

# תרגיל בית 1 – דו"ח הגשה

מגישים: יונתן קויפמן, מעיין אייטק, דניאל מריח

ת.ז.: 316111442 ,206713612 ,212984801

מועד הגשה: 11/05/2023

קישור לגיטהאב: [https://github.com/maayan-aytek/DS\\_LAB\\_hw1](https://github.com/maayan-aytek/DS_LAB_hw1)

# תוכן עניינים

מבוא	3
ניתוח וחקירת נתונים	3-8
תיאור הפיצ'רים ובדיקת התפלגויות	3-5
קורלציות וקשרים	5-6
ערכים חסרים והטיפול בהם	7
חוסר איזון בנתונים	7-8
הנדסת פיצ'רים	8-10
בחירת פיצ'רים עבור המודל	8-9
טרנספורמציות על פיצ'רים	9-10
חיזויים	10-14
הקדמה	10
XGBOOST	11-12
Fully Connected Neural Network	12-13
Random Forest	13-14
סיכום	14
ביבליוגרפיה	15
נספחים	16-19

- חקרנו את הנתונים מכל ההיבטים – סטטיסטיקה תיאורית, התפלגויות, קורלציות חשובות בין פיצ'רים שונים או בין הפיצ'רים ללייבל, פילוח של חולים/לא חולים בתתי-קבוצות למיניהן.
- טיפלו בערכים החסרים בצורה מאוד יצירתית, המתחשבת בין היתר בהתפלגות הערכים הקיימים ובאחוזי ה-outliers של כל פיצ'ר, וגם בחוסר האיזון שיש בדאטה בין כמות החולים לבין כמות הבריאים.
- קראנו מאמרים מרתקים על אלח דם ועל האופן שבו מתמודדים עימו בעולם, ובמסגרת הקריאה מצאנו מספר מדדים שקשורים למחלה – SOFA, SIRS, qSOFA, ו-BUN/CREATININE ratio – מה שהוביל אותנו לבצע טרנספורמציות ויצירת פיצ'רים חדשים, אשר קיבלו ציון גבוה במדד ה-feature importance של המודלים. כמו כן, עשינו מבחן השערות על אחד מהפיצ'רים הנ"ל.
- ניסינו מגוון מודלים ואלגוריתמי למידה – מבוססי עצים, boosting ומבוססי רשתות נוירונים ולמידה עמוקה, וחקרנו איזה פיצ'רים והיפר-פרמטרים מניבים את תוצאות החיזוי האופטימליות. יתרה מזאת, הערכנו את הביצועים של המודלים על סמך מספר מדדים ואף ניתחנו את תהליך קבלת ההחלטות והשיקולים שלפיהם סיווגו את הדאטה.

## 2. ניתוח וחקירת נתונים

### 2.1. תיאור הפיצ'רים ובדיקת התפלגויות:

להלן תיאור ותרשימי התפלגויות רק של הפיצ'רים שבחרנו בהם עבור המודל, כפי שיפורט בסעיף 3.1.

**ניתן למצוא בנספחים 1 ו-7 את תרשימי ההתפלגויות של כלל הפיצ'רים ואת הטבלה המלאה (בהתאמה).**

סוג פיצ'ר	שם פיצ'ר	טיפוס	תיאור	סטטיסטיקה תיאורית
סימנים חיוניים	HR (Heart Rate)	Float	כמות פעימות הלב בדקה	מינימום: 20, מקסימום: 280 ממוצע: 84.66, חציון: 84
	O2Sat	Float	ריוויין חמצן בדם	מינימום: 20, מקסימום: 100 ממוצע: 97.19, חציון: 98
	Temp	Float	טמפרטורת הגוף (צלזיוס)	מינימום: 20.9, מקסימום: 50 ממוצע: 36.9, חציון: 36.5
	SBP	Float	לחץ הדם בעת כיווץ שריר הלב	מינימום: 20, מקסימום: 299 ממוצע: 123.51, חציון: 80
	MAP	Float	ממוצע משוקלל של לחצי הדם לאורך מחזור קרדיאלי אחד	מינימום: 20, מקסימום: 300 ממוצע: 82.29, חציון: 80
ערכי מעבדה	BaseExcess	Float	מדד לעודף הביקרבונאט	מינימום: -32, מקסימום: 49.5 ממוצע: -0.74, חציון: 0
	HCO3	Float	כמות הביקרבונאט	מינימום: 0, מקסימום: 55 ממוצע: 24.08, חציון: 24
	FiO2	Float	אחוז החמצן הנצרך	מינימום: -50, מקסימום: 4000 ממוצע: 0.58, חציון: 0.5
	pH	Float	רמת החומציות	מינימום: 6.62, מקסימום: 7.78 ממוצע: 7.37, חציון: 7.38
	PaCO2	Float	הלחץ של פחמן דו-חמצני בדם	מינימום: 10, מקסימום: 100 ממוצע: 41.16, חציון: 40
	SaO2	Float	ריווי החמצן בדם העורקי	מינימום: 23, מקסימום: 100 ממוצע: 92.51, חציון: 97
	AST	Float	ריכוז האנזים AST בדם	מינימום: 5, מקסימום: 9961 ממוצע: 264.5, חציון: 40
	BUN	Float	ריכוז האוריאיה בדם	מינימום: 1, מקסימום: 268 ממוצע: 23.55, חציון: 17
	Alkalinephos	Float	מדד לאנזים הכבד	מינימום: 7, מקסימום: 2528 ממוצע: 101.3, חציון: 73
	Calcium	Float	כמות הסידן בדם	מינימום: 1, מקסימום: 27.9 ממוצע: 7.54, חציון: 8.25

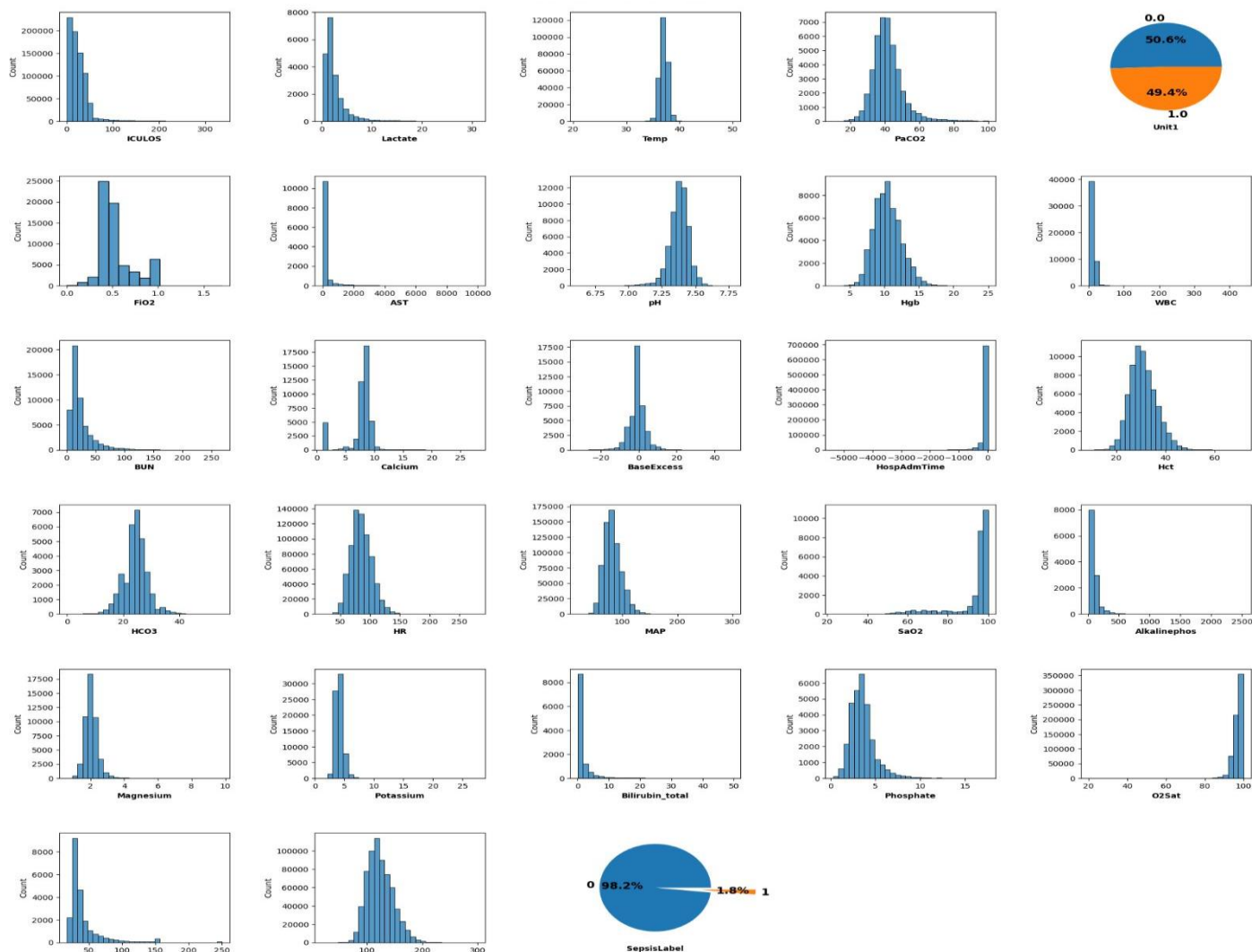
מינימום: 0.2, מקסימום: 31 ממוצע: 82.68, חציון: 1.88	כמות הלקטוז בדם	Float	Lactate	
מינימום: 0.6, מקסימום: 9.8 ממוצע: 2.04, חציון: 2	כמות המגנזיום בדם	Float	Magnesium	
מינימום: 0.3, מקסימום: 17.6 ממוצע: 3.54, חציון: 3.3	כמות הפוספט בדם	Float	Phosphate	
מינימום: 1.3, מקסימום: 27.5 ממוצע: 4.13, חציון: 4.1	כמות הפוטסיום בדם	Float	Potassium	
מינימום: 0.1, מקסימום: 49.6 ממוצע: 1.98, חציון: 0.8	כמות הבילירובין הכוללת בדם	Float	Bilirubin_total	
מינימום: 8.8, מקסימום: 71.7 ממוצע: 30.81, חציון: 30.3	כמות תאי הדם האדומים בדם	Float	Hct	
מינימום: 2.6, מקסימום: 25 ממוצע: 10.43, חציון: 10.3	כמות ההמוגלובין בדם	Float	Hgb	
מינימום: 17.1, מקסימום: 250 ממוצע: 41.3, חציון: 32.4	הזמן בשניות עד להיווצרות כריש דם	Float	PTT	
מינימום: 0.1, מקסימום: 440 ממוצע: 11.45, חציון: 10.3	מספר כדוריות דם לבנות	Float	WBC	
ערך שכיח-0, שכיחות-50.6%	אגף לטיפולים רפואיים במחלקת הטיפול הנמרץ	int (בינארי קטגוריאל)	Unit1	נתונים דמוגרפיים
מינ': -5366.8, מקס': 17.34 ממוצע: -50.95, חציון: -5.95	הפרש השעות בין זמן הרישום בבי"ח לזמן הרישום בטיפול הנמרץ	Float	HospAdmTime	
מינימום: 1, מקסימום: 336 ממוצע: 26.57, חציון: 21	משך שהייה בטיפול הנמרץ	int (שלם חיובי)	ICULOS	
ערך שכיח-1, שכיחות-98.2%	האם המטופל סובל מאלח דם	int (בינארי קטגוריאל)	SepsisLabel	תוצאה

