**עבודה בקורס מבוא לבינה מלאכותית – תשפ"א, 2021**

מרצה: פרופ' יורם לוזון , מתרגלת: שובל פרידמן

**תאריך הגשה: 22.7.21**

כפיר הוא חובב איסוף פטריות בטבע בשעות הפנאי. מאחר והוא חוקר מסודר, הוא סיווג כל פטריה לפי המשפחות ונתן לכל פטריה מאפיינים כלליים (גודל, ריח....).

כפיר נעלם והשאיר אוסף של פטריות לסווג. הנתונים נמצאים בקובץ - "mushrooms\_data.txt".

שלושת השוליות של כפיר (אופק, חיים וליאור) שהינם תתרנים, מתחרים על התפקיד שהתפנה במפתיע, אך לא היה להם את הידע העצום שהיה לו. לכן, הציעו שלושת העוזרים של כפיר שלוש גישות אפשריות לנתונים, כאשר התיוג הוא לפי פיצ'ר הריח (odor):

* גישה ראשונה: לאשכל את הנתונים ולראות את ההתאמה בין כל אשכול לתיוג של הפטריות.
* גישה שניה: לבנות מכונה שלומדת על סמך המאפיינים את משפחת הפטריה.
* העוזר השלישי אמר שיש קודם כל צורך בבניית מאפיינים יותר משמעותיים ורק אז בביצוע הגישות של שני העוזרים האחרים.

אנא עזרו לשוליות של כפיר להתמודד עם היעלמותו של כפיר מהמעבדה.

עבור כל גישה, אנא פרטו את השיטות שהשתמשתם בהם, את הכלים להשוות בין השיטות וצרו ויזואליזציה­­ של התוצאות שלכם בצורה גרפית.

כדי לזכות בתפקיד עליהם להגיש דוח בlatex (overleaf) וקוד ב github. הדו"ח לא צריך להיות ארוך יותר מ6-7 עמודים, מחולק לפרקים הבאים:

תקציר

הקדמה על הבעיה

שיטות

תוצאות

ומסקנות לגבי השיטה העדיפה.

ל­­אחר שליאור השוליה זכה בתפקיד הנכסף מצא במגירה נתונים שנמחקו חלקית בגלל רטיבות (מסומנים ב "-"). אנא עזרו לליאור לסווג את הנתונים ולבדוק כמה מדויק הסיווג שלו. (וגם את זה כמובן תוסיפו לדוח). נתונים אלה נמצאים בקובץ ("mushrooms\_data\_deleted.txt")

**רק למי שעושה בזוגות!!**

כל שיטה שלכם צריכה להיות מלווה במבחן סטטיסטי (אתם יכולים להיעזר בקישור הזה <https://www.scribbr.com/statistics/statistical-tests/> וכמובן לחפש באינטרנט statistical tests ולמצוא תוצאות רבות אחרות)

בהצלחה וחופשת קיץ נעימה☺