

ניתוח וחיזוי ביטולי ניתוחים מתוכננים: דוח פרויקט מסכם

תאריך: 16.07.2025

מגישים: נועה חמו ומקסים זוטובסקי

מנחה: ד"ר שרון ילוב הנדזל

פרק 1 : מבוא

בפרויקט זה בנינו מודל למידת מכונה המבוסס על נתונים מבית החולים מאיר, שמטרתו לחזות האם מטופל צפוי לא להתייצב לניתוח שתואם לו מראש. תחזית כזו תוכל לאפשר לבית החולים להיערך מבעוד מועד, לשפר את ניצול חדרי הניתוח והמשאבים, ולהפחית את מספר הביטולים הבלתי צפויים.

ביטול של ניתוח מתוכנן יוצר עומס תפעולי ניכר על מערכת הבריאות : הוא מוביל לבזבוז של חדרי ניתוח שהוקצו מראש, משבש את לוחות הזמנים של צוותים רפואיים, ופוגע בניצול היעיל של ציוד ומשאבים. ההשלכות הכלכליות של ביטולים אלו עשויות להיות משמעותיות, במיוחד כאשר אין מספיק זמן למלא את המשבצת שהתפנתה. תופעת הביטולים נפוצה במיוחד בתחום הרפואה הציבורית ומהווה אתגר מתמשך בתכנון וניהול משאבים. מעבר להשפעה על המערכת עצמה, לביטול ניתוח יש גם השלכות ישירות על המטופל : העיכוב עשוי להאריך את זמן ההמתנה לטיפול הדרוש ולהוביל להחמרה רפואית במקרים מסוימים.

"בכ-24% מהתורים שלא בוטלו [לרפואה יועצת] המבוטח לא הגיע אליהם, שהם כ-3.5 מיליון תורים... בשנה"

- דו"ח מבקר המדינה, "זמני המתנה לקבלת שירותי רפואת מומחים", 2023

פרק 2: תהליך העבודה ומתודולוגיה

2.1 איסוף והבנת הנתונים

בסיס הנתונים שעליו התבסס הפרויקט התקבל מבית החולים מאיר. הוא כלל תיעוד נרחב של כ-63000 ניתוחים מתוכננים, מתוכם כ-20% ניתוחים שבוטלו.

הנתונים הגולמיים כללו מגוון רחב של פרמטרים עבור כל ניתוח, ביניהם:

מידע אדמיניסטרטיבי: תעודת זהות מותממת של המטופל, קוד ניתוח, והמחלקה המנתחת.

מידע לוגיסטי: תאריכי פתיחת הבקשה והניתוח המקורי, אתר הניתוח (חדר ניתוח) וסיבת הביטול.

מידע דמוגרפי: גיל, מין, עיר מגורים ומצב משפחתי של מטופל.

מידע קליני: סוג הרדמה, סיבת הניתוח, אבחנות רקע, רשימת תרופות קבועות ומדדי מעבדה כגון, BMI, המוגלובין, ואשלגן.

אחד המאפיינים של בסיס הנתונים זה היותו בלתי מאוזן, כ-80% מהרשומות מייצגות ניתוחים שהתקיימו, בעוד שרק כ-20% מייצגות ניתוחים שבוטלו. חוסר איזון זה משפיע ישירות על בחירת מדדי ההערכה למודל. שכן מודל שיחזה שניתוח הוא תמיד לא מבוטל יגיע ל-80% דיוק, אך למעשה הוא חסר כל ערך מעשי. לכן, הפרויקט שלנו מתמקד בממדים הרגישים לחוסר איזון.

רגישות (Recall)

מה הוא אחוז הביטולים האמיתיים שהמודל הצליח לזהות. זהו הממד הקריטי ביותר להצלחת הפרויקט.

דיוק החיזוי (F1-Score)

ממוצע הרמוני Precision ו Recall המספק תמונה מאוזנת על ביצועי המודל

2.2 הנדסת תכונות Feature Engineering

הנתונים הגולמיים עברו תהליך עיבוד והכנה מקיף במטרה להפוך אותם למתאימים למודל למידת מכונה ולהפיק מהם תובנות בעלות ערך. תהליך זה כלל ניקוי נתונים, טיפול בערכים חסרים, ובעיקר – הנדסת תכונות, במסגרתה פותחו משתנים חדשים בעלי פוטנציאל חיזוי גבוה.

