

Reddiment: Reddit Sentiment-Analyse

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

I. EINLEITUNG

Die GameStop-Aktie (NYSE: GME) hatte Anfang 2021 eine längere Periode mit schnellen, drastischen Kurschwankungen 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. Im Fall der GameStop-Aktie wurde der Subreddit von Wang und Luo [1] in Bezug auf Vorhersagen durch Sentiment-Analyse betrachtet. Häufig wird Sentiment-Analyse dazu verwendet, um zu bestimmen, ob ein Text negative, positive oder neutrale Emotionen enthält.

In den weiteren Abschnitten des Konzeptpapiers stellen wir *Reddiment* vor – ein webbasiertes Dashboard zur Sentiment-Analyse von Subreddits. 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 vom 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 Beiträgen auf Reddit in Bezug auf künftige Marktbewegungen ist etwas besser als bei der Analyse von Zeitungen.

Die kommerzielle Plattform Brandwatch [3] ermöglicht die Analyse des Volumens der Gespräche bis hin zum Sentiment über ein Dashboard. So kann man sehen, welche Themen häufig besprochen werden oder welcher Subreddit aktiv ist. Es lassen sich Abfragen zu beliebigen Begriffen erstellen und mithilfe von booleschen Operatoren können diese 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. Selektion von Subreddits

Als Benutzer möchte ich Subreddits auswählen können, damit ich basierend auf der Selektion weitere Operationen ausführen kann. Akzeptanzkriterien sind:

- Eingabefeld für Subreddits
- Subreddits werden als Namen oder URL des Subreddits eingegeben
- Button für Suchauftrag

B. Eingrenzung der Suche

Als Benutzer möchte ich Schlagwörter spezifizieren können, welche den Suchauftrag eingrenzen. Akzeptanzkriterien sind:

- Eingabefeld für Schlagwörter
- Schlagwörter können als beliebige Strings eingegeben werden
- Suchauftrag agiert ausschließlich mit diesen Schlagwörtern
- Zeitliche Eingrenzung der Suche durch vordefinierte Zeitintervalle (z. B. 7 Tage).

C. Zeitlicher Verlauf des Hypes

Als Benutzer möchte ich einen Graph zum Hypeverlauf, damit ich die Zunahme bzw. Abnahme an Interesse an einem Thema visuell erfassen kann. Akzeptanzkriterien sind:

- Die Aktualität muss sich einstellen lassen auf 1 Tag, 1 Woche, 1 Monat, 3 Monate, 6 Monate.
- Die y-Skala muss als absoluten und prozentualen Wert einstellbar sein.
- Der Graph muss so skaliert sein, dass der Platz (mind. 90% vom Graph-Feld) ausreichend genutzt wird.
- Die Aktualität wird jede Stunde abgefragt.
- Aggregierter Verlauf des Hypes (Zusammengesetzt aus Erwähnungsrate und Sentiment der Erwähnungen).

D. Verknüpfung mit Aktienkurs

Als Benutzer möchte ich den zeitlichen Verlauf des Hypes mit dem Aktienkurs verknüpft haben, damit

- Überlagerung des Hypeverlaufes mit dem Aktienkurs.

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 populiert.

F. Testabdeckung

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

- Die Abdeckungsrate liegt bei mindestens 50%.

G. 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.

H. (Optional) Bereitstellung in der Cloud

Als DevOps-Engineer möchte ich die Cloud-kompatible Anwendung in der Bereitstellung der Anwendung. Akzeptanzkriterien sind:

- Die prinzipielle Erreichbarkeit ist sichergestellt, sodass bei der Abschlusspräsentation eine Demo der Anwendung in der Cloud möglich ist.
- Die eingeschränkte Erreichbarkeit der Anwendung in der Cloud ist dokumentiert.

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. Ein Node.js-Webserver mit ExpressJS im Server-seitigen Backend stellt die Funktionalität der im Frontend angebotenen Aktionen bereit. 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] 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.
- [2] M. Lubitz, "Who drives the market? Sentiment analysis of financial news posted on Reddit and Financial Times," *University of Freiburg Publications*, 2017.
- [3] 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/>.