**Twitter Sentiment Analyse für Cryptowährung Dogecoin**

**Prüfungsleistung Data Exploration Project**

im Studiengang WWI-19DSB

an der Dualen Hochschule Baden-Württemberg Mannheim

von

**Luca Fenucciu, 2512023**

**Kilian Ebi,**

**Miguel Sarasa-y-Zimmermann,**

**Sebastian Hildenbeutel,**

Abgabedatum 12.07.2021

# Projektbeschreibung

Für die Prüfungsleistung der Vorlesung „Data Exploration Project“ wird von der Gruppe eine Sentimentanalyse in Bezug auf die Tweets bezüglich der Cryptowährung Dogecoin durchgeführt. Hierzu bezieht die Gruppe einen geeigneten Datensatz mit Hilfe der Twitter API. Dieser Datensatz wird von der Gruppe bereinigt und zur Verarbeitung vorbereitet.

Für die Durchführung einer Sentimentanalyse wird ein Modell zur Vorhersage des Sentiments trainiert. Als Trainingsdaten dienen Trainings-Tweets des nltk Pakets in Python. Das Modell wird von der Gruppe gemäß den Machine Learning Standards trainiert, evaluiert und optimiert.

# Umsetzung

## Datenbeschaffung

## Datenverarbeitung

Die Tweets beinhalten eine Reihe an Informationen, welche für die spätere Date Auswertung von Bedeutung sein könnten (mögliche Korrelationen). Aus diesem Grund ist es zumindest anfänglich von Vorteil die scheinbar überflüssigen Informationen nicht zu verwerfen. Hierfür empfiehlt sich die Arbeit mittels eines DataFrames dessen Spalten sich einfach auswerten/bearbeiten lassen.

Bevor die Daten in das Modell eingeschleust werden können, sollten sie bereinigt werden. Hierfür betrachtet man zuerst den Texten und macht jegliche Zeichenkombination aus welche man herausfiltern möchte. Oft sind das Twitternamen, URL´s… Obwohl auf den ersten Blick Emojis nicht dazu gehören werden diese auch herausgefiltert da das Modell mittels Standard Tweets trainiert wird, die Crypto Twitter-community jedoch ihre ganze eigenen Emojicon Sprache besitzt. Aus demselben Grund sollte man auch alle nicht englischen Tweets löschen da das Modell auf einem rein englischen Datensatz trainiert wurde.

## Erstellung des Modells

## Auswertung

# Fazit und Ausblick

Die Durchführung des Projekts hat mehrere Probleme einer Sentimentanalyse im Bezug auf Tweets offenbart. Ein Problem ist, dass ein Machine Learning Algorithmus Sarkasmus, Memes und Humor nur schlecht erkennen kann. Dadurch können manche Tweets durch den Algorithmus nicht richtig zugeordnet werden. Des Weiteren ist es in diesem Modell nicht möglich Bilder zu interpretieren, was bedeutet, dass viele Tweets nicht analysiert werden können.

Die Gruppe sieht in dem Modell Potential. Das Modell kann noch erweitert werden, um beispielsweise Bilder noch zu erkennen und Sarkasmus auch teilweise besser interpretieren zu können. Dies ist aufgrund der begrenzten Durchführungszeit allerdings noch nicht möglich gewesen. Im Bezug auf den wirtschaftlichen Nutzen ist es möglich mit dem Modell das Sentiment in Korrelation zu dem aktuellen Kurs zu setzen. Hierzu ist es möglich eine Webapp zu programmieren in dem automatisch täglichen Tweet gezogen werden und das Sentiment automatisch analysiert wird. Dies wiederum bietet dann die Möglichkeit durch die dadurch gewonnen Erkenntnissen ein Modell zur Vorhersage des Kurses zu trainieren.