

פרויקט גמר
מבוא ללמידת מכונה
2025 שנה "ל התשפ"ה

שלב ראשון – אקספלורציה

בתחילת העבודה על הפרויקט קראתי את מסמך ההיכרות עם הדאטה ונתתי השערות משלי על איזה פיצ'רים יהיו משמעותיים יותר ואיזה פחות. טענתי את הדאטה למחברת והתחלתי לבדוק התפלגויות של העמודות ואת היחס שלהן עם עמודת Label בעזרת היסטוגרמות וcountplots (דוגמה בנספח 1.1).

תחילה מצאתי כמה פיצ'רים קטגוריאליים שעבור אחד או יותר מהקטגוריות שלהן, יש יותר בקשות הלוואה שאושרו מאשר נדחו. אפשר לראות את הפערים בנספח 1.2.

בנוסף, אפשר לראות שקיים קשר בין הפיצ'רים שמתייחסים לזהות מגיש הבקשה ולזהות מגיש הבקשה הנוסף. אפשר לראות שרוב האנשים לא מגישים בקשה להלוואה עם אדם נוסף אבל אלו שכן מבקשים עם אדם נוסף, לרוב יהיו בעלי אותה עדה וגזע ומינים הפוכים (טבלאות הקורלציה נמצאות בנספח 1.3). אני אשתמש במידע הזה בהמשך כדי למלא את הערכים החסרים בפיצ'רים האלה.

בפיצ'רים המספריים אפשר לראות קורלציה גבוהה יחסית בין כמה פיצ'רים: `population & C`, `applicant_income` & `loan_amount` ועוד. (נספח 1.4) אני אשתמש במידע הזה בהמשך כדי למלא את הערכים החסרים בפיצ'רים האלה.

בבדיקה שעשיתי לפיצ'ר ID התגלה שקיימים הרבה יותר בקשות הלוואה עם מספר ID שגדול מ-400,000. בדקתי את ההתפלגות של label ביחס ל ID וגיליתי שרוב ההלוואות שלא אושרו קיבלו ID שגדול מ-400,000 ורוב ההלוואות שאושרו קיבלו ID קטן מ-400,000. באופן אינטואיטיבי חשבתי בהתחלה להוריד את הפיצ'ר הזה כשנצטרך למדל את הדאטה אבל הקשר של ID ללייבל כל כך מובהק שהחלטתי להשאיר אותו למידול. בהמשך העבודה, בהליך הורדת המימדים, הוחלט שפיצ'ר ID הוא הפיצ'ר החשוב ביותר לניבוי המודל. (נספח 1.5)

בפיצ'רים `applicant_income` ו-`loan_amount` יש אוטליירים רבים שנצטרך להתמודד איתם בעיבוד המקדים (נספח 1.6). בנוסף בהתפלגות של label ביחס ל-`applicant_income` אפשר לראות את גובה השכר שממנו יותר מאשרים הלוואות מאשר דוחים אותן (נספח 1.7).

בניתי מיפוי של שני הפיצ'רים האלו וצבעתי כל נקודת דאטה בהתאם לlabel שלה (נספח 1.8). אפשר לראות שקיים יחס מסוים בין גודל ההלוואה להכנסת המלווה שלפיו לרוב מאשרים או דוחים את ההלוואה. יצרתי גרף עמודות שמציג את כמה אחוזים ההכנסה של מגיש הבקשה מתוך גודל ההלוואה שביקש (נספח 1.9). נוצרה התפלגות שנראית כמעט נורמלית באזור 25%. נראה שרוב האנשים מבקשים הלוואה בגודל של בערך פי 4 מגובה השכר שלהם.

בנספחים 1.10 ו-1.11 אפשר לראות את ההיסטוגרמות של שני הפיצ'רים `Minority_population` ו-`hud_median_family_income` ביחס לערך label שלהם. בשניהם אפשר לראות את השלב שבו אחוז מיעוטים או גובה הכנסה מתחיל להשפיע על מספר ההלוואות שמקבלות אישור.

עבור הפיצ'רים האנונימיים, בדקתי את ההתפלגות שלהם וראיתי שעבור פיצ'ר A ופיצ'ר B מדובר בהתפלגות נורמלית (נספח 1.12). עבור פיצ'ר C נדובר שראינו קורלציה גבוהה בינו לבין פיצ'רים אחרים כמו `number_of_owner_occupied_units`. אפשר לראות את הקורלציה הזו בנספח 1.13. אני אשתמש בהתפלגויות ובקורלציה הזו כדי למלא את הערכים החסרים בעמודות האלו.

שלב שני – עיבוד מקדים

1. השלמת הנתונים החסרים:

בדקתי באופן מרוכז כמה אחוזים מכל עמודה חסר (נספח 2.1). לכל עמודה בחרתי בשיטה שונה למלא את הנתונים החסרים.

