Retweet Prediction

Arpit Agrawal

AGENDA

1.Problem Statement

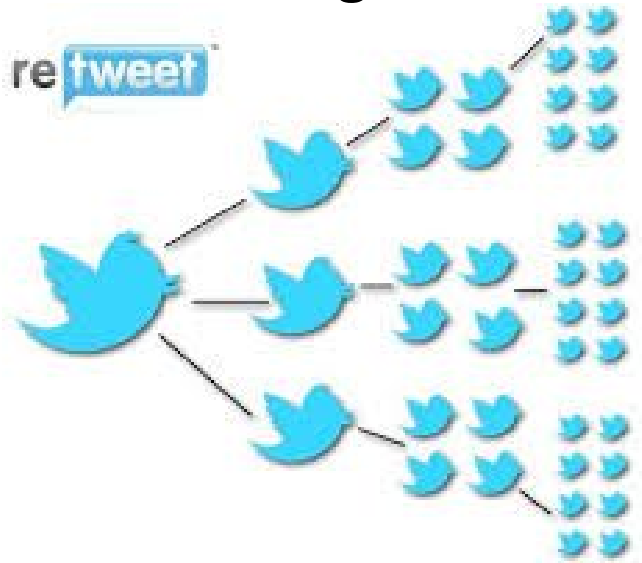
2.Dataset Description

3.Exploratory Data Analysis

4.Modelling and Results

PROBLEM STATEMENT

Case: Predict total number of Retweets a Tweet will get



Importance of Retweets: It extends the reach of a message

DATASET DESCRIPTION

Dimensions:

No of observations = 42368 No of variables = 42

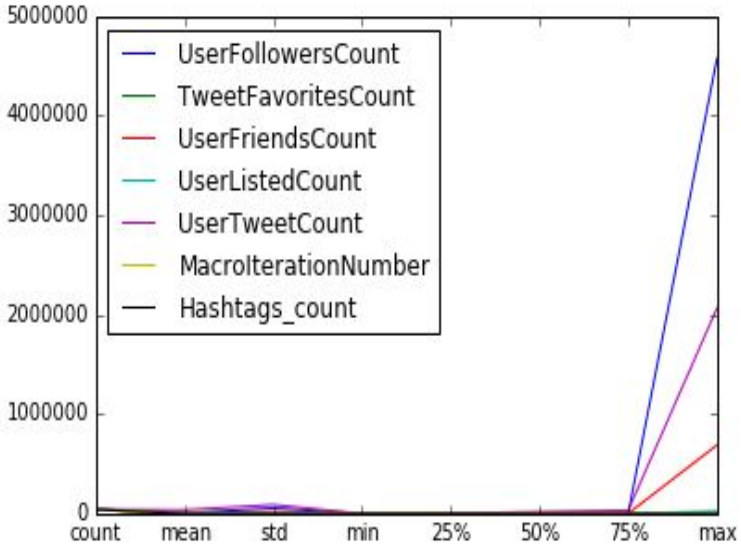
Dataset contains:

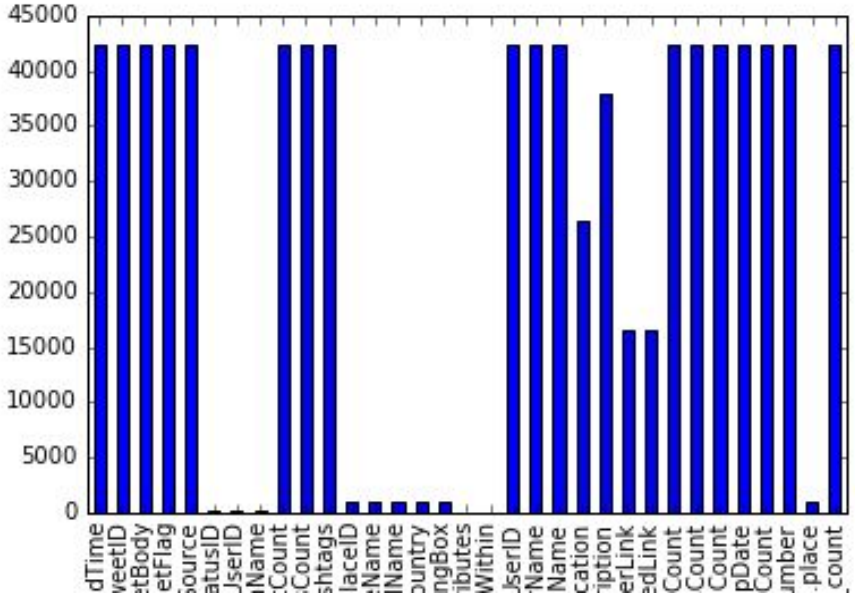
Detail about Tweet(time, place, id, source, retweet count, hashtags, location , source etc)

Details about the user(id,place,name,friends count, description, link, expanded link, tweet count etc)

