

MOOD DETECTOR



in social media tweets

OSTRYCHARCZYK.KACPER@GMAIL.COM

MOTIVATION

- we spend more and more time in social media
- what we tweet reflects our mood and feelings
- every year 800.000 people die due depression suicide (WHO)

PROBLEM STATEMENT

- we aim to use artificial intelligence (AI) to check whether a tweet is characterized with a specific mood (good or low / bad)
- thanks to that, we can check how our tweet may be received by the others
- or what a tweet says about someones current feelings

DEPRESSION WORD BANK

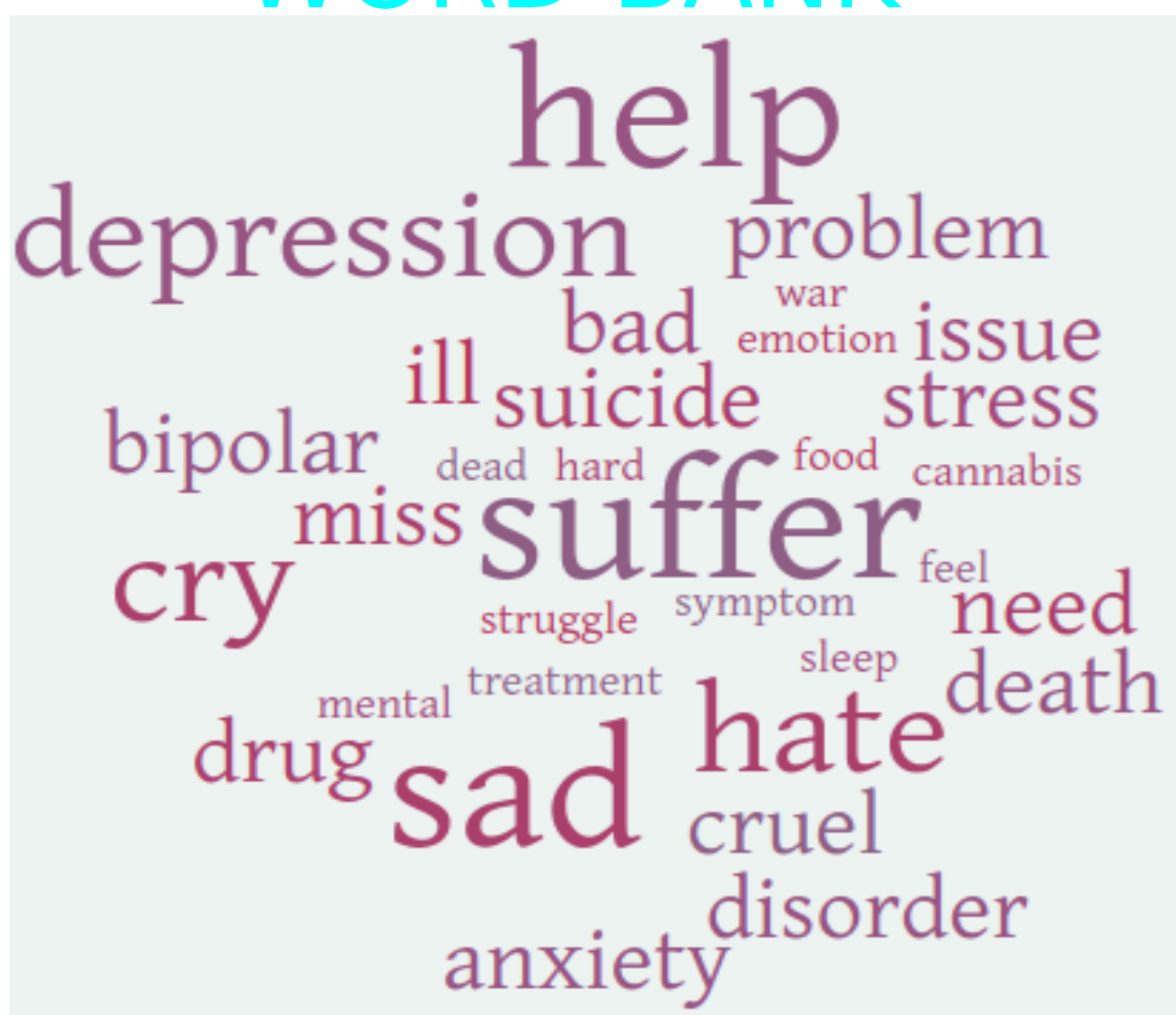


Figure 1: example of keywords for depressive mood

HAPPY MOOD WORD BANK



Figure 2: example of keywords matching good mood

DATA

TRAIN/TEST

1.6M labelled tweets (positive or negative)

language cleaning 'she's' to 'she is' etc.

