

Twitter Sentiment Analyzer with Machine Learning

Background

Obtaining feedback from a few of your customers is easy. However, imagine that you could ask the whole world! You would know immediately the general opinion and could react accordingly. This machine learning solution examines twitter data in real-time for their tonality. Combined with a query for your product or service this yields an up-to-the-minute dashboard of your customers' satisfaction. The classification, whether a tweet is positive or negative, i.e. the tonality, is done by a deep neural network.

The idea to train a deep neural network already exists since 1990. As machine learning technology progresses, this idea transforms into reality, what is known as deep learning today.

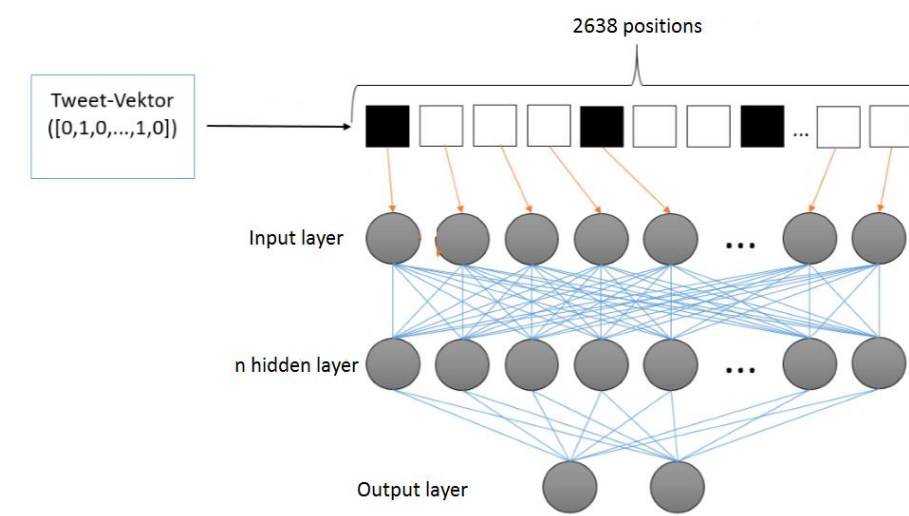
Dataset

- Tweet-sentiment140-dataset of Stanford University
- 1.6 million preclassified tweets

0	1	2	3	4	5
87374637645	Mon Nov 06 22:19:45 PDT 2009	NO_QUERY	HeatherChia	is upset that he can't update his Facebook by texting it... and might cry as a result School today also. Blah!	
39297387288	Tue Dec 07 15:20:17 PDT 2010	NO_QUERY	surajram	Watching I couldn't bear to watch it. And I thought the UA loss was embarrassing	
38738625000	Wed Jan 08 18:05:19 PDT 2011	NO_QUERY	schwendinger	@ianjanylin I wish I got to watch it with you! I miss you and @iamlinnicki how was the premiere?!	
22376376266	Fri Apr 09 09:56:06 PDT 2011	NO_QUERY	clintredwine	@happydayvintage Thank! I guess it feels morning out to my tweet fan...	
28974388388	Sat Oct 10 01:00:55 PDT 2012	NO_QUERY	jillybichn	@happydayvintage Thank! I guess it feels crazy that I'm just now getting around to it, being 30 years old and all	
00399933377	Sun Sep 11 19:29:33 PDT 2010	NO_QUERY	electromarkie	just bought a little yellow outfit for Chelsea	

