VIEW POINT

ANNOTATIONS FOR BETTER CUSTOMER INSIGHTS

Abstract

The impact of Covid-19 has energized customer expectations for products and services delivering technology-led experience, convenience, and comfort. Understanding and responding to these fast-changing expectations demands speedy, almost-'live' processing of the vast amounts of customer data available today. While several businesses are already utilizing artificial intelligence for this purpose, this paper presents how market success depends on rightly annotating the data so that it yields better customer insights.

Infosys®
Navigate your next

The demand for data insights, at speed

As the saying goes "data is king," and there is no denying this fact with the amount of data created every second. Today, for almost every business, the ever-increasing gargantuan amount of data being generated holds the key to building strong customer insights and enabling better decision making. A proper analysis of this data can help to understand customer behavior, demographic-driven expectations, market changes, as well as the readiness, adaptability, sentiment, and value delivery of a business to serve its customer better.

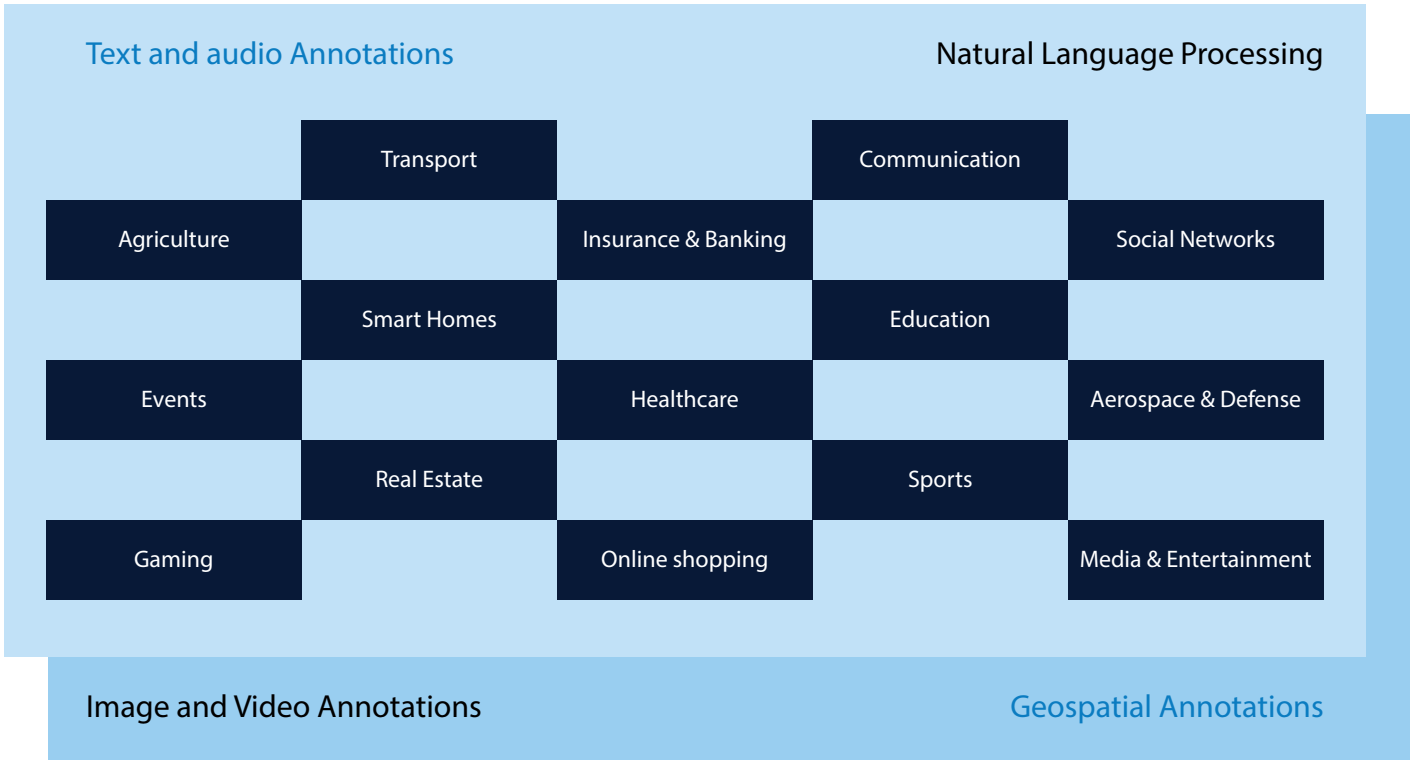
Typically, a large volume of data needs both time and a large workforce to execute data processes. Also, because of the impact of COVID-19, customer expectations for technology-led experience, convenience, and comfort have only grown. However, understanding and responding to these changing expectations demands speedy processing of data and businesses using only manual data interventions would not achieve the desired customer insights in time to capitalize on them. This is because newer market entrants are mostly leveraging technology developments to analyze data quicker, rapidly provide unique value propositions, and gain the first mover advantage. Resultantly, established players are forced to scramble to develop equivalent offerings within an even more limited timeframe.

