

**המכללה האקדמית תל-חי**

**החוג למדעי המחשב**

**למידה עמוקה בפייתון**

**תרגיל 3**

# Multi-Class Text Classification

**מוגש לידי: ד"ר בוריס יאזמיר**

**מוגש על ידי:**

**יובל צל-ציון | 305768871 | yuvalzelzion@gmail.com | 052-8514116**

**גל רייקין | 308188853 | raykingal@gmail.com | 052-4501581**

**סמסטר ב׳**

**2021**

# תיאור האלגוריתם:

האלגוריתם המתואר במאמר הוא אלגוריתם לסיווג טקסט למחלקות. האלגוריתם מקבל פיסת טקסט באורך כלשהו ומסווג את הטקסט אל המחלקה שבה הטקסט עוסק.

אימון המודל מתבצע על מסד נתונים בשם DBpedia. מסד נתונים זה מכיל 6,30,000 פיסות טקסט שמשתייכות ל-14 מחלקות שונות.

# תיאור שלבים במימוש:

# **תחילה כמו תמיד יש לייבא את הספריות הנדרשות, ספריית** torchtext **ותתי-ספריות שלה.**

# **לאחר מכן מגדירים פרמטרים** NGRAMS **ו-**Batchsize**, ואז מורידים את מסד הנתונים מ-**DBpedia**.**

# **יוצרים מחלקה בשם** TextSentiment **שהיא למעשה תהווה את המודל. מאתחלים את ההיפר-פרמטרים הנחוצים** VOCAB\_SIZE**,** EMBED\_DIM**,** NUN\_CLASS**, ויוצרים אינסטנט של המודל.**

# **מגדירים פונקציות שיוצרות את קבוצת האימון ואשר מפעילות את האימון ואת הבדיקה של המודל.**

# **מפעילים את האימון של המודל (**5epochs**), וכאשר האימון נגמר מריצים בדיקה אשר מראה את ההפסד והדיוק.**

# **לבסוף מייצרים את 14 המחלקות, ואז אפשר להתחיל להריץ בדיקות.**

# **במאמר מבצעים 3 בדיקות של סיווג פיסות טקסט וניתן לראות שהטקסט מסווג ל-3 מחקות שונות** NaturalPlace**,** Plant**,** Animal**.**

# מסקנות והצעות לשיפור:

# **לאלגוריתם ביצועים מעולים. אפשר להגיע לתוצאות מדויקות מאוד בלי צורך בעיבוד מקדים גדול בכלל.**

# **ביצענו אימון על מספר גדול מאוד (5,60,000) של מופעים בפחות מ-5 דקות.**

# **הצעה לשיפור שחשבנו עליה בזמן העבודה עם האלגוריתם היא סיווג של טקסט ליותר ממחלקה אחת.**

# **ייתכן כי טקסט מסוים מסווג למחלקה אחת בדיוק חלקי, ולעוד מחלקה נוספת בדיוק חלקי אחר. במקרה זה, אולי אפשר לשפר את האלגוריתם שיתריע כי קיימת מחלקה נוספת שהטקט יכול להשתייך אליה, יראה את כל המחלקות האפשריות ואף את רמות הדיוק שהטקסט יכול להתאים לכל אחת מהן.**

# תוצאות הרצות:

[קישור](https://github.com/yuvalzelzion/DeepLearning/tree/main/HW3) ל-GitHub עם הקוד.

[קישור](https://drive.google.com/file/d/1f98qY51zb6unrHbVNLC_O4tHNK4imoYa/view?usp=sharing) לסרטון של תוצאות הרצה.

תמונות לדוגמא:Text

Description automatically generated with medium confidence

Graphical user interface, text, application, email

Description automatically generated