Toxic Comment Classification

By Minh Ngoc Pham
Data Science Career Track

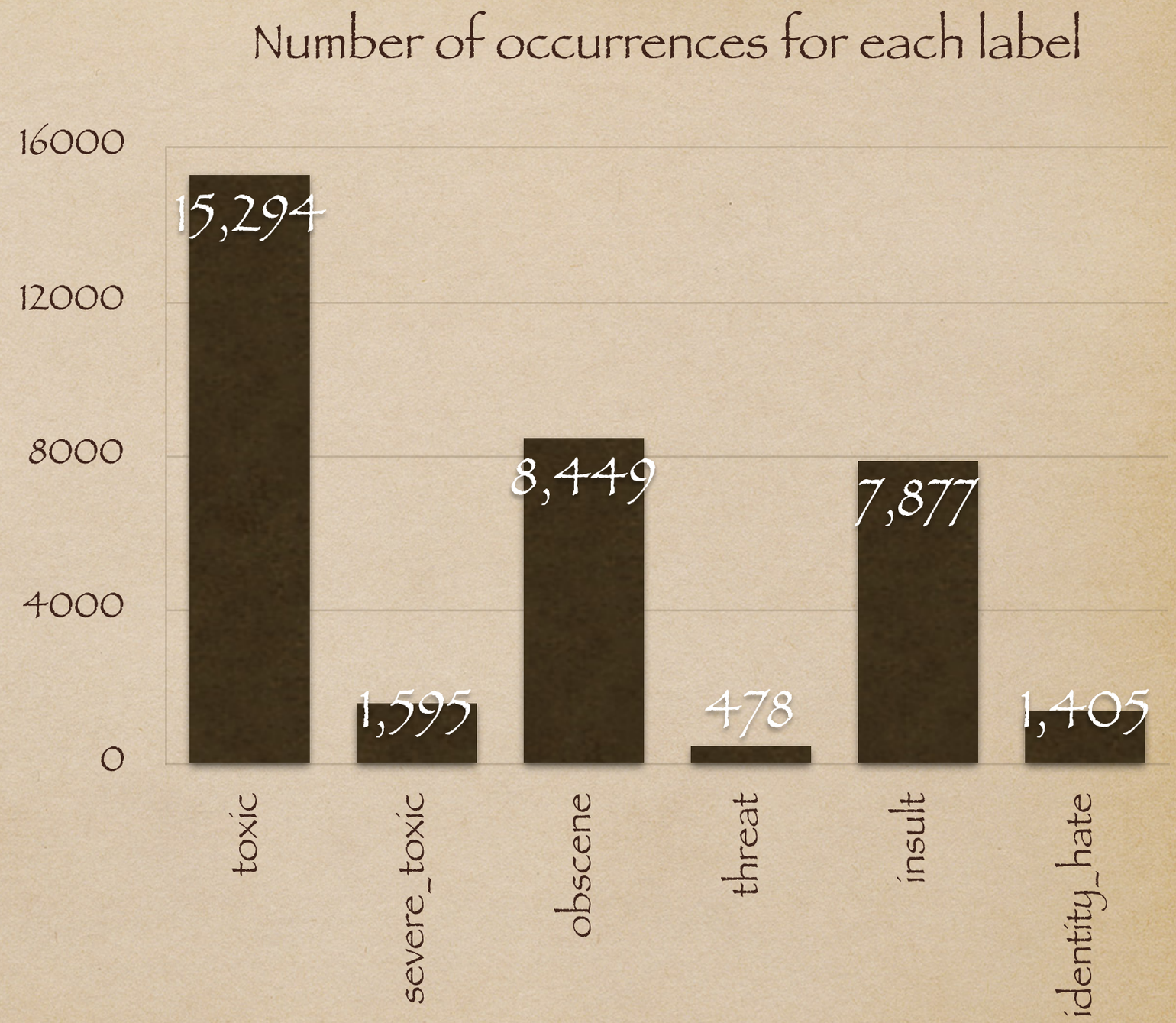
