



Predictive Maintenance Case Study

Predict and anticipate maintenance tasks to
prevent failures and increase machines uptime



This case study is an example of a company that uses many machines to build their final products.

As the supply chain is stopped every time a machine breaks, the manager asked a consulting firm to build **a predictive model** that finds which **machine is going to break next and why.**

As we explored the data (we got the dataset from Kaggle), we understood that the company is using 1000 machines. A machine has an average lifetime of 55 weeks, with some brand new machines and others that are running since almost two years. In our dataset almost **40 % of the machines broke in the past two years.**

For each machine which corresponds to a row in our dataset (1000 rows), we have 6 variables: “lifetime” indicates number of weeks since the machine has been used, “broken” which is our target variable (Yes or No), then we have 3 numeric variables related to temperature, pressure and moisture, and 2 variables related to the team using the machine and the machine’s provider.

We used MyDataModels to build a predictive model which will allow the manager to predict which machine will break and why, goal being to anticipate the maintenance tasks to prevent that machine to actually break and stop the supply chain.

Preparation of the dataset file

The response variable Broken (Yes or No) value was derived from the metadata. Rows were then randomized such that the order of samples in the rows was ensured to be random.

Results in less than 10 minutes

Generation of an explanatory model, took just under 10 minutes. When evaluated against an independent hold out set of 300 samples (actually the test subset derived from the 1000 samples dataset, **see Fig 2** comments), the resulting explanatory model was 85% accurate with a True Negative Rate of 84%, and a True Positive Rate of 87%.

The True Negative Rate represents the percentage of samples predicted Negative (not Broken in our case) and being actually Negative in the test subset. True Positive Rate represents the percentage of samples predicted Positive (Broken in our case) and being actually Positive in the test subset.

In this business case, it is absolutely critical for the manager to get a 0% FNR (False Negative Rate) to avoid unexpected machine stops, we have used a unique and innovative MyDataModels feature which allows to adapt resulting models to specific business cases, and in this case to get a 0% FNR. **See figure 4.**