filtering based on external word bank

EVALUATION

twitter stream based on popular #hashtags

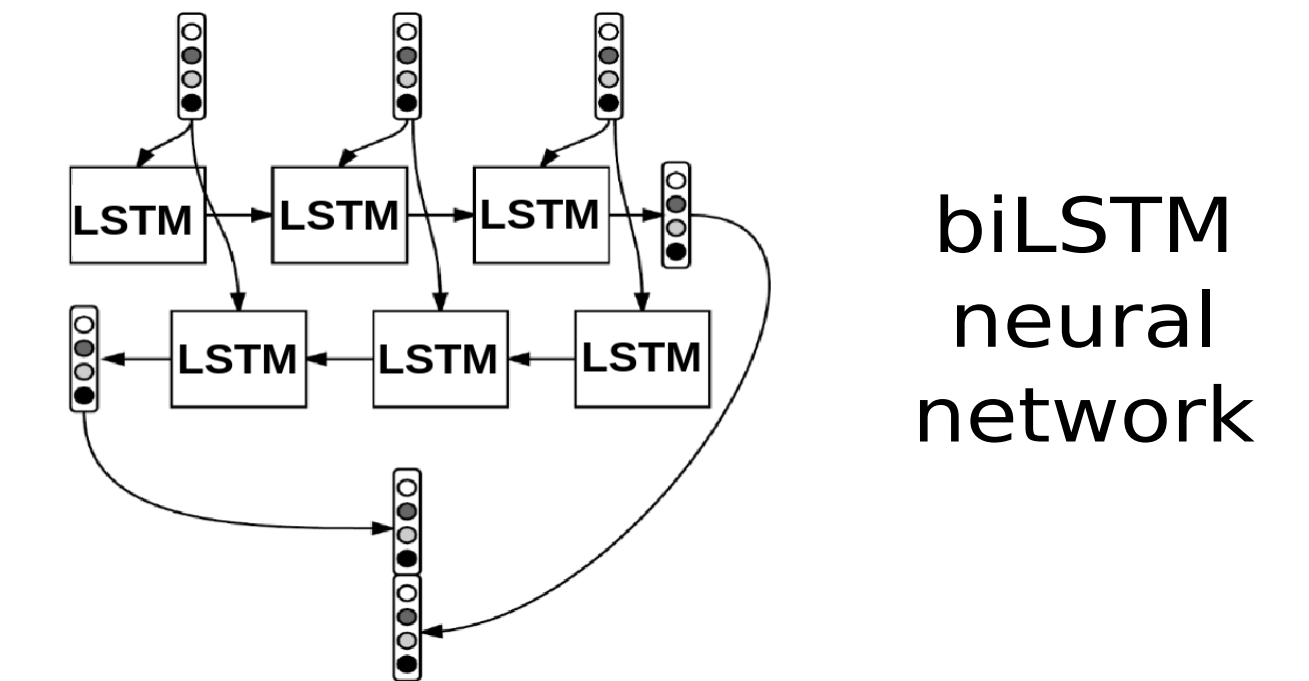
language cleaning

CLASSIFIER

Figure 3: ways of gathering and preprocessing data used for classification

MODEL

I → [] fasttext encoding
 feel → []
 #sick → []



model evaluation on test set with accuracy > 81%

	positive	depressive
positive	0.81	0.19
depressive	0.19	0.81

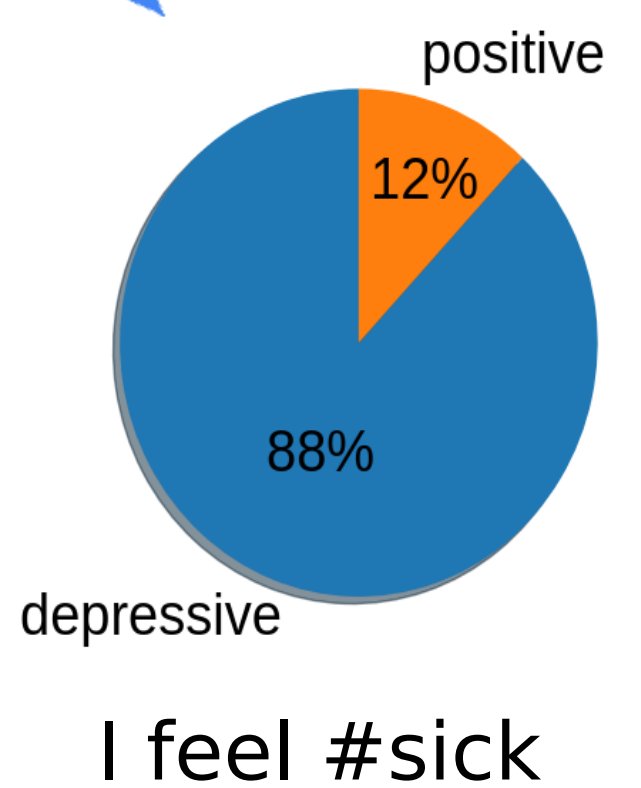


Figure 5: visual explanation of how model works: a tweet ('I feel #sick') is encoded and given to classifier which returns the probability of one of two classes - higher probability labels tweet as 'low mood'

RESULTS

Percentage of tweets with # for one of two classes

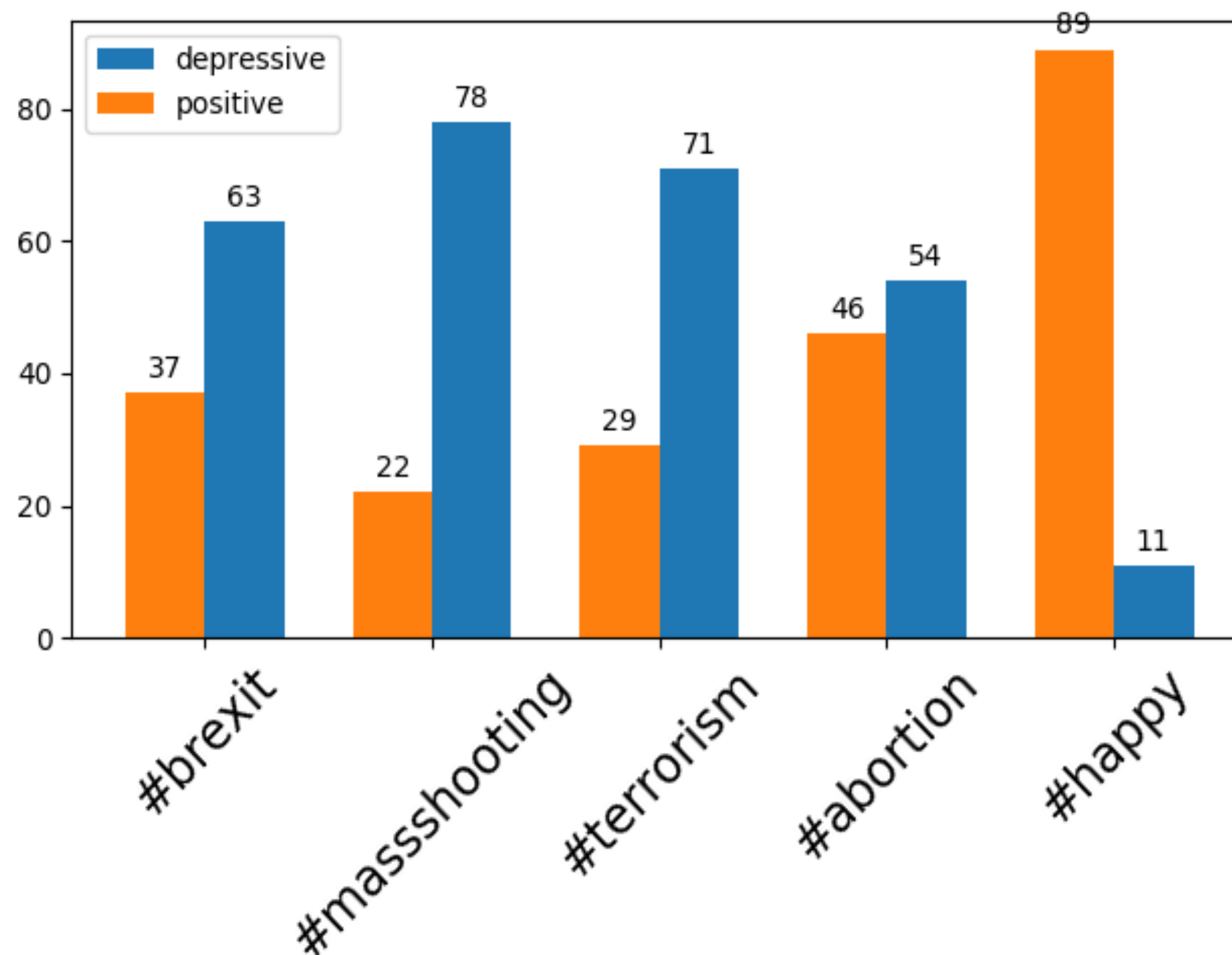


Figure 4: presentation of classifier results classifying popular tweets into one of two categories connected with mood - we can see that topics such as mass shooting or terrorism are negatively characterized, more controversial topics as brexit or abortion differs less in overall; #happy was added as a simple to classify showcase

SIDE EFFECTS

- given classifier (neural network) works well with keywords from word banks, but behaves poorly on neutral statements (returns similar probabilities)
- therefore it is difficult to classify many political statements since many times they are not emotional
- I decidet to reject tweets with less than 15% of probability difference in prediction

BUSINESS

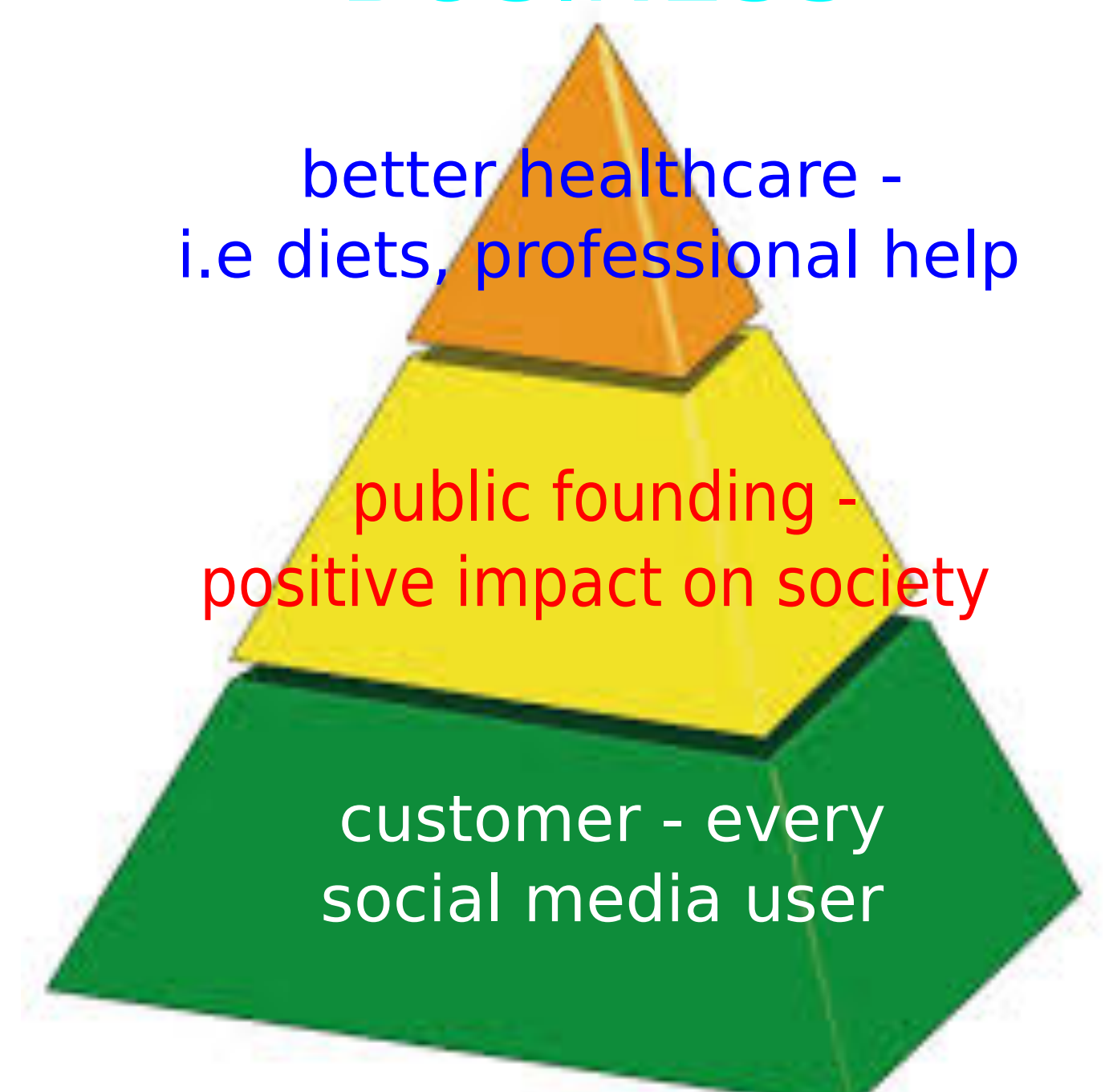


Figure 6: topics regarded monetization of the product issues