**Proposal 1: Public opinion Social media and historical compared to Social media stock over time**

**Defining your problem (deadline: Tuesday)**

You’ll spend the first couple of days building a problem definition. An important part of this is understanding the limitations and possibilities of the platform that you will be building your software on.

The result of this phase is a *proposal document* that thorougly describes the folowing aspects of your application:

* **what features will be exposed to the end user, what problem will be solved for the user**

De gebruiker krijgt inzicht in de relatie van de publieke opinie en de prijs van volatiele aandelen

* **a preliminary sketch of what the application will look like**
* **what data sets and data sources will you need, how you will get the data into the right form for your app**

I will use the folowing sources to extract data:

Yahoo finance: historical stock prices

* **what separate parts of the application can be defined (decomposing the problem) and how these should work together**
* **how the platform will facilitate creating your application, and what external components (APIs) you need to make certain features possible**
* **potential problems that may arise during development and what possibilities you have to overcome these**
* **a short review of similar applications or visualizations in terms of features and possibly technical aspects**

The proposal document should be well-written and clearly formatted. Do not forget to include a title, your name, and a paragraph summary of the application goals at the top.

Although we call the document a proposal, you are expected to refine and discuss the contents with fellow students and the teaching crew *before* handing it in.

**Setting up your development environment (deadline: Tuesday)**

While working on your problem definition, you can start creating your development environment. From the pre-proposal, you should know what platform you will be using and how to test the software during development and thereafter.

If you have not done so already, you should setup *git* to share your code with the teaching staff on a daily basis.

Make sure you have a basic project on GitHub containing all boilerplate code that is generated by the framework that you are using. Also make sure that you can easily test your application so you will not be slowed down during development.

**Submitting**

You will need to submit your boilerplate code by *push*ing it to GitHub. Then go to the GitHub website and put your proposal document on the project Wiki.

Voorstel 1: Social media stock

**Social media stock price fluctuations by twitter information**

**Jeroen Meijaard**



(bron: https://img.youtube.com/vi/oGhau1Y8vXQ/hqdefault.jpg?h=90&w=120&sigh=\_\_MyZcsWZJdkxO3vi6A0zpf5cS6rY= , geraadpleegd op 27 mei 2015)

Doelen:

* Inzicht verschaffen in de relatie tussen het sentiment over een bedrijf op twitter en prijsontwikkeling van een social media aandelen (TWITTER, TESLA, APPLE etc..).
* Een overzicht te geven in belangrijke historische gebeurtenissen voor het bedrijf.

1. Features voor gebruikers.

De gebruiker zal inzicht krijgen in de relatie tussen het sentiment over een bedrijf op twitter en prijsontwikkeling van een social media aandelen (TESLA, APPLE etc..). Een overzicht te geven in belangrijke historische gebeurtenissen voor het bedrijf.

De gebruiker kan interactief in de kaart de historische gebeurtenissen opzoeken, zodat de impact op de aandeelprijs vergeleken kan worden.

2. Schets van de visualisatie

De visualisatie zal de gebruiker een lineplot laten zien waarin de data geplot is over een bepaalde tijdsperiode. Op de x-as waar de tijd is aangegeven zullen ook de historische gebeurtenissen weergeven zijn, waar de gebruiker op kan klikken om extra informatie te krijgen.

De y-as zal gebruikt worden om de prijs van het aandeel en het sentiment te plotten.

3. Data

Voor het verkrijgen van de benodigde twitter data kunnen de volgende bronnen worden gebruikt.

Data voor het sentiment kan worden verkregen via de api: <http://www.followthehashtag.com> of http://snapbird.org/

Eventueel kan er nog terug gevallen worden op de twitterdata verkregen uit januari 2012 ten behoeve van het scriptie onderzoek op de onderstaande link is gepost.

<https://archive.org/details/archiveteam-twitter-stream-2012-01>

Historische social media aandelen prijzen.

<http://finance.yahoo.com/q/hp?s=YHOO>

Voor het bepalen van het sentiment kan de code uit onderstaande scriptie gebruikt worden.

The Impacts of Information on Stock Prices Assessed by Social Media Sentiment, Guangxue Cao

4. Onderdelen visualisatie

- Data transformatie via de voorontwikkelde script van bovengenoemde scriptie t.b.v. sentiment.

- Maken lineplot in D3 of canvas (afhankelijk van de grootte van de te tijdsperiode)

- Line plotten voor koers ontwikkeling en sentiment

- Toevoegen historische gebeurtenissen

- Optioneel: Keuze menu voor het laden van meerdere datapunten.

5. Platform

Zoals in de paragraaf data al besproken zal ik gebruik maken van followthehastag.com om de data aangeleverd te krijgen. Dit platform geeft de mogelijkheid om excel documenten uit te geven.

Hierover dient nog een visualisatie laag gebouwd te worden met d3. De historische data zullen worden gescraped met behalve van de DOM API.

6. Potentiële problemen

Een mogelijk probleem is dat ik maar een uit een korte tijdsperiode (2 maanden) data van twitter kan verkrijgen. Dit is afhankelijk van de api die ik gebruik. Dit kan overkomen worden door de toevoeging te doen om voor meerdere aandelen de visualisatie te maken. Hierdoor kan men, danwel over een korte tijdsperiode, de onderlinge verwantschap en sterkte van prijsfluctuaties in de aandelen vergelijken.

Een ander mogelijk probleem bij het gebruik van d3 is het aantal punten. Echter wordt door het bepalen van de sentiment waarde in het python script al samengevoegd voordat het in javascript ingeladen dient te worden.

7. Overview gerelateerde en bestaande visualizaties

Onderstaande site weergeeft tweets op een geografische kaart weer maar gebruikt geen sentiment om aan de aandeelprijs te relateren.

<http://tweetmap.mapd.com/desktop/>

Hoewel ook het ook niet echt in de buurt komt van het voorstel laat deze visualisatie het sentiment zien bij het invoeren van een bepaalde key.

<http://www.csc.ncsu.edu/faculty/healey/tweet_viz/tweet_app/>

Alleen betaalde tools kunnen visualisaties maken die vergelijkbaar kunnen zijn. Echter zal deze applicaties zich vooral focussen op de historische prijzen in plaats van real time datafeeds.

<http://www.pentaho.com/resource/analyzer-on-mongodb-video>