Cognilytica Research
iMerit

Briefing Note

Doc. ID: CGBN157

8th Apr, 2019

ABSTRACT

In order for many machine learning algorithms to be trained, especially supervised learning algorithms, they need to be fed relevant data that has been appropriately "labeled" with the required output that needs to be learned. However, there is a chicken-and-egg problem with systems automatically being able to label images (if they could automatically label the image, then why would you need to train them based on labeled images), all data labeling solutions are by their very nature human labor oriented. Humans must use their cognitive power to label and annotate images in such a way that machines can use those labels and annotations as part of the training process. **iMerit** provides "technology-enabled services" as they have tools to help humans provide the right annotations as well as verify and audit the annotations. They also recently partnered with with Amazon Web Services (AWS) to provide data labeling services on AWS SageMaker Ground Truth.
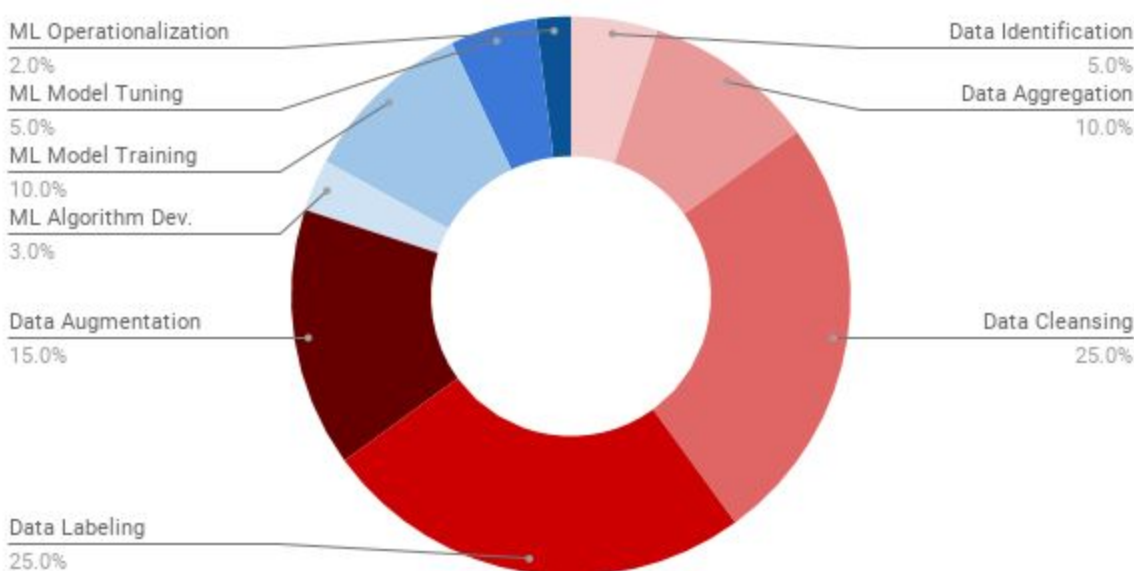
Market & Industry Context

In order for many machine learning algorithms to be trained, especially supervised learning algorithms, they need to be fed relevant data that has been appropriately "labeled" with the required output that needs to be learned. Algorithms must be trained on hundreds to thousands of images that have been precisely marked in the data labeling process. For example, image recognition systems that use deep learning neural network approaches need large volumes of clean, normalized image data where the image has been properly labeled as the desired output to train the system over multiple training iterations to build a model that can generalize properly to recognize future images. Such labeling needs to happen for any supervised learning application.

Data preparation activities, including data labeling, takes up a significant amount of the time for most AI and machine learning projects. See below:

Percentage of Time Allocated to Machine Learning Project Tasks

Source: Cognilytica

- ML Operationalization 2.0%
- ML Model Tuning 5.0%
- ML Model Training 10.0%
- ML Algorithm Dev. 3.0%
- Data Augmentation 15.0%
- Data Labeling 25.0%
- Data Identification 5.0%
- Data Aggregation 10.0%
- Data Cleansing 25.0%

However, there is a chicken-and-egg problem with systems automatically being able to label images (if they could automatically label the image, then why would you need to train them based on labeled images), all data labeling solutions are by their very nature human labor oriented. Humans must use their cognitive power to label and annotate images in such a way that machines can use those labels and annotations as part of the training process.

