

עבודת גמר

Decision Support Systems

Azure חיזוי ע"י תוכנת

שם המרצה: פר' אופיר בן אסולי

שם	ת.ז
חן אמסלם (ראש הקבוצה)	037092137
בשיר פארס	316373679
רותם ביטי	319056453
דמאס פיקדה	328954821

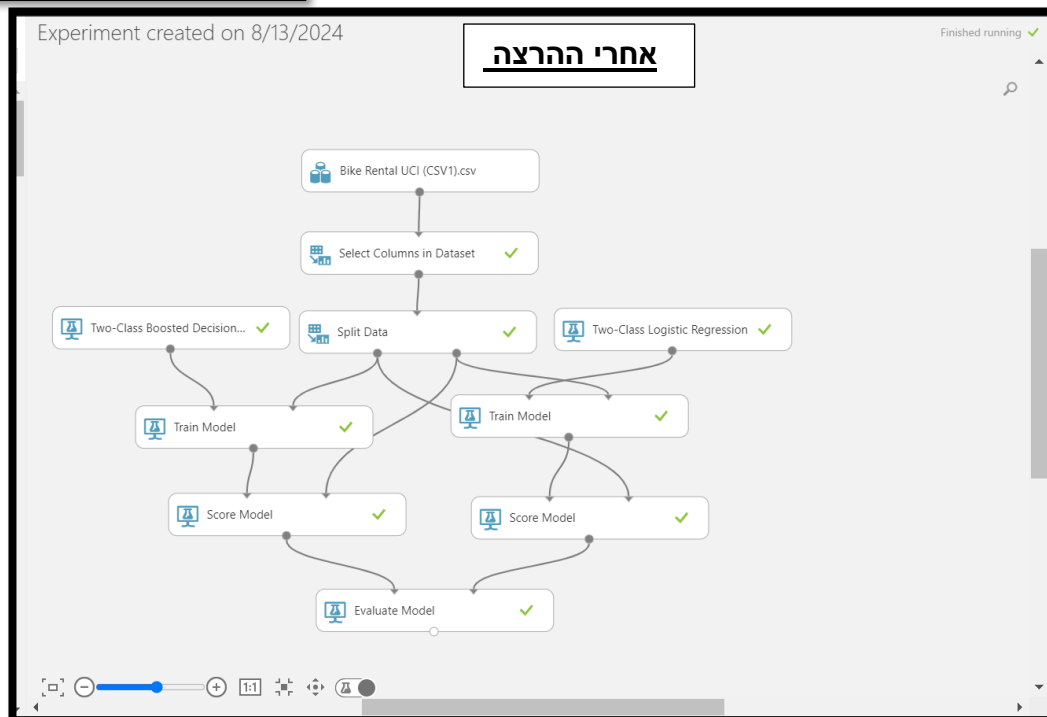
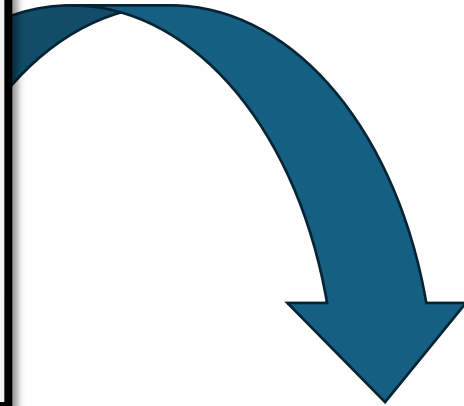
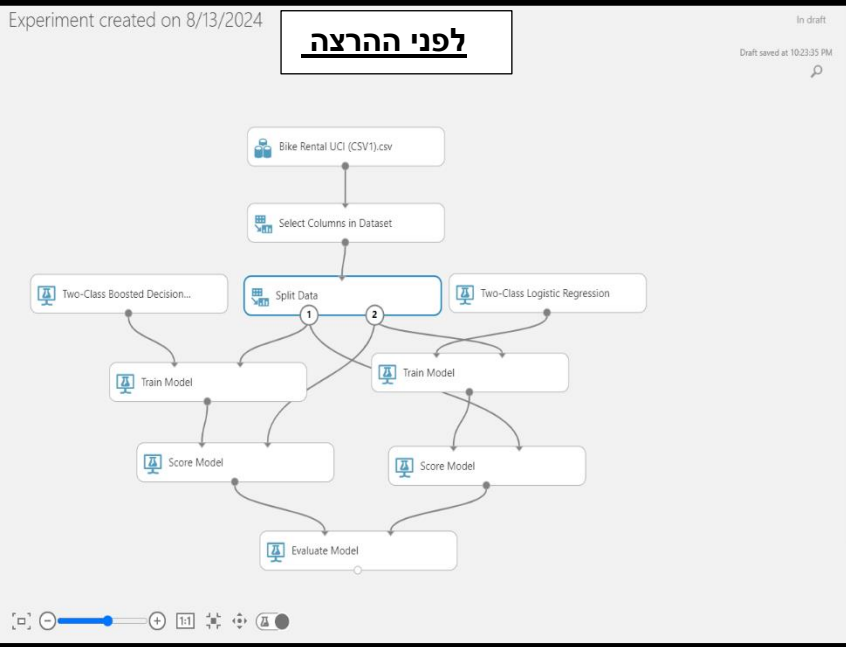
ניתוח ההבדלים בין תוצאות החיזוי של המודלים

Two-Class Boosted Decision Tree



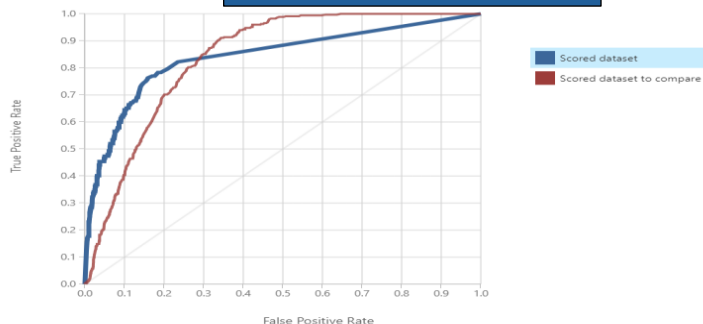
Two Class Logistic Regression

חלוקה של 70% אימון ו- 30% בדיקה



ROC PRECISION/RECALL LIFT

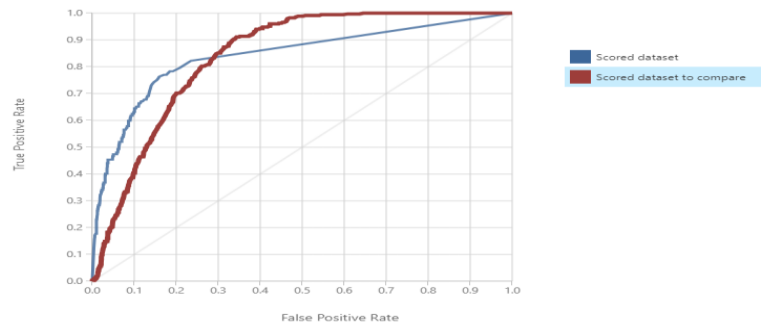
Two-Class Boosted Decision Tree



True Positive	False Negative	Accuracy	Precision	Threshold	AUC
73	79	0.894	0.490	0.5	0.862
False Positive	True Negative	Recall	F1 Score		
76	1234	0.480	0.485		
Positive Label	Negative Label				
1	0				

ROC PRECISION/RECALL LIFT

Two Class Logistic Regression



True Positive	False Negative	Accuracy	Precision	Threshold	AUC
1	348	0.901	0.250	0.5	0.840
False Positive	True Negative	Recall	F1 Score		
3	3186	0.003	0.006		
Positive Label	Negative Label				
1	0				

פיענוח תוצאות החיזוי והשוואה ביניהם

AUC	F1 score	Precision	Recall	Accuracy	NL	TN	FN	PL	FP	TP	MODEL
0.862	0.485	0.490	0.480	0.894	0	1234	79	1	76	73	Two-Class Boosted Decision Tree
0.840	0.006	0.250	0.003	0.901	0	3186	348	1	3	1	Two Class Logistic Regression

מודל 1:

Two-Class Boosted Decision Tree

AUC (0.862): •

מודל זה מקבל ערך AUC גבוה יותר, מה שמעיד על יכולת טובה יותר להבחין בין המעמדות.

