

יישומי למידת מכונה ברפואה – מטלת אמצע

אתם חלק מצוות הפיתוח של חברה שמספקת פתרונות למידת מכונה בתחום הבריאות, ומתמודדת מול חברות אחרות על פרויקט חשוב שמטרתו לפתח מודלים לחיזוי מוקדם של תמותה במחלקות טיפול נמרץ.

במטלה זו עליכם לבצע ניתוח נתונים גישושי (Exploratory Data Analysis) שיהווה בסיס לפיתוח מודל למידת מכונה לחיזוי מוקדם של תמותה בעת שהייה במחלקה לטיפול נמרץ (death during the ICU stay) על סמך נתונים קליניים.

נתונים:

נתוני mimic 3_demo dataset.

[/https://physionet.org/content/mimiciii-demo/1.4](https://physionet.org/content/mimiciii-demo/1.4)

אוכלוסייה: יש להגדיר את אוכלוסיית המטרה (חולים בעלי תנאי מסוים, כל החולים) עבורם תוצאות המחקר יהיו תקפות.

משתנה תוצאה: death during the ICU stay

משתנים חוזים (Predictors) יש לכלול לפחות 2 משתנים מכל אחד מסוגי המשתנים הקליניים הבאים:

א. משתנים דמוגרפיים

ב. דיאגנוזות

ג. בדיקות מעבדה או סימנים חיוניים

הנחיות:

1. לצורך בחירת המשתנים החוזים, יש להתבסס על הספרות המחקרית בנושא. בחרו מאמר לפי חיפוש ב-google scholar עם מילות המפתח הבאות:
machine learning, mimic III, mortality prediction.
2. הכנת קובץ הנתונים: יש לשלוף את הנתונים מהטבלאות לקובץ csv, ולבצע בדיקות מקדימות של המשתנים בקובץ (לבחון את הנתונים, Unique ID, סוגי נתונים, נתונים חסרים, לייצר את משתנה התוצאה וכו').
3. עיבוד מוקדם (Preprocessing):
 - 1) למשתנים עם מדידות מרובות (multiple measurements), יש לבחור וליישם שיטה לייצוג מדידות במספר נקודות זמן (למשל, first, max, average, count, etc.). להסביר באיזה טווחי זמן התמקדתם ומדוע בחרתם בשיטה זו.
 - 2) למשתנים עם נתונים חסרים (missing data), יש לבחור וליישם שיטת imputation להתמודדות עם נתונים חסרים עבור כל אחד מהמשתנים. להסביר מדוע בחרתם בשיטה זו (במידה ואין נתונים חסרים במשתנים שבחרתם, צרו באופן מלאכותי 2 דוגמאות).
4. ניתוח נתונים גישושי:
 - 1) ויזואליזציה - יש להציג גרפים שמציגים מידע שיתמוך בפיתוח מודל חיזוי (למשל, התפלגויות המשתנים, איתור נתונים חריגים, קשרים בין המשתנים לתוצאה הסופית, קשרים בין המשתנים). הגרפים צריכים להיות בהתאם לכלל כללי ההצגה המקובלים, ולהעביר באופן ברור את המסר.
 - 2) מבחנים סטטיסטיים - יש לבחון מובהקות קשרים בין המשתנים החוזים למשתנה המטרה באמצעות מבחנים סטטיסטיים מתאימים.
5. הגשה: מצגת שתציג באופן ממוקד וממצה את עיקרי **תהליך העבודה, ממצאים והמלצות לגבי שימוש במשתנים שבחנתם במודל החיזוי**. משך ההצגה בכיתה עד 7 דקות. יש לכלול במצגת פרק נספחים שיכלול חומרים שלא הצגתם במסגרת הזמן הנתונה אך אתם יכולים להישאל לגביהם.
6. קהל המטרה שלכם הוא מנהל הפיתוח ומנהלי המוצר.
הגשה במודל: מצגת+ מחברת פייתון מסודרת.