Angry Tweets

Predict the sentiment of microblog posts

by Evgeny Melnikov

HattoryChan

Introduction

Финал проекта

Анализ и выбор инструментов

Подготовка данных

Chapter Four
The Spirits of Ashenvale

Two days later, along the borderlands of Ashenvale forest...

Постановка задачи

Выбор темы проекта

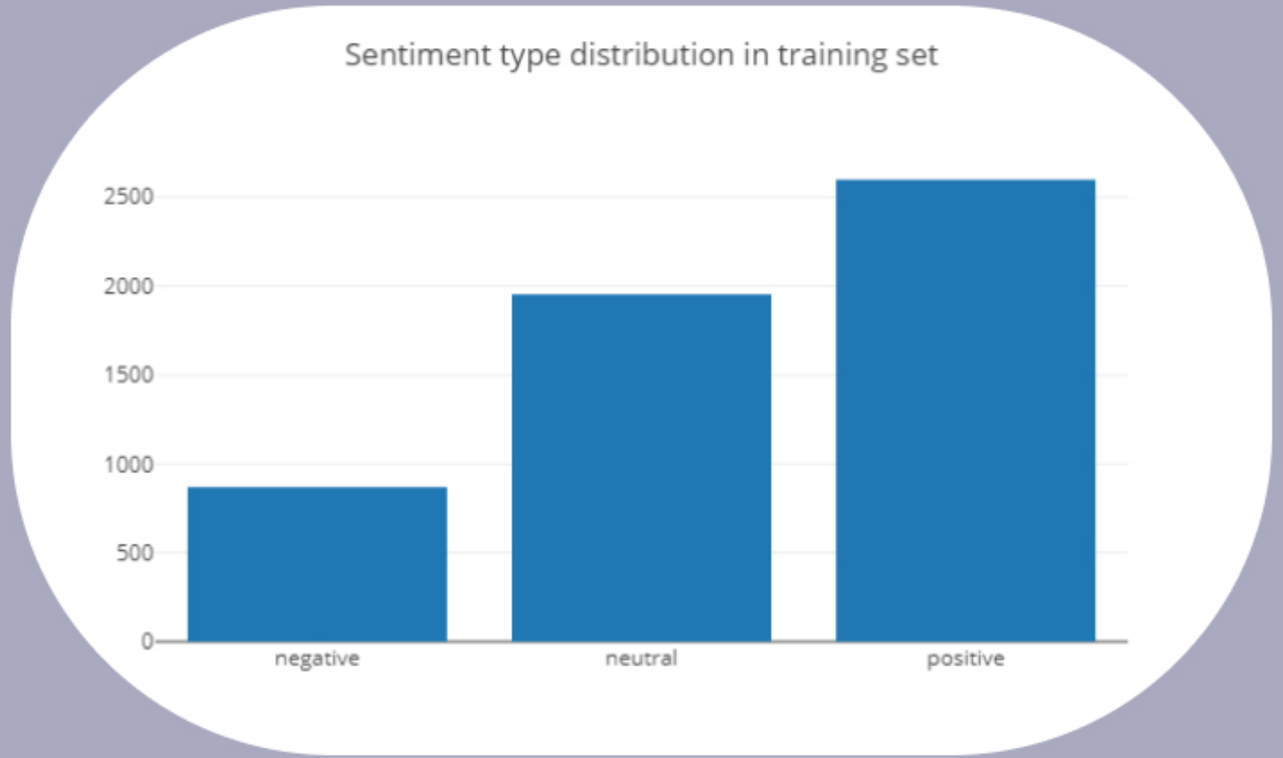