F1 Score (0.485): •

ציון F1 הוא ממוצע משוקלל של Precision ו Recall, ומשקף איזון בין דיוק לבין זיהוי נכון. כאן הוא נמצא בטווח סביר, אך עדיין לא גבוה.

Precision (0.490): •

Precision מציין כמה מהתחזיות החיוביות נכונות. המודל הזה מצליח לנבא די טוב במונחים של דיוק חיובי, אך עדיין יש מקום לשיפור.

Recall (0.480): •

Recall מראה את היכולת לזהות נכונה את המקרים החיוביים. הערך כאן נמוך יחסית, כלומר המודל לא מצליח לתפוס את כל המקרים החיוביים.

Accuracy (0.894): •

המדד הזה גבוה מאוד, כלומר רוב התחזיות של המודל נכונות, אך זה יכול להיות תלוי באיזון הנתונים.

מדדי: TN, FN, FP, TP: •

True Negatives (TN = 1234): •

False Negatives (FN = 79): •

False Positives (FP = 76): •

True Positives (TP = 73): •

מודל 2:

Two-Class Logistic Regression

- **AUC (0.840):**
AUC נמוך יותר מהמודל הקודם, כלומר הוא פחות טוב בזיהוי נכון של המעמדות.
- **F1 Score (0.006):**
זהו ציון נמוך ביותר, מה שמעיד על כך שהמודל לא מצליח לאזן בין דיוק (Precision) לזיהוי (Recall) בצורה טובה.
- **Precision (0.250):**
דיוק נמוך מאוד, כלומר המודל מצליח בניבוי חיובי רק במעט מהמקרים.
- **Recall (0.003):**
Recall נמוך ביותר, המודל לא מצליח לזהות כמעט בכלל מקרים חיוביים.
- **Accuracy (0.901):**
דיוק כללי טוב, אך ככל הנראה נובע מהטיה בגלל שיש יותר מקרים שליליים (מאוזן יותר לטובת התחזיות השליליות).
- **מדדי: TN, FN, FP, TP**
 - **True Negatives (TN = 3186):** זיהוי נכון של מקרים שליליים גבוה יותר, אך זה לא מצביע על ביצועים כלליים טובים.
 - **False Negatives (FN = 348):** כמות גבוהה מאוד של טעויות בזיהוי מקרים חיוביים.
 - **False Positives (FP = 3):** מספר נמוך של טעויות בתחזיות חיוביות שגויות.
 - **True Positives (TP = 1):** המודל זיהה רק מקרה חיובי אחד נכון.

מסקנות:

- **Two-Class Boosted Decision Tree** הוא המודל הטוב יותר מבין השניים. למרות ש-Precision ו-Recall אינם מושלמים, הוא מצליח לזהות מקרים חיוביים (TP) בצורה טובה הרבה יותר מאשר ה-Logistic Regression.
- Two-Class Logistic Regression נכשל בניבוי מקרים חיוביים ומקבל ציון נמוך מאוד ב-Precision. למרות שהוא מקבל Accuracy גבוה, זה נובע מכך שיש יותר מקרים שליליים, והמודל מצליח לזהות אותם היטב, אך כמעט שלא מזהה חיוביים.