Thus, the requirements of the market for speedy data processing and at scale is very high, and it is here that artificial intelligence and machine learning (AI/ML) with its ability to process large volumes of data is playing an effective role. Many companies are now adapting to the changing times and embracing these digital means to understand their data better. As an outcome, AI/ML has certainly made life easier across multiple industries for both customers and companies through enabling quick, efficient, effective decision making and a whole new customer experience.

A starring role for annotation services

Where AI/ML are used to analyze massive volumes of data, training datasets play a very critical role in building the overall effectiveness of these machine learning systems. For training the system, annotation services help categorize and structure the data into datasets in a way that helps the machine learning algorithms achieve business outcomes more efficiently.

Using quality training data, with proper data annotations and labelling, helps enhance the training datasets and builds robust models that can handle large volumes of data with relative ease in an almost 'live' manner. This enhances the business' visibility into expected outcomes and enables better foresight of the project execution cycle when building machine learning programs.



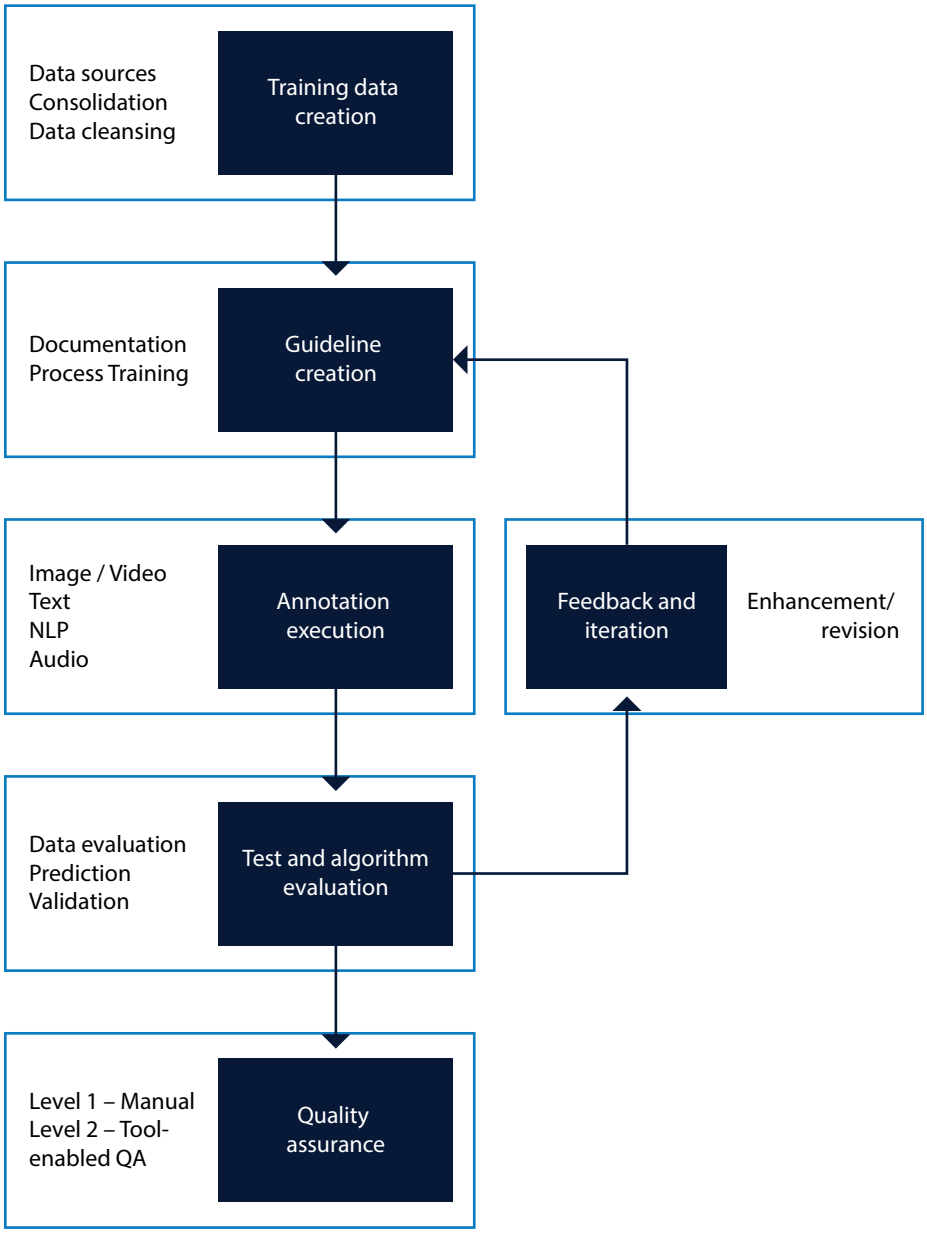
By now, most industries have realized that machine learning algorithms demonstrate heightened efficiency when provided with high quality data inputs. While the initial investments required to improve data quality are considerably high, the benefits outweigh the costs in the long run when these models are deployed effectively to understand and deliver what customers really want. Undoubtedly, companies using quality data for their AI/ML initiatives to enhance their services inculcate a culture of building exemplary customer experience and reap exponential growth.

