Developer Network

Developing Analytics Solutions with the Data & AI Global Practice

Can a machine learning product be trusted?

★★★★★

November 13, 2018 by Michele Usuelli // 0 Comments

Author: Michele Usuelli, Lead Data Scientist, Microsoft Enterprise Services

The data scientist develops products based on machine learning models, having as a result an actionable answer. In other words, a machine is advising an end user about some actions to perform. Why would we trust the advice of a machine?

Before designing a machine learning product, it's crucial to keep in mind its application.

This article covers the key aspects to take into account to have a reliable product. Specifically, we will look into predictive models, aiming to foresee future events.

To be more concrete, we will be utilizing three real-life examples

- Predictive maintenance: the engines of one or more machines break down frequently. The predictive model foresees whether there will be an engine failure in the near future.
- Credit scoring: a bank provides its customers with loans. Having a new customer requesting a loan, the bank estimates the likelihood of having a default.
- Sales forecast: a retailer is interested in predicting the sales volume of one or more items.

Collecting lessons learned from data science projects, common key aspects are the following.

- Interpretability: capability of explaining the rationale behind a machine learning model.
- Measurability: assessment of the performance and accuracy of the model.
- Actionability: influence on the end user to perform a specific action.

Each aspect is covered by the related section.

Interpretability

Predictive models foresee the breakdown of an engine, the credit score of a customer or the sales volume of a new product. In either case, how do we know if we can trust the outcome of a model? Also, can we check if it complies to rules and abides by ethical principles?

To enable that, the first step is ensuring that the model is explicable. Let's look into our examples We might predict that an engine will break down because it's running slower than usual. A bank customer will likely default because of a high credit card debt. The sales volume will raise because of a commercial promotion. These example are close to reality, but often the set of useful information is wider. Still aiming to explain the rationale behind its outcome, a machine learning model combines and utilizes more useful information.

Machine learning models are based on data sources, so tracking all the information being used is crucial. In this way, we can easily spot ethical issues and legislation incompliances. The next step is to identify which specific information is being used by the model and what is being discarded. Ultimately, applying the predictive model on new data, the information being used is a small highly-relevant subset.

Depending on the technique, the methodology might be different. It's usually straightforward to explain tree-based models such as the decision forest, as shown by my article. Likewise, linear models can be easily interpreted. However, some techniques, e.g. neural networks, are too complex and obscure. Even in those cases there are workarounds such as LIME and SHAP, showing what would have changed if the data differed.

Measurability

Let's assume that a predictive model is already explainable. The thought process to define its outcome is fully transparent and we can track all the information being used. What will happen applying the model in a real-life situation? Will it be able to foresee future events accurately?

If we predict engine breakdowns, we want to know how many of them we can predict and the number of false positives. The same applies to credit risk. If we are predicting the sales volume, a metric could be the average error, whichever way the "average" is defined.

Being able to assess the reliability and accuracy of the outcome of a model is a key to be trustworthy. Clearly, it's impossible to know how a model will perform in the future. However, it's possible to know how the same methodology would have performed in the past.

Predicting something that is already known is trivial. Therefore, a common pitfall it to "cheat", either voluntarily or involuntarily. To prevent that, the key is to replicate the real life application. Some important aspects are the following.

- Separate training and test set entirely.
- Take into account the time: in most cases, we are predicting and event in time. A common pitfall is to predict the past utilizing future information.

In addition, the outcome of the testing should reflect the broader context. Key aspects are the following.

- Business focus: in addition to measuring the accuracy, are we also translating it into measurable business KPIs?
- Benchmark: if we weren't utilizing a predictive model, by how much would the forecast be worse?

My colleague Ajay published a series of two articles, showing how to measure the performance of classification and regression models.

Actionability

Last but not least, even if we can explain and measure the outcome of a machine learning product, to have a tangible impact it will trigger an action.

In the predictive maintenance scenario, the action is to either send an engineer to check the machine or to provide confidence that the machine doesn't need any maintenance. In the loan use-case, the action is to either approve or reject the loan. Predicting sales helps optimizing the production volume.

Prior to start experimenting and building the product, having in mind the end action helps defining the direction of the project. After having delivered the product, keeping track of the actions and feedback allows to improve the product further.

Actionability applies to both interpretability and measurability.

- Interpretability: knowing the root causes allows to define an action to fix them. Not all the root causes are necessarily actionable. Knowing what can be changed is an enabler to define an action. For instance, in the predictive maintenance scenario, the action is to fix the root cause leading to a failure. If the root cause is having the last maintenance too far back in time, the solution is to send an engineer to maintain the machine.
- Measurability: the estimated error is as important as the prediction. Knowing the actions that can be taken enables the definition of an accuracy metric accordingly. In the sales forecast use-case, the accuracy metric might be related to the way the forecasting will be consumed.

Actionability is the starting point and the end point of most of the data science projects.

Conclusions

This article covered interpretability, measurability and actionability, key aspects of each data science project. One remark is about other important topics, such as ethics, privacy and compliance, that will be enabled by these.

analysis Center of Excellence Centre of Excellence CoE Data Data Science Data Scientist Role Decision forests Event Hubs k-fold Machine Learning Measurability Model Goodness Modelling Power BI PowerShell ML Scoring R Random Projection R Services Scikit SQL Server 2016 SQL Server R Services Stream Analytics Visualizations Whitepaper

Archives

November 2018 (1)
All of 2018 (4)
All of 2017 (3)
All of 2016 (11)
All of 2015 (11)

Join the conversation

Add Comment