D&B Survey Finds Data Practices Lacking, Creating Opportunity for Data Preparation Firms

By AI Trends Staff

Businesses are missing revenue opportunities and losing customers due to bad data practices, according to a new underline report from Dun & Bradstreet, based on a survey of 500 decision-makers in the US and the UK.

The survey found 20 percent of businesses have lost a customer, some declining to sign new contracts, due to incomplete or inaccurate information about the customer.

Some 17 percent of organizations offered too much credit to a customer due to a lack of information about them, and lost money as a result, the report found.

Compliance is twice as big of a concern in the UK than the US (31 percent vs. 16 percent), which may reflect the challenge of meeting the requirements of the General Data Protection Regulation (GDPR), in effect in the European Union. Already, over 10 percent of organizations under the new rules report having been fined for data issues.

The way that data is structured appears to be a significant barrier in many organizations, with indications that data is often poorly structured, difficult to access and out of date. Nearly half of business leaders (46 percent) say that data is too siloed to make any sense of it, with the biggest challenges to making use of data being:

 – protecting data privacy (34 percent)

- having accurate data (26 percent)
- and analyzing/processing that data (24 percent).

 This lack of structure may reflect the fact that 41 percent of business leaders say that no one in their organization is responsible for the management of data. This absence of ownership may also be why 52 percent of business leaders said they haven't had the budget to implement data management practices within their organizations.

 "Businesses must make data governance and stewardship a priority," said Monica Richter, chief data officer, Dun & Bradstreet, in a press release. "Whether leaders are exploring AI or predictive analytics, clean, defined data is key to the success of any program and essential for mitigating risk and growing the business."

The study does reveal a growing recognition that responsibility for data should be a priority for the C-suite. However, business leaders are divided as to who in the leadership team owns the data and what that will look like in the future. One thing all business leaders agree on is that the CEO has had, currently has and will have ultimate responsibility for data – more so than even the CTO or CIO.

Commenting on the findings, Anthony Scriffignano, Ph.D, chief data scientist at Dun & Bradstreet, said, "Information has always been critical for businesses, but over the past decade, the volume of data, the types of information available and the ability to do new things with that data have expanded enormously. It's not surprising that many business leaders feel they are still catching up and their organizations are yet to make the most of data – and some have even been fined or lost customers due to incomplete or 'dirty' data."

Monica Richter, chief data officer, Dun & Bradstreet

The survey of 510 UK and US business decision makers was conducted by Censuswide in March 2019. The businesses ranged in size from sole traders to those with over 500 employees and came from a wide range of industries, including finance, manufacturing, retail, marketing, and IT.

Opportunity Seen in Data Preparation

To prepare data for successful AI projects requires time and investment. The market for AI and ML data preparation solutions was over $500M in 2018, and is expected to grow to $1.2 billion by the end of 2023, according to a report from Cognilytica Research.

Anthony Scriffignano, Ph.D, chief data scientist at Dun & Bradstreet

The reports finds about 80 percent of AI project time is spent on aggregating, cleaning, labeling, and augmenting data to be used in ML models. Just 20 percent of AI project time is spent on algorithm development, model training and tuning, and ML operationalization.

Humans play a role in labeling AI data and in AI quality control. Mike Riegel, Chief Revenue Officer of CloudFactory, a data preparation firm, advised companies to deploy its most expensive human resources – data scientists and ML engineers – strategically.

[In response to a few related questions from AI Trends, CloudFactory's Director of Business Development Philip Tester sent the following:]

What is the role of humans in data preparation for AI projects?

The development of AI systems requires humans in the loop, and that won't change anytime soon. People will continue to play a critical role in data preparation for AI projects. People structure the datasets used to feed the machine learning (ML) algorithms that train AI systems, making it possible for machines to "learn" how to navigate a self-driving vehicle through city streets or review a legal contract, for example.

It takes trained people to clean, label and structure massive amounts of data quickly and accurately to deliver AI breakthroughs or scale a critical data process fast. At CloudFactory, our managed-team approach combines technology and people to provide a workforce in the cloud for data labeling. In our work with organizations over the last 10 years, we've learned that communication between data labelers and AI project teams is critical, so we provide a closed feedback loop between "CloudWorkers" and our client teams. This allows clients to incorporate iterations and improvements into their workflows, processes and tools over time.

Philip Tester, Director of Business Development at CloudFactory.
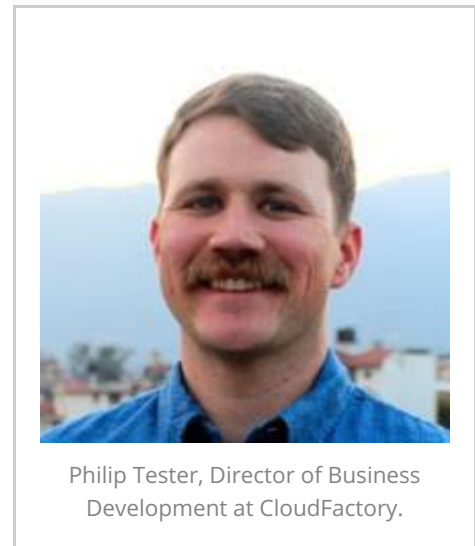
Is labeling of AI data to be used in machine learning models a manual process?

Data preparation and engineering tasks represent <u>over 80%</u> of the time consumed in most AI and ML projects. As businesses seek to apply AI to innovate customer experience and launch disruptive products, they need to label massive amounts of text, images and/or videos to create production-grade training data for ML models. This is a time-consuming, manual process, and many organizations aren't prepared for the management burden and cost that come with spinning up a data labeling team in-house. Skilled humans in the loop and data labeling tools that maximize labeling quality are necessary.

What does data preparation look like going forward – is it likely to be more automated?

Developing high-performance ML models requires a strategic combination of people, tools and processes. Smart leaders will assign to people the tasks that require domain expertise, context and adaptability and automate the tasks that require repetition, measurement and consistency. Our clients use open source and commercial tooling to automate the process

and break the work down into smaller tasks. Thinking strategically about the entire data production line will maximize data quality, optimize worker productivity and limit the need for costly re-work.

Read the source posts: report from Dun & Bradstreet, the Cognilytica Research, site for CloudFactory.