Social Media, Recommendation Engines and Real-Time Model Execution:

A Practical Case Study

Phil Winters Phil.Winters@KNIME.COM
Michael Zeller Michael.Zeller@Zementis.com

Table of Contents

Introduction: Hype vs. Quantifiable Reality

There is a lot of theory and hype around the topics of social media, recommendation engines and real time modeling, but until now not many practical examples that can be measured in terms of ROI. KNIME AG and Zementis have joined together to provide this white paper, which summarizes a practical case study that combines all three topics, and delivers a measured and solid business case.

Social media is of course very hot even if no longer a "new" topic. In our context, social media provides not only a mechanism for individuals to communicate and share over common touchpoints of their choice but also to collect that information so that it can be used to provide a better experience.

Recommendation engines are not new – they take forms from market basket analysis in retail to advanced analytic systems providing next best offer or next best activity suggestions. They are also very popular for making suggestions – or recommendations – of similar items when a particular product or offering is selected. Amazon is probably the most famous example that uses recommendation engine analytics. In the past, all types of recommendation-based analytics were quite difficult to automate as the data preprocessing and model creation and maintenance required were complex and generally needed different systems working together. KNIME, as a robust analytics platform, simplifies this process by providing all core functionality in one workflow necessary to create a recommendation engine.

Real-Time Model Execution has a number of different connotations. In general, it moves the use of statistics from a fixed activity on historical data to a process of collecting "real time data", running the appropriate analytics, and then immediately taking appropriate action based on those new results. The classic examples of real-time model execution include looking for fraudulent transactions when using a credit card online and generating trend information on the stock market. In other areas, the advantages of real-time have not always been as clear.

For many end consumers, the perception that something was done immediately to generate a specific or personalized offer action is what is important. Real-time execution can be implemented extremely well by combining the KNIME analytics platform and the Zementis ADAPA real-time decision engine.

The Predictive Model Markup Language (PMML) is the de-facto standard for the exchange of data mining models supported by all major commercial vendors and open source platforms. In a first step, we use KNIME to generate initial models in PMML. The PMML models are then passed to ADAPA for on-demand execution with real-time data as it appears.

Case Study: Online Music Store

Let's imagine we are running a business that sells downloadable music online. One of our differentiators is being able to show a selected music artist, but also the popularity of that music artist and which other music artists are also listened to alongside the chosen artist. This is a classic recommendation topic that adds value to a listener (and possible purchaser). At the same time as providing our recommendations, we can offer a "good price" to entice viewers to buy from us rather than going somewhere else.

Both the Recommendations and the Pricing are classic topics for advanced analytic modeling, and both offer possibilities for real-time scoring, as the more accurate our models, the more accurate both the recommendations and the prices we can offer. Whereas there ARE examples of recommendation engines that need to be real-time scored (and the KNIME / ADAPA combination is very capable of doing that), our case study only occasionally gets the complex data needed for calculating recommendations and therefore will focus on real-time pricing.

Pricing can be very volatile and fast changing. Real-time pricing decisions are therefore extremely important to the organization trying to sell a product as well as show related artists. Our company's goal of course is to always maximize both total sales AND margin.

Pricing elasticity is well documented in the literature. When an item is desired, there is more of a willingness to pay a premium price. When an item is less desired, the price will play an important part in the decision to move from "undecided" to "purchase". Discounts are a classic way of helping customers not only choose a particular supplier, but to help a customer move from undecided to purchase.

At the same time, discounts are expensive. They eat into the profit a company makes. In an ideal world, we would use an understanding of the desirability of a product, and use discounts appropriately.

For desirable items, such as in our example focusing on the top selling music artists, we would offer a competitive (but not discounted) price. For medium selling items, we would give a small discount to "nudge" customers into purchasing. For items that were not selling well at all, we would offer a heavy discount – a sale with a small margin being preferable to no sale at all.

In our example, we have recent (but historical) sales data for each of the artists available. We can use clustering to determine three categories of musicians, and provide the relevant message for each: "every day low price", "10% discount" or "15% discount". For the exact moment when we created and ran our model, these decisions would always have maximized our revenue and profit.

But today's online buyers are very savvy and price sensitive, and the "popularity" of an item as well as the price at a competitor's site might change quickly. We would want to respond to those external factors so that we could continue to

a) Maximize our sales and
b) Maximize our margin

This is a classic case for real-time decisioning. If we constantly receive the new sales data, we can automatically reclassify each musician dynamically into the appropriate discount class, and therefore ensure that we are constantly maximizing not only sales but margin. Here is how we would do that using KNIME and ADAPA using social media data publicly available from www.Last.FM.