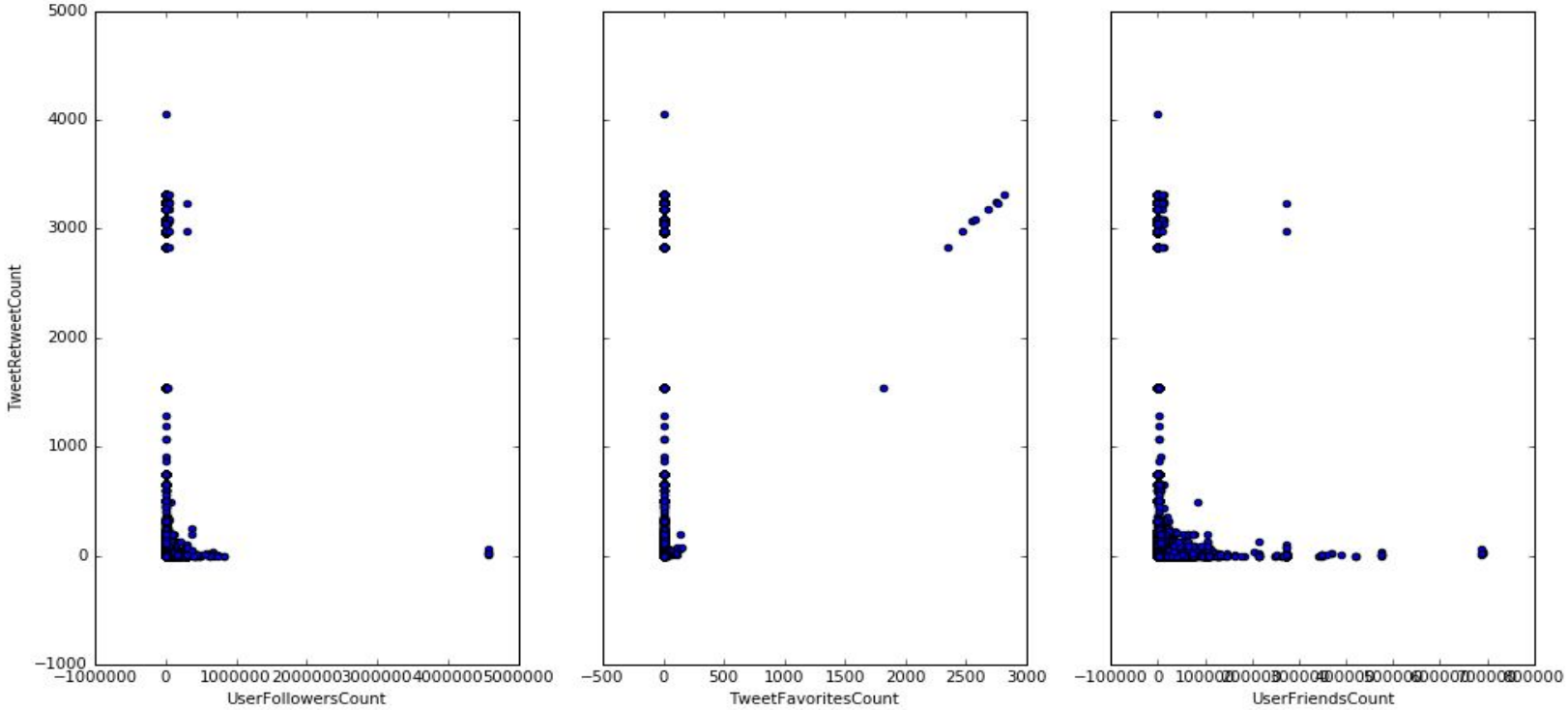
EXPLORATORY DATA ANALYSIS

Carried out univariate analysis for each variable (mean, median, count, min, max, null values)

