

# Vertica In-Database Machine Learning

In today's data-driven world, creating a competitive advantage depends on your ability to transform massive volumes of data into meaningful insights. Companies that use advanced analytics and machine learning are twice as likely to be top quartile financial performers and three times more likely to execute effective decisions. Whether your organization is trying to reduce customer churn, predict mechanical equipment failures or bring entirely new products to market, you need an advanced analytics platform that can keep up with the new speed and scale of business.

## Vertica In-Database Machine Learning

Vertica in-database machine learning enables companies to incorporate predictive algorithms into their SQL analytics database for model building, scoring and evaluation. By building these functions directly into Vertica's core—with no need to download and install separate packages—companies can break down the barriers preventing predictive analytics projects from getting off the ground and into production, including:

- Legacy tools can't handle the scale of today's data volumes, forcing companies to down-sample and compromise model accuracy
- Multiple platforms are required for data storage, SQL analytics, data preparation, and statistical analysis, increasing the time and cost required to deploy predictive analytics projects
- Resource constraints and skills gaps prevent data analysts from having a hand in the creation and execute of predictive analytics models

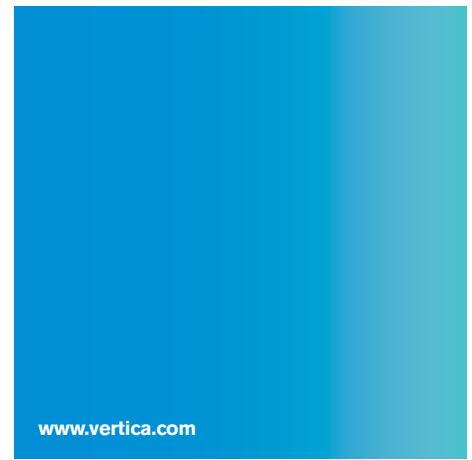
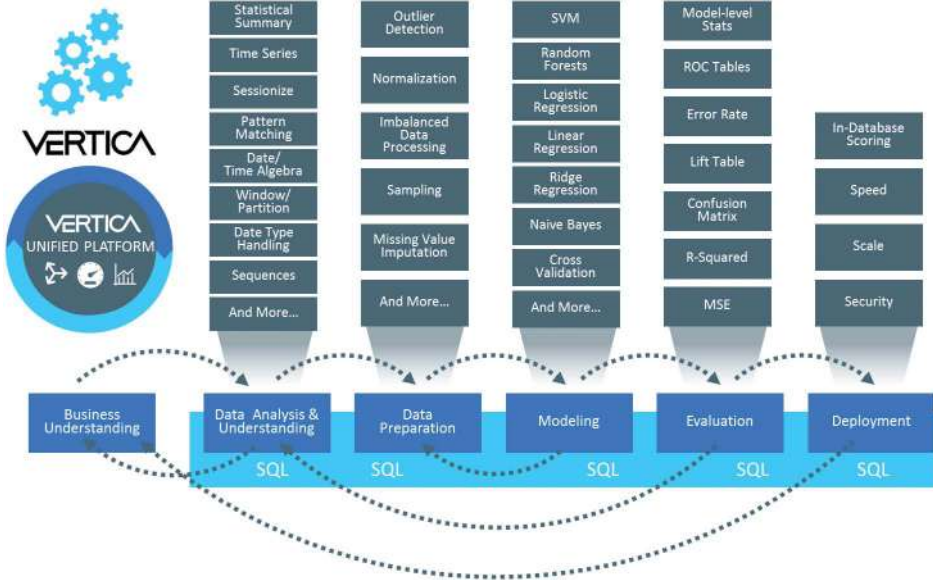
## Vertica Analytics Platform In-Database Machine Learning Functions

1. **Linear Regression**—use to predict continuous numerical outcomes in linear relationships along a continuum. Vertica supports Linear Regression by modeling the linear relationship between independent variables, or features, and a dependent variable, or outcome.
2. **Logistic Regression**—use to model the relationship between independent variables, or features, and some dependent variable, or outcome. The outcome of logistic regression is always a binary value.
3. **K-Means**—use to cluster data points into k different groups based on similarities between the data points. This unsupervised machine learning algorithm has a wide number of applications, including: Search engines, spam detection, cybersecurity
4. **Naïve Bayes**—use to classify your data when features can be assumed independent. The algorithm uses independent features to calculate the probability of a specific class. This supervised machine learning algorithm has a wide number

## Quick View

Vertica's in-database machine learning supports the entire predictive analytics process with massively parallel processing and a familiar SQL interface, allowing data scientists and analysts to embrace the power of Big Data and accelerate business outcomes with no limits and no compromises.

- **End-to-end Machine Learning Management**—Prepare data with functions for normalization, outlier detection, sampling, and more—then create, train and score machine learning models on massive data sets.
- **Massively Parallel Processing (MPP) Architecture**—Build and deploy models at Petabyte-scale with extreme speed and performance on a unified advanced analytics platform
- **Simple SQL Execution**—Manage and deploy machine learning models using simple SQL-based functions to empower data analysts and democratize predictive analytics
- **Familiar Programming Languages**—Create and deploy C++, Java, Python or R libraries directly in Vertica with user-defined functions



### Vertica Headquarters

150 Cambridgepark Drive  
Cambridge, MA 02140

Learn more at: [www.vertica.com](http://www.vertica.com)

of applications, including: spam filtering, classifying documents, and image classification

5. **Support Vector Machines**—use to predict continuous ordered variables based on the training data. This supervised learning method has a number of applications, including: predicting time series, pattern recognition, and function estimation
6. **Random Forest**—use to create an ensemble model of decision trees. Each tree is trained on a randomly selected subset of the training data. This supervised learning method has a number of applications, including: prediction genetic outcomes, financial analysis, and medical diagnosis

**“Vertica’s new in-database machine learning capabilities are like gold! We are extremely excited to train our Machine Learning models on our data in Vertica and ship them with our platform to run on our customers’ clusters. This is something that is much harder with any other tool. Vertica’s in-database machine learning will improve our ability to offer new predictive analytics features quickly and easily to our growing customer base. It will improve our competitive positioning!”**

**ABHISHEK SHARMA**  
Data Scientist  
Fidelis Cybersecurity

### Industry Use Cases

- **Financial services organizations** can discover fraud, detect investment opportunities, identify clients with high-risk profiles, or determine the probability of an applicant defaulting on a loan.
- **Government agencies**, such as public safety and utilities, can use machine learning to help detect fraud, minimize identity theft, and analyze data from smart meters to identify ways to increase efficiency and save money.
- **Communication service providers** can leverage a variety of network probe and sensor data to analyze network performance, predict capacity constraints, and ensure quality service delivery to end customers.
- **Marketing and sales organizations** can use machine learning to analyze buying patterns, segment customers, personalize the shopping experience, and implement targeted marketing campaigns.
- **Advertising Technology companies** can optimize audience targeting, analyze visitor behavior through A/B and multivariate testing, and predict user engagement patterns.
- **Manufacturers** can identify product defects, predict equipment maintenance needs, optimize supply chain planning and forecast demand.

### Learn More At

- [Product Page](#)
- [White Paper](#)
- [Documentation](#)