ההתפלגויות של הפיצ'רים הרציפים חושבו על סמך כל הרשומות, בעוד שהפיצ'רים הקטגוריאלים – מגדר ו-UNIT1 – חושבו

פר מטופל, מפני שמדובר בערכים קבועים לכל מטופל. באשר ללייבל, רצינו לבדוק מה היחס בין כמות המטופלים שאובחנו

כחולים לבין כמות המטופלים הבריאים, ולכן גם יחס זה חושב פר מטופל ולא לפי כמות רשומות.

Distributions of the columns

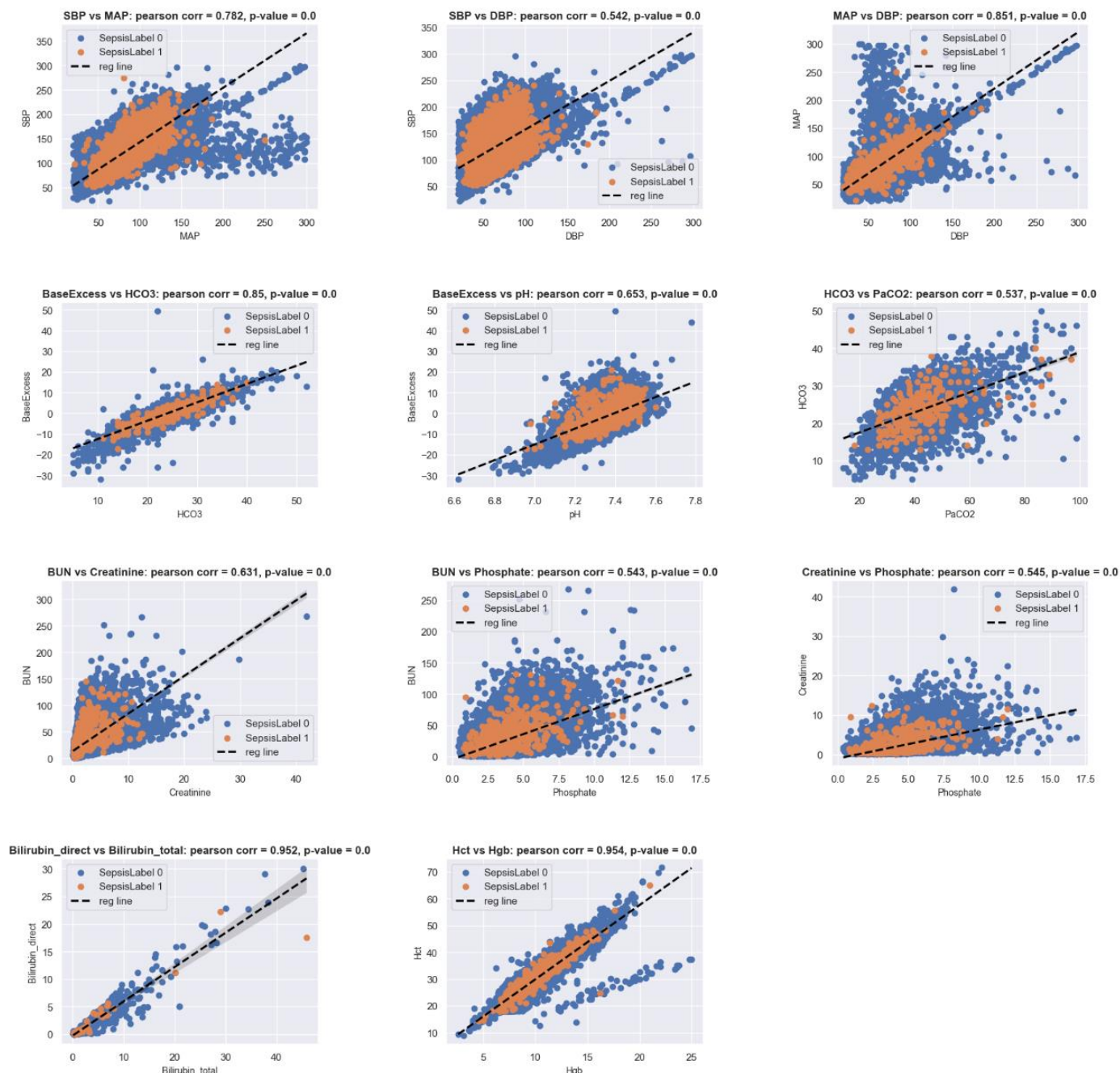


ניתן לראות כי חלק מהפיצ'רים מתפלגים (בקירוב) נורמאלית או חי-בריווע, וניתן אף לשים לב לחוסר האיזון המובהק בדאטה בין המטופלים החולים לבין המטופלים הלא חולים בתרשים העוגה הרלוונטי.

