

TEXT MESSAGE NLP ANALYSIS

Snizhana Kurylyuk



SPAM TEXT MESSAGES

For your chance to WIN a
FREE Bluetooth Headset
then simply reply back
with "ADP"

500 free text msgs. Just
text ok to 80488 and we'll
credit your account

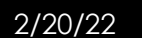
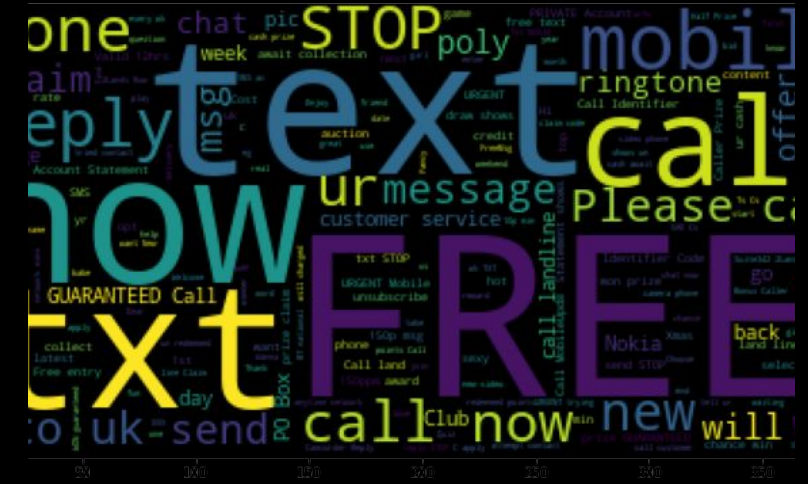
You have 1 new
voicemail. Please call
08719181513.

Text BANNEDUK to 89555
to see! cost 150p
textoperator g696ga 18+
XXX

Are you unique enough?
Find out from 30th August.
www.areyouunique.co.uk

Refused a loan? Secured
or Unsecured? Can't get
credit? Call free now
0800 195 6669 or text
back 'help' & we will!

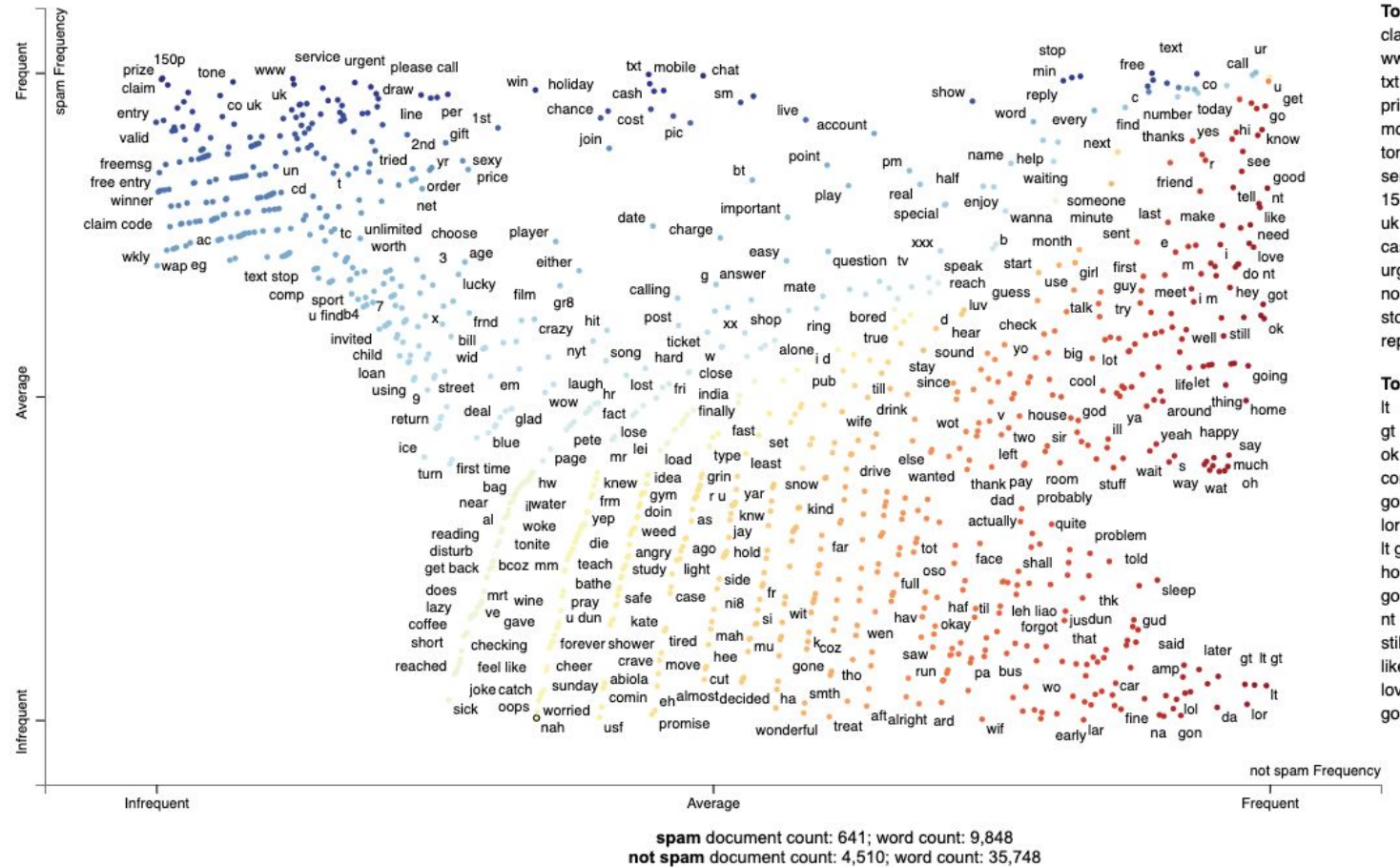
Get a FREE mobile video
player FREE movie. To
collect text GO to 89105.
Its free! Extra films can be
ordered t's and c's apply.
18 yrs only