להלן התכונות המרכזיות שפותחו:

זמן המתנה לניתוח - Wait days

חושב כהפרש הימים בין תאריך פתיחת הבקשה לניתוח לבין מועד הניתוח המתוכנן. משתנה זה נועד לבחון את ההשערה כי זמן המתנה ארוך יותר מעלה את הסיכוי לביטול.

מרחק מהמרכז הרפואי - Distance_km

פותח על ידי המרת עיר המגורים של המטופל למרחק גיאוגרפי (בק"מ) מבית החולים. משתנה זה מכמת את המאמץ הלוגיסטי הנדרש מהמטופל להגיע לניתוח.

משתני מורכבות רפואית: num_medications, num_diagnoses

במקום להשתמש ברשימות הטקסט הארוכות של תרופות ואבחנות, יצרנו שני משתנים מספריים: האחד סופר את כמות התרופות הקבועות, והשני את כמות אבחנות הרקע. משתנים אלו מהווים מדד פשוט למורכבות הרפואית של המטופל.

משתנים מבוססי זמן

מתאריך הניתוח המתוכנן חולצו משתנים קטגוריאליים המציינים את העונה בשנה (קיץ, סתיו, חורף, אביב) ואת היום בשבוע. זאת במטרה לבדוק האם יש דפוסים עונתיים או יומיים בביטולי ניתוחים.

2.3 הכנת הנתונים (Data Preparation)

בשלב הסופי של הכנת הנתונים, ולפני הזנתם למודלים, בוצעו שתי פעולות עיבוד חיוניות:

טיפול בערכים חסרים (Missing Values)

רשומות שהכילו נתונים חסרים טופלו באופן שיטתי. בהתאם לאופי המשתנה וכמות הערכים החסרים, הפעולה שנבחרה הייתה הסרת הרשומה כולה או השלמת הערך החסר באמצעות ערך סטטיסטי, כגון הממוצע של אותה עמודה.

קידוד משתנים קטגוריאליים (Categorical Encoding)

משתנים המכילים מידע טקסטואלי, כדוגמת "מחלקה מנתחת" ו"סוג הרדמה", הומרו לייצוג מספרי. תהליך זה הכרחי על מנת לאפשר לאלגוריתמים של למידת מכונה לעבד את המידע ולהתאמן עליו.

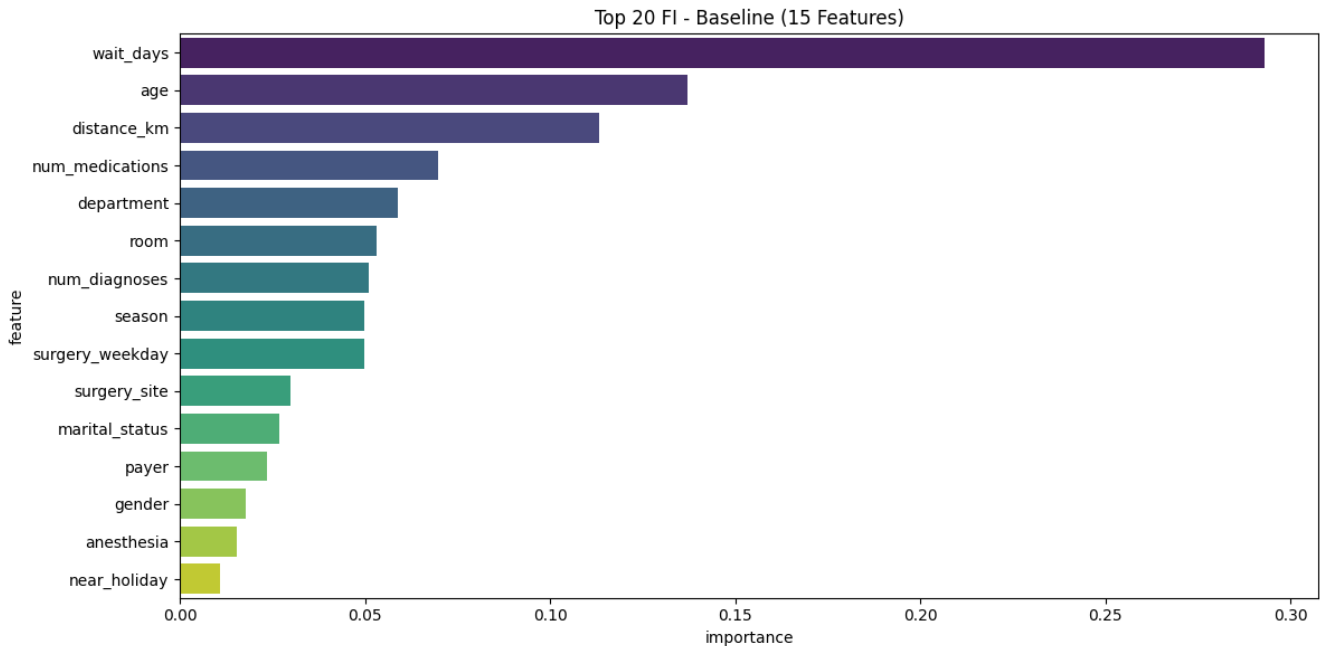
תהליך הנדסת התכונות היה שלב קריטי שאפשר למודלים ללמוד דפוסים מורכבים יותר מהמידע הגולמי, ושיפר באופן משמעותי את יכולת החיזוי.