## 2.2. קורלציות וקשרים:

חקרנו ולמדנו תופעות מדעיות-ביולוגיות מעניינות וקשרים שעשויים להיות בין הפיצ'רים השונים (שהם מדדים ביולוגיים של מטופל). בהתחשב בכך שערכי הפיצ'רים הינם רציפים, ביצענו מבחן פירסון לקורלציות וקיבלנו את התוצאות הבאות:

Correlations between features

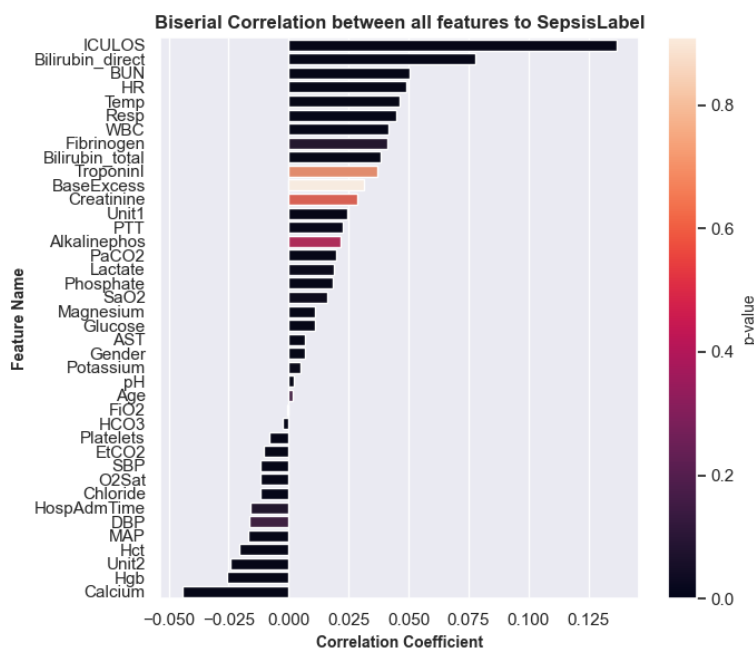


התוצאות שקיבלנו הגיוניות, הרי שישנם קשרים מדעיים מובהקים בין חלק מהפיצ'רים. כך למשל – MAP, SBP ו-DBP קשורים ללחץ הדם של המטופל ובמאמר [1] אף מציגים נוסחה לאמידת MAP על ידי SBP ו-DBP (ראה נוסחה בנספח 3), מה שיכול גם להעיד על ייתורם עבור למידת המודל. כמו כן,  $\text{HCO}_3$  ו-pH BaseExcess מספקים מדד לתפקודי

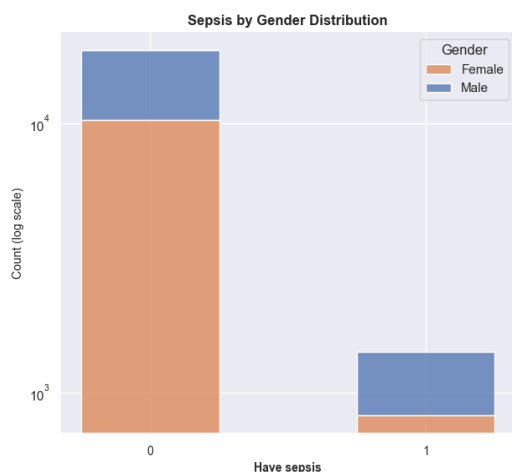
הכליות. בנוסף, Bilirubin\_direct הינו אחד המרכיבים של Bilirubin\_total (סכום של direct + indirect), ולפי מאמר [2], מאחר שתאי הדם האדומים (HCT) מהווים כלי קיבול עבור המוגלובין (HGB), קיים קשר ישיר בין השניים.

זאת ועוד, ניתן לראות בגרפים תופעה מעניינת נוספת שאינה קשורה לקורלציות דווקא, והיא אופן פיזור התצפיות של מטופלים אשר אובחנו עם אלח דם – קל לראות שבכל הגרפים הפיזור של התצפיות הנ"ל הינו מאוד מזערי וכי רוב התצפיות מקובצות באזור יחיד, מה שעשוי להעיד על קיומם של מאפיינים וערכים ספציפיים המצביעים על היווצרותו של אלח דם.

בנוסף לכך, בעקבות זיהוי התופעה המתוארת קודם לכן בתרשימים של הקורלציות, בדקנו האם קיימת קורלציה בין הפיצ'רים (כל אחד לכשעצמו) לבין קיומו של אלח דם, אולם קיבלנו קורלציות נמוכות מאוד, כאשר הגבוהה ביותר עמדה על 12.5 אחוזים. להלן התרשים –



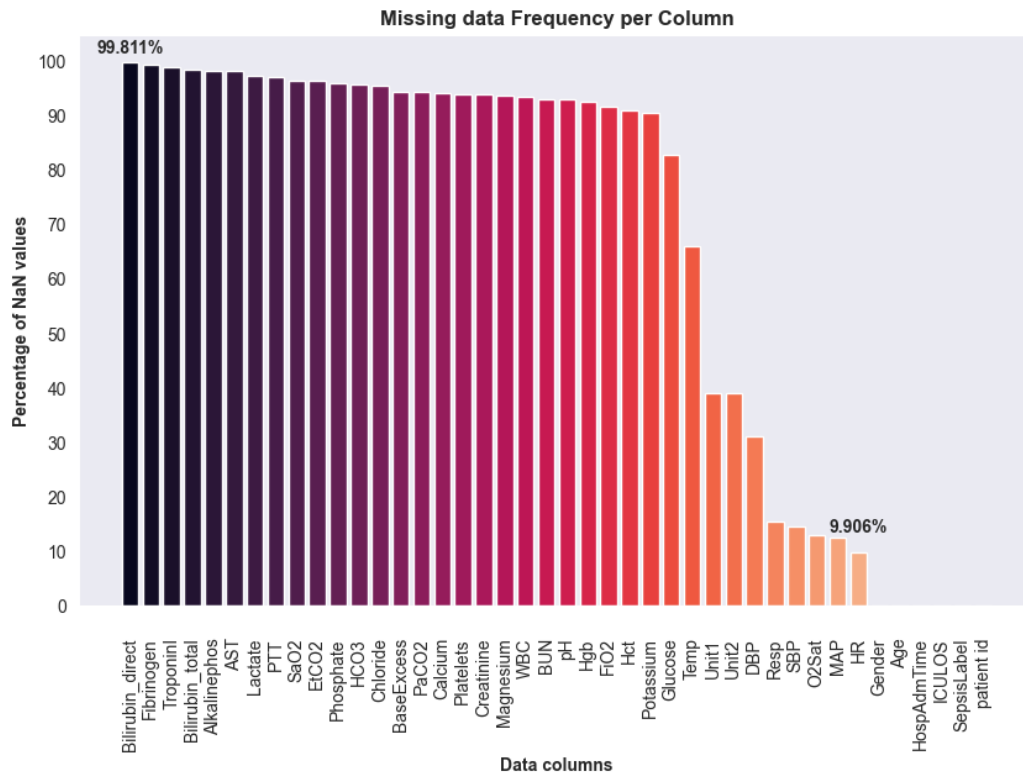
חקרנו גם את פילוח אוכלוסיית החולים והלא חולים על סמך נתונים דמוגרפיים, כגון מגדר. מצורף תרשים שבו ציר X מתבסס על עמודה שיצרנו, המכילה 1 אם המטופל חלה באלח דם (קרי, לפחות רשומה אחת מכילה ליבל 1) ו-0 אחרת, ובהתאם לכך התרשים מציג נתונים ללא כפילויות, אלא ככמות המטופלים. כמו כן, בציר Y עשינו סקאלת לוג על מנת שהגרף יהיה קריא, לאור הפער בין גודל הברים. להלן התרשים –



ניתן לראות בבירור כי פילוח האוכלוסייה של החולים והלא חולים שונה – בקרב הלא חולים ישנו רוב משמעותי של נשים לעומת גברים, ואילו באוכלוסיית החולים בדיוק ההפך.

### 2.3. ערכים חסרים והטיפול בהם:

בדקנו תחילה את אחוזי החסרות של כל אחד מהפיצ'רים בדאטה, כלומר, בכמה רשומות מתוך כלל הרשומות ערך הפיצ'ר חסר, וקיבלנו את התרשים הבא –



בגלל אחוזי החסרות הגבוהים בערכים של חלק מהפיצ'רים, ובמטרה לבצע השלמת נתונים חסרים עם ערכים פחות "רועשים", אימצנו תחילה את השיטה שתוארה במאמר [3] להתמודדות עם הנתונים החסרים והשגויים שהושאו במאגר – השמטת מטופלים עם פחות מ-36 רשומות, ששקולות ל-3 משמרות באורך 12 שעות למשמרת של מדידות; **מדובר בהשמטה שנעשתה לשם השלמה עם ערכים "מהימנים" יותר בלבד**, כלומר, מטופלים אלה כן הוחזרו בשלב האימון של המודל. לאחר מכן, ובהתחשב בכך שהרשומות של כל מטופל מסודרות לפי רצף כרונולוגי, יישמנו את שיטת ההשלמה backward/forward כשהתאפשר (כלומר, כאשר הערך של הפיצ'ר לא היה חסר ברשומה אחת לפני/אחרי). בהקשר הזה, הפעלנו תחילה את שיטת ה-back ורק אז את שיטת ה-forward מכיוון שהיה משמעות לסדר יישום שתי השיטות הנ"ל, ובסדר הזה הצלחנו להשלים יותר רשומות מאשר באופן הפוך. בהמשך לכך, עבור הרשומות שלא התאפשר להשלים בשיטה הנ"ל, השלמנו עם הממוצע או החציון לפי כלל ההשלמה הבא – אם אחוז ה-outliers בפיצ'ר מסוים היה קטן מ-5%, השלמנו עם הממוצע, אחרת השלמנו עם החציון (שאינו רגיש ל-outliers כמו ממוצע – מה שעשוי להבטיח השלמה עם ערך "תקין" יותר).

