Quantifying Hype

Predicting Crypto Trading Behavior

MACHINE
LEARNING
TOKYO

Suzana Ilic | machinelearningtokyo@gmail.com

media

social media

price

social media

Twitter

1400 annotated examples

spam, pos, neg, hype, doom

counts	word
1490	bitcoin
978	rt
637	blockchain
625	crypto
617	cryptocurrency
517	ethereum
473	ico
402	btc
332	airdrop
322	eth

	Timestamp	Username	Label	Tweet	URL
0	March 14, 2018 at 09:47PM	@RizkyF_404	spam	RT @kickcity_io: Token Kart added KickCity to ...	http://twitter.com/RizkyF_404/status/973903509...
1	March 14, 2018 at 09:47PM	@exchangebutler	spam	RT @Denaro_io: Our support team is here to hel...	http://twitter.com/exchangebutler/status/97390...
2	March 14, 2018 at 09:47PM	@Bilalbinsaqib	pos	$eth trendline broke\nnext support @600 and th...	http://twitter.com/Bilalbinsaqib/status/973903...
3	March 14, 2018 at 09:47PM	@ahsen_soydan	spam	RT @ahsen_soydanvip: 💎 High Quality Partner ✔ ...	http://twitter.com/ahsen_soydan/status/9739035...
4	March 14, 2018 at 09:47PM	@mourneopened	hype	RT @izx_io: IZETEX is on the Battle of the Cry...	http://twitter.com/mourneopened/status/9739035...

Multi-class classification system with 5 classes

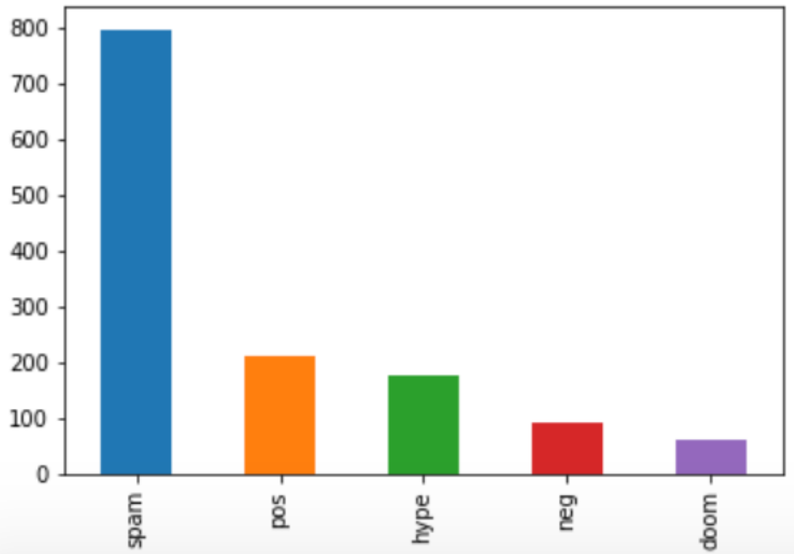
In [7]:
```
print(len(data.Label.unique()))
print(len(data))
```

```
5
1338
```

In [8]:
```
print(data.groupby('Label').size())
%matplotlib inline
import matplotlib as plt
data['Label'].value_counts().plot(kind="bar")
```

```
Label
doom      62
hype     177
neg       91
pos      211
spam     797
dtype: int64
```

Out[8]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1a1ea72a90>
```



Model	F1 score	Features/Pre-training/ Hyperparameters
SVM	60.2%	TF-IDF
	Best score: 0.622429906542056 1 Best C: 1000 Best Kernel: rbf Best Gamma: 0.001	
BiLSTM	62.3%	300-dimensional Glove embeddings Unknown vectors initialized as zeros
		Adam (lr=0.005) 1 Embedding 1 BiLSTM 2 Dropout Dense Activation function=softmax Loss=categorical_cross_entropy

Data	Bigger dataset 10x size of labeled data
Model	Stricter cleaning Transfer learning: Sentiment Analysis Different embeddings (e.g. Glove Twitter) Attention