ניתוח ההבדלים בין תוצאות החיזוי של המודלים

Two Class Neural Network



Two Class Support Vector Machine

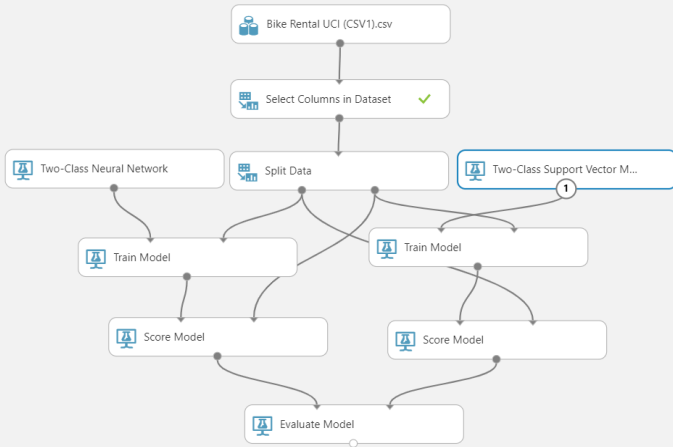
חלוקה של 80% אימון ו- 20% בדיקה

Experiment created on 8/13/2024

לפני ההרצה

In draft

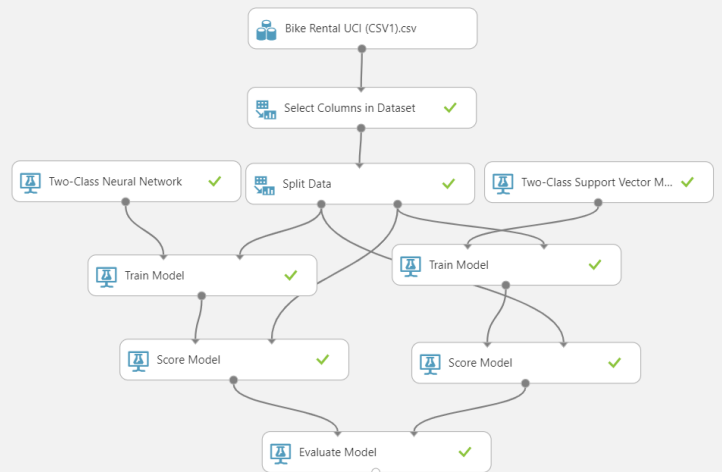
Draft saved at 10:26:14 PM

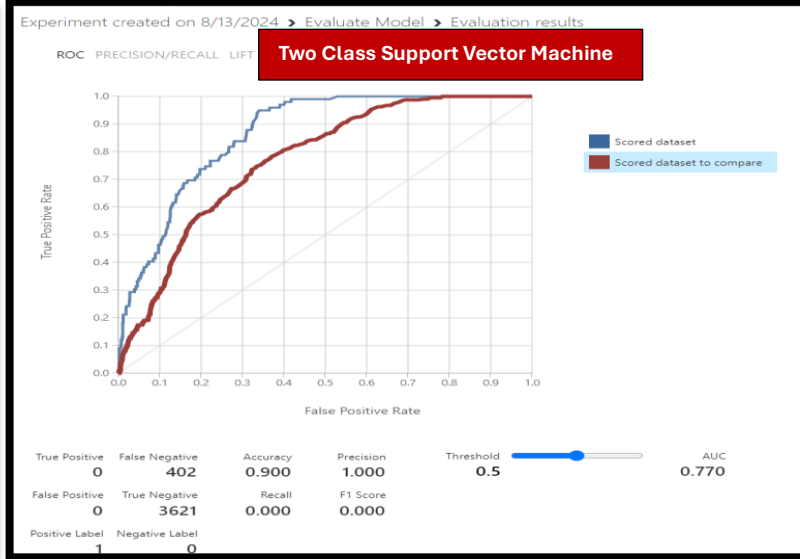
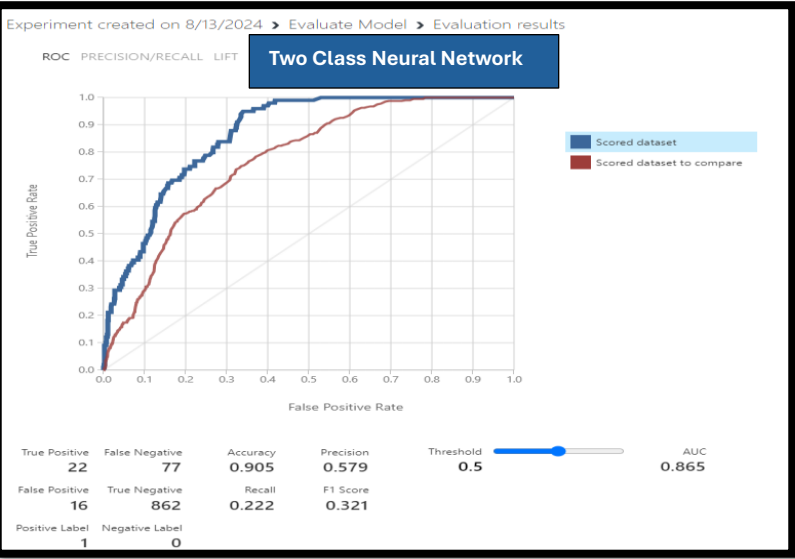


Experiment created on 8/13/2024

אחרי ההרצה

Finished running ✓





פיענוח תוצאות החיזוי והשוואה ביניהם

AUC	F1 score	Precision	Recall	Accuracy	NL	TN	FN	PL	FP	TP	MODEL
0.865	0.321	0.579	0.222	0.905	0	862	77	1	16	22	Two-Class Neural Network
0.770	0.000	1.000	0.000	0.900	0	3621	402	1	0	0	Two Class Support Vector Machine

מודל 1:

Two-Class Neural Network

- AUC (0.865): ערך AUC טוב מאוד, מה שמעיד על כך שהמודל יכול להבחין בצורה טובה בין המעמדות.
- F1 Score (0.321): ציון F1 נמוך יחסית, מה שמעיד על כך שיש חוסר איזון בין Precision ל-Recall.
- Precision (0.579): דיוק ממוצע. כ-57.9% מהתחזיות החיוביות היו נכונות, כלומר, למודל יש דיוק סביר בזיהוי מקרים חיוביים.
- Recall (0.222): Recall נמוך, מה שמצביע על כך שהמודל מצליח לזהות רק 22.2% מהמקרים החיוביים. זהו ביצוע חלש בזיהוי המקרים החיוביים האמיתיים.
- Accuracy (0.905): דיוק כללי גבוה, אך שוב, זה יכול לנבוע מהטיה בגלל כמות גבוהה של מקרים שליליים בזיהוי.

- מדדי: TN, FN, FP, TP:
 - True Negatives (TN = 862): 862 מקרים שליליים.
 - False Negatives (FN = 77): 77 טעויות בזיהוי מקרים חיוביים.
 - False Positives (FP = 16): 16 תחזיות חיוביות שגויות.
 - True Positives (TP = 22): 22 מקרים חיוביים.

מודל 2 :

Two-Class Support Vector Machine

- AUC (0.770):
- AUC נמוך יחסית, מה שמעיד על יכולת פחות טובה של המודל להבחין בין המעמדות.
- F1 Score (0.000):
- זהו ציון F1 נמוך מאוד, מה שמצביע על כך שהמודל נכשל לחלוטין באיזון בין Precision ל- Recall.
- Precision (1.000):
- Precision גבוה מאוד, אבל זה מטעה, מכיוון שיש מעט מאוד זיהויים חיוביים אמיתיים ($TP = 0$).
- Recall (0.000):
- Recall שואף ל-0, מה שמעיד על כך שהמודל נכשל לחלוטין בזיהוי מקרים חיוביים.
- Accuracy (0.900):
- Accuracy גבוה, אך זה שוב נובע מכמות גדולה של מקרים שליליים.
- מדדי: TN, FN, FP, TP:
- True Negatives (TN = 3621): 3621 מקרים שליליים.
- False Negatives (FN = 402): 402 טעויות בזיהוי מקרים חיוביים.
- False Positives (FP = 0): 0 טעויות חיוביות שגויות.
- True Positives (TP = 0): 0 זיהוי נכון של מקרים חיוביים.

מסקנות:

- **Two-Class Neural Network** מצליח טוב יותר מאשר ה-Support Vector Machine (SVM) למרות שה Recall-נמוך יחסית, לפחות הוא מצליח לזהות מקרים חיוביים, ($TP = 22$) ולמרות שיש טעויות, ($FP = 16$) הוא מתפקד טוב יותר באופן כללי בהשוואה ל-SVM.
- Two-Class Support Vector Machine נכשל בזיהוי מקרים חיוביים, ($TP = 0$) ולכן המודל הזה לא שימושי במיוחד במקרה הזה. למרות שדיוקו הכללי (Accuracy) גבוה, זה נובע אך ורק מזיהוי נכון של מקרים שליליים.

