

Reddiment: Reddit Sentiment Analyse

Tobias Bauer, Fabian Beer, Daniel Holl, Adrian Imeraj, Konrad Schweiger, Philipp Stangl, und Wolfgang Weigl

I. EINLEITUNG

Die GameStop (NYSE: GME) Aktie hat Anfang 2021 eine längere Periode mit schnellen, drastischen Kursschwankungen erlebt. Der Subreddit, ein Unterforum von Reddit, r/wallstreetbets spielte dabei eine Rolle. r/wallstreetbets, auch bekannt als WallStreetBets oder WSB, ist ein Subreddit, in dem über Aktien- und Optionshandel spekuliert wird. Der Subreddit ist bekannt für seine profane Art und die Vorwürfe, dass Nutzer/innen Wertpapiere manipulieren. Dieser Fall wurde von Wang und Luo [3] anhand von Sentiment Analyse betrachtet. Häufig wird Sentiment Analyse dazu verwendet, ob ein Text negative, positive oder neutrale Emotionen enthält. Im Rahmen der Modularbeit in *Big Data and Cloud Computing* stellen wir Reddiment vor - eine webbasiertes Dashboard zur Sentiment Analyse von Subreddits.

Die weiteren Abschnitte des Konzeptpapiers sind wie folgt aufgebaut: In Abschnitt II wird eine Auswahl verwandter Arbeiten vorgestellt. In Abschnitt III werden die Anforderungen in Form von User Stories wiedergegeben. Abschließend wird kurz auf die Methoden in Abschnitt IV eingegangen, gefolgt von den Literaturverzeichnis am Ende.

II. VERWANDTE ARBEITEN

Beispielsweise hat sich Lubitz [2] mit möglichen Treibern von Kapitalmärkten beschäftigt. Anhand von Sentiment Analyse wurden Finanznews die auf Reddit und Financial Times erschienen sind verglichen. Die Vorhersagekraft von Reddit auf künftige Marktbewegungen ist etwas besser als bei der Analyse von Zeitungen.

Die kommerzielle Plattform Brandwatch [1] ermöglicht die Analyse vom Volumen der Gespräche bis zum Sentiment anhand eines Dashboards. So kann man sehen, welche Themen am häufigsten besprochen werden oder welcher Subreddit am aktivsten ist. Es lassen sich Abfragen zu beliebigen Begriffen aufsetzen und mithilfe der Booleschen Operatoren können die Abfragen umfassend und spezifisch sein.

III. ANFORDERUNGEN

In der Anforderungsanalyse wurden drei Stakeholder identifiziert: Benutzer, Entwickler, und DevOps Engineer. Deren Anforderungen werden in diesem Abschnitt in Form von User Stories beschrieben.

A. Subreddit Suche

Als Benutzer möchte ich Subreddits nach Schlagwörtern durchsuchen können, damit ich . Akzeptanzkriterien sind:

- Weboberfläche mit Suchfelder
- TODO

B. Hypeverlauf

Als Benutzer möchte ich einen Graph zum Hypeverlauf, damit ich Akzeptanzkriterien sind:

- TODO
- Überlagerung des Hypeverlaufes mit dem Aktienkurs.

C. Anomalieerkennung

Als Benutzer möchte ich benachrichtigt werden wenn Anomalien in bestimmten Themen auftreten, damit ich Akzeptanzkriterien sind:

- TODO

D. Monorepo

Als Entwickler möchte ich ein Monorepo, damit ich zu jeder Zeit einen konsistenten Stand des Gesamtprojekts habe. Akzeptanzkriterien sind:

- Der Code von Backend und Frontend liegt vollständig im Repository vor.
- Building-, Test- und Linting-Tools sind konfiguriert und funktionieren.
- Im Repository ist eine Anleitung zur lokalen Projekteinrichtung hinterlegt.

E. Entwicklungsdatenbank

Als Entwickler möchte ich eine Entwicklungsdatenbank, damit ich die Funktionalität abgesondert von der Produktionsumgebung umsetzen, als auch evaluieren kann. Akzeptanzkriterien sind:

- Das Datenschema der Entwicklungsdatenbank entspricht der in der Produktion eingesetzten Datenbank.
- Die Entwicklungsdatenbank wird mit realen Reddit Daten populierte.

F. Cloud-Kompatibilität

Als DevOps-Engineer möchte ich eine Cloud-kompatible Anwendung für eine Bereitstellung der Anwendung. Akzeptanzkriterien sind:

- Die einzelnen Teile der Anwendung sind in Containern.
- Eine grundlegende Secrets-Verwaltung zur sicheren Aufbewahrung von Zugangsinformationen ist eingerichtet.
- (Optional) Die prinzipielle Erreichbarkeit ist sichergestellt, sodass bei der Abschlusspräsentation eine Demo der Anwendung in der Cloud möglich ist.
- (Optional) Die eingeschränkte Erreichbarkeit des Spiels in der Cloud ist dokumentiert.

G. Testabdeckung

Als Entwickler möchte ich eine ausreichende Testabdeckung, damit Fehler frühzeitig erkannt werden. Akzeptanzkriterien sind:

- Die Abdeckungsrate liegt bei mindestens 50%.
- TODO

IV. METHODEN

Für die Repräsentation der Anwendung im client-seitigen Frontend wird ein geeignetes Framework verwendet. Zur Speicherung der Daten kommt voraussichtlich Elasticsearch zum Einsatz. Zwischen den beiden Instanzen befindet sich ExpressJS, mit Node.js als Laufzeitumgebung, im serverseitigen Backend. Die Kommunikation zwischen Frontend und Backend wird über eine RESTful-API abgewickelt. Um eine fehlerfreie Anwendung zu entwickeln wird zum einen TypeScript als projektweite Programmiersprache verwendet. Dadurch sollen Fehler bereits zur Kompilierzeit identifiziert werden können. Zum Anderen wird ein geeignetes Test-Framework für Unit-Tests verwendet.

LITERATUR

- [1] J. Boyd. (2022). "Reddit Top 6 Tools für die Subreddit Analyse." [Online], Adresse: <https://www.brandwatch.com/de/blog/reddit-top-6-tools-fuer-die-subreddit-analyse/>.
- [2] M. Lubitz, "Who drives the market? Sentiment analysis of financial news posted on Reddit and Financial Times," *University of Freiburg Publications*, 2017.
- [3] C. Wang und B. Luo, "Predicting \$ GME Stock Price Movement Using Sentiment from Reddit r/wallstreetbets," in *Proceedings of the Third Workshop on Financial Technology and Natural Language Processing*, 2021, S. 22–30.