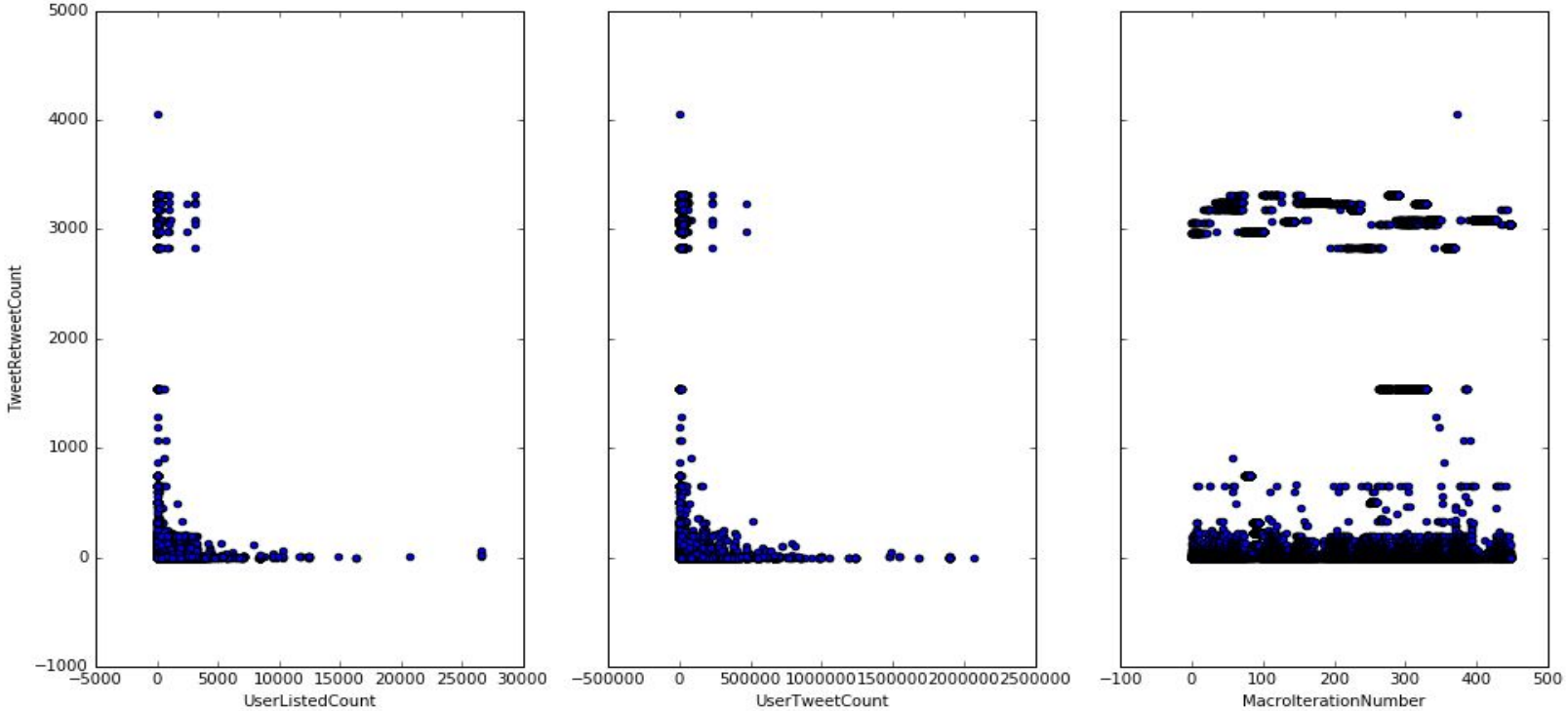




EXPLORATORY DATA ANALYSIS



EXPLORATORY DATA ANALYSIS



MODELLING AND RESULTS

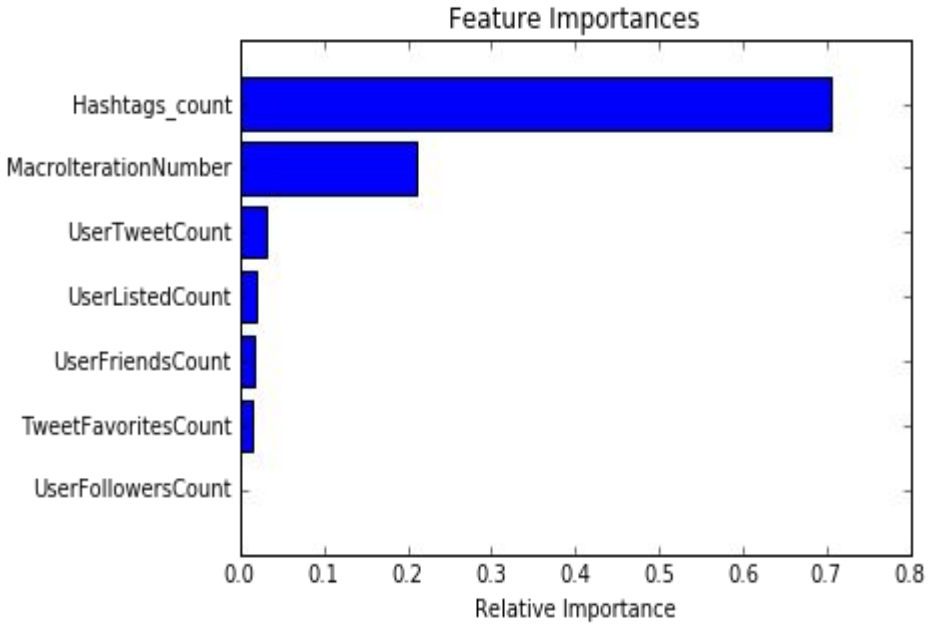
Used seven features to build the model (UserFollowersCount, TweetFavoritesCount, UserFriendsCount, UserListedCount, UserTweet Count, MacroIteration Number, Hashtags_count)

Model	RMSE(root mean square average)	Accuracy(%)
1. Linear Regression	1152	34.87
2. Random Forest	120	98.2

Hashtags_count was an additional feature which was added to both model but random forest showed better results

MODELLING AND RESULTS

This graph shows the relative importance of the features used for modelling



Thankyou!!!