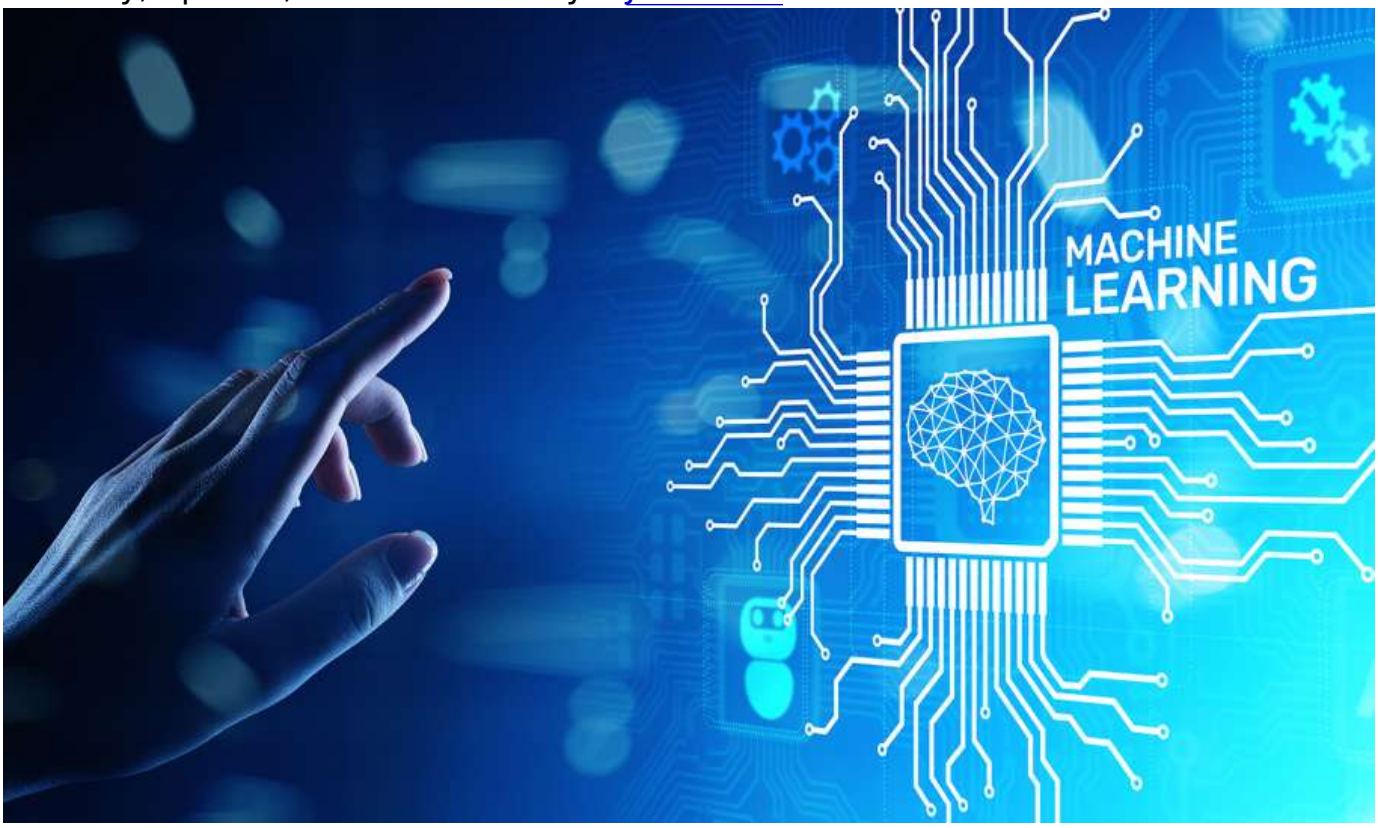
Is Your Sales Data Clean Enough for a Machine-Learning World?

Tuesday, April 21, 2020 - Written by: Jeff Kalter



Today’s companies run on data. We track trends, assess results and analyze feedback. Data helps us manage risk, plan for future growth and allocate resources.

Data is the foundation of our lead lists, sales pipelines and connections with customers. We collect contact records, integrate industry insight and profile companies based on their technographics (tech stack) and firmographics (company demographics).

At their core, sales metrics, key performance indicators (KPIs), sales projections and transactions are all data.

According to Salesforce’s “State of Sales” study:

- Today’s sales teams are twice as likely to prioritize leads based on data insights than gut instincts.
- Half of sales teams leverage data in their sales forecasting, and top performers are one-and-a-half times as likely to rely on data-driven forecasting.
- 81% of reps think it is vital to have an integrated view of data throughout the buyer’s journey.

