

The Wisdom of Crowds: Best Practices for Data Prep & Machine Learning derived from Millions of Data Science Workflows

Ingo Mierswa
RapidMiner
Boston, MA
imierswa@rapidminer.com

ABSTRACT

With hundreds of thousands of users, RapidMiner is the most frequently used visual workflow platform for machine learning. It covers the full spectrum of analytics from data preparation to machine learning and model validation. In this presentation, I will take you on a tour of machine learning which spans the last 15 years of research and industry applications and share key insights with you about how data scientists perform their daily analysis tasks. These patterns are extracted from mining millions of analytical workflows that have been created with RapidMiner over the past years. This talk will address important questions around the data mining process such as: What are the most frequently used solutions for typical data quality problems? How often are analysts using decision trees or neural networks? And does this behavior change over time or depend on the users experience level?

Keywords

Visual Workflow; Machine Learning Tools; Analytics; Data Visualization; Wisdom of the Crowds.

Bio

Dr. Ingo Mierswa is the founder and CTO of RapidMiner and as such responsible for the strategic innovation of the open source machine learning platform. Ingo is an industry-veteran data scientist since starting to develop RapidMiner at the Artificial Intelligence Division of the TU Dortmund University in Germany. Ingo, the scientist, has authored numerous award-winning publications about predictive analytics and big data. Ingo, the entrepreneur, has been leading the company with growth rates of up to 300 percent per year over the first seven years. In 2012, he spearheaded the go-international strategy with the opening of offices in the U.S. as well as the UK and Hungary. After two rounds of fundraising, the acquisition of Radoop, and supporting the positioning of RapidMiner with leading analyst firms like Gartner and Forrester, Ingo takes a lot of pride in bringing the world's best team to RapidMiner.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author. Copyright is held by the owner/author(s).

KDD '16, August 13-17, 2016, San Francisco, CA, USA

ACM 978-1-4503-4232-2/16/08.

<http://dx.doi.org/10.1145/2939672.2945365>