

Sentiment Analysis based on social media

Diese Bachelorarbeit fokussiert sich auf die Frage, ob die Einschliessung von Sentimentindikatoren in Prognosemodellen zur Verbesserung der binären Klassifikationsgenauigkeit (Accuracy) von Aktienprognosen führt. Die Sentimentindikatoren werden basierend auf Titelzeilen aus Onlinebeiträgen erstellt, in welchen ein Begriff zum Unternehmen vorkommt, über dessen Aktienkurs eine Prognose gemacht wird.

Dieses Projekt verwendet keine Sentimentindikatoren von Drittanbietern, sondern zeigt eine detaillierte Vorgehensweise zur Konstruktion von Sentimentindikatoren aus API-Schnittstellen im Web. Vor der Evaluation der Prognosemodelle wird eine ausführliche deskriptive Sentimentanalyse geführt, um den Aufbau von Sentimentindikatoren zu erläutern.

Logistische Regression und Feedforward-Künstlich-Neuronale Netzwerke werden verwendet, um Prognosemodelle zu entwickeln. Dabei werden sie entweder durch Finanzindikatoren oder durch eine Kombination aus Finanz- und Sentimentindikatoren geschätzt.

Die Hypothese gilt als bestätigt, wenn alle Prognosemodelle, in denen Sentimentindikatoren verwendet wurden, eine bessere Performanz erzielen als jene, welche ausschliesslich aus Finanzindikatoren bestehen.

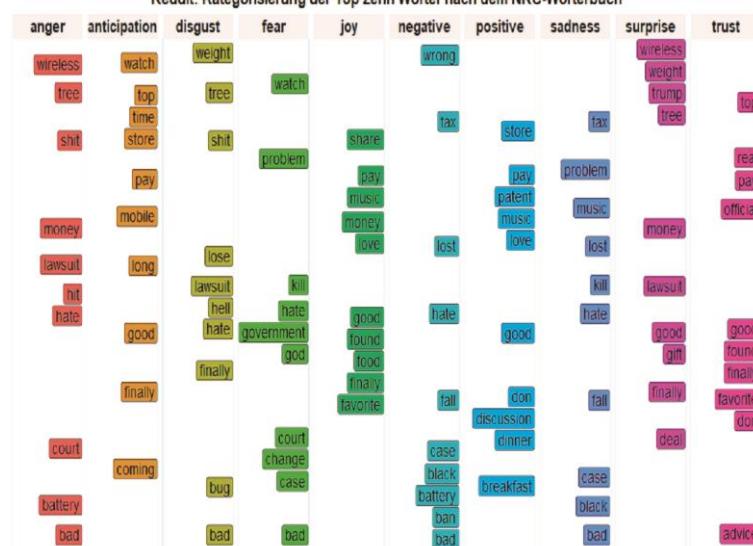
In der Abbildung sieht man die zehn häufigsten Wörter kategorisiert nach den Basis-Emotionen, welche in den Titelzeilen von Onlinebeiträgen auf dem Sozialen Netzwerk Reddit.com zum Begriff 'Apple Inc.' vorkommen.



Diplomand
Zeqë Salihaj

Dozent
Marc Wildi

Reddit: Kategorisierung der Top zehn Wörter nach dem NRC-Wörterbuch



Die zehn häufigsten Wörter in Onlinebeiträgen zum Begriff "Apple Inc." kategorisiert nach den Basis-Emotionen