**Microsoft Security Bulletin Analysis**

***דוח הערכה מוצג על ידי בר כהן וסהר חיים יעקב.  
  
דוח זה מסכם את שלב ההערכה בפרויקט, ובוחן האם התוצאות שהתקבלו בתהליך המידול עומדות ביעדים שהוגדרו בתחילת הדרך – הן ברמה האנליטית והן ברמה העסקית. הדוח כולל ניתוח של התאמת המודלים למטרות הפרויקט, הצגת הממצאים המרכזיים, והסקת מסקנות באשר לאפקטיביות הכללית של הפתרון. בנוסף, נבחן תהליך העבודה עצמו, תוך זיהוי נקודות חוזקה והצעות לשיפור עתידי***. ***​***

***שם מרצה : מר אבי זכאי.***

***שם מנחה : מר חנן לב.***

***מגישים :***

***בר כהן 208110254***

***סהר יעקב 314741851***



*תוכן עניינים*

[1. הערכת התוצאות : - 3 -](#_Toc197723301)

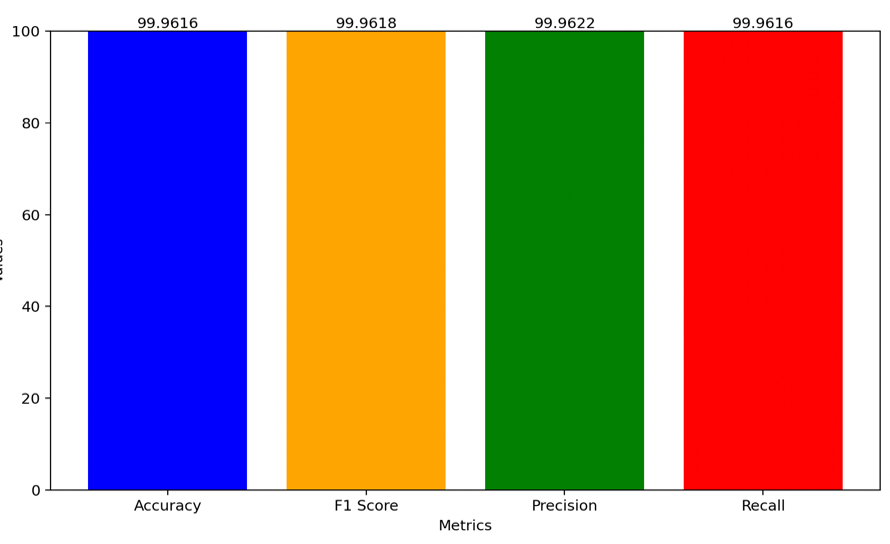
[2. תהליך הסקירה : - 6 -](#_Toc197723302)

# **הערכת התוצאות :**

בשלב זה של הערכת תוצאות המודל, נבחן את המטרה של הפרויקט מבחינת קריטריוני ההצלחה. המטרה של הפרויקט היא חיזוי מדויק של חומרת תקיפה עבור הארגון, באופן שיאפשר תגובה לרמת האיום, הפרדה של תקיפות חזקות יותר והערכות בהתאם.

לאחר שהצגנו את דרך ההתמודדות שלנו, חקרנו את הפתרונות האפשריים, ביצענו ניסויים מרובים עם מגוון מודלים תוך התאמה של פרמטרים לכל אחד מהם. התוצאות שהתקבלו מראים כי ניתן להגיע לרמות דיוק גבוהות מאוד בחיזוי חומרת התקיפה, ובפרט כאשר נעשה שימוש במודלים כמו .CatBoost

בנוסף בחנו גם מדדים נוספים כמו F1, Precision ו ,Recall -וכן התייחסנו ליכולת של המודלים להבחין בין רמות חומרה שונות – קלות, בינוניות וקריטיות – כיעד עסקי חשוב ביותר עבור הארגון.  
  
**ההתפלגות עבור המודל שנבחר :**



איור 1 : השוואת הדיוק בין ארבעת המודלים המובילים

התוצאות שהתקבלו ממודל CatBoost וממודלים נוספים שבחנו במהלך הפרויקט מציגות רמה גבוהה מאוד של ביצועים, בהתבסס על דיוק ורגישות. הנתונים מראים בצורה כמעט מושלמת, אפקטיבית ויעילה שהמודלים מצליחים לחזות את ערך המטרה - חומרת התקיפה.

במהלך ניתוח המודלים זיהינו כ 2-3 תכונות שהשפעתן על תוצאת החיזוי היא גבוהה ביחס לאחרות. תכונות אלו בלטו במודל ותרמו להסבר ולשיפור יכולת הניבוי שלו. בנוסף המודל חושף אילו יחידות בארגון ניתנים לפריצה ולתקיפה וזקוקים לשיפור ותחזוקה גבוהה יותר.