Project Introduction

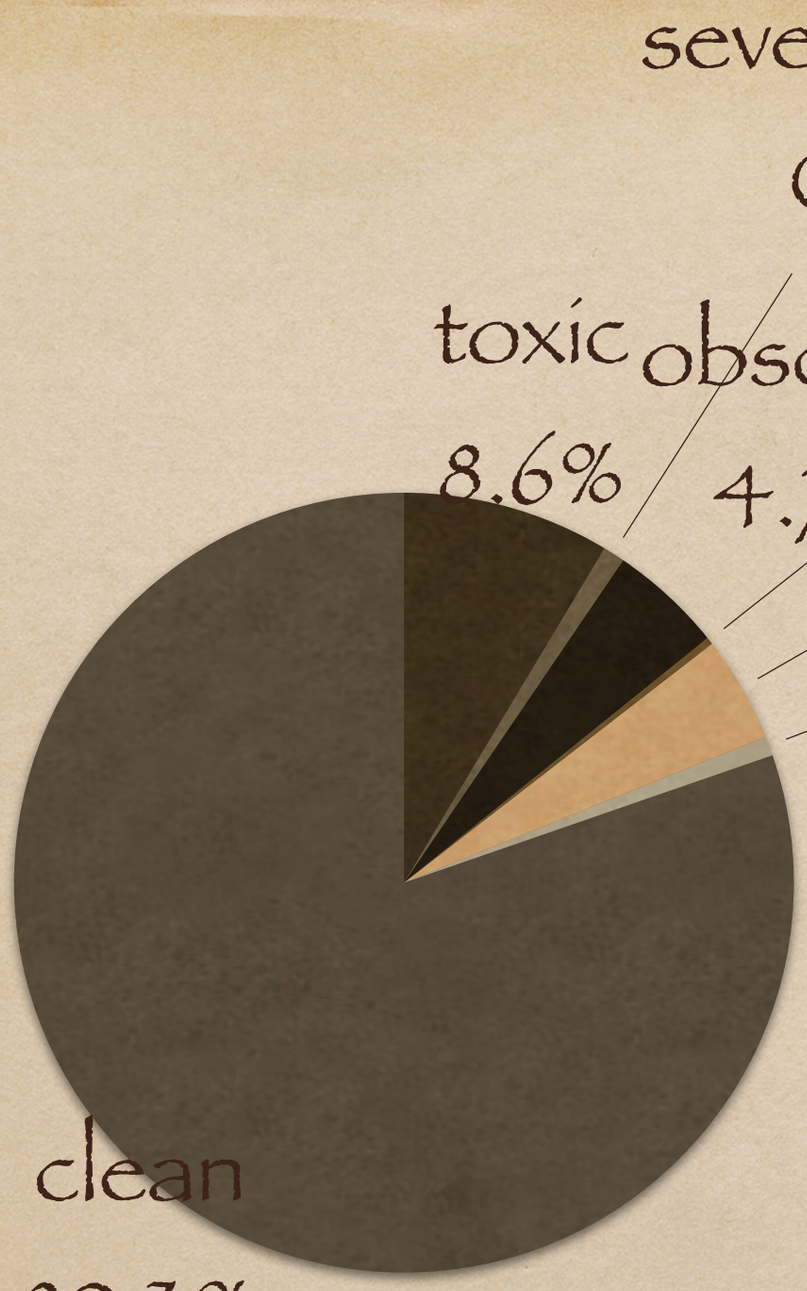
- Toxic Comment Classification Challenge
- Published in 2018

