SENTIMENT ANALYSIS ON TWEETS IN A FINANCIAL DOMAIN
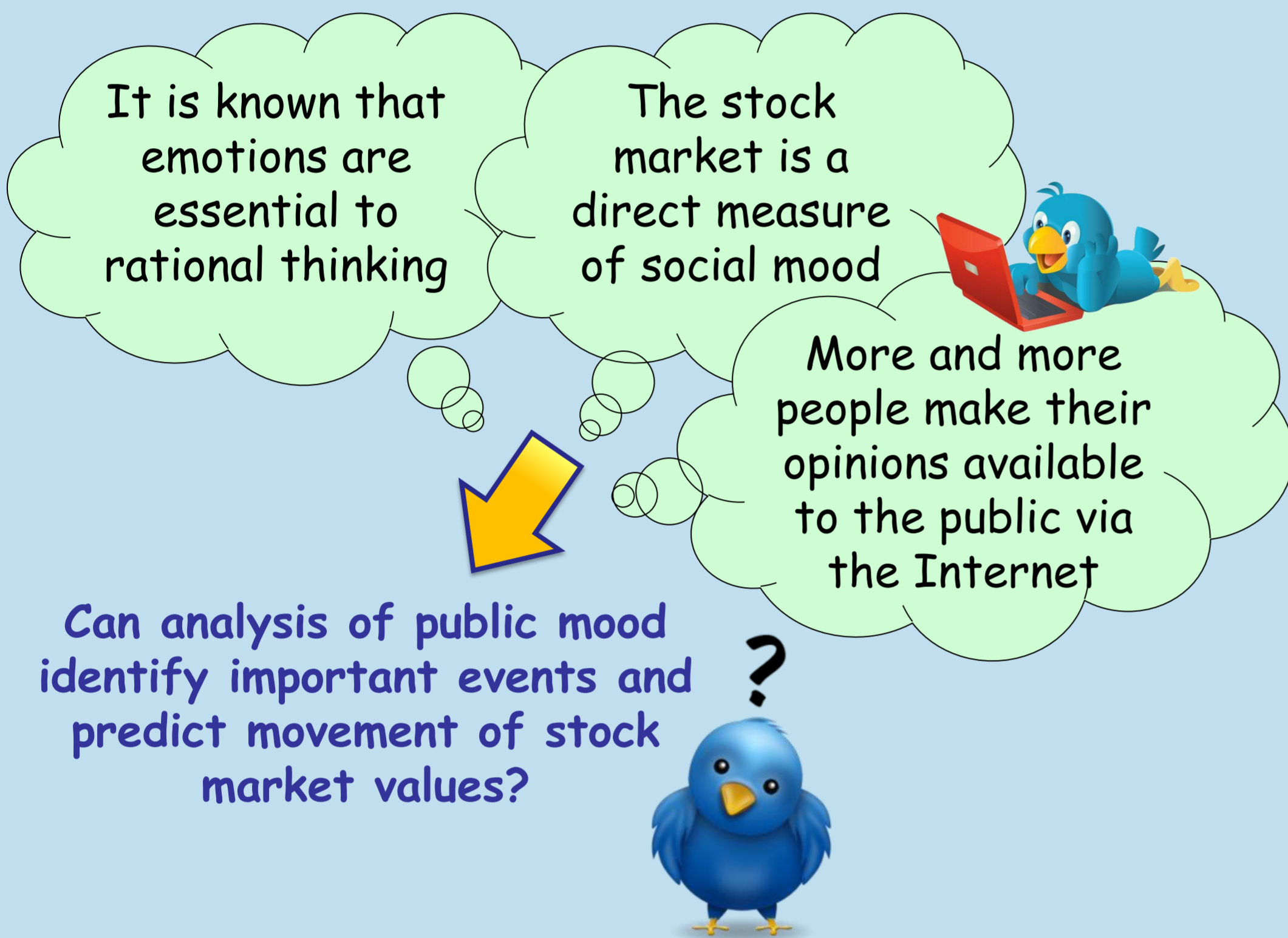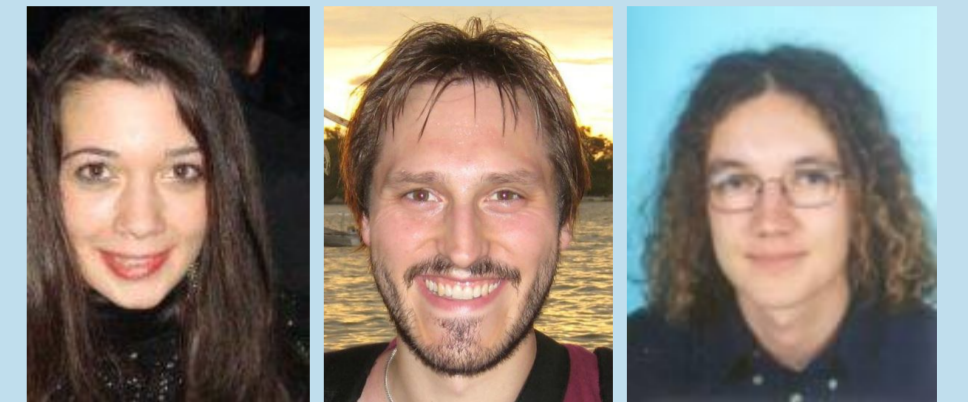
JASMINA SMAILOVIĆ[1], MIHA GRČAR, MARTIN ŽNIDARŠIČ

[1]Study programme: Information and communication technologies,
[1]Jožef Stefan International Postgraduate School
[1]SUPERVISOR: doc. dr. Martin Žnidaršič, prof. dr. Nada Lavrač
Inštitut Jožef Stefan, Jamova cesta 39, 1000 Ljubljana

It is known that emotions are essential to rational thinking

The stock market is a direct measure of social mood

More and more people make their opinions available to the public via the Internet

Can analysis of public mood identify important events and predict movement of stock market values?

