Machine learning methods are highly dependent on the quality of the data they receive as input, but data preparation and cleaning can be an unwieldy task, taking up roughly 60% of the time of data scientists and analytics professionals.

To address this need, SparkCognition™ has developed the Darwin™ platform, an automated machine learning product that accelerates data science at scale, enabling you to assess the quality of your dataset and advising you on how to fix problems to make it suitable for the model-building process. Darwin then automates time-consuming tasks that range from model creation and optimization to model deployment and continuous maintenance.

Assessing the Overall Quality of Your Data

Darwin scores your data on its usefulness for the data science process.

- *Columns that can be directly used are marked in green*
- *Columns that will require pre-processing are marked in yellow*
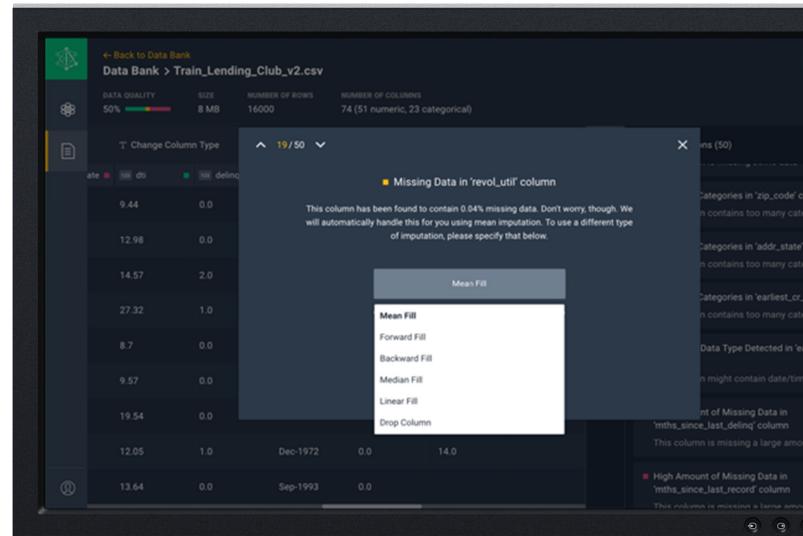- *Columns that will be dropped are marked in red*
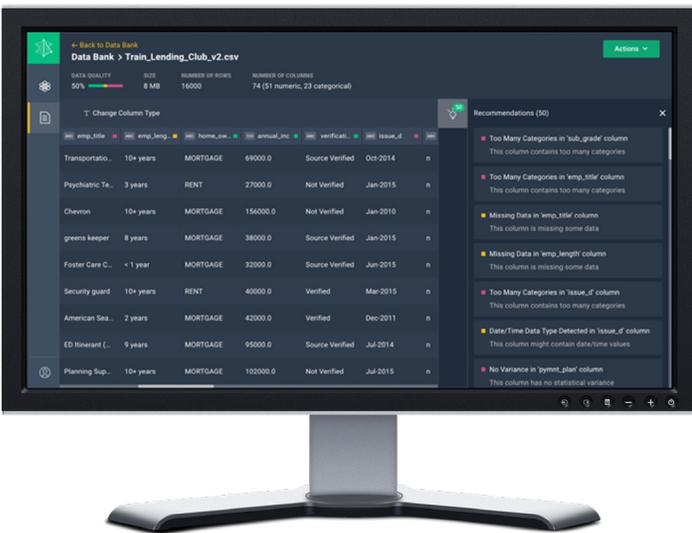
Getting Your Data Ready for Machine Learning

Darwin's guided data preparation workflow discovers and recommends solutions for potential problems in your data set, such as:

- *Missing data*
- *Columns with low variance*
- *Columns with too many categorie*

Quality Data = Quality Models

Next, Darwin uses its patented blend of evolutionary algorithms and deep learning methods. This method specializes in discovering novel, elegant network architectures, while also supporting hyperparameter search for common algorithms such as Random Forest and XGBoost. Darwin automates the following steps:

- *Execution of the data cleaning profile*
- *Feature generation to enrich the dataset*
- *Construction of a supervised or unsupervised model*

Darwin uses an iterative genetic process to scratch build model topologies that are optimized with each passing generation, creating unique solutions that correctly and accurately reflect your data and output higher-quality predictions.

Quality Models = Faster Operationalization

Darwin's automated workflows around data quality and model creation allow a faster turnaround of use cases, enabling organizations to operationalize the output of data science and innovation teams faster. This approach effectively transforms organizations into factories of use cases that efficiently operate on their data to positively impact what matters: the bottom line.

How Can You Experiment with Darwin?

Take the next step in your machine learning journey with Darwin's automated model building approach.

Evaluate Darwin today at:
 https://www.sparkcognition.com/darwin-trial/