Data mining techniques in Real-time Marketing[1]

Vladimir Gromov
Software Engineering School
National Research University Higher School of Economics
Moscow, Russia
v.gromov@hotmail.com

Scientific Advisor: Prof. Sergey Avdoshin
Software Engineering School
National Research University Higher School of Economics
Moscow, Russia
savdoshin@hse.ru

Abstract – **This paper gives an overview of the concept of a new system to support CRM in real-time using data-mining techniques. To ensure that in the modern world of dynamic companies remain in the leaders of their industry they need to continually monitor activity of their customers. Such activities are performed by analysts. Nevertheless, people are unable to handle huge amounts of data, which are encountered by such organizations as banks or mobile operators daily. In this situation information systems come to help. Software Engineering Department Higher School of Economics, in collaboration with IBM Company is conducting research in this area; the interim results have been examined in this work.**

I. INTRODUCTION

Different people use different products and services for different reasons. However, they are all part of the company's customer base and demand a handling that best suits their characteristics. Nowadays companies tend to build long-term relationship with the customers. In order to maintain such relationship companies need to discover the customer's needs. However today companies often have huge customer base, so it is unprofitable to handle each customer apart from others. Another problem is that if company wants to promote some specific product among its customers, including all customers into the target group would be either expensive for a company, or irritating for them, so they may turn to competitors.

In such situation data mining comes to help. With aid of historical data company can identify features of customers that lead to acceptance of some product or offer so that it would be possible to narrow the targeted audience. It could be also useful identify group of customers that require specific handling in order to ensure their loyalty.

II. CURRENT SITUATION IN CRM

Nowadays marketing is automated mainly with use of CRM systems. There are several types of CRM systems, but we are interested in the analytical systems.

They provide the following capabilities:

- Classification of customers by some basis
- Analysis of market situation and competitors
- Analysis of choice and price of goods
- Analysis of conducted sales
- Analysis of purchases and supplies
- Accounting and evaluation of marketing campaigns

However, traditional systems are unable to apply data mining techniques in order to make predictions about customer's behavior and aid company in making marketing decisions. Moreover they are unable to handle huge amounts of data in real-time. The proposed solution is aimed at overcoming these issues.

III. PROPOSED METHODS

The proposed method implies usage of several types of models [2]:

- Response Modeler — Increases ROI on acquisition campaigns by identifying and allowing to target those prospects who are most likely to respond to direct mail campaign.

- Cross Seller — helps maximize sales of products and services to existing customer base by identifying those customers that are most likely buy other products and services (e.g., additional services or products based on related product and service purchases) as well as the specific products/services a particular customer would be most interested in.
- Segmenter and Profiler — helps analyze and better understand customers for more one-on-one marketing by segmenting them into homogeneous groups (using natural clustering methods, segmentations generated from response, cross-sell or customer valuation models, or using manually defined segmentations) and profiling them. The results of these groups can be used to enhance customer acquisition programs (by finding prospects that are similar to existing customers) as well as retention programs (by making a focused offer that will appeal to a specific group).
- Customer Valuator — predicts the spending level or profitability of your customers over a specific time period to help forecast demand, strategize acquisition campaigns, and identify most valuable customers. These models can be used in many ways to optimize marketing efforts.

Provided the data mining models are properly built, they can uncover groups with distinct profiles and characteristics and lead to rich segmentation schemes with business meaning and value [1].

IV. A CUSTOMER SCENARIO FOR THE PROPOSED SYSTEM

Let us consider a customer that wants to offer advertisements to consumer who visits CNN.com.

We have several advertisements at the disposal for offer. So, our extreme opportunities are:

- Bid low on everything
- Bid high on only those you are most confident of a hit.

The second option is available if we could build up an "anonimized profile" of a consumer based on:
- Transactions from this customer
  - Cardholder since YYYYMM
  - Average transaction value
      - Monthly transaction value
      - Categories purchased
      - Brands purchased
- Descriptive
  - Age
  - Gender
  - Family situation
  - Zip code
- Interactions
  - Web registration
  - Web visits
  - Customer service contacts
  - Channel preference
- Attitudes
  - Satisfaction scores
  - Shopper type
  - Eco score

All these data allow us identify consumer as well as predict his response on one or another advertisement. The data will be processed in real-time and decision will be made once consumer enters the site. All decisions are considered and used for further scoring.

So now let us assess the how such system can be implemented using the sample data.

V. SYSTEM STRUCTURE

The sample data represents statistics of bank loans decisions. There is no information about input data, as it is covered and changed to meaningless symbols and number. The only field we have information about is final decision: positive or negative.

