Ethnicity prediction

Nirmal kumar Ravi

Can we predict ethnicity of twitter users using popular twitter accounts?

Method

In this project we use popular twitter accounts (politicians, movie stars, sports stars) to predict ethnicity of twitter users
Step 1: Get the most popular twitter accounts which people belong to same ethnicity follows
Step 2: Label the followers with that ethnicity
Step 3: Get friends of the followers
Step 4: Construct a Boolean matrix with users and accounts they follow
Step 5: find classification boundaries and make predictions

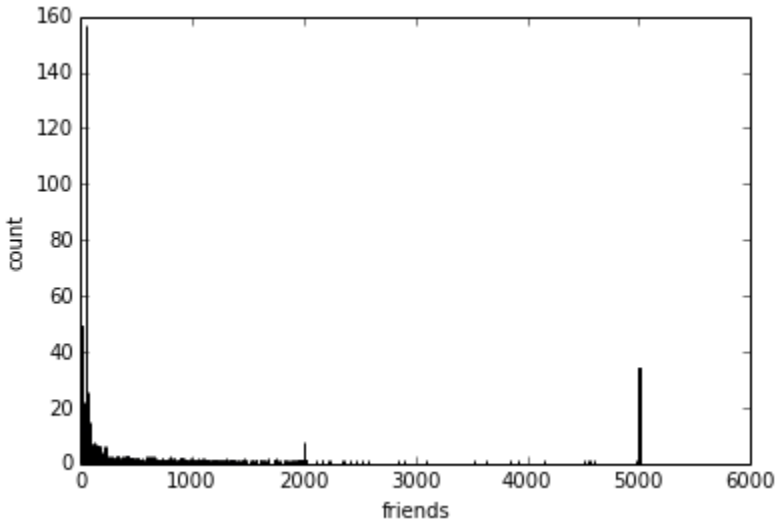
Ethnicity predicted

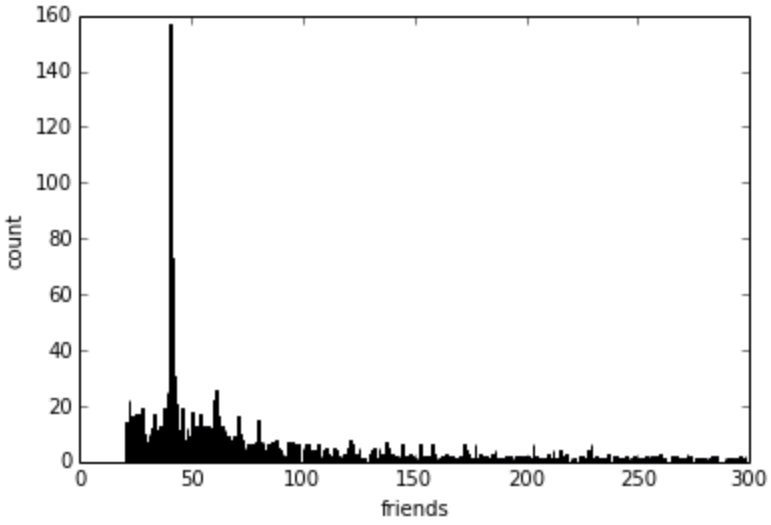
- Indian
- Canadian
- Brazilian

Prime Accounts (Popular Accounts)

Twitter Id	screen name	Description	# of followers
18839785	narendramodi	Narendra Modi prime minister of India	300
135421739	sachin_rt	cricket player from India	300
132385468	BeingSalmanKhan	salman khan actor from India	300
131574396	dilmabr	Dilma Rousseff politician from Brazil	300
158487331	neymarjr	Neymar Jr soccer player from Brazil	300
150697226	aliceb_oficial	Alice Braga actress from Brazil	300
7401202	Stephen Harper	Stephen Harper prime minister of canada	300
581437101	dedication9trey	foot ball player from canada	300
731379950	David_Acer	David Acer comedian canada	300

Data Visualization




160
140
120
100
80
60
40
20
0
count
0
50
100
150
200
250
300
friends

I used **logistic regression** to fit the data

Accounts with top coefficients

Indian

SrBachchan
aamir_khan
PMOIndia

Canadian

CBCNews
rickmercer
CP24

Brazilian

DavidLuiz_4
Cristiano
KAKA

Accuracy on training set

93%

Accuracy on test set

78%

Questions