**ניתן למצוא בנספח 4 טבלה המציגה את אחוז ה-outliers בכל אחד מהפיצ'רים שנבחרו עבור מודל החיזוי.**

### 2.4. חוסר איזון בנתונים:

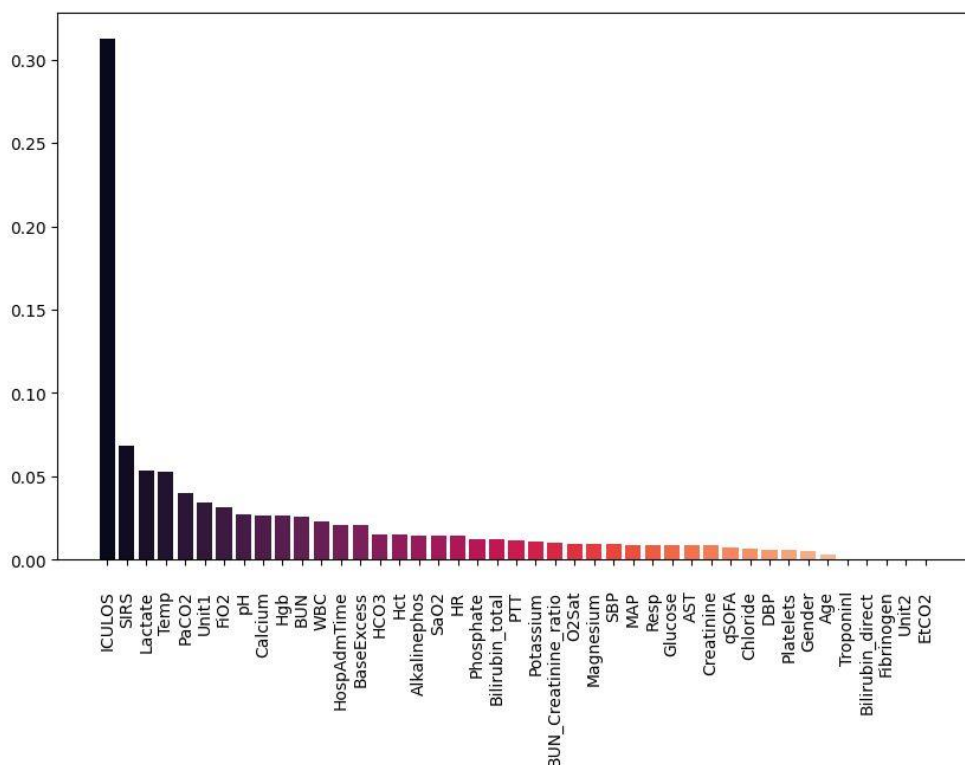
במסגרת העבודה עם הנתונים וחקירתם, הבחנו בכך שמאגר הנתונים אינו מאוזן וכי כמות המטופלים שאובחנו עם אלח דם קטנה באופן משמעותי לעומת כמות המטופלים הבריאים – יחס מקורב של 13 מטופלים בריאים למטופל אחד עם אלח דם. כפתרון לסוגיה הזו, ניסינו קודם לבצע גם Under-sampling של מטופלים בריאים, שזה להשמיט **באופן אקראי** רשומות של מטופלים בריאים וגם Over-sampling של מטופלים שאובחנו עם אלח דם, שזה להגדיל את כמות הרשומות של מטופלים שאובחנו עם אלח דם בצורה מלאכותית – הן ע"י יצירת רשומות סינתטיות על

סמך הרשומות הקיימות במאגר הנתונים והן ע"י דגימה מחדש. בדיעבד, לאחר שמצאנו כי ה-Over-sampling פוגע בחיזויים של המודל במקום לסייע, ביצענו רק Under-sampling וצמצמנו את היחס בין כמות המטופלים הבריאים לבין כמות המטופלים שאובחנו עם אלח דם מ-13:1 ל-5:1.

### 3. הנדסת פיצ'רים

#### 3.1. בחירת פיצ'רים עבור המודל

בחלק הקודם בחנו את כל הדאטה והפיצ'רים שעומדים לרשותנו בצורה סטטיסטית וכללית. כעת, נרצה לחקור לעומק ולמצוא את הפיצ'רים הרלוונטיים ומעניינים (על סמך מאפיינים ביולוגיים ומדעיים שעשויים להצביע על התהוותו של אלח דם) עבור החיזויים של המודל. במסגרת הדברים, קראנו את מאמר [4] שדן בהרחבה על המדדים SIRS ו-SOFA (ראשי תיבות של Systemic Inflammatory Response Syndrome ו-Sequential [Sepsis-related] Organ Failure Assessment, בהתאמה) – מדדים שנועדו לנסות לסייע באבחון מוקדם של אלח דם. במשך שנים רבות נעשה שימוש במדד SIRS מתוך אמונה שאלח דם נבע מתגובת SIRS לזיהום בגוף, וקביעת ערך המדד עבור מטופל התבססה על הפיצ'רים הבאים: Temp, HR, Resp, PaCO<sub>2</sub> ו-WBC. בהתאם לכך, ראינו בפיצ'רים הנ"ל כמועמדים רלוונטיים עבור מודל החיזוי. במעלה הדרך, ומתוך הבנה כי מדד SIRS לא היה אינדיקטור מספיק טוב, נוצר מדד SOFA – מדד שהעלייה בו נמצאה קשורה לסבירות גבוהה יותר של תמותה, ובפרט, מהווה אינדיקציה טובה יותר להיווצרותו של אלח דם. על כן, הפיצ'רים שמרכיבים את המדד הנ"ל מועמדים גם הם להיות רלוונטיים עבור המודל – FIO<sub>2</sub>, Creatinine, Bilirubin, Platelets ו-MAP. על מנת לבסס את בחירת הפיצ'רים שלנו ככל שניתן ולמצוא את הפיצ'רים החשובים ביותר עבור המודל, ביצענו feature importance על סמך מודל החיזוי XGBOOST (המודל שנבחר בסוף) וקיבלנו את התוצאה הבאה –

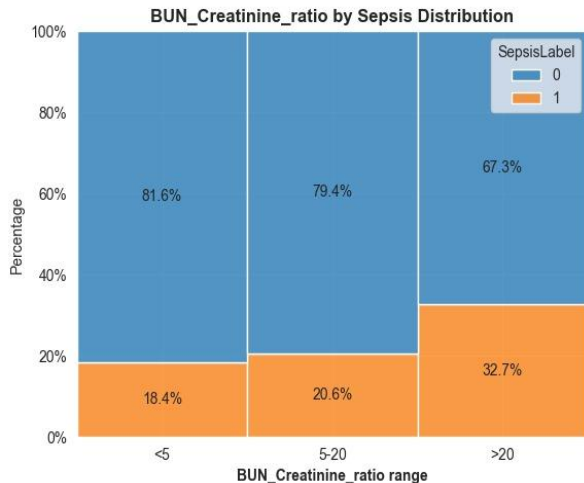


בהתבסס על כל הממצאים לעיל ומעט ניסוי וטעיה, בחרנו ב-29 הפיצ'רים המובילים במדד ה-feature importance, כלומר, מ-ICULOS ועד Resp כולל, מכיוון שקיבלנו את התוצאות הכי טובות איתם.



- הוספת מדד של היחס BUN/CREATININE, שלפי מאמר [5], עשוי להצביע על מחלה/נזק בכליות, דימום במעי – שהם סימפטומים שעלולים להופיע גם בעת התהוותו של אלח דם – ובעיות מערכתיות נוספות. במאמר הנ"ל, צוין כי הטווח התקין של הממד הינו בין 5 לבין 20, כך שממדים מתחת ל-5 או מעל 20 מצביעים על היתכנותה של מחלה או אי-תקינות מערכתית בלתי נראית לעין. בהתאם לכך, בדקנו את פילוח החולים

והבריאים בטווחים הללו וקיבלנו את התרשים הבא –



על אף שהמדד הנ"ל לא מצביע באופן ישיר על התהוותו של אלח דם, אלא כאמור מצביע על איזושהי אי-תקינות מערכתית שאולי קשורה לאלח דם, ניתן עדיין לראות בתרשים לעיל הלימה חלקית בין טווחי הממדים לבין אחוז החולים באלח דם בקרב המטופלים שנמצאים באותו הטווח.

