

## פרויקט חלק ב' – עיבוד שפות טבעיות בטכניות למידה عمוקה

תאריך הגשה - 15/01/2026

בחלק זה, בדומה לחלק א' תבנו מערכת ליזיוי וגישה בטקסט תוך שימוש בטכניות של למידה عمוקה. המערכת תקבל משפטים קצרים (תגיות מטוייר) ותשוג כל משפט לאחד מששת סוגי הרגשות הבאים:  
0: sadness, 1: joy, 2: love, 3: anger, 4: fear, 5: suprise

המטרה היא לתרגל בניה, אימון והערכתה של מודלים לעיבוד שפה טבעית ולהשוות בין ביצועי ארכיטקטורות שונות. בחלק א' השתמשנו במודלים מסוג LSTM ו GRU ואילו בחלק זה נרצה להשתמש במודלים pre-trained מבוסס Transformers.

מטרת חלק זה של הפרויקט הינה להשתמש בשיטות הפופולריות היום לפתור בעיות בעולם עיבוד השפה הטבעית. כמו כן, עולם ה-ML וה-DL מבוסס מחקר עשיר וכן אנו מאמינים מאוד בקריאת מאמרם והבנתם לצורך יישום.

### דרישות

- עליכם לישם 3 מודלים, ניתן לישם מודלים שנלמדו בכיתה (למשל BERT, ELECTRA) או מודלים אחרים שתואימים לדרישות.
- עליכם להציג השוואה בין שלושת המודלים בהיבט זמן ריצה, גודל המודל וכן השוואה של תוצאות המודלים בהיבט של דיק ומדדי ביצוע נספחים לרלוונטיים בעיניכם למשימה (שים לב שיש חסר איזון בDATA).
- עבור המודל הטוב ביותר - עליכם לבצע **Model Compression** בשתי דרכים לפחות ולבצע השוואה ודרוג של המודלים המכוונים לעומת המודלים הגדולים וمسקנות הנובעות מהතוצאות.

### קריטריוניים להערכתה

- Functionality - על הקוד למלא את דרישות המטרה באופן מלא, המודלים צריכים להתכנס לדיק של 80% לפחות.
- Readability - על הקוד להיות קריא וברוח.
- Quality - על הקוד להיות מאורגן, מודולרי וכזה המאפשר לשזרר את תוצאותיו באופן פשוט.
- שים לב!** הקוד צריך להיות בניו בר שיהיה אפשר לעמוד את המשקלות שהתקבלו מסט האימון עבור המודל הטוב ביותר ולישם על סט טסח חיצוני. מימוש את הפונקציה הבאה:  
**run\_inference(weights, csv) → predictions**
- CSV של טקסטים, ומפיקה קובץ תחזיות.
- Comparison and Research - השוואה בין המודלים והשיטות השונות ואזכורם לרלוונטיים מהספרות.
- Results analysis - ניתוח התוצאות בצדקה מעמיקה ומפורטת. השוו בין מודדים אפשריים שונים, בחרו את המודדים הרלוונטיים לבעה הלא מאוזנת והסבירו את בחירת המודדים.
- Article - מבנה, כתיבה רהוטה והסבירו על חלקי הפרויקט, בחירותיכם, והניסויים שערכתם. מה שלא בתוכם במאמר לא יחשב בציון הסופי.

### קבצים להגשתה

- קובץ/מחברת Python המכלים את הקוד המלא והערות לרלוונטיות.
- קובץ עם המשקלות שהתקבלו מהמודל שביתן לטעינה בקוד ביחד עם קובץ הטסט.
- דו"ח באורך של עד 6 עמודים, בפורמט של מאמר אקדמי המסכם את עיקרי התהילה, הניסויים והתוצאות בפורמט PDF.