## **Model Based App Rank Forecasting**

**MSBA 2016 – Data Mining in R | Assignment 7**

**Team Members:**

* Sid Reddy
* Ashish Soni
* Samrat Kulkarni
* Nirmal Yuvraj
* Pradeep Menon

**Objective:**

       Forecasting the rank of an APP.

**Key Assumptions:**

       The time-series analysis will be done on one mobile app per genre.

       Based on a sample of the available data, the window of time-series will be from Jan 2015 to July 2015.

       Based on the available data, the apps are assumed to be ranked with a rank range between 0 to 199.

**Choice of Model types :**

* SVM (cost=10,gamma=0.01)
* MARS
* Custom work flow (50% Multiple Linear Regression and 50% RPart)
* Random Forests (ntree=500)
* Multiple Linear Regression
* RPart

**Workflow common parameters:**

* 'timeseriesWF',
* predictor='predict',
* type="grow",
* relearn.step=30,

**Performance Metrics can be chosen from:**

Relative error metrics such as Theil, name, nmse and few absolute error metrics.

**Baseline:**

Since APP Rank is timeseries data and ‘theil’ metric is typically used in time series tasks. The used baseline is the last observed value of the target variable.

**About the Shiny App:**

|  |  |  |
| --- | --- | --- |
| **Panel Name** | **Details** | **Features to highlight** |
| Start | App Data set needs to be loaded. It takes approximately 1 minute to load the dataset. Till then the other panels are disabled. | Using conditional flags other panels are disabled. |
| Select Mobile App | Select App Genre, Publisher name, App name.  Possibility to plot motion chart of App Rank.  The output console displays if any data belonging to the app is missing or not.  Only the available data can be used for rank forecasting and comparison of model performance. | The drop down menu content are reactive and the relevant data gets populated as the selections are made. |
| Model Performance Estimation | Choose model type, training depth and evaluation metric for model comparison. | Training set, Hold out set and the predicted data are plotted dynamically updated.  Different tabs display the data for clarity.  Model performance is compared and plotted. |
| Model based App Rank Forecast | Select Model type, number of days to forecast, plot of historical and forecasted rank. There is also an option to plot the technical indicators. | Forecasting utilizes different technical indicators that are relevant to any time series although mainly used in financial analysis. |

**R Libraries required:**

performanceEstimation

quantmod

e1071

randomForest

DMwR

earth

rpart

dplyr

plyr

shiny

shinyBS

googleVis

dygraphs

**Model output**

       Given a app rank time-series for ‘x’ number of days, forecast of the rank for ‘x+n’ days would be computed.

       ‘x’ and ‘n’ will be parameterised and can be changed based on user-input.

