

Simon Fischer, Ingo Mierswa (Eds.)

**Proceedings of the  
5th RapidMiner World (2014)**

Shaker Verlag  
Aachen 2014



# Preface

In its fifth year, a lot is new at the RapidMiner Community Meeting and Conference. First and most obviously, it comes with a new name, RapidMiner World, but it is not only the name that has changed! Also, for the first time, RapidMiner World takes place in the US, and it is our pleasure to welcome you to four exciting days from August 18–21, 2014 in Boston where RapidMiner’s new headquarters are situated.

Again, we have selected many excellent technical papers for presentation at the conference, introducing a variety of new extensions and techniques. But in addition to those, we will also have dedicated sessions focusing on exciting use case studies and applications of RapidMiner in industry projects. Furthermore, we are looking forward to the Industry Analyst Session led by Shawn Rogers of Enterprise Management Associates and the panel discussion on Tuesday which will provide insights into the latest research and opinions about the analytics market.

But not all is new. Of course we are keeping the tried and tested ingredients of a vibrant RapidMiner user meeting: You can win fantastic prizes in two challenges, the “Boston Tea Party Challenge” hackathon as well as our classic, the live data mining game show “Who Wants to be a Data Miner?”, where you can compete by using RapidMiner to solve problems it was never supposed to solve.

Training courses surround the main programme: the RapidMiner 101 on Monday for those who are not yet familiar with RapidMiner and the RapidMiner 201 on Thursday. Again, those who are interested to become a certified RapidMiner analyst can take the certification exam at the end of the meeting.

Also, we are glad to be able to present the latest news from the RapidMiner product suite in various formats, including a product roadmap presentation by Giuseppe Taibi, a demonstration of the latest and upcoming features, a Radoop Big Data workshop by Zoltan Prekopcsák who recently joined the team, and a partner showcase by Sheamus McGovern and David Weisman.

Last but not least, we maintain the tradition of exceptional social events, starting with the Harpoon Brewery Tour on Monday and continuing with an event you will never forget, the Boston Duck Tour, and dinner on Tuesday.

We hope that you enjoy the programme and the presentations and utilize the time between the talks for exchanging ideas and establishing new collaborations.

July 2014

Simon Fischer  
Senior Vice President Engineering  
RapidMiner GmbH



# Contents

<b>Part I – New RapidMiner Extensions</b>	<b>1</b>
<b>1 Data Mining with Background Knowledge from the Web</b> <i>Heiko Paulheim, Petar Ristoski, Evgeny Mitichkin, Christian Bizer .</i>	<b>1</b>
<b>2 An User Interface for Big Data with RapidMiner</b> <i>Marcelo Beckmann, Nelson F. F. Ebecken, Beatriz Pires de Lima, Myrian A. Costa . . . . .</i>	<b>15</b>
<b>3 MetahAtt: Metaheuristic Based Optimization Extension</b> <i>Sandro Radovanović, Milan Vukićević, Elena Milovanović, Milena Popović . . . . .</i>	<b>33</b>
<b>4 A Basic Network Analytic Package for RapidMiner</b> <i>Balázs Kósa, Márton Balassi, Péter Englert, Gábor Rácz, Zoltán Pusztai, Attila Kiss . . . . .</i>	<b>47</b>
<b>5 Defining and Executing Process Mining Workflows with Rapid-ProM</b> <i>Ronny Mans, Wil van der Aalst, Eric Verbeek . . . . .</i>	<b>61</b>
<b>Part II – Text Mining</b>	<b>73</b>
<b>6 Semi-Supervised Text Classification: A New Extension for RapidMiner</b> <i>Purvi Rekh, Amit Thakkar . . . . .</i>	<b>73</b>
<b>7 Similarity Assessment and Resume Analysis Using Clustering and Cosine Similarity Measures in RapidMiner</b> <i>Surbhi Lodha, Santosh Vishwakarma, Kamaljit Lakhtaria . . . . .</i>	<b>89</b>
<b>8 Sentiment Analysis on Arabic Tweets Using RapidMiner</b> <i>Salha al Osaimi, Khan Muhammad Badruddin . . . . .</i>	<b>103</b>
<b>Part III – Use Cases</b>	<b>113</b>
<b>9 Design and Implementation of Data Mining Case Studies in RapidMiner</b> <i>Matthew North . . . . .</i>	<b>113</b>

<b>10 Massive Data Analysis on an Aircraft Fleet Reliability Process</b>	
<i>Thiago Donato, André Noronha . . . . .</i>	<b>123</b>
<b>11 Proposal of Assembly Work Plans to Accelerate High-Variance Product Emergence Processes</b>	
<i>Ralf Klinkenberg, Regina Wallis, Olga Erohin, Jochen Deuse, Frank Stromberger, Simon Fischer . . . . .</i>	<b>133</b>
<b>12 Wanted: Big Data for Predicting Chemical Behavior</b>	
<i>Izacar Martinez . . . . .</i>	<b>145</b>
<b>13 Analysis of How Meteorological Events Are Impacting Microwave Radio Communications Networks</b>	
<i>Daryl O' Toole . . . . .</i>	<b>159</b>