, yield and earnings.

This data flows in, completing Selvam's application and providing visibility of his farming history to any potential lenders to evaluate credit risk, they use the combination of geo-location along with Aadhar to extract Selvam's farm credit history and details of all existing and completed loans. The automated process allows a majority of the lenders to approve/reject the loan online within minutes. Artificial Intelligence flags issues needing a clarifying phone call.

Comparing different offers transparently, allows Selvam to choose wisely and he chooses a lender that pays the seed supplier directly, crediting the balance loan into Selvam's regular bank using the UPI interface, and creating an auto-debit for the month after harvest. The entire process is digital. The bank also remits a fee to the Start-up for providing Selvam's cropping history.

Back in 2020, Selvam recalls filling several loan applications individually, wasting money travelling to Madurai and loan sanctions took an average of two months. The IAP has reversed and democratised the process, giving him the power - and lenders now bid for his loan.

Use Case 2: Translating Agri Data into Actionable Insight (Govt. Decision Making)

Picture this real-life scenario:

A Bihar Government bureaucrat is trying to forecast the tomato's post-harvest prices to avoid the severe crisis last season when prices crashed due to a sudden glut. Tomatoes dumped on the road and farmer suicides attracted bad press. He pores over submissions from each district, showing crop-wise acreage and sowing week. From experience, he knows this data could be 2-3 months old, while tomato's harvest in 3-4 months. He observes that gross crop acreage varies 15-25% across submissions, and he suspects some districts fill in data without stepping out of the office.

Based on this, how can he forecast prices or take action?

Should he believe the input suggesting Tomato acreage is 15% lower than the previous cycle? How would the forecasted winter rains and colder weather impact tomato yield? He wishes he could advise farmers better because staggering sowing and harvesting by 1-2 weeks play a significant role in smoothening market price swings.

But for that, he needs automated, real-time data.

startup for its farm data, aiding its revenue stream.

To solve this problem, an enabling framework like the IAP will use Agri Data-sets from multiple sources (government, enterprises, startups), seamlessly translating it into information and then into precisely actionable Insight to be leveraged for varied Agri (and non-Agri) use cases.

For instance, digital crop signature from satellites, combined with AI, can reveal crop-wise acreage under plantation by district within 4-6 weeks of planting. As the crop matures, it estimates crop yield and then combined with acreage and processing capacity in proximity - likely post-harvest prices. The IAP also provides an alternate forecast by analysing aggregated seed sales data, district wise, from suppliers to predict acreage under tomatoes. Both estimates are correlated for accuracy.

Early price forecasts enable faster tactical actions avoiding price crashes, like helping stagger harvesting or tying up additional quantities with processing plants in advance.

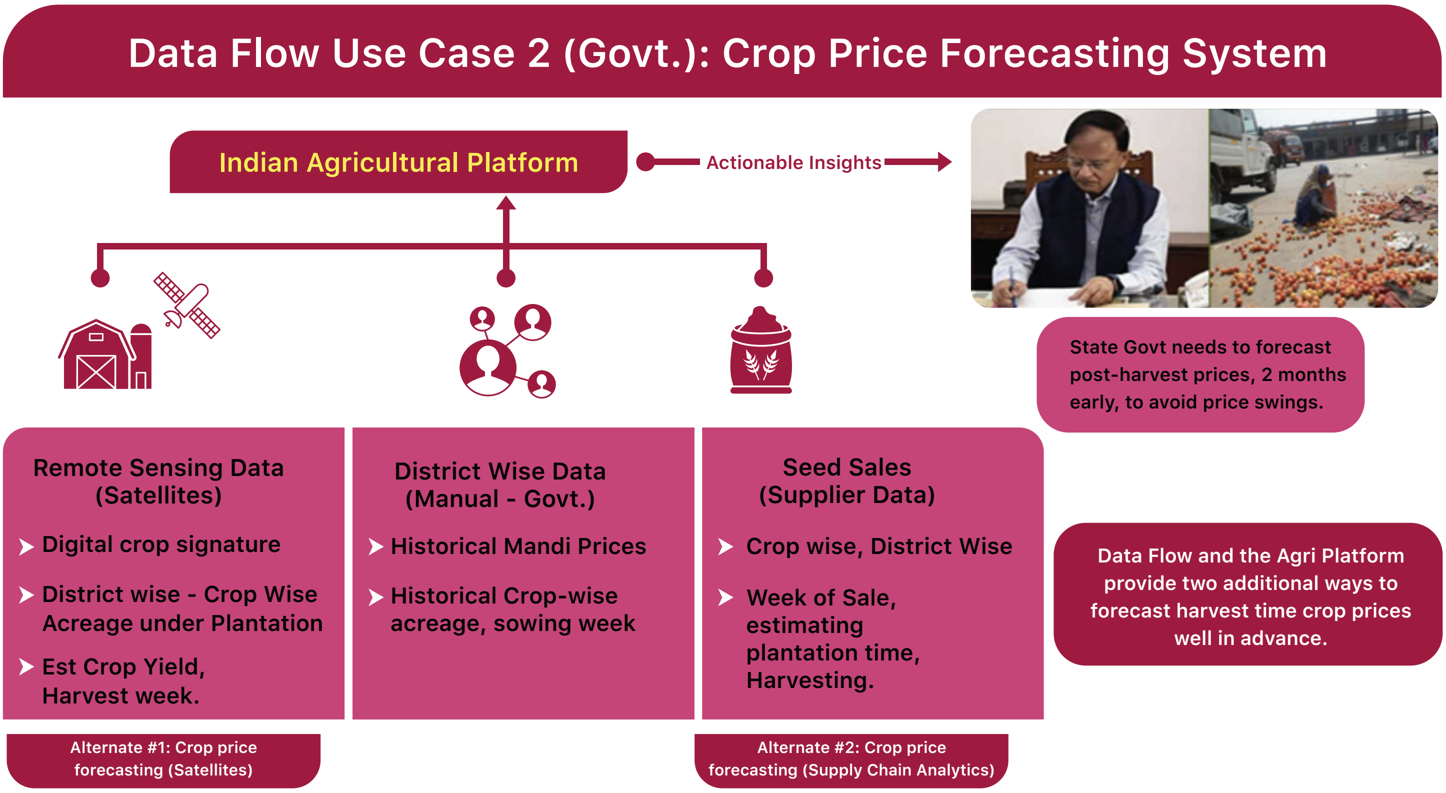
This kind of application integration requires seamless data interoperability across the Agri ecosystem, with due conformance to Data privacy and usage policies. Currently, there is a vacuum in Agri Data standardisation, calibration and certification. Disaggregated and non-standardised data is deemed un-trustworthy and rendered ineffective for further processing.