kaggle

- 159,571 comments in training set
- 153,164 comments in test set

- Highest count in toxic comments
- There are a lot of clean comments

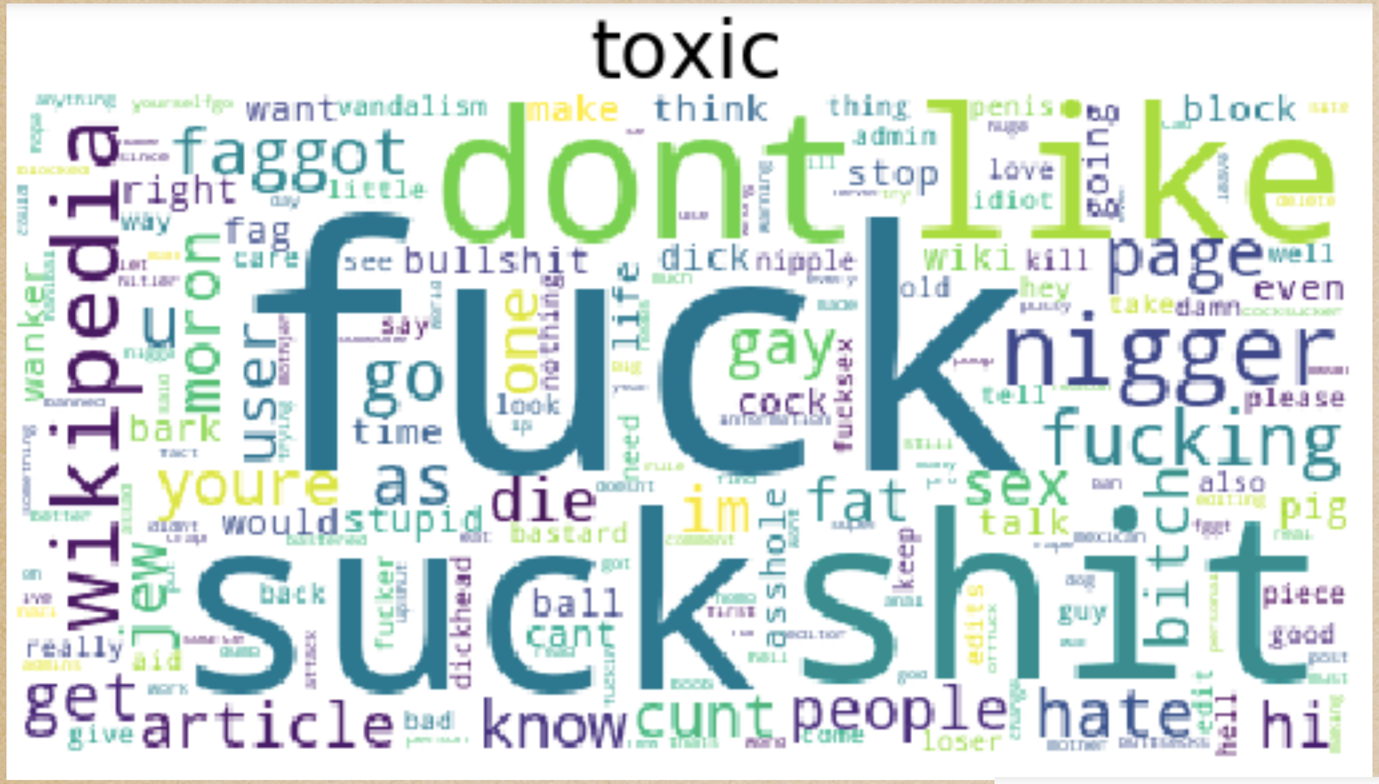