- א. העמודות של race_name_2/3 החלטתי שלא משתלם להתעסק איתם בגלל שחסר בהם יותר מידי מידע ולכן מחקתי אותם.
- בתכנון המקורי תכננתי למחוק גם את עמודת ID אבל בעקבות המסקנות מהאקספלורציה החלטתי להשאיר אותה. ההחלטה הזו תתגלה כמשמעותית בתוצאות הניבוי של המודלים.
- ב. עמודות Bi A מציגות שתי התפלגויות שנראות נורמליות. לא מצאתי שהן מראות קשר חזק לאף פיצ'ר אחר. לכן החלטתי למלא את הנתונים החסרים בעזרת הממוצע והחציון.
- ג. בעמודה D מילאתי את הערכים החסרים עם הערך הנפוץ ביותר.
- ד. עבור כמה עמודות, היה נראה לי שאמור להיות קשר כמעט חד חד ערכי מערך אחד לערך שני לדוגמה: census tract number & minority, census tract number & population, county & county code, population, msamd, msamd & tract_to_income ועוד. עבור כל האפשרויות האלה בדקתי עד כמה באמת אפשר להכריע בעזרת עמודה אחת על עמודה אחרת. בדקתי כמה ערכים קיימים בעמודה 2 עבור אותו ערך בעמודה 1. קיבלתי את התוצאות שאפשר לראות בנספח 2.2. אפשר לראות שמלבד הקשר החד חד ערכי בין county code-י county code, הקשרים בין העמודות הם לא חד חד ערכיות. יחד אם זאת, אפשר לראות שברובם קיים יחס חד חד ערכי עבור יותר מ-75% מהערכים. לכן החלטתי בשביל הערכת הנתונים החסרים שאניח הנחה נאיבית עבור העמודות עם מעל 75% ערכים חד חד ערכיים שקיים קשר חד חד ערכי ב-100% מעמודה אחת לעמודה השנייה ואשלים את הנתונים החסרים אחד בעזרת השני. העמודות שמילאתי כך ויחס המילוי ביניהם מופיע גם בנספח 2.2. לכן בסוף התהליך מילאתי את האיברים הריקים בעמודות: county, county_code, census_tract_number, population, minority_population ו-msamd.
- ה. עבור עמודה loan_amount, חישבתי את החציון של העמודה בחלוקה לשלושת הקטגוריות לסוג ההלוואה (loan_purpose). מילאתי את הערכים החסרים עם החציון לפי הערך שהיה בעמודת loan_purpose באותה שורה. את השורות בהן לא היה גם ערך בloan_purpose מילאתי בחציון הכללי של loan_amount.
- ו. עבור כל העמודות שאין להן קשר חד חד ערכי עם עמודה אחרת כתבתי פונקציה שמחפשת את הערך הנפוץ ביותר בהינתן ערך מעמודה דוגמת. הפונקציה מקבלת רשימה של עמודות אותן היא צריכה למלא ועמודה נוספת שתשמש כאינדקס לבחירת הערך המשלים. בנוסף כדי להשתמש בפונקציה בהכנת הנתונים בסט test, יש אפשרות בפונקציה להכניס target_df ו-train_df. כך אפשר להשלים את הנתונים בסט test לפי הנתונים של סט train. בעזרת הפונקציה הזו השלמתי את הנתונים החסרים בעמודות: loan_type, loan_purpose (according to loan_amount), agency, property_type, hud_median_family_income, tract_to_msamd_income, number_of_owner_occupied_units (according to census_tract_number), and C (according to number_of_owner_occupied_units). הפונקציה נקראת fill_null_with_mode.
- ז. עמודות שיש בהן אפשרות של Not applicable או information not provided, מילאתי את הערכים החסרים בהם עם האפשרות הזו. ייתכן שיש להם משמעות עבור ערכים חסרים ולכן יש קטגוריה לכך. לדוגמה, יכול להיות שאדם שבוחר שלא להצהיר על מוצא או מין יקבל תוצאה שונה מאדם שכן.
- ח. בעמודות שמתייחסות ל-Co applicant מילאתי את הערכים החסרים בצורה הבאה- אם באחד העמודות מופיע הערך "No co-applicant" אז כל הערכים שקשורים ל-Co applicant גם יקבלו את הערך "No co-applicant".
- אם באחד העמודות מופיע ערך אחר, אז הערכים של applicant_race_name_1 ו-applicant_ethnicity יהיו הערכים שיופיעו במקומות החסרים בעמודות המקבילות שמתייחסות ל-Co applicant. הערך שייכנס למקומות החסרים ב-co_applicant_sex יהיה הערך ההפוך מהערך שמופיע ב-applicant_sex.