Standardisation will help improve "data trust", furthering automation using AI models, also avoiding real-world biases creeping into AI prediction.

India Agricultural Platform will incubate new "Data Partnerships", Business models and revenue streams:

Innovation impacts the entire agricultural value-chain: (Soil testing, Crop selection, Sowing, Irrigation, Yield estimation, Harvesting, Farm Equipment, Farm operations & management including weed control & pesticide application, Price discovery, Sorting and Food processing). This value-chain is rapidly becoming digital, leading to an increasing amount of live and real-time data being generated. Physical maps being digitised is an example, while another is through the multiplicity of payment and Agri trading platforms. In addition, there is a rapid increase in "machine-data" generated by precision farming applications: field sensors (for soil moisture etc.), satellites (for yield estimation, early pest warning etc.), robotics (for water and nutrient injection, harvesting, high precision weed removal etc.), mobile cameras (for pest attack, nutrient deficiencies), and drones (for real-time farm monitoring, surveying, 3D modelling etc.).

The IAP enables a technical and commercial framework to harness this dataflow and facilitate "data partnerships" between the Government, Startups, Corporates, Research, Academia based on either direct or indirect business benefit. Once AI processed data is available to be leveraged, new, innovative business models will emerge, helping monetise data contributing to improved productivity and profitability of Agriculture and other sectors. Given that



similar challenges exist across the developing world, India will establish itself as a globally recognised Agri innovator.

Powered by Artificial Intelligence (AI) & Data Analytics, IAP helps tactical and strategic decision making, leveraging multi-year, multi-source information aggregated from the farms to state/national levels. It processes huge data flows and uses tools like video, voice, and vernacular translation to facilitate farmer engagement. The platform is hosted on a Cloud and reduces duplication by integrating data sources and an extensive backend of new and existing applications: Govt's eNam, ITC's eChoupal, NCDEX's NeML, APEDA's TraceNet etc. related to logistics, weather, supply-chain, warehousing, assaying, recommendation engines, etc.

About the Author

Nipun Mehrotra is the Founder of "The Agri Collaboratory", a Not - for Profit, Food and Agriculture "Think and Do Tank, committed to taking a holistic, collaborative approach to agriculture by leveraging innovation at scale and enabling an open Agri platform. The Agri Collaboratory is an open alliance of like-minded partners across interest groups working on a mission to reposition India as a leading global agricultural Innovator by 2030.

Nipun is a transformational business leader with 35 years of diverse experience in Wipro and IBM. He led strategic initiatives and large operational businesses across the IBM Asia Pacific Region for twelve years and was the Chief Digital Officer, India/SA, until leaving IBM (2019). He led IBM's thrust on digital platforms and innovation in several industry domains - including important societal areas like agriculture, health - with AI, Cloud, IoT, Blockchain, Remote Sensing data, etc.

His expertise lies in managing large complex businesses, balancing constant transformation with operational execution. At four distinct occasions across India and Asia, he helped reshape large multi-hundred-million-dollar businesses with thousands of employees, facing strong headwinds due to industry upheavals and aligning them to market growth.

Fair AI for Satellite Data Applications

Krishna G. Namboothiri,
Product Manager – Data Science
SatSure

In the previous edition of The SatSure Newsletter, we spoke about the Explainability of Artificial Intelligence. The authors argued that the **AI explainability paradigm needs to be overcome before venturing into its ethics.** Explainability and interpretability is the concept that humans can understand the predictions made by an AI tool in stark contrast to a "blackbox" model. While interpretability is about how accurate a machine learning model can associate a cause to an effect, explainability talks about how different parameters, often hidden in Deep Nets, justify the results.

This article picks up from the above by assuming that the explainability problem has been tackled. At SatSure, we use AI models on satellite data for use in agricultural sector. Let us consider a hypothetical model that uses satellite information to calculate farm credit score for financing. Thanks to the explainable model, we understand how different factors like soil moisture, temperature, and historical cropping patterns influence this prediction. Before employing this model in the real world, we should stop and ask if the model is actually **"Fair".**

Fair AI – the Opposite of Biased AI.

In a machine learning system, data is used to train a model that provides predictions used to make specific decisions. In any of these steps, if one category is discriminated against another, intentionally or unintentionally, and these predictions reinforce the already existing population bias, the model becomes an unethical AI application. For example, in a given population, we see that women have historically lower wages than men. If a machine learning algorithm learned this data and used it to predict a recruits salary and used gender as a parameter, this would be unethical. The distinction should be clear. Truthful reporting of the data does expose systemic discrimination. But when the same bias is utilised for future decision making, this only perpetuates the bias. **A fair machine learning algorithm aims to eliminate this vicious loop.**

AI Applications Using Satellite Data

Earth observation satellites capture images at high resolution at high frequencies, and AI can transform how data is processed both inspace and on Earth. This changes how we can deliver insights to customers. The applications include object detection of elements like buildings, road infrastructure, industrial area, forest land etc. There can also be dynamic detection of land cover changes that can help us monitor agriculture, deforestation, water reserves etc. Many satellite data-based algorithms employ complex AI architectures, which have the scope for biases to creep in either at the detection phase or when it is translated to policymaking.

