

המחלקה להנדסת תעשיה

פרויקט קורס כריית ידע מתקדם 0572-5448-01

איתור חשד לזיהום בפגים   
בהתבסס על נתוני פגיה בבית חולים איכילוב

מרצה: פרופ' עודד מימון

עוזר הוראה: ד"ר רפאל הוד

קבוצה DM07:

גל לוי

כפיר סיטון

מילה אורלובסקי

תוכן עניינים

[חלק א' - בחירת הארגון ואיסוף הנתונים 4](#_Toc485991357)

[1. רקע 4](#_Toc485991358)

[2. תרומת כריית הידע לארגון 4](#_Toc485991359)

[3. בחירת הנתונים 5](#_Toc485991360)

[חלק ב'- עיבוד מקדים של הנתונים (DATA PREPARATION) 5](#_Toc485991361)

[4. הגדרת תכונת המטרה 5](#_Toc485991362)

[5. בחירת התוכנה 6](#_Toc485991363)

[6. פורמט הנתונים הקיים 6](#_Toc485991364)

[7. השלמת נתונים 6](#_Toc485991365)

[8. מניפולציה ראשונית על הנתונים 7](#_Toc485991366)

[9. יצירת שדות חדשים 9](#_Toc485991367)

[חלק ג – כריית הידע והסקת מסקנות 10](#_Toc485991368)

[10. בחירת אלגוריתם לכריית ידע 10](#_Toc485991369)

[11. יישום האלגוריתמים 12](#_Toc485991370)

[12. ניתוח התוצאות 19](#_Toc485991371)

[13. השוואת אלגוריתמים 21](#_Toc485991372)

[14. השוואה לספרות 21](#_Toc485991373)

[15. משמעות התוצאות 22](#_Toc485991374)

[16. מסקנות 23](#_Toc485991375)

[ביבליוגרפיה 24](#_Toc485991376)

[נספחים 25](#_Toc485991377)

[1. נספח 1 - סכמת הטבלאות 25](#_Toc485991378)

[2. נספח 2 - תיאור המשתנים (Attributes) 26](#_Toc485991379)

[3. נספח 3 – התפלגות ציון Apgar לפי שבוע לידה, לפני השלמת נתונים 30](#_Toc485991380)

# רשימת טבלאות

[טבלה 1. טבלת דירוג חום 8](#_Toc486250329)

[טבלה 2. התפלגות אוכלוסייה לפי סיווג SGA 9](#_Toc486250330)

[טבלה 3. רשימת Attributes במודל עץ החלטה 14](#_Toc486250331)

[טבלה 4. רשימת Attributes במודל רגרסיה לוגיסטית 16](#_Toc486250332)

[טבלה 5. תוצאות רגרסיה לוגיסטית 17](#_Toc486250333)

[טבלה 6. Confusion matrix רגרסיה לוגיסטית 17](#_Toc486250334)

[טבלה 7. סיכום מודל SVM 19](#_Toc486250335)

[טבלה 8. סיכום תוצאות שלושת המודלים 21](#_Toc486250336)

# רשימת איורים

[איור 1. תהליך טיוב חום 8](#_Toc485991573)

[איור 2. מפת חום 12](#_Toc485991574)

[איור 3. Boxplots עבור משתנים רציפים 13](#_Toc485991575)

[איור 4. תרשים עץ החלטה 15](#_Toc485991576)

[איור 5. תרשים ROC עץ החלטה 16](#_Toc485991577)

[איור 6. תרשים ROC רגרסיה לוגיסטית 18](#_Toc485991578)

# חלק א' - בחירת הארגון ואיסוף הנתונים

## רקע

בית חולים ליס ליולדות (איכילוב) הינו בית היולדות השני בגודלו בארץ, בו מתרחשות כ 12,000 לידות בשנה. מחלקת טיפול מיוחד בילוד (פגיה) מטפלת הן בתינוקות שנולדו פגים טרם זמנם והן בתינוקות שנולדו במועד אך נזקקו לטיפול מיוחד, קרי טיפול נמרץ. דוגמא לכך היא ילודים עם מומים מולדים או הסתבכות שקשורה בלידה, כגון שאיפת מי שפיר. בפגיה מתאשפזים כ 800 יילודים בשנה שמהווים כ 7 אחוזים מכלל הלידות החיות בבית החולים.

במחלקה 3 חדרי טיפול עיקריים: חדר 1 טיפול נמרץ בו 16 מיטות עם מכשירי ניטור לכל ילוד, חדר 2 עם 6 מיטות ו"חדר מתאוששים" עם 7 מיטות, בהם מאושפזים ילודים שאינם נמצאים עוד בסכנת חיים ממשית והינם קרובים לשחרור. צוות המחלקה כולל 8 רופאים מומחים, 5 מתמחים, 60 אחיות ו 7 אנשי כח עזר.

במחלקה נעשה שימוש בשתי מערכות מידע בשני תחומים:

* אדמיניסטרציה – ERP SAP (נמ"ר) המאפשרת רישום של החולים, ניוד בין מחלקות, התחשבנות מול קופת החולים, הזמנת ייעוציים ועוד
* קליניקה - IMDSOFT , המערכת מאפשרת ניטור סימנים חיוניים ותיעוד אירועים נשימתיים, תיעוד תרופות, אבחנות, פרוצדורות, כלכלה ועוד.

נתונים משתי המערכות נאספות ונשמרות במחסן הנתונים של בית החולים.

## תרומת כריית הידע לארגון

כאמור, כמות הנתונים הנאגרת על כל ילוד רבה, וביכולתה לסייע לצוות הרפואי בקבלת החלטות קליניות. כיום, המערכות הינן תפעוליות בלבד, כלומר מאפשרות ניטור והזנת נתונים אך לא הפקת ידע , כך שבכדי לקבל החלטה על טיפול, הרופא צריך לאסוף נתונים ממספר מקורות ולבצע את האינטגרציה בראשו.

השיקול בבחירת פרויקט בתחום הילודים בכלל ופגים בפרט הוא היות והנתונים ייחודיים כיוון שלאוכלוסיית מטופלים זו אין נתוני רקע ממקורות אחרים (כגון קופות חולים, היסטוריה רפואית), ובכך הם מאפשרים ניתוח נתונים שפחות חשוף לרעשים.

פגים הינם מטופלים בעלי מערכת חיסונית מוחלשת ונתונים לסיבוכים רבים ואף התדרדרות חמורה עד כדי מוות. כיום, הרופא בעת ביקורו עובר על תיקו של הילוד ובהינתן סט פרמטרים, אותם הוא מנתח במוחו הוא מקבלת החלטה טיפולית. כפי שנפרט בהמשך, בחרנו להתמקד בסיבוכים שניתן לאתר באמצעות לקיחת תרבית דם. אולם נטילת תרבית היא שלב בו מתעורר חשד ממשי לזיהום או מחלה, כלומר זהו שלב מאוחר יחסית, בו עד שתאושש תוצאת התרבית ויתחיל טיפול, הילוד יכול להתדרדר במהירות.   
נרצה לבחון מצבים חשודים לסיבוך outcome late onset, כלומר תרבית שנלקחה החל מהיום השלישי באשפוז, שכן תרבית מוקדמת יותר יכולה להצביע על סיבוך שהתרחש קודם לכן – בלידה או ברחם, ולא כתוצאה מטיפול במחלקה.

התרומה בכריית ידע בנושא זה שאת תוצאותיה ניתן ליישם במחלקה ולנסות לזהות מוקדם יותר, לעתים אף טרם הופעת הסימפטומים המעידים עליה, את ההתדרדרות.