- ט. בעמודה של `applicant_income` יצרתי אומד של ההכנסה שמעריך את ההכנסה של `applicant` בעזרת הממוצע בין חציון ההכנסה לפי מין מגיש הבקשה לבין חציון ההכנסה לפי מוצא מגיש הבקשה. כל הערכים החסרים בעמודה מקבלים את הערך לפי הנוסחה הזו.
2. איחוד ערכים דומים בעמודות הקטגוריות:
- בעמודות של `applicant` ו-`co applicant` איחדתי את הנתונים שכתוב בהם "Information not provided..." ואת הנתונים שכתוב בהם "Not applicable" לכדי נתון אחיד: "Not provided".
3. הסרת Outliers:
- בדאטה קיימים נתונים חריגים. הנתונים האלה לא מייצגים את רוב הדאטה ולכן לא נרצה שהמודלים ילמדו עליהם. היו נתונים חריגים בכמות קיצונים בעמודות: `loan_amont`, `applicant_income` ו-`tract_to_msamd_income`. אפשר לראות בנספח 2.3 את המצב לפני ואחרי הורדת הנתונים החריגים.
4. יצירת פיצ'רים חדשים:
- במקום להשתמש בנתונים של העמודות `MSAMD`, `county` ו-`Census_tract_number`, החלטתי ליצור מהם פיצ'רים חדשים. העמודות האלו מכילות המון נתונים ייחודיים וקטגוריים ולכן שימוש ישירות בהם במודל עלול ליצור מימדיות יתר. החלטתי להשתמש בעמודות האלה כדי ליצור עמודות שיציגו במקום את שם המחוז או מספר סידורי, את אחוזי ההצלחה של כל מחוז או מספר סידורי בעמודות האלה.
5. קידוד פיצ'רים קטגוריים:
- צריך לקודד את הפיצ'רים הקטגוריים לכדי מספרים כדי שיהיה אפשר לאמן את המודלים עליהם. אני בחרתי להפוך את הפיצ'רים למספרים בשני דרכים בהתאם לכמות הקטגוריות בכל אחד מהם: אם יש בעמודה אחת יותר משלוש קטגוריות, נשתמש ב-`OrdinalEncoder`, אם יש בעמודה שלוש קטגוריות או פחות, נשתמש ב-`dummies` של `pandas`. הרעיון מאחורי זה הוא למנוע יצירה של יותר מידי עמודות חדשות ולנסות לשמור על כמות קטנה כמה שאפשר של פיצ'רים מקודדים חדשים. אפשר לראות בנספח 2.4 את כמות הקטגוריות בכל פיצ'ר ואת הקו המכריע לסוג הקידוד.
6. נרמול הדאטה:
- הדאטה מכיל הרבה סוגים של ערכים מספריים בשלב הזה. חלק ממנו מייצג משכורות של מגישי בקשות להלוואה, חלק ממנו מייצג לאיזה סוכנות הוגשה הבקשה. בנוסף האיברים המספריים נמצאים בטווחים שונים. אנחנו נרצה לנרמל את הדאטה בשביל לבנות מודלים כמו רגרסיה לינארית, KNN, רשת נוירונים ועוד. השתמשתי ב-`StandardScaler` כדי לנרמל את כל הדאטה. הנרמול מאפשר למודלים להתייחס לכל הפיצ'רים באותה סקאלה ובכך מונע מפיצ'ר אחד עם מספרים גדולים להטות את המודל כולו.
7. הקטנת כמות המימדים:
- מימדיות דאטה גדולה מידי גורמת לכל נתון בדאטה להיות ייחודי מאוד ומרוחק יותר משאר הנתונים בדאטה. מימדיות כזו יכולה להקשות על המודלים שלנו למצוא תבניות ולגרום להם ל-`Overfitting`. ההחלטה אם אני צריך לבצע הורדת מימדים או לא התקבלה אחרי סקירת המצב של המודלים עם ובלי הורדת מימדים.
- הגעתי למצב שיש לי 34 פיצ'רים בדאטה. בהתחלה ניסיתי להוריד מימדים עם PCA. עם PCA, אם אני משאיר 95% מהשונות של הדאטה, אני מקבל 27 פיצ'רים חדשים של PCA, כלומר מקטין רק 7 מימדים. כדי להכריע אם הורדת המימדיות הזו משתלמת עבור המודלים שלי אני בדקתי את `accuracy` של המודלים עם חלוקה ל-`train` & `validation` של הדאטה המקורי ושל הדאטה עם הגודל המוקטן. ברוב המודלים היו פערים גם בדיוק של `train` וגם בדיוק של `Validation` לטובת ההחלטה לשמור על כל הפיצ'רים. החלטתי שלפני שאני מכריע שכדאי לשמור על כל הפיצ'רים אני אנסה עוד שיטה אחת להקטין את כמות המימדים. החלטתי להשתמש במודל של `Random Forest` ולבחור את הפיצ'רים ששומרים על 95% מהחשיבות. בשיטה הזו נשארו 25 פיצ'רים שהם עדיין הפיצ'רים המקוריים של הדאטה. כשהרצתי את המודלים עם הקטנת המימדים הזו, הערכים של `accuracy` עלו עבור כל המודלים ביחס לגודל הדאטה המקורי. לכן בחרתי ב-25 הפיצ'רים האלו (נספח 2.5).

8. הכנת הדאטה בקובץ test:

בשביל להכין את הדאטה מקובץ test לניבוי בסוף התהליך, העברתי את הדאטה באותם שלבים 1-7 שמוצגים לעיל אבל בלי לקחת בחשבון את הנתונים שנמצאים ב-test. כל הנתונים בשלב הזה התבססו על מה שנלמד מהדאטה train בלבד.

שלב שלישי – בניית המודלים וטיוב ההיפרפרמטרים

המודלים שבחרתי הם רגרסיה לוגיסטית, KNN, Multi-Layer Perceptron ו-Random Forest.

כדי לכוון את ההיפרפרמטרים של כל מודל, השתמשתי בכלי אופטימיזציה שלא נלמד בקורס בשם Optuna. Optuna פריימוורק אופטימיזציה בייסיאנית שלומדת באופן דינמי אילו היפרפרמטרים מניבים תוצאות טובות יותר. היא לומדת לאורך תהליך הטיוב איזה פרמטרים משפיעים על המודל לחיוב ופועלת לבדוק וריאציות שלהם במקום של היפרפרמטרים שמשפיעים פחות. בזכות Optuna יכולתי לקבל דיוק גבוה יותר להיפרפרמטרים שלי וניצלתי את משאבי החישוב שלי בצורה טובה יותר. הגדרתי עבור כל מודל את ההיפרפרמטרים שאני רוצה Optuna יטיב והנחתי אותו שיבחר את הפרמטרים שמביאים את ה-Accuracy הגבוה ביותר על דאטה האימון. אלה הפרמטרים שנבחרו לכל מודל:

1. רגרסיה לוגיסטית

- a. penalty: רגורליזציה לפי לאסו או לפי רידג'. ההיפרפרמטר הזה מקטין את השונות ומגדיל את ההטייה של המודלים. בטיוב של Optuna, נבחר להשתמש ברגורליזציה רידג'.
- b. C: האילוץ של הרגורליזציה. ככל שנבחר אילוץ קטן יותר כך השונות תקטן וההטייה תגדל. בטיוב של Optuna בדקתי ערכים בין 0.0001 לבין 1 ונבחר להשתמש באילוץ קטן 0.000354.