There are only a few ways to get access to well-labeled data for machine learning purposes:

- *Internal, Self-Managed Human Labor* - Use internal workforce to do labeling.
- *Outsourced, Self-Managed Human Labor* - Use third-party providers to provide contract labor while company manages the work pool and quality of output.
- *Third Party Managed Labeling Providers* - Use third parties specializing in labeling to provide labeling workforce as well as manage workers and quality of work output.

- **User-Driven Labeling** - Large user base companies such as Google, Facebook, Amazon, Netflix, Quora, and Yelp, are using their users to do their image and other labeling work for them, perhaps without their explicit knowledge that they're using them for labeling. For example, CAPTCHA.
- **Pre-Trained Models and Existing Labeled Data Sets** - Access already trained machine learning models that can be extended via Transfer Learning and other approaches, or access the underlying labeled data set to provide a starting point for machine learning projects. Existing labeled data sets include ImageNet, MNIST, Visual Genome, many government and non-profit sources. Many vendors offer pre-trained models for specific application domains.

Customers have the data, but they don't have the resources to label large data sets, nor do they have a mechanism to insure accuracy and quality. Raw labor is easy to come by, but the assurance of quality is not easy to guarantee. In addition, labeling projects involve multiple steps as well as requiring human subjective decision making. Third party managed labeling solution providers address this gap by providing the labor force to do the labeling combined with the expertise in large-scale data labeling efforts and an infrastructure for managing labeling workloads and achieving desired quality levels.

Company / Solution Overview

iMerit was founded in 2012 as a for profit social enterprise to employ marginalised people who have been trained in computer literacy and soft skills. The company saw the potential of the Indian talent pool and created a for-profit company with a double bottom-line: its financial / company performance and its societal impact. The majority of their 2000+ employees are in India, but the company recently opened an office in New Orleans, Louisiana, USA following the same social model they follow in India. The people work full time for iMerit at their secure facilities and are not remote or outsourced workers.

iMerit has rapidly evolved into data services, particularly with the explosion of Data Science and AI. They saw that information was growing in quantity but the main issues were around quality. Customers come to them when they are trying to do supervised learning machine learning and have experienced challenges in labeling image, video, and text data at the desired scale and consistency and automation, crowdsourcing, or pre-trained models no longer work.

Labeling can involve multiple steps as well as requiring human subjective decision making and iMerit is able to deliver on this. iMerit provides "technology-enabled services" as they have tools to help humans provide the right annotations as well as verify and audit the annotations. They provide the human workforce to power the solutions while also providing security, scalability, auditing capabilities to see where it's going wrong and fix things that are being misunderstood and misinterpreted, and adaptability especially around edge cases.

The company can work on their customers' platforms, third-party platform, or any other platform that the customer requires, and can also access information from their customer sites through their customers' APIs. The company is a technology-enabled services company, but they aren't focused on being an exclusively technology focused company. They call this a "Service Delivery Platform"- it's the iMerit People Platform (iMPP).

Recently, the company partnered with AWS to provide data labeling services on SageMaker Ground Truth. SageMaker Ground Truth is a fairly new capability of Amazon SageMaker that makes labeling large datasets easier, faster, and more accurate. This partnership allows iMerit the ability to provide labeling services for the many customers that use SageMaker providing more exposure and creating an additional revenue stream for the company.

Customer Use Cases & Value Propositions

This is an enterprise-level service with enterprise-level pricing, support, and quality.