## בחירת הנתונים

הנתונים לפרויקט מגיעים ממחסן הנתונים של מחלקת פגיה, המבנה רלציוני ונשמר ב 2014 MS SQL, הנתונים קיימים משנת 2011 עד היום, ובסך הכל כ 5000 רשומות.  
סכמת הטבלאות מובאת [בנספח 1](#נספח1).

סיכום הטבלאות:

* טבלה ראשית Neonatal – זוהי טבלה שטוחה המכילה את כל האוכלוסייה, כל כניסה למחלקה מהווה שורה. מקור הנתונים הינו מערכת IMDSOFT אך חלק מהנתונים היו חסרים בנתוני מקור והושלמו ממערכות אחרות (למשל נתוני חדר לידה).   
  בטבלה 5023 שורות, 21 Attributes מסוגים שונים.
* טבלת טמפרטורה ואירועים נשימתיים – טבלאות Log בהן כל אירוע הוא שורה

פירוט הAttributes בשלושת הטבלאות [בנספח 2](#נספח2)

# חלק ב'- עיבוד מקדים של הנתונים (DATA PREPARATION)

## הגדרת תכונת המטרה

תחילה, הוגדרה מטרת העל כצמצום שיעור סיבוכים של פגים. מכאן נגזרה מטרת הפרויקט והיא איתור מוקדם של סיבוך, ע"י חיזוי הסיכוי של ילוד לפתח סיבוך. לאור ריבוי המצבים האפשריים, וחוסר היכולת לבודד באופן חד משמעי את הסימפטומים לכל סיבוך, בחרנו להתמקד בשני סיבוכים, אשר את החשד להם ניתן לבסס באמצעות לקיחת תרבית דם. הראשון, SEPSIS שהוא תגובה חיסונית של הגוף לנוכחות זיהום חמור בדם (ספרות רפואית משייכת כרבע מתמותת פגים לSEPSIS), והשני NEC - דלקת מעי המאפיינת פגים בה חלקים מהמעי נמקים, סיבוך תדיר בקרב אוכלוסיית פגים. החשד מתממש אם בתרבית צומח חיידק, קרי התרבית חיובית.   
יש לציין כי היות התרבית חיובית עדיין מחייבת הסתכלות של זיהומולוג לקביעה ודאית. כיוון שקביעה זו לא ממוחשבת, נתייחס לתרבית חיובית כ Proxy indicator לזיהום. מהלך זה סוכם עם הקלינאים, כך שאת משמעותו הם מבינים ומקבלים.

מכאן, שתכונת המטרה שלנו מתבססת על שדה PositiveCulture3daysוע"י הפיכתו לבינארי PositiveBinar אנחנו מבקשים לחזות האם תתקבל עבור הילוד תרבית דם חיובית לאחר 3 ימי שהייה במחלקה.

## בחירת התוכנה

מטרתנו הייתה בניית מודל Supervised learning כדי לחזות אירוע בינארי. לשם כך בוצעו עיבודים נוספים בנתונים והשלמות נתונים רבות, הכנת הנתונים לכרייה בוצעה במחסן הנתונים באמצעות תהליכי ETL ב SQL. בחינתם והבנתם, הצגתם לרופאים לצרכי הגדרות קליניות, בוצעה באקסל.

בשלבי בניית המודלים נזקקנו לתוכנה שמצד אחד יודעת להתמודד עם מורכבות ורב מימדיות, ומנגד תאפשר פיתוח סינטקסים קלים והורדת חבילות מובנות שלא יצריכו כתיבת קוד סבוכה.   
השתמשנו בפייתון בסביבת Anacondaו- Spyder.

## פורמט הנתונים הקיים

מבנה הנתונים הינו רלציוני, כפי שתואר בסעיפים קודמים. לצורך התאמת הנתונים לכרייה נצרך חישוב וקטגוריזציות להתאמה לכרייה:

* משקל בלידה – כיוון שמדובר ברצף נומרי רציף, יש לחלק אותו לקטגוריות בעלות משמעות קלינית זהה. לאחר התייעצות עם קלינאים חילקנו לקבוצות במרווח של 250 גר: /<750 gr/751-1000/1001-1250/1251-1500/1501-1750//1751-2000/>2000.  
  נבנה שדה חדש Weight\_group, ו Weight\_group\_code (קטגוריות בערכים נומריים)
* שבוע לידה – כנ"ל [ <26 (מצב של פגות חמורה) 27-29/30-32/33-35>=36/ (לידה במועד)  
  נזנה שדה חדש Pregnancy\_Week\_group ו Pregnancy\_week\_group\_code (קטגוריות נומריות)
* ציון אפגר – מצב חמור ציון 0-5, מצב בינוני 6-8, מצב טוב 9-10.

## השלמת נתונים

כאמור, הנתונים מגיעים למחסן ממספר מקורות, וכתוצאה מכך קיימים לא מעט ערכי missing וערכים לא הגיוניים כתוצאה ממילוי ידני.   
הרצת frequencies על כל משתנה סייע לאיתור השיבושים בנתונים, והניע לפעולות השלמה ותיקון הבאות:

* **כפילויות** הוצאו מקרים בהם תינוק אחד נכנס יותר מפעם אחת לפגייה. מקרים אלה נדירים (38 תינוקות, כ 0.9% מהמקרים), חלקם כתוצאה מרישום לקוי.
* **שבוע לידה ומשקל לידה** היו ערכי missing רבים (כ 20%), לצורך השלמתם חוברו נתוני הפגים לטבלת יולדות (המתועדות במערכת Chameleon – מערכת תפעולית בחדר לידה). למי שנותר ללא נתון (4 מקרים) בוצעה קריאת תיקים, תינוקות שנולדו במועד השלמנו ערך default 40 וקטגוריית משקל <2000. למקרים אלה עודכן AGA=1 (כלומר משקל ממוצע לגילם)
* **ציון Apgar** חסר ל 199 מקרים, מתוכם 141 נולדו במועד. הנחת העבודה הייתה שאם נולדו במועד, ציון האפגר שלהם היה כמו הציון הנפוץ לקבוצת גיל ומשקל זו. בהתבסס על התפלגות ציון אפגר לפי שבוע לידה, הערך השכיח ביותר היה 9, לכן למי שנולד בשבוע 33 ומעלה עודכן 9, בשבוע 30-32 עודכן ציון 8, בשבוע 26-29 – ציון 5 ומתחת לשבוע 26 ציון 1. התפלגות מפורטת בנספח 3.
* **מין** – ב- 148 מקרים היה חסר מין יילוד, הנתון הושלם מנתוני נמ"ר (אדמיניסטרציה)
* **גיל** בשני מקרים הוזנה למערכת בטעות ת.ז של התינוק בשדה ת.ז של האמא ולכן גיל האם חושב לפי גיל תינוק, תיקון בוצע ע"י לקיחת הנתון ממערכת קמיליון – מערכת של חדרי לידה.
* **DischargeDate** – טעויות רישום/אי דיוק בהזנת הנתונים גרמו לכך שנמצאו ילודים שהיו מאושפזים במשך יותר משנתיים, מה שכמובן אינו נכון. איתרנו מקרים אלה ע"י frequency על משך שהייה (תאריך שחרור-תאריך כניסה). לשם פתרון התחברנו לטבלת תנועות ב נמ"ר, נתונים יותר מדויקים ובמקרים בהם היה חסר או שהנתון לא הגיוני (הפרש של יומיים או יותר בין המערכות) ביצענו החלפה של מקור הנתונים.

