

# 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

	Acct 1	Acct 2	Acct 3	Acct 4	Acct 5	Acct 6	Acct 7	Class
User 1	1	1	1	0	1	0	0	C1
User 2	1	1	0	0	0	0	0	C1
User 3	1	1	0	0	1	1	0	C1
User 4	1	0	0	0	0	0	1	C1
User 5	1	1	1	0	0	0	0	C1
User 6	0	1	0	1	1	1	0	C2
User 7	0	0	1	1	0	1	0	C2
User 8	0	0	0	1	1	1	1	C2
User 9	0	0	1	1	1	1	0	C2
User 10	0	0	0	1	1	0	0	C2

## Data

Twitter

5000 followers \* 12 brands \* 5000 friends = 60000 requests =

41 days / 10 twitter accounts = **4.1 days**

## **Related work**

[Understanding the Demographics of Twitter Users](#) .

In this US census data has been used to predict the ethnicity. They have used names of twitter account to predict the ethnicity.

## **Bias Correction**

**Not everyone loves politics.**

We must choose popular twitter accounts carefully, to avoid bias to particular set of population.

## **Timeline**

October 22:	Proposal Presentation
October 29:	Data collection
November 5:	Method implementation
November 12:	Analysis of results
November 19:	Enhancing results
November 26:	Conclusions and reviews
December 3:	Presentation



### **Future work**

Using **hash tags** with combination of popular twitter accounts and make our predictions more accurate

**Questions?**