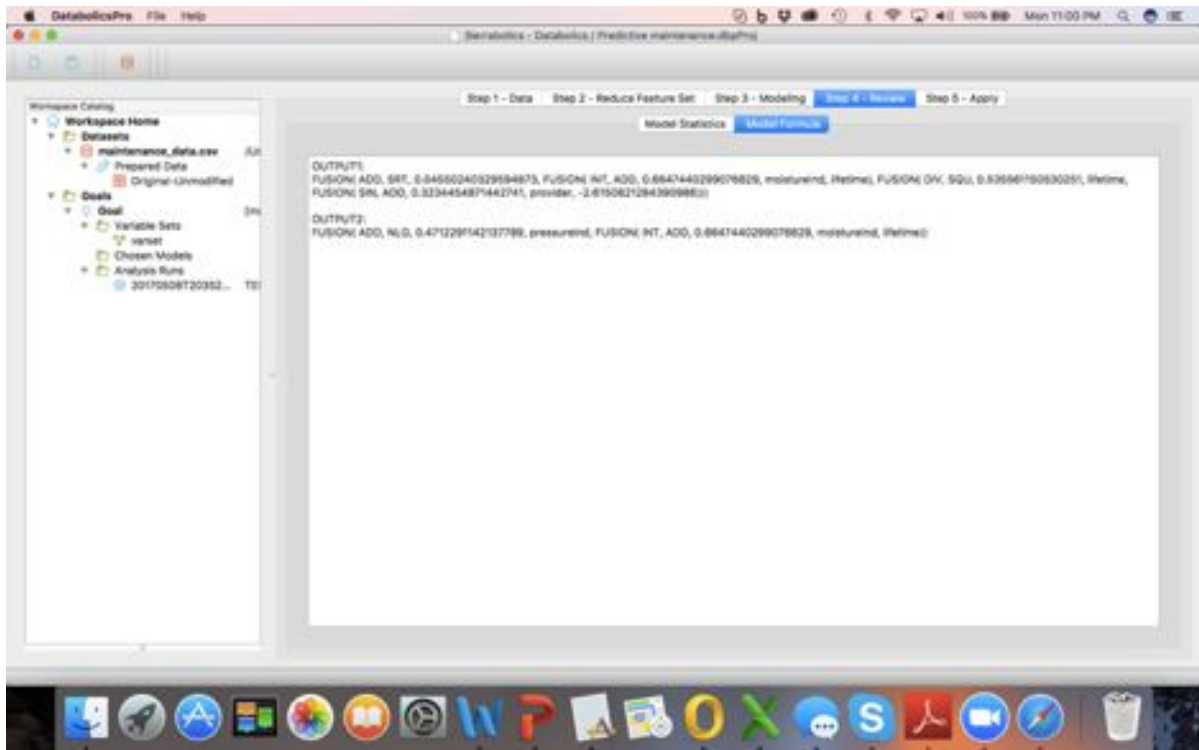


Figure 1: resulting model (Model ID: 20170508T203529.718336). It uses 4 variables (lifetime, provider, moisture and pressure), several operators and constants. In a binary classification model, model is made of 2 equations, one calculating probability of true prediction, the other calculating probability of false prediction. Upon computing the value of each equation, the greater value yields the case predicted by the model. The difference between these 2 numbers gives the prediction's confidence score.

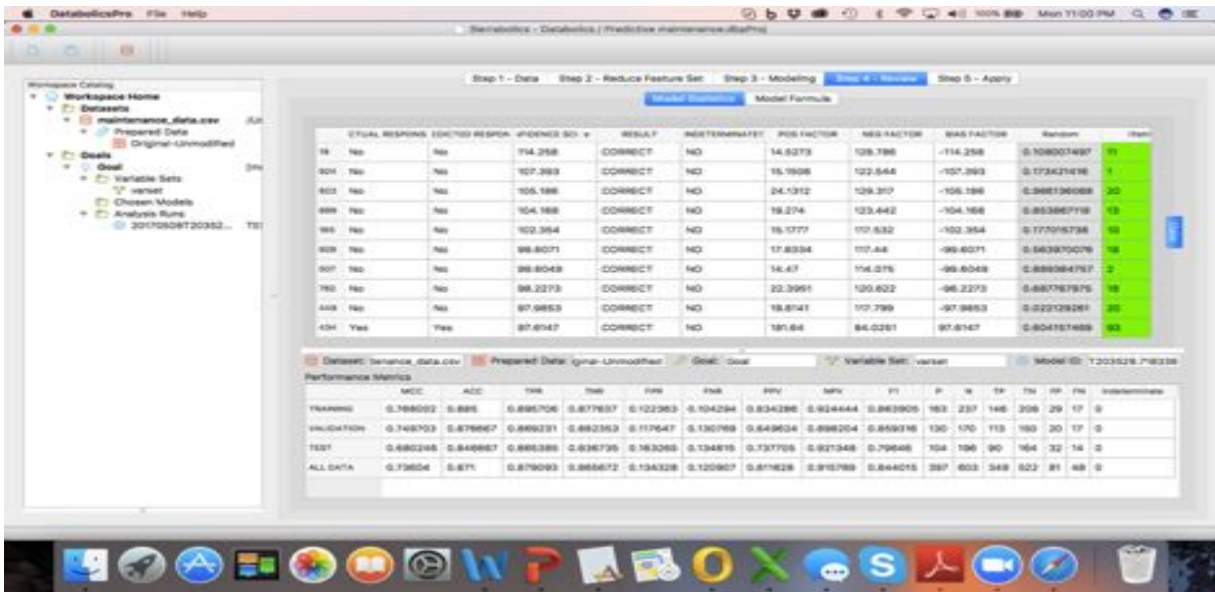


Figure 2: Review of best model statistics. One can see the 87% accuracy obtained on the entire dataset and on the 3 subsets (training, validation and test, training and validation are used for model creation, the test set is an independent hold out used to provide a final accuracy assessment by the final model against new data not available to the modeling process) used to build hundreds of potential models and finally select the best model.