- **Autonomous Vehicles** - One of the largest market segments for iMerit are tech-forward divisions of large automotive firms and emerging autonomous vehicle companies. The primary problem being solved is around image recognition labeling and annotation. iMerit sees image and sensor data, which the company segments and annotates within a scene, including tracking objects, which then become inputs to the machine learning algorithms used by the AV companies. These workflows can be quite complex at over an hour an image and require strong infrastructure and deep expertise.
- **Aerial Imagery -** In this market segment, iMerit can derive insights and analysis from aerial imagery including drone and satellite for a wide range of applications. They see "alternative data" that are not directly obvious to humans on the ground. Applications include detecting deforestation, usage of natural resources, building and other analysis. Insurance companies are using this data to develop better risk analysis models.
- **Speech and Text Applications** - Customers use iMerit to help train their speech systems using audio data from a wide range of sources. iMerit assists with tagging content for things like sentiment analysis, adding more industry-specific vocabulary, building knowledge graphs for contents, and extracting more information in this space.
- A partnership with **Amazon Web Services (AWS)** was recently announced for their AWS SageMaker Ground Truth platform, with additional customers mentioned in the press release for that announcement.
- Additional customers include **eBay, New England Historical Genealogy Society, Crowd Reason**, and others for a wide range of use cases.

iMerit sells service engagements typically based on hours or months and does not license its tools. However they do support customizations or integrations for a fee.

Customers can work with IMerit as a full time experience, or for a short time frame to augment an existing team. Pricing for these services can be transactional (per labeling) or monthly. Within the scope of the project, some of the people are allocated to the labeling and some to QC to ensure quality and accurately.

Company / Solution Profile

Company & Solution Profile (as of Apr. 2019)	
Company Name	**iMerit**
Founded	**2012**
Company Stage / Funding	**Growth Stage, $4.5+ Mil. Funders include Omidiyar Network, Khosla Impact fund, Michael and Susan Dell Foundation, and more (source: Crunchbase)**
CEO	**Radha Basu**
Contact information	**14435C Big Basin Way, #256 Saratoga, CA, United States 95070 Jai Natarajan, jai@imerit.net www.imerit.net**

Cognilytica Market Positioning
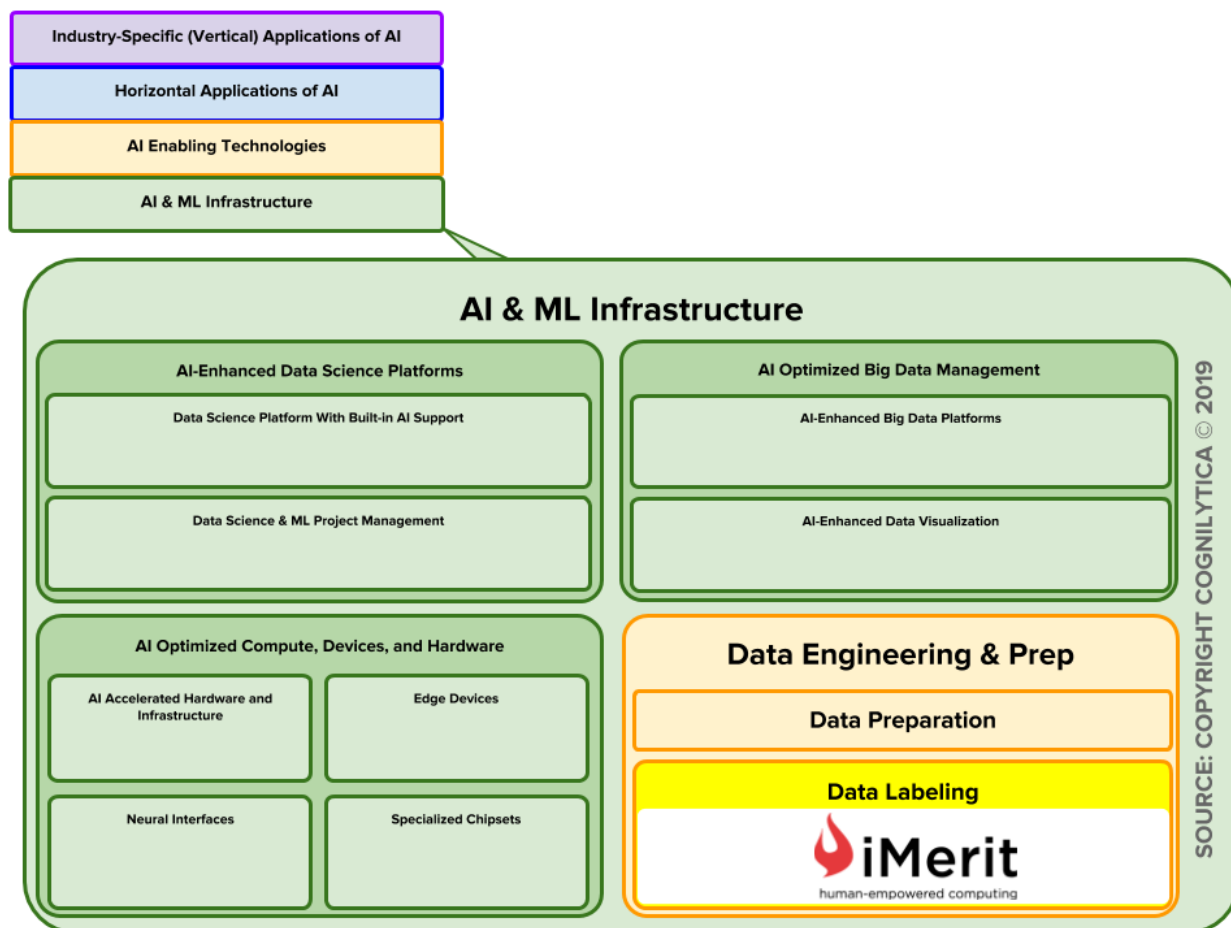
The Cognilytica Take

