**Twitter Bot Detection using Machine Learning**

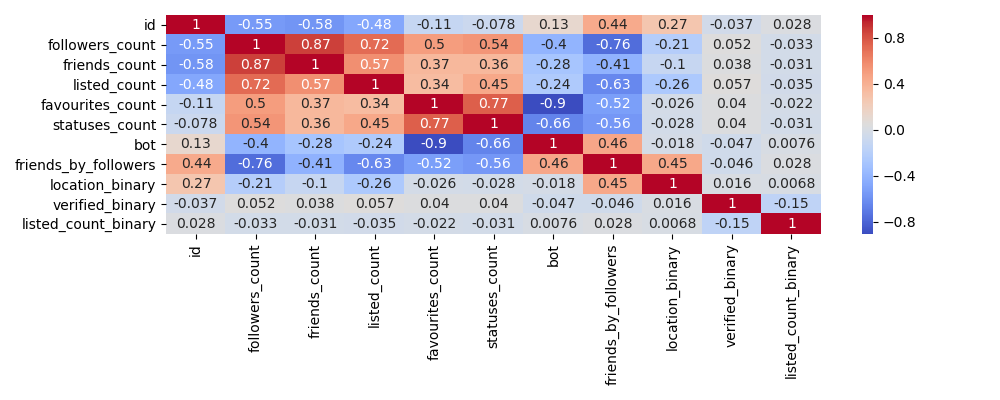
Twitter being one of the initial online social media platforms to allow bots to autonomously perform operations like tweeting, re-tweeting, liking, following and much more. An estimate of two-third of the tweeted links in popular websites are posted by these bots. In order to ensure the information shared across the platform through conversations is credible, there is a need to focus on identification of spam bots. This will be an effort to prevent malicious automation from disrupting user’s experience on twitter. Motivated by the need to identify and filter out the spam bots we have come up with an approach to distinguish between genuine accounts and spam twitter bots.

**Dataset:**

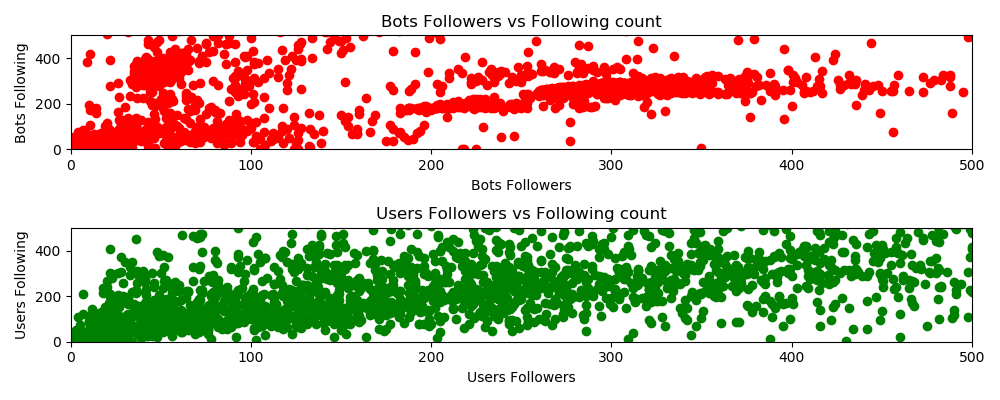
The dataset used for this project is available at <https://botometer.iuni.iu.edu/bot-repository/contribute.html>. I used cresci -2017 data which consist of 3,474 genuine accounts, 4,912 social spambots and 1,100 traditional spambots details such as number of followers, following count, listed count, created date. There are 2,839,362 tweets by humans and 3,457,344 tweets by bots.

