

Task 1 Stopwörter

- Stoppwörter kommen am häufigsten vor auf dem ganzen Trainingsdaten (Label 1). Wir können keine Informationen ableiten, worum die ganze Argumenten gerade gehen.
- Wenn ohne Stoppwörter, kann man nachvollziehen, welche Themen von Argumenten abgedeckt sind: children, education, goverment, money...
- Die Verteilung zwischen Lemmatisierung (3) und Bild 1 sind ähnlich. (Positionen ein paar Wörter sind getauscht). Man kann davon ableiten, dass die Argumenten oft in dem Grundformen geschrieben sind.

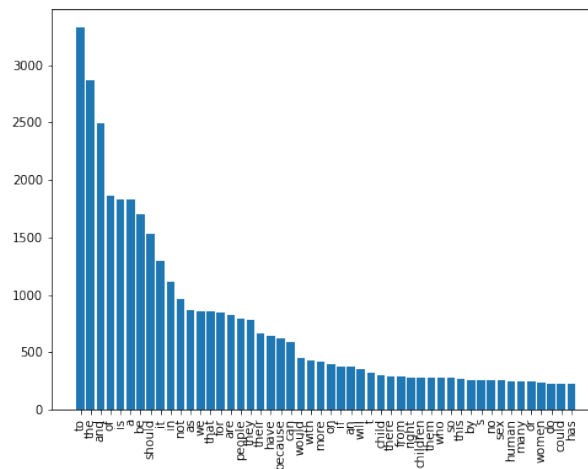


Figure 1: Mit Stoppwörter

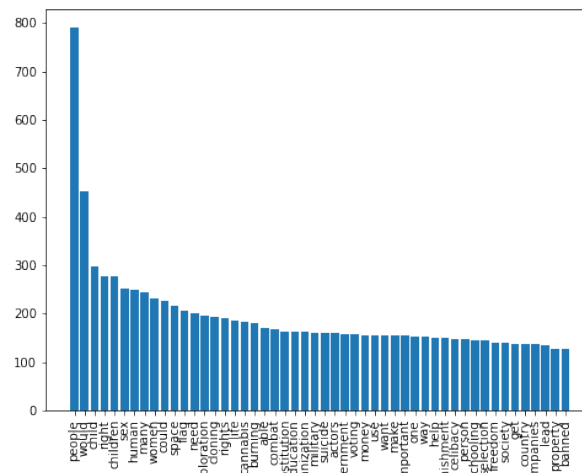


Figure 2: Arguments ohne Stoppwörter

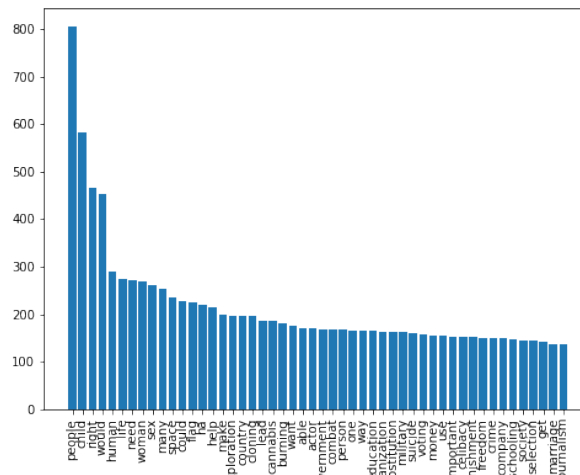


Figure 3: Arguments sind lemmatisiert ohne Stopwörter

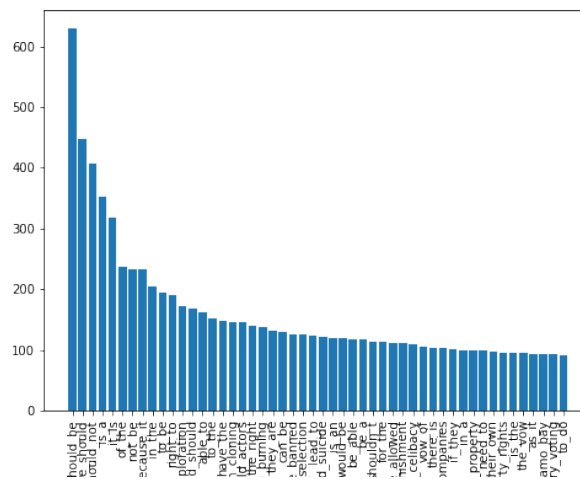


Figure 4: Bigrams

Task 2 Bi-Grams

- Bild 4 und 5: Stoppwörter kommen ebenfalls sehr häufig bei den N-Grams vor. - Triviale Kombinationen: should be, it a, it is, to be, able to, would be kommen vor - Es gibt keinen signifikanten Unterschied zwischen Bi-Grams ohne und mit Lemmatisierung wenn Stopwörter nicht entfernt wurden.
- Wenn Texte lemmatisiert sind und Stoppwörter entfernt sind, kommen einige interessante Kombinationen in die häufigsten Kombinationen vor: human right, space exploration, intellectual property, ...