כחלק מתהליך העבודה, הרצנו גרסאות שונות של מודלים תוך שינוי פרמטרים כמו עומק העצים, קצב הלמידה, כמות האיטרציות, ובחנו כיצד השינויים האלו משפיעים על איכות החיזוי. ניסויים אלו הובילו לתובנות חשובות ושיפורים מדידים בביצועים. דירגנו כל מודל בהתאם לביצועיו, כך שיכולנו לזהות בצורה ברורה את המודל המוביל.

המודל יהווה יתרון עבור הארגון מכיוון שהוא יאפשר לארגון תיעדוף וטיפול בתקיפות חמורות יותר, דבר שיוביל להקצות להן משאבים בצורה חכמה יותר, ולבנות תוכנית תגובה מבוססת סיכון, ובכך ישנו שיפור ממשי בניהול האיומים ובהגנה על נכסי הארגון, מה שתומך בקבלת החלטות אסטרטגיות מדויקות יותר.

**איך נדרג את המודלים?**

המודלים ידורגו לפי היכולת שלהם לעמוד ביעדים העסקיים, תוך התמקדות בכמה פרמטרים -

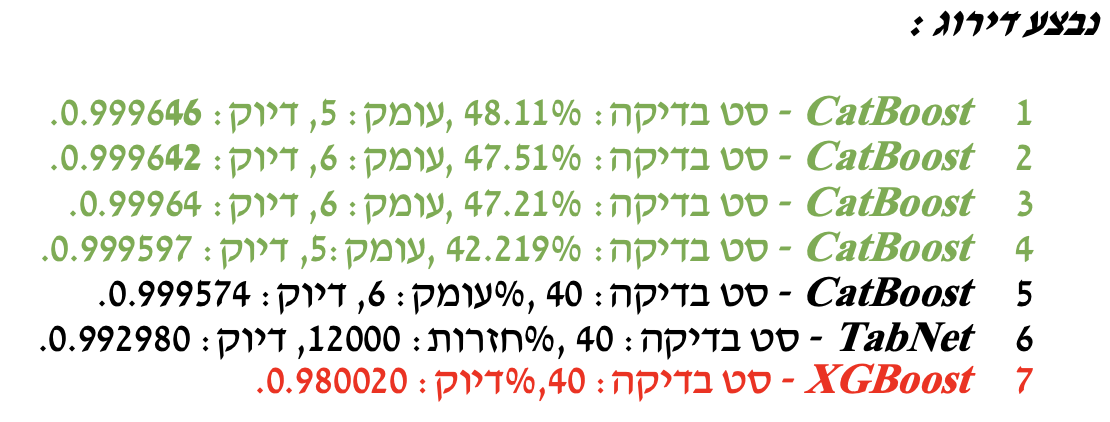
***יכולת חיזוי:*** המודל שיצליח לחזות את רמות הפגיעות ורמות החומרה בצורה המדויקת ביותר -ידורג גבוה יותר.

***פרמטרי המודל:*** נרצה להימנע ממודלים שיכולים לסווג תקיפות ״קלות״ כקשות, בכדי להימנע מהצמדות יתר - קבענו פרמטרים קבועים קטנים אך נשמור על דיוק גבוה.

הקריטריונים העסקיים להצלחה היו דיוק בזיהוי תקיפות - קבענו דיוק של מעל 90% למודל אמין ,אפקטיבי ,בעל זמן תגובה מהיר ,בעל הפחתת התראות שגויות - יכולת הפחתה של שיעור התחזיות השגויות שנחזו כלא מזיקות או להפך, ובעל חיסכון במשאבים.

השתמשנו במודלים שונים לצורך השוואה ובחינה של איזה מהם הכי יעיל ומתאים ליעד העסקי שלנו ומתוכם בחרנו את המודל שנתן את הביצועים הטובים ביותר.  
  
 דירוג המודלים מתבצע ע״פ רמת דיוקם לזיהוי ההתקפות בקבוצת הבדיקה וכך יתרמו ליעדים העסקיי בהמשך.

