

תרגיל 1 להגשה בקורס Big Dat

דגשים כללים

1. ההגשה במייל של כל המשימות (בהתאם לפרוט כאן ובסילבוס) עד ליום שהוגדר- 25.5.2024 לשליחת כל המשימות במייל.
2. ההגשה והעבודה הינם אישיים בלבד.
3. **לא יתקבלו הגשות מעבר לתאריך זה.** הגשה במייל מעבר לתאריך שנקבע (לפני מועד ההגנה) – כאילו לא הוגש התרגיל – (הסטודנט ית לא עמדה בדרישות הקורס – לא ניתן לגשת לבחינה).
4. קבצי קוד – יש לשלוח בשמירה בפורמט של Jupiter notebook (**ipynb**) **בלבד** – כולל **הפליטים וההרצה** + קבצי הקלט בפורמט שהורץ בקוד (לא יתקבל פורמט אחר).
5. קובץ הקלט לתרגיל:
 - i. לצורך העבודה יש להשתמש בקובץ ה DATASET שממוקם באתר הקורס במודל (ראה datasetsקובץ)
 - ii. יש להשתמש ב DATASET לפי ספרת הביקורת של ת.ז. (הספרה האחרונה בת.ז.)
 - iii. לדוגמה, במידה והספרה האחרונה בת.ז. הינה 3 - יש להשתמש ב dataset שמסתיים בספרה זו בלבד (dataset_N3)
 - iv. יש לבצע את כל ההוראות בהתאם לדרישות התרגיל ודרישות הקורס המפורטות בסילבוס הקורס
6. יש לשלוח הסבר ופרוט על המשימות ולשלב כהערות בקוד, כולל ניתוח (כגון: שלב X המבצע שאליתה למציאת XXXX, הפלט מציג YYY והמשמעות הינה ZZZZ.. וכו'). הרצה ללא ניתוח, הסבר ו/או משמעות של הפליטים – לא תתקבל.
7. **חשוב להוסיף בהגשה במייל את הפרטי המלאים של מגיש/ה.**
8. בהתאם לסילבוס – ההגשה/ההגנה הינה חובה - במידה וסטודנט ית לא נכח.ה בהגשה/ההגנה ו לא שלח.ה במועד את כל הקבצים בהתאם לדרישה = המשימה לא הוגשה (ציון אפס =0). הרישום להגשה הינו דרך המודל בלבד (כפי שהיה בהגשה של תרגיל 1).
9. ציונים יפורסמו לאחר סיום הקורס.
10. הניקוד למשימות – מתואר/מוצג בסיום המשימות בהמשך (הציון הסופי כולל את ההגשה, ההגנה ומשימה במהלך ההגשה).
11. אין מועד ב' להגשה/הגנה.
12. ההגשה וההגנה הינם אישים (העבודה חייבת להיות מקורית ואישית). הן הניתוח והן המשמעויות יכולות להשתנות (גם לאותו dataset) בהתאם לזווית העבודה, האפיון, היישום של כל סטודנט ית.
13. ההגשה בהתאם לכל דרישות הקורס (ראה סילבוס), כולל דרישות חובת נוכחות בקורס וכו'.

תיאור כללי של המשימה

- המשימה בתרגיל זה עוסקת ביכולת ההכנה של המידע, הניתוח הראשוני, בחירת מודל מתאים ויכולת החיזוי והניבוי לתרחישים שהסטודנט ית מציע.ה (כפי שנלמד והוצג בקורס).
- לצורך התרגיל – שהסטודנט ית יבצע.תבצע עיבוד וניתוח מקיף לארגון (ארגון לא אמיתי) לטובת ההבנה של הנתונים, בחירת מודל מתאים, חיזוי עתידי וקבלת החלטות ומשמעויות של החיזוי והצגה והסבר של כל התהליך (כחלק מההגשה).
1. יש להכין את datasets (ספציבי לפי הת.ז. של כל סטודנט.ית) באופן **מלא ומקיף**, בכלים הרלוונטיים לקורס זה (20 נקודות)
 2. יש להציג את התכונות כמרכזיות של ה dataset ולציין מידע מקיף המכיל פרטים על כל ה dataset, על התכונות השונות, מידע מקיף על כל feature, קשרים בין ה features והסבר ומסקנות בהתאם. (20 נקודות).
 3. יש לבצע מספר שאילתות (בהתאם ליכולות שנלמדו בקורס) המציגות חתכים שונים על התכונות של ה feature השונים, בחיתוכים שונים ומגוונים. יש להציג **לפחות 5 שאילתות** עם יכולות מרוכבות שנלמדו בקורס, כולל ניתוח והסבר על התוצאות לטובת הארגון (30 נקודות)
הערות:
 - השאילתות חייבות להקיף מספר תכונות features עם חיתוכים ואיחודים מורכבים לצורך למידה והבנה, כגון, יכולות של בניית מידע חדש על בסיס הניתוח של מידע קיים, יכולות קיבוץ בהתאם לתנאים שונים וכו'
 - שאילתה ללא ניתוח, ו/או ללא הסבר על התוצאות והמשמעות – תוריד מהציון הכללי – יש להתייחס לנתונים הכוללים בניתוח ולהבין את המשמעויות לטובת הארגון. 4. יש לתכנן use case לצורך בניית מודל לטובת ניתוח וחיזוי בארגון (על בסיס מה שנלמד בקורס + מידע נוסף שקשור ליכולות הסביבה הנלמדת, מודלים אפשריים וכו') ולבנות חיזוי מלא (בכלים שנלמדו בקורס) לצורך התרחבות עסקית ו/או יכולות חדשות שהארגון מציע ו/או ניתוח חדש של מאפיינים חדשים לצורך שירותים חדשים בארגון. (30 נקודות)
 - a. יש להסביר באופן מלא ומפורט את ה Use Case.
 - b. יש להסביר את התהליך מבחינת הקוד, הפלטים, המודל, דיוק ומידע נוסף.
 - c. יש לפרט את המסקנות והמשמעויות לארגון.