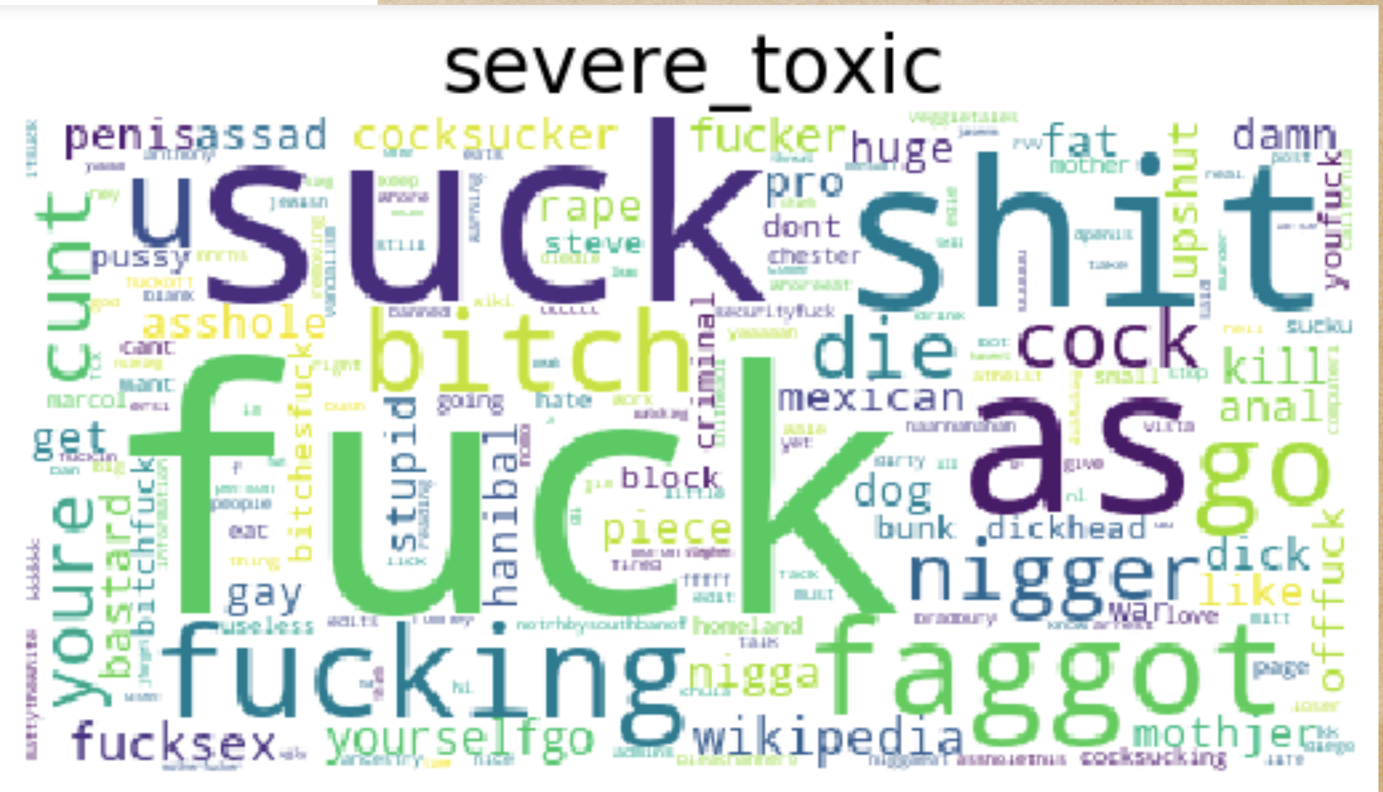


- 80.3% of the comments in the training test are clean
- Class imbalance issue

- High number of comment with 0 label
- There are only 31 labels classified as all of the 6 labels