פרק 3: פיתוח מודלים והערכת ביצועים

3.1 מודל הבסיס (Random Forest)

כנקודת ייחוס להשוואה פותח ואומן מודל מסוג random forest אשר מבוסס על סט של 15 משתנים שנבחרו לאחר שלב מקדים של הנדסת משתנים וניסויים ראשוניים. משנתים אלו זוהו כבעלי פוטנציאל החיזוי הגבוהה ביותר מתוך כלל המידע ולכן מהווים את נקודת המוצא שלנו לשיפור המודל.

בחינת תרשים חשיבות המשתנים במודל זה גילתה כי המשתנים שהשפיעו במידה הרבה ביותר על החיזוי היו 'זמן המתנה לניתוח', 'גיל המטופל' ו'המרחק בקילומטרים' של המטופל מבית החולים.

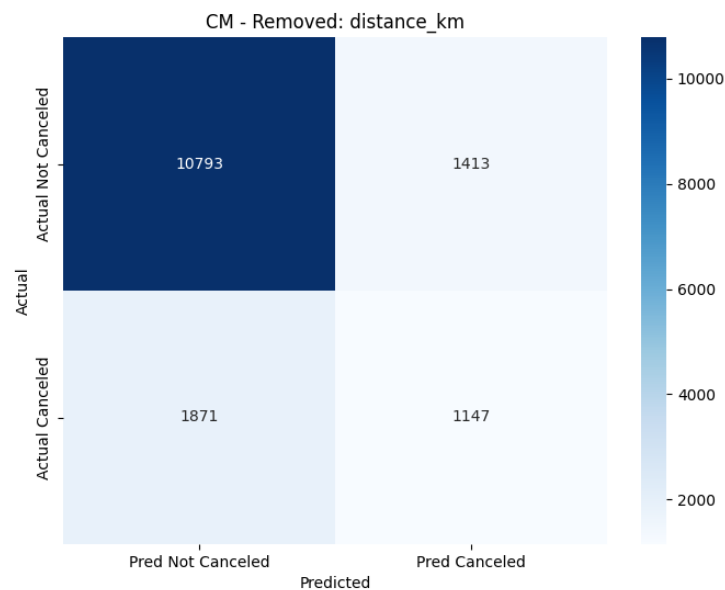
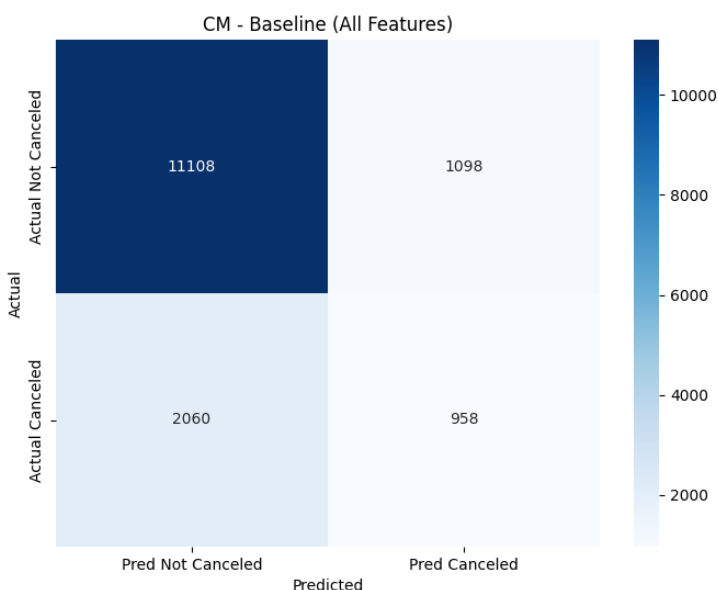