Biases in Machine Learning Applications of Satellite Image

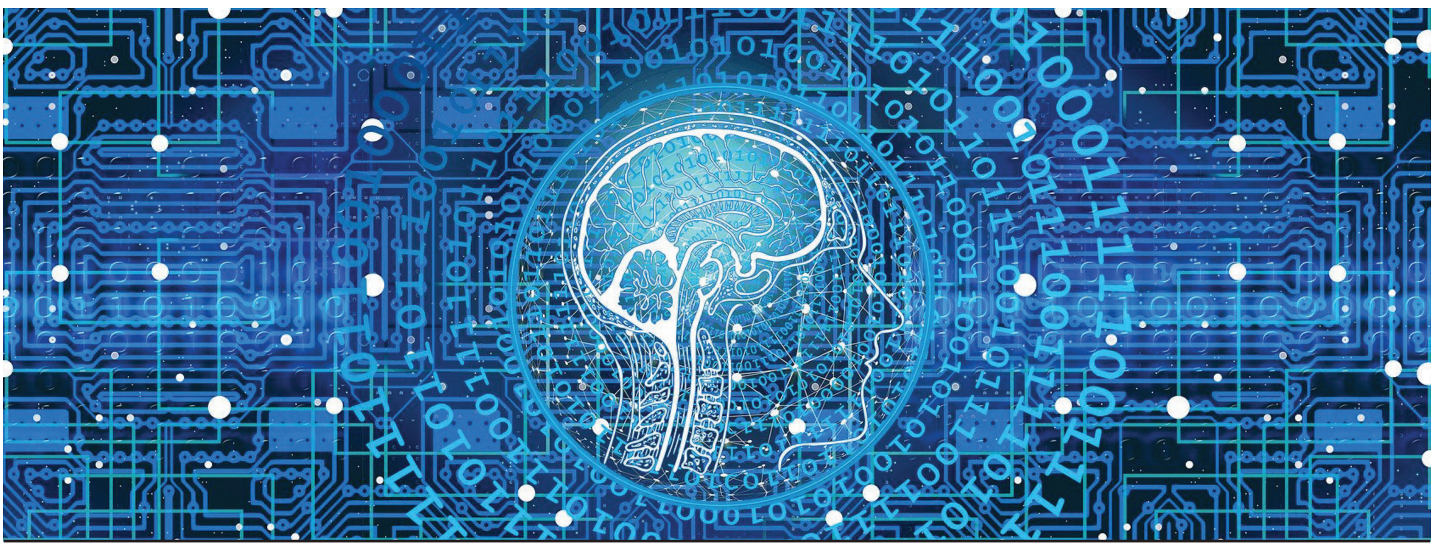
We can categorise the potential biases of applying machine learning on satellite images into four main areas:

1. Radiometric Biases

Satellite image-based remote sensing is highly dependent on the sensor specifications, the topology, features of the land, illumination factors, and meteorological conditions both before and during the acquisition of an image. Technical biases could result during the process of generating a remote sensing product. A common challenge faced in this regard is the presence of cloud cover in optical data. Another is the classification of mountainous terrain, which becomes problematic because of variations in the sun illumination angle resulting in biased reflectance data. While these are indeed significant hurdles to overcome and should be considered while modelling satellite data, this analysis is best left to remote sensing technologists and hence not included in the scope of this article.

2. Spatial and Geographical Bias

A simple example of this is that a driverless car trained over asphalt roads will fail to navigate grasslands. A quick search on crop classification models using deep learning will take you to a highly Amerocentric and Eurocentric world of datasets and publications[1]. These models are constantly touted as benchmarks and become global standards for solving machine learning problems on remote sensing data. However, the land-use patterns differ significantly in



developing countries where these insights could have the maximum impact. Generating diverse datasets comes at a cost which is currently the most significant obstacle towards solving this problem. A way to address this bias is through the use of suitable open data sets like OpenStreetMap[2].

It is not just the datasets that are a victim of this bias. The existing paradigm also influences some of the metrics used to evaluate our hypotheses. For example, there have been studies to assess regional wealth using night light data[3] , an urban-centric model that will inherently classify an agriculturally thriving community as poor. However, later studies addressed this, and the authors recommended that other metrics like agricultural maps be used in wealth mapping[4]. This step is the right direction that identifies the source of bias and attempts to mitigate it.

3. Imbalanced Data

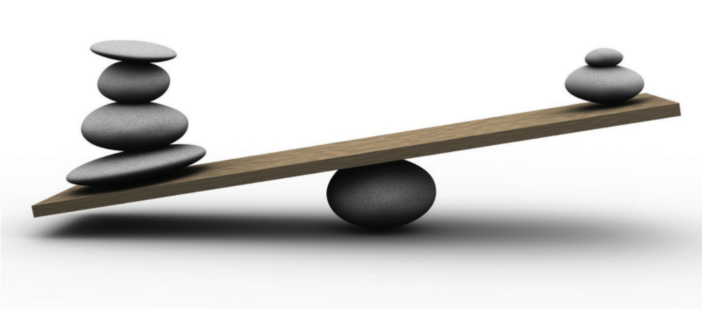


Fig a: Class imbalances causes the model to be biased towards majority class

Consider a binary classification problem with 99% of data belonging to one class and 1% belonging to the second class, which leads to challenges for the model to learn the characteristics of the minority class and hence difficulties in prediction. In one such example in our area of interest, the authors detect oil spills from satellite images[5] . One of the common methods of mitigating this issue is to upsample the minority data samples or downsample the majority of data samples to create a class balance. But also, one has to be careful in modifying the algorithm and metrics used in testing. For example, classifying all samples in the above mentioned hypothetical dataset as a majority class will give us a 99% accuracy even without any training. This is more of a statisticalproblem,and Bayesian statisticians have devised numerous methods to identify and tackle it systematically [6].

4. Reinforcing Biases

We would be discussing this particular bias in detail as it is probably the most dangerous of all biases. Reinforcing bias is central to the example discussed in the introduction paragraph on women employee wages and how that can become a vicious loop. This is illustrated in Fig b, where a real-world bias is propagated in data, modelling and policymaking to further the original bias. As humans, we can see how this is ethically wrong, but machines are not equipped with this "intelligence". The missing ingredient is causal reasoning which helps us disengage correlation from causation, while for an AI, the relations are purely associative.