We investigate whether sentiment analysis of daily posts from Twitter can identify important events and predict a rise or fall in closing price before a change happens.
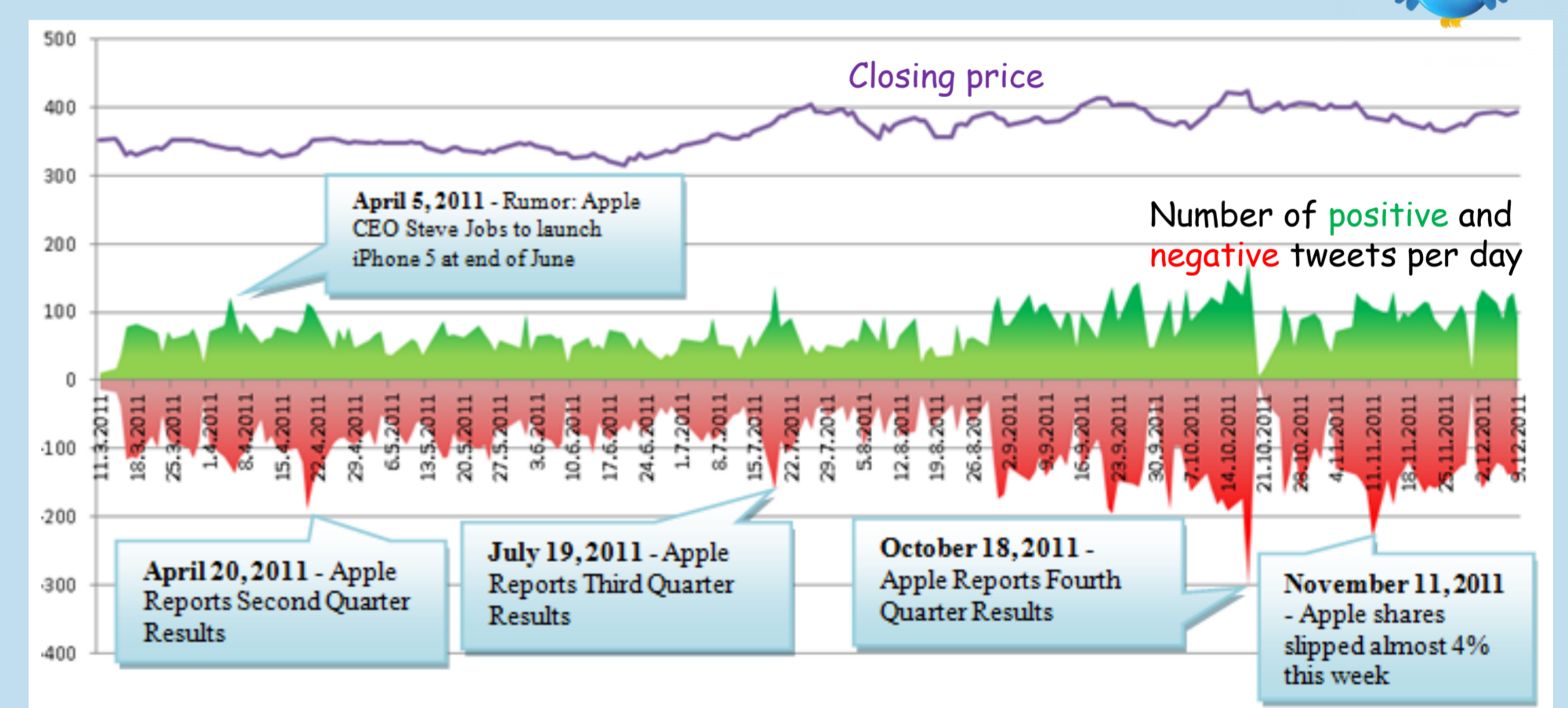
DATA PREPROCESSING SETTINGS FOR THE SVM CLASSIFIER

Machine learning approach

Training set: collection of 1,600,000 tweets

Testing set: manually labeled 177 negative and 182 positive tweets

Several parameters were changed and based on accuracy the best classifier was chosen.

Maximum N gram length	Minimum word frequency	Replace usernames with a token	Replace web links with a token	Remove letter repetition	Accuracy	Precision/Recall
2	2	No	Yes	Yes	81.06%	81.32%/81.32%
2	2	No	No	Yes	78.83%	77.60%/81.87%
2	2	Yes	No	Yes	78.55%	75.86%/84.62%
2	2	Yes	Yes	Yes	78.27%	76.53%/82.42%
2	3	No	No	Yes	76.88%	77.97%/75.82%
1	2	No	No	Yes	76.32%	72.99%/84.62%

CLASSIFYING FINANCIAL TWEETS

We analyzed posts that discussed Apple stocks in the period from March 11 to December 9, 2011.

Number of positive and negative tweets for each day was counted.

Peaks in graph→ people intensively talk about Apple → important events.

Next, we applied The Granger causality test to check whether we can predict future movement of Apple closing price. The test indicates that positive sentiment probability could predict stock price movements, as we got a significant result in our dataset for a two day lag.

Lags	p-value
1	0.4855
2	0.0565
3	0.0872

Changes in values of positive sentiment probability with a delay of two days could predict a similar rise or fall in closing price.