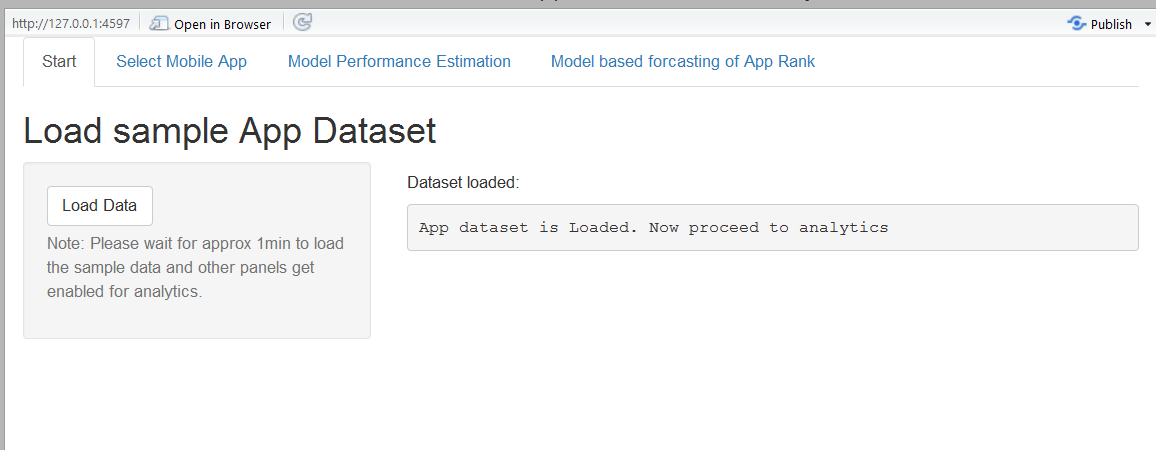
**Hypothesis**

 For apps that are a part of a large publisher family (Publisher wih 5+ Apps), we expect them to rank higher than a publisher with a single app. We would assume that the accuracy of predicting the rank of this group would be higher.

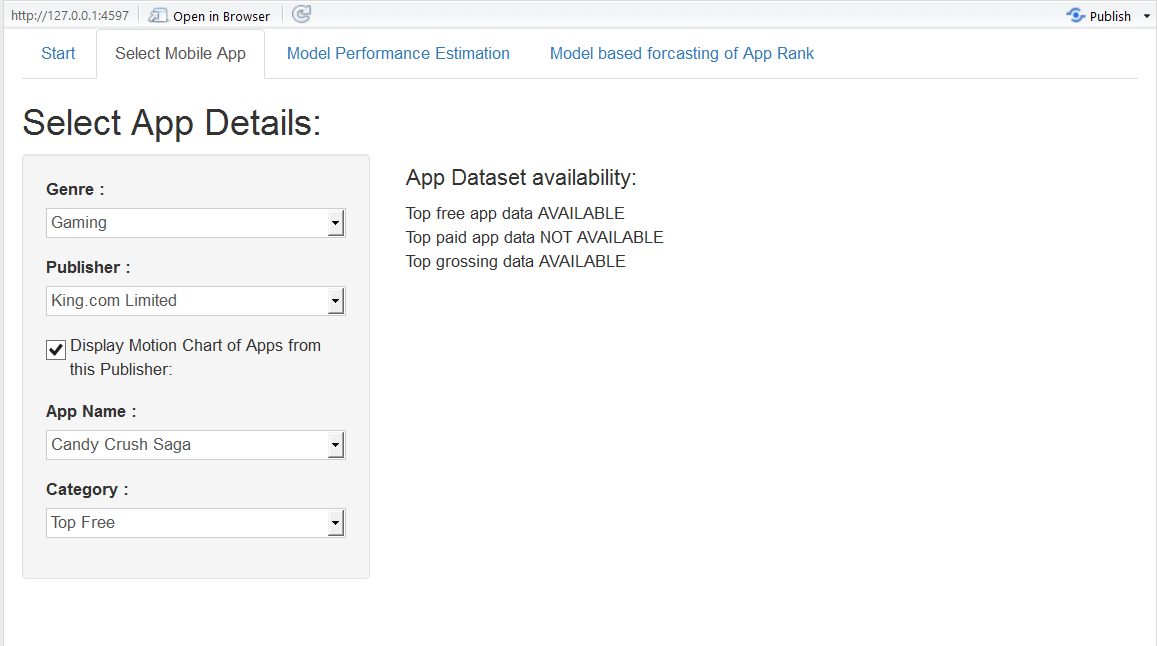
        Testing model accuracy on different genres – We expect a more “data” populous genre such as “Games” to be easier to predict.