The Basics: Recommendation Engine and Pricing Model

For our example, we will use Social Media data from the popular music site www.last.fm* and use a predictive analytics technique to make music preference recommendations for the top artists in the form "Others who like X also like….".

First, we transform the Social Media data to make it suitable for association, perform an advanced association analysis, and utilize the resulting statistics to select lists of artists and recommendations, combine this list with overall facts about the sample and enhance the artist data with pictures to create a dynamic multi-media report.

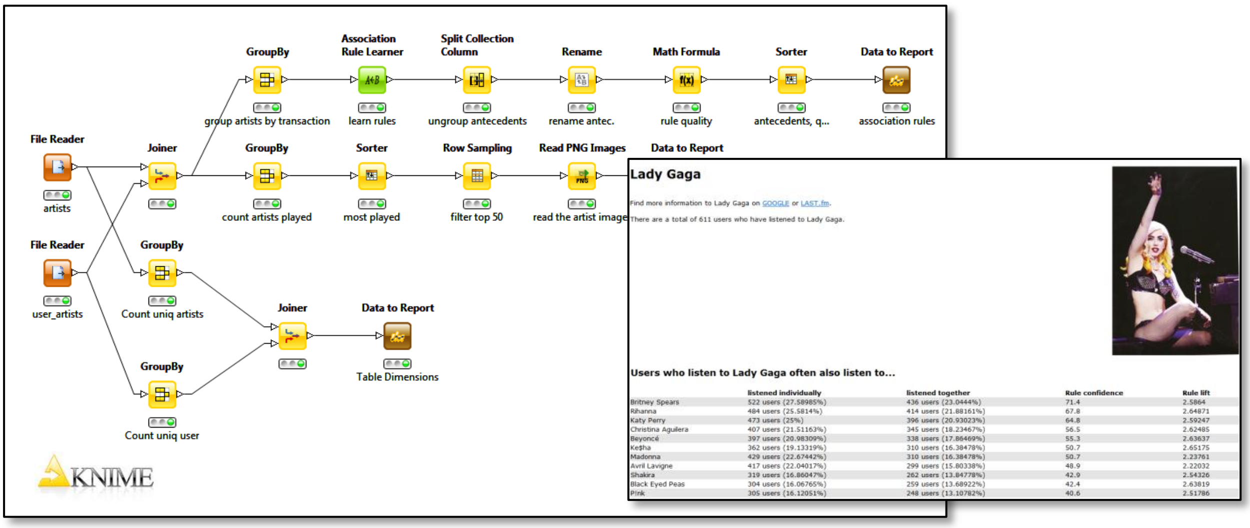


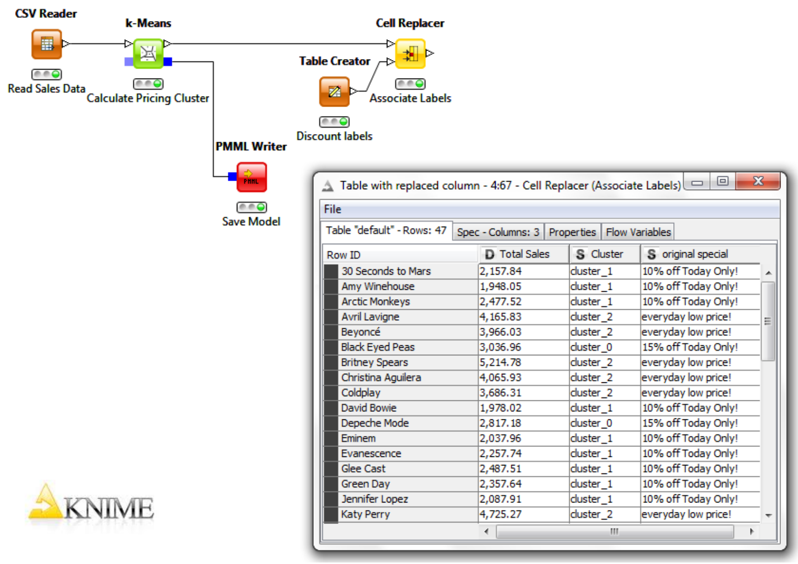
Figure1: Recommendation Engine Example

This workflow and a corresponding report are available at http://knime.org/node/52703

*This dataset contains social networking, tagging, and music artist listening information from a set of 2K users from Last.fm online music system. http://www.last.fm

The dataset is released in the framework of the 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec 2011) http://ir.ii.uam.es/hetrec2011 at the 5th ACM Conference on Recommender Systems (RecSys 2011) http://recsys.acm.org/2011