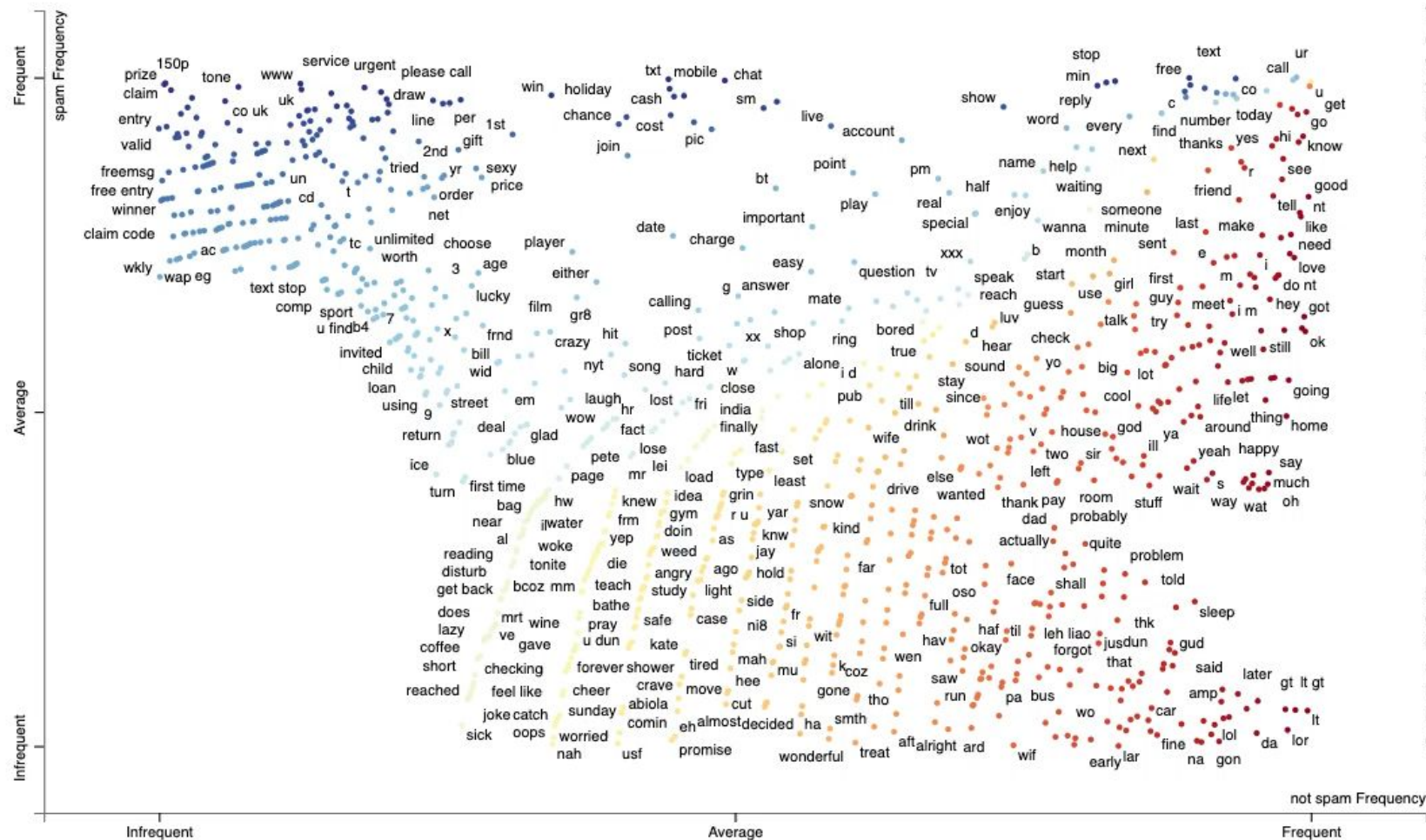


TOPIC MODELING

Topic 0	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9
<ul style="list-style-type: none"> •Gt •Lt •It gt •gt min •It gt min •Min •like It gt •like It •gt It •It gt It 	<ul style="list-style-type: none"> •Know •Let •Don't •let know •Want •Need •dont know •Sure •Pls •guy 	<ul style="list-style-type: none"> •Ok •Lor •ok lor •Thanx •Said •Ask •Leave •Yup •Ya •ur 	<ul style="list-style-type: none"> •Come •want •Home •want come •Tomorrow •come home •come tomorrow •Pick •Tell •online 	<ul style="list-style-type: none"> •Free •Ur •Text •Send •Stop •Txt •Reply •Message •phone •mobile 	<ul style="list-style-type: none"> •Got •Da •Sorry •Yeah •Wat •Later •Say •Oh •Work •hey 	<ul style="list-style-type: none"> •Good •Day •Love •Hope •Great •Morning •Happy •Dear •Night •babe 	<ul style="list-style-type: none"> •Like •Tell •Don't •Feel •Think •like It gt •like It •feel like •lol •yeah 	<ul style="list-style-type: none"> •Going •Home •Lor •Way •Dinner •going home •work •lunch •sleep •today 	<ul style="list-style-type: none"> •Time •Wat •Want •wat time •tell •make •week •long •soon •im

SCATTER TEXT





spam document count: 641; word count: 9,848