בחינת השפעת המשתנים על ביצועי המודל

כדי לבחון את חשיבותו של כל משתנה ולשפר את ביצועי המודל, נערך תהליך שיטתי של בחינת השפעת המשתנים. תהליך זה כלל אימון מחדש של המודל מספר פעמים, כאשר בכל פעם הוסר באופן יזום משתנה אחר מתוך סט המשתנים המקורי. מטרת כל אימון הייתה לבחון כיצד הסרתו של אותו משתנה בודד משפיעה על יכולת החיזוי של המודל, בהשוואה למודל הבסיס.

מסקנה: המודל האופטימלי לזיהוי ביטולים

ניתוח התוצאות שהתקבלו מתהליך זה הוביל למסקנה חד-משמעית. המודל שהציג את הביצועים הטובים ביותר למטרת המחקר, דהיינו זיהוי מקדים של ביטולי ניתוחים, הוא המודל שאומן ללא המשתנה 'המרחק בקילומטרים' של המטופל מבית החולים. הסרתו של משתנה זה הביאה לשיפור משמעותי ביכולת המודל לזהות נכון מקרים של ביטולים, וזאת מבלי לפגוע באופן קריטי במדדים אחרים.



השוואה כמותית בין מודל הבסיס למודל המשופר

מדד הרגישות – (Recall)

זהו הממד החשוב ביותר למטרה שלנו, שכן הוא בוחן איזה אחוז מכלל הביטולים שהתרחשו בפועל, המודל אכן הצליח לזהות. בממד זה נרשם השיפור המשמעותי ביותר. המודל המשופר הציג רגישות של 38%, לעומת 31.7% בלבד במודל הבסיס. המשמעות היא יכולת טובה בהרבה לאתר מטופלים בסיכון.

מדד הדיוק בחיזוי - (Precision)

מדד זה עונה על השאלה "כאשר המודל חזה 'ביטול' באיזו מידה הוא צדק?". כאן נרשמה ירידה קלה, מ-46.6% במודל הבסיס ל-44.8% במודל המשופר. ירידה זו מייצגת את ה"מחיר" ששולם עבור השיפור ברגישות, המודל החדש מייצר מעט יותר התראות שווא.

F1-score

זהו מדד משולב, המציג את האיזון בין הרגישות לדיוק. העובדה שממד זה עלה באופן משמעותי מ-0.377 ל-0.411 מוכיחה באופן סופי את עדיפות המודל החדש.

XGBoost 3.2

מודל זה תוכנן במיוחד כדי למקסם את יכולת זיהוי הביטולים, על ידי בחירה מכוונת של סף ההחלטה נמוך (0.15). התוצאה היא מודל "אגרסיבי", המציג את יכולת הזיהוי הגבוהה ביותר מבין המודלים שנבחנו. השימוש במודל זה ולאחר מכן כיוולו המדויק, הניב את המודל האפקטיבי והמאוזן ביותר למטרת המחקר.

הישגו המרכזי והמשמעותי ביותר של המודל הוא ביכולת הזיהוי הגבוהה שלו. המודל הצליח לזהות נכונה 71.5% מכלל הביטולים האמיתיים. זיהוי של יותר משני שלישים מהמקרים שבהם צפוי ביטול מאפשר לבית החולים לפעול מראש ולבצע התערבות רחבה שמסייעת בצמצום היקף התופעה.