Next we need to utilize music artists' sales data to create our 3-tier pricing strategy "every day low price", "10% discount" and "15% discount". To do this, we have taken a file of fictitious "sales data" for our artists and run a k-means clustering algorithm to assign each of our artists to a cluster and to assign a label to each of our clusters. The resulting flow and data would look like this:



Row ID	Total Sales	Cluster	original special
30 Seconds to Mars	2,157.84	cluster_1	10% off Today Only!
Amy Winehouse	1,948.05	cluster_1	10% off Today Only!
Arctic Monkeys	2,477.52	cluster_1	10% off Today Only!
Avril Lavigne	4,165.83	cluster_2	everyday low price!
Beyoncé	3,966.03	cluster_2	everyday low price!
Black Eyed Peas	3,036.96	cluster_0	15% off Today Only!
Britney Spears	5,214.78	cluster_2	everyday low price!
Christina Aguilera	4,065.93	cluster_2	everyday low price!
Coldplay	3,686.31	cluster_2	everyday low price!
David Bowie	1,978.02	cluster_1	10% off Today Only!
Depeche Mode	2,817.18	cluster_0	15% off Today Only!
Eminem	2,037.96	cluster_1	10% off Today Only!
Evanescence	2,257.74	cluster_1	10% off Today Only!
Glee Cast	2,487.51	cluster_1	10% off Today Only!
Green Day	2,357.64	cluster_1	10% off Today Only!
Jennifer Lopez	2,087.91	cluster_1	10% off Today Only!
Katy Perry	4,725.27	cluster_2	everyday low price!

Figure 2: Pricing Model Example

At this point, we have both a solid recommendation engine and a solid pricing model that we can use on our website. And for that first day, our pricing model will be as accurate as possible. However, over the coming days and months, our pricing categorization will not necessarily be entirely accurate, as the popularity of an artist and the effect of our competitions pricing will be constantly changing purchase behavior. For example, if we take 30 days of musicians' sales data and look at just one musician's sales numbers for each, we see a lot of variation:

Without accounting for this variation in popularity in our pricing strategy, we will either not sell music that we could have sold, because the perception is that our price is "too high", or sell music with a too-high discount that, being popular, would have sold anyway.

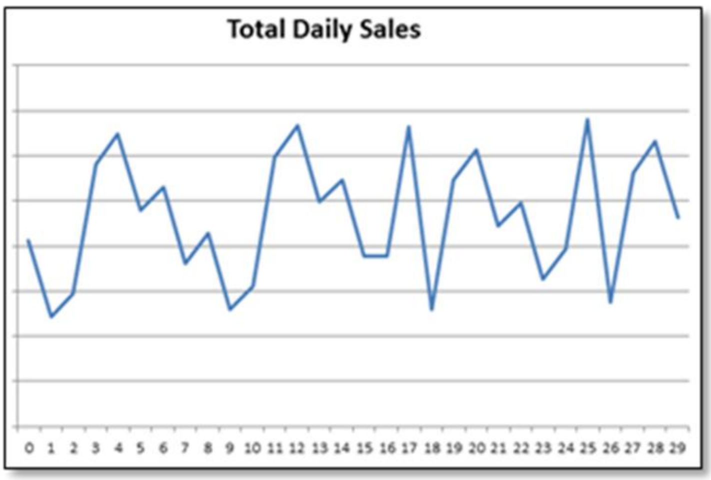


Figure 4: Daily Sales Variation for one Artist over 30 days.

THIS is where the power of real-time comes in to play!

Real-time in Action

In Figure 3, you will notice that we also created and stored a PMML model of our clustering. This accurate model can be rerun within KNIME as new data becomes available. However, the PMML model can also be used by the ADAPA Real-time Engine to not only rescore our musicians but instantly take appropriate action based on that new scoring. That is what we will do here.

The power of ADAPA standards-based real-time scoring engine is in the execution. When new data is available – daily, hourly, by the minute or even triggered instantaneously – the ADAPA scoring engine used the PMML model to recalculate instantly, either in-database, on-site or in the cloud.

For our example, we have new sales data coming in for each musician daily, and we use the KNIME-created PMML model to instantly recalculate and rescore the price we will offer the potential customer. Our sales data could be coming from a database, a real-time feed or another data source. For demonstration purposes, we will have the data coming from an Excel Spreadsheet that is made available each day.