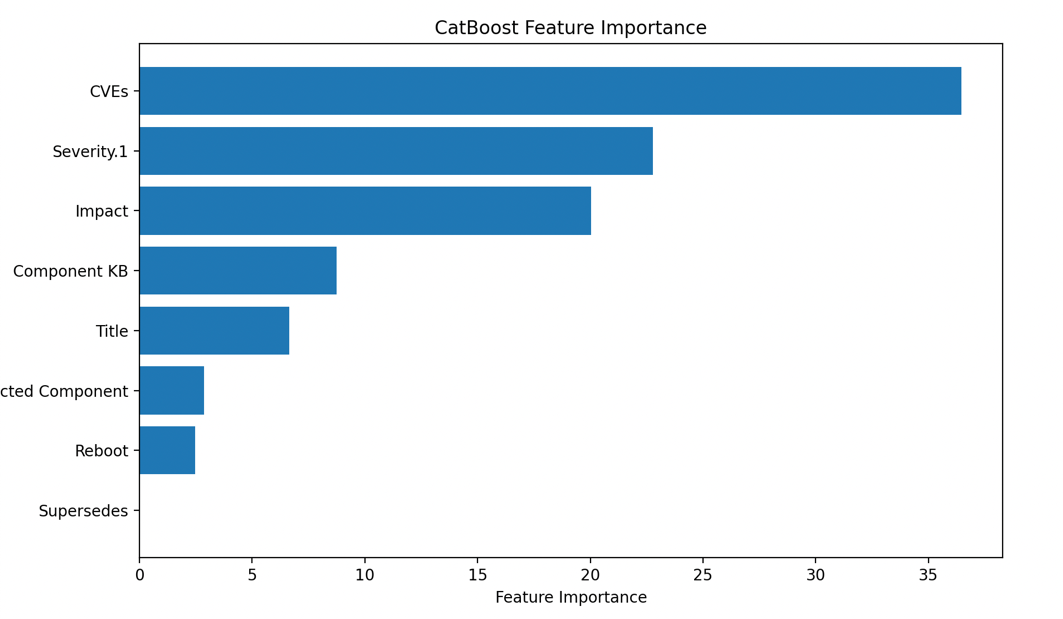
**אלו הם 7 המודלים החזקים ביותר שתורמים למטרה העסקית:**



**עם זאת במהלך פיתוח המודלים השונים למענה על המטרות התפתחו שאלות כגון :**

איור 2 : 7 המודלים המובילים עם פרמטרים הטובים ביותר

* האם ניתן לחזות את חומרת התקיפה לפי השפעת הבעיה על המערכת?
* מה מידת ההשפעה של מידת החומרה ראשונית על החיזוי המשני, האם יעילה?
* עד כמה מספר זיהוי (ID) של בעיית האבטחה הוא רלוונטי לחומרת התקיפה?
* האם כל תקיפה (חזקה או חלשה) מצריכה הפעלה מחדש לאחר העדכון?
* מה הקריטריונים שמאפיינים תקיפה חזקה?
* האם ניתן לחזות את הרכיב המושפע או המוצר שהושפע (מאותה תקיפה) ע״פ התכונות?



איור 3 : גרף המציג את התכונה המשפיעה ביותר

***לסיכום,***   
  
בשלב זה הערכנו את ביצועי המודלים לחיזוי חומרת תקיפות הסייבר, במטרה לאפשר תגובה מותאמת לסיכון ותיעדוף משאבים שיובילו ליעד העסקי. מודלים כמו **CatBoost**  הראו דיוק גבוה מאוד, במיוחד בזיהוי תקיפות חמורות וזוהו מספר מאפיינים קריטיים שתורמים לחיזוי. דירוג המודלים נעשה ע״פ דיוק וזמן התגובה.    
   
בנוסף הועלו מס' שאלות להמשך חקירה כגון : מהי השפעת החומרה הראשונית? מהי חשיבות מזהי תקיפות? והאם חקירה עתידית של משתנה אחר יכולה להועיל?.

שאלות אלו מהוות בסיס לפעולה עתידית במסגרת תהליך שיפור המודל, בדגש על שיטות Feature Engineering מתקדמות, ניתוח סדרות זמן, ושילוב מידע הקשרי נוסף, שיכול לתרום ליצירת מערכת ניבוי חכמה, מהירה ומדויקת יותר.

# **תהליך הסקירה :**

בסעיף זה נסקור את תהליך הלמידה שעברנו במהלך העבודה על הפרויקט ועל חברת מיקרוסופט. נבחן את ההחלטות והפעולות שביצענו – מה נעשה בצורה טובה, ומה ניתן היה לבצע טוב יותר. מטרת הסקירה היא להפיק לקחים משמעותיים אשר יסייעו לנו בפרויקטים עתידיים. נציין את התחומים שבהם ניתן היה להשתפר – החל משלב בחירת הנושא ועד להצגת הדוחות לאורך הדרך.

*כעת נעבור דוח – דוח וננתח את השלבים שביצענו וננסה להסיק מסקנות לפרויקטים הבאים.* ***• דוח בחירת הנושא***

בחירת הנושא היא שלב מרכזי וקריטי בפרויקט בתחום מדעי הנתונים, שכן מדובר בהחלטה אסטרטגית שמשפיעה על כלל שלבי הפרויקט – החל מהשגת הנתונים, דרך ניתוחם, ועד להסקת המסקנות. בפרויקט זה, בחרנו לעסוק בחיזוי רמת החומרה של תקיפות סייבר, בהתבסס על נתוני אבטחה שנאספו מארגון Microsoft.

מה הצליח?

* רלוונטיות - בחרנו נושא אקטואלי, רלוונטי ומעורר עניין.
* שילוב - הנושא משלב עולם תוכן עסקי וטכנולוגי – אבטחת מידע, ניתוח נתונים, מודלים חכמים.
* אמינות - הדאטה שנבחר מגיע ממקור אמין ומכיל מעל 20 אלף תצפיות.

