FlashPoint – build a realtime sentiment analysis processing pipeline

2-LINER Description: Design and develop a real-time processing pipeline and dashboard for the sentiment analysis of tweets and posts about our companies.

[Description of the assignment]

- About ToThePoint and our subcompanies (OnTheSpot, OneDot and ZeroSum) is regularly tweeted about on many channels. We want to check these posts for sentiment analysis so we can have a visual map about how favorable or unfavorable these messages were.
- We want to both check Facebook and Twitter messages where certain hashtags/keywords are used and we want to 'catch' events on a Kafka message bus (LinkedIn) for further processing.
- After the pre-processing of the collected events, we want to stock the messages on a Hadoop backend so we can initiate a sentiment analysis on these messages. Thereafter we want to visualize this sentiment analysis on our giant leaderboard screen in our offices.
- Your assignment is the design and build of this challenge

[Goals]

- Iteratively design and develop a machine learning model
- Develop a data processing pipeline
    - Pre-processing steps
    - Sentiment analysis
- Put into production your application
- Build a live real-time visualization of the results you've found

[What will you gain?]

- Learn how to design an end-to-end data processing pipeline
- AND put this in production
- Gain knowledge about steam processing
- Gain knowledge and experience in machine learning
- You will get to know Hadoop
- You will gain experience in powerful visualization libraries such as D3.js
- That lovely feeling you get knowing your design will be effectively used in production

[What do you need]

- You have a shown interest in a challenging but instructive assignment
- You'd like to explore Machine Learning and stream processing techniques
    - Using Spark, Python or Scala does not scare you at all
- You know what ReactJS is, or are eager to learn
- You like to learn about data visualization
- You like to learn a heck of a lot on a relatively short period of time

[Location of your assignment]

- Veldkant 33B, 2550 Kontich

[Your mentor]

- Kevin Smeyers – Technical Lead Machine Learning

[Technologies you'll be using]

- Spark Connect / Spark Streaming
- Hadoop
- Scala
- Kafka
- ReactJS
- D3.js