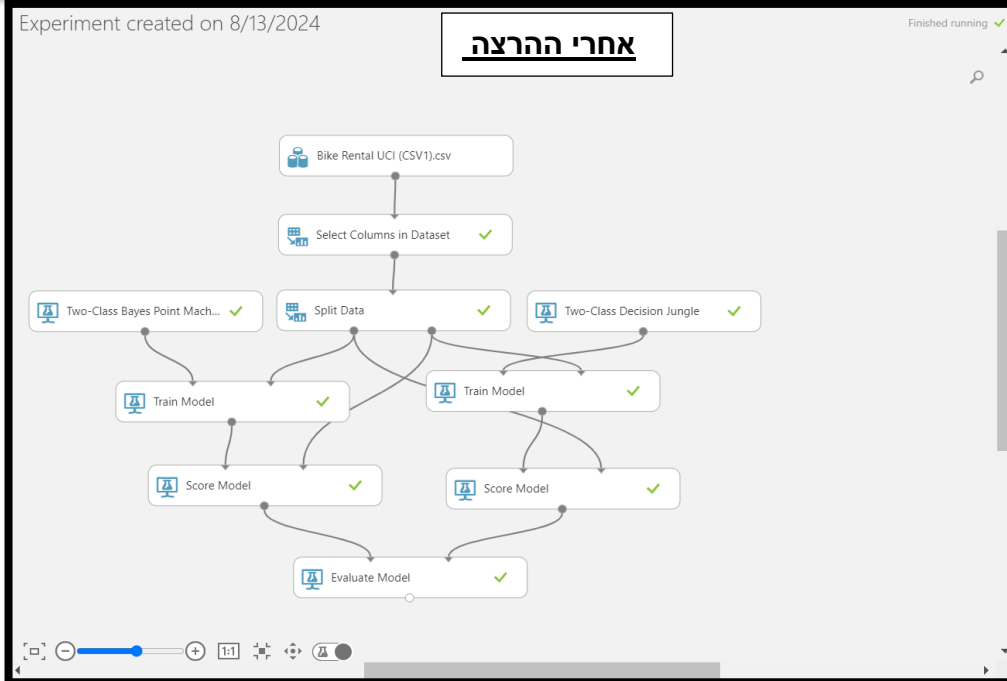
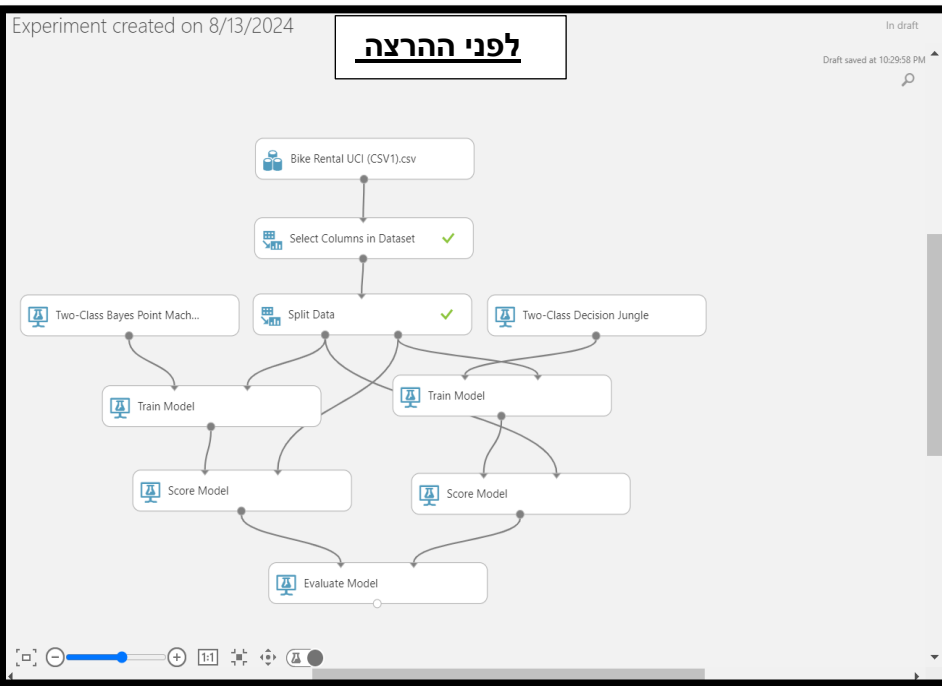
ניתוח ההבדלים בין תוצאות החיזוי של המודלים

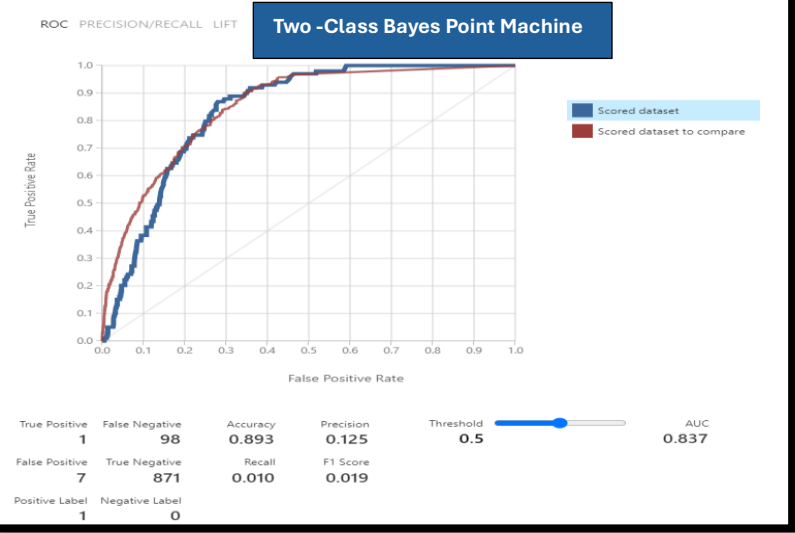
Two -Class Bayes Point Machine



Two -Class Decision Jungle

חלוקה של 80% אימון ו- 20% בדיקה





פיענוח תוצאות החיזוי והשוואה ביניהם

AUC	F1 score	Precision	Recall	Accuracy	NL	TN	FN	PL	FP	TP	MODEL
0.837	0.019	0.125	0.010	0.893	0	871	98	1	7	1	Two-Class Bayes Point Machine
0.851	0.196	0.687	0.114	0.906	0	3600	356	1	21	46	Two Class Decision Jungle

מודל 1

Two-Class Bayes Point Machine

- AUC (0.837):** ערך AUC סביר אך נמוך בהשוואה למודלים קודמים, מה שמעיד על כך שהמודל מתקשה להבחין בצורה טובה בין המעמדות.
- F1 Score (0.019):** ציון F1 נמוך מאוד, המצביע על כך שהמודל מתקשה באיזון בין Precision ל-Recall.
- Precision (0.125):** דיוק נמוך מאוד (12.5%), כלומר רוב התחזיות החיוביות היו שגויות.
- Recall (0.010):** Recall נמוך מאוד (1%), מה שמעיד על כך שהמודל כמעט ואינו מזהה מקרים חיוביים.
- Accuracy (0.893):** Accuracy גבוה, אך גם זה נובע מהטיה לזיהוי שלילי, שכן רוב המקרים הם שליליים.
- מדדי: TN, FN, FP, TP:**
 - True Negatives (TN = 871):** המודל זיהה נכון 871 מקרים שליליים.
 - False Negatives (FN = 98):** 98 מקרים חיוביים שלא זוהו.
 - False Positives (FP = 7):** 7 תחזיות חיוביות שגויות.
 - True Positives (TP = 1):** רק מקרה חיובי אחד זוהה נכון.

מודל 2:

Two-Class Decision Jungle

- **AUC (0.851):**
- AUC גבוהה יותר מהמודל הקודם, המצביע על יכולת מעט טובה יותר להבחין בין המעמדות.
- **F1 Score (0.196):**
- ציון F1 נמוך אך טוב יותר בהשוואה למודל הראשון, מה שמצביע על כך שהמודל מצליח לאזן במעט בין Precision ל-Recall.
- **Precision (0.687):**
- דיוק גבוה יחסית, מה שמעיד על כך שכאשר המודל מנבא מקרה חיובי, הוא נוטה להיות מדויק ב-68.7% מהמקרים.
- **Recall (0.114):**
- Recall נמוך (11.4%), המודל מצליח לזהות מעט מאוד מקרים חיוביים.
- **Accuracy (0.906):**
- Accuracy גבוה, אך גם כאן ייתכן והוא נובע מהזיהוי הנכון של המקרים השליליים.
- **מדדי: TN, FN, FP, TP**
- **True Negatives (TN = 3600):** 3600 מקרים שליליים.
- **False Negatives (FN = 356):** מספר גבוה של טעויות בזיהוי מקרים חיוביים.
- **False Positives (FP = 21):** 21 תחזיות חיוביות שגויות.
- **True Positives (TP = 46):** 46 מקרים חיוביים זוהו נכון.