מה ניתן לשפר?

* ניתן היה לבחון נושאים נוספים או לשלב כמה תתי נושאים וחיזוי מספר מטרות (כגון: חיזוי חומרה של תקיפה וגם זמן תגובה).
* כדאי היה לבדוק מראש אילו מדדים קיימים לצורך הערכת הצלחה עסקית ,בכדי להתאים את מטרות הניתוח.

***• דוח עסקי***

הדוח העסקי מקיף ומסביר את הארגון ממנו לקוחה הבעיה שבנתונים.

מה הצליח?

* הסבר על הארגון בצורה מפורטת ומקיפה, התעמקות במגוון רחב של נושאים של הארגון ומאפייניהם, הסבר כולל על הפרטים של הארגון הגדול.
* חיברנו את הניתוח למקרה עסקי ברור: תיעדוף תקיפות, חיסכון בזמן/כסף, שיפור תגובה ארגונית.

מה ניתן לשפר?

* ארגון מיקרוסופט הוא גוף גדול ומרכזי, אשר מושך אליו קהילות ענק של מדעני נתונים. חיפוש ממוקד ועמוק יותר אחר מקורות נתונים היה עשוי להוביל למציאת דאטה איכותי ורלוונטי אף יותר לפרויקט.
* מיקרוסופט פועלת בתחומים מגוונים ורבים. בדיעבד, היינו יכולים לצמצם את המיקוד ולהתעמק ביחידה מסוימת בתוך הארגון – לדוגמה, מחלקת Azure, מחלקת Microsoft Defender (המתמקדת באבטחת מידע), מחלקת Windows, מחלקת Microsoft 365, או מחלקת GitHub (שבבעלות מיקרוסופט). מיקוד כזה היה מאפשר להעמיק את ההבנה ולהפיק תובנות ממוקדות ומדויקות אף יותר.

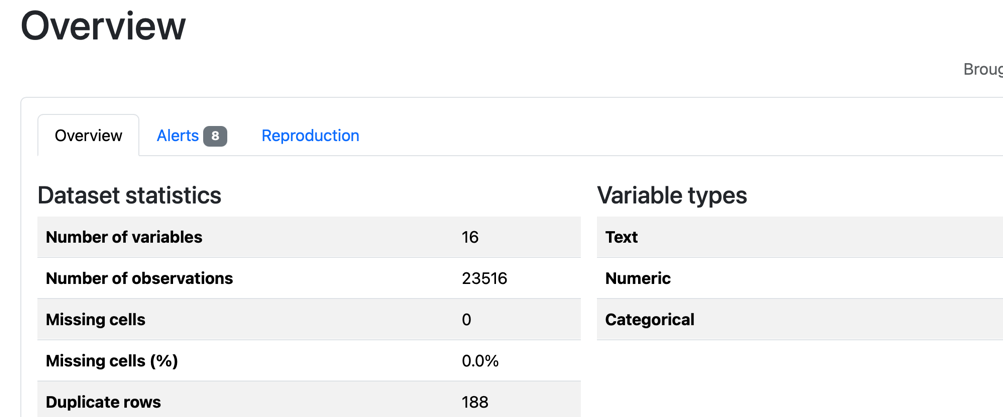
***• דוח הבנת נתונים***

דוח זה עוסק בהבנת מאפייני מערך הנתונים ששימש אותנו בפרויקט. הוא כולל ניתוח של איכות הנתונים, תיאור של השדות והמשתנים השונים, הסבר על המשמעויות שלהם, וכן סקירה של קשרים ותבניות שנמצאו בין משתנים. מטרת הדוח היא לספק תמונה ברורה ומעמיקה של הנתונים לפני תחילת שלבי הניתוח והבניית המודלים.

מה הצליח?

* זיהנו את סוגי הפיצ׳רים וקשרים שונים למשתנה המטרה.
* ביצענו ויזואליזציות וגרפים שונים שמסבירים את הנתונים.

מה ניתן לשפר?

* יכולנו להשתמש בספריות אוטומטיות לחקר נתונים כמו import ProfileReport - שזוהי ספרייה שיוצרת דף HTML מותאם עבור ה -.Data Frame



***הסבר -***

- import ProfileReportהוא דוח המיועד לניתוח והבנה של פרופיל הנתונים המיובאים לארגון או למערכת. דוח זה מספק תובנות לגבי הנתונים שנכנסים למערכת, כולל מאפיינים, איכות, והקשרים בין הנתונים השונים.

***• דוח הכנת הנתונים***

דוח זה מתאר את שלבי ההכנה והעיבוד שבוצעו על מערך הנתונים במטרה להתאימו ללמידת מכונה. תהליך ההכנה כלל טיפול בערכים חסרים (NA), תיקון שגיאות, סינון משתנים לא רלוונטיים, המרת טיפוסי נתונים, והנדסת תכונות במידת הצורך. שלבים אלה בוצעו על מנת לשפר את איכות הדאטה ולהבטיח שהמודל יתבסס על מידע נקי, עקבי ומשמעותי.

