**הוראות לפרויקטון**

1. תאריך ההגשה: 09/02/2019
2. מטרת הפרויקט לבחון את יכולת הסטודנטים לממש אלגוריתם מתקדם מתוך קריאה של מאמר מקצועי בתחום.
3. דוגמא חלקית של עבודה נמצאת בתיקיה Example.
4. המאמרים יבחרו מתוך רשימת המקורות שמפורסמת באתר הקורס או על ידי בחירה חופשית של הצוות מתוך מאמרים שהתפרסמו ב-5 השנים האחרונות בכנסים הבאים:

|  |  |
| --- | --- |
| https://dblp.uni-trier.de/db/conf/kdd/ | KDD |
| https://dblp.uni-trier.de/db/conf/nips/ | NIPS |
| https://dblp.uni-trier.de/db/conf/icml/ | ICML |
| https://dblp.uni-trier.de/db/conf/icde/ | ICDE |
| https://dblp.uni-trier.de/db/conf/icdm/ | ICDM |

1. כל זוג צריכים לבחור מאמר אחד
2. סטודנט בודד שרוצה להגיש עבודה לבד רשאי לעשות זאת.
3. כל קבוצה (זוג או בודד) יגישו לנתי שמעוני (nathanie@post.bgu.ac.il) רשימה של שלושה מאמרים שהם מעדיפים על פי סדר העדפה.
4. הקצאת המאמרים לקבוצות על פי עקרון First Come First Served.
5. לאחר אישור בחירת המאמר, הסטודנטים יממשו את **כל** האלגוריתמים במאמר בסביבת בסביבת R או Python.
6. לחלק מהמאמרים ניתן כבר למצוא מימושים באינטרנט. ניתן להשתמש במימושים אלה כ- REFERENCE **בלבד**. **אין להעתיק את הקוד**.
7. יבצעו ניסוים באלגוריתמים על גבי רשימת **של 10** בסיסי הנתונים מתוך אתרים מוכרים כגון:
   1. Kaggle
   2. UCI Repository
   3. Harvard Dataverse
   4. Amazon Public Data Sets
   5. Academic Torrents
   6. Awesome Public Datasets
   7. DATA.GOV
   8. Datahub
   9. Scientific Data Repository
   10. Google Dataset Search
   11. Keel
   12. OpenML
   13. Penn-ML

ניתן להריץ על בסיסי נתונים ממאגרים אחרים באישור המתרגל

1. יש לדווח על המדדים המקובלים לאיכות התוצאות (למשל AUC) וכן זמן הלימוד (באלפיות שניה)
2. לכתוב דו"ח מסכם במבנה של הדוגמא. (חשוב **אין** להעתיק נוסחאות או פסאודוקוד מהמאמר – אלא יש לכתוב אותם מחדש תוך שימוש בסימונים שמקובלים בקורס – ראו סעיף 5 בדוגמת הפרויקט )
3. מה משפיע על הציון?

* שלמות – כל מרכיבי הפרויקט קיימים.
* נכונות.
* בהירות - כתיבה מובנת של הדו"ח
* מקוריות.
* יעילות המימוש.
* תכולה - מורכבות הפרויקט.
* אבלואציה מקיפה של האלגוריתם (בחירת מדדים מתאימים, קונפיגורציות של האלגוריתם ע"פ מספר פרמטרים, בחירת בייסלייינים מתאימים, גיוון ב-datasetים.
* הסקת מסקנות מעניינות הנתמכות בתוצאות הניסויים.
* תיעוד הקוד
* עזרה למשתמש

1. כל עבודה המוגשת על-ידכם (דו"ח, תרשים, נוסחה, שקף, קוד תוכנה וכו') אמורה להיות עבודתכם האישית אלא אם כן צוין אחרת ע"י הפניה מפורשת למקור המידע. מותר להעתיק קטעים מהמאמר, אך אז יש לשים את הקטע המסומן בגרשיים. ניסיון להגיש עבודת אחרים כעבודתכם האישית (plagiarism) עלול להביא לציון אפס בקורס. אם העבודה משמשת גם לקורס אחר או משתמשת בסיס לעבודת הפרויקט הסופי – יש לציין עובדה זו **במפורש** וכן להדגיש אלו תוספות נבעו מקורס הנוכחי.
2. התוצרים הבאים יוגשו כקובץ ZIP או RAR למתרגל
3. קוד המקור
4. הוראות מפורטות להרצה של הקוד – עם פירוט של הפרמטרים של האלגוריתם
5. תוצאות מפורטות של הניסויים
6. דוח על העבודה על פי הדוגמא שצורפה