

עבודת סוף קורס – למידת מכונה וסטטיסטיקה בישומים רפואיים

העבודה בזוגות.

ביחרו דاطה סט רפואי עליו תבצעו אנווליזה לפי ההוראות בהמשך (ניתן להיעזר במאגרי המידע שהעלתי לתאר המודל, למשל דאטה של מיקרוביום, דאטה של גנים וכדומה) ועדכנו בקובץ האקסל שהועלה למודל את שמות חברי הקבוצה ואת הדאטה סט שבחרתם.

עבודה זו מורכבת מ 2 שלבים:

שלב 1:

הציג עיקרי העבודה, שאלת החקירה, שיטות החקירה, האתגר בדאטה איתם עליכם להתמודד, תוצאות ראשוניות, מסקנות ראשוניות ומה מתקווים להרחבת בבהמשך.

תאריך הגשה: XX נא להשتبץ באקסל שהעלותי למודל.

הציגות יועברו פרונטלית בכיתה על ידי הזוג בליווי מצגת וברפים תומכים. משך זמן הציגה כ 5 עד 7 דקות. אין לחזור מעבר ל 10 דקות.

יש להגיע פרונטלית למפגש המסכם של הקורס, אנא כבדו את החברים שלכם.

שלב 2:

הגשה כוללת של העבודה במבנה של מאמר (יוגש בשפה האנגלית כpdf) כולל את ראש הפרקים הבאים (שםו לב אורך העבודה המקורי 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 האנווליזה המלאה שנעשתה בעבודה. יש להעלות אף את הדאטה לגיטוב.

שםו לב שאתם מספרים סיפור ולא רק מבצעים גיבוב של מבחנים סטטיסטיים בלי קו מקשר ביניהם.

הבהרה:

זמן ניתוח הפרויקט יש להתייחס למושגים הבאים (במידה וROLONNTI לשאלת המחקר), נא דאגו שהשאלה תהיה מקופה ולהשתמש בכמה שיטות כלים ומושגים מהקורס כדי לענות עליה:

מה הדאטה איתו אתם עובדים, מה מאפייני התפלגותם שלו, מה האתגרים שנובעים מסוג הדאטה זהה, מה האתגרים שנובעים מהמשימה הספציפית שאתם רוצים לבצע, האם הדאטה טבלאי, תומונתי, טקסט, האם הפיטרים תלויים, בלתי תלויים, האם processingprocessing שאתם מבצעים לדאטה מתאים להנחות של המודלים שאתם משתמשים בהם (נירמול, סקלות ערכיים, אי תלות, מולטי קוילינאריות, רגישות למרחקים). בנוסף – יש להתייחס לשיטות processingprocessing שנלמדו בקורס, מודלי הלמידה שנלמדו בקורס (מודלים לדאטא טבלאי פשוט, מודלים לתמונות CNN, מודלים לגרפים (GAT, GCN), שימושות המבנה של הדאטה אל מול הערכיהם אם רולוננט). אימון מודלים, לוס, רגולריזציה, overfitting, חלוקה נכונה של הדאטה לקבוצות אימון, ולידציה,

mbach, שיטות הערכה של המודל תוך שימוש דגש על מטריקות מדידה שונות כמו (r^2 , squared, correlation, mistakes, confusion matix, accuracy, AUC, Precision, Recall, F1 score) שימוש ב מבחנים סטטיסטיים מתאימים ויזואליות מתקינות. בדיקת השערות, היפוטזה, רוח סマー, רמת מובהקות, טעות סוג ראשון, טעות סוג שני והיחסים ביניהם, עצמת המבחן, מבחן דו צדי, מבחן צדי, מבחן אחד, מבחן לזוג מדגים בלתי תלויים, מבחן למדגים מזוגים, התפלגיות ובחינותם, מבחן ח' בריבוע, מבחן F, מבחנים פרמטריים, מבחנים אפרמטריים, מתאימים, תיקוני למידות רבות, השוואות, וכן כל מושג שהזכירנו במהלך הרצאות או התרגולים ולא ציין מפורשות.

חשוב מאד – שהפרויקט לא יציג **כואס מבחנים ובדיקות בלי קשר**, אלא **סתיפור**. אתם באים לספר את הסיפור של הדטה שלכם תוך **חיקוי התהילה** שמחשנו בקורס של בחירת הכלים המיטבים להראות זו ושים כלים קיימים במידה יש בכר צור.

בנוסף – על הגрафים להיות קריאים, כתוב קרייא, צבעים משמעותיים, כוורות לציריהם, caption figure מובנים.

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

חלוקת הציון בפרויקט הסופי:

30% - שלב א (פרנציה)

70% - שלב ב (העבודה הסופית) + מבחן הגנה על העבודה (הגנה בבירור, עדכן שימוש לב-שיינטנו נקודות עד 10) גם ליצירות והתרומות כללית, מעבר לנכונות המתמטית של העבודה.

סומכת عليכם ☺ **שייהיה מעניין...**

בהצלחה!