## מניפולציה ראשונית על הנתונים

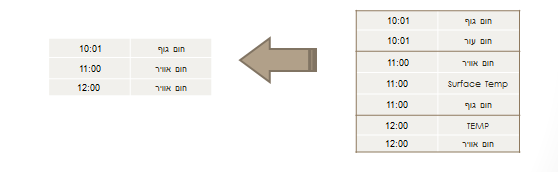
לחלק מהשדות היה צורך לבצע מניפולציה או טרנספורמציה של נתונים בכדי להשמישם.

* **חום** - כאשר התינוק מגיע למחלקה, הוא מחובר למוניטור שמודד את חום גופו, אך גם מודד את חום האוויר באינקובטור, כמו כן חיישן נוסף מודד את חום עורו ומד חום מודד את חום האוויר בחדר. כל המדידות הנ"ל נרשמות באותה טבלת חום, כך שייתכן מצב בו באותה נקודת זמן תהיינה מדידות עם ערכים שונים ממקורות שונים. לשם כך, ובסיוע קלינאים, הוכנה טבלת דירוג של חום, מיהו החום הקובע במידה וקיימת יותר ממדידה אחת בו זמנית. בעצם, תוצר שלב זה היא טבלת חום מטוייבת לאחר הוצאת השורות הלא רלוונטיות. בטבלה להלן ניתן לראות את המדרג שהוכן לחום

|  |  |  |
| --- | --- | --- |
| parameterid | parameter\_order | parameter\_name |
| 11492 | 1 | חום גוף |
| 23872 | 2 | חום עור-ליס |
| 23873 | 3 | חום אוויר-ליס |
| 24447 | 4 | Surface Temp-ליס |
| 24448 | 5 | Set Point Temp-ליס |
| 24079 | 6 | TEMP-ליס |
| 24446 | 7 | Core Temp-ליס |

טבלה 1. טבלת דירוג חום

ובתרשים הבא ניתן לראות דוגמא לתהליך הטיוב, בהתאם לטבלה לעיל:



איור 1. תהליך טיוב חום

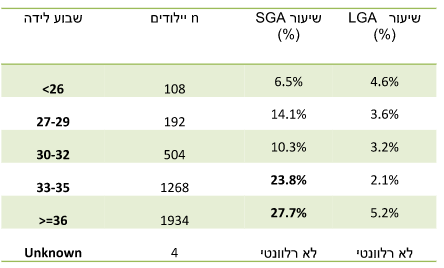
* **יחידות**– חום מוצג למשתמש במעלות צלזיוס, אך לאחר בחינת הנתונים גילינו שהמדידות נשמרות בפרנהייט. לשם כך התבצעה המרת יחידות.
* **ערכים לא פיזיאביליים** – מדידת החום נשמרת כל העת, גם אם הערכים לא פיזיאביליים (למשל 17 מעלות או מטה). בבירור במחלקה, מדידות אלה מתרחשות כאשר התינוק מוצא לרחצה או החייאה, או אם המדחום נפל, לכן ערכים אלו מוצאים מהחישוב. לפי הגדרת הקלינאים ערכי החום הרלוונטיים הם מעל 30 מעלות או מתחת ל- 43.
* **תדירות אירועים נשימתיים** – כדי לבדוק כיצד תדירות אירועים נשימתיים משפיעה על לקיחת תרבית, בוצע מיצוע של אירועים ליממה קודם ללקיחת תרבית, ההשערה היא שככל שתדירות האירועים עולה, כך עולה הסיכוי להתפתחות סיבוך.

## יצירת שדות חדשים

לצורך חישובים, הרחבת התכנים והתאמת לתכונת מטרה, נוצרו מספר שדות חדשים במאגר הנתונים, חלקם נבנו מעיבוד טבלאות log והוספתם לטבלתNeonatal הראשית (יצירת פורמט Flat File)

* מדדי גודל יחסי של ילוד: SGA  (Small gestational age) LGA) Large gestational age) או AGA (Average gestational Age). החישוב כולל גיל הלידה ומינו של הילוד. לחישוב שלנו התבססנו על אחוזונים אמריקאי [1]
* אם משקל לידה < אחוזון 10 בהתאם לשבוע ומין => SGA
* אם משקל לידה > אחוזון 90 בהתאם לשבוע ומין = > LGA
* אם (10th percentile≤weight≤90th percentile) => AGA

להלן התפלגות האוכלוסייה לפי הסיווג המחושב:



טבלה 2. התפלגות אוכלוסייה לפי סיווג SGA

* + שדות מחושבים יומיים, כאמור מטבלאות log לתוך טבלת Neonatal. הרעיון הוא לשייך עליה בתזמון האירועים לתזמון לקיחת תרבית
    1. Temperature\_arrivalday\_over38\_cnt ספירת מספר מדידות חום מעל 38 מעלות ביום הגעתו למחלקה. 38 זהו סף מקובל בספרות הרפואית לתהליך המחייב התייחסות.
    2. Temperature\_firstday\_over38\_cnt ספירת מספר מדידות חום מעל 38 מעלות ביום הראשון לאחר הגעתו למחלקה.
    3. Temperature\_secondday\_over38\_cnt כנ"ל ביום השני
    4. Temperature\_thirdday\_over38\_cnt כנ"ל ביום השלישי
    5. Events\_arrivalday\_cnt מספר אירועים נשימתיים ביום הגעתו למחלקה
    6. Events\_firstday\_cnt מספר אירועים נשימתיים ביום הראשון לאחר הגעתו למחלקה
    7. Events\_secondday\_cnt כנ"ל ביום השני
    8. Events\_thirdday\_cnt כנ"ל ביום השלישי

PositiveBinar – שדה המטרה בעבודתנו הוא שדה חדש שנוצר מתוך שדה קיים PositiveCulture3days. במקום ספירת כמות תרביות חיוביות במהלך האשפוז הפכנו לשדה בינארי שמשמעותו: האם הייתה לפחות תרבית דם אחת חיובית ביום השלישי באשפוז ואילך.

# חלק ג – כריית הידע והסקת מסקנות

## בחירת אלגוריתם לכריית ידע

אנו מעוניינים לנבא אירוע בינארי (ניבוי תרבית דם חיובית לאחר 3 ימי אשפוז), לשם כך ביצענו קלסיפיקציה (סיווג) באמצעות מודלי Supervised Learning (למידה מונחית).