מסקנות:

- Two-Class Bayes Point Machine ביצועי המודל נמוכים מאוד, בעיקר בגלל Recall נמוך מאוד (0.01) וציון F1 נמוך. למרות שיש לו Accuracy סביר, הוא לא מזהה מקרים חיוביים בצורה אפקטיבית.
- **Two-Class Decision Jungle** מודל זה מתפקד טוב יותר, במיוחד ב-Precision (68.7%) - אך יש לו Recall נמוך מאוד. הוא מצליח לזהות יותר מקרים חיוביים מאשר Bayes Point Machine, אך עדיין עם חוסר איזון.

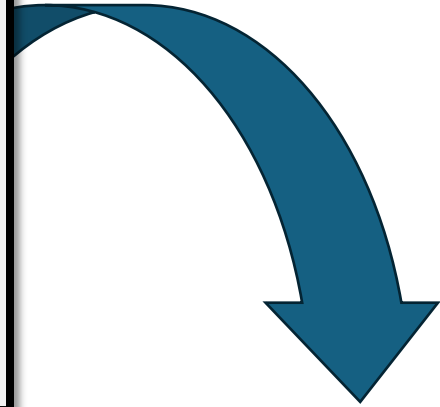
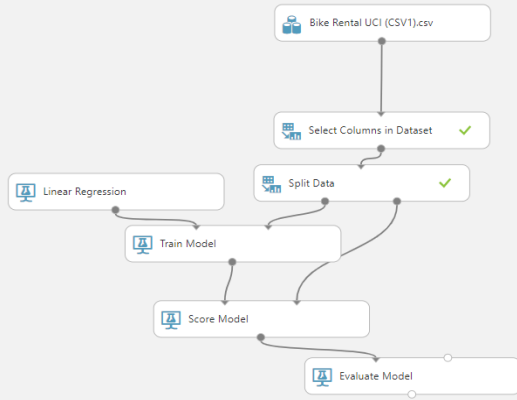
Linear Regression רגרסיה לינארית

Experiment created on 8/13/24

לפני ההרצה

In draft

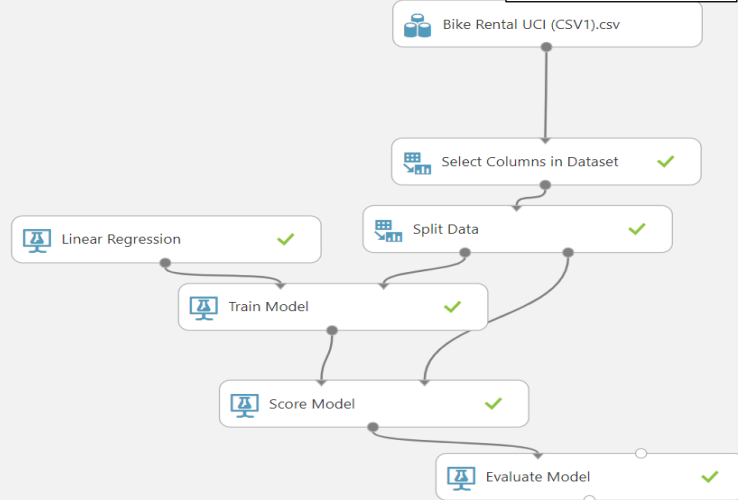
Draft saved at 4:32:12 PM



Experiment created on 8/13/2024

אחרי ההרצה

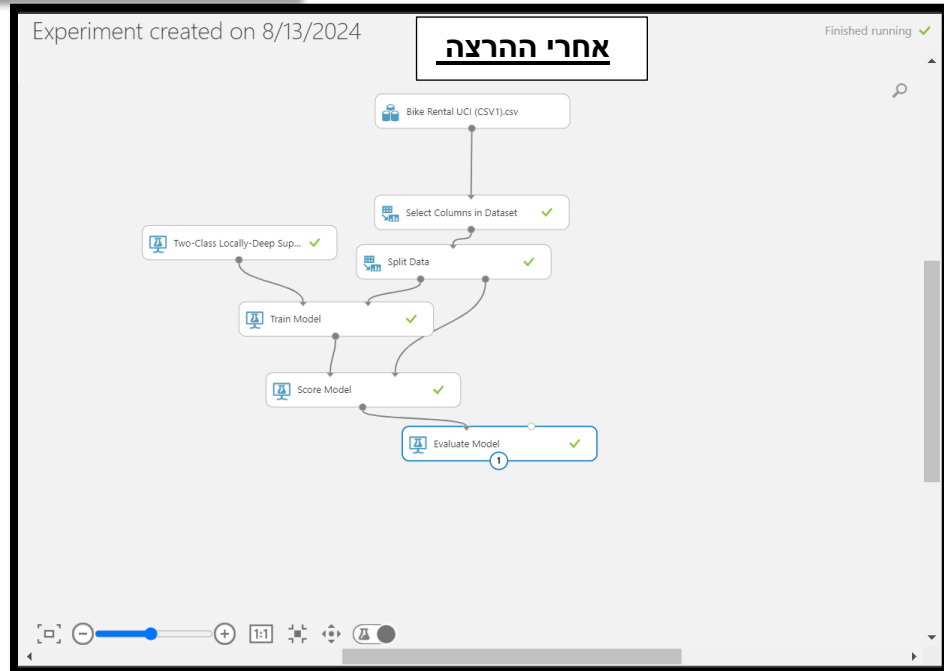
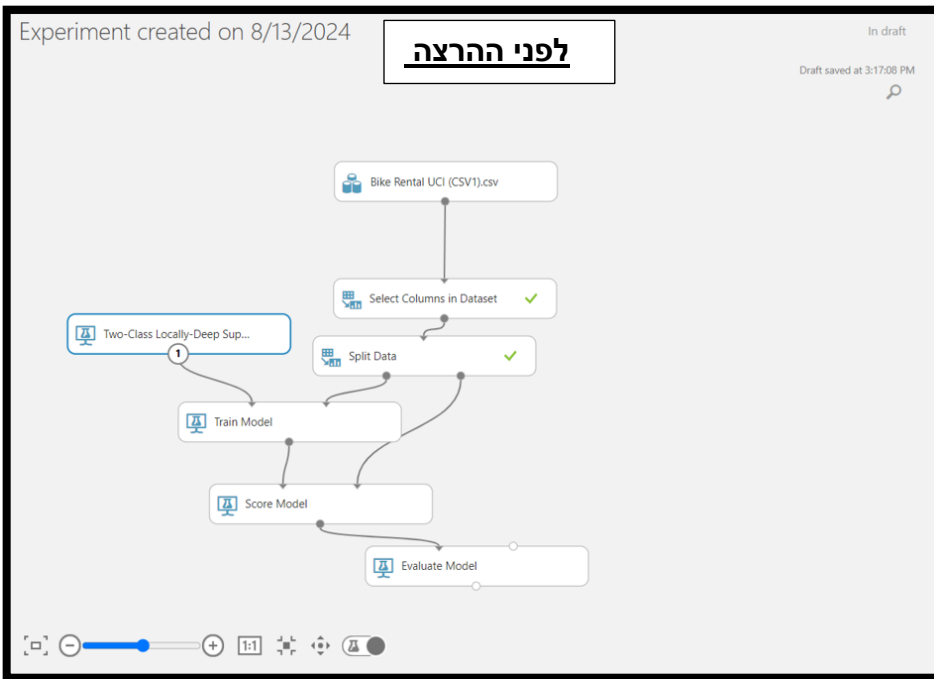
Finished running ✓