Let us go back to our original explainable and interpretable model for farm credit. And let us assume that a financial enterprise uses this model to get recommendations on potential areas for investment. We understand how soil moisture contributes to crop performance, and an AI model will recommend that it makes sense to focus our attention and investment on regions with higher soil moisture. However, the model

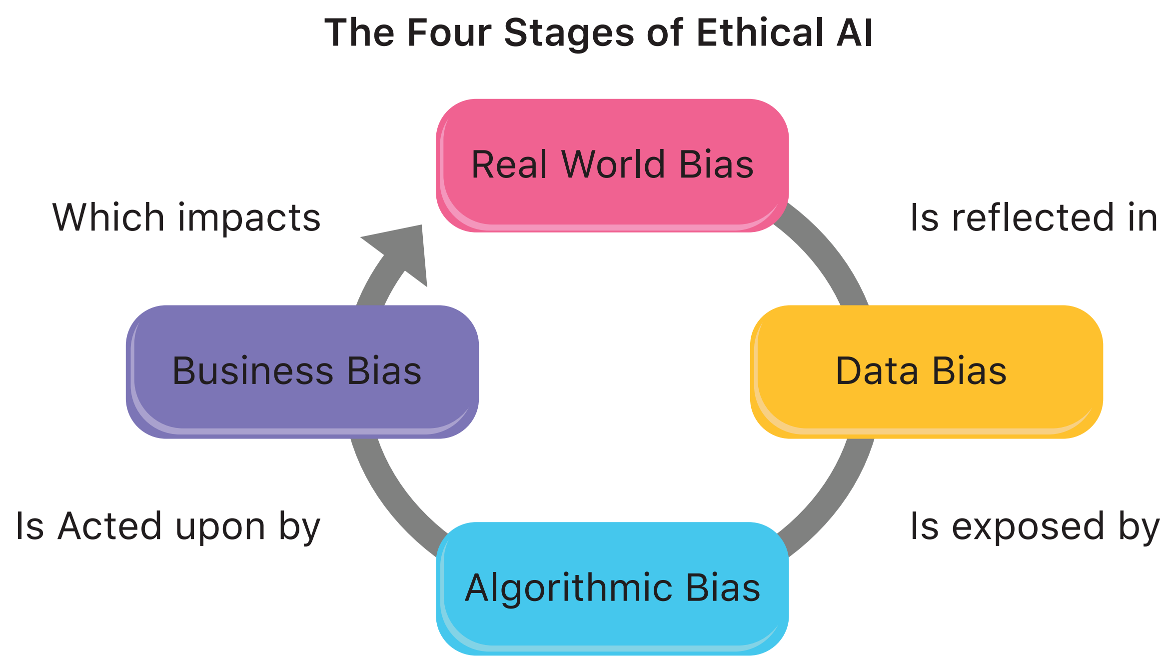


Fig b: Four stages of bias propagation

misses in this analysis that soil moisture is heavily a function of rainfall and irrigation facilities. Irrigation facilities, in turn, is a measure of the existing infrastructure facilities and wealth of a region. Thus, we could be investing in an already developed agricultural area, making it even more prosperous while a systemically neglected region would continue to underperform due to lack of attention.

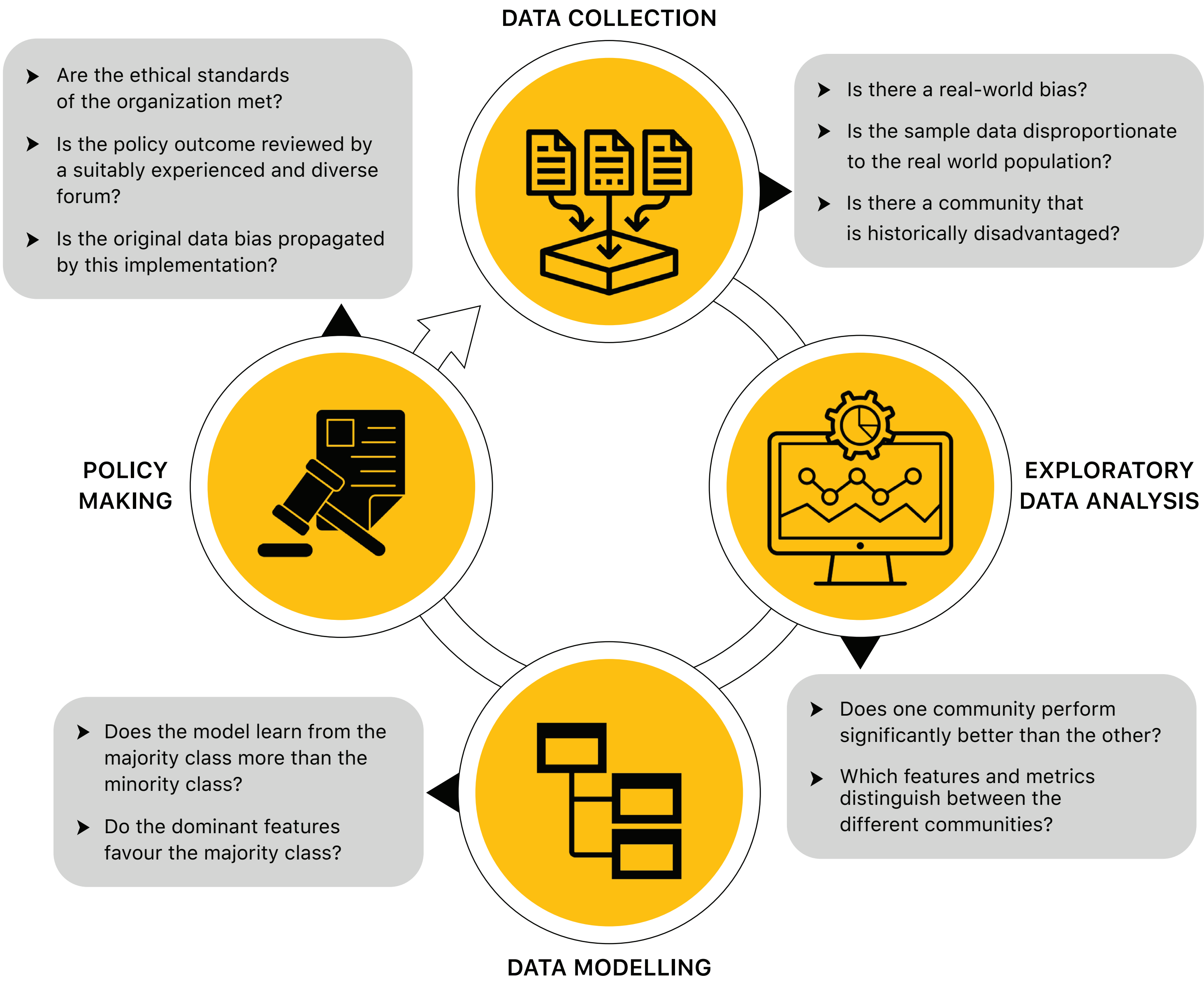
Let’s say we also assume we use the farmers profile as a feature in the model. Considering our women’s wage example and knowing that traditionally women have lesser assets and income than men, this will lead to discrimination against women farmers. We see this taking place in countries where female labour is essentially invisible, and in an agricultural family, it is usually the man who is labelled the farmer[7]. Thus as the biases add up, a woman farming in a historically impoverished region has higher chances of continuing to be poor while a man doing the same job in a rain-fed area is more likely to prosper.