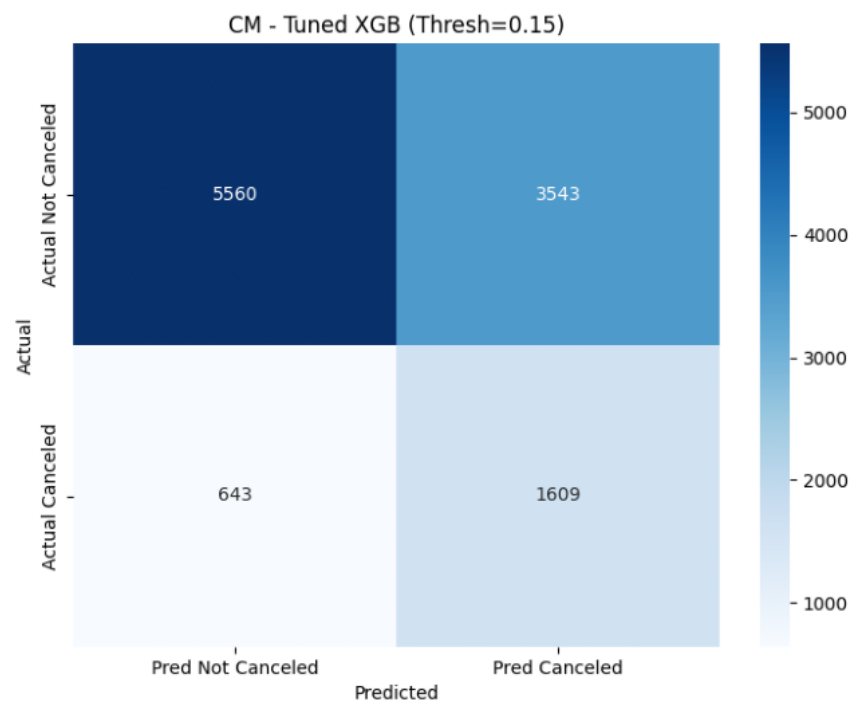
המחיר: ריבוי התראות שווא.

כצפוי ממודל "אגרסיבי" כל כך, ההצלחה בזיהוי גבתה מחיר משמעותי של התראות שווא. המודל חזה באופן שגוי 3,543 מקרים כביטולים, בעוד שבפועל המטופלים הגיעו. משמעות הדבר היא שכאשר המודל התריע על ביטול, הוא טעה ב-68.8% מהמקרים. זהו מחיר תפעולי גבוה, הדורש משאבים לבדיקת התראות רבות שאינן נכונות.

סיכום:

הבחירה במודל זה מייצגת החלטה אסטרטגית, המעדיפה לאתר כמה שיותר ביטולים אמיתיים, גם אם הדבק כרוך בבדיקת מספר גדול של התראות שגויות. ניתוח חשיבות המשתנים מראה כי המודל מסתמך בעיקר על נתוני 'המחלקה המנתחת' ו'זמן המתנה לניתוח' כדי לקבל את החלטותיו.

Confusion Matrix



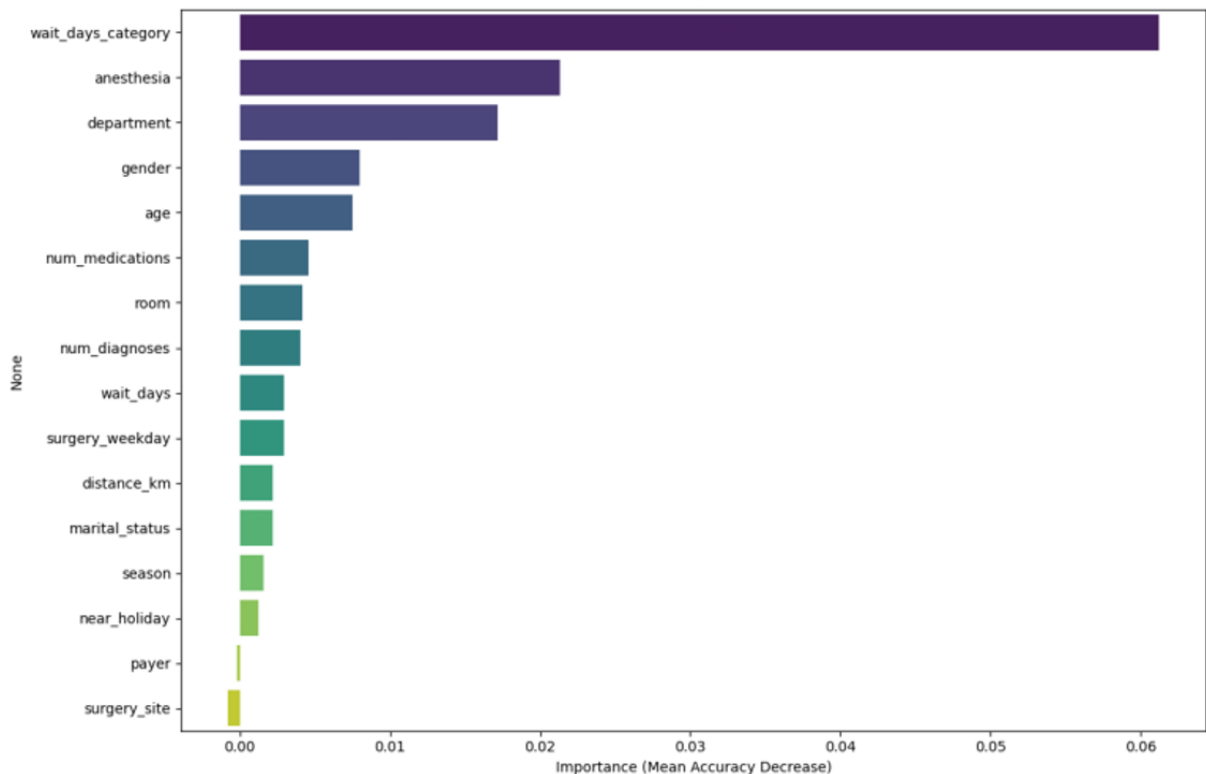
3.3 רשת נוירונים (Neural Network)