To execute ADAPA, we will use ADAPA in the cloud via the free EXCEL add-in, available under https://myadapa.zementis.com/adapaconsole/. First, we load our model into ADAPA and make it available for execution:

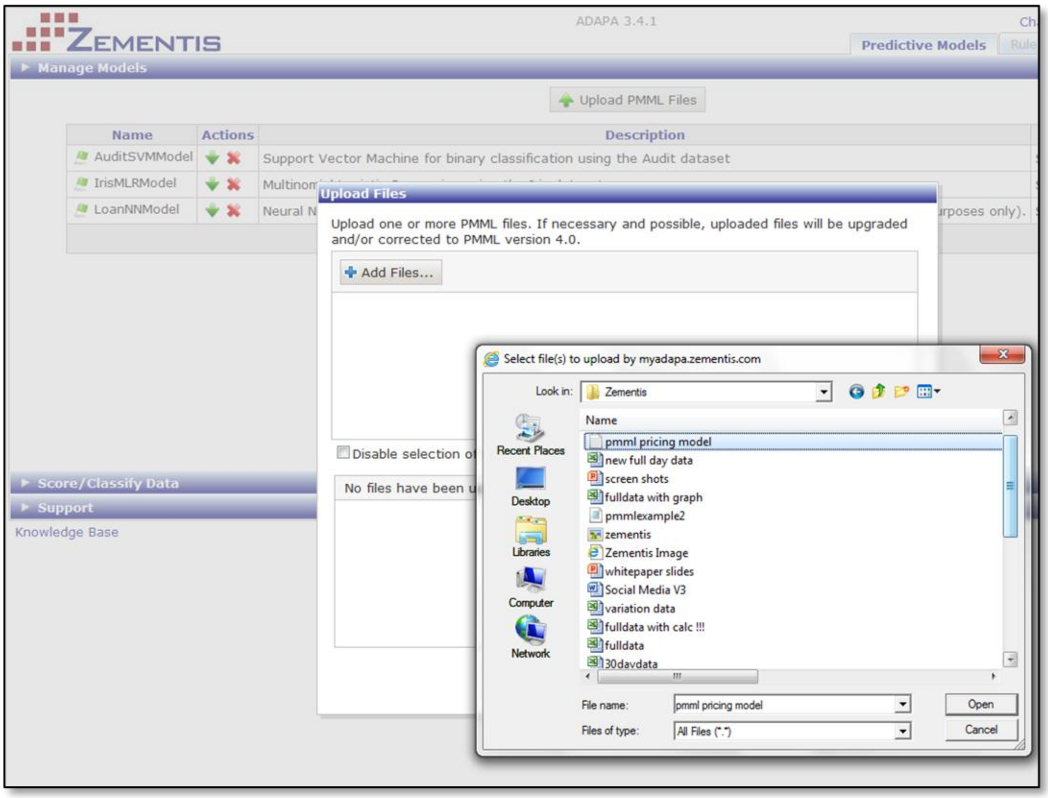


Figure 5: Loading the PMML model into ADAPA

In a production environment, this transfer of PMML model between KNIME and ADAPA could be automated.

We then point ADAPA to the source of our new sales data and automatically rescore that data.