⇐ **מבחן השערות:** נשער כי תוחלת הממד הנ"ל בקרב המטופלים הבריאים שונה מהתוחלת של המטופלים החולים, כאשר השערת האפס היא שהתוחלות שוות. כלומר, שערכי הממד אינם תלויים בקיומו או אי-קיומו של אלח הדם. על מנת לבחון את השערתנו, השתמשנו ב-T-test דו-צדדי, רק לאחר שווידאנו כי התנאים להפעלתו מתקיימים. תחילה, ברור כי אין תלות בין מטופלים שונים. לאחר מכן, הסתכלנו על התפלגות **לוג** הערכים ב-QQ Plot וקיבלנו כי הם מתפלגים נורמאלי (ראה **תרשים בנספח 6**), ומכיוון שמדובר בטרנספורמציה חד-חד ערכית ומונוטונית עולה ממש, הרי שתוצאות המבחן על לוג הערכים חלות גם על הערכים ללא הלוג.

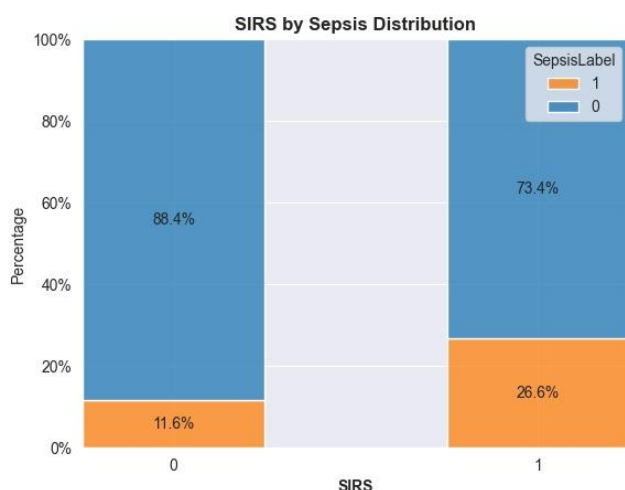
**תוצאות המבחן:** דוחים את השערת האפס ברמת מובהקות 95%.

$$P - Value = 1.84512 \cdot e^{-17} \approx 0$$

כלומר, ההשערה שלנו מתקיימת.

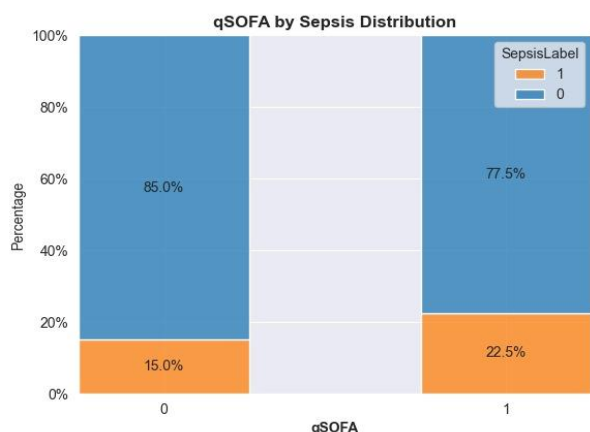
- הוספת מדד SIRS כפי שכבר תואר **בסעיף 3.1**, שחושב ע"פ הקריטריונים **שבנספח 2**, כאשר פצינט מקבל 1 אם מתקיימים לפחות שניים מהקריטריונים הנ"ל. בדומה לממד הקודם, גם כאן בדקנו את כמות החולים

והבריאים בכל אחת מהקבוצות וקיבלנו את התרשים הבא –



ניתן לראות בצורה מובהקת כי אחוז החולים שסווגו כלא חולים קטן יותר מאחוז החולים שסווגו כחולים לפי המדד, אך בצפוי המדד סיווג בצורה מחמירה גם מטופלים לא חולים כחולים – מה שמחזק את הנכתב בסעיף 3.1 בנוגע להיותו אינדיקטור לא מספיק טוב (בעיקר מחמיר יתר על המידה).

- הוספת מדד qSOFA שלפי מאמר [4], מהווה תחליף ל-SOFA ונועד לספק תוצאות מהר יותר (ומכאן השם Quick SOFA), אם כי מדויקות פחות בדרך כלל. המדד חושב לפי הקריטריונים שבנספח 5, כאשר פציינט מקבל 1 אם הוא מקיים את שני הקריטריונים הנומריים. בדקנו גם עבור מדד זה את התפלגות המטופלים החולים והבריאים בכל אחת מהקבוצות וקיבלנו את התרשים הבא –



גם כאן, הממצאים דומים לממצאים של מדד ה-SIRS, אולם פחות טובים בהתחשב בכך שאחוז גדול יותר של חולים סווג כלא חולים ע"פ המדד הזה. לא השתמשנו במדד זה בסוף לאור הציון הנמוך שקיבל ב-feature importance.

- כחלק מניסיונותנו לחלץ ולזקק את הנתונים והערכים האקספלנטיביים ביותר במטרה לסייע למודל בחיזוי ולמזער את הרעש, ניסינו עבור כל מטופל להחליף את העמודות של הפיצ'רים בעמודות שמכילות את הממוצע ואת השיפוע של ישר הרגרסיה על סמך 6 השעות האחרונות. כלומר, בכל רשומה יופיע עבור כל פיצ'ר שיפוע וממוצע שחושב לפי 6 הרשומות שלפניו. הבחירה ספציפית ב-6 שעות הייתה בהתאם לדרישת המשימה שהמודל יחזה התהוותו של אלח דם 6 שעות לפני. אולם, טרנספורמציה זו לא הניבה תוצאות מספקות, לצערנו, ולכן לא המשכנו איתה.

#### 4. חיזויים

- ביצענו אגרגציה על הרשומות כך שכל מטופל יוצג ע"י רשומה אחת בלבד. ספציפית, בדאטה של האימון השארנו לכל מטופל את הרשומה הראשונה שהייתה עם לייבל 1 במידה והייתה כזו, אחרת את הרשומה האחרונה. לעומת זאת, בדאטה של הטסט השארנו את הרשומה האחרונה.
- ב-Post Analysis של 3 האלגוריתמים, הרצנו את המודלים על 4 תתי-הקבוצות הבאות: גברים, נשים, מטופלים עד גיל 50 ומטופלים מגיל 50.
- במטריצות הבלבול שמצורפות בהמשך, ציר Y הוא ציר הלייבל האמיתי וציר X הוא ציר החיזוי.
- מצורפים מטה שלושת המודלים שהניבו את מדדי החיזוי הטובים ביותר, זאת לאחר שניסינו מודלים נוספים, כגון Bagging, Adaboost, Decision Tree ו-MLP (Multi-layer Perceptron).

i. The Algorithm: XGBOOST

ii. Hyperparameters & Regularization:

- n\_estimators = 200
- max\_depth = 50
- eta = 0.05
- gamma = 0.1
- reg\_lambda = 0.8
- min\_child\_weight = 2

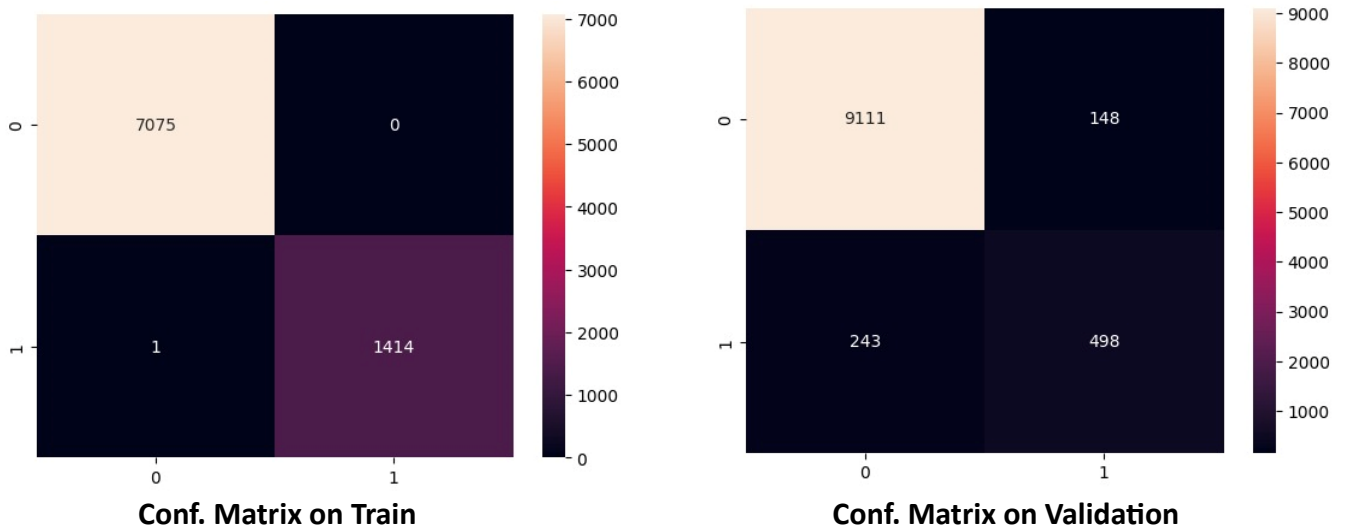
iii. Training & Validation Results:

**Training Results:**

- F1 Score: 0.9996
- Recall: 0.9993
- Precision: 1.0
- AUC-ROC: 0.9996

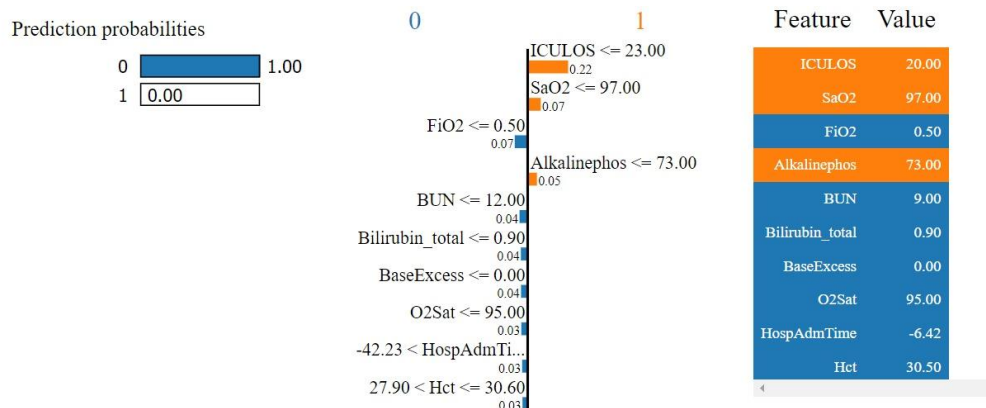
**Test Results:**

- F1 Score: 0.7181
- Recall: 0.6721
- Precision: 0.7709
- AUC-ROC: 0.828



iv. Post Analysis:

- F1 Score on Women: 0.7148
- F1 Score on men: 0.7203
- F1 Score on Above 50 Y.O: 0.7195
- F1 Score on Less or Equal 50 Y.O: 0.7134
- **Model's Interpretability:**



The picture above portrays how the model predicts a patient's label – the table on the right consists of the patient's measurements' values; The line in the middle shows the threshold values for each feature and the weight it contributes to the model in deciding the patient's label; The bars on the left are the probabilities of this patient getting 0 or 1 as label. For example, we can see that the model is 100% sure that this patient's label is 0.

## 4.2. אלגוריתם 2

### i. The Algorithm: Fully Connected Neural Network

### ii. Hyperparameters & Regularization:

- Input Layer  $\in \mathbb{R}^{29 \times 58}$
- Two Hidden Layers  $\in \mathbb{R}^{58 \times 116}$
- Output Layer  $\in \mathbb{R}^{116 \times 2}$
- Activation Function: ReLU
- Loss Function: Cross Entropy
- Optimizer: ADAM

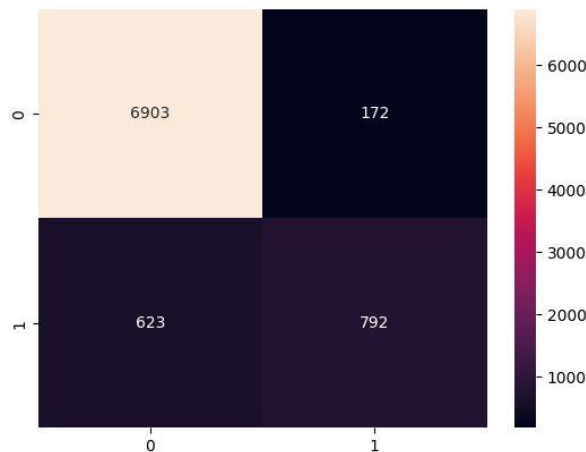
### iii. Training & Validation Results:

#### Training Results:

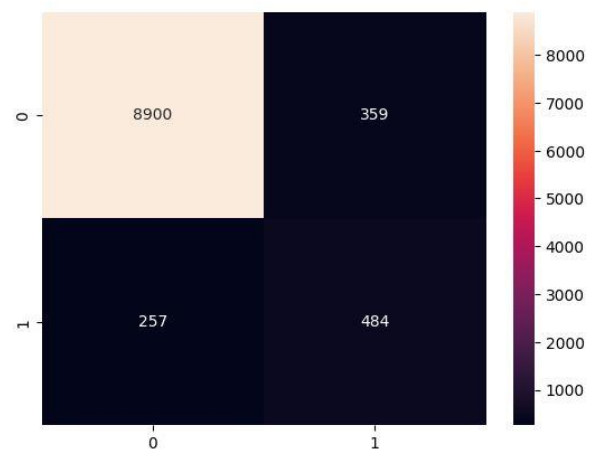
- F1 Score: 0.6658
- Recall: 0.5597
- Precision: 0.8216
- AUC-ROC: 0.7677

#### Validation Results:

- F1 Score: 0.6111
- Recall: 0.6532
- Precision: 0.5741
- AUC-ROC: 0.8072



Conf. Matrix on Train



Conf. Matrix on Validation

The results above were obtained after 2 EPOCHS.

### iv. Post Analysis:

- F1 Score on Women: 0.6021
- F1 Score on men: 0.6105
- F1 Score on Above 50 Y.O: 0.6169
- F1 Score on Less or Equal 50 Y.O: 0.5765
- **Model's Interpretability**: Based on what we've learned in ML2 and upon what is widely known, there are no scientific-based explanations nor proofs about neural networks' learning process; All we know is that it improves its predictions' results by updating weights

and minimizing gradients iteratively according to a given loss function. Thus, the model's interpretability is rather vague and unexplainable.

### 4.3. אלגוריתם 3

i. The Algorithm: Random Forest

ii. Hyperparameters & Regularization:

- n\_estimators = 200      - max\_depth = 25      - min\_samples\_leaf = 5

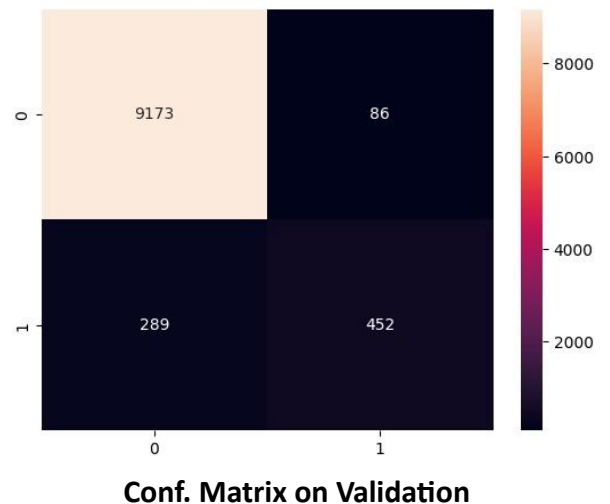
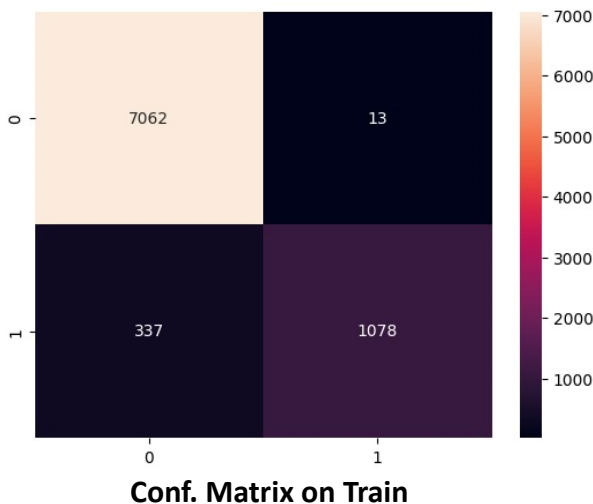
iii. Training & Validation Results:

**Training Results:**

- F1 Score: 0.8603      - Recall: 0.7618  
- Precision: 0.988      - AUC-ROC: 0.88

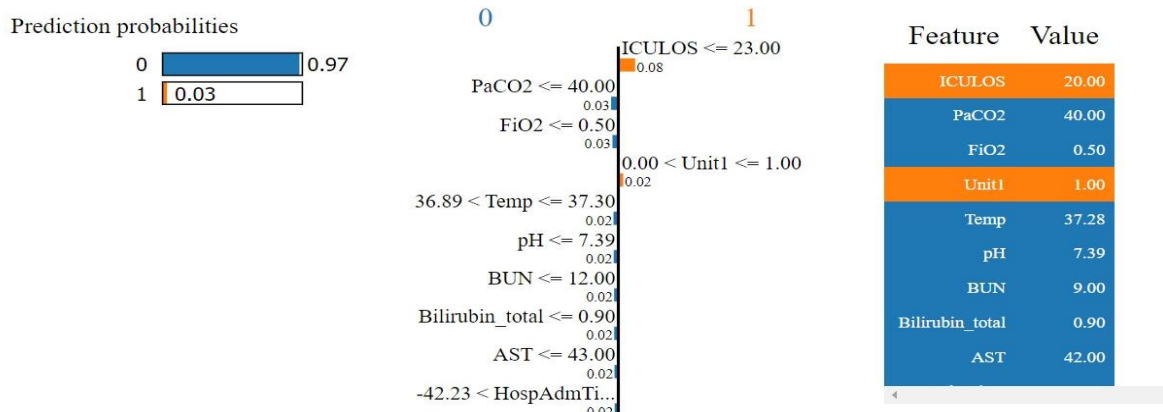
**Validation Results:**

- F1 Score: 0.7068      - Recall: 0.61  
- Precision: 0.8401      - AUC-ROC: 0.8003



iv. Post Analysis:

- F1 Score on Women: 0.7031      - F1 Score on men: 0.7093  
- F1 Score on Above 50 Y.O: 0.7049  
- F1 Score on Less or Equal 50 Y.O: 0.7129  
- Model's Interpretability:



The same explanation about XGBOOST's interpretability applies here as well. Nevertheless, we can notice here that in the example above the model is 97% sure that the patient's label is 0 and 3% that the label is 1.