First of all, a model is built in Modeller that encapsulates the above mentioned techniques (fig. 1).
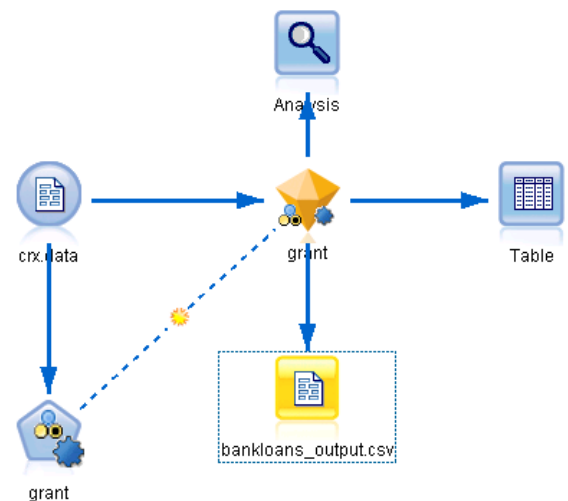

Figure 1. SPSS Modeller model

On the figure 1 there is a model that consists of data source node that imports data into model.

After that auto classifier comes to work. It uses several data mining techniques and selects 3 decision trees that provide highest confidence. In future scoring will be performed and prediction will be maid according to the prediction with the highest confidence.

After that the model is trained on some historical data. Another portion of historical data is used for evaluation of obtained model.

Next step is exporting of model and integrating it into IBM InfoSphere Streams operator (fig. 2). IBM InfoSphere Streams is a runtime environment that allows easy distribution of load between nearly unlimited number of computational nodes.
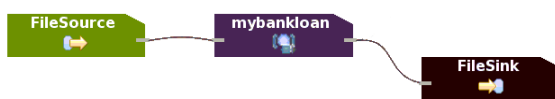

Figure 2. IBM InfoSphere Streams application graph

On figure 2 there is an application graph that denotes structure of InfoSphere Streams program. In this application data is read from an input file, than it is passed to an operator that applies SPSS Modeller model to the data and finally scoring results are put to an output file. The model is executed in the Streams program by special operator that calls the SPSS Modeller Solution Publisher, with the help of special API.

The training never stops. Every time new data arrives, model is trained once again. So, each time a transaction is made by customer, the system estimates whether he is likely or not to accept some marketing proposal and depending on estimation makes offer.
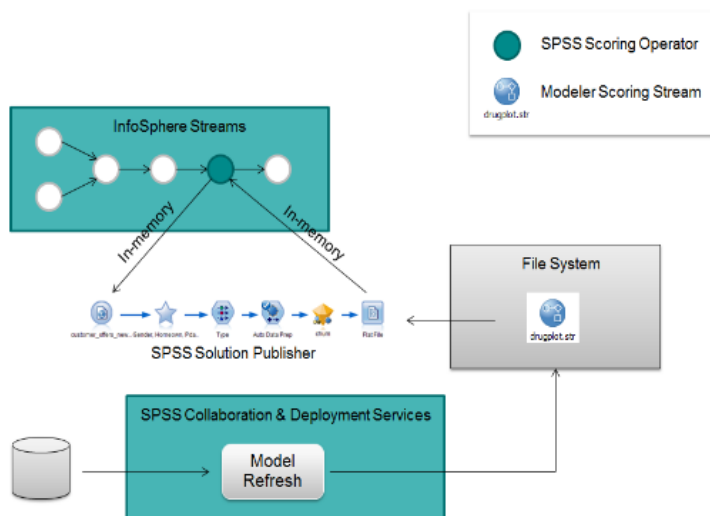
As it is seen from the figure 3, the model is refreshed in real time by SPSS Colloboration and Deployment Services.

VI. CONCLUSIONS

In this work data-mining techniques and information system implementing these techniques are presented. There are some tasks that are already implemented, like integration of Infosphere Streams and Modeller. The next step would be identification of set of models that would be used to analyze input data and transform it to output. The architecture of Infosphere streams allows to split the system among several computational nodes and thus it can be easily scaled in order to meet business needs.

REFERENCES

[1] K. Tsiptsis and A. Chorianopoulos, *Data Mining Techniques in CRM* 1st ed., John Wiley & Sons, Ltd. 2009

[2] Ing. Vladek Šlezingr. Campaign Management system Technical Proposal For: Home Credit International a.s. 2011

Figure 3. Operation of model.