PRESS ANY KEY TO CONTINUE

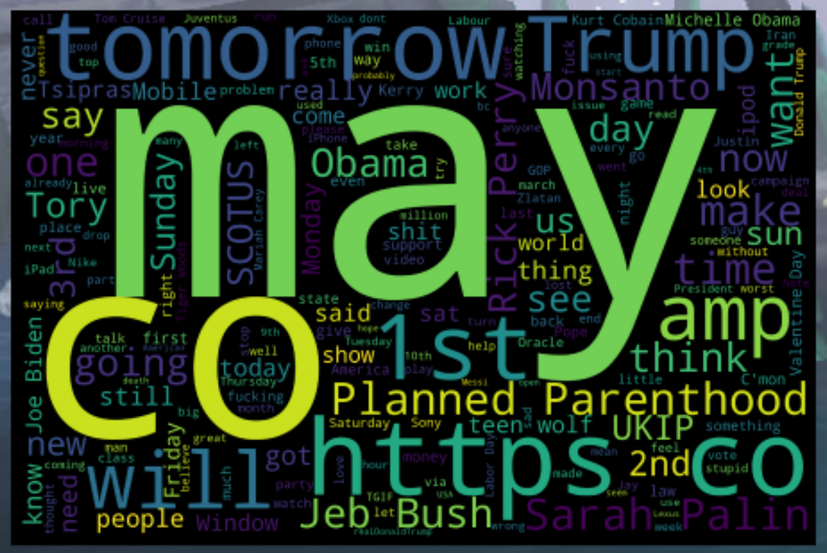
Data preprocessing

train.csv (5971 tweets)
test.csv (4000 tweets)

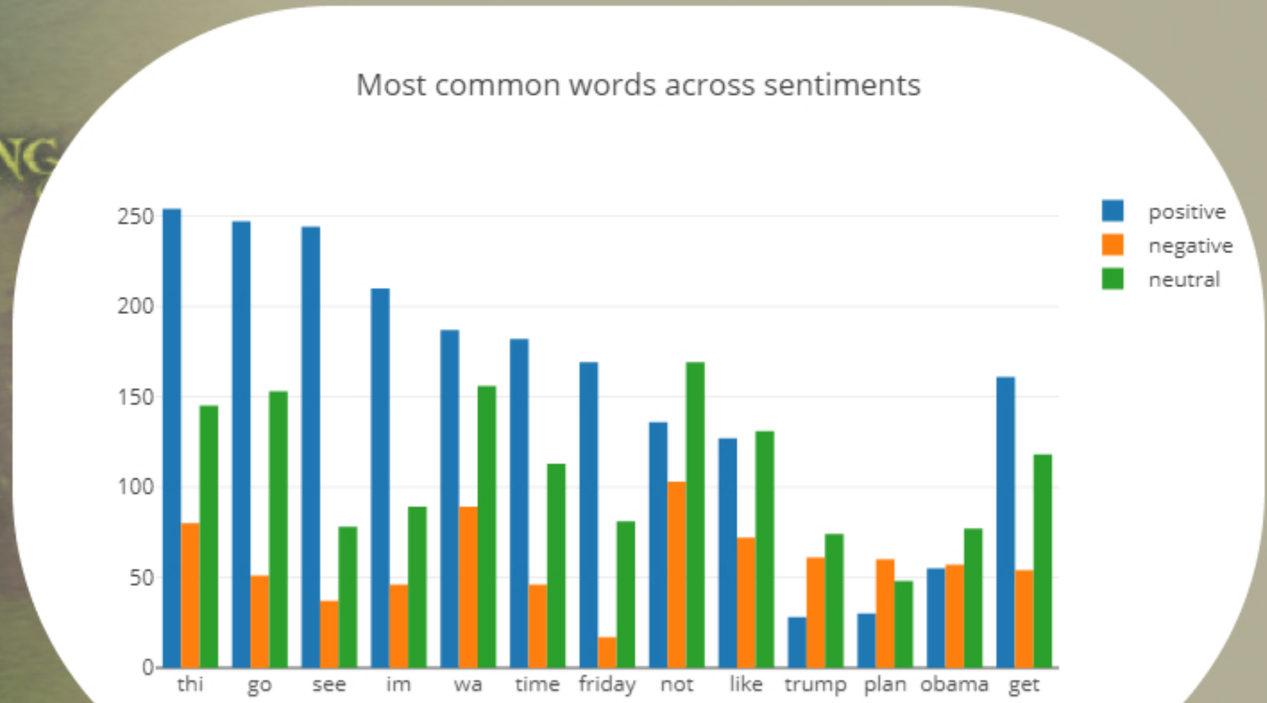
#CrossSkyHigh is
going IOS #saturday. For now try
this http://t.co/Jrc6ktK9rT
#indiedev @GamerRTer
@ShoutGamers
http://t.co/bkWUjKEmju