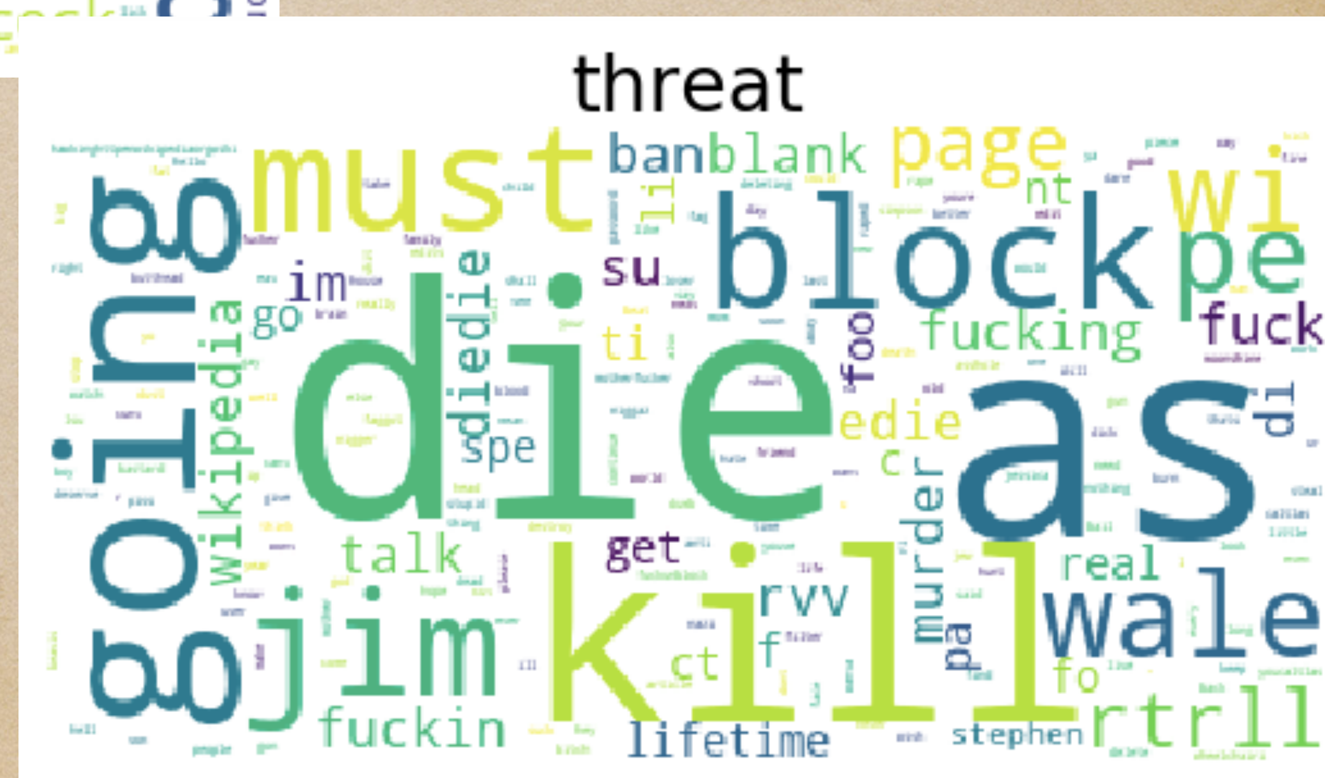
Word Cloud


toxic
fuck
suck
shit
dont
like
wikipedia
nigger
fucking
faggot
page
youre
as
die
bitch
hate
people
cunt
know
article
get
hi
moron
user
go
gay
cock
sex
fat
im
jew
stupid
would
ball
bullshit
dick
nipple
block
think
want
vandalism
make
thing
penis


severe_toxic
fuck
suck
shit
fucking
faggot
bitch
as
go
cunt
youre
nigger
cock
die
kill
anal
asshole
penis
assad
cocksucker
fucker
huge
fat
damn
pro
dont
rape
steve
pussy
upshut
youfuck
mexican
stupid
hanibal
bastard
bitchfuck
bitchesfuck
criminal
piece
dog
dick
dickhead
like
bunk
offfuck
nigga
homeland
wikipedia
yourselfgo
fucksex
mothjer
gay
block
get
page
love
war

Word Cloud


obscene
fuck
shit
suck
faggot
nigger
fucking
bitch
cunt
go
u
as
dont
wikipedia
like
page
bullshit
asshole
dick
user
fat
die
article
hate
stupid
youre
block
bastard
nipple
dickhead
im
get
know
people
penis
cocksucker
buttsecks
idiot
fucker
wiki
think
piece
life
pussy
talk
hanibal
nothingboob
jew
love
kill
guy
well
dog
said
one
gay
edits
stop
want
going
thing
see
say
little
mothjer
bastered
fuck
nigga
rape
hey
give
poop
even
big
fag
damn
right
would
make
edit
come
eat
time
back
huge
cant
mexican
take
yourselfgo


threat
die
as
kill
going
go
must
block
jim
wale
rrtrll
ban
blank
page
wi
pe
fuck
fucking
wikipedia
diedie
im
su
ti
edie
murder
talk
get
real
lifetime
fuckin
stephen
rvv
ct
spe
foo

Word Cloud


insult
fuck
suck
nigger
fucking
faggot
fat
shit
bitch
go
as
cunt
hate
like
dont
youre
moron
jew
get
u
hi
cocksucker
bastard
asshole
loser
stupid
wikipedia
fucker
dickhead
penis
pussy
cock
nigga
idiot
block
mexican


identity_hate
nigger
jew
fat
gay
fuck
faggot
die
fucking
nigga
huge
suck
shit
stupid
like
bitch
cunt
mexican
hate
jewish
steve
bunk
ancestry
people
licker
tommy
fuckoff
kill
spanish
get
dont
go
as
u