הניסיון לבחינת השימוש במודל רשת נוירונים הניב תוצאות מעורבות אך מעוררות מחשב, המדגימות את היתרונות והחסרונות של גישה זו. באופן כללי, ביצועי המודל במדדים כמו accuracy היו נמוכים יותר, אך הוא הציג הצלחה יוצאת דופן במשימה החשובה ביותר : זיהוי ביטולים.

הצלחה יוצאת דופן בזיהוי ביטולים

המודל הצליח לזהות 64.7% מכלל מקרי הביטול האמיתיים לעומת 38% של המודל הקודם. מבחינה תפעולית זהו שיפור קריטי המאפשר לאתר קבוצה גדולה משמעותית של מטופלים בסיכון. אולם, ההצלחה בזיהוי ביטולים גבתה מחיר כבד של "זהירות יתר". כאשר המודל התריע על מטופל שעומד לבטל ניתוח ב-66% מהמקרים הוא טעה, המטופל בפועל הגיע לניתוח.

לסיכום, מודל רשת הנוירונים אינו מהווה שיפור כולל על פני מודל Random Forest הבחירה בו תהיה תלויה בשאלה האם בית החולים מוכן "לשלם" את המחיר התפעולי של בדיקת התראות שווא רבות, בתמורה ליכולת לאתר מספר גדול יותר של ביטולים אפשריים. מעניין לציין שהמודל העניק חשיבות גבוהה יותר למשתנים אחרים כמו 'זמן המתנה' ו'סוג הרדמה', מה שמראה כי הרשת נוירונים למדה דפוסים שונים.



3.4 רשת נוירונית עם שכבות הטמעה

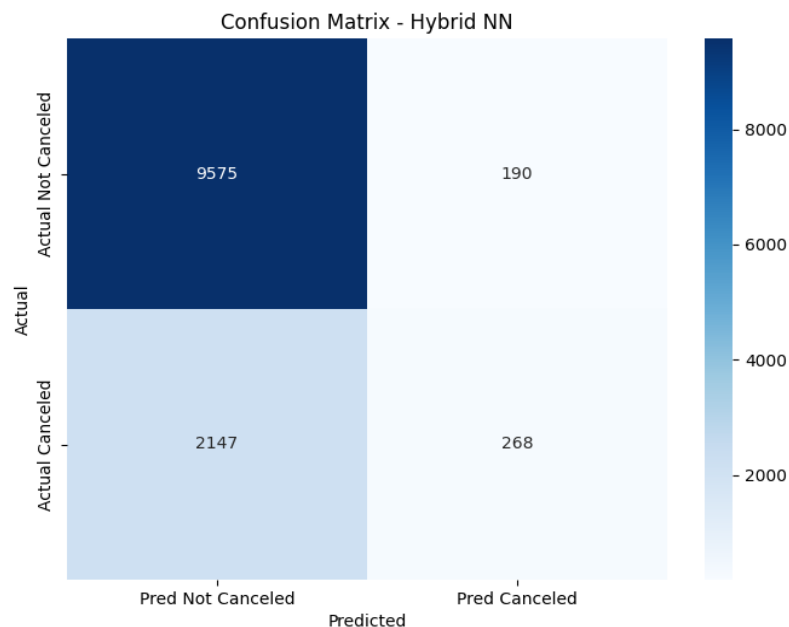
ארכיטקטורת רשת נוירונית עם שכבות הטמעה היא גישה מתקדמת להתמודדות עם משתנים קטגוריאליים (כמו שם מחלקה או סוג ניתוח). במקום להמיר משתנים אלו למספרים בודדים וחסרי הקשר, שיטה זו מתרגמת כל קטגוריה ייחודית לייצוג מספרי רב-ממדי, המכונה וקטור הטמעה. הייחוד בגישה זו הוא שערכי הווקטור אינם נקבעים מראש, אלא נלמדים על ידי המודל עצמו במהלך תהליך האימון. תהליך הלמידה מאפשר למודל לזהות ולכמת קשרים מורכבים ודמיון בין קטגוריות שונות, בהתבסס על השפעתן על תוצאת החיזוי. באופן זה, הרשת יכולה להגיע להבנה עמוקה יותר של הנתונים.

