

Expressing the value of data science in an ROI framework

September 19 2016

Data science is rapidly becoming woven into the fabric of organizations of all sizes and types, and is driving significant societal and economic impact. Organizations are increasingly becoming data driven, investing in infrastructure, people and processes to embrace the data science journey.

In a recent paper published in *EPJ Data Science*, University of Notre Dame researchers study how organizations can quantify decision making in data science. Doctoral student Saurabh Nagrecha and his adviser, Nitesh Chawla, the Frank M. Freimann Professor of Computer Science and Engineering and director of iCeNSA, advocate that data science is a process and present a solution to quantifying the value of data acquisition and modeling in a return on investment (ROI) framework.

"An ROI-based valuation means that organizations can budget for existing strategies better, readily compare vastly different data strategies, and in a budget-constrained environment even answer the tough questions like 'To achieve the desired outcomes, should I invest money in more data acquisition or more complex modeling or both?'" Chawla said. "We have developed an ROI-based modeling framework, called NPV model in this paper, that can begin to answer such questions."

The NPV model enables users to translate a [machine learning](#)-based [predictive model](#)'s performance over time from traditional empirical measures into dollar values by combining machine learning, [data acquisition](#), operational costs, and investment parameters.

"Typically, success for machine learning models is expressed in accuracy, precision, recall, ROC and other such metrics," Chawla said. "Facets of costs should be incorporated in evaluation, when available, as false negatives might be more costly than false positives, for example. Our paper expands this cost-sensitive classification framework by incorporating costs to acquire external data, modeling costs and operational costs, all of which are essential for the real-world deployment of these machine learning models. Moreover, these predictions don't just happen at once, but instead occur over a timeline—where it is important to consider the time-based valuations under constraints."

Chawla pointed out that a data-driven organization may make predictions on millions of instances of streaming data every day using an in-house predictive model. They have an idea of the cost of a correct prediction, a false positive, a false negative, operational costs, cost of capital for the team, etc. Using the NPV model, they can now ascribe a value to their entire data [science](#) operation and strategize for the future.

"If organizations want to investigate the possibility of tying in external data into their operations, they can use our technique, run it on their current data alongside their in-house data, and get the value of the new model," Nagrecha said. "If this new value, minus the switchover costs, is greater than that of the current model, then it means that over time, it is worth getting external data. Using the same process, they can evaluate competing bids for external data, multiple machine learning techniques, etc., on the same strategy board—all on the basis of their respective NPVs, and select the best ones given expected outcomes and budget."

The team's approach is generally applicable to all organizations as they face the decision of becoming increasingly more data-driven and yet constrained for resources. This paper provides a strategy board for organizations to develop a budget and allocate resources on various

activities along the [data science](#) process. It starts with answering a basic question, "How valuable is the external data that I can acquire today to my future operations?"

More information: Saurabh Nagrecha et al. Quantifying decision making for data science: from data acquisition to modeling, *EPJ Data Science* (2016). [DOI: 10.1140/epjds/s13688-016-0089-x](https://doi.org/10.1140/epjds/s13688-016-0089-x)

Provided by University of Notre Dame

Citation: Expressing the value of data science in an ROI framework (2016, September 19) retrieved 25 April 2024 from <https://phys.org/news/2016-09-science-roi-framework.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.