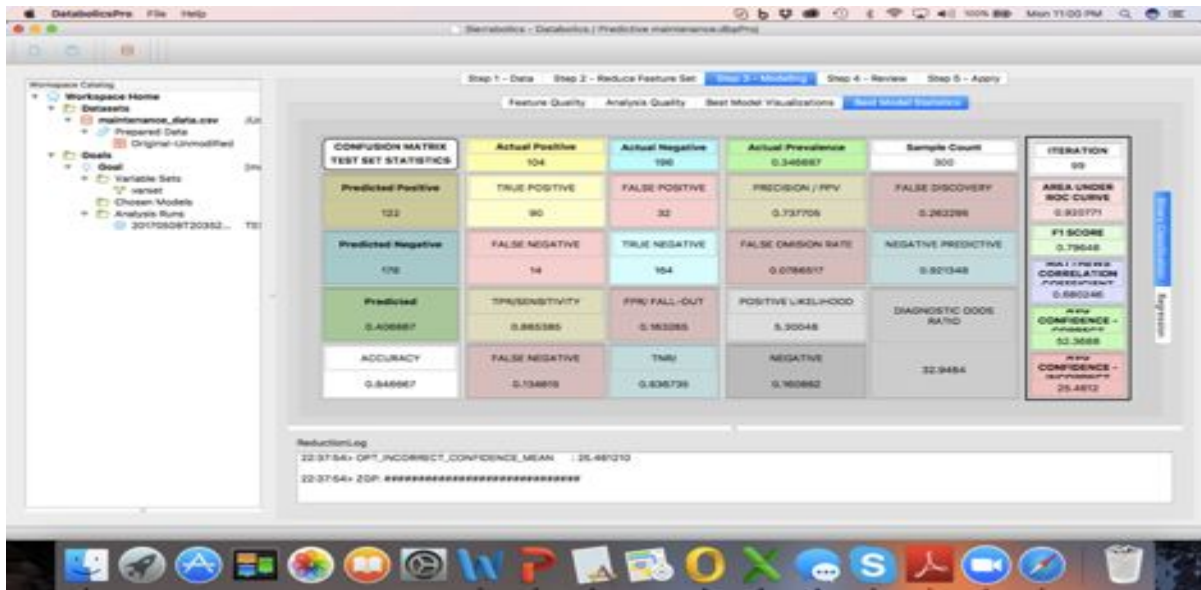


Figure 3: When evaluated against an independent hold out set of 300 samples, the resulting explanatory model was 85% accurate with a true negative rate of 84%, and a true positive rate of 87%.

