CODR Bits & Bytes

VOLUME 2 ISSUE 1 Feburary 10, 2021

EDITORIAL COMMENT

A robust Data Engineering process is very important in creating Data Science solutions

For any government department which is dependent on a huge amount of data it generates regularly, data analytics becomes extremely important for better decision making. However, there are several challenges that can impede policy maker's ability to use data analytics to arrive at a quick policy decision.

With the introduction of big data, and the advent of data driven organisations across all the departments, policy makers are often overwhelmed with the amount of accumulated data. Every bit of information is collected, leaving the Analysts with a huge number of interlocking datasets to perform data analytics.

Here is where the need of a data engineering process that automatically collects and organises the relevant information becomes very important. Data Engineers need to dive into the huge pile of data to decide what is important and what is not, and then create the scripts to automate the process. But this can only be done after careful analysis of the objectives and end results of the system – or multiple systems. The challenge does not end there. Every addition of a new report, or change in an existing report, may also result in a change in the data engineering process.

If one examines any successful data analytics and AI/ML implementation carefully, they will notice that about 70% of the time is consumed in the data engineering process, 20% in data analysis & visualisation and only 10% in creating AI/ML models. This clearly shows the importance of the data engineering process in creating Data Science solutions.

INTERVIEW

Mrinlani Kabbur

Consultant

Public Affairs Centre

Q. What are the challenges, based on your experience, in the application of data science when dealing with governance issues?

A: *To develop data science applications, defining the objective and outcomes is critical. It has been a challenge to understand the pain points or the problems concerning governance. Secondly, most of the data available with the state departments is aggregated at the district level. For a deeper analysis, individual or household level data is necessary. It is also seen that the disaggregated data at the taluka and district offices is still not digitised. Finally, there is no unique individual or household identifier in the data collected by various departments. This is very important to bring the disparate datasets together.*

Q: According to you, what is a priority area of data usage for the government?

A: *Currently, PAC-CODR is focusing on analysing data relating to Sustainable Development Goals (SDGs) and Special Development Plan (SDP). It has taken up data analysis on key indicators in aspirant and performer category of SDGs -1, 8 and 10, 2, 3, 4 and 5. SDGs-1, 8 and 10 is developing a macro economic model for*

CODR Bits & Bytes

VOLUME 2 ISSUE 1 Feburary 10, 2021

poverty alleviation and inequality reduction. SDG-2 has calculated the potential yield of dominant crops in each district and analysing its determinants. The factors influencing child, maternal mortality and universal immunisation outcomes are analysed in SDG-3. The SDG-4 team is analysing the household characteristics leading to secondary drop outs while SDG-5 is developing a Gender Inequality Index. We are also performing data analytics on the SDP indicators for the upliftment of backward talukas with a special focus on human development challenges in North Karnataka. To ground-truth the results of the quantitative analysis of SDGs and SDP, we have planned a qualitative analysis based on primary data.

Q: How do you think PAC can contribute to Government of Karnataka with reference to data analytics?

A: *The government officers have a good understanding of ground realities, be it the problematic geographies or the critical issues. Data analytics can augment this understanding by providing evidence based rationale. It can also quantify the actual-normative gap, provide trends, patterns, incidence and rate of occurrence for focused interventions and resource allocation decisions. The application of data science can also contribute to identifying missing data gaps and propose surveys to collect the same.*

Conceptualised & Designed by

www.pafglobal.org

PUBLIC AFFAIRS FOUNDATION
Partnership for Better Governance

UPDATES

24 Dec 2020

Conducted virtual meeting with Health Department officials to give a brief introduction about the Data Integration project and the current project plan.

28 Dec 2020

Conducted another virtual meeting with Health Department officials to discuss more in detail about the datasets available with the Health Department with respect to People, Money, and Land & Assets.

30 Dec 2020

A meeting was chaired by Dr. Shalini Rajneesh, ACS, Planning and PAC where the PAC team presented the findings of the first set of reports on SDP, SDG 1, 8, 10, 2, 3, 4 & 5. Among other things, suggestions to the current approach for some SDGs and next steps to be taken were discussed. A detailed MoM on this meeting is already sent.

08 Jan 2021

- A presentation and a demo of HRMS system was conducted by the HRMS team.
- Conducted a meeting with the Health Department to get a demo of the eHealth system.

12 Jan 2021

Dr. Narendar Pani, the SME suggested by the Planning Department, conducted a virtual session on all aspects of SDGs and SDP for the PAC team, during which he addressed the broad conceptual framework and made suggestions on the next steps.