הניסיון להשתמש במבנה היברידי מורכב יותר, המשלב רשת נוירונים עם שכבות הטמעה, הניב תוצאות שהן כמעט תמונת ראי למודל רשת הנוירונים הקודם. על פניו, המודל נראה מוצלח עם רמת דיוק כללית גבוהה, אך ניתוח מעמיק חושף כי הוא אינו עונה על המטרה המרכזית של המחקר, ובמובנים רבים אף נכשל בה.

שמרנות יתר': כישלון בזיהוי המטרה'

עולה כי המודל ההיברידי הצליח לזהות 9.6% בלבד מכלל מקרי הביטול האמתיים. מודל אשר מפספס 9 מתוך כל 10 מטופלים שמבטלים אינו יכול לשמש ככלי יעיל להתערבות מקדימה.

לסיכום, המודל ההיברידי מהווה דוגמה הפוכה למודל רשת הנוירונים. בעוד שהמודל הקודם היה "פזיז" וזיהה מקרים רבים במחיר של טעויות רבות, המודל ההיברידי הוא "שמרן" יתר על המידה, ומתוך רצון להימנע מטעות, הוא מחמיץ כמעט את כל המקרים שהוא נועד לאתר. תוצאה זו מדגימה כי הוספת מורכבות למודל אינה מבטיחה שיפור בביצועים, ובמקרה זה אף הובילה למודל שאינו שימושי למטרה שהוגדרה.



3.5 השוואת ביצועי המודלים

מדד	Random Forest	Hybrid NN	XGBoost
Accuracy	78.43%	80.80%	63.14%
ROC AUC	0.75	0.75	0.72
PR AUC	0.42	0.41	0.37
Precision	44.80%	60.00%	31.23%
Recall	38.01%	10.00%	71.45%
F1-score	0.4113	0.1600	0.4346

הטבלה שלפנינו משווה בין שלושת המודלים שנבחנו במטרה לחזות האם מטופל לא יתייצב לניתוח. מבין השלושה Random Forest הציג את הביצועים המאוזנים ביותר, הוא מצליח לאתר חלק משמעותי מהמטופלים שלא יגיעו, תוך שמירה על רמת דיוק טובה וללא התראות שווא רבות.

מודל XGBoost מזהה יותר מקרים של אי-הופעה, אך במחיר של ריבוי טעויות. הוא יכול להתאים כאשר יש רצון לבצע התערבות רחבת היקף, כמו שליחת תזכורות אוטומטיות או מסרונים למספר רב של מטופלים.

מודל רשת נוירונים לא הצליח להתמודד עם האתגר, למרות אחוז דיוק כללי גבוה הוא כמעט ולא זיהה בפועל את מי שלא יגיע, ולכן אינו מתאים למשימה.

פרק 4: סיכום, מסקנות והמלצות

4.1 סיכום כללי

הפרויקט הנוכחי התמודד עם האתגר המשמעותי של ביטולי ניתוחים מתוכננים, תופעה בעלת השלכות כלכליות, תפעוליות ורפואיות נרחבות. באמצעות ניתוח של בסיס נתונים מקיף המונה כ-63,000 רשומות מבית החולים מאיר, פותחו, אומנו ונבחנו מספר מודלים של למידת מכונה במטרה לחזות מראש מטופלים הנמצאים בסיכון גבוה לאי-התייצבות לניתוח. הממצא המרכזי של המחקר הוא שאף מודל אינו מספק פתרון מושלם בפני עצמו, אך שילוב אסטרטגי בין מודלים שונים מאפשר יצירת מערכת חיזוי יעילה ומאוזנת.