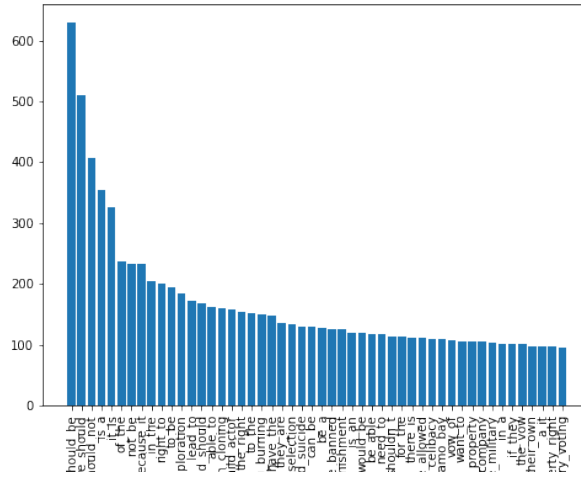


Figure 5: Lemma Bi Grams

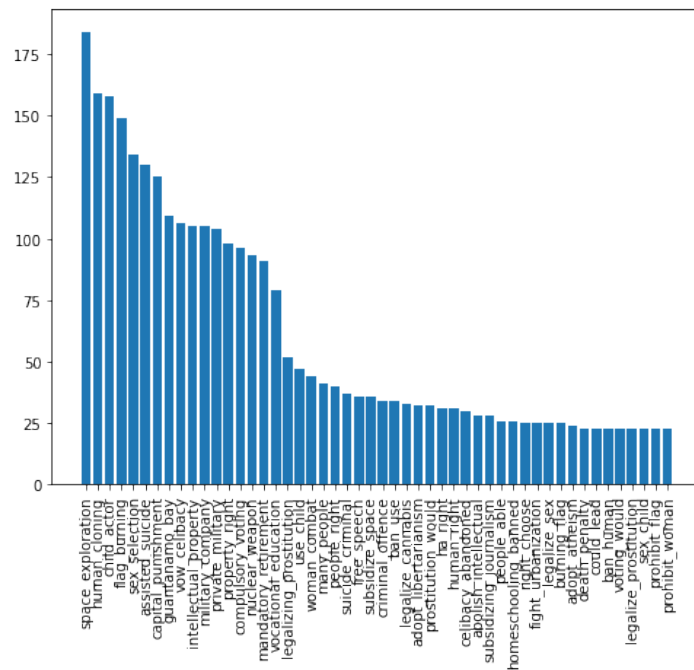


Figure 6: Lemma ohne Stoppwörter

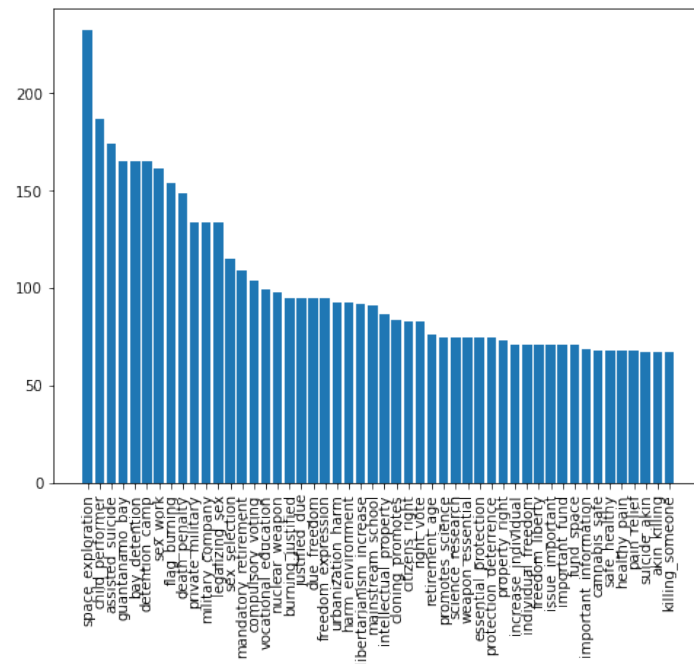


Figure 7: Bi-Grams von Keypoints

Task 3 Bi-Grams zwischen Arguments und Keypoints

Mittels Bi-Grams können wir sehen, dass die interessanten Kombinationen von Wörtern in den Arguments und in den Keypoints gleich sind. → Man benutzt gleiche oder semantisch ähnlichen Wörter für Arguments und Keypoints: space exploration, human right, citizen right, nuclear weapon...