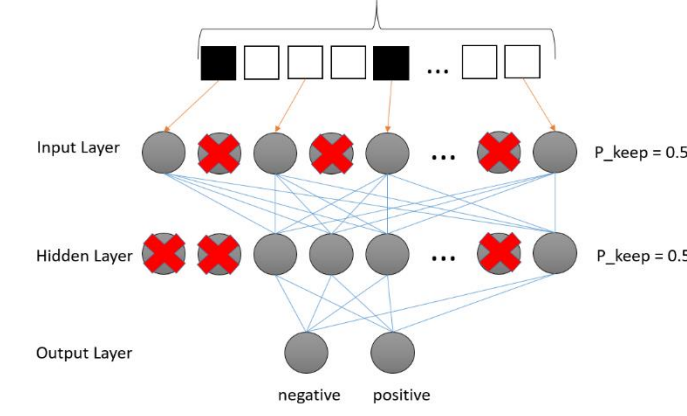
Define neural network

- Deep learning framework TensorFlow
- Feedforward network
- One hidden layer
- Input layer: 2638 neurons = length of lexicon
- Output/softmax layer: 2 neurons (positive or negative)



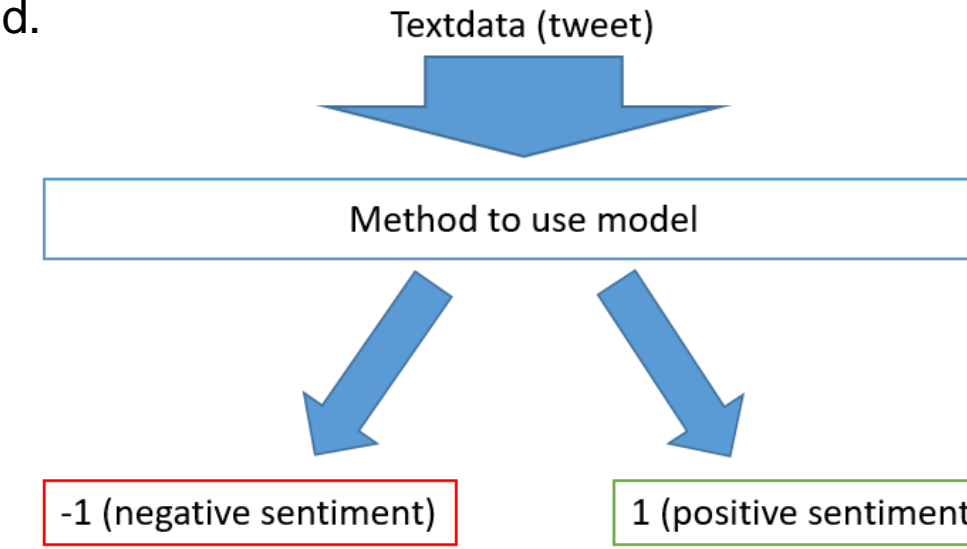
Optimize the training

- L2 Regularizer
- Dropout: Dropping out units in the neural network to reduce overfitting



Use model

The training process of an ML model uses the training data to learn from. The model is an artifact and an output of this process. For using the trained model the weights and parameter will be restored.



