תרגיל - ניתוח טקסט - סיווג סנטימנט

מטרת התרגיל היא לסווג סנטימנט שלילי/חיובי מתוך תגובות בדף הפייסבוק של הנשיא ריבלין.

ראשית, יש להיכנס לקישור הבא:

<https://github.com/omilab/Neural-Sentiment-Analyzer-for-Modern-Hebrew>

בקישור הנ״ל אפשר למצוא שני jupyter notebooks שבשלב זה לא רלוונטיים, אפשר להתעלם מהן. מה שכן נדרש שקריאת ההסבר בקובץ readme וקריאת המאמר עצמו.

לאחר השלמת החלק התיאורטי, נדרש לכתוב מסווג פשוט המבוסס על TF-IDF אך לא רק ואפשר להשתמש ברעיונות נוספים כמו למשל POS. הרעיון, לממש מסווג פשוט ולא רשתות עמוקות, למשל אפשר להשתמש ב-Logistic Regression Classifier.

### References

[Introduction to Information Retrieval](https://nlp.stanford.edu/IR-book/pdf/06vect.pdf)

[Tf-idf weighting](https://nlp.stanford.edu/IR-book/html/htmledition/tf-idf-weighting-1.html)

[How to Use Tfidftransformer & Tfidfvectorizer - A Short Tutorial](https://kavita-ganesan.com/tfidftransformer-tfidfvectorizer-usage-differences/#.YLu8nhSA7Fo)

[sklearn.feature\_extraction.text.TfidfVectorizer — scikit-learn 0.24.2 documentation](https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html)