**NOTE:**

1. We didn't use the accuracy measure, because in an imbalanced data set, assessing and improving a learning model based upon accuracy will lead the model to classify all data with the majority's label, guaranteeing a high accuracy score (that is equal to the percentage of data that is labeled with the majority's label).
2. Although XGBOOST has a higher type I error score on the validation set in comparison with Random Forest, it has a lower type II error score, which explains XGBOOST's higher F1 score. We chose XGBOOST and were willing to tolerate such tradeoff at the price of a higher F1 score.

**5. סיכום**

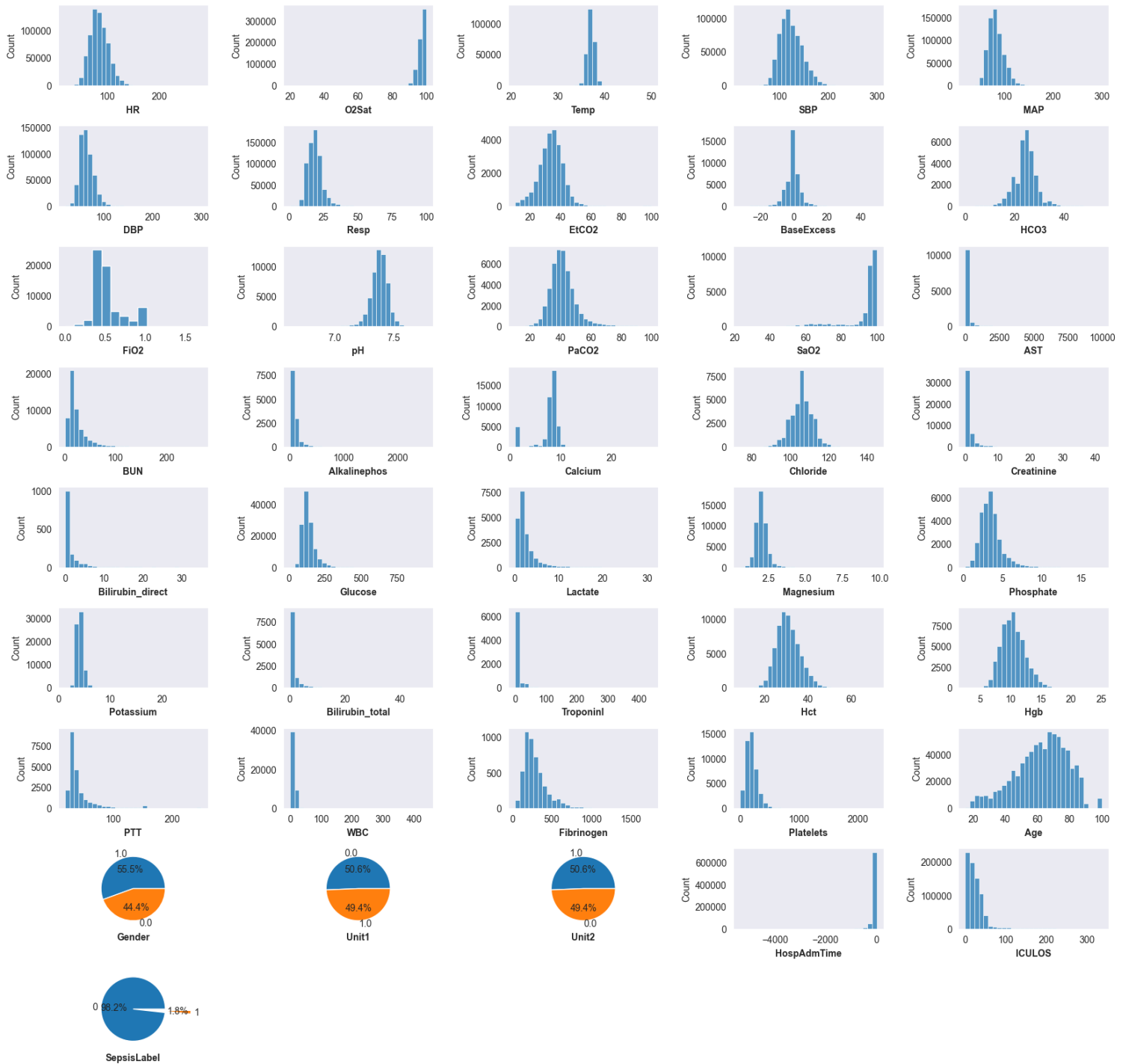
במסגרת העבודה שלנו, קיבלנו תוצאות המצביעות על קשר חזק בין משך השהייה במחלקת הטיפול הנמרץ לבין אלח דם, וגם בין מדד SIRS לבין המחלה. כמו כן, נראה שמודלי למידה מבוססי עצים התמודדו טוב יותר עם משימת הסיווג בהשוואה לשאר סוגי המודלים (הן רשתות עמוקות והן Boosting), ובמיוחד, המודל הסופי שבחרנו בו – XGBOOST – אשר הגיע לתוצאות דיוק טובות, באופן יחסי.

יתרה מזאת, חוסר האיזון בדאטה, אחוזי החסרות הגבוהים בפיצ'רים והמחסור במדדים נוספים שהיו צריכים להילקח מהמטופלים (כגון הקריטריון השלישי הנדרש לחישוב מדד ה-SOFA בצורה מדויקת) היוו מגבלות למחקר שלנו, וסביר להניח כי אילולא מגבלות אלו, יכולנו לקבל תוצאות טובות ומהימנות יותר.

לאור כל הנאמר לעיל, נמליץ כצעדי המשך על מחקר נוסף שיתבסס על דאטה שלם יותר ומגוון יותר (למשל, תחת ההנחה שהדאטה שקיבלנו מגיע מבית חולים אחד, רצוי לאסוף דאטה ממספר בתי חולים), לאסוף מדדים נוספים מהמטופלים ולהתמקד בעיקר במודלים מבוססי עצים.

- [1]** Sun J, Yuan J, Li B. SBP Is Superior to MAP to Reflect Tissue Perfusion and Hemodynamic Abnormality Perioperatively. *Front Physiol.* 2021
- [2]** WHAT IS THE CONNECTION BETWEEN HEMOGLOBIN AND HEMATOCRIT? (Hyperlink)
- [3]** Cruz, M.F., Ono, N., Huang, M. et al. Kinematics approach with neural networks for early detection of sepsis (KANNEDS). *BMC Med Inform Decis Mak* 21, 163 (2021).
- [4]** Singer, Mervyn, et al. "The third international consensus definitions for sepsis and septic shock (Sepsis-3)." *Jama* 315.8 (2016): 801-810
- [5]** BUN/Creatinine Ratio: High & Low Levels + Normal Range (Hyperlink)

Distributions of the columns



[1]

Two or more of:

Temperature  $>38^{\circ}\text{C}$  or  $<36^{\circ}\text{C}$

Heart rate  $>90/\text{min}$

Respiratory rate  $>20/\text{min}$  or  $\text{PaCO}_2 <32 \text{ mm Hg}$  ( $4.3 \text{ kPa}$ ) White blood cell count  $>12000/\text{mm}^3$  or  $<4000/\text{mm}^3$  or  $>10\%$  immature bands

[2]



$$\text{MAP} = \text{DBP} + 1/3 \times (\text{SBP} - \text{DBP})$$

[3]

Outliers %		Outliers %	
FiO2	21.98	WBC	3.31
HospAdmTime	13.91	pH	2.62
Creatinine	12.02	Temp	1.90
PTT	10.53	SepsisLabel	1.76
BUN	7.35	MAP	1.67
Lactate	6.81	Hct	0.91
Calcium	5.66	HR	0.87
PaCO2	5.47	Hgb	0.76
Glucose	5.32	Age	0.16
ICULOS	4.27	Unit1	0.00
Phosphate	3.94	Gender	0.00
Platelets	3.40	patient id	0.00