2. KNN

- a. n_neighbors: כמות השכנים שהמודל יכריע בעזרתו את הסיווג. ככל שכמות השכנים קטנה יותר ככה הסיכוי שהמודל יעשה overfitting גדלה יותר ולהפך. כלומר ככל שהכמות קטנה יותר ככה השונות גדלה וההטייה קטנה, ככל שהכמות גדולה יותר השונות קטנה וההטייה גדלה. בטיוב של Optuna, בדקתי ערכים בין 5 ל-55. הכמות שנבחרה היא 40 שכנים.
- b. p: שיטת מדידת המרחק בין שכנים. השיטה לא משפיעה על השונות או על ההטייה ישירות אבל היא כן משפיעה על איך נבחרים השכנים שמכריעים את הסיווג של המודל. שיטת המדידה שנבחרה היא 1 (מרחק Manhattan).

3. MLP

- a. hidden_layer_sizes: כמות הנוירונים בשכבה. ככל שיש יותר נוירונים ככה המודל מסוגל לתפוס יותר דפוסים בדאטה והשונות גדלה וההטייה קטנה. בבדיקה שלי הכנסתי בדיקות על כמות של 20 עד 100 בקפיצות של 20. המבנה שנבחר הוא 100.
- b. alpha: ערך שמשפיע על הרגורליזציה. ככל שהערך באלפא גדול יותר ככה הרגורליזציה חזקה יותר מה שמגדיל את ההטייה ומקטין את השונות. ככל שהערך קטן יותר הרגורליזציה נחלשת מה שמגדיל את השונות ומקטין את ההטייה. מבין ערכים בין 0.00001 לבין 0.1 הערך שנבחר הוא 0.000864.
- c. learning_rate_init: ערך הלמידה ההתחלתי של פונקציית האופטימיזציה. ההיפרפרמטר הזה לא משפיע ישירות על השונות וההטייה אבל משפיע על כמות הזמן שלוקח למודל להתכנס. אם הערך קטן מידי, עולה הסיכוי להיתקע במינימום מקומי זה עלול לגרום להטיה גבוהה יותר. אם הערך גדול מידי, אנחנו עלולים לפספס את המינימום לגמרי מה שיכול להכשיל את כל המודל. מבין ערכים בין 0.00001 לבין 0.01 הערך שנבחר הוא 0.00348.
- d. activation: פונקציית האקטיבציה של המודל. בדקתי את הפונקציות relu ו-tanh. הפונקציה שנבחרה היא tanh. ייתכן שהפונקציה הזו נבחרה בגלל שהערכים בפיצ'רים מנורמלים עם תוחלת 0 ופונקציית האקטיבציה הזו שומרת על הסימן מהערך בקלט במעבר משכבה לשכבה ברשת.

4. Random Forest:

- a. `n_estimators`: כמות העצים ביער. ככל שיש יותר עצים ביער ככה השונות קטנה אבל ההטיה לא גדלה. ההטיה לא גדלה בגלל שכל אחד מהעצים מביא שונות גדולה והטיה קטנה והממוצע של ההכרעות שלהם רק יקטין את השונות בלי להגדיל את ההטיה שמושפעת ממבנה העץ. בדקתי כמות עצים בין 100 לבין 300 והכמות שנבחרה היא 291 עצים.
- b. `max_depth`: עומק מקסימלי של כל עץ. ככל שיש לעץ יותר פיצולים ככה הוא יותר מורכב ובעל שונות גדולה והטיה קטנה. אם נגביל את העץ נוכל לשלוט על השונות ובכך למנוע `overfitting`. בדקתי ערכים בין 10 לבין 60 והכמות שנבחרה היא 31.
- c. `max_features`: כמות הפיצולים שמשקלים עבור כל פיצול בעצים. ההיפרפרמטר הזה מייצר אקראיות בבחירת הפיצולים בפיצולים של העצים. ההיפרפרמטר הזה עוזר לייצר מגוון של עצים שמסתמכים על פיצולים שונים וכך אנחנו נמנעים מהסתמכות מוגזמת על פיצור או פיצורים יחידים. בדקתי את האפשרויות `log2`, `sqrt` ו-0.5. הערך שנבחר הוא 0.5 (כל פעם נבחרים חצי מכמות הפיצורים ונבדקים ביניהם)

לא בדקתי היפרפרמטרים נוספים למודל הזה בגלל כמות הזמן הארוכה במיוחד שלקח למחשב לנסות לטייב. אני מודע לכך שיתכן וטיוב נוסף של היפרפרמטרים כמו `min_samples_leaf` ו-`min_samples_split` היו יכולים לדייק את המודל עוד יותר אבל החלטתי להשתמש בערכי ברירת המחדל שלהם כדי לקצר בזמן הטיוב.

כל הפרמטרים המלאים של המודלים נמצאים בנספח 3.1.

בנספח 3.2 מופיעים גרפים שמציגים את תהליך טיוב ההיפרפרמטרים בכל מודל ואת החשיבות של כל היפרפרמטר למקסום הדיוק של המודל. אפשר לראות שברגרסיה לינארית החשיבות הייתה בעיקר על הערך של C ופחות על הרגולריזציה. ב-KNN ההיפרפרמטר החשוב יותר היה כמות השכנים. ב-MLP ההיפרפרמטר החשוב היה ערך הלמידה ההתחלתי ואחר כך הערך של אלפא וההיפרפרמטרים הנוספים.

שלב רביעי – הערכת המודלים

אחרי בניית המודלים ביצעתי הערכה למודלים בעזרת `confusion matrix` ובעזרת K-Fold Cross Validation.

