

Assignment 7

Angewandte Modellierung 25

Carl Colmant

June 16, 2025

Exercise 1.

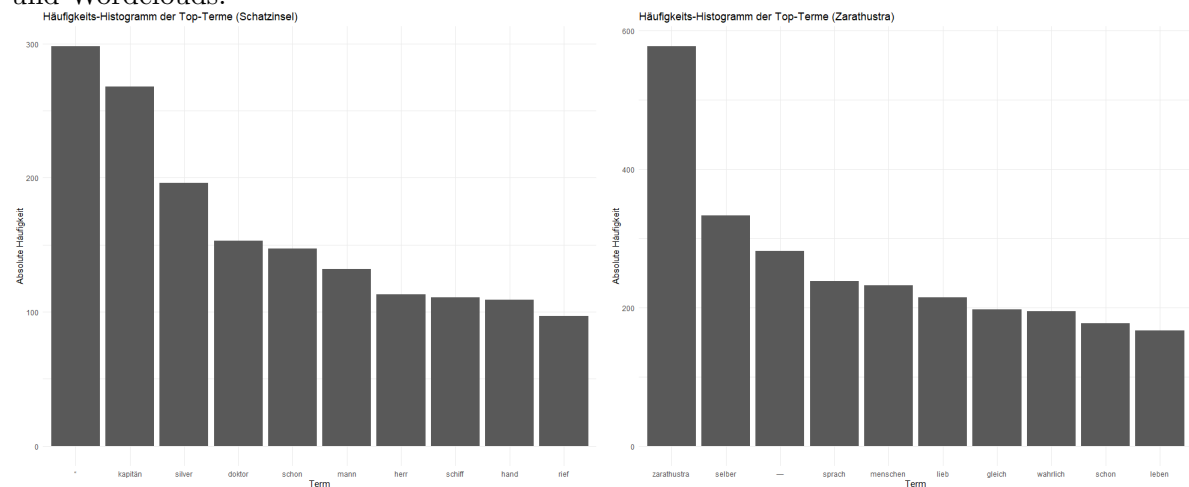
Ich habe mich für die Bücher "Also Sprach Zarathustra" und "Die Schatzinsel" entschieden in Deutscher original Sprache (ein großer Fehler). Ich habe direkt zwei gewählt weil ich später an probleme gestoßen bin und ich so die Möglichkeit habe zu vergleichen ob es an der Sprache liegt oder an den Texten.

2.

Zu erst habe ich die text dateien in R studio geladen. Zun Dann habe ich die Stopppwords raus genommen heir für kann man die sprache der erkannten stopp wörter angeben. Da zu kommt aber noch dass die Texte beide sehr alt sind (Die Schatzinsel ist von 1883 und Also Sprach Zarathustra auch von 1883) das führt dazu dass stopp wörter vor kommen die heute nicht mehr gebräuchlich sind und somit auch nicht in der Stopppwörter liste sind. ("dass", "diess", "'s", "wer", "—", "the", "konnt", "ganz", "mehr", "sagt", "gutenbergTM", "project"). Nun entferne ich befor ich die Stopppwords aus dem Text entferne alle kommata, punkte, gänsefüße usw. und änder den Text in Kleinbuchstaben.

3.

Nun werden noch die Frequenzen der Wörter gezählt und dann kann auhc schon die Wordcloud erstellt werden. Daraus ergeben sich dann folgende Frequency Histogramme und Wordclouds:



Wie man sieht ist bei der Schatzinsel immer noch ein hochkomma im Text enthalten, das passiert weil das ein extra sonderzeichen ist welches ich nicht gefunden habe. In zarathustra gibt es das gleiche Problem mit dem Wort "—" welches ich nicht gefunden habe.

