

רקע: סוכרת מסוג 2 היא מחלה מטבולית שמתאפיינת באי היכולת של אינסולין, הורמון שמופרש על ידי הבלב, לתרום להחדרת גלוקוז (סוכר) לתאי הגוף - מצב שנקרא תנגודת לאינסולין. סוכרת, לפחות בשלבים המוקדמים שלה, היא מחלה א-סימפטומטית והחולה אינו חש באי-נוחות, כאב או סובל מתסמינים אחרים. אולם, כאשר לארוך זמן רמת הסוכר בדם אינה מאוזנת, עלולים להופיע סיבוכים כגון: רגל סוכרתית, פגיעה בכליות, פגיעה בראייה ובכלי גם קטנים, והן סיכון מוגבר להתקפי לב ולמקרי שבץ.

תיאור האתגר: מטרת האתגר הינה זיהוי מוקדם של הידרדרות כלייתית (נפרופתיה) בקרב חולי סוכרת. האינדיקציה להידרדרות כלייתית בפאנל הנתונים המצורף הינה האבחנה של כשל כלייתי כרוני (chronic renal failure) אצל החולים הסוכרתיים. בעיקרון, הבעיה כאן היא בעיית סיווג (classification), כשהמשתנים המסבירים (features) מתארים את מצבו של חולה בזמן אבחון במחלת הסוכרת. המשתנים מכילים מידע סוציו-דמוגרפי, מידע קליני (תחלואת רקע) וערכים של מספר סוגי בדיקות דם שנלקחו סמוך לזמן האבחנה. בעת ביצוע התחזית, נרצה לחזות את ההידרדרות בשלושה אופקי זמן הבאים: תוך שנתיים, תוך חמש שנים ותוך עשר שנים. לכל אופק זמן יישוב ציון F1 (ממוצע $k=10$), cross-validation, וסט פתרונות בעל ממוצע ערכי F1 הגבוה ביותר יזכה את הפותר במקום הראשון. הפתרון יכיל: תיאור של שיטת המימוש (הסבר או/קוד), ותוצאות מספריות של ביצועי המודל. מודל לוגיסטי בסיסי שהוא מוחל על כלל המשתנים המצורפים אמור להניב תוצאת AUC של כ-0.73 – נסו להתעלות מעבר לזה! בנוס משמעותי יינתן לזיהוי ואפיון אוכלוסיות בסיכון (תכל"ס מה שמעניין את הקופה ☺)

אתגר שבאתגר: האתגר במשימה נעוץ בשלושה ממדים. 1- בדיקות הדם, שאמורות להסביר בצורה משמעותית את השונות בין החולים השונים, אינן תדירות ולמספר לא מבוטל מהחולים חסרות בדיקות (ערכי NA). חוסר זה אינו אקראי (not missing at

random) והוא עשוי לנבוע ממספר מגוון של גורמים כגון: מצבו הבריאותי של החולה ודפוסי התנהגות שלו. 2- לכל מבוטח נרשם זמן שונה עד שאבחנה (נתון זה קיים בנתונים המצורפים). נסו להבין האם זה משפיע על פתרון האתגר? 3- בהמשך ל-2, יש אוכלוסיה גדולה של מבוטחים שאובחנו בכשל כלייתי כרוני בזמן גילוי מחלת הסוכרת. האם הם אוכלוסיה שונה (תת-אוכלוסיה) או מקרה פרטי של האוכלוסיה הכללית?

דרישות תוכנה: ניתן להשתמש ב-R, פייטון או ג'וליה (Julia)