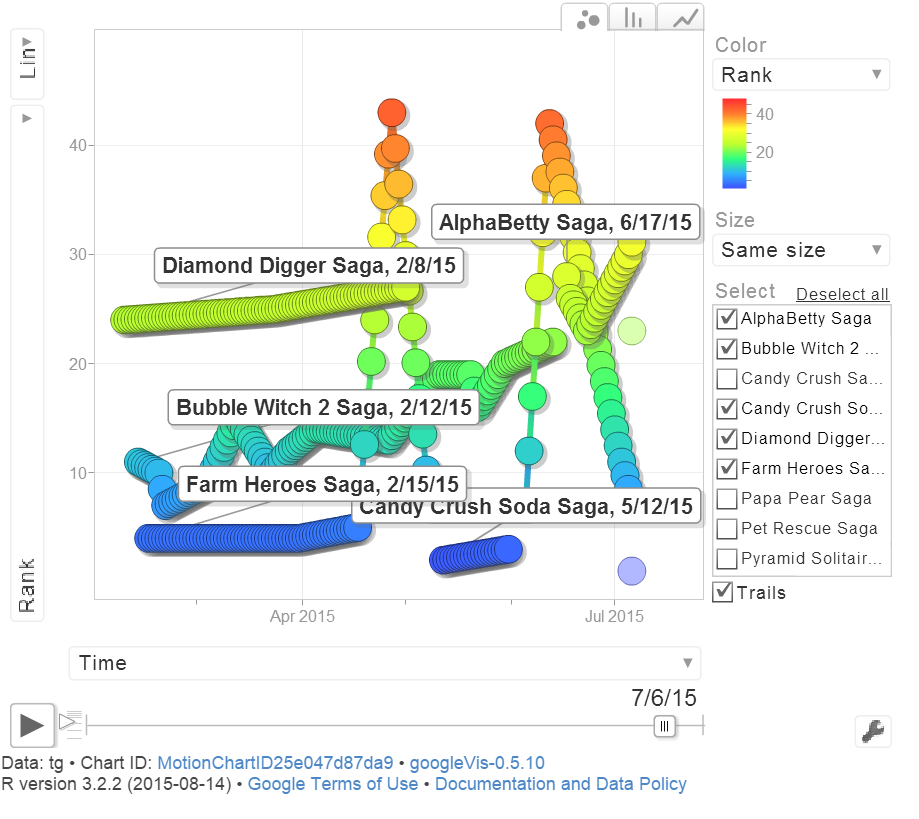
**Sample Output**

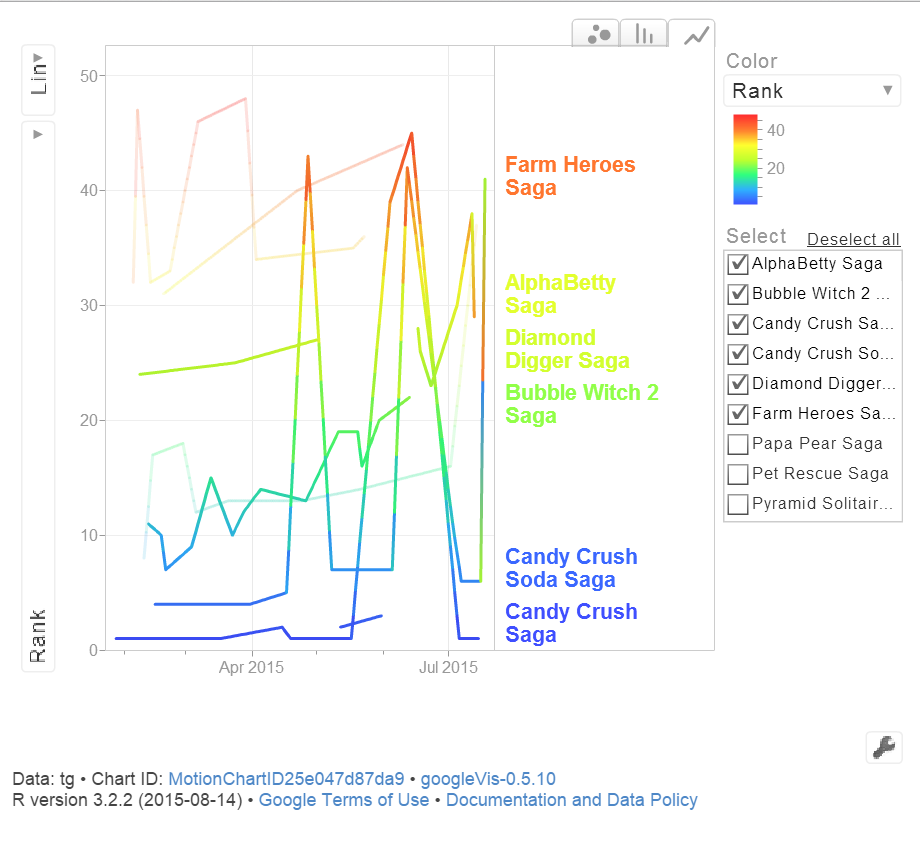
1. **Load Sample App Data**

