**הסבר על הקבצים**

**TD-IDF/BM25:**

מחולק ל-2 קבוצות קבוצה 1 על קובץ clean וקבוצה שניה על קובץ lemma .

(על מנת שלא נקבל מטריצה גדולה מדי הורדו stopwords ומילים שמופיעות פחות מ-5 פעמים.)

**: Word2vec/glove**

מחולק ל-2 קבוצות קבוצה אחת של מודל glove וקבוצה אחת של מודל Word2vec (300 dim)

כל מודל מחולק ל-2 קבוצות קבוצה אחת על קובץ clean וקבוצה שניה על קובץ lemma.

כל קבוצה clean/lemma חולקה ל- 4 מקרים:

* הורדו stopwords [בלי משקולות].
* הורדו stopwords וגם הוכפל כל וקטור של מילה במשקל שלה(IDF).
* לא הורדו stopwords [בלי משקולות].
* לא הורדו stopwords וגם הוכפל כל וקטור של מילה במשקל שלה(IDF).

**Doc2vec:**

מודל Doc2vec על קבצי המקור (300 dim).

**Bert:**

מודל bert על קבצי המקור (768 dim ) כל וקטור של מילה הוכפל במשקל שלה (IDF)

מודל bert על קבצי המקור (768 dim ) [בלי משקולות]

מודל Sentence-bert על קבצי המקור (384 dim) .