בחרנו ליישם שלושה מודלים מוכרים בתחום זה:

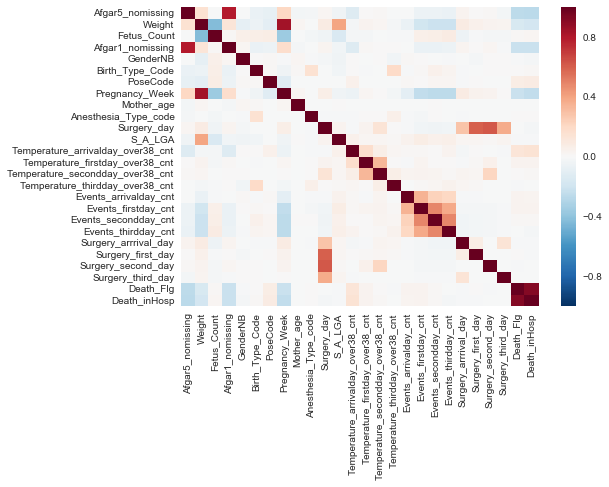
1. **עץ החלטות** – המודל הנפוץ לקלסיפיקציה, בעבודתנו נשתמש באלגוריתם מבוסס CART, עץ המאפשר לעבוד עם משתנים קטגוריאליים ובינאריים.   
   העץ מורכב מצמתי החלטה, כאשר כל אחד מהם מתפצל על בסיס תנאי של מאפיין מסוים. בנוסף מורכב מעלים המכילים את הערך מטרה החזוי עבור התצפית המתאימה בהתאם למסלול שמוביל אליהם בעץ. המודל מבצע חזרות רקורסיביות על המשתנים המסבירים ומפצל אותם לקבוצות הומוגניות בהתבסס על מדד Information Gain, שהוא תוצר ישיר של אנטרופיה עבור כל Attribute.  
   היתרון במודל זה הוא היכולת להציגו בצורה ברורה, כך שניתן לבצע התערבות בהתאם לכללים מובנים. החיסרון הוא מורכבותו ונטייתו להגיע לעומקים גדולים גם אם הוא עובר pruning(גיזום). חשוב להבין את הפרמטרים של המודל ולהתנסות בהם כדי לא להציג תוצאות מנוונות או לחילופין ליצור עץ שהוא overfitted(נוטה לתאימות יתר).
2. **רגרסיה לוגיסטית** – מודל שמתאים מאוד לבעיה של קלסיפיקציה בינארית. מאפשר לבנות מודל שיחזה את ההסתברות להתרחשות של אירוע. התוצר יאפשר לזהות קשרים בין משתנה מסביר למשתנה מטרה, את מובהקות וחוזק הקשר ואת טיב מודל החיזוי (R2). את הצגת תוצאות המודל ניתן להראות טבלאית ובנוחות, תוך ציון מקדם "תוספת סיכון" של כל גורם על התוצאה (OR – Odds Ratio). החיסרון בו הוא קושי בהתמודדות עם כמות רבה של משתנים קטגוריאליים או כאשר טווח הנתונים רחב (span)
3. **SVM – Support Vector Machine** – מודל שמאפשר איתור של וקטורים המפרידים בין הקלאסים השונים ומאפשרים חיזוי בהתאם לכל המשתנים התלויים, ניתן להשתמש גם במשתנים רציפים וקטגוריאליים ובטווח נתונים רחב. המודל יחסית פשוט ליישום. החיסרון שלו הוא הקושי בתרגום הפלט והצגתו להנהלה, וגם הסיכון של overfitting. מבחינת השלכות עסקיות נוכל לבצע חיזוי בלבד באמצעות המודל, אך לא להציג אותו כמערכת תומכת החלטה.

קיימים מודלי קלסיפיקציה נוספים שלא נבחרו לעבודה זו, לדוגמא Random Forest ו- Neural Network לא נבחרו עקב מספר תצפיות מועט מדי לשימוש המודל. כמו כן, המודל KNN (K-Nearest Neighbors) לא נבחר עקב הצורך בנרמול כמעט כל המשתנים לצרכי שימוש בפונקציות מרחקים במודל, ומעצם היותו מודל Lazy שלאו דווקא מגיע לפתרון אופטימלי אבסולוטי אלא מקומי.

## יישום האלגוריתמים

צעדים מקדימים:

* בניית משתני dummyלמשתנים קטגוריאליים. מבנה זה יסייע הן למודל רגרסיה לוגיסטית והן ל- SVM. שימוש בפקודה pd.get\_dummies
* בחינת קורלציות בין המשתנים, ע"י בניית מטריצת קורלציה (באמצעות מקדם Pearson), הצגתם ע"י מפת חום:



איור 2. מפת חום

קורלציה גבוהה נמצאה בין זוג המשתנים גיל ומשקל, וכמו כן בין שני משתני ציון Apgar. בהמשך לכך הוחלט להשאיר במודל את שבוע הלידה בלבד, כיוון ששבוע לידה זה פרמטר ידוע מראש אף טרם הלידה וניתן לבצע מעקב אחר היולדת במידת הצורך כבר בהגעתה למיון יולדות. בין זוג המשתנים הנוסף הוחלט להשאיר את ציון Apgar אחרי 5 דקות, מתוך הגיון קליני שככל שהמדידה רחוקה יותר מהלידה, זה מצביע על מצב מייצג יותר מאשר מצב רגעי בזמן לידה. כלומר, במידה והציון לאחר 5 דקות נמוך קיים גורם סיכון גדול יותר מאשר במדידה לאחר דקה אחת.

* טיפול ב Outliers – הכנת boxplots למשתנים רציפים לבחינת outliers

|  |  |
| --- | --- |
| Pregnancy week | Weight    ערך אחד של 8810 outlier הוצא מהקובץ |
| Mother age    ערך אחד של 99 (משמעות שאין ערך ) הוצא מהמדידה |  |

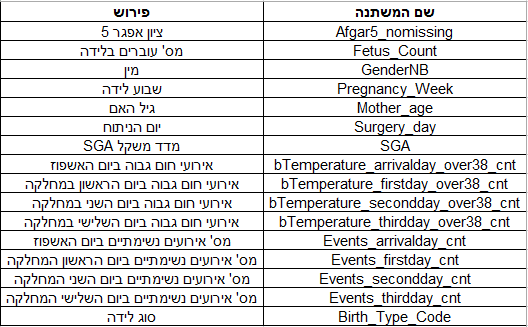
איור 3. Boxplots עבור משתנים רציפים

* התייחסות ל Imbalanced classes- – נבחרה שיטה מובנית בPython -. בכל מודל שיישמנו התבצע שימוש בפרמטר של המודל class\_weight = balanced. השיטה ממשקלת את הקלאסים לפי יחס גודל המדגם ותדירות של כל קלאס במדגם באופן הבא:

בפרויקט שלנו, משקל גבוה ניתן לקלאס החיובי ביחס לשלילי:

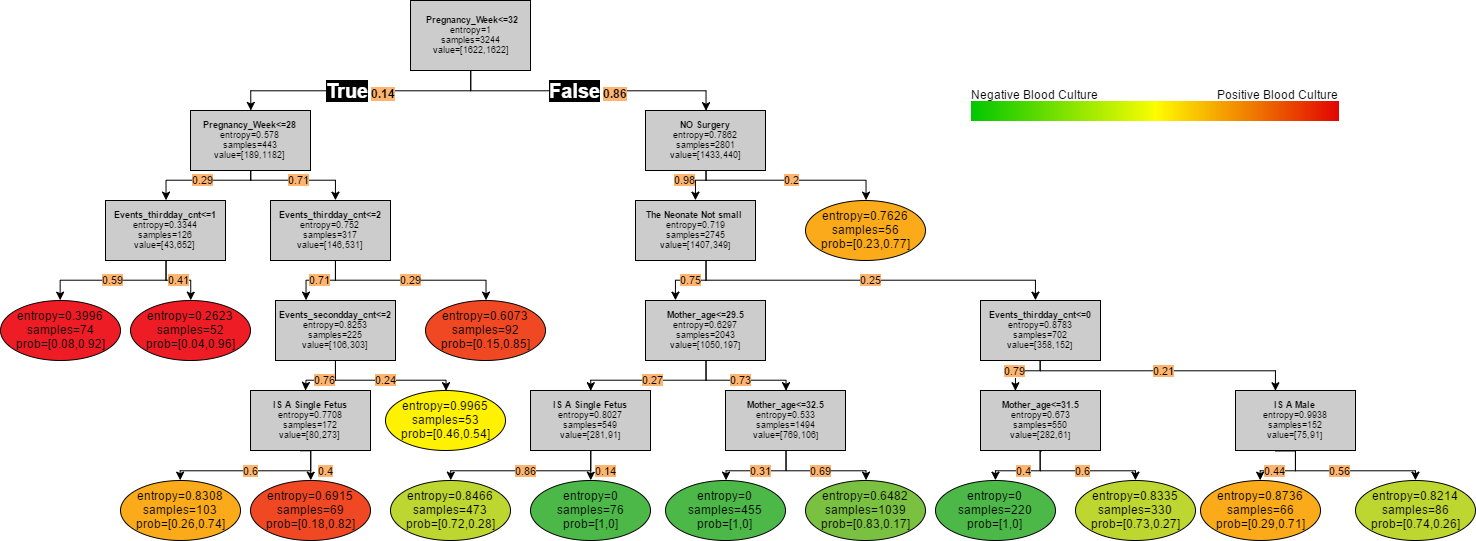
**עץ החלטה**

למודל עץ החלטה הוכנסו 16 Attributes להלן:



טבלה 3. רשימת Attributes במודל עץ החלטה

תחילה התקבל עץ מורכב וגדול, בוצע תהליך pruning באמצעות הגדרת היפר פרמטר min\_samples\_leaf=50 ועומק עץ מקסימלי 5.

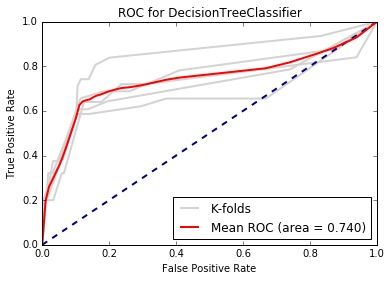
בסכמה להלן מתואר העץ שהתקבל

איור 4. תרשים עץ החלטה

פיצול הענפים הוגדר לפי קריטריון אנטרופיה. העץ מאפשר לזהות תחומים בהם יש סיכוי גבוה לתרבית חיובית (עלים אדומים) לבין תחומים בהם סיכוי לכך נמוך (עלים ירוקים). דוגמא במעבר בעץ: אם שבוע לידה נמוך מ 28 (extreme prematurity) והתרחש לפחות אירוע נשימתי אחד במהלך היום השלישי לאשפוז, קיים סיכוי של 0.96 להמצאות תרבית דם חיובית.

לסיכום Attributes שמשתתפים בעץ ומשפיעים על משתנה ההחלטה: שבוע לידה, גיל האם, ניתוח, SGA, אירוע נשימתי, מספר עוברים בלידה (תאומים), מין הילוד (זכר).

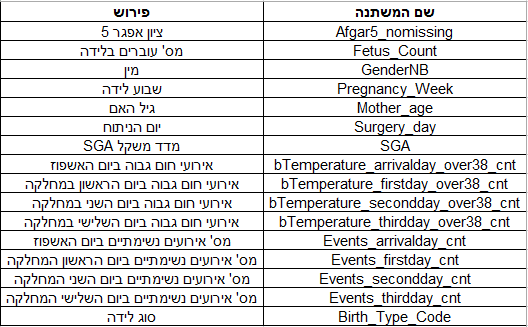
לצורך בדיקת מהימנות המודל בוצע Cross validation. בחרנו k=5. ולהלן התוצאה:



איור 5. תרשים ROC עץ החלטה

**רגרסיה לוגיסטית**

למודל הוכנסו 16 Attributes להלן:



טבלה 4. רשימת Attributes במודל רגרסיה לוגיסטית

מתוכם 6 משתנים נמצאו מובהקים, להלן טבלת המשתנים המובהקים:

|  |  |
| --- | --- |
| משתנה | OR |
| חום גבוה (+38) ביום ההגעה למחלקה | 4.1 |
| ניתוח בזמן הגעה למחלקה | 2.3 |
| מס' העוברים בלידה (תאומים) | 2.0 |
| מין (זכר) | 1.9 |
| מס' אירועים נשימתיים ביום השלישי לאשפוז | 1.2 |
| שבוע הלידה | 0.9 |

טבלה 5. תוצאות רגרסיה לוגיסטית

ניתן לראות כי הקשר החזק ביותר נמצא בין משתנה חום ביום ההגעה, ולאחר מכן ניתוח ביום הגעה למחלקה. עוד נמצא כי תאומים וזכרים בסיכון כפול לפיתוח תרבית דם חיובית. שבוע לידה נמצא בקשר שלילי, כלומר ככל ששבוע הלידה עולה כך קטן סיכון לתרבית דם חיובית.

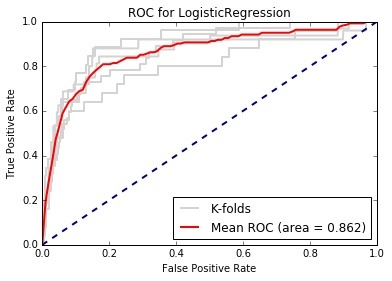
ציון R2 של המודל עומד על 0.7, נתון הגיוני שכן במודלים ביולוגיים צפוי ציון R2 נמוך יחסית עקב רמות אי ודאות גבוהות בתהליכים ביולוגיים.

לאחר הרצת המודל התקבלה Confusion matrix הבאה:

|  |  |  |
| --- | --- | --- |
| P 1 | P 0 |  |
| 227 | 1333 | A 0 |
| 29 | 9 | A 1 |

טבלה 6. Confusion matrix רגרסיה לוגיסטית

לצורך בדיקת מהימנות המודל בוצע תהליך Cross validation. בחרנו k=5. ולהלן התוצאה:



איור 6. תרשים ROC רגרסיה לוגיסטית

**מודל SVM**

תחילה, בוצעה חלוקה רנדומלית לנתוני training/test ביחס 80/20%בהתאמה. כמו כן נעשה שימוש ב- random state קבוע כדי שהסט עצמו לא ישתנה כתלות בהרצה. החלוקה האקראית נועדה ליצירת אי תלות במגמות בנתונים עצמם.

*train\_test\_split(X, y, test\_size=0.2, random\_state=42)*

כיוון שהקלאסים לא מאוזנים, השתמשנו בפרמטרים הבאים:

* [class\_weight="balanced"] מבצע משקול של התוצאות ובכך מביא לאיזון הקלאסים.
* C=13] ] תחילה נבחר מקדם C יחסית קטן לבחינת המודל. המודל צריך להיות מאוד מחמיר עם FN (סיווג False Negative), כך של- C תפקיד חשוב, בתור מקדם העונש של המודל.
* [kernel = linear] תחילה הוגדרה פונקציית קרנל ליניארית. בהמשך המודל הורץ גם עם פונקציית קרנל מעריכית (ברירת המחדל) אך בפועל פונקציה כזו תקשה עלינו להבין את התוצאות ויכולה להביא ביתר קלות ל- overfitting.