This relevance of annotations to business growth is reflected by the growth of the market for its services, as noted in several analyst reports from NASSCOM. The global data annotation tools market size was valued at $494 Mn in 2020 and expected to grow at a CAGR of 27.1% for the period till 2028, while India as a market alone is expected to exceed $7 Bn by 2030. While ~75% players are in the initial and growth phase, 70% of global players are in the intermediate to advanced phase. Also, managed services contribute to 65-70% of the market size. [1]

[1] https://nasscom.in/knowledge-center/publications/data-annotation-billion-dollar-potential-driving-ai-revolution

Iterating for data quality

The machine learning process is an iterative one, requiring multiple levels of enhancements to achieve the desired results. Also, the outcomes of any model are often fuzzy with cross-collaborations across multiple teams necessary to enhance the process and define outcomes as they evolve. Thus, appropriate data sets should be generated/used while building the algorithm to ensure bias and data fairness aspects are accounted for.



Apart from this, there are other important facts to consider while performing annotations. Adequate guidelines need to be provided for data cleansing and classification, since every use case is different, and individual industry requirements are unique. Also, with nuances within similar industry being high, and methods to build efficient machine learning process quite unique, one size does not fit all. There is no "one way" of doing annotations and it all depends on the models being developed.

Yet, it is important to remember that quality data and the ability to provide enhanced delivery output enables an efficient machine learning model. Thus, manual data quality reviews need to be performed, building on consensus from multiple user inputs.

Using the right talent matters

The 2.5 quintillion bytes of data being generated every day, only increases the challenges for businesses to glean the insights required to enhance value for end customers. This requires not only identifying the right data, but also skilled professionals with the domain knowledge, data handling, technical skills, and analytical expertise to structure the data and generate a refined training dataset that is suitable for machine learning.

Such quality training data when sourced through efficient data handlers helps machine learning processes to learn quickly and effectively implement a solution in half the actual time that would otherwise be required. Quality data services are hence the crux of the machine learning process, and both the people as well as methods and models leveraged for execution need to be considered as investments towards seamless product functionality.