The Responsibility of a Fair AI

The question we have before us on whose shoulder does the responsibility of building and using a Fair AI model lie? One might argue that the onus is on the consumer of the AI product. They would need to take into account the social inequalities and let that not penalise the underprivileged. But what would be the incentive to do so?

Or is the onus on the Data Science or Machine Learning teams handling the data and generating the AI model? But how can a Computer Scientist or Decision Scientist sitting miles away from the regions of interest and has little subject knowledge on the underlying socio- economic patterns know how to account for social biases? It takes a vast amount of experience, both academically and socially, to identify the unintentional consequences. For this reason, it is usually recommended to have a diverse team to work towards a responsible solution. Going back to the above example of farm credit calculation, if the reviewing team had an urban planner, an economist, an agricultural scientist, and a farmers representative to advise the computer scientist, the data modelling would have taken a different approach. A gender- balanced team would have advocated for improving the opportunities for women. Instead of focusing on the peripheral features, the attempt would have been to identify the overarching patterns. **In short, in an ideal world, a machine learning model and the team that develops it should be representative of ground truth.**

The Four Stages of Mitigating Bias in an AI Model



1 Shankar, Shreya, et al. "No classification without representation: Assessing geodiversity issues in open data sets for the developing world.preprint : 1711.08536(2017).
2 https://www.openstreetmap.org/#map=4/21.84/82.79
3 Jean, Neal, Rachel Luo, and Jae Hyun Kim. "Nighttime Light Predictions from Satellite Imagery."
4 Doll, Christopher NH. CIESIN thematic guide to night-time light remote sensing and its applications."
Center for International Earth Science Information Network of Columbia University, Palisades, NY (2008).
5 Kubat, Miroslav, Robert C. Holte, and Stan Matwin. "Machine learning for the detection of oil spills in satellite radar images." Machine learning 30.2 (1998): 195-215.
6 Sun, Yanmin, Andrew KC Wong, and Mohamed S. Kamel.Classification of imbalanced data: A review."
International journal of pattern recognition and artificial intelligence 23.04 (2009): 687-719.
7 Narayanan, Sudha, and Sharada Srinivasan. No Country for Young Women Farmers: A situation analysis of India. Working Paper, Indira Gandhi Institute of Development Research, 2020.

Open Data Cube in SatSure

Thaiseer Parammal,
Team Manager - Platform Development,
SatSure

Remote Sensing is nothing but observing and analysing objects from a distance, using a sensor usually installed on a satellite called earth observation satellites, and recording this information captured on a timely basis to study the earth. We call it **Earth Observation (EO).**

Some of the freely available earth observation data are ESA's Sentinel sensors based satellite images, NASA's Landsat or Modis based data products. These freely available data products have made access to earth observation data for industry experts, academic researchers, and institutions much simpler. This is making way for breakthrough developments in how earth observation data is perceived and it's usage and applications. These data and the applications or insights they bear have tremendous potential to deliver breakthrough decision intelligence which can have a high impact on agriculture, infrastructure, climate change, or any other verticals.

With each passing year, the increase in EO satellite data volume is rapidly increasing along with their global coverage. The applications and insights built on top of these are mainly driven by technological advances and open data policies adopted by governments and space agencies. Availability of these data is no longer the major bottleneck for this data's potential to deliver these high impact insights. The pipeline of data connections from the satellites to the insights engine to the end-user resulted in most chunks of the earth observation data being underutilised.

Under the Committee on Earth Observation Satellites (CEOS) leadership, the Open Data Cube (ODC) initiative was started to address the above problem. Founded in 1984, this international organisation acts as the primary forum for global coordination of space-based Earth observation. The Open Data Cube initiative aims to provide a solution in the form of a data architecture that adds value to its users and increases EO satellite data's impact.

The Open Data Cube (ODC) is an Open Source Geospatial Data Management and Analysis Platform that helps you harness the power of Satellite data. According to The Open Data Cube Initiative, its main objective is to increase the impact of satellite data by providing an open and freely accessible exploitation tool and to foster a community to develop, sustain and grow the breadth and depth of applications1. It presents a common analytical framework composed of a series of data structures and tools that facilitate the organisation and analyse large gridded data collections.

In SatSure, we harness Open Data Cube's power to scale the insights-delivery pipeline of our geospatial analysis derived from the historical EO data on a pixel level granularity.

Why Open Data Cube?

