

Laxman Dutt Degala

Vishnu Gude

Ayyappa

Akhileshwar

 | KDM – CS560 | April 17, 2014

CAAM for Smart Energy

Increment 3

**Please find below our updates for Increment 3:**

1. Recap (Increments 1 and 2)
2. Selection of most suitable model for the given data
3. R source code for making the application real time
4. Using Console application and visual studio for the real time processing
5. Final steps
6. **Recap (Increments 1 and 2):**
7. Data has 100000 records and 1350 attributes (Columns). Decision on attributes has been made, Out of 1350 attributes we are considering 25 attributes which could decide the energy consumed by a household for a month and predicts the $ amount for the energy consumed.
8. Decision has been made to use R and Rattle (**R** **A**nalytics **T**ool **T**o **L**earn **E**asily), which is a graphical Data Mining Application.
9. Description about various patterns/visualizations (Boxplot, Histogram) has been provided.
10. Different models have been created using various existing algorithms like Decision Tree, Random Forest, Linear, and Neural Net.
11. **Selection of most appropriate model for the given data:**

Models have been created using different existing algorithms like Decision Tree, Random Forest, Linear, and Neural Net. Score reports are generated for all these models and compared the target values of the Test/Train/validation data with the actual existing data. Model which gives the best result is selected for our data. Model that we figured out is Decision Tree model.

Now that the model has been built, this has to be used for generating new predictions for the user inputs.

1. **Real time processing using Big data:**

As we already know, we use Hadoop/Mahout for processing big data; as a Batch process. Thus, we would need a database to store the results from Mahout, a statistical analytics tool for Big Data.

In our case, we have SOLR or HBase wherein we store the results and later retrieve data from the same. This is not a real time processing directly from the user inputs. Even if we try to run it in real time scenarios, user has to wait for indefinite time to get his results.

Thus, we planned to use R script in order to make it real time. Visual basic (c#) supports R script, by just importing R related package and a few lines of R coding based on the data that we have.

**R source code for real time processing –**

Steps required for the real time processing:

1. Once the model is created using Rattle, all the R specific code can be found in the log.
2. Take the code and create .R script file with some additional code changes to customize the model for real time processing.

**Constraint & Resolution:**

Main constraint when huge data is involved or when predictions are involved is that it might take indefinite time to process the request, thus application could even hang-up.

But when used R-script for running the model whenever user inputs some data, It takes seconds to run the model and help user in prediction based on variety of inputs that user types.

Also, the input is not confined to limited values, this enables the application to run for variety of inputs from the user.

1. **Console Application:**
2. Create a Console application in C# to generate .exe file which would run the R script dynamically whenever the user inputs the details on UI of the application.
3. Whenever user types input, it’s dynamically loaded and the model that we created using rattle/R, is run for the input. This can also be assumed as the validation data. Instead of the validation data that we give while checking/testing the accuracy of the model, we would give a single record of data based on the user input. This makes a real time application.

Please find the attached R-script in Github.

1. **Final steps:**

With some more changes to be done for the console application, we are left with creating a front end (User Interface) for the application.

There is also a plan to insert the Histograms, Box plots in our application to make it user friendly. Our thought is that a visual picture can give a clear picture to the user.

This should be done in a couple of days.