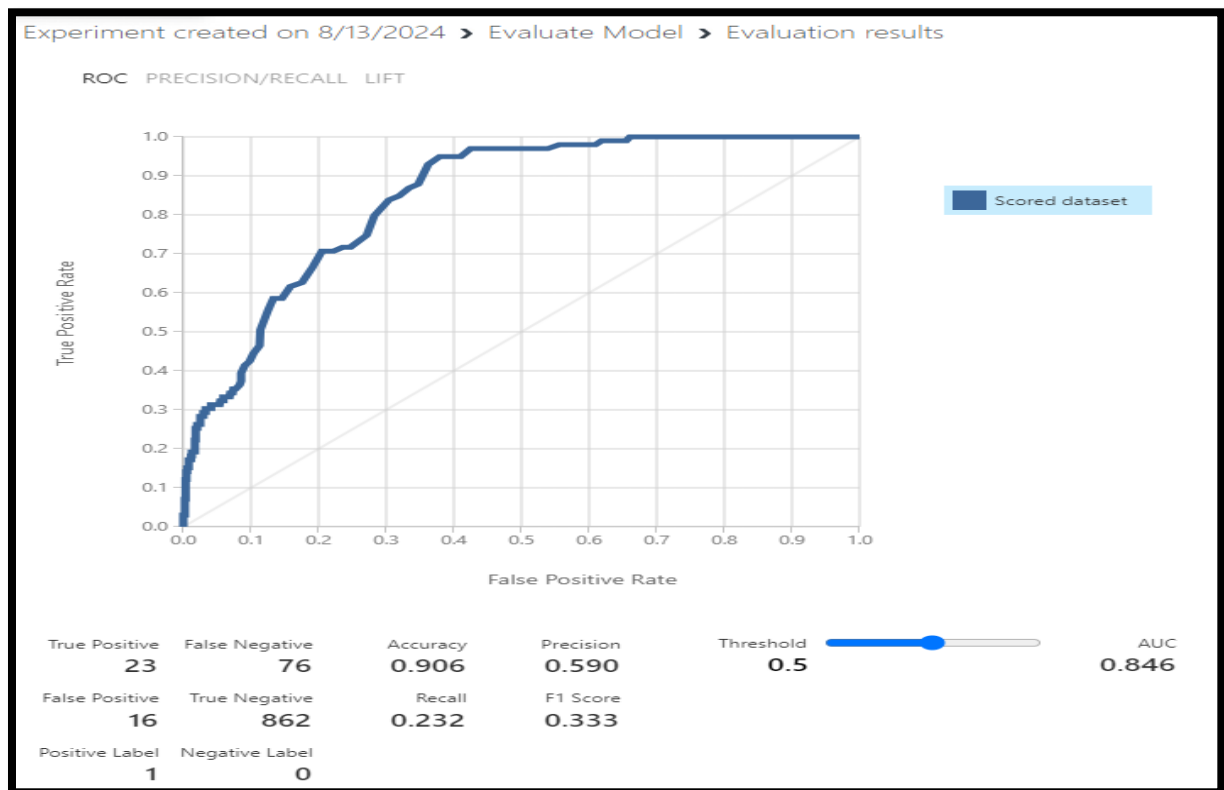


מסקנות:

המודל מציד דיוק סביר ($MAE=0.127$) המעיד על מרחק ממוצע נמוך בין החיזויים לערכים האמיתיים.
 (RMSE=0.181) נותן משקל גבוה לטעויות מה שמעיד על קיום טעויות משמעותיות.
 (Determination=0.639) מעיד שהמודל מסביר כ-64% מהשונות בנתונים, מה שמצביע על התאמה בינונית של המודל.

Two-Class Locally-Deep Support Vector Machine





פיענוח תוצאות החיזוי

AUC	F1 score	Precision	Recall	Accuracy	NL	TN	FN	PL	FP	TP	MODEL
0.846	0.333	0.590	0.232	0.906	0	862	76	1	16	23	Two Class Locally Deep Support Vector Machine

תוצאות המודל:

Two Class Locally Deep Support Vector Machine

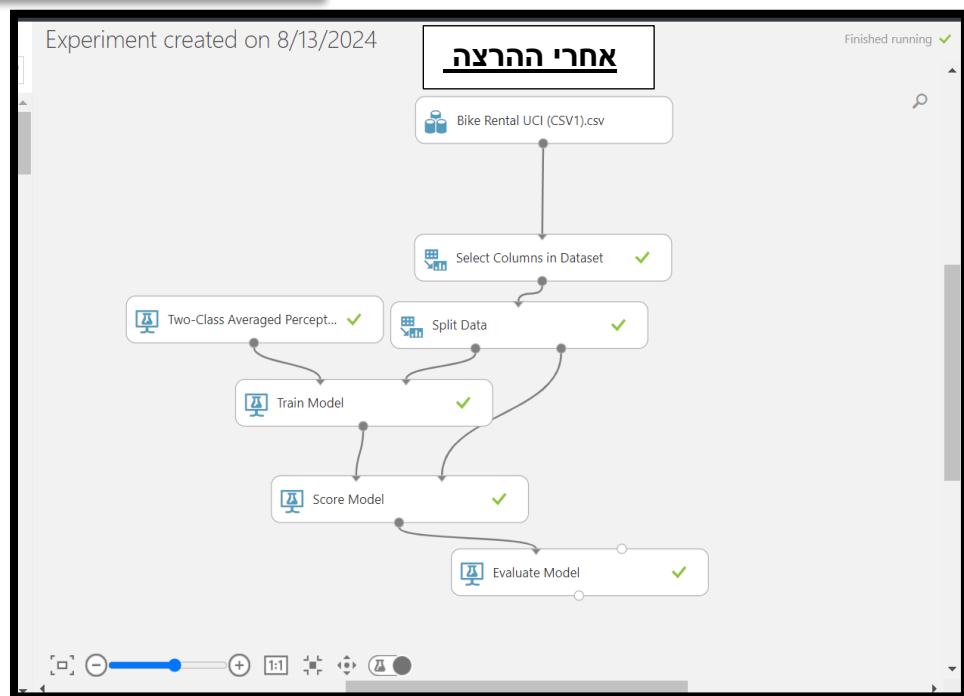
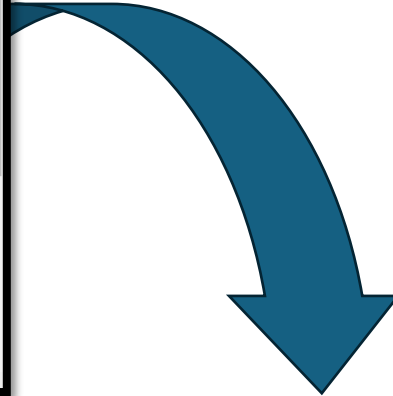
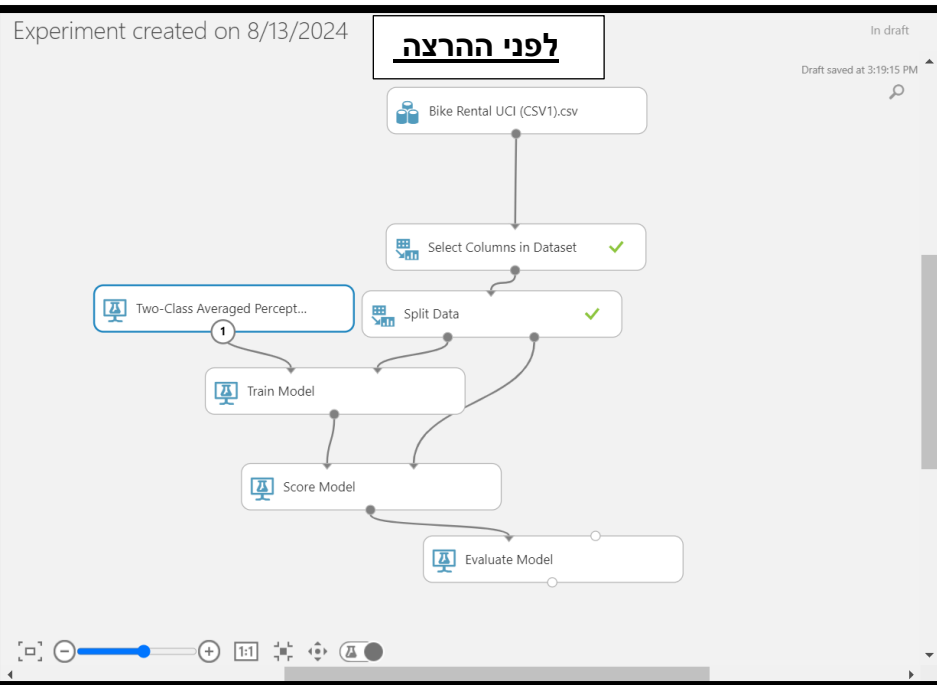
- AUC (0.846):**
 - ערך AUC גבוה יחסית, המצביע על יכולת טובה של המודל להבחין בין המעמדות.
- F1 Score (0.333):**
 - ציון F1 נמוך, מה שמעיד על איזון לא טוב בין Precision ל-Recall.
- Precision (0.590):**
 - דיוק טוב יחסית (59%), כלומר כאשר המודל מנבא מקרה חיובי, הוא נוטה להיות מדויק ב-59% מהמקרים.
- Recall (0.232):**
 - Recall נמוך יחסית (23.2%), המודל מצליח לזהות כמות נמוכה של מקרים חיוביים.
- Accuracy (0.906):**
 - Accuracy גבוה (90.6%), אך זה יכול לנבוע מזיהוי נכון של כמות גדולה של מקרים שליליים.