מה הצליח?

* ביצענו ניקוי כולל עם שלבים ברורים ומוגדרים.
* בדקנו קשרים חשובים למשתנה מטרה.
* בחרנו את העמודות הרלוונטיות

מה ניתן לשפר?

* תהליך ניקוי הנתונים היה יכול להתבצע בדרכים נוספות, כגון שימוש בשיטות חכמות יותר למילוי ערכים חסרים (כמו חישוב ממוצעים לפי קבוצות, שימוש במודלים לחיזוי ערכים חסרים, או טכניקות איטרפטיביות), במקום הסרה או מילוי גורף.
* אופן ההתמודדות עם רשומות חלקיות – לדוגמה, רשומות הכוללות ערכים חסרים – היה יכול להיבחן לעומק: במקרים מסוימים ייתכן שהיה עדיף להשאיר את הרשומות ולמלא ערכים חסרים, ולא להסירן, על מנת לשמר כמה שיותר מהמידע המקורי.

***• דוח המידול***  
  
דוח זה מתאר את תהליך הבנייה וההערכה של אלגוריתמים שנועדו לחזות את תכונת המטרה שנבחרה במסגרת הפרויקט. הוא כולל סקירה של האלגוריתמים שנבחרו, הסבר על אופן הבחירה בהם, תיאור של שלבי האימון והבדיקה של המודלים, וכן ניתוח ביצועים באמצעות מדדים מתאימים (כגון: דיוק, F1, AUC, ועוד). מטרת הדוח היא להציג כיצד תהליך המידול בוצע בפועל, מהן התוצאות שהתקבלו, ומה ניתן להסיק מהן לצורך חיזוי יעיל ואמין.

מה הצליח?

* בפרויקט זה בחנו מספר מודלים מתקדמים, בהם: CatBoost, LightGBM, XGBoost ואחרים.
* השווינו בין המודלים באמצעות מדדי ביצועים שונים, וביצענו ניסויים עם פרמטרים מגוונים על מנת לשפר את תוצאות החיזוי. התהליך כלל הערכת ביצועים, כוונון היפר-פרמטרים ידני, והשוואה שיטתית בין האלגוריתמים.

מה ניתן לשפר?

* + ניתן היה לשלב כלים אוטומטיים מתקדמים כגון Optuna – ספרייה לאופטימיזציית פרמטרים המבצעת חיפוש חכם ומהיר אחר שילובי הפרמטרים הטובים ביותר למודלים, ובכך לייעל את תהליך הכוונון ולחסוך זמן.
  + כמו כן, ייתכן והיה ניתן לשלב שיטות למידת Ensemble נוספות (כגון: Stacking או Blending - שילוב תוצאות של כמה מודלים בעזרת מודל נוסף "עליון" (meta-model)) כדי לשפר את הביצועים הכלליים של המודל.
* בנוסף, שילוב של כלי AutoML – אשר מבצעים אוטומציה לכל שלבי בניית המודל, מבחירת האלגוריתם ועד לכוונון פרמטרים – עשוי היה להוביל לפתרונות מדויקים ויעילים יותר, תוך חיסכון משמעותי בזמן ובמשאבים.

*כלים של AutoML לבחינה עתידית:*

במהלך המשך העבודה והעמקת היכולות בתחום חיזוי ואופטימיזציה של מודלים, ניתן לשקול שימוש בכלים מתקדמים של AutoML שמבצעים אוטומציה לתהליך המלא של בניית מודלים. להלן מספר כלים בולטים:

* **H2O AutoML**

תומך במשימות של רגרסיה, סיווג וסדרות זמן. משלב מודלים מתקדמים כגון Random Forest ו־XGBoost, וכולל תהליך הרכבה אוטומטית (Ensembling) שמאפשר שיפור ביצועים על ידי שילוב מספר מודלים.

* **Auto-sklearn**

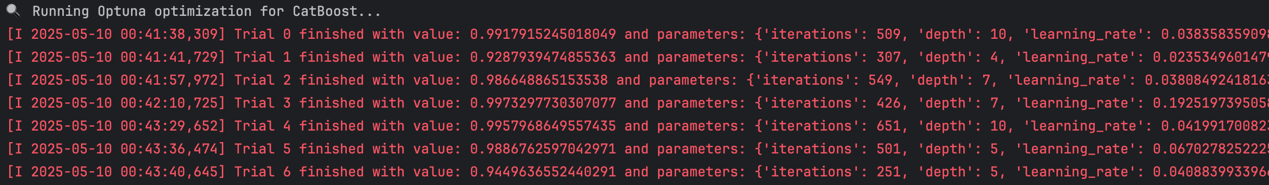
מבוסס על ספריית scikit-learn, ותומך ברגרסיה וסיווג. משתמש באלגוריתמים מוכרים כגון Random Forest ו־SVM, ומבצע חיפוש חכם אחר שילוב המודלים והפרמטרים האופטימליים.