Schauen Abbildung: 6 und 7

Task 4 Tri-Grams zwischen Topics, Arguments und Keypoints

- Topics sind meistens kurz geschrieben, deshalb gibt es nicht viele Kombinationen von Wörtern
- Tri-Grams sind weniger gleich zwischen Topics, Arguments und Keypoints, aber es gibt immer noch die Gemeinsamkeit: legalize sex selection, guantanamo bay detention, intellectual property, ...
- Anhand des Topics verwendet man auch oft die gleichen oder ähnlichen Begriffen, die in dem Topic benutzt hat. Die Wahrscheinlichkeit, dass die Wörter in dem Argument und Keypoints wieder verwendet werden, ist hoch
- Spezifische Kombinationen sind oft im Topic, in den Arguments und Keypoints verwendet.

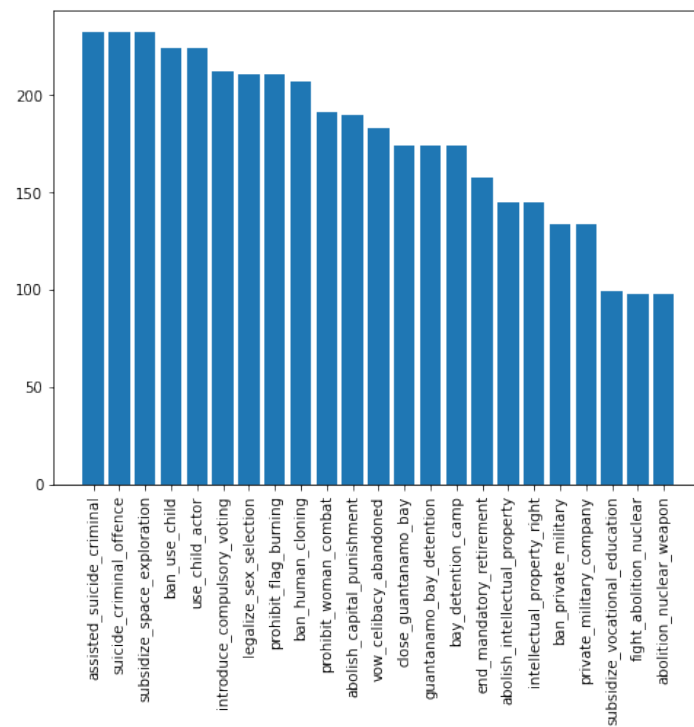