להלן טבלה מסכמת של confusion matrix בכל הרצה וציון המודל חיזוי.

|  |  |  |
| --- | --- | --- |
| **Hyper parameters** | **Confusion Matrix** | **Score** |
| Kernel linear C=13 | |  |  |  | | --- | --- | --- | | **P 1** | **P 0** |  | | 128 | 813 | A 0 | | 24 | 4 | A 1 | | 0.87 |
| Kernel Polynomial C=13 | |  |  |  | | --- | --- | --- | | **P 1** | **P 0** |  | | 1 | 940 | A 0 | | 28 | 0 | A 1 | | 0.99 |
| Kernel linear C=1000 | |  |  |  | | --- | --- | --- | | **P 1** | **P 0** |  | | 146 | 795 | A 0 | | 24 | 3 | A 1 | | 0.84 |

טבלה 7. סיכום מודל SVM

ניתוח רגישות למודל:

1. שינוי של פונקציית קרנל למעריכית: ציון גבוה מאוד, אך כנראה כתוצאה מ overfitting
2. שינוי ל C=1000 קרנל ליניארי. ציון יורד אך גם FN יורד, כלומר אנו מוכנים "לספוג" יותר טעויות של FP ולבצע בדיקה "מיותרת" מאשר לפספס זיהום פוטנציאלי.

כאמור, במודל זה לא נציג ROC. טיב המודל לכאורה נמדד ע"י Accuracy Score אך ההחלטה התקבלה בשקלול עם אופי הייחודי של התוצאות הרצויות.

## ניתוח התוצאות

נמצאו פרמטרים משותפים במודל רגרסיה לוגיסטית ועץ החלטות. אלה פרמטרים הגיוניים כגון שבוע לידה, מספר עוברים בלידה, אירועים נשימתיים. כל המצבים האלה מחלישים מאוד את הגוף של הילוד ומהווים גורם סיכון להתפתחות זיהום.   
במודל הרגרסיה הלוגיסטית נמצא גם כי ניתוח ביום ההגעה למחלקה משלש את הסיכון לתרבית דם חיובית לאחר 3 ימים. גם ממצא זה מתיישר עם ההיגיון הקליני, שכן ניתוח הוא טראומה לגוף, ובעקבותיו הגוף מוחלש וחשוף יותר לזיהומים. אם נוסיף לכך גוף קטן וחלש שטרם סיים להתפתח, סביר שהסיכון גדל מאוד.   
מין הילוד נמצא כגורם מובהק, שכן זכרים הם בסיכון כפול לתרבית דם חיובית, ממצא זה מסתדר עם ההיגיון הקליני, שכן זכר בעת לידתו הוא בעל מערכת פחות מפותחת מהנקבה. מספר ילודים בלידה (תאומים) מכפיל גם הוא את הסיכון לתרבית דם חיובית. ממצא זה הגיוני כיוון שתאומים נמצאים בסיכון יתר ללידה מוקדמת ומשקל לידה נמוך.  
חום מעל 38 לפחות פעמיים במהלך יום ההגעה למחלקה (לרוב יום הלידה), נמצא כגורם שמעלה את הסיכון לתרבית דם חיובית פי 4. ממצא זה הגיוני, אך יש לקחת בחשבון שחום גבוה יכול להיווצר מעצם זה שתינוק נולד בשבוע מוקדם, כאשר החום הוא תסמין אופייני עקב חוסר יכולת ויסות חום הגוף (בלוטת ההיפותלמוס טרם התפתחה דיו).

כל המשתנים הנ"ל שנמצאו משפיעים מסתדרים עם ההיגיון, אך יש לקחת בחשבון שאנחנו לא יודעים מה גרם למה, האם העובדה שמדובר בתאומים גרם ללידה מוקדמת וכתוצאה מכך שהתינוקות פחות מפותחים הם חשופים לזיהומים או שמדובר בגורם ישיר. יש להתייחס לתוצאות בהבנה וזהירות, ואין להסיק סיבתיות ישירה.

בעץ החלטה נמצא קשר חזק ביותר בין תרבית דם חיובית לשבוע לידה, לאחר מכן קיימת התייחסות לאירועים נשימתיים, עצם ביצוע ניתוח בימים הראשונים באשפוז ואירועים נשימתיים. כל הגורמים הנ"ל נמצאו שכגורמי סיכון לתרבית חיובית, אבל גם כאן יש לבחון קשר של כל אחד מהם בנפרד, שכן אמנם בנתונים לא נמצאה ביניהם קורלציה אך בפועל קיים קשר בין כל הגורמים. גוף מוחלש חשוף ליותר בעיות מלכתחילה.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **F2** | **F1** | **Recall** | **Precision** | **Accuracy** | **Model** |
| 33.5% | 18.2% | 76.3% | 10.3% | 83.7% | Decision Tree |
| 35.5% | 19.7% | 76.3% | 11.3% | 97.6% | Logistic Regression |
| 44.2% | 25.1% | 89.3% | 14.3% | 84.6% | SVM |

## השוואת אלגוריתמים

בוצעה השוואה בין האלגוריתמים לבחינת המודל המיטבי.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |

טבלה 8. סיכום תוצאות שלושת המודלים

ניתן לראות ששלושת המודלים הם בעלי יכולת חיזוי ברמת דיוק טובה. במקרה שלנו קריטריון החלטה לבחירת המודל הוא Recall שציון גבוה שלו ממזער את FN. בתחום שלנו, נעדיף Sensitivity גבוה יותר, מקרים שלFP מאשרFN (עדיף לקחת בדיקה מיותרת מאשר לפספס בדיקה לילוד עם זיהום, ולהביא לסכנת חיים). לפי ציון זה מודלSVM עדיף. בנוסף, גם ציון F2 קריטריון לבחירה מאותם טעמים, וגם לפיו כדאי לבחור ב SVM.

אולם בפועל לא נבחר בו, כיוון שלא ניתן לשקף אותו לצוות הרפואי כמודל שממו ניתן להסיק מסקנה. המודל מאפשר לבצע פרדיקציה בהתאם לווקטורים במרחב, אך בפועל לא ניתן ליישם אותו כפרוטוקול טיפולי.   
לצורך הצגת התוצאות לצוות הרפואי נבחר להשתמש בעץ החלטה, שמאפשר תצוגה ויזואלית ובהירה של הצעדים לפעולה או הקריטריונים לקבלת החלטה ללקיחת תרבית דם.

## השוואה לספרות

הספרות הרלוונטית בהקשר של ניתוח גורמים המשפיעים על סיבוכים באוכלוסיית פגים, אינה רחבה. מחקרים בנושא זה בעייתיים עקב מורכבות רבה בהגדרתSEPSIS באופן מדויק, ולפי כך רוב המחקרים מתבססים על איסוף נתונים ידני ועל גדלי מדגם קטנים.   
רוב הספרות מתמקדת בסיבוך "פשוט" למדידה – תמותה, או לחילופין ב- Early onset outcomes. מאמר של Stoll et al משנת 2002 בוחן את היארעות Sepsis ביחס למשקל לידה, ולאחר מכן את ה השלכות הזיהום. גם מאמר ישראלי (et al Makhoul 2002) מתמקד בהשפעת משקל לידה נמוך על הופעת הזיהום. בהקשר של NEC נמצא קשר בין גיל הלידה ומשקל הלידה להיות התינוק מונשם (Mohamed and Shah 2012). גם אנחנו בחרנו להתייחס להנשמה אך באופן מתקדם יותר, כלומר במקום לבחון רק את עובדת היותו מונשם בדקנו את תדירות אירועים נשימתיים היומיים. למשל אם הייתה עליה בתדירות Apnea (הפסקת נשימה רגעית) יש להניח שהתינוק מתחיל לפתח בעיה כלשהי.

מחקר שתוצאותיו דמו לתוצאות מודל הרגרסיה שלנו בדק פרמטרים פיזיקליים כגון משקל לידה וגיל. כמו במודל שלנו גם במחקר זה נמצא קשר בין מין זכר להופעת SEPSIS (Fanaroff et al 1998). בנוסף נבדקו נתוני מעבדה, אך בעבודה זו לא השתמשנו בנתוני מעבדה.