* **PyCaret**

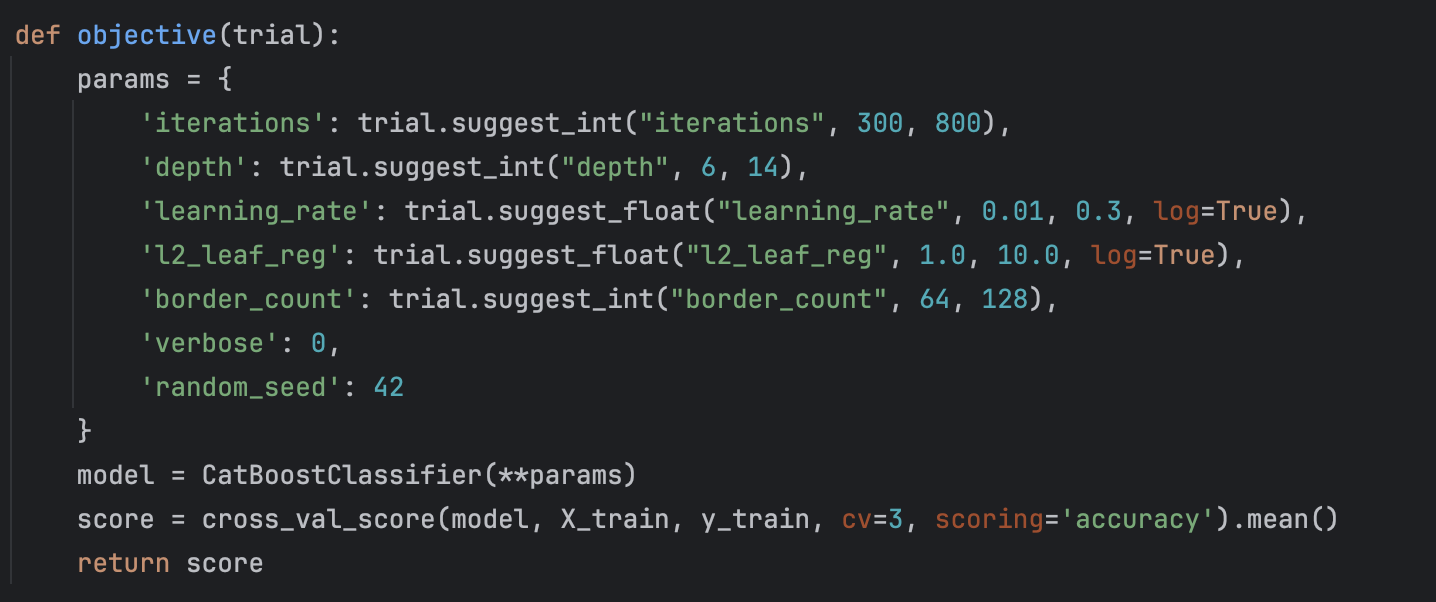
מספקת ממשק פשוט ונוח ליישום משימות של סיווג, רגרסיה וניתוח שפה טבעית (NLP). עושה שימוש באלגוריתמים כמו LightGBM ו־XGBoost, ומבצעת את כל תהליך ההשוואה והכוונון של מודלים בפקודות בודדות.

* **FLAML (Fast and Lightweight AutoML)**

מתמקדת בהפחתת עלויות חישוב וזמן ריצה. תומכת ברגרסיה, סיווג וסדרות זמן, ומבוססת בעיקר על מודלים כגון LightGBM ו־ XGBoost. אידיאלית לפרויקטים שדורשים פתרונות יעילים בזמן קצר.

* תהליך optuna ופלט מלא:

איור 4 : תהליך ההרצה של Optuna



איור 5 : הגדרת טווח הפרמטרים לחיפוש

בתמונה ניתן לראות את Optuna בשלבי הריצה שלה , כל שורה אדומה היא הרצה של מודל CatBoost או מודל לבחירה עם טווח הפרמטרים שניתנו למודל, לדוגמא בשורה הראשונה מתקבל מדד value והפרמטרים של המודל. Optuna מריצה מספר מודלים בהתבסס לקלט המשתמש או אם אין שינוי מהותי בvalue והמודל לא משפר את עצמו, Optuna מסתיימת ומראה את התוצאות הרצויות של המודל בעל הvalue הטוב ביותר. עבור דאטה רב, Optuna מריצה מספר רב של מודלים עם ביג דאטה.

ב-"**Trial 0**", המודל הוגדר עם 509 עצים, עומק מקסימלי של 10, שיעור למידה של 0.038, רגולציה של 0.135 וגבול של 148. הערך של פונקציית המטרה עבור ניסיון זה היה 0.9917915245018049. כל ניסיון מסתיים עם סטטוס "finished", המציין שהניסיון הושלם בהצלחה.

