Report for Sunbase Data Assignment

# Preprocessing the Data:

The basics, I converted all the columns to floats and integers. So that the data can be fed. Then I tried adding other features like “val2money” but the data was not corelated to the churn whatsoever that any feature I added proved useless.

But that did not yield better results, so I decided to comment it. Now I Scaled down all the data using StandardScaler

so that the model can yield a result faster. I also used PCA for Feature Extraction to remove the curse of dimensionality so that the Model could be even faster

# Model Selection:

I could not come in a decision with one specific Model so I employed 4 different kinds of models that gave me a satisfactory output.

DecisonTree Model

RandomForest Model

LogisticRegression Model

Neural Network Model

All of them gave me an accuracy of about 50%, since the data was so un corelated, this was the best I was going to get.

# Model Deployment

I used Flask and AWS EC2 engine to deploy the Model I had made. Along with some interactive front end.

# Data Corelation

The biggest flaw for the accuracy of the models was that the data was not co related whatsoever even if the data was slightly Co Related with important parameters like Monthly Bill or Usage GB. The Machine Learning models would have yielded better results