Spatial Data Infrastructures (SDI) were created to properly discover, access, manage, distribute, reuse, and preserve storage for these big data sets. Some of the most commonly used and popular platforms among them are:

- Open Data Cube (ODC),
- Google Earth Engine (GEE),
- System for Earth Observation Data Access,a Processing and Analysis for Land Monitoring (SEPAL)
- OpenEO

In the article, ***An Overview of Platforms for Big Earth Observation Data Management and Analysis***, these platforms were evaluated on the key metrics of the platforms, ODC outruns the others. Fig. 2 is an illustration of processing scalability vs the storage scalability of the platforms. All platforms perform high on storage

Table 1: A comparison of the four platforms on a scale of low, medium and high on different factors:

Key Metrics / Platform	ODC	GEE	SEPAL	OpenEO
Processing abstraction	**Medium:** Xarray and celery	**Medium:** Predefined pixel-wise functions	**Low:** User runs his own code	**Medium:** User-Defined Functions, Process graphs and Jobs
Open Governance	**High:** Defined governance process	**Low:** Proprietary software, closed source software	**Medium:** Only open source repository	**Medium:** Only open source repository
Infrastructure replicability	**High:** Open source code, docker containers and documentation available	**Low:** Proprietary closed source software	**Medium:** Open source code with basic documentation available	**Undefined:** Dependent on the backend used
Processing scalability	**Medium:** A template application available (Python and Celery)	**High:** Code automatically executed in parallel using a Map Reduce approach	**Low:** User runs his own code	**Undefined:** Dependent on the backend used
Storage scalability	**High:** Distributed File System, S3 and HTTP	**High:** Google storage services	**High:** Google storage services	**Undefined:** Dependent on the backend used
Data access interoperability	**High:** OGC Services	**Medium:** Tile service	**Low:** Without any ease	**High:** OGC Services
Extensibility	**High:** Open source and modular code	**Low:** Proprietary closed source software	**High:** Open source	**Medium:** open source software with proprietary software

performance, scalability, replicability etc., with ODC outperforming all of them.

On comparing the data access interoperability and extensibility (Fig. 1) of scalability, whereas ODC and GEE are on medium and high in terms of processing scalability. Upon comparing the Open Governance and extensibility in Fig. 3, of all these platforms, ODC is ahead of others by far.

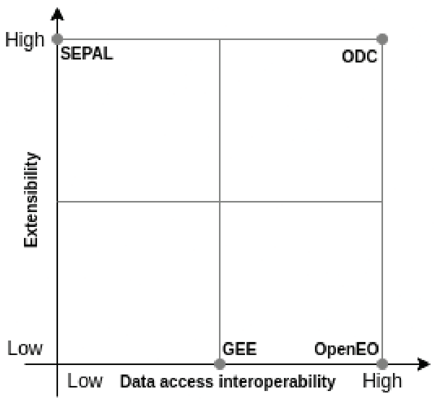


Fig. 1 - Extensibility vs Data Access Interoperability

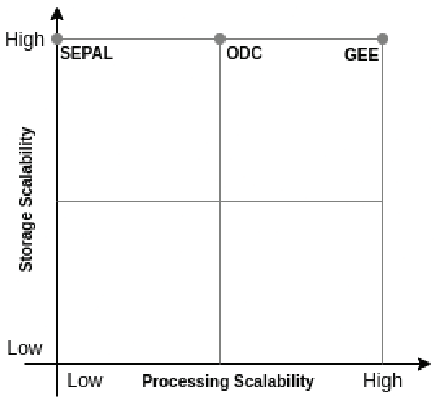


Fig. 2 - Storage Scalability vs Processing Scalability

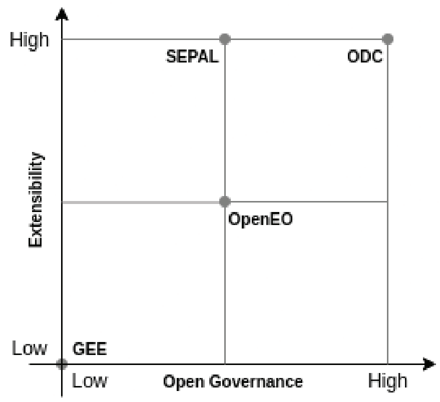


Fig. 3 - Extensibility vs Open Governance

Overall, ODC was found to be the one that is best suited for the requirements of SatSure based on the parameters analysed.

