

# Inferring Political Opinion from Social Media Data

UG4 Project Meeting 1

Angus Scott

# Project Task

- Build classifier(s) to predict whether a Twitter user was more likely to be a Yes or No supporter in the Scottish Independence Referendum
- Using data obtained from both the Edinburgh Twitter Corpus and my own mining
- Use classification techniques as described in related papers (e.g. Sentiment Analysis, Clustering, Graph/Network Algorithms) and then adapt/improve

# Current Progress

- Annotated sample of data for early experiments, continuing to annotate data a little at a time.
- 300 Tweets tagged, 114 Yes, 44 No, 142 Undeterminable
- 38% of Yes tweets contain ‘#voteyes’, 0% of No tweets contain ‘#voteyes’
- Collection of keywords largely identify the political leaning, but no single one.

whichever  
nationalism  
yessers worked  
amp  
please  
memoriams  
hoping offers christians stay great sanity alexsalmond  
total wake win delighted historicday tories/@david magic  
scared voting voting  
freedom friends  
whatever  
country  
@mandybs  
never facts follow  
destroyed aye  
anachronism @barbjonn @serb unitedkingdom  
terrible thing  
lies responding yes  
scotlandindependence  
must @uk nationality  
@uk  
@pjpaton  
economy  
tomorrow  
positivity  
think http://t.co/ctdtgqtcep  
fight better  
breath sketch  
obama foreign sturgeon climbing  
decided  
sentimental http://t.co/bmxeglpqne  
thanks division  
united

http  
2.5  
c'mon  
like  
sense  
scottish  
salmond  
forever  
worldwide business accused @conservativeccch  
blocking noscotland leave @gerrymcgarvey  
xenophobia  
remember see  
concessions lying  
bile lad years  
merrier wonderful  
18 keep eu  
vs  
good  
negative back https://t.co/5gzwxr4glj  
d11 brushes close oconnell  
@nicholaswatt



# Future plans

- Find Keywords using Statistical Synonyms/Topic Models (began this process already using Dice's coefficient and EMIM, plan to use LDA)
- Tweets lack context, investigate urls posted with tweets
- Investigate when people went public with their opinion, or if anyone switched/why they switched

# Questions?