****

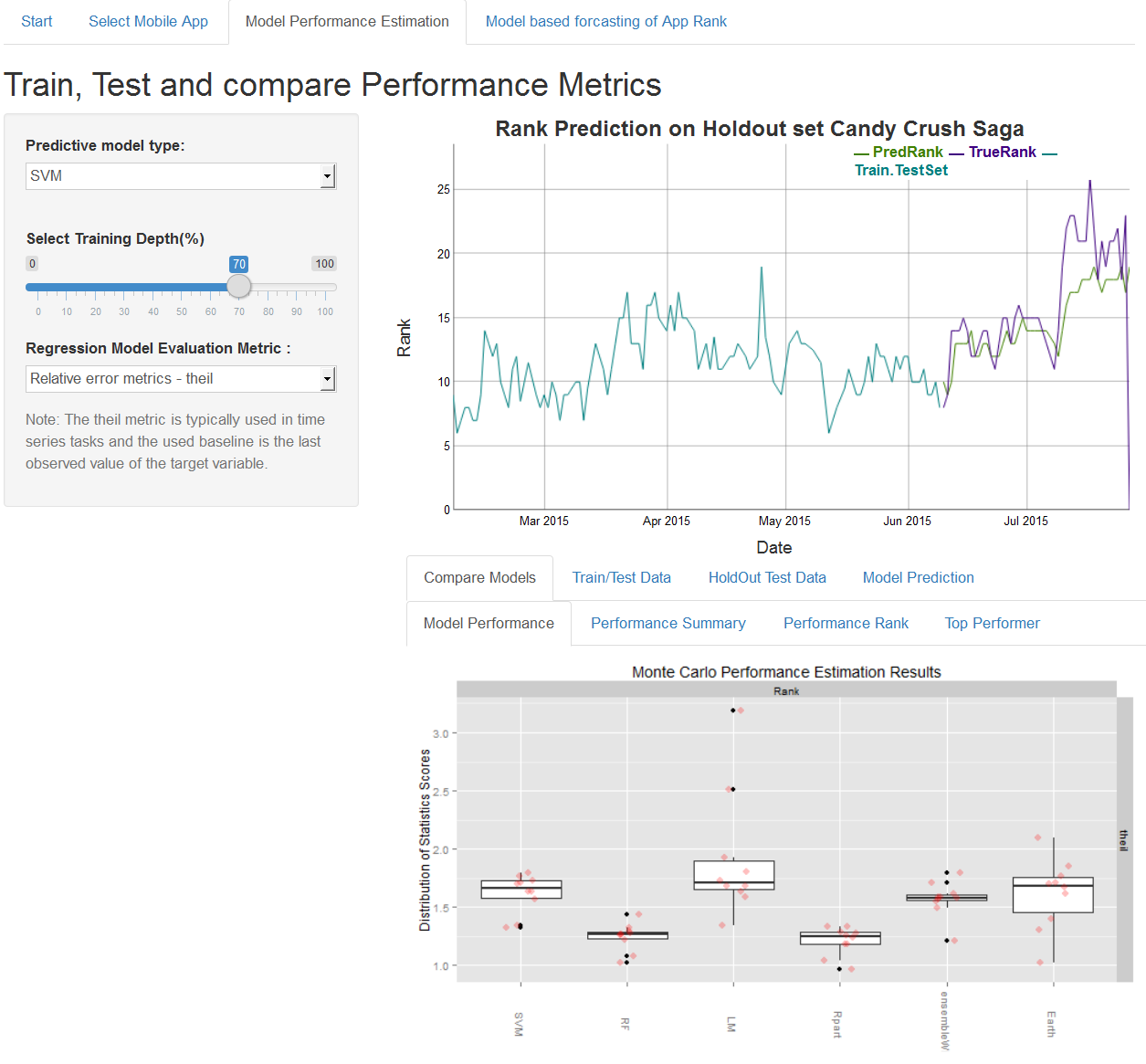
1. **Select App Details as well as Motion chart of Rank for App from same publisher**