ODC - Under the Hood

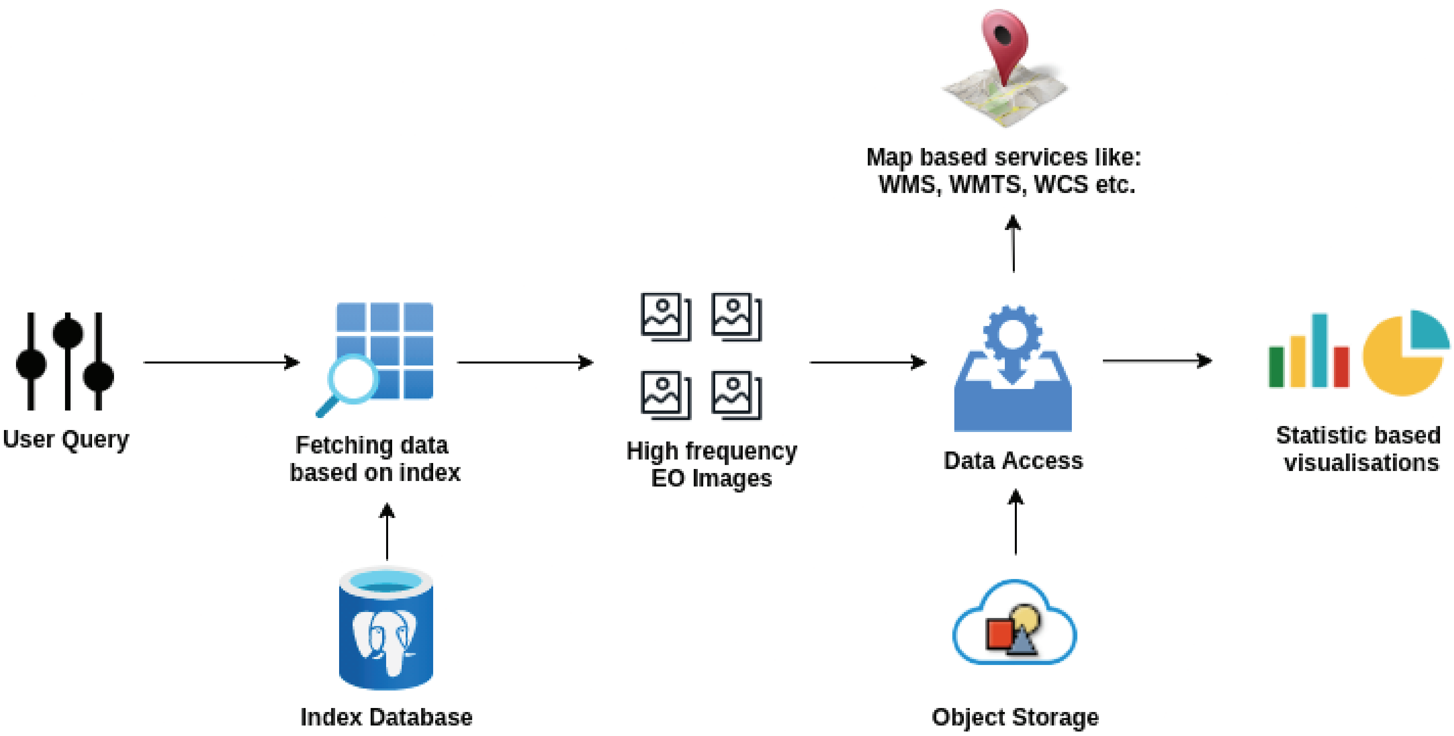


Fig. 4 - Open Data Cube Query Flow

ODC uses standardised schemas to define the products/datasets available, using this information to index the datasets available for a product and keep the metadata in a structured format in the index database. The index database is powered by

PostgreSQL. These metadata will have information about the actual location of the Cloud Optimized GeoTIFF (COG) corresponding to each dataset.

Based on the user query, ODC makes a lookup in the index database to determine which datasets correspond to the query and groups them based on the request. Upon receiving this information, it requests the cloud storage to fetch the actual COGs and generates the required images or statistics.

The requests can be in terms of map services like WMS, WMTS, WCS or statistics based WPS requests. Using ODC in the backend, users will be able to tune the queries for specific requirements and fetch the underlying data from a vast collection of datasets available underhood with ease.

The queries can be modified to fetch the products like below on a real-time basis:

- Point or boundary-based extraction
- Applying resampling techniques
- Applying different projections
- Fetching based on time information

For example, a query optimised to fetch the underlying image of a specific product/region/ time on a visualisation layer will look like the following.