מחקר מקיף יותר בוצע כדי לנסות לזהות פרמטרים שיכולים לנבא מוקדם יותר את ההסתבכות, מחקרים אלה בדקו מגמת סימנים חיוניים ביממה לפני הסיבוך ומצאו קשר לסימנים כגון דופק, קצב נשימה, סטורציה (רוויון החמצן) בדם. נתונים אלה קיימים בבסיס הנתונים אך שליפתם ועיבודם מורכב ולא בוצע במסגרת עבודה זו.

לסיכום, ראינו הלימה מסוימת בין תוצאות המודלים שלנו לגורמים שנמצאו בקשר חזק עם סיבוך. אמנם לא התייחסנו למשקל לידה, ובמודל שלנו נמצא קשר בין שבוע לידה לסיבוך, במחקרים מצאו קשר בין משקל בלידה. מהיות המשתנים קורלטיביים ניתן להסיק בוודאות מסוימת שיש קשר בין כולם ובין תכונת המטרה.

## משמעות התוצאות

תחום ניתוח נתונים קליניים הוא מורכב, ולכן יש להתייחס לתוצאות שהתקבלו ולהבינם לעומק. תחילה נציין כי עקב גורמי אי ודאות רבים מודלים ביולוגיים מתאפיינים בדיוק נמוך. התוצאות שהתקבלו במודלים שלנו מפתיעים לטובה.  
טבעי להבין ששבוע לידה משפיע רבות על הסיכון להעראות סיבוכים, כיוון שהגוף של הילוד לא התפתח דיו ומערכת חיסונית שלו פגיעה מאוד. לממצא זה ולממצאים נוספים מהמודלים שלנו, כגון לידת תאומים וחום בקבלה יש להתייחס באופן ספקני, שכן לא ברור לנו מהתוצאות שקיבלנו האם הקשר שנמצא הוא קשר ישיר בין משתנה מסביר למשתנה מטרה או קשר עקיף באמצעות משתנה מתווך. שבוע לידה מוקדם אפריורית חושף את הילוד לסיבוך, ייתכן שהחום התפתח כתוצאה מהשבוע לידה המוקדם כיוון שמערכת ויסות חום (הומיאותרמיה) טרם התפתחה ולא היה פרמטר בלי תלוי עצמאי. כנ"ל לידת תאומים -ידוע כי לידת תאומים יכולה לגרום ללידה מוקדמת יותר ובמשקל נמוך יותר, לכן לא ניתן להסיק ודאית מה גורם למה.   
מה שכן נוכל להסיק מהתוצאות הוא שקיים קשר בין פרמטרים התחלתיים תלויי ילד שלא ניתנים להשפעה (כגון שבוע לידה) ופרמטרים רציפים בהם הרופא יכול לטפל (לדוגמא חום) למשתנה המטרה. חשוב לבודד את שני סוגי המשתנים הנ"ל כדי לנסות לבנות מודל שיוכל לתת המלצות להנחיות עבור רופאים (Guidelines) על בסיס פרמטרים שניתנים לשליטה (טיפול), תוך התייחסות לפרמטרים שלא ניתנים לשליטה כגון משקל לידה.   
תוצאה מעניינת אחרת היא פרמטרים שלא יצאו משפיעים, בעיקר ציון 5 Apgar. בתחילת הפרויקט הנחת העבודה הייתה שציון זה מהווה גורם סיכון משמעותי להתפתחות תרבית דם חיובית, כיוון שתינוק שגם אחרי 5 דקות מלידתו היה במצב פיזיולוגי ירוד, קיים סיכוי כי הבעיה תשפיע על המשך מצבו, אך לא נמצא קשר באף מודל. ניתן לשקול את נחיצותו של הציון בקרב פגים, ולנסות לאתר ציון אחר שיכול לבטא באופן טוב יותר את מצב הילוד הפג.

## מסקנות

תחום קליני, כאמור, בעל רמות אי ודאות גבוהות. ביצוע פרויקט חיזוי בתחום לא יכול לספק תוצאות חד משמעיות, אולם יכול להסב תשומת לב ולעורר דיון מקצועי במחלקה, שיאפשר גיבוש פרוטוקולי טיפול המתחשבים במסקנות אלה.   
לתחום חקר הביצועים הקליני קשה להתאים מודל מתמטי, כיוון שתוצאות הטיפול לא תמיד תלויות בתהליכי העבודה אלא הרבה במטופלים עצמם, מה שיוצר הרבה מאוד יוצאי דופן, ומקרים בהם גם התקנון המורכב ביותר לא נותן מענה, ואילו המשך הרחבה יוצר overfitting.

העבודה אתגרה אותנו מאוד. תחילה היה עלינו להבין כיצד להתמודד עם נתונים לא מושלמים שדרשו גוון עיבודים, השלמות וחשיבה. הבנת הנושא הקליני היה מפתח עיקרי להתקדמות בעבודה, ולכן מאוד חשוב לעבוד בצמוד לקלינאים ולקבל מהם הנחיה והכוון קליני.   
אתגר נוסף היה Low event rate או בשמו המקובל יותר imbalanced classes, כאשר האירוע אותו אנו רוצים לחזות מתרחש בכ- 3-4% מהמקרים. נברנו בספרות והתייעצנו עם בעלי ניסיון להבין את הנושא ואת דרכי ההתמודדות.   
אתגר אחרון הוא התמודדות עם סביבת עבודה חדשה, לכולנו הייתה היכרות ראשונה עם סביבת Python. החוויה הייתה מעניינת ומסקרנת.

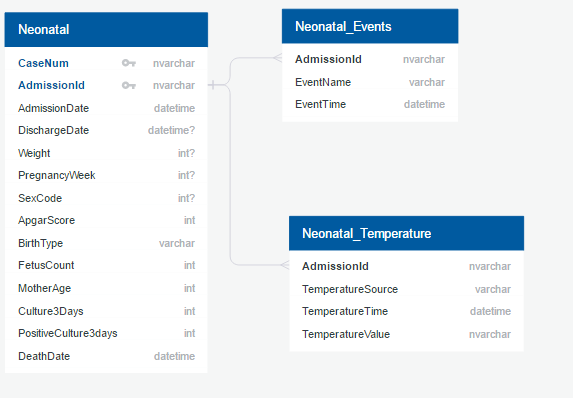
כיוון שהמודלים שנבנו הם רק הסתכלות ראשונית על הנתונים (עד לאחרונה נתונים אלה לא היו זמינים במחסן הנתונים), יש לבצע מחקר נוסף והפעם לנסות לשלב פרמטרים פחות מובנים, ואף לנסות לשלב מדדים פיזיולוגיים, על מנת שניתן יהיה להפוך את המודל לכלי תומך החלטה בזמן אמת. כלומר, בהינתן זיהוי התדרדרות המבוססת על נתוני עבר, המערכת תוכל להתריע לרופא על צורך בהתערבות.