Weighing the options

Like any other process, most annotation data projects come with their own challenges. It gets even more challenging when the tasks are complex and with sensitive data involved. Thus, multiple factors contribute to any company's decision-making on how these projects are to be executed and need to be carefully analyzed prior to any decision. They include the volume and sensitivity of data involved, budget allocated, process complexity, tool dependencies, timelines to achieve outcomes, data security requirements, people dependencies, the expertise and the time required, and other such factors.

After factoring in these details, the deliberations typically involve the following choices:

- Should the project be driven with an inhouse talent pool to reduce cost implications?
- Or, should a dedicated partner with adequate experience be leveraged to handle the projects for seamless execution?
- Or, should crowd sourcing methods be leveraged to execute projects for fast data handling at a large scale?

Often, once these questions are answered, companies begin evaluating the risk through smaller execution cycles, either through pilots or short-term contractual engagements. Then during the execution phase, the projects go through changes and some unique requests develop as they progress, becoming a continuously evolving process.

Modeling for success

Very large annotation projects with high data volumes and either simple or complex annotation activities to be performed are generally handled better in a managed services model. The higher data volumes — and with some projects needing multiple resources handling homogeneous data to predict quality model outcomes — require multiple iterations, and hence there is a need for a strong team to be involved to add value to projects.

Also, most such project executions are time-bound, requiring a short-term approach to the execution and development of appropriate models, to enhance outcomes. In these cases, service providers using an Agile methodology would be key to seamless execution and handling change efficiently.

For example, annotation services are very important when working on computer vision projects, helping in enhancing the data quality and providing better rules for the machine to learn. The volume of data in such projects is enormous and deriving meaningful insights to help achieve outcomes that make a difference to all stakeholders would require focusing a minute lens on cost, effort, and time.

Thus, while processes that are mostly simple can be delivered through crowd sourcing, processes that are complex need to be executed through third-party vendors or managed services. On the other hand, critical projects that are sensitive and require high skill are often better done with in-house resources.

Partnering for the journey

As we've seen, especially for larger more complex projects, selecting a right partner with expertise and experience in handling annotation services would be more rewarding than going it alone or crowd sourcing. Handling the massive amounts of data involved would otherwise require recruiting and deploying quite a large workforce. Additionally, the costs — in case of any need to rework in-house or crowd-sourced projects — would be quite high. Another advantage of working with an external service provider is that project scale-up timelines would be minimal, amidst the competitive market pressures to quickly deliver.

However, realizing true business value would require close collaboration and participation from the service provider throughout the iterative process. Also, selecting a competent provider would require evaluating their capabilities to handle sensitive data and personally identifiable information, as well as data security requirements such as GDPR. The provider would also need to be able to meet customers' expectations of localization, and have expertise in handling the necessary tools, technology, and infrastructure.

In any case, whether annotating in-house, through a service provider, or through crowdsourcing, getting it done right will determine any organization's success in meeting the implicit, unmet, and often misunderstood actual needs of customers.

Authors

Mothiraj R

Senior Practice Lead - Utility Squad, Digital interactive Service, Infosys BPM

Mothiraj has over 18 years of experience spanning customer service, operations management, center of excellence and solution design. In his current role, he supports service delivery for clients in the AI/ML annotation and GIS services. He manages a team that executes critical client project and is also responsible for the center of excellence activities that include upskilling resources on domain, identifying best practices and implementation in these services.

Mothiraj holds a bachelor's degree in business management and a post graduate diploma in business administration.

Sathish Kasthuri

Senior Solution Design Head, Digital Interactive Services, Infosys BPM

Sathish is responsible for GTM strategy and implementation of new solutions across the Digital Interactive Services portfolio. Sathish has been with Infosys since 2010, and has held key positions across sales, transition, operations, and solution design. Prior to joining Infosys, he has been with several niche Media & Entertainment companies including animation studios. He has more than 25 years of experience across several geographies including the US.

Sathish has a Master's in Engineering from College of Engineering, Guindy, and has done his Bachelor's in Engineering from PSG College of Technology.

For more information, contact infosysbpm@infosys.com

Infosys®
Navigate your next