**Data cleaning:**

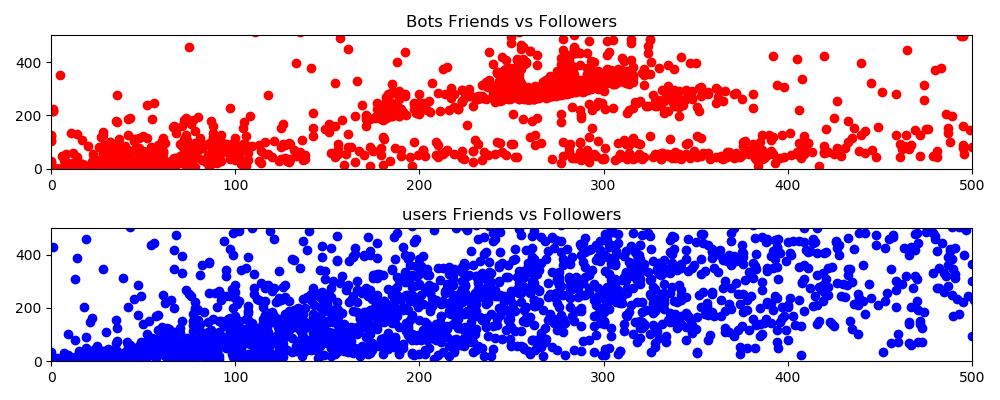
The data from traditional spam bots and social spambots are combined and a target variable “bot” is created with value 1 and for genuine accounts its 0. I used only numeric variables along with few categorical varibales for my analysis. The variables used are id, screen\_name, location, followers\_count, friends\_count, listed\_count, favourites\_count, created\_at, verified, statuses\_count, lang, default\_profile, bot, name and description. The screen\_name, location, verified is converted to Boolean variable. Number of followers for bots are much less when compared to users. The correlation between the target variables is given below:



The users have number of followers much less then following count. The below graph shows that.



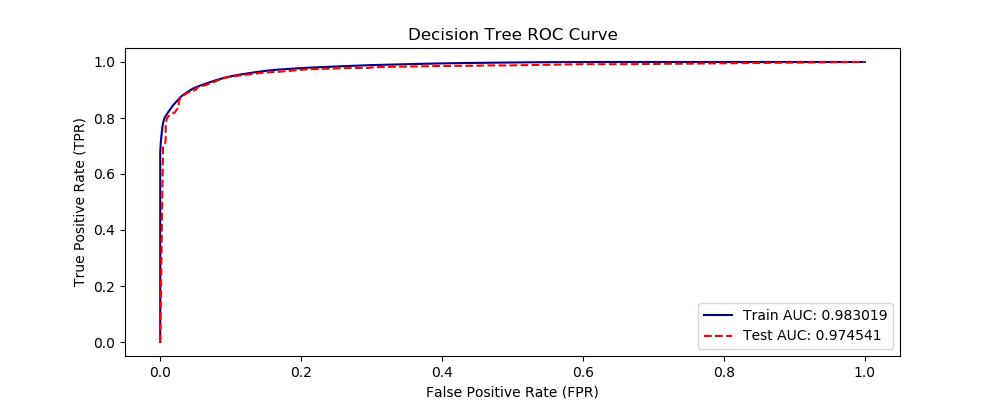
So, I created a new variable number of followers / numbers of following and used this variable in further analysis. Now let’s look at plot for number of friends’ vs followers for both users and bots. Number of friends for bots is much less when compared to actual users. The below graph infers the same



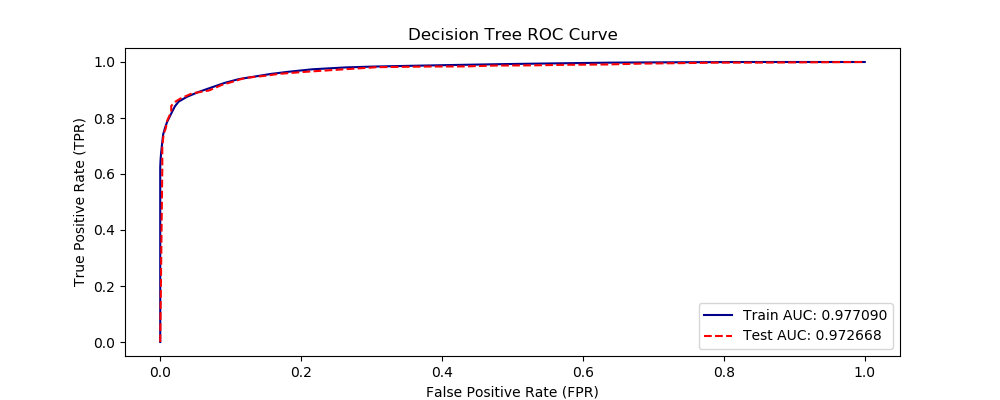
As there is lot of text in location variable, I created a binary class, whether location of user or bot is given and used in further analysis. The final dataset consists of verified, location\_binary, followers\_count, friends\_count, statuses\_count, listed\_count\_binary, bot. This dataset is split into training and validation data. Decision Tree, Logistic Regression, SVM are used for classification of target variable.