[4]

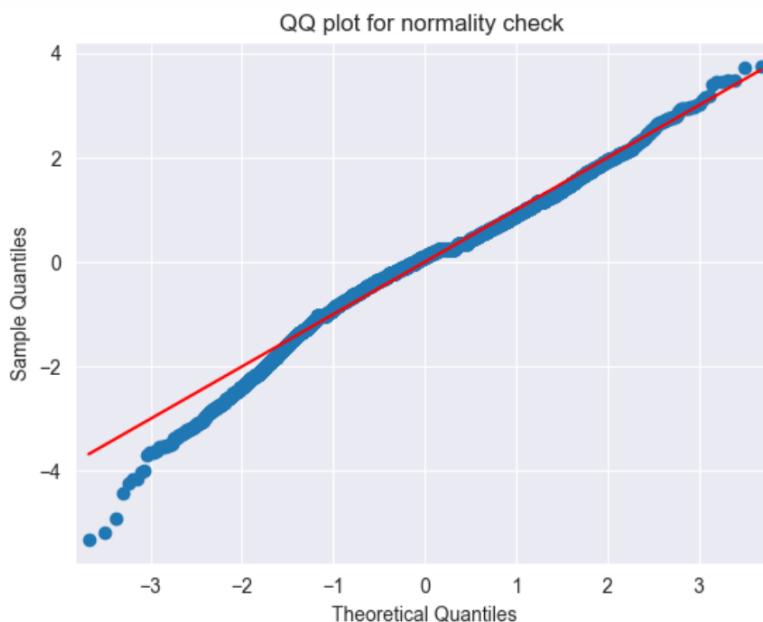
#### Box 4. qSOFA (Quick SOFA) Criteria

Respiratory rate  $\geq 22/\text{min}$

Altered mentation

Systolic blood pressure  $\geq 100 \text{ mm Hg}$

[5]



[6]

סוג פיצ'ר	שם פיצ'ר	טיפוס	תיאור	סטטיסטיקה תיאורית
סימנים חיוניים	HR (Heart Rate)	Float	כמות פעימות הלב בדקה	מינימום: 20, מקסימום: 280 ממוצע: 84.66, חציון: 84
	O2Sat	Float	ריוויין חמצן בדם	מינימום: 20, מקסימום: 100 ממוצע: 97.19, חציון: 98
	Temp	Float	טמפרטורת הגוף (צלזיוס)	מינימום: 20.9, מקסימום: 50 ממוצע: 36.9, חציון: 36.5
	SBP	Float	לחץ הדם בעת כיווץ שריר הלב	מינימום: 20, מקסימום: 299 ממוצע: 123.51, חציון: 80
	MAP	Float	ממוצע משוקלל של לחצי הדם לאורך מחזור קרדיאלי אחד	מינימום: 20, מקסימום: 300 ממוצע: 82.29, חציון: 80
	DBP	Float	לחץ הדם בעת הרפיית שריר הלב	מינימום: 20, מקסימום: 300 ממוצע: 63.78, חציון: 62
	Resp	Float	מספר הנשימות בדקה	מינימום: 1, מקסימום: 100 ממוצע: 18.75, חציון: 18
	EtCO2	Float	כמות הפחמן הדו-חמצני הנפלט	מינימום: 10, מקסימום: 100 ממוצע: 33.03, חציון: 33
ערכי מעבדה	BaseExcess	Float	מדד לעודף הביקרבונאט	מינימום: -32, מקסימום: 49.5 ממוצע: -0.74, חציון: 0
	HCO3	Float	כמות הביקרבונאט	מינימום: 0, מקסימום: 55 ממוצע: 24.08, חציון: 24
	FiO2	Float	אחוז החמצן הנצרך	מינימום: -50, מקסימום: 4000 ממוצע: 0.58, חציון: 0.5
	pH	Float	רמת החומציות	מינימום: 6.62, מקסימום: 7.78 ממוצע: 7.37, חציון: 7.38
	PaCO2	Float	הלחץ של פחמן דו-חמצני בדם	מינימום: 10, מקסימום: 100 ממוצע: 41.16, חציון: 40
	SaO2	Float	ריווי החמצן בדם העורקי	מינימום: 23, מקסימום: 100 ממוצע: 92.51, חציון: 97
	AST	Float	ריכוז האנזים AST בדם	מינימום: 5, מקסימום: 9961 ממוצע: 264.5, חציון: 40
	BUN	Float	ריכוז האוריאיה בדם	מינימום: 1, מקסימום: 268 ממוצע: 23.55, חציון: 17
	Alkalinephos	Float	מדד לאנזים הכבד	מינימום: 7, מקסימום: 2528 ממוצע: 101.3, חציון: 73

מינימום: 1, מקסימום: 27.9 ממוצע: 7.54, חציון: 8.25	כמות הסידן בדם	Float	Calcium	
מינימום: 74, מקסימום: 145 ממוצע: 105.88, חציון: 106	כמות הכלוריד בדם	Float	Chloride	
מינימום: 0.1, מקסימום: 41.9 ממוצע: 1.49, חציון: 0.92	כמות הקריאטינין בדם	Float	Creatinine	
מינימום: 0.01, מקסימום: 35 ממוצע: 1.58, חציון: 0.42	כמות בילירובין ישיר בדם	Float	Bilirubin_direct	
מינימום: 10, מקסימום: 952 ממוצע: 137.07, חציון: 127	כמות הגלוקוז בדם	Float	Glucose	
מינימום: 0.2, מקסימום: 31 ממוצע: 82.68, חציון: 1.88	כמות הלקטוז בדם	Float	Lactate	
מינימום: 0.6, מקסימום: 9.8 ממוצע: 2.04, חציון: 2	כמות המגנזיום בדם	Float	Magnesium	
מינימום: 0.3, מקסימום: 17.6 ממוצע: 3.54, חציון: 3.3	כמות הפוספט בדם	Float	Phosphate	
מינימום: 1.3, מקסימום: 27.5 ממוצע: 4.13, חציון: 4.1	כמות הפוטסיום בדם	Float	Potassium	
מינימום: 0.1, מקסימום: 49.6 ממוצע: 1.98, חציון: 0.8	כמות הבילירובין הכוללת בדם	Float	Bilirubin_total	
מינימום: 0.01, מקסימום: 440 ממוצע: 8.34, חציון: 0.3	כמות הטרופונין בדם	Float	TroponinI	
מינימום: 8.8, מקסימום: 71.7 ממוצע: 30.81, חציון: 30.3	כמות תאי הדם האדומים בדם	Float	Hct	
מינימום: 2.6, מקסימום: 25 ממוצע: 10.43, חציון: 10.3	כמות ההמוגלובין בדם	Float	Hgb	
מינימום: 17.1, מקסימום: 250 ממוצע: 41.3, חציון: 32.4	הזמן בשניות עד להיווצרות כריש דם	Float	PTT	
מינימום: 0.1, מקסימום: 440 ממוצע: 11.45, חציון: 10.3	מספר כדוריות דם לבנות	Float	WBC	
מינימום: 35, מקסימום: 1760 ממוצע: 6285.69, חציון: 248	חלבון כרישה המיוצר בכבד	Float	Fibrinogen	נתונים דמוגרפיים
מינימום: 2, מקסימום: 2322 ממוצע: 196.72, חציון: 182	כמות טסיות הדם	Float	Platelets	
מינימום: 15, מקסימום: 100 ממוצע: 61.66, חציון: 63.31	גיל המטופל	Float	Age	
אחוז נשים: 44.5, אחוז גברים: 55.5	מגדר (0 לאישה ו-1 לגבר)	int (בינארי קטגוריאל)	Gender	
ערך שכיח-0, שכיחות-50.6%	אגף לטיפול רפואיים במחלקת הטיפול הנמרץ	int (בינארי קטגוריאל)	Unit1	
ערך שכיח-1, שכיחות-50.6%	אגף לטיפולים כירורגיים במחלקת הטיפול הנמרץ	int (בינארי קטגוריאל)	Unit2	
מינ': -5366.8, מקס': 17.34 ממוצע: -50.95, חציון: -5.95	הפרש השעות בין זמן הרישום בבי"ח לזמן הרישום בטיפול הנמרץ	Float	HospAdmTime	
מינימום: 1, מקסימום: 336 ממוצע: 26.57, חציון: 21	משך השהייה בטיפול הנמרץ	int (שלם חיובי)	ICULOS	תוצאה
ערך שכיח-1, שכיחות-98.2%	האם המטופל סובל מאלח דם	int (בינארי קטגוריאל)	SepsisLabel	

[7]