

פרויקט חלק א' – עיבוד שפות טבעיות בטכניקות למידה עמוקה

תאריך הגשה - 11/12/2025

מטרת הפרויקט

בפרויקט זה תבנו מערכת לזיהוי רגשות בטקסט תוך שימוש בטכניקות של למידה עמוקה. המערכת תקבל משפטים קצרים (תגובות מטוויטר) ותסווג כל משפט לאחד מששת סוגי הרגשות הבאים:
0: *sadness*, 1: *joy*, 2: *love*, 3: *anger*, 4: *fear*, 5: *surprise*

המטרה היא לתרגל בנייה, אימון והערכה של מודלים לעיבוד שפה טבעית ולהשוות בין ביצועי ארכיטקטורות שונות.

Dataset

ה-dataset* איתו תעבדו מכיל ציטוטים של תגובות מטוויטר המסווגות לאחת מבין ששת סוגי הרגשות המצוינים למעלה. ישנם שני קבצים, קובץ אחד הוא קובץ האימון איתו תאמנו את הרשת וקובץ נוסף שהוא קובץ הולידציה עליו תראו את התוצאות. הקוד צריך לתמוך בהרצה של קובץ test.csv שתיבדק על ידי.

דרישות

בחלק זה תיישמו רשת נוירונים מסוג GRU ו-LSTM על מנת לבצע משימת Emotion Detection ותדרשו לבנות מערכת יציבה בה ניתן לשנות פרמטרים על מנת לבצע השוואה ולקבל החלטה לגבי המודל המיטבי. עבור כל רשת יש ליישם את אחת משיטות ה-Embeddings שנלמדו בקורס (למשל GloVe או Word2Vec) בנוס – תוכלו לנסות ליישם שיטת Embedding משלכם.

עבור כל רשת יש לבחון את השפעת שינוי ה-Hyperparameters השונים ולהציג את ההבדלים בתוצאות (אם ישנם) על גבי קובץ ה-validation.

קריטריונים להערכה

- הקוד צריך למלא את דרישות המטלה באופן מלא וכן להיות כתוב באופן קריא וברור
- עליכם להציג באופן ברור את ההשוואה בין רשתות הנוירונים השונות וכן השפעת שינוי ה-Hyperparameters.

*בחלק זה לא תימדדו על תוצאות המודלים אלא על אופן היישום אך שימו לב שהרשת כן צריכה להתכנס ולקבל תוצאות סבירות, השתדלו להגיע לדיוק של 75% ומעלה.

קבצים להגשה

- קובץ/מחברת Python המכילים את הקוד המלא והערות רלוונטיות.
- דו"ח באורך של עד 3 עמודים המסכם את עיקרי התהליך, הניסויים והתוצאות.

*Reference : CARER:Contextualized Affect Representations for Emotion Recognition