דו"ח פרוייקט- עקרונות שפות תיכנות

משימת הפרוייקט:

היום רוב החדשות עוסקות בחדשות מפחידות, מלחיצות או סתם סיפורים רעים שקורים בעולם, כי לצערינו, פחד מביא רייטינג(קוראים). רצינו ליצור פלטפורמה שתסנן את החדשות השליליות ובכך תאפשר למשתמש חוויה טובה בקריאת החדשות.

ולכן החלטנו שהמשימה תיהיה ליצור אתר שאוסף חדשות חיוביות באמצעות ML.

דרכי פעולה:

אחרי קריאה בכמה מאמרים בנושא ראינו שדרך נפוצה לזהות אם משפט הוא חיובי או שלילי הוא ניתוח באמצעות n-gram, ולכן החלטנו להשתמש ב- 1 gram ו- 2 gram.

החלטנו להשתמש בNltk על מנת לנתח את המשפטים בקלות, בנוסף במעבדה השתמשנו בכלי זה כדי לבצע משהו דומה.

Datasets:

מצאנו kaggel הוא מכיל כ – 10Kמשפטים חיובים ושלילים בהתפלגות של כ- 50%. בחרנו במאגר זה מאחר והמאגר הכיל משפטים בשפה גבוהה, שמתאים לכתבות בעיתונים ואתרי חדשות.

מקור - <https://www.kaggle.com/chaitanyarahalkar/positive-and-negative-sentences>

תוצאות:

הגענו לתצאות של 75% accuracy ואתר שמראה בעיקר כתבות חיוביות או נטרליות.

מסקנות:

לא פשוט כ"כ לנתח משפטים רוב הכתבות שזוהו ככתבות חיוביות הן כתבות ספורט, אנחנו בחרנו שבמקרה ויש כתבה שזוהתה כנטרלית לא נציגה אותה. ולכן במידה והלמידה תהיה יותר טובה אולי גם יהיה ניתן להציג כתבות נטרליות ולא לחשוש מכתבה שלילית