תהליך האופטימיזציה מנסה למצוא את הצירוף הטוב ביותר של פרמטרים כדי לשפר את הביצועים של המודל, כאשר כל ניסיון מציג את הערכים השונים של הפרמטרים ואת התוצאה המתקבלת. זהו תהליך אטרקטיבי שמטרתו למצוא את המודל האופטימלי עבור המשימה הנתונה.

**כעת ננתח את שלבי הריצה של Optuna:**

מה קורה בכל הרצה?

בכל הרצה Optuna בוחרת ערכים להיפר-פרמטרים כמו מספר העצים, עומק העצים , גודל הדגימה ועוד. המודל מאומן עם הערכים שנבחרו, מבוצעת הערכה של הביצועים ונשמרת תוצאה. על סמך תוצאות אלו, Optuna לומדת אילו שילובים משתלמים יותר ומעדכנת את הבחירות שלה בניסויים הבאים – כך נבנית אופטימיזציה חכמה.

מה קורה עם ביג דאטה?

כאשר מדובר בביג דאטה, משך הזמן של כל הרצת מודל מתארך משמעותית. לכן, מומלץ לבחור במדדי ביצוע כמו -Cross Validation בנוסף, ייתכן שיידרשו יותר ניסויים כדי להגיע למודל האופטימלי. במצבים כאלה חשוב לנצל תשתיות כמו GPU או ריבוי ליבות, ולהפעיל מנגנון שמפסיק הרצת מודלים שלא מראים שיפור מובהק.

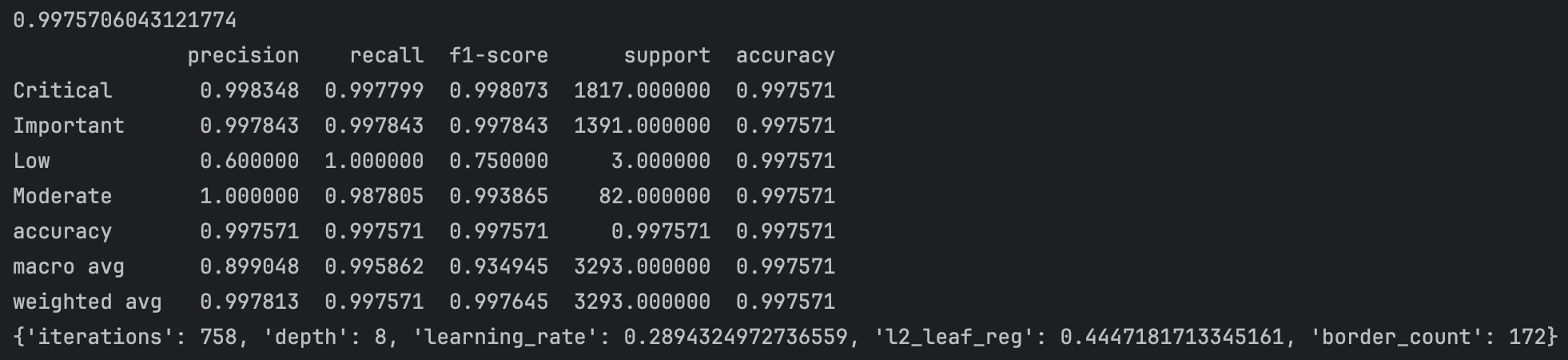
איך Optuna יודעת מתי לעצור?

ניתן להגדיר מראש מספר ניסויים מקסימלי (n\_trials) או הגבלת זמן כוללת (timeout). מעבר לכך, Optuna עשויה להפסיק את תהליך האופטימיזציה מוקדם יותר, אם היא מזהה שאין שיפור משמעותי בתוצאה – לדוגמה, אם 20 הרצות אחרונות לא הביאו לשיפור במדד היעד.

בסיום התהליך, Optuna מדפיסה את המודל עם התוצאה הטובה ביותר, יחד עם ערכי ההיפר-פרמטרים שהניבו אותה. ניתן לשמר את המודל, לצייר גרפים של ביצועים לאורך הזמן, ולהשוות בין מודלים שונים.

תהליך Optuna **אינו מתאים לביג דאטה** ויכול לקחת מספר שעות לאמן על ביג דאטה, עם זאת נריץ על הדאטה המקורי שכולל מעל 20,000 רשומות.

עבור 15 הרצות שיפור קבלנו את הנתונים הבאים , accuracy = 0.9975 :



איור 6 : מדדי Optuna

*הסבר פלט פרמטרים :*   
  
במהלך תהליך הטיוב (Hyperparameter Tuning) של מודל Random Forest, התקבלו הפרמטרים הבאים אשר סיפקו את הביצועים הטובים ביותר:

**iterations**: מספר האיטרציות (עצים) שנבנו במהלך האימון של המודל. ערכים גבוהים יותר

מציינים מודל מורכב יותר.