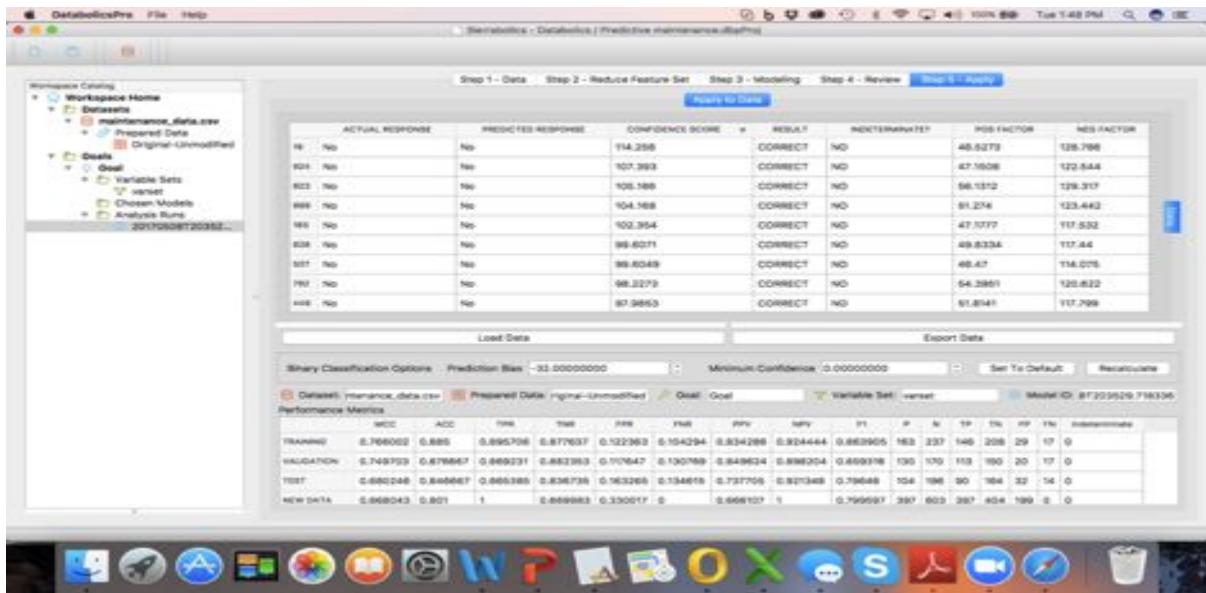
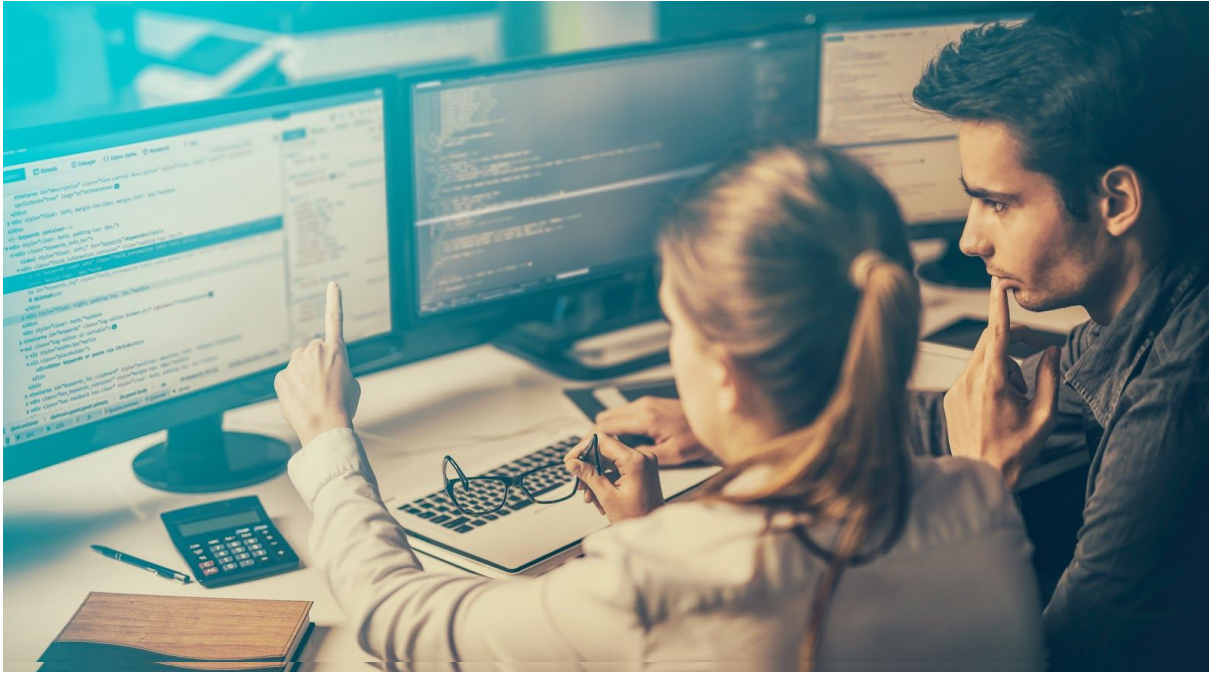


Figure 4: You can see the prediction bias set to -32 and the FNR obtained is 0% on the New Data dataset instead of 13% on the Test subset. We decided to set the prediction bias to -32 because the occurrence of the False Negative sample with the highest Confidence Score had a bias equal to -31.8. By setting the bias to -32, we were able to remove all subsequent False Negative samples with lower Confidence Scores. The trade off was that accuracy was 80% instead of 85% but the model was more in line with business case objectives of the manager.



MYDATAMODELS

The Automated Machine Learning Company

MyDataModels is a software company that develops and markets an Automated Machine Learning 2.0 Platform.

Automated Machine Learning 1.0 opened the path few years ago. It automates most of the predictive modeling process, but still requires some coding and Machine Learning skills.

MyDataModels goes much further. Our Automated Machine Learning 2.0 service is powered by our unique Machine Learning engine, inspired by evolutionary algorithms. It is a one click data-in model-out service which enables all professionals to build predictive models without coding or Machine Learning skills.