**Decision Tree:**

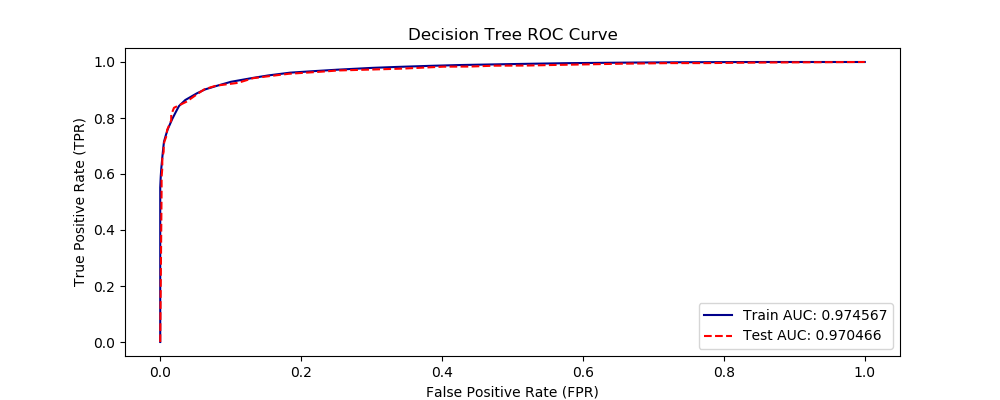
Decision Tree with entropy criteria, minimum sample tree = 50 and min sample split = 10 is built. The training accuracy is 93.34% and accuracy on test data is 93.25%. receiver operating characteristic curve is shown below:



Decision Tree with entropy criteria, minimum sample tree = 100 and min sample split = 20 is built. The training accuracy rate is 92.28% and accuracy rate on test data is 91.98%. receiver operating characteristic curve is shown below:

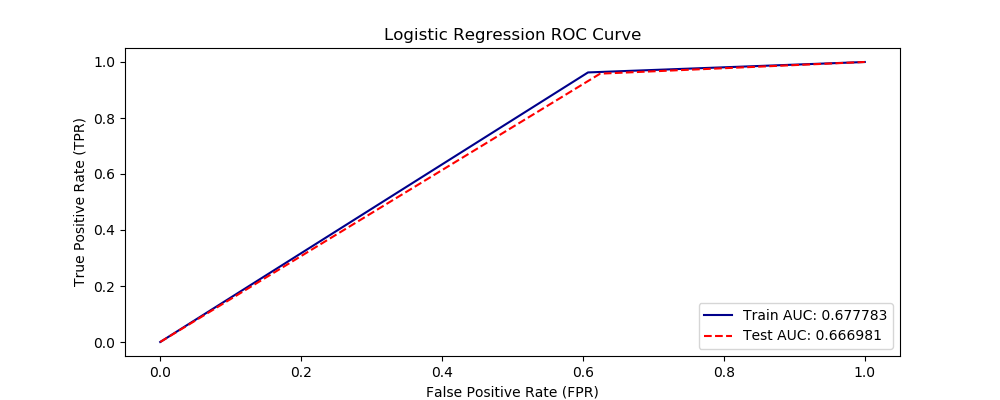


Decision Tree with **gini** criteria, minimum sample tree = **100** and min sample split = **20** is built. The training accuracy rate is 91.98 % and accuracy rate on test data is 91.44 %. receiver operating characteristic curve is shown below:



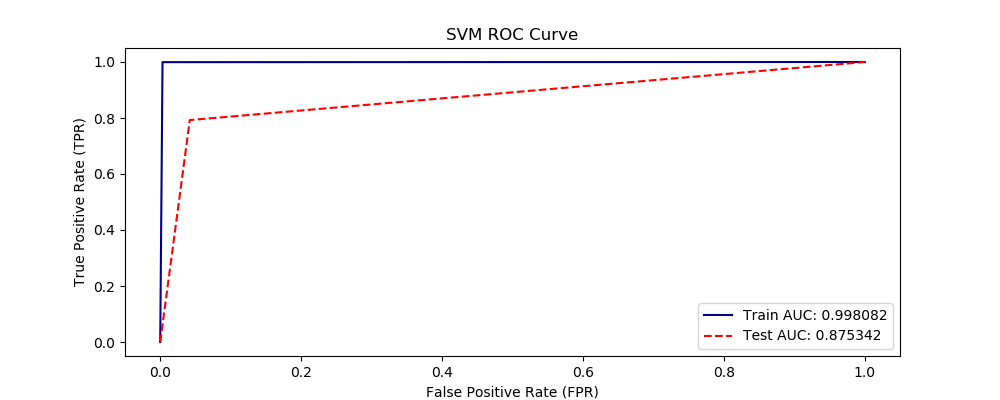
**Logistic Regression:**

Logistic Regression on accuracy of 78.17% training data and 77.77% on test data. The receiver operating characteristic curve is shown below:

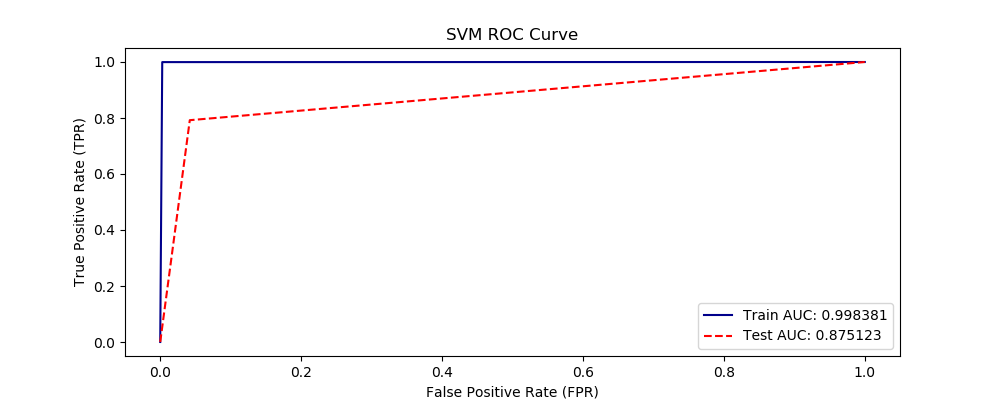


**Support Vector Machine:**

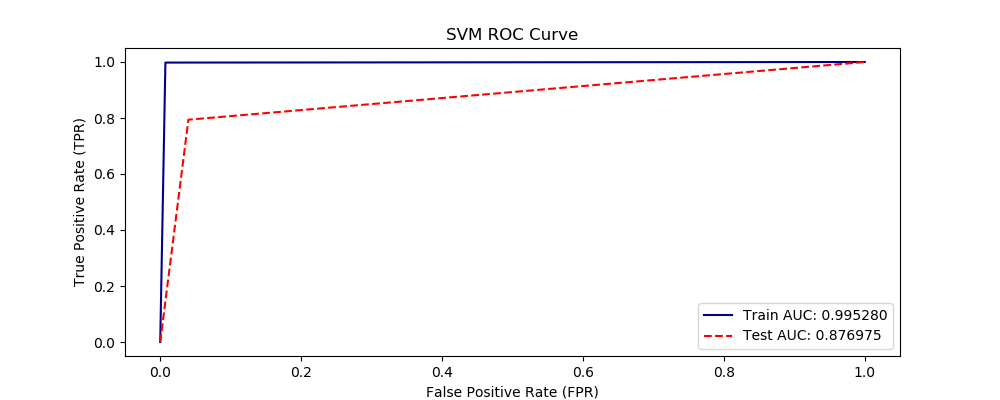
Support vector machine with gamma =0.001 and regularization parameter C = 100 the accuracy of the model on train model = 99.85% and on test data = 84.39%. The receiver operating characteristic curve is shown below:



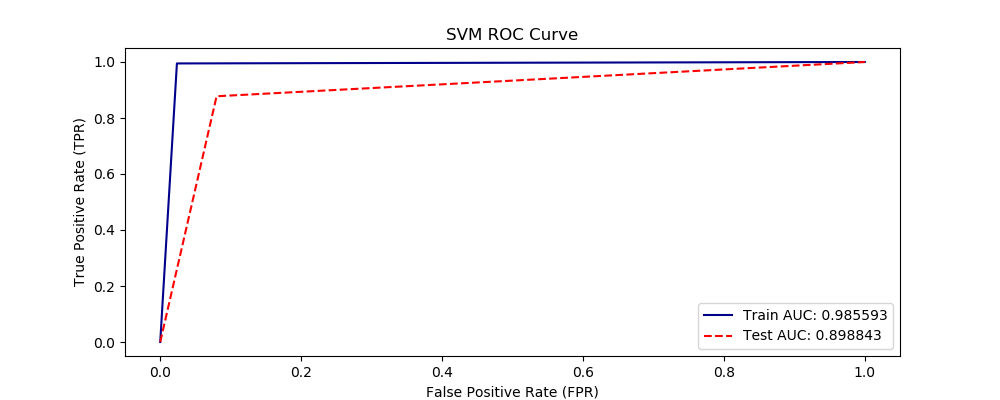
Support vector machine with gamma =0.001 and regularization parameter C = 150 the accuracy of the model on train model = 99.88% and on test data = 84.36%. The receiver operating characteristic curve is shown below:



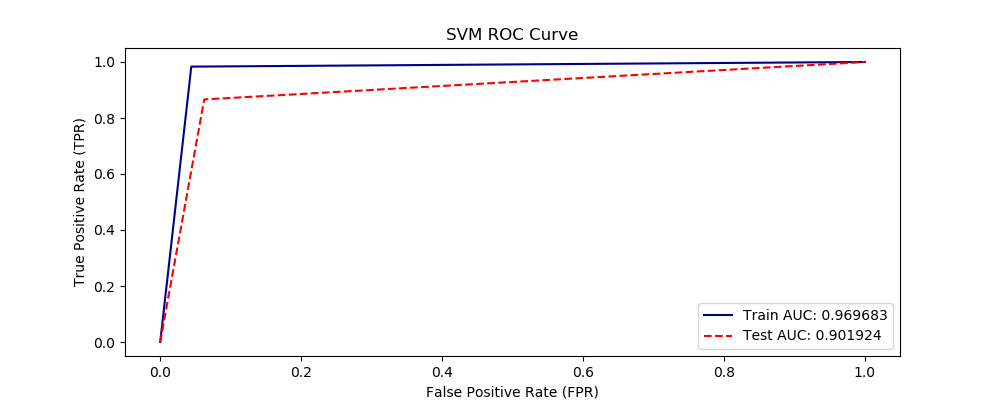
Support vector machine with gamma =0.001 and regularization parameter C = 150 the accuracy of the model on train model = 99.62% and on test data = 84.54%. The receiver operating characteristic curve is shown below:



Support vector machine with gamma =0.0001 and regularization parameter C = 100 the accuracy of the model on train model = 98.89% and on test data = 89.08%. The receiver operating characteristic curve is shown below:



Support vector machine with gamma =0.0001 and regularization parameter C = 1 the accuracy of the model on train model = 97.47% and on test data = 88.83%. The receiver operating characteristic curve is shown below:



Decision tree is the champion model with accuracy on test data 93.5%.

**Classify bot based on tweets:**

I merged all the social spam bot tweets and genuine accounts tweets and created a target variable whether the tweet is made by bot or human

**Data preparation** :

Removing the nan texts and tokenizing the texts. Then I used Keras to tokenize and took a maximum length of 250 for each tweet by padding zeros posteriorly. Using a 200,000 sample data for training and testing on 40,000 sample data. I used 5 epochs for training the model. The model seems has an accuracy of 93% after only being trained on the 3.8% of data, the accuracy on test data is 93.14%.