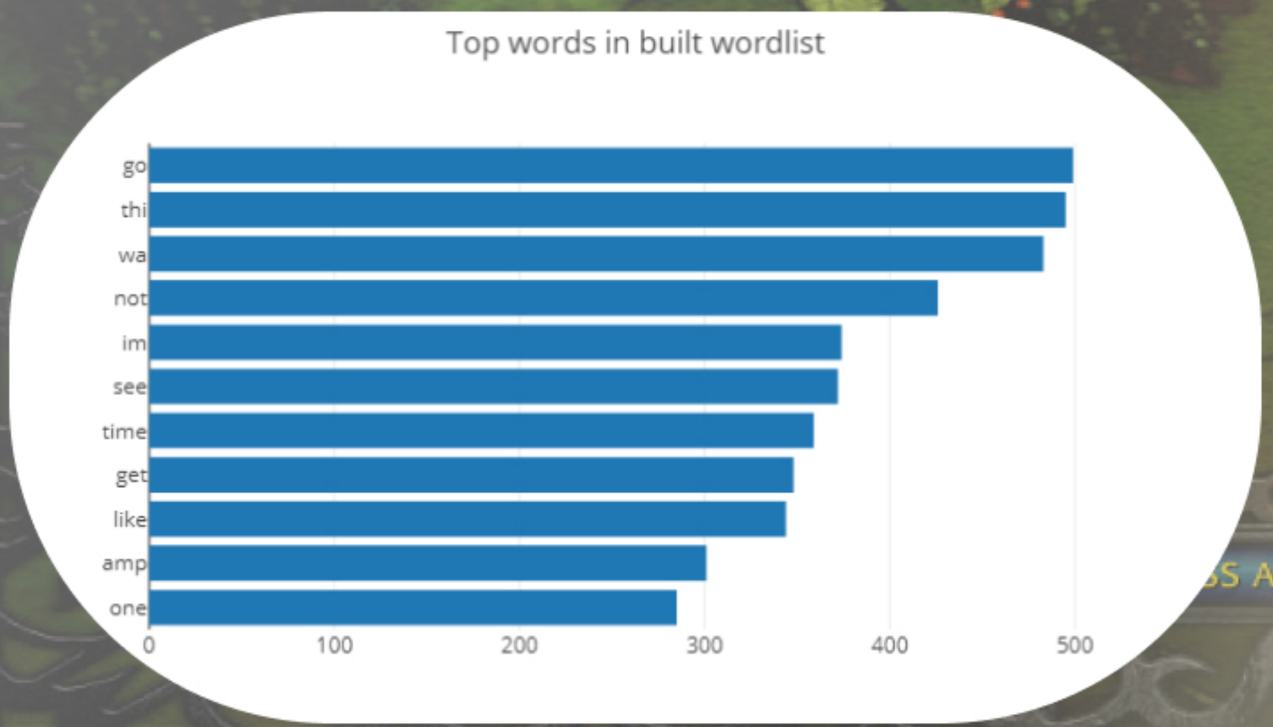
Data preprocessing

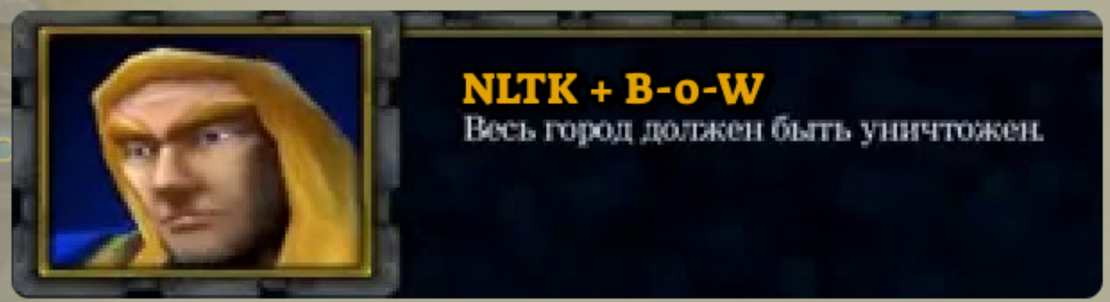




Cleaning and Building

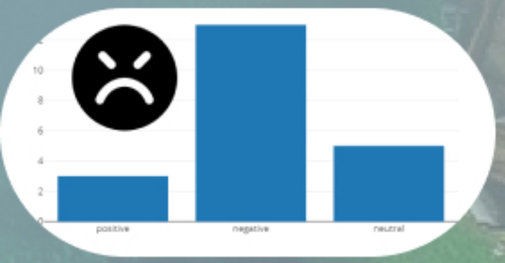


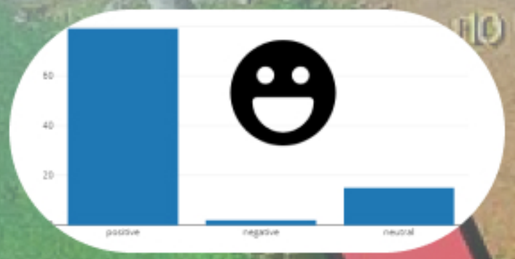


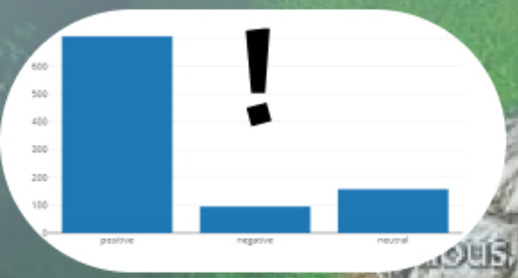


Classification

Additional features







Bag-of-Words + Naive Bayes

Results

	Negative	Neutral	Positive
F1	[0.38949672	0.45214221	0.71411765]
Precision	[0.45408163	0.4853229	0.65978261]
Recall	[0.34099617	0.42320819	0.77820513]

Accuracy 0.5802089735709896

Average accuracy: 0.4324276684

Extended features + Random Forest

Accuracy: [0.54786451 0.49557522 0.38200
0.51698671 0.55473373]
Average accuracy: 0.4582582267978166

СЛОЖНА

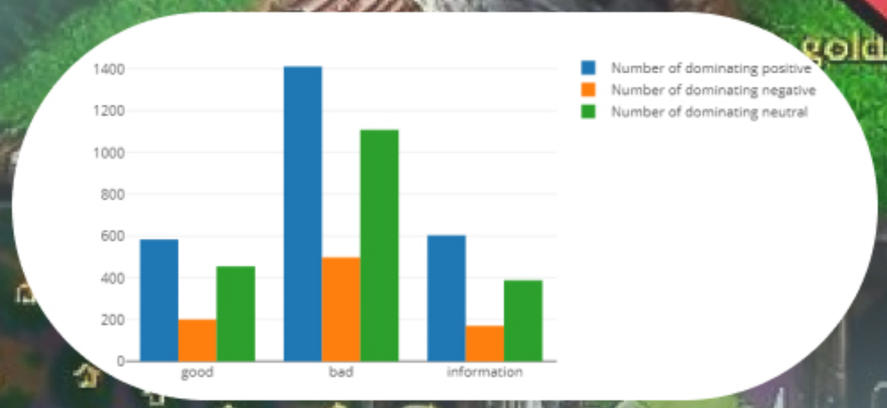
Classification

Word2Vec

GoogleNews
-vectors-
negative300

Final data model

extra text features (number of: !, ?, :-) etc)
word2vec similarity to key words
word2vec 200 dimension averaged representation
bag-of-word representation of a tweet



Extended features + XGBoost

Results

	Negative	Neutral	Positive
F1	[0.16716418	0.41090555	0.69450317]
Precision	[0.37837838	0.47845805	0.59082734]
Recall	[0.10727969	0.36006826	0.84230769]

Accuracy 0.5507068223724647

Average accuracy: 0.559825169

Finding best parameters for XGBoost

Cross validation score: 0.55810

ОЧЕНЬ

Final of final project

Full model + Random Forest

```
                Results
          Negative     Neutral     Positive
F1       [0.02973978 0.4581749   0.72633213]
Precision[0.5        0.51716738 0.60884649]
Recall   [0.01532567 0.4112628  0.9        ]
Accuracy 0.5820528580208973

Average accuracy: 0.5485980068408269
```

Full model + XGBoost

```
                Results
          Negative     Neutral     Positive
F1       [0.28808864 0.51471825 0.72065728]
Precision[0.52       0.50746269 0.66450216]
Recall   [0.19923372 0.5221843  0.78717949]
Accuracy 0.5974185617701291

Average accuracy: 0.51745273
```

Finding best parameters for XGBoost

```
Accuracy 0.5974185617701291
```