**depth**: העומק המקסימלי של כל עץ במודל. ערכים גבוהים יותר מציינים עצים מורכבים יותר.

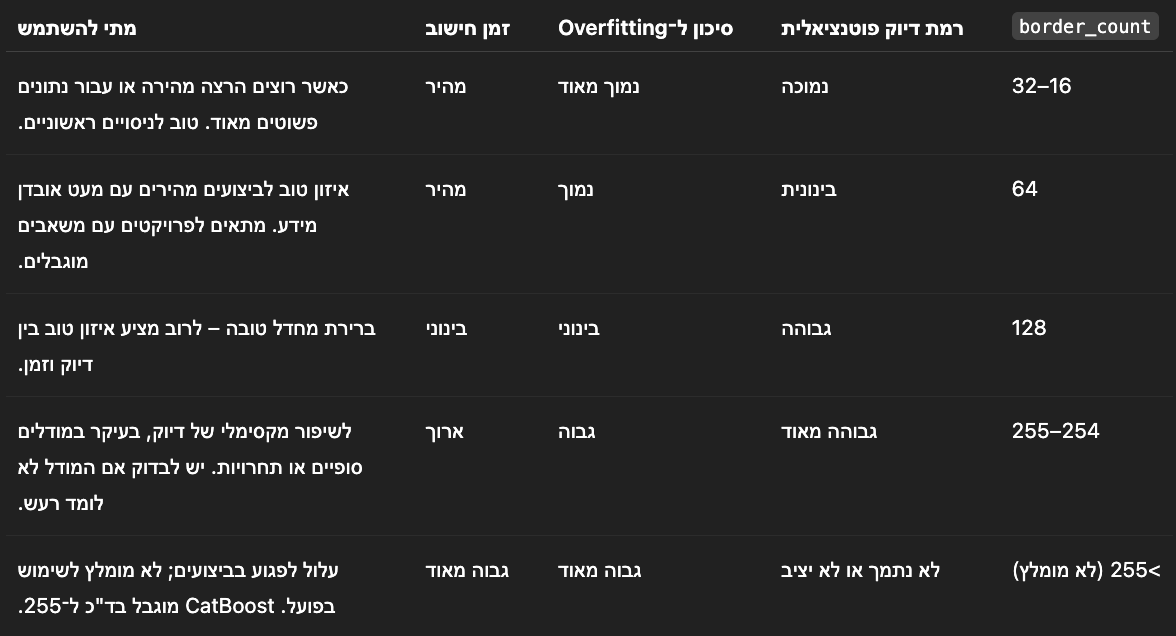
**learning\_rate**: קצב הלמידה של המודל, המשפיע על כמה כל עץ חדש תורם לתיקון השגיאות

של העצים הקודמים. ערכים נמוכים יותר מציינים תיקון איטי יותר, אך יכולים להוביל לביצועים טובים יותר.

**l2\_leaf\_reg**: פרמטר רגולריזציה L2 עבור העלים בעצים. ערכים גבוהים יותר מציינים

רגולריזציה חזקה יותר, מה שיכול למנוע התאמה יתר (overfitting).

**border\_count**: מספר הגבולות (split points) המשמשים לחלוקת התכונות בעצים. ערכים גבוהים יותר מציינים חלוקות מורכבות יותר.



איור 7 : תכונת border count והערכה של ערכים שלו

עם קידוד נכון של הנתונים אפשר להפעיל את תהליך Optuna על מספר מודלים שונים ביניהם XGBoost , lightGBM ,Random Forest ועוד.

ניתן להניח כי שימוש בתהליך Optuna יראה תוצאות אופטימליות יותר, הפחתת רעש והפחתת התאמת יתר הן תוצאות ישירות של השימוש ב-Optuna לאופטימיזציה של היפר-פרמטרים. הפחתת רעש יכולה להתרחש כאשר אנו מתאימים את הפרמטרים בצורה חכמה, במקום להסתמך על ניסויים אקראיים או הערכות ידניות של פרמטרים. חיפוש חכם אחרי ערכים אופטימליים בעזרת Optuna מצמצם את האפשרות להגיע לפרמטרים לא מתאימים או לא אופטימליים שמובילים לרעש בתוצאות, בכך שהוא מקטין את השפעת הסטיית-תקן של תחזיות המודל (variance). בנוסף, הפחתת התאמת יתר מתבצעת כיוון ש-Optuna מבצע אופטימיזציה

תוך שימוש במדדי ביצועים. ולרוב אנחנו משתמשים גם בנתונים נפרדים כדי לבדוק את הביצועים על נתוני מבחן.

Optuna מבצע אופטימיזציה תוך כדי שימוש ב-validation על כל פרמטר, כך שנמנע מהמודל ללמוד יותר מדי על הנתונים (overfitting) ולהתאים את עצמו בדיוק יתר על המידה לנתונים ספציפיים של קבוצת האימון, ובכך מצמצם את הסיכון של התאמת יתר.