במטריצת בלבול (`confusion matrix`) מוצגים לנו הסיווגים שהמודל ביצע לעומת הסיווגים של הנתונים בפועל. חילקתי את הנתונים ל-Train ו-Validation כך ש-80% מהנתונים אימנו את המודלים ו-20% בדקו את המודלים. מטריצה הבלבול הציגה את הסיווגים של כל מודל על הנתונים ב-Validation. המטריצה של מודל Random Forest מופיעה בנספח 4.1.

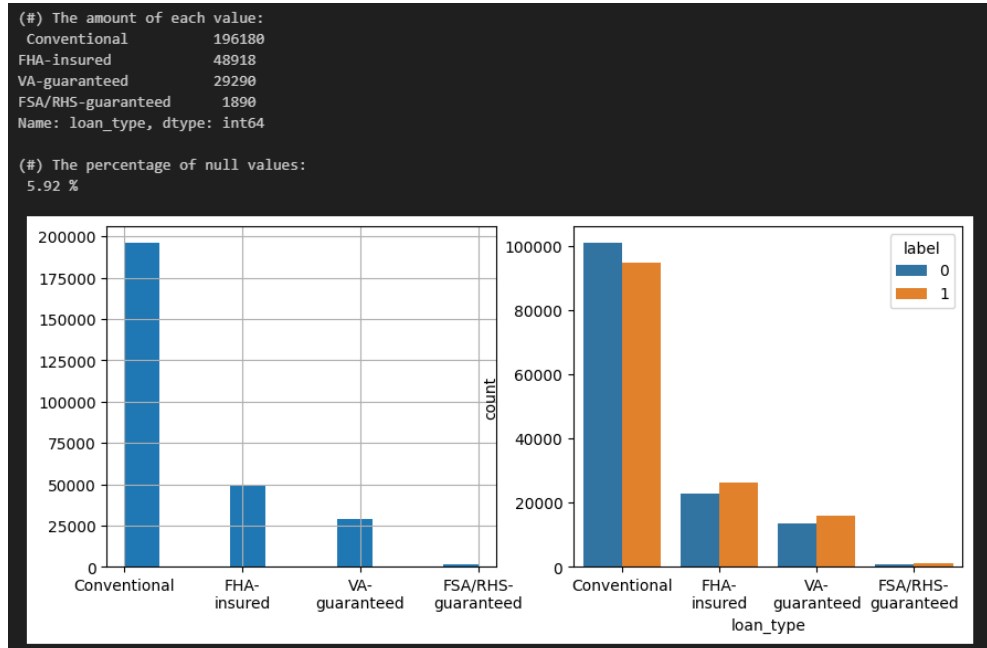
בעזרת המטריצה מצאתי שיש למודל `accuracy` 0.898, `Recall` 0.902 ו-`Precision` 0.892. אפשר ללמוד מכך שהמודל מדויק ב-89.8% מהפרדיקציות שלו, וטרייד-אוף מאוזן בין ה-`recall` וה-`Precision`.

מצורף בנספח 4.2 תרשים של עקומות ROC עבור כל Fold שביצעתי במסגרת K-Fold Validation.

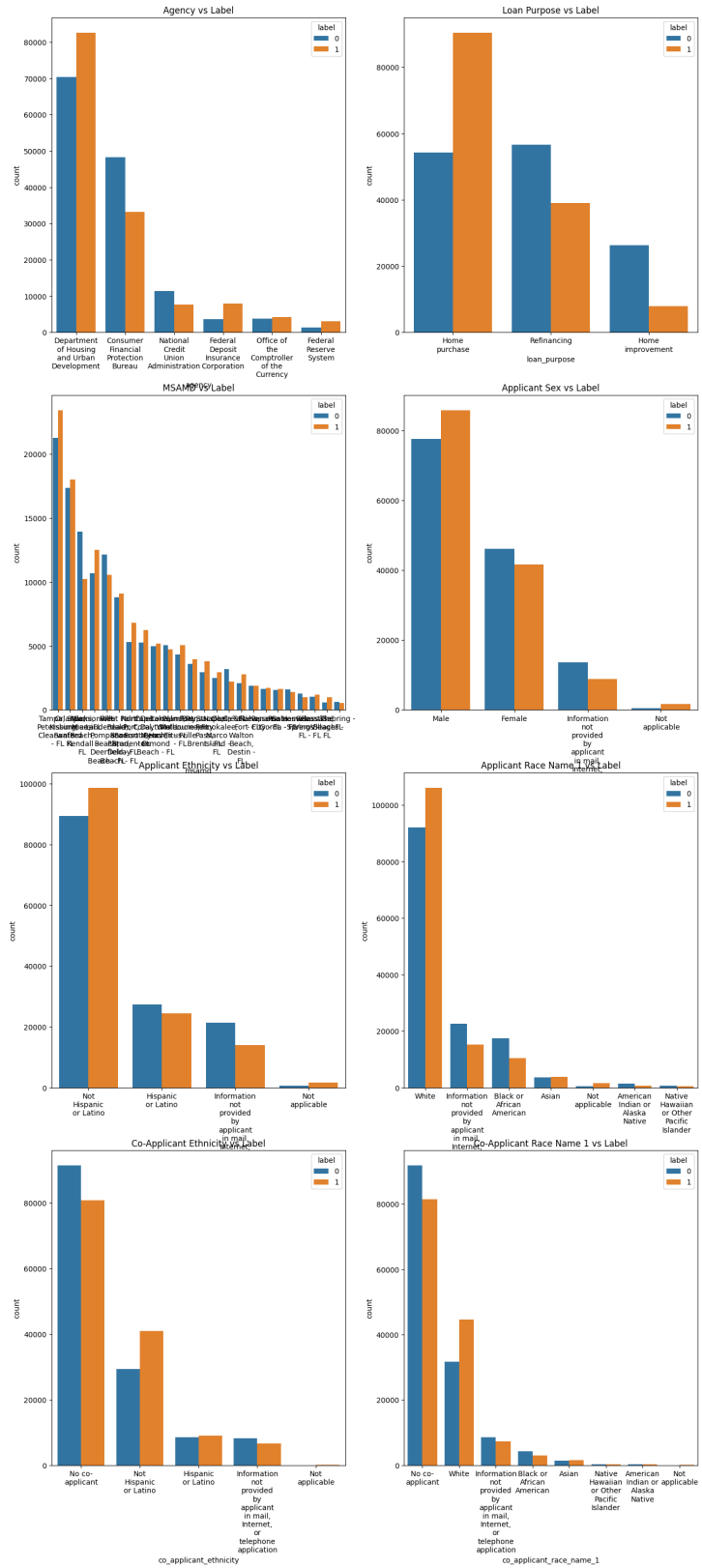
בנוסף, בזמן חישוב עקומות ROC חישבתי גם עבור כל Fold את `AUC` של `train` ובדקתי את הפערים בין `AUC` הממוצע של `train` לעומת `AUC` הממוצע של `validation`. אפשר לראות את הנתונים בנספח 4.3. אפשר להסיק אם המודלים הם `overfitted` בעזרת הפער ב-`AUC` בין `train` ל-`validation`. ככל שהפער יותר קטן ככה המודל מכיל בצורה טובה יותר. במודל של MLP ו-Logistic Regression נראה שיכולת ההכללה היא הטובה ביותר. ב-KNN נראה שקיים מעט `overfitting` אבל עדיין הפער קטן מאוד בין `train` ל-`validation`. כדי להתמודד עם `Overfitting` צריך להתאים את ההיפרפרמטרים של המודל כך שלא ישנן את הנתונים של `train` אלא יזהה את הדפוסים. במילים אחרות צריך להגדיל את ההטיה על חשבון השונות הגדולה שהמודל מכיל.