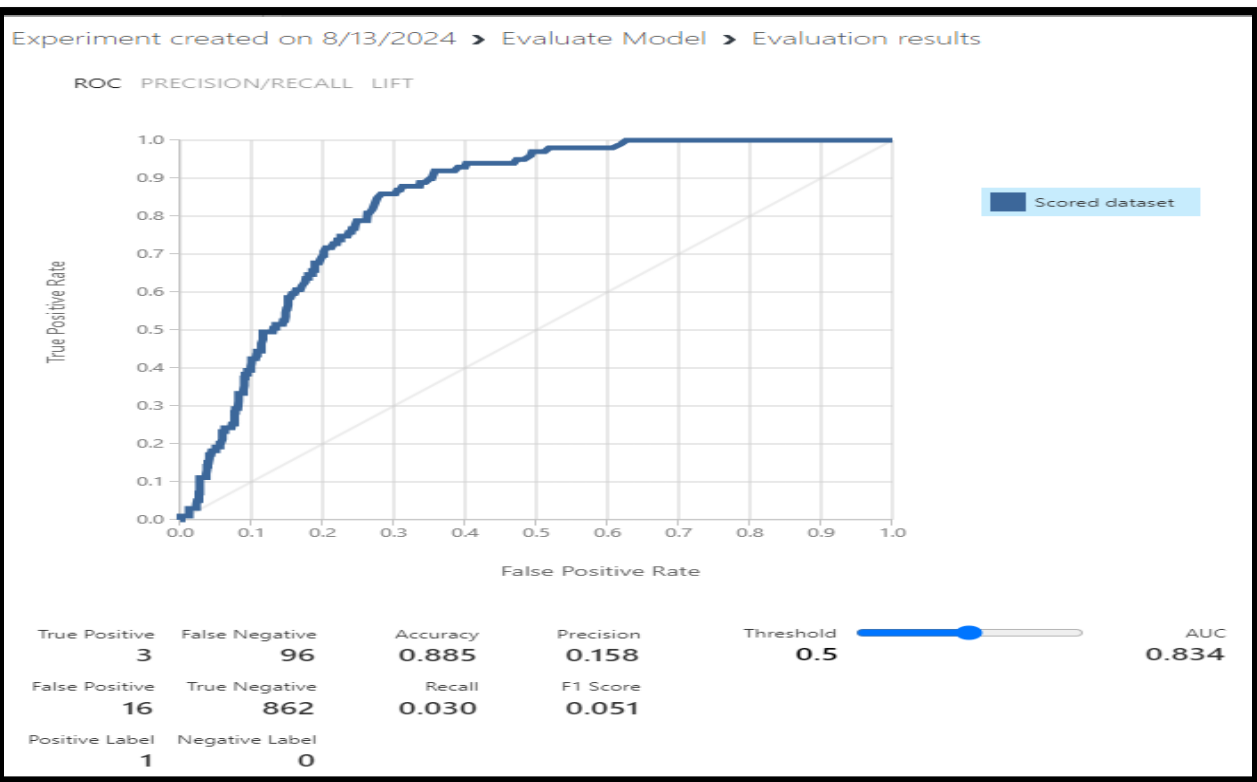
• מדדי: TN, FN, FP, TP

- **True Negatives (TN = 862):** המודל זיהה נכון 862 מקרים שליליים.
- **False Negatives (FN = 76):** 76 מקרים חיוביים שלא זוהו נכון.
- **False Positives (FP = 16):** 16 תחזיות חיוביות שגויות.
- **True Positives (TP = 23):** המודל זיהה נכון 23 מקרים חיוביים.

כלי חיזוי שני

Two Class Averaged Perceptron





פיענוח תוצאות החיזוי

AUC	F1 score	Precision	Recall	Accuracy	NL	TN	FN	PL	FP	TP	MODEL
0.834	0.051	0.158	0.030	0.885	0	862	96	1	16	3	Two Class Averaged Perceptron

תוצאות המודל:

Two Class Averaged Perceptron

- AUC (0.834):** ערך AUC סביר, מה שמעיד על כך שהמודל יכול להבחין בצורה בינונית בין המעמדות.
- F1 Score (0.051):** ציון F1 נמוך מאוד, מה שמעיד על איזון גרוע בין Precision ל-Recall.
- Precision (0.158):** דיוק נמוך מאוד (15.8%), כלומר המודל לא מדויק במיוחד בזיהוי מקרים חיוביים.
- Recall (0.030):** Recall נמוך מאוד (3%), כלומר המודל כמעט ואינו מזהה מקרים חיוביים.
- Accuracy (0.885):** Accuracy סביר אך הוא נובע כנראה מרוב המקרים השליליים שזוהו נכון.
- מדדי: TN, FN, FP, TP:**
 - True Negatives (TN = 862):** 862 מקרים שליליים.
 - False Negatives (FN = 96):** 96 מקרים חיוביים שלא זוהו.
 - False Positives (FP = 16):** 16 תחזיות חיוביות שגויות.
 - True Positives (TP = 3):** 3 מקרים חיוביים נכון.