4.2 מסקנות ותובנות מרכזיות

אין פתרון קסם אחד: האיזון בין רגישות לדיוק (Recall vs. Precision)

ניתוח השוואתי הראה באופן מובהק את הטרייד-אוף הקיים בין היכולת לזהות כמה שיותר ביטולים אימיתיים לבין המחיר של יצירת התראות שווא. מודלים אגרסיביים הצטיינו בזיהוי, אך במחיר תפעולי של בדיקת התראות שווא. ממצא זה מדגיש את הצורך בבחירת מודל המתאים למטרת התערבות ספציפית.

חשיבותה של הנדסת תכונות

המשתתפים שפותחו בתהליך הנדסת התכונות, כגון "מרחק מבית חולים", "מספר תרופות" ו"זמן המתנה לניתוח". התגלו כבעלי כוח החיזוי הגבוה ביותר באופן עקבי בכל המודלים. תובנה זו מאששת את ההשערות הראשוניות ומדגישה כי הגורמים המשפיעים ביותר על ביטול קשורים ללוגיסטיקה של המטופל ולמורכבותו הרפואית.

מורכבות אלגוריתמית אינה מבטיחה הצלחה

הניסוי עם מודל רשת הנוירונים ההיברידית, שהיה המורכב ביותר מבחינה ארכיטקטורה, הניב את התוצאות החלשות ביותר בממדד החשוב ביותר ביטולים Recall המודל הפך שמרני מדי וכמעט לא זיהה כלל. מסקנה זו מלמדת כי הוספת מורכבות למודל אינה מבטיחה שיפור בביצועים, ולעיתים פתרונות פשוטים ויציבים הם האפקטיביים ביותר.

4.3 המלצות להמשך מחקר ופיתוח

טיפול בחוסר האיזון בנתונים

האתגר המרכזי במחקר היה היחס הלא-מאוזן בין הניתוחים שהתקיימו לבין אלו שבוטלו. כדי לאפשר למודלים ללמוד בצורה אפקטיבית יותר את הדפוסים של קבוצת המיעוט (הביטולים), מומלץ לפעול לאיסוף רשומות נוספות של ניתוחים שבוטלו. הגדלת כמות הדוגמאות הללו תאפשר יצירת בסיס נתונים מאוזן יותר, מה שעשוי להוביל לשיפור ישיר במדדי הריגושת והדיוק של המודלים.

הוספת מידע חדש

ניתן להעמיק את יכולת החיזוי על ידי שילוב נתונים ממקורות נוספים. לדוגמה, הוספת משתנים סוציו-אקונומיים עשויה לחשוף קשרים ודפוסים נסתרים המשפיעים על התנהגות המטופלים, ובכך לשפר משמעותית את ביצועי המודל.

4.4 המלצות אופרטיביות ליישום במרכז הרפואי

מסקנות המחקר מצביעות על כך שאין מודל אחד שהוא מושלם, אך ניתן לרתום את החוזקות של כל מודל כדי ליצור אסטרטגיית התערבות דו-שלבית, חכמה ויעילה. במקום לבחור במודל אחד בלבד, ההמלצה היא לשלב בין מודל Random Forest לבין מודל XGBoost.

שלב 1: התערבות רחבה באמצעות מודל XGBoost

מודל זה שהוכיח יכולות זיהוי מרשימות (מעל 71%) ישמש כמנגנון הסינון הראשון. מודל זה יופעל על כלל רשימת המטופלים ויאתר את הקבוצה הגדולה של מטופלים הנמצאים בסיכון לביטול. למטופלים אלו תתבצע התערבות ברמה נמוכה כגון שליחת מסרון תזכורת על התור בזמן שהוגדר מראש.

שלב 2: התערבות ממוקדת באמצעות מודל Random Forest

מודל זה שהוכיח דיוק חיזוי גבוה יותר ישמש לזיהוי קבוצת המטופלים בסיכון הגבוה והוודאי ביותר. כאשר מודל זה מתריע על ביטול, הסבירות שהוא צודק גבוהה יותר. לכן, עבור קבוצה מצומצמת זו מומלץ לבצע התערבות בעלות גבוהה יותר, כגון שיחת טלפון אישית עם המטופל.

שילוב אסטרטגי זה מנצל את הטוב משני העולמות: הוא מבטיח כיסוי רחב לזיהוי מקרי ביטול רבים ככל האפשר, תוך מיקוד המשאבים היקרים (זמן כוח אדם) באותם מקרים בהם ההתערבות צפויה להיות האפקטיבית ביותר.