With so much riding on data, we need to be able to rely on what it’s telling us. At the most basic level, we expect phone numbers to be accurate, job titles to be current and customer information to be complete.

But that’s no longer enough.

As businesses integrate sophisticated applications into their operations, we’re relying on automated systems running AI-enhanced algorithms. Our systems are using data to learn—tracking and analyzing metrics, looking for patterns and tapping into past results to predict future outcomes.

Clean, reliable data is more critical than ever. So it may be time to review your data management strategies to ensure your data is clean enough for this machine-learning world.

What is Machine Learning and Why Does It Matter?

In simplest terms, machine learning occurs when AI algorithms process and analyze large volumes of data to provide insight not previously known.

If you want to see machine learning in action, look no further than Amazon, Audible and Netflix. All are analyzing data in real-time—learning from your searches and purchases and comparing them with millions of other transactions to provide you with suggestions and recommendations based on your interests, needs and tastes. These “smart” systems are providing the same services to everyone on these websites simultaneously.

Automated systems that use machine learning can take over many mundane processes while providing customers with lightning-fast response rates for services that only a few years ago would have been impossible, such as 24/7 online

customer service.

Inside the sales department, machine learning is at the heart of “guided selling” strategies that score sales opportunities by likelihood to close, the value of the deal, projected lifetime value and more. It’s at the center of predictive sales forecasting, finding new market opportunities, documenting best practices that will improve sales reps’ productivity and developing optimal pricing strategies.

Machine learning replaces intuition with insight. But it all requires clean data.

Steps Companies Can Take to Deliver Clean Data

How clean does your data need to be? The simple answer is the cleaner, the better. As the saying goes, “Garbage in, garbage out.”

Whether or not you believe data is the new oil, there’s little doubt that clean data is your competitive edge. It can deliver better customer insights, more accurate forecasting and highly predictive lead scoring. Also, it can provide prescriptive sales processes that cut the time spent chasing leads and focus reps on serious prospects. Thus, clean data can allow your salespeople to spend more time building relationships with prospects and customers.

Traditionally, dirty data has referred to any data that is incomplete, rife with misspellings and missing fields, out of date, or includes many duplicates. But we need to expand our thinking to consider the data itself. For example, will it add to a machine’s knowledge base? Which data sets are necessary for machine learning? What do you do with outliers, for example, values that fall way outside the norm?

And then there is the question of how data is structured and managed. It’s not only critical for the best machine-learning results. It’s also about security and privacy—both within your company and to comply with GDPR and the California Consumer Privacy Act.

While your data engineers are likely responsible for the work to set up data management, governance and cleaning protocols, you must have a sense of the processes involved. Ultimately, the results will have a direct impact on your sales.

Here’s a brief list of what it takes to make data clean enough for a machine-learning world:

- **Data governance**: Before your data engineers can get down to specifics, your company needs to set the policies and standards for managing data assets internally—everything from collection and availability to integrity and security.
- **Preprocessing data**: To select the correct data for an algorithm, you need to understand the objective. Which data sets, when processed together, will provide appropriate insight or accurate forecasting? It's about having the right data and enough of it to be statistically sound.
- **Data quality assurance**: Your data scientists will probably need to undergo some exploratory data analysis to understand the data, measure its overall quality and develop the hypothesis at the heart of a systems algorithm. Does it work? Does the outcome enable the system to automate some of your sales processes effectively?
- **Structural integrity**: In addition to having a process for culling duplicate data and irrelevant observations, you need protocols for managing data. Do incomplete fields need to be filled in? Is data formatting consistent? How will you catch and correct typos and misspellings? Do you have a policy for dealing with outliers? Are those outliers the result of missing values, weak data sets or legitimate considerations?
- **Point of entry**: Consider the source of your data. If someone enters it manually, for example, you may need a set of protocols to ensure it is free of human error. If it's automated, you may need to reformat specific fields to match similar fields in other data sets.
- **Augmented data management**: This is the future of data management—using AI and machine learning to automate and standardize your data management practices. Not only can machine learning help automate elements of data engineering that have been manual in the past, but also augmented systems can help you get the maximum value out your data. For instance, these systems can help you to discover new connections between data sets and opportunities to draw new data insights.

The integrity of the data cleansing process and the development of machine-learning algorithms are critical to your business. The more you understand about the role of machine learning in assisting your sales and the data that can help

achieve accurate predictions and prescriptions, the better you can work with your data engineers to help boost your sales.

For help meeting your sales goals, call us at +1 813-320-0500 (US) or +39 06 978446 60 (EMEA), or contact us online.

Published in
List Development & Cleansing