נספחים:

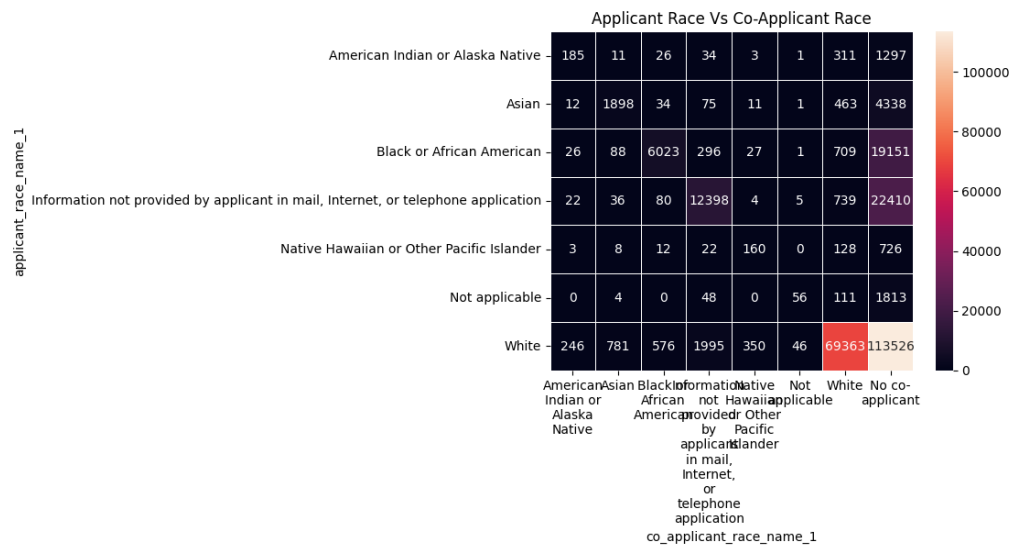
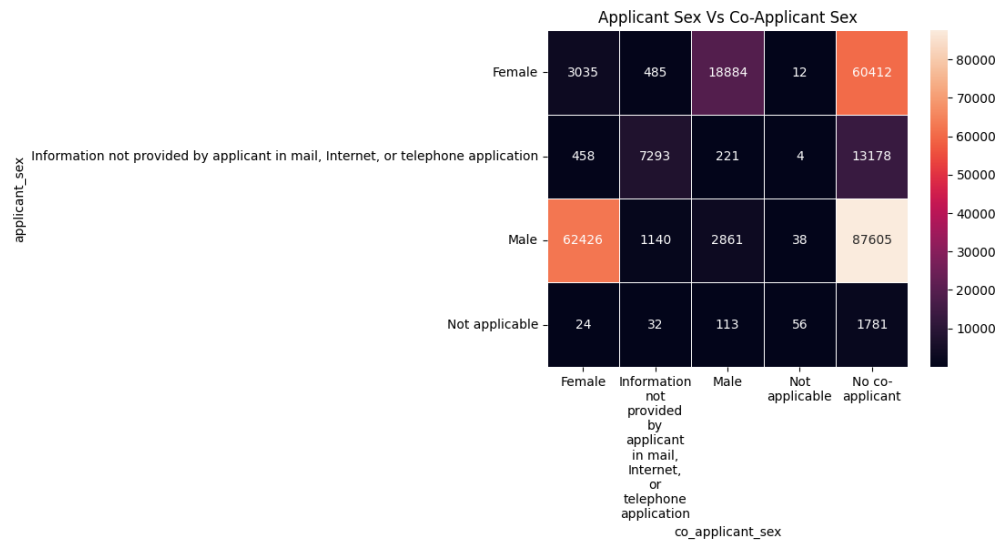
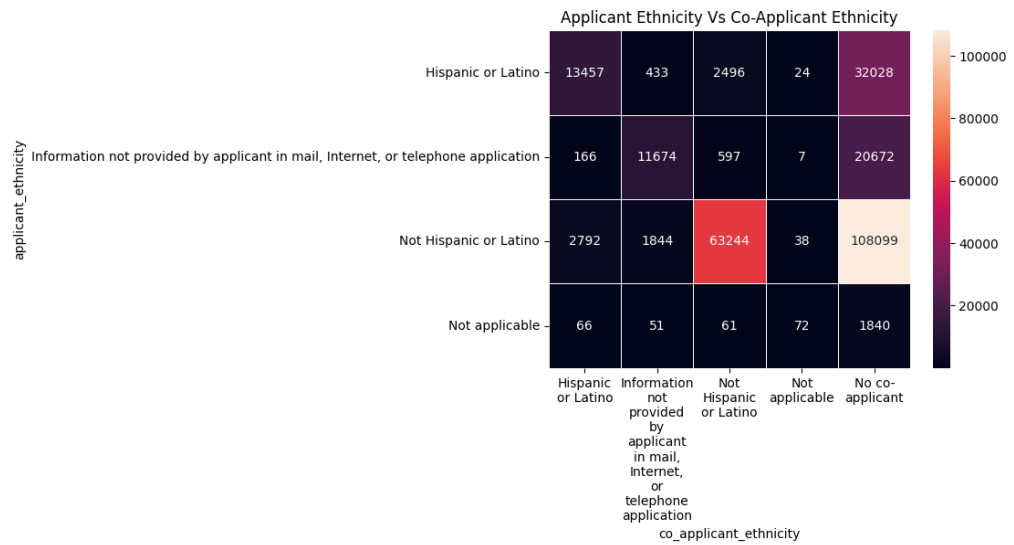
1.1 – גרפים בסיסיים לכל העמודות:



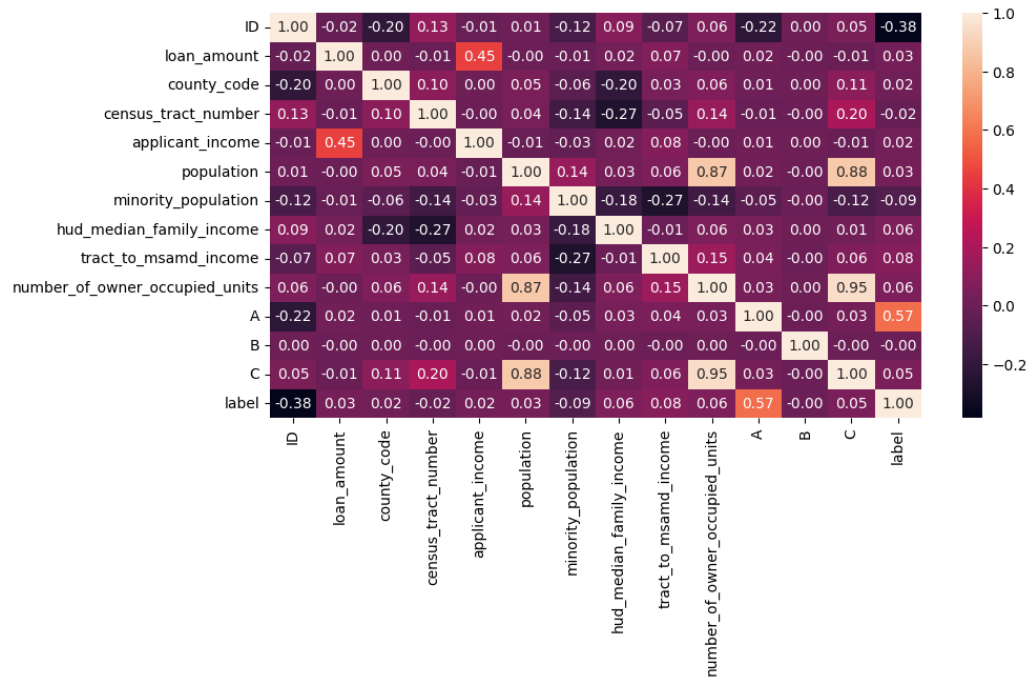
1.2 פורים באישור הלוואות בקטגוריות מסוימות בפיצ'רים מסוימים (תא 10):



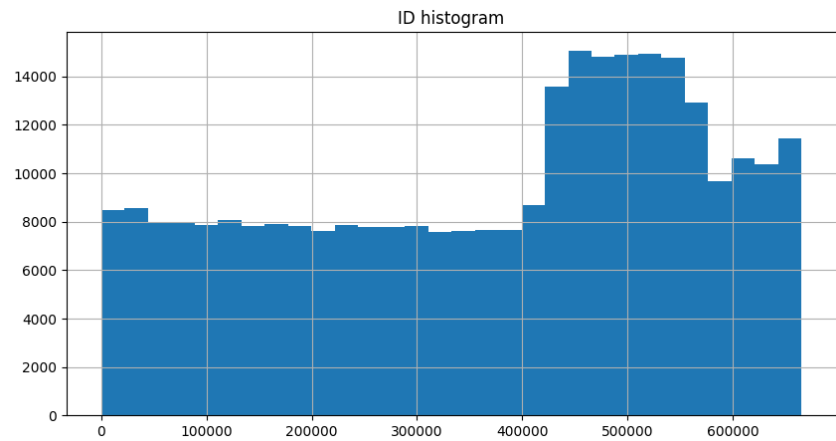
1.3 קורלציה בין פיצ'רים של מגיש הבקשה לבין מגיש הבקשה הנוסף (תאים 12,14,16):

