

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Datensatz und Problemstellung</b>	<b>1</b>
2.1	Initialer Datensatz . . . . .	1
2.2	Bereinigung des Datensatzes . . . . .	1
2.3	Exploration des bereinigten Datensatzes . . . . .	1
2.4	Klassifikationsproblem . . . . .	1
<b>3</b>	<b>Statistische Methoden</b>	<b>2</b>
3.1	Repräsentationen der Wörter und Sätze (mit preprocessing der tokens . . . .	2
3.1.1	Bag-Of-Words . . . . .	2
3.1.2	Term Frequency Inverse Document frequency . . . . .	2
3.1.3	Sequentielle Darstellung . . . . .	2
3.1.4	Word-To-Vec Überwacht, Summe von Word-To-Vec . . . . .	2
3.1.5	Glove Embeddings . . . . .	2
3.2	Algorithmen und Verfahren . . . . .	2
3.2.1	Extreme Gradient Boosting . . . . .	2
3.2.2	Random Forest . . . . .	2
3.2.3	Neuronales Netz: Multi-Layer-Perception . . . . .	2
3.2.4	Neuronales Netz: Convolutional Neural Net . . . . .	2
3.2.5	Neuronales Netz: Long-Short-Term-Memory Neural Net . . . . .	2
<b>4</b>	<b>Statistische Auswertung</b>	<b>2</b>
4.1	Vorauswahl der Verfahren . . . . .	2
4.2	Anwendung der Modelle . . . . .	2
4.2.1	Performanzmaße . . . . .	2
4.2.2	Explaining der besten Modelle . . . . .	3
4.2.3	Anpassung des besten Modell auf den gesamten Datensatz . . . . .	3
<b>5</b>	<b>Zusammenfassung</b>	<b>3</b>
5.1	Ergebnisse . . . . .	3
5.2	Fazit und Ausblick . . . . .	3