# ביבליוגרפיה

1. A Revised Birth Weight Reference for the United States, Duryea. L et al. The American College of Obstetricians and Gynecologists.2014
2. Stoll, B.J., Hansen, N., Fanaroff, A.A., Wright, L.L., Carlo, W.A., Ehrenkranz, R.A., Lemons, J.A., Donovan, E.F., Stark, A.R., Tyson, J.E. and Oh, W., 2002. Late-onset sepsis in very low birth weight neonates: the experience of the NICHD Neonatal Research Network. Pediatrics, 110(2), pp.285-291.
3. Makhoul, I.R., Sujov, P., Smolkin, T., Lusky, A. and Reichman, B., 2002. Epidemiological, clinical, and microbiological characteristics of late-onset sepsis among very low birth weight infants in Israel: a national survey. Pediatrics, 109(1), pp.34-39.
4. Mohamed, A. and Shah, P.S., 2012. Transfusion associated necrotizing enterocolitis: a meta-analysis of observational data. Pediatrics, pp.peds-2011.
5. Fanaroff et al . 1998. Incidence, presenting features, risk factors and significance of late onset septicemia in very low birth weight infants. The National Institute of Child Health and Human Development Neonatal Research Network.. Pediatric Infectious Disease Journal17(7) pp 593-598
6. Griffin, M.P. and Moorman, J.R., 2001. Toward the early diagnosis of neonatal sepsis and sepsis-like illness using novel heart rate analysis. Pediatrics, 107(1), pp.97-104.
7. Apgar, Virginia (1966). "The Newborn (Apgar) Scoring System: Reflections and Advice". Pediatric clinics of North America 13: 645–650

# נספחים

## נספח 1 - סכמת הטבלאות



## נספח 2 - תיאור המשתנים (Attributes)

טבלה Neonatal

|  |  |  |  |
| --- | --- | --- | --- |
| שם שדה | תיאור השדה | סוג שדה | טווח ערכים |
| RN | מספר אשפוז | מספר רץ אוטומטי במערכת | מספר אוטומטי רץ |
| AdmissionDate | תאריך כניסה למחלקה | datetime |  |
| DischargeDate | תאריך שחרור מהמחלקה | datetime |  |
| Weight | משקל בלידה | שדה טקסטואלי מספרי, תיעוד המשקל בלידה במערכת IMDSOFT | 400-8810 (בגרמים) |
| Pregnancy Week | שבוע הלידה | שדה טקסטואלי מספרי, תיעוד שבוע הלידה במערכת IMDSOFT. | 21-42 |
| GenderNB | מגדר | קטגוריאלי | 1. זכר 2. נקבה |
| Apgar1 | ציון אפגר בלידה [בדיקה הנעשית בתינוק מיד בתום הלידה, במטרה להעריך את מצבו הגופני של תינוק [4] | שדה טקסטואלי מספרי | 0 – מצב וגיטטיבי  10 – תינוק במצב תקין  התפלגות ערכים אצלנו: |
| Apgar5 | ציון אפגר לאחר 5 דקות מהלידה |  |  |
| BirthType | סוג לידה | קטגוריאלי | 1. וגינלי 2. קיסרי 3. שולפן ריק (לידת ואקום) |
| PoseDesc | מצג הלידה | קטגוריאלי |  |
| AnesthesiaType | סוג הרדמה בלידה | קטגוריאלי |  |
| FetusCount | מספר עוברים בלידה | טקסט חופשי שדה מספרי | 1-3   |  |  | | --- | --- | | Fetus\_Count | N | | missing | 212 | | 1 | 3114 | | 2 | 1117 | | 3 | 52 | |
| MotherAge | גיל האם | טקסט חופשי שדה מספרי | כל מספר חיובי.  בנתוני היסטוריה 16-57 |
| IsFirstPregnancy | האם מדובר בלידה ראשונה | בינארי | 1 לידה ראשונה  0 לידה שניה ומעלה |
| Culture3Days | האם נלקחה תרבית דם לאחר 3 ימים מהאשפוז | בינארי | 1 – נלקחה תרבית  0- לא נלקחה תרבית דם |
| PositiveCulture3days | האם התרבית דם שנלקחה אחרי 3 ימים יצאה חיובית (צמחית חיידק) | בינארי | 1 – תרבית חיובית  0- תרבית שלילית |
| Surgery\_arrrival\_day | האם עבר ניתוח ביום ההגעה למחלקה | בינארי | 1. כן 2. לא |
| Surgery\_first\_day | האם עבר ניתוח ביום ההגעה למחלקה | בינארי | 1. כן 2. לא |
| Surgery\_second\_day | האם עבר ניתוח ביום ההגעה למחלקה | בינארי | 1. כן 2. לא |
| |  |  |  |  | | --- | --- | --- | --- | | Surgery\_third\_day | Surgery\_first\_day | Surgery\_second\_day | Surgery\_third\_day | |  |  |  |  | | האם עבר ניתוח ביום ההגעה למחלקה | בינארי | 1. כן 2. לא |
| Death\_Flg | האם נפטר | בינארי | 1. כן 2. לא |
| Death\_inHosp | האם הפטירה היתה בבית חולים | בינארי | 1. כן 2. לא |

תיאור שדות בטבלת Neonatal Events – טבלה המתארת אירועים נשימתיים.   
ההשערה היא שכמות גדלה של אירועים אלה בשעות קדם ההתדרדרות יכולה להצביע על התקרבותה

|  |  |  |  |
| --- | --- | --- | --- |
| שם שדה | תיאור השדה | סוג שדה | טווח ערכים |
| AdmissionId | מספר ייחודי במערכת IMDSOFT (כל כניסה למחלקה מקבלת מספר ייחודי) | מספר רץ אוטומטי במערכת | 5 ספרות |
| EventName | שם אירוע נשימתי | קטגוריאלי | * כחלון * אפניאה * ברדיקרדיה * טכיקרדיה * דסטורציה |
| EventTime | זמן בו תועדה התרחשות האירוע | datetime |  |

Neonatal Temperature

|  |  |  |  |
| --- | --- | --- | --- |
| שם שדה | תיאור השדה | סוג שדה | טווח ערכים |
| AdmissionId | מספר ייחודי במערכת IMDSOFT (כל כניסה למחלקה מקבלת מספר ייחודי) | מספר רץ אוטומטי במערכת | 5 ספרות |
| Temperature Source | מקור מדידת חום | קטגוריאלי | * חום גוף * חום אויר * חום עור |
| TemperatureTime | זמן בו תועדה התרחשות האירוע | datetime |  |
| TemperatureValue | ערך נמדד של חום | שדה טקסט חופשי מספרי | 32-657\*  \*תהליך טיוב מפורט בהמשך |

## נספח 3 – התפלגות ציון Apgar לפי שבוע לידה, לפני השלמת נתונים

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ציון APGAR** | **<26** | **27-29** | **30-32** | **33-35** | **>=36** |  | **Unknown** |  | **סה"כ** |
| 0 | 1 |  |  | 3 | 10 |  |  |  | 14 |
| 1 | 29 | 11 | 16 | 16 | 72 |  |  |  | 144 |
| 2 | 15 | 17 | 15 | 16 | 63 |  |  |  | 126 |
| 3 | 13 | 12 | 11 | 12 | 65 |  |  |  | 113 |
| 4 | 11 | 13 | 15 | 12 | 72 |  |  |  | 123 |
| 5 | 34 | 54 | 77 | 79 | 138 |  |  |  | 382 |
| 6 | 11 | 41 | 58 | 67 | 152 |  |  |  | 329 |
| 7 | 4 | 26 | 86 | 136 | 190 |  |  |  | 442 |
| 8 | 7 | 27 | 186 | 479 | 288 |  |  |  | 987 |
| 9 | 2 | 5 | 134 | 761 | 1261 |  |  |  | 2163 |
| 10 |  |  |  |  | 2 |  |  |  | 2 |
| NULL | 13 | 16 | 8 | 17 | 141 |  | 4 |  | 199 |
| **סה"כ** | **140** | **222** | **606** | **1598** | **2454** |  | **4** |  | **5024** |