Figure 8: Topic 3 Grams

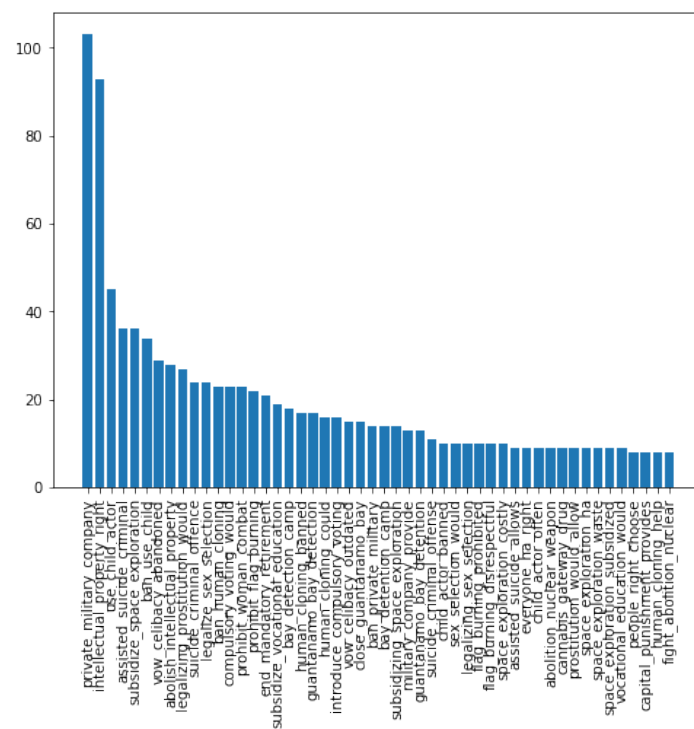


Figure 9: Arguments 3 Grams

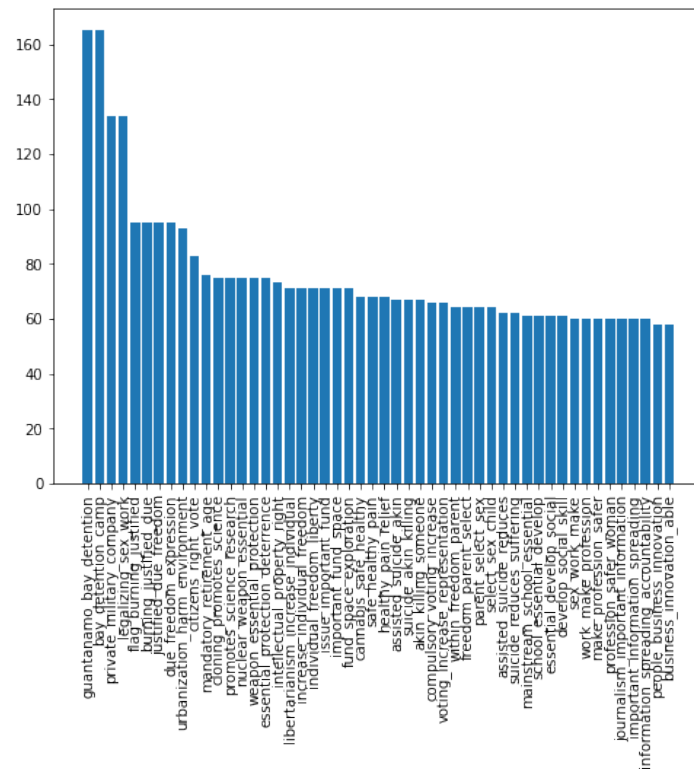


Figure 10: Keypoints 3 Grams

Task 5 Anzahl von Wörtern in den Arguments und Keypoints

Bei den Argumenten sind die Wörter vielfältiger. Man benutzt mehr Wörter, um ein Argument zu schreiben.

Wortschatz bei den Keypoints sind weniger. Alle Keypoints haben die Argumenten 'zusammengefasst'. Deshalb sind sie einfacher mit weniger Wörtern ausgedrückt.

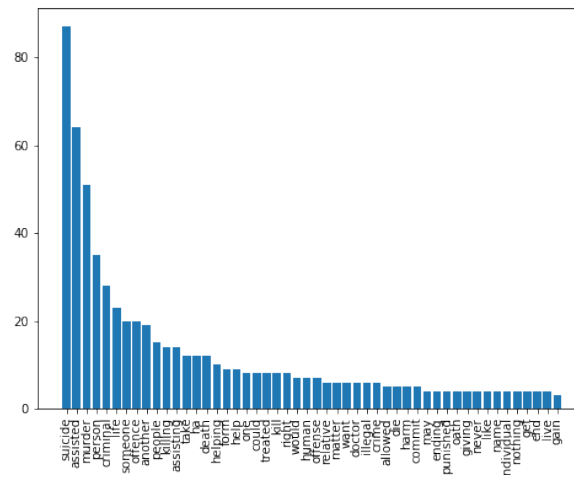


Figure 11: Wortverteilung von den Arguments eines Topics

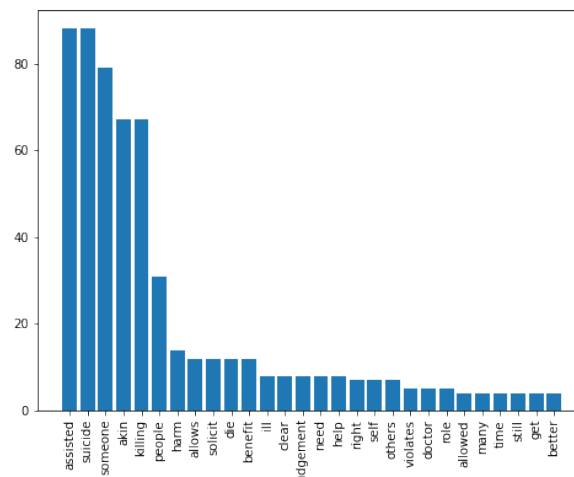


Figure 12: Wortverteilung von den Keypoints eines Topics per Stance