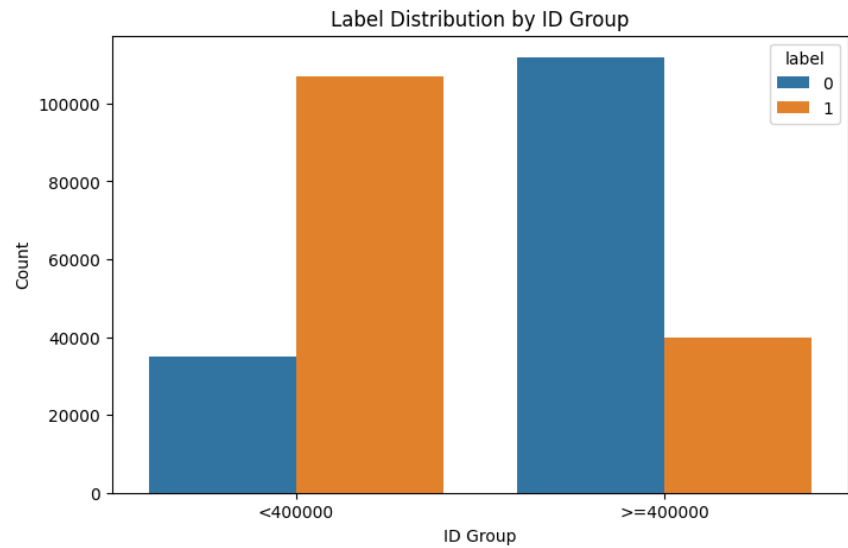


1.4 קורלציה בפיצ'רים המספריים (תא 19):



1.5 היסטוגרמה והתפלגות של ID ביחס לlabel (תאים 22 ו24):

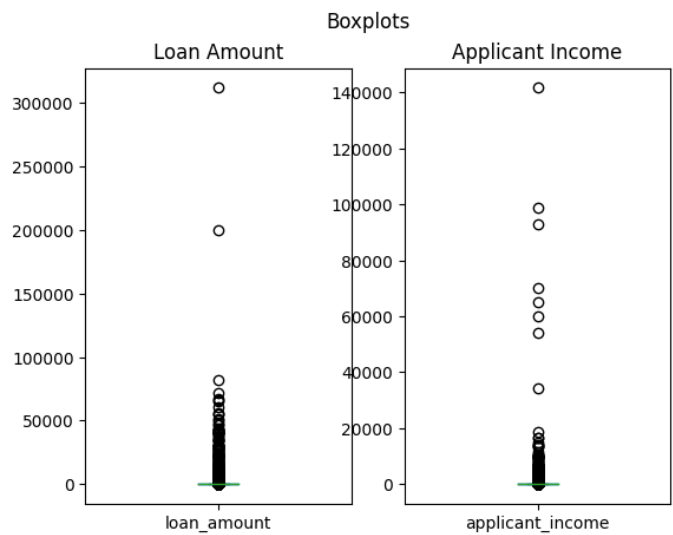




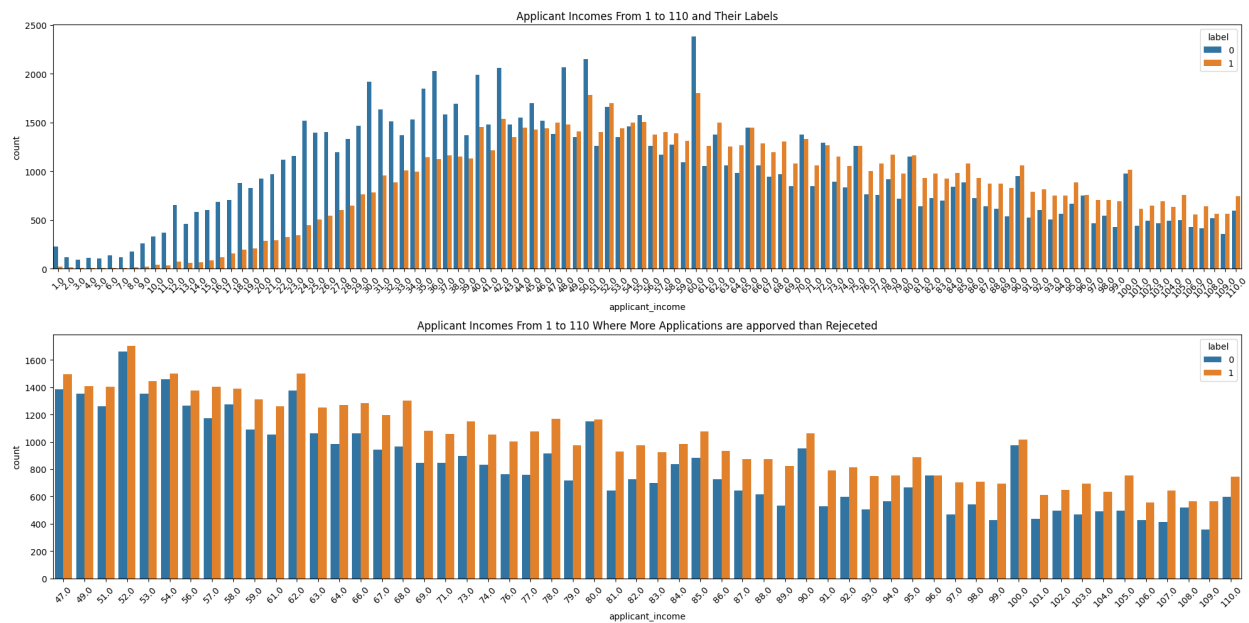
1.6 כמות הנתונים החריגים בLoan_amount וב-applicant_income (תא 28):

Number of outliers for loan_amount: 10991

