Proceedings of the PMML Workshop 2011

Held in conjunction with the
KDD-2011 Conference
August 21, 2011

Workshop Organizers
Rick Pechter (MicroStrategy), Chair
Robert Grossman (Open Data Partners & University of Chicago)
Christoph Lingenfelder (IBM)
Ashok Savasere (SAS)
Michael Zeller (Zementis)

Sponsored by
Edited by

Rick Pechter
MicroStrategy

Robert Grossman
Open Data Partners & University of Chicago

and

Michael Zeller
Zementis
# Table of Contents

- Preface .......................................................................................................................... 4
- Workshop Agenda ........................................................................................................ 5
- PMML Conformance Progress Report – Five Years Later
  by Rick Pechter ........................................................................................................... 6
- Scorecard Element in PMML 4.1 Provides Rich, Accurate Exchange of Predictive Models for Improved Business Decisions
  by Andrew Flint and Alex Guazzell ............................................................................ 16
- Massively Parallel In-Database Predictions using PMML
  by Konstantinos Stathatos, Kaushik K. Das, Eugene Fratkin, Aleksander Gorajek and Maulin Gajjar ........................................................................................................ 22
- Comprehensive PMML Preprocessing in KNIME
  by Dominik Morent, Konstantinos Stathatos, Wen-Ching Lin and Michael R. Berthold ........................................................................................................ 28
- The PMML Path towards True Interoperability in Data Mining
  by Alex Guazzelli, Tridivesh Jena, Wen-Ching Lin and Michael Zeller ...................... 32
- Enhancing A Production Rule Engine With Predictive Models Using PMML
  by D. Sottara, P. Mello, C. Sartori and E. Fry .............................................................. 39
- An Extension of the PMML Standard to Subspace Clustering Models
  by Stephan Günnemann, Hardy Kremer and Thomas Seidl ..................................... 45
- Background Knowledge and PMML - First Considerations
  by Tomas Kliegr, Stanislav Vojir and Jan Rauch ...................................................... 54
Preface

Over the past decade, the Predictive Model Mark-up Language (PMML) has emerged as the *de facto* standard for predictive model specification that is vendor neutral, application agnostic and a true enabler for cross-platform interoperability. As PMML has matured, a broader variety of applications and deployment mechanisms are being used. These create opportunities for future growth but they also create challenges for the standard to remain flexible enough to meet the demands of new users and new solutions. Related fields such as data grids, web services, and the semantic web have also developed standards-based infrastructures and services relevant to KDD. These standards and standards-based services and platforms have the potential to change the way data mining is being used.

Based on the structure of similar standards workshops that were part of past KDD conferences, this workshop aims to bring together practitioners from industry and academia to discuss various applications and foster a dialog about data mining models, model interoperability, data transformation and post-processing workflows and related topics. In addition, the workshop will include talks on future requirements and the future direction for data mining services and platforms. As part of the KDD-2011 conference, we hope to expose the broader community to this area, and conversely, to collect requirements for new versions of the PMML standard. Included in this year’s workshop are eight papers that cover a variety of topics related to PMML, not only how it is used, but also how it can be extended while remaining the *lingua franca* for predictive models. The workshop will also include a panel discussion on PMML where participants address questions and debate future directions for the standard.

As predictive analytics and data mining continue to gain momentum in industry and business applications, open standards like PMML provide an essential foundation to share predictive models across vendors and to disseminate model-related information within organizations. It is only through the contributions of many people from industry and academia that PMML standard continues to grow in adoption and evolve in capability. The workshop organizers would like to thank all those who give generously of their skills and expertise in order to make PMML the standard that it is today.

Robert Grossman  
Rick Pechter  
Michael Zeller  

June 24, 2011
<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Presenter(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:00 to 8:30</td>
<td>Welcome &amp; Introductions</td>
<td>Rick Pechter</td>
</tr>
<tr>
<td>8:30 to 8:45</td>
<td>PMML Conformance Progress Report – Five Years Later</td>
<td>Rick Pechter</td>
</tr>
<tr>
<td>8:45 to 9:00</td>
<td>Scorecard Element in PMML 4.1 Provides Rich, Accurate Exchange of Predictive Models for Improved Business Decisions</td>
<td>Andrew Flint, Alex Guazzelli</td>
</tr>
<tr>
<td>9:00 to 9:15</td>
<td>Massively Parallel In-Database Predictions using PMML</td>
<td>Konstantinos Stathatos, et al</td>
</tr>
<tr>
<td>9:15 to 9:30</td>
<td>Comprehensive PMML Preprocessing in KNIME</td>
<td>Dominik Morent, et al</td>
</tr>
<tr>
<td>9:30 to 9:45</td>
<td>The PMML Path towards True Interoperability in Data Mining</td>
<td>Alex Guazzelli, et al</td>
</tr>
<tr>
<td>9:45 to 10:00</td>
<td>Enhancing A Production Rule Engine With Predictive Models Using PMML</td>
<td>D. Sottara, et al</td>
</tr>
<tr>
<td>10:00 to 10:30</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>10:30 to 10:45</td>
<td>An Extension of the PMML Standard to Subspace Clustering Models</td>
<td>Stephan Günneemann, et al</td>
</tr>
<tr>
<td>10:45 to 11:00</td>
<td>Background Knowledge and PMML - First Considerations</td>
<td>Tomas Kliegr, et al</td>
</tr>
<tr>
<td>11:00 to 11:45</td>
<td>Panel Discussion</td>
<td>Organizers</td>
</tr>
<tr>
<td>11:45 to 12:00</td>
<td>Wrap-up and Conclusions</td>
<td>Rick Pechter</td>
</tr>
</tbody>
</table>