not spam document count: 4,510; word count: 35,748

Term: free

spam frequency:
583 per 25,000 terms
225 per 1,000 docs
Some of the 190 mentions:

not spam frequency:
53 per 25,000 terms
12 per 1,000 docs
Some of the 56 mentions:

Top spam Characteristic

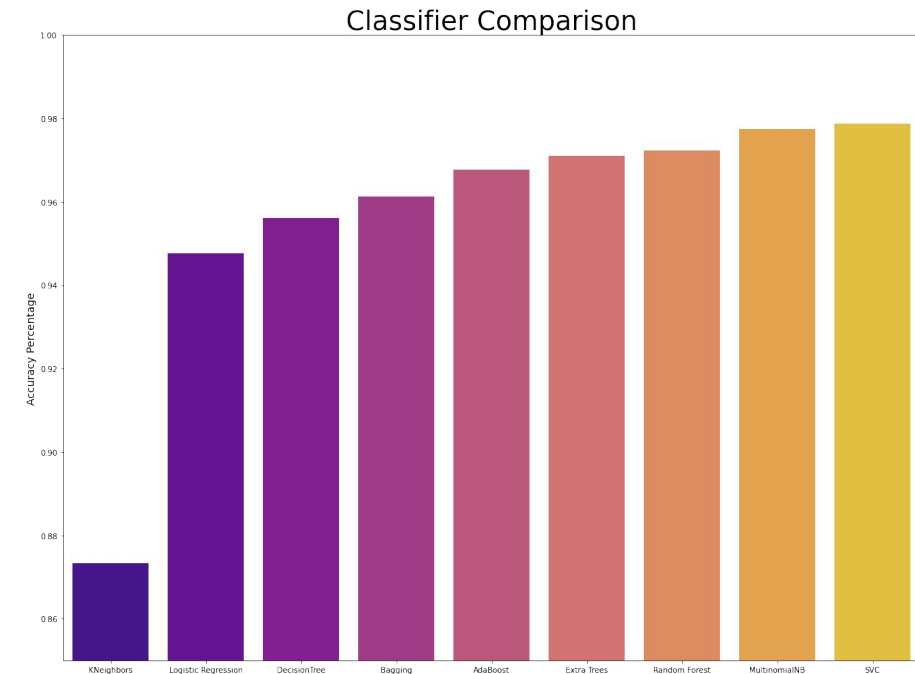
claim	chikku
www	boytoy
txt	oredi
prize	freemsg
mobile	frnd
tone	lor
service	frnds
150p	rakhesh
uk	thk
cash	custcare
urgent	savamob
nokia	aight
stop	txts
reply	goodmorning

Top not spam

It	knw
gt	mayb
ok	juz
come	abiola
got	bcoz
lor	haf
lt gt	muz
home	tmr
going	leh
nt	liao
still	gon
like	tomo
love	wkly
good	pls

INITIAL CLASSIFICATION MODELS

Classifier	Accuracy_Score
KNeighbors	0.873385
Logistic Regression	0.947674
DecisionTree	0.956072
Bagging	0.961240
AdaBoost	0.967700
Extra Trees	0.970930
Random Forest	0.972222
MultinomialNB	0.977390
SVC	0.978682



HYPERPARAMETER TUNING

Random Forest Model

Accuracy: 0.97
Precision: 0.93
Recall: 0.84
F1: 0.89

Multinomial NB Model

Accuracy: 0.98
Precision: 0.94
Recall: 0.88
F1: 0.91

SVC Model

Accuracy: 0.98
Precision: 0.98
Recall: 0.86
F1: 0.91



Streamlit

Text Message Classifier

Enter a message for spam evaluation

Made with Streamlit



QUESTIONS