

## **עבודת סוף קורס – למידת מכונה וסטטיסטיקה בישומים רפואיים**

### **העבודה בذוגות.**

ביחרו דאטא סט רפואי עלי'ו תבצעו אנליזה לפי ההוראות בהמשך (ניתן להיעזר במאגרי המידע שהעלתי לתאר המודל, למשל דאטא של מיקרוביום, דאטא של גנים וכדומה) וערכו בקובץ האקסל שהועלה למודל את שמות חברי הקבוצה ואת הדאטא סט שבחרתם.

עבודה זו מורכבת מ2 שלבים:

#### **שלב 1:**

הציג עיקרי העבודה, שאלת החקירה, שיטות החקירה, האתגר בדאטא אליו אתם לתרמו, תוצאות ראשוניות, מסקנות ראשוניות ומה מתכוונים להרחבת בהמשך.  
תאריך הגשה: XX נא להשתלב באקסל שהעלותי למודל.  
הציגות יועברו פרונטלית בכיתה על ידי הזוג בilioוי מצגת וגרפים תומכים. משך זמן ההציגה כ5 עד 7 דקות. אין לחזור מעבר ל10 דקות.  
יש להציג פרונטלית למפגש המסכם של הקורס, אנא כבדו את החברים שלכם.

#### **שלב 2:**

הגשה כוללת של העבודה במבנה של מאמר (וgesch בשפה האנגלית cpdf) הכולל את ראש הפרקם הבאים (שםו לב אורך העבודה המקסימלי הוא **8 עמודים**):

- Abstract – main background to the question, main results and conclusion
- Introduction- literature review on the subject
- Results-presenting the main hypothesis and answering them using the tools of the course and new tools. You should support your claims with tables, figures, visualizations.
- Methods – describing the tools and calculations used during the project, such as-algorithms, statistical tests, etc, there is no need for explaining the theoretical methods you used, for example, if you used a specific feature selection, there is no need to define what is feature selection that report the details of the specific feature selection you used (dimension, filtering criteria, etc.)
- Discussion-the conclusions and limitations of your analysis

בנוסף, יש לצרף קישור לGitHub מסודר עם README מפורט והוראות הפעלה מדויקות כך שייהי ניתן לשחזר את האנליזה המלאה שנעשתה בעבודה. יש להעלות אף את הדאטא לגיטהב.

**שםו לב שאותם מספרים סיפור ולא רק מביצעים גיבוב של מבחנים סטטיסטיים בלי קו קשר ביניהם.**

#### **הבהרה:**

בזמן ניתוח הפרויקט יש להתייחס למושגים הבאים (במידה ורלוונטי לשאלת המחקר), נא דאגו שהשאלה תהיה מקיפה ולהשתמש בכמה שיטות כלים ומושגים מהקורס כדי לענות עליה:  
מה הדאטא אותו אתם עובדים, מה מאפייני התתפלגות שלו, מה האתגרים שנובעים מסוג הדאטא זהה, מה האתגרים שנובעים מהמשמעות שאתthem רוצים לבצע, האם הדאטא טבלאי, טקסט, האם הפיטיצרים תלויים, בלתי תלויים, האם processing שאתthem מביצעים לדאטא מתאים להנחות של המודלים שאתthem משתמשים בהם (נירמול, סקלות ערכיים, אי תלות, מולטי קוילינאריות, רגישות למרחקים). בנוסף – יש להתייחס לשיטות הprocessing שנלמדו בקורס, מודלי הלמידה שנלמדו בקורס (מודלים לדאטא טבלאי פשוט, מודלים לתמונות CNN, מודלים לגרפים (GCN, GAT), שימושות המבנה של הדאטא אל מול הערכיהם אם רלוונטי. אימון מודלים, לוס, רגולריזציה, overfitting, חלוקה נכונה של הדאטא לקבוצות אימון, ולידציה,

mbach), שיטות הערכה של המודל תוך שימוש דגש על מטריקות מדידה שונות כמו (r squared, correlation, mistakes, confusion matix, accuracy, AUC, Precision, Recall, F1 score) שימוש במבחנים סטטיסטיים מתאימים ויזואלייזציות מתאימות. בדיקת השערות, היפותיה, רוחם סマー, רמת מובהקות, value-k, טעות סוג ראשון, טעות סוג שני והיחסים ביניהם, עצמת המבחן, מבחן דו צדדי, מבחן חד צדדי, מבחן למדגים אחד, מבחן לזוג מדגים בלתי תלויים, מבחן למדגים מזוגים, התפלגיות ובוחנותם, מבחן ח' בריבוע, מבחן F, מבחנים פרמטריים, מבחנים אפרמטריים, מתאימים, תיקוני למידות מרובות, השוואות, וכן כל מושג שהזכירנו במהלך הרצאות או התרגולים ולא צוין מפורשת.

חשוב מאד – שהפרויקט לא יציג כאוסף מבחנים ובדיקות בלבד קו מקשר, אלא סיפור. אתם יכולים לספר את הסיפור של הדadata שלכם תוך חיקוי התהילה שמחשנו בקורס של בחירת הכלים המיטבים להראות זו וסבירול **כלים קיימים** במידה ויש לכך צורך.

בנוסף – על הגרפים להיות קריאים, כתב קראי, צבעים משמעותיים, כוורות לצירים, caption figure מובנים.

תאריך הגשה של השלב השני של העבודה: XX לא ניתן הארכות מעבר.

#### חלוקת הציון בפרויקט הסופי:

30% - שלב א (פרנצטזה)

70% - שלב ב (העבודה הסופית) + מבחן הגנה על העבודה (הגנה בבירור, עדכן)

שים לב – שיינטו נקודות (עד 10) גם ליצירות והתרשומות כללית, מעבר לנוכנות המתמטית של העבודה.

סומכת عليكم ☺ שייהי מעניין...

**בהצלחה!**