Two-Class Locally-Deep Support Vector Machine

מסקנות:

מהתוצאות של המודל של **Two-Class Locally Deep Support Vector Machine**, ניתן להסיק כמה תובנות חשובות:

1. **AUC (0.846):**

ערך ה-AUC גבוה יחסית, מה שמעיד על כך שהמודל מצליח להבדיל בצורה טובה בין המעמדות החיוביים לשליליים. ככל שה-AUC גבוה יותר, כך למודל יש יכולת טובה יותר להבחין בין מקרים חיוביים לשליליים. זהו אינדיקטור לכך שהמודל יכול להיות יעיל במקרים שבהם חשוב להבחין בין המעמדות.

2. **F1 Score (0.333):**

הציון F1 נמוך (0.333) מעיד על חוסר איזון בין ה-Precision ל-Recall. ה-F1 משמש אינדיקטור לאיזון בין דיוק (Precision) למידת ההחזר (Recall), ולכן התוצאה הנמוכה מצביעה על כך שהמודל מתקשה לשמור על איזון טוב בין שני המדדים.

3. **Precision (0.590):**

ערך ה-Precision מראה כי כאשר המודל מנבא מקרה חיובי, הוא מדויק ב-59% מהמקרים. זה אומר שכאשר המודל מזהה מקרה חיובי, יש סיכוי סביר שהוא צודק, אבל עדיין קיימת טעות ב-41% מהמקרים. זה מעיד על בעיות בזיהוי נכון של מקרים חיוביים.

4. **Recall (0.232):**

ה-Recall נמוך יחסית (23.2%) מעיד על כך שהמודל לא מזהה את רוב המקרים החיוביים. זהו חסרון משמעותי במיוחד במקרים שבהם חשוב לזהות כמה שיותר מקרים חיוביים, כמו בזיהוי מחלות או תקלות במערכות קריטיות.

5. **Accuracy (0.906):**

ה-Accuracy הגבוה (90.6%) יכול להטעות, מאחר והוא מושפע בעיקר מהזיהוי הנכון של מקרים שליליים. כשהקבוצה השלילית הרבה יותר גדולה מהקבוצה החיובית, Accuracy גבוה לא בהכרח מעיד על מודל טוב, מאחר והמודל עלול לפספס מקרים חיוביים רבים.

6. **מדדי TN, FN, FP, TP:**

- **TN (862):** המודל זיהה נכון את רוב המקרים השליליים, ולכן Accuracy גבוה.
- **FN (76):** יש 76 מקרים חיוביים שהמודל לא הצליח לזהות, מה שמוביל ל-Recall נמוך.
- **FP (16):** כמות השגיאות החיוביות היא קטנה יחסית, מה שמוביל ל-Precision טוב.
- **TP (23):** המודל הצליח לזהות רק 23 מקרים חיוביים מתוך כלל המקרים החיוביים, שוב מדגיש את הבעיה ב-Recall.

מסקנות כלליות:

המודל מצליח להבחין היטב בין המעמדות (AUC גבוה), אך חסר איזון בין היכולת לזהות מקרים חיוביים לבין הדיוק בזיהוי. המודל מתמקד בעיקר בזיהוי מקרים שליליים, מה שמוביל ל-Accuracy גבוה אך Recall נמוך. אם המטרה היא לזהות מקרים חיוביים בצורה טובה, המודל זקוק לשיפור, ייתכן בעזרת איזון מחדש של המעמדות, שימוש באסטרטגיות נוספות כמו oversampling של הקבוצה החיובית או שינוי במטרות האימון של המודל.

Two Class Averaged Perceptron

מסקנות:

התוצאות של מודל **Two-Class Averaged Perceptron** חושפות מספר בעיות משמעותיות בתפקודו:

1. **AUC (0.834):**

ערך ה-AUC סביר, המצביע על כך שלמודל יש יכולת בינונית להבחין בין המעמדות החיוביים לשליליים. למרות זאת, ערך זה אינו מציין את איכות ההמלצות במדויק במקרה של מיעוט מקרים חיוביים, כפי שנראה בהמשך.

2. **F1 Score (0.051):**

ציון ה-F1 נמוך מאוד מעיד על בעיה רצינית באיזון בין ה-Precision ל-Recall. ה-F1 Score מחושב כהרמוניה של שני המדדים, מה שמעיד שהמודל לא מצליח לאזן בין זיהוי נכון של מקרים חיוביים לבין הדיוק בזיהוי זה.

3. **Precision (0.158):**

דיוק המודל נמוך מאוד (15.8%), מה שאומר שכאשר המודל מנבא מקרה חיובי, רק 15.8% מהמקרים הם אכן נכונים. זה מעיד על כך שהמודל נוטה לנבא מקרים חיוביים שגויים.

4. **Recall (0.030):**

ערך ה-Recall נמוך ביותר (3%), כלומר המודל כמעט ואינו מזהה מקרים חיוביים. זהו חסרון קריטי כאשר המטרה היא לזהות מקרים חיוביים כמו בזיהוי של מחלות או תקלות קריטיות. המודל פשוט מחמיץ כמעט את כל המקרים החיוביים.

5. **Accuracy (0.885):**

ה-Accuracy סביר, אך הוא נובע מכך שהמודל מצליח לזהות נכון את רוב המקרים השליליים. במקרה שבו יש חוסר איזון בין מספר המקרים החיוביים למקרים השליליים, Accuracy יכול להטעות ולהציג את המודל כטוב יותר ממה שהוא בפועל, כפי שנראה כאן.

6. **מדדי TN, FN, FP, TP:**

- **TN (862):** המודל מצליח לזהות נכון 862 מקרים שליליים, וזה מסביר את ה-Accuracy הסביר.
- **FN (96):** המודל לא זיהה 96 מקרים חיוביים, מה שמצביע על בעיית Recall חמורה.
- **FP (16):** המודל עשה מעט שגיאות בזיהוי מקרים חיוביים (FP), אך זה לא מתקן את העובדה שהוא כמעט ולא מזהה מקרים חיוביים נכונים.
- **TP (3):** המודל זיהה רק 3 מקרים חיוביים נכונים, מה שמסביר את ה-F1 וה-Recall הנמוכים.

מסקנות כלליות:

המודל **Two-Class Averaged Perceptron** סובל מבעיות חמורות בזיהוי מקרים חיוביים. למרות שיש לו Accuracy סביר, הוא אינו מתאים למשימות שבהן חשוב לזהות מקרים חיוביים, במיוחד כאשר ה-Recall וה-F1 נמוכים בצורה קריטית. במקרה של בעיות מאוזנות יותר או אם המטרה העיקרית היא לזהות מקרים שליליים, המודל עשוי להיות שימושי. עם זאת, למשימות הדורשות זיהוי טוב של מקרים חיוביים (כמו גילוי הונאה, אבחון רפואי, וכו'), יש צורך בשיפור משמעותי, למשל באמצעות אימון נוסף, טכניקות לאיזון הקטגוריות, או בחירה במודל אחר שיכול להפיק ביצועים טובים יותר.