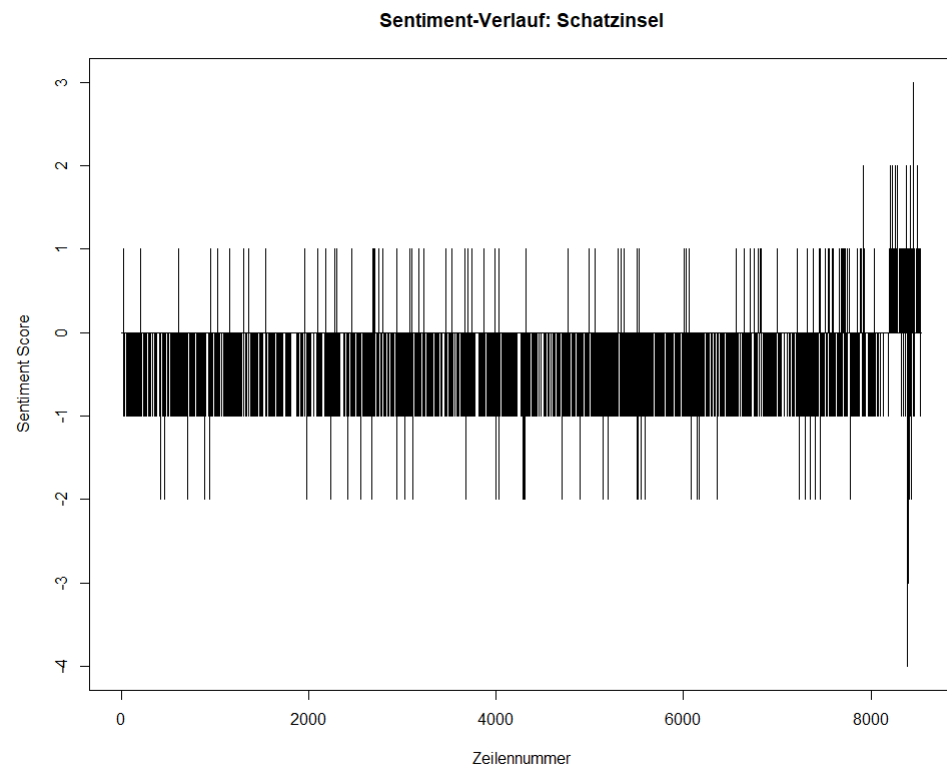


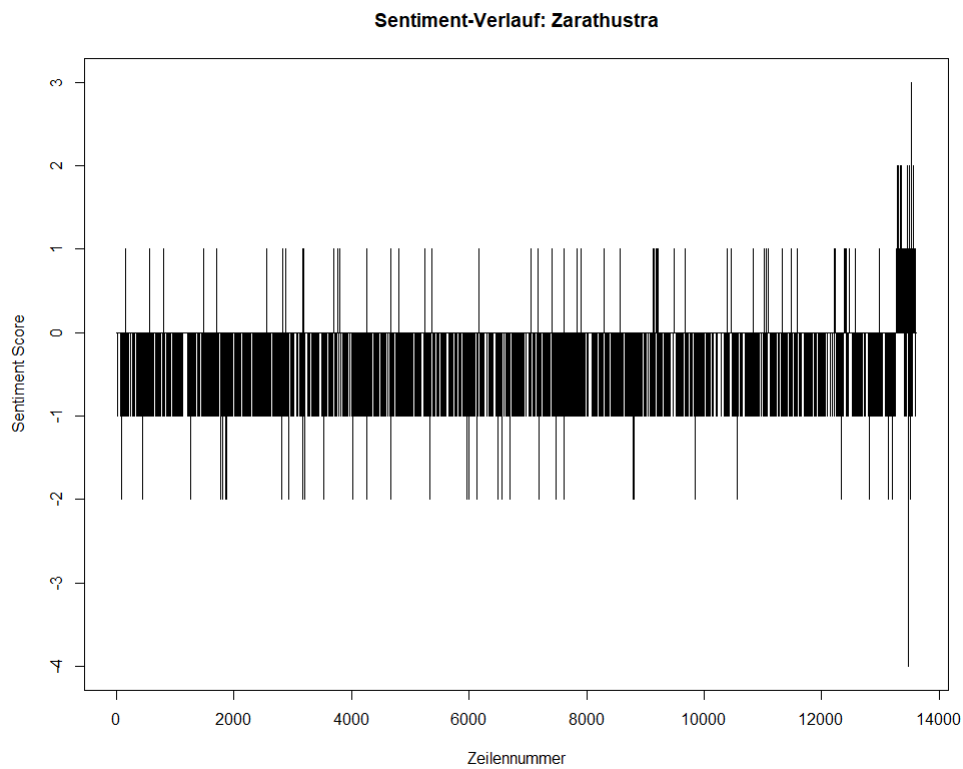


Wie man sieht beschäftigt sich die Schatzinsel als Roman vor allem mit seinen Charakteren und every day life so zu sagen. Zarathustra hingegen ist ein philosophisches Buch mit philosophischen Themen z.B. Leben und Tod und dem Geist und der Seele.

4.

Nun kommen wir aber an Probleme die sich ab hier durch die Aufgaben ziehen, Sentiment analysis ist sehr komplex und die modelle die es gibt sind oft für die englische Sprache trainiert. daher bekommt man hier:





Man sieht das sich die Stimmung der Texte kaum unterscheidet und kaum ausschlägt das ist besonders deswegen komisch weil Zarathustra ein mitreißendes Buch was mit vielen Emotionen geschrieben ist.

Exercise 2.

Github

Wie immer sind alle meine benutzten Dateien auf meinem Github zu finden.