The market for data labeling is growing and becoming very important as more and more companies need clean, accurate, well-labeled data for ML training. The crowd-sourced solutions to labeling are now considered the amateur, first-mover approach. The maturing market is demanding more scalable, proven repeatable solutions. This is why there is a growth of third-party labeling firms which requires a particular skill set and technology enablement that is not trivial to make labeling work at scale, without reduced quality.

There is some competition in the market including **Scale.AI, CloudFactory**, and to a limited extent, **Figure Eight** (which was recently acquired by **Appen**). However, the primary competitive advantage that iMerit has is its delivery excellence and quality of its solution architects. They also have greater flexibility on technology as they aren't rigid on the required platform of interaction and work with customer's existing technologies and platforms. Their recent partnership with AWS also shows the high quality of the product they deliver. iMerit's unique in their impact mission of helping people and they have a diverse workforce (50%+ women) which are competitive strengths. The work of data annotation and labeling is fairly intense and the need is only growing. When looking for quality data labeling enterprise-grade providers in the market **iMerit** should be considered a leading option for organizations looking for data labeling and preparation capabilities.

Related Research

➢ Data Engineering, Prep, and Labeling for AI 2019 Report (CGR-DE100)
➢ Paxata QuickTake (CGQT143)

About Cognilytica

Artificial Intelligence (AI) and related technologies will impact all industries and all corners of the world. Without insight into how AI will impact you and your business, you risk being left behind. Cognilytica is an analyst firm that provides real-world, industry and adoption focused market research, intelligence, advisory on Artificial Intelligence (AI) and related areas.

● Cutting through the Hype by Focusing on Adoption — Cognilytica cuts through the noise to identify what is really happening with adoption and implementation of AI in public, private, and academic settings. We focus on the usage of AI in the real world, not the buzzword hype.
● Industry-Leading Market Research — Market-level research on application, use cases, and comparative research on the state of AI adoption in the industry. Focusing on real-world adoption of AI technology and cutting-edge application.
● Advisory with Knowledgeable Experts — Get access to knowledgeable research analysts that spend their time immersed in the world of AI implementation and adoption.
● Research through Conversation — Cognilytica generates its research through direct conversation with industry thought-leaders, technology practitioners, and business decision-makers. We ignore the press releases and skip the hype to produce unique, original research through direct engagement.
● Deep Engagement Opportunities — Connect with peers, industry leaders, experts, and influential practitioners at subscriber-only exclusive workshops, events, and seminars aimed at advancing the state of your AI adoption.

Cognilytica analysts publish research reports, white papers, and briefing notes at regular intervals that are available to our annual subscribers as well as for one-off purchase. Cognilytica offers advisory time with analysts virtually or on-site Analysts are also available for commissioned research projects, white papers for internal or external consumption, and speaking engagements at client events or public conferences. If you have an AI market intelligence or research need that can be fulfilled with our industry knowledge, body of research, methodology, and expertise, Cognilytica analysts are available to assist. Visit www.cognilytica.com for more information on opportunities you can take advantage of.