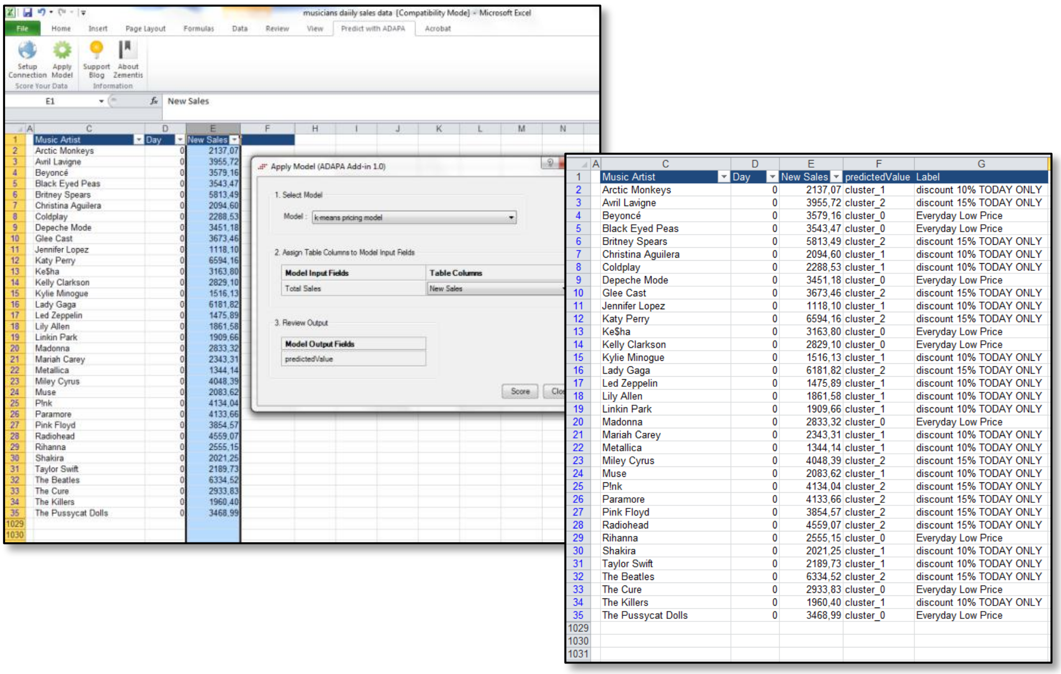


Figure 6: Rescoring our pricing model on a new day's data

In our case, ADAPA automatically reruns the scoring every day and instantly makes those new pricing decisions available and, in our case, could push them to the web application for immediate use.

Here is how it might look for a user accessing the website on day 1 and again on day 7 for the same artist:

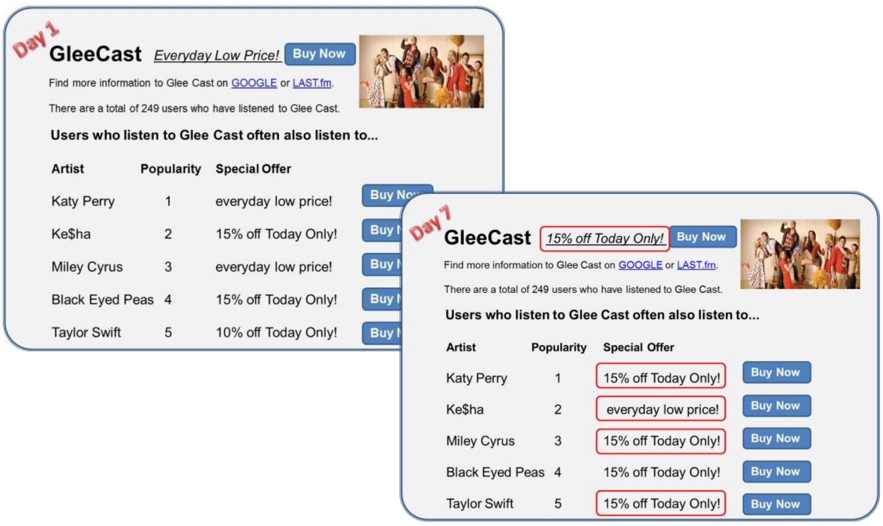


Figure7: Special Offers for different days based on real-time scoring of sales data

Note that the discounts provided have changed based on the sales data we had. ADAPA ensures that we are using our KNIME created models to maximize not just sales but also our margins daily. If we had up-to-date recommendation data available to us, we could generate a recommendation engine PMML from KNIME and use the same ADAPA approach to not only give the best price, but to also give the current accurate list of “Others who listened to X also listen to…”.

Why is this all important? Because it means not only a better personalized user experience for our customers and prospects, but gives us a solid increase in turnover and margin!

The Bottom Line: Increased Sales AND Margin

What does the combination of the two platforms actually bring our online music company?

If we assume in our example that our margin on downloadable music is 5%, the average purchase rate via recommendations is 2% for popular, 1% for not so popular and .5% for unpopular and that our “discounts” are 0, 10% and 15% respectively, we can run both our “fixed model” (always keeping the discounts as they were in the original learning excursive) and our PMML model, which is dynamically changing daily, over a 30 day period and compare the results.

For our example, this would result in a **12.8%** increase in overall sales as well as an impressive **26.8%** increase in margin. Why? Because the combination of KNIME and ADAPA has ensured that:

1) We don’t accidently price an artist too high, driving the buyer away or to another site
2) We do use price elasticity in the form of discounts correctly thereby ensuring that the appropriate discount gains us as many sales as possible
3) We don’t “give away” excess margin by providing a discount either where one was not necessary or where a larger discount was not required to gain as many sales as possible.

Our case study is just one example as to how advanced analytics combined with real-time execution has real world benefits for organizations. Regardless of whether a requirement to control risk, increase personalization with the customer or maximize sales and margin exists, the combination of KNIME and ADAPA are ideal for leveraging the power of data by providing an end-to-end solution, from model development to operational deployment and real-time execution within any business process.

For more information on KNIME, please go to www.KNIME.com.

For more information on Zementis and ADAPA, please go to www.Zementis.com

