Open Source Data Mining

by [OSOE Project](#).
▼ Details

Open Source Data Mining with R and ERP5

▼ Details

Data Model and Report Model

▼ Details

The first thing to understand with ERP5 is that the data model, i.e. how data input by the user is stored in the Object Database, and the reporting model i.e. how data is structured in tables, are completely independent.

ERP5 departs from traditional object relational mapping (ORM) – a technology which is known not to work since the early 90s - to object relational indexing (ORI) – a technology which is extremely scalable and flexible. ORI is one of the secret weapons of ERP5 which makes it way more scalable than many ERPs, and much easier to handle when it comes to reporting and data mining.

In ERP5, it is possible to configure tables in the so-called catalog, in such a way that document properties extracted from the object database are structures in such a way that reporting can be extremely fast.

In a sense, ERP5 has its own built-in data-warehouse. New tables can be added any time, without interruption of a production system, if specific analyses are required and current tables are not sufficient.

Therefore, the first step in data mining with ERP5 is to feed the catalog with appropriate data.

PALO: Pivot Table on Steroids

▼ Details

Before starting with the data mining within ERP5 we will present three open source solutions which from our point of view are worth to take a look at in our case.

The first solution is JEDOC PALO which is a kind of "dynamic pivot table on steroids". This tool allow you to perform simple analytics and data mining reports very easily thanks to it ability to dynamic pivot tables.

This solution must be considered if you wish to have beautiful reports with graphs and if you want to use only menu driven reports.

Rapid-I: Open Source Analytics

▼ Details

Rapid-I is a little bit more complete than PALO, but also a little bit more complex. You will be able to do much more than pivot tables and this solution is used by big companies to do machine learning, or to predict trends or behavior from recorded data.

In order to use it with ERP5 you will have to extract data from the transactional database then you will be able to use them.

R: Open Source Statistics

http://www.agrocampus-ouest.fr/math/useR-2009/slides/Wijffels.pdf

▼ Details

R is a statistical language which is used for example by "Thomas Cook" a travel agency to predict the prices of the flight tickets for the last minute booking. For that they are using mathematical theorems.

R can be used with all scientific mathematics and this language comes from the statistical language rather than from the

informatics (such as Rapid-I). This is why you will find more statistical libraries with R than with the other.

The cons about R is that it requires you to know how to program it and use it. It uses less menus and more the programming skills of the user.

You should always evaluate Rapid-I and R while you have a Data Mining project. They will always have pros and cons for both of them, depending of the characteristics of your project.

ERP5: Charts, Forms and Gadgets

▼ Details

Of course we couldn't finish the presentation of the data mining tools without speaking a little bit about the report engine which is build in ERP5.

In ERP5 you can by default create any report you want using the SQL and the python language. Those reports can then either be displayed as a standard ERP5 form, or you can also use the build in Open Office engine to use the results and create graphs with them.

This engine is already really powerful and is sufficient most of the time.

You can even go further by creating a custom dashboard with the custom gadgets rendering your custom reports. By doing so, you have everything you need to measure the performance of your company and to take good management decisions.

In today's lesson we will use those capabilities which will be improved with the use of the R language.

In order to be able to use R, we added a library into your instances which let us use the R language and libraries within a python shell.

A Time Series Tutorial

http://www.statoek.wiso.uni-goettingen.de/veranstaltungen/zeitreihen/sommer03/ts_r_intro.pdf

▼ Details

The statistical method which will be used is named the "time series". The purpose is not to learn this method but to see how it can be used in the modern information system used by large companies to predict the future. If you are interested in learning the method, we recommend you to follow the above link.

This method is capable from the observation of the past data to predict new ones. So in order to test it, we will enter a function which will create sale order of a certain quantity of product. The quantity will follow a function which will be predefined. Once those data will be setted up, we will try to predict the future of the sale order. If it works properly, we should be able to find the function we used to create the data.

Data Maning with R and ERP5

▼ Details

First of all we will review or learn what is python and how to use it. The purpose is not to teach you this new language but it is to let you understand the code we will use during the lesson.

Second we will use the SQL wizard to create a report. Then finally we will work on the arima model with a series of simple sale orders.

Open Source Data Mining

▼ Details

If you don't have any more questions, we can do the tutorials.