****

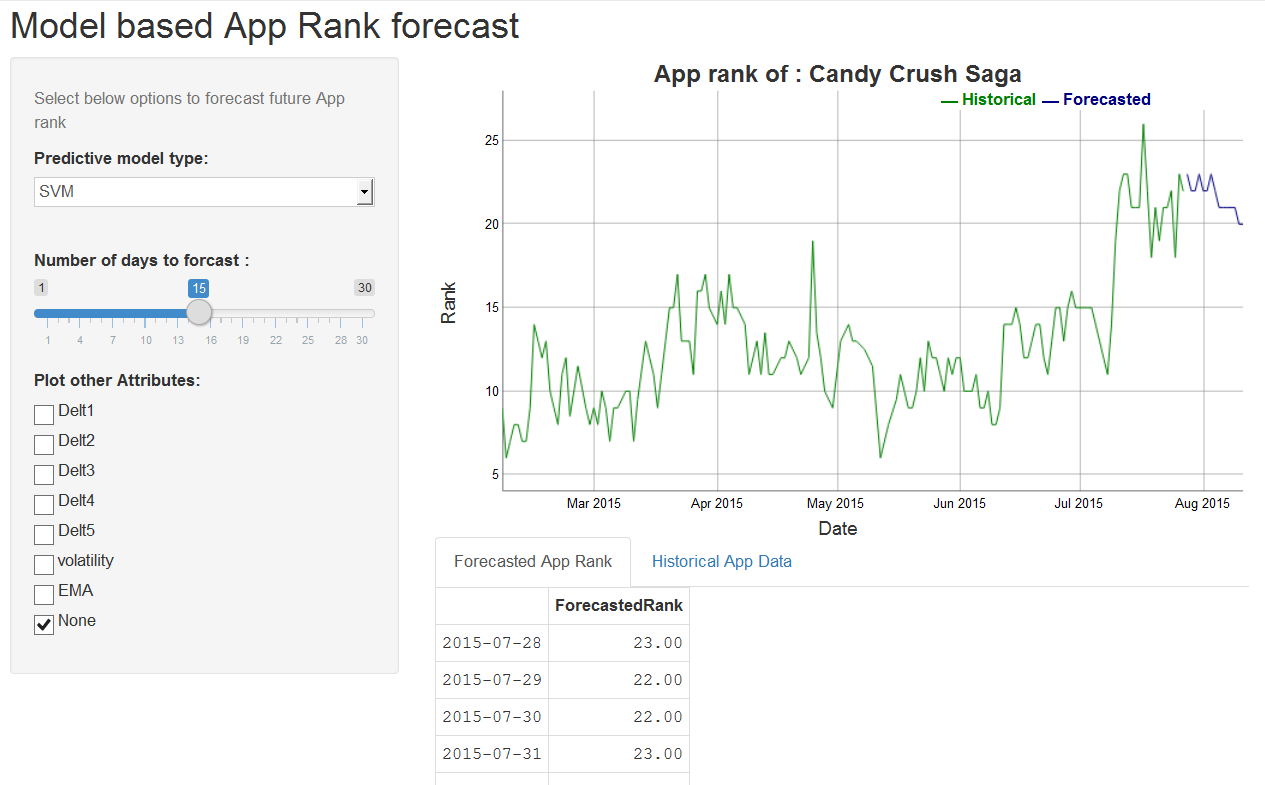
****

****

1. **Compare Model Performance on holdout set**

****

1. **Model based Forecasted Rank**

****

**With Technical Indicators selected:**

****