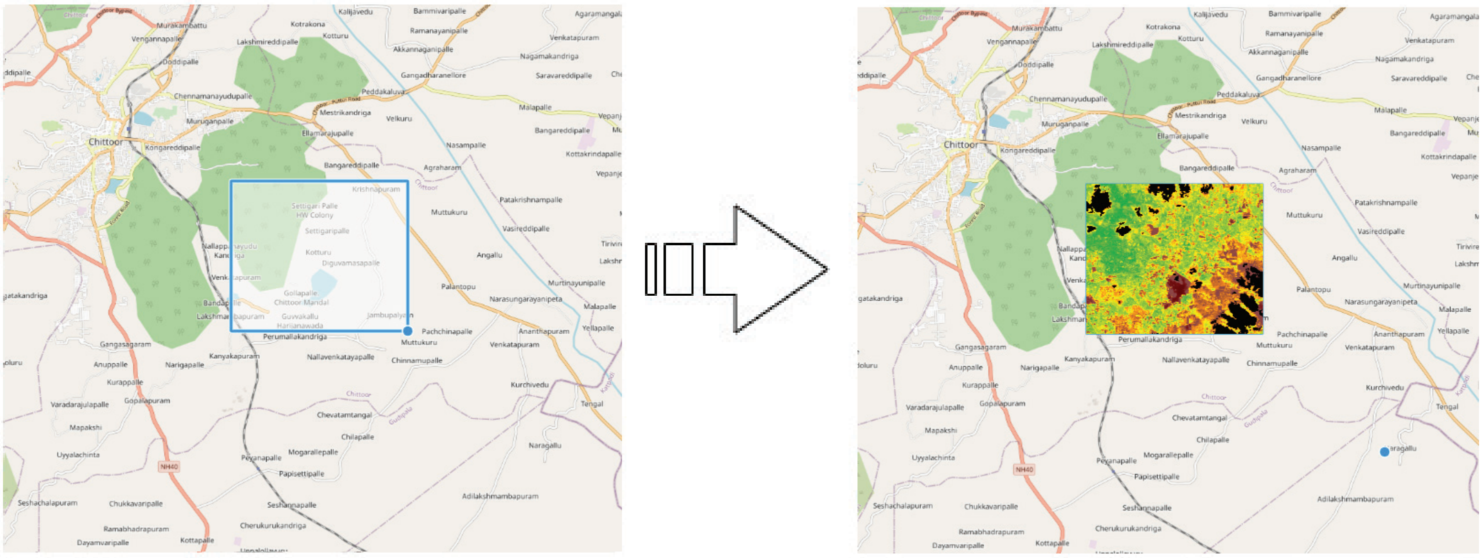


Fig. 5 - Fetching the map for a specific product/ boundary based visualization

ODC - In SatSure

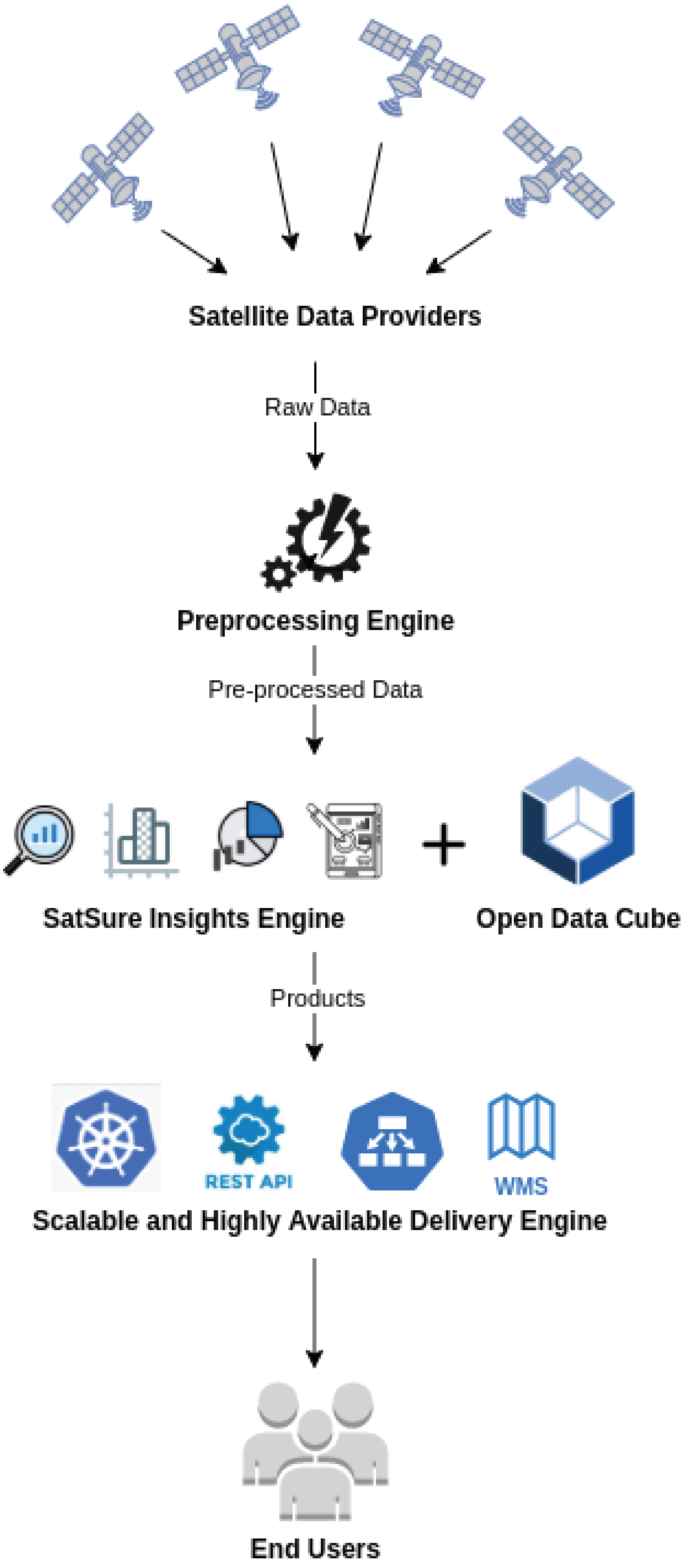


Fig. 6 - High-Level Architecture of SatSure ODC Implementation

In SatSure, we have satellite data from different sources like ESA's Sentinel sensors and NASA's Modis products. We pull these data from their source archives into our AI/ML analytics engine. Before this is fed into our engine, the data undergoes a preprocessing series to be analysis-ready.

In the insights engine, there are proprietary machine learning models and remote sensing semi-automatic algorithms. They generate insights and transform these earth observation data into directly actionable products for solving high impact environmental, economic and social challenges. These products are delivered to end-users in different sectors using various technologies, including Insight Dashboards, RestAPIs, WMS layers, etc., in our delivery engine. This engine is powered using high availability and highly scalable modern architectures of Kubernetes services in our cloud, as described above in Fig. 5.

The Future Ahead

In the wake of five years into the UN's Sustainable Development Goals (SDGs), the transformational vision and new data requirements called for to realise the 2030 Agenda has only been partially realised. The extent of this challenge has been underestimated and is further amplified by knowledge and availability of geospatial data. The SDGs are highly dependent on geospatial information and Earth observations as the primary data for relating people to their location and place.

In the 11th meeting of the Inter-agency and Expert Group on SDG Indicators Working Group on Geospatial Information (WGGI), a vision was created to see geospatial and location-based information to be recognised and accepted as official data for the SDGs and their global indicators.

At SatSure, by harnessing Open Data Cube's power, we are ready to scale the United Nations' vision in achieving SDGs by enabling people's access to geospatial information at various levels and by democratising modern location awareness requirements using EO datasets.

References:
https://www.opendatacube.org/
https://www.swissdatacube.org/index.php/edsproject/bringing-open-data-cube-into-practice
Gomes VCF, Queiroz GR, Ferreira KR. An Overview of Platforms for Big Earth Observation Data Management and Analysis. Remote Sensing. 2020; 12(8):1253.
https://doi.org/10.3390/rs12081253
https://ceos.org/
https://unstats.un.org/sdgs/files/meetings/iaeg-sdgs-meeting-11/6b.%20WGGI%20SDGs%20Geospatial%20Roadmap%20Outline.pdf

CREDITS

Cover page & Articles

Image under license from Shutterstock.com

Article 1: THE INDIAN AGTECH PARADIGM

Page 5

Image from Shutterstock

Page 7

Created by IconMark from the Noun Project

Created by Gan Khoon Lay from the Noun Project

Created by Colourcreatype from the Noun Project

Created by Flatart from the Noun Project

Article 2: The “India Agricultural Platform”

Page 10

Image from Freepik

Page 13

Created by Matt Caisley from the Noun Project

Created by glyph.faisalovers from the Noun Project

Created by Made by Made from the Noun Project

Created by ProSymbols from the Noun Project

Article 3: Fair AI for Satellite Data Applications

Page 16

Photo from unsplash

Page 19

data collection by Becris from the Noun Project

Data Analysis by Vectors Point from the Noun Project

Data Model by Tomas Knopp from the Noun Projec

Policy by Lars Meiertoberens from the Noun Projectt

Article 4: Open Data Cube in SatSure

Page

Image from Shutterstock