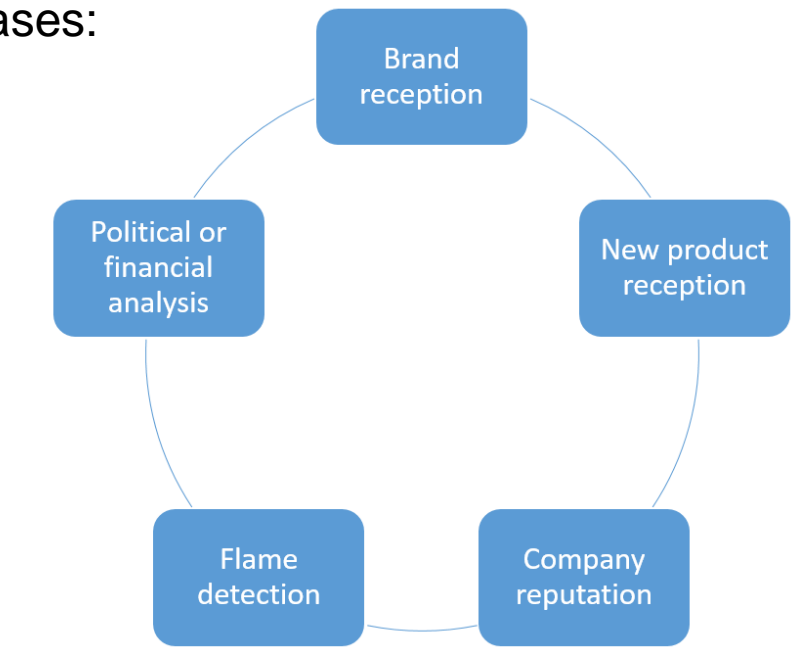
Generate predictions

In this application a real-time prediction is generated, so the front-end can use the results interactively. Another option would be a batch prediction, when you want to obtain predictions for the observations all at once.

To feed the application synchronously with real data the pipeline has to be connected to a streaming API or stream processing framework.

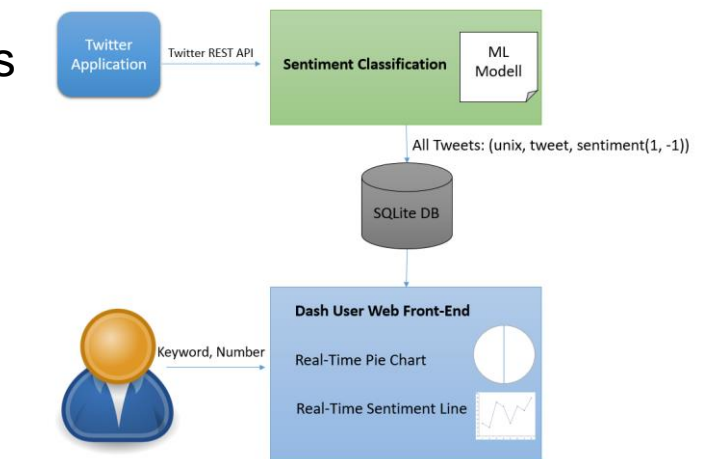
Sentiment Analysis

- Type of data mining that measures the inclination of people's opinions
- Quantify the general public's sentiments in Twitter
- Typical use cases:

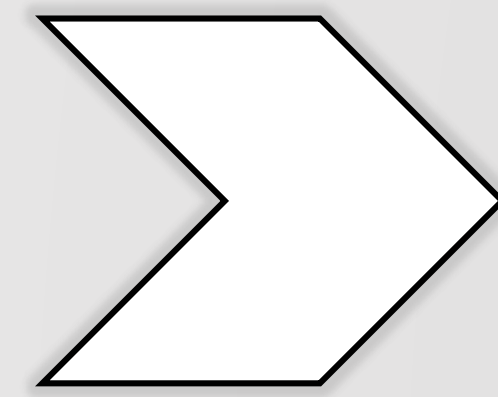


Visualize via web-front-end

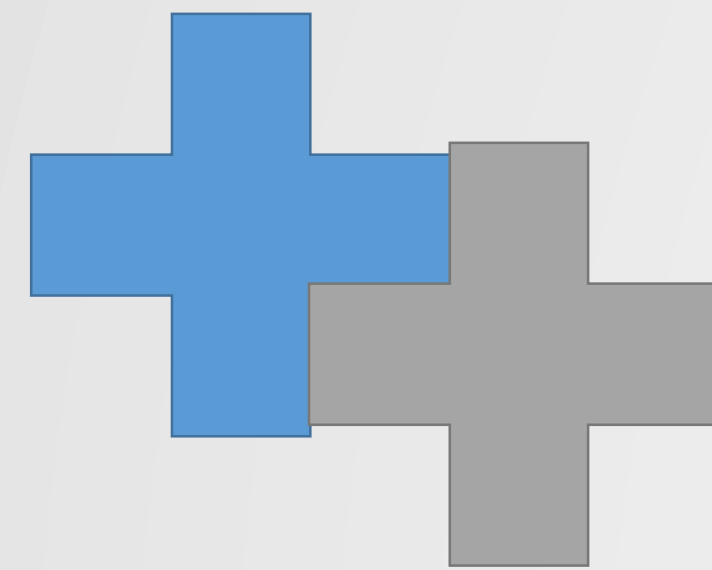
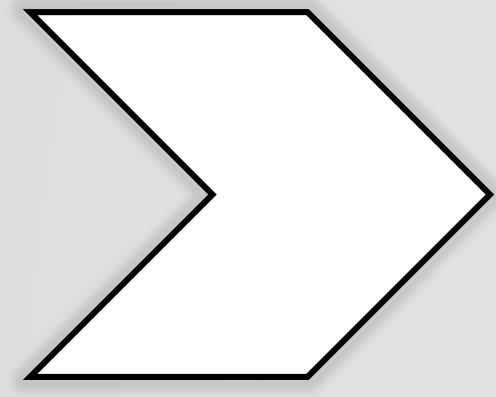
- Read the last classified tweet out of the database
- Process the user input
- create statistics and diagrams



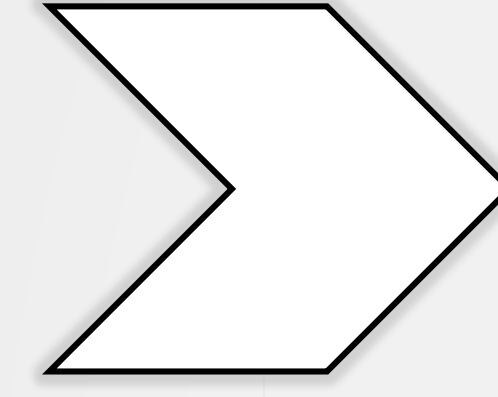
DATA



TRAIN MODEL



INTEGRATE



USE CASE

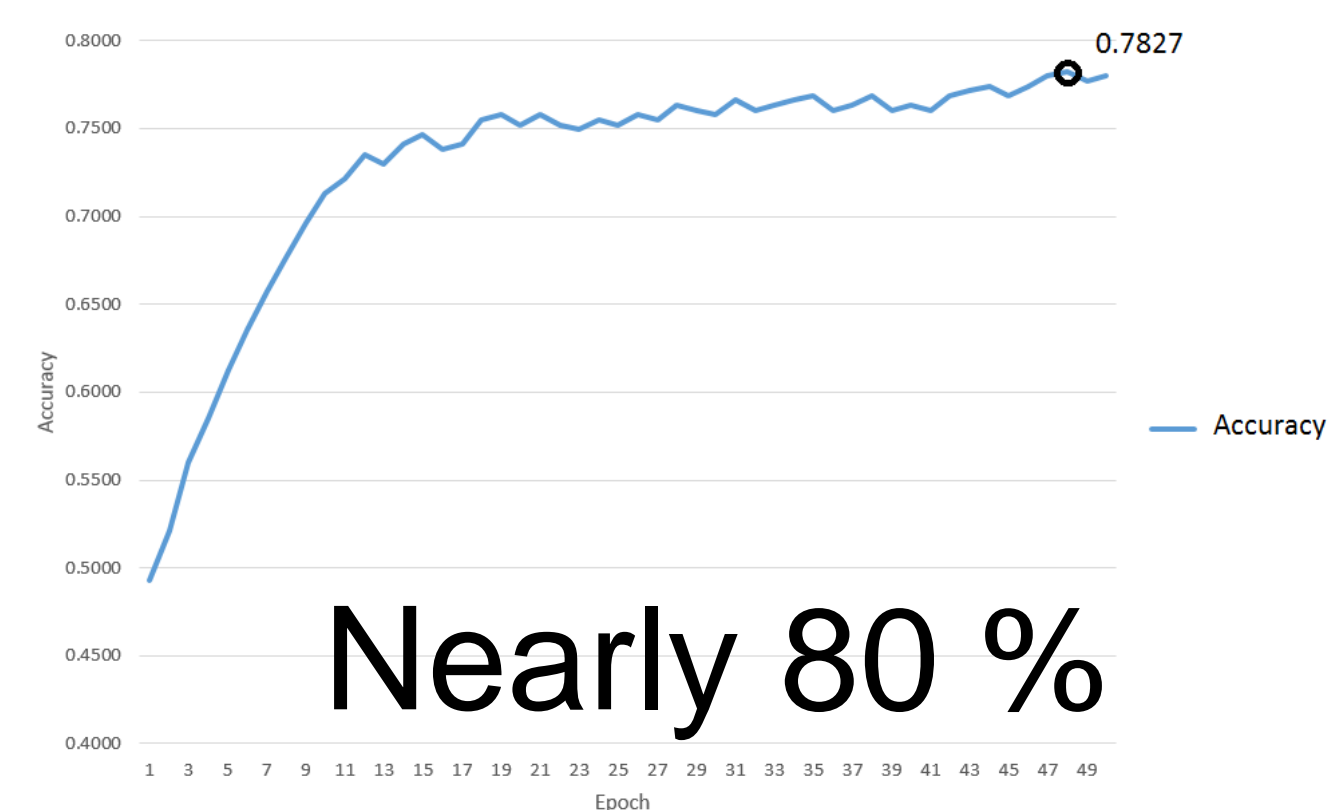
Clean & format data

- Each tweet has positive or negative polarity
- Lexicon-based numerical vector transformation
- Example to create a vector (Bag-of-Words-Model):

Lexicon	[dog, cat, house, flower]
Tweet	The hungry dog runs into the house.
Vector	[1,0,1,0]

Evaluate the training

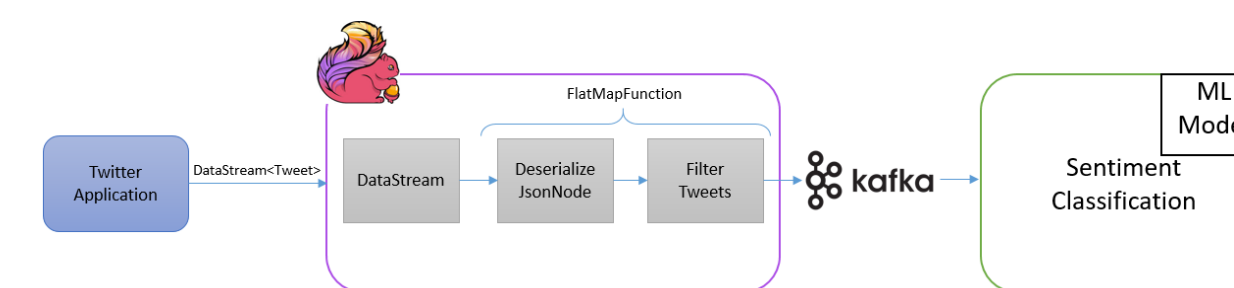
- GPU-based (NVIDIA Tesla K80)
- Input data: 50 epochs, 100 000 tweets per epoch
- Running time: about 1 day!



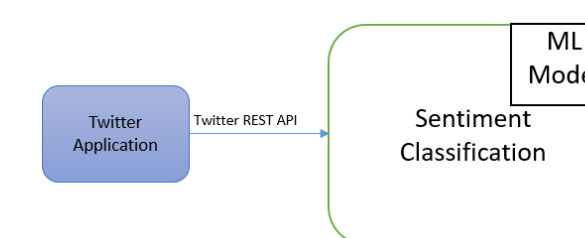
Stream the live data

In this case there are two ways to implement a twitter stream:

1. Streaming pipeline with Apache Flink and Kafka



2. Twitter REST API



Present the results

- Real-time pie chart
- Real-time sentiment line
- Last five tweets
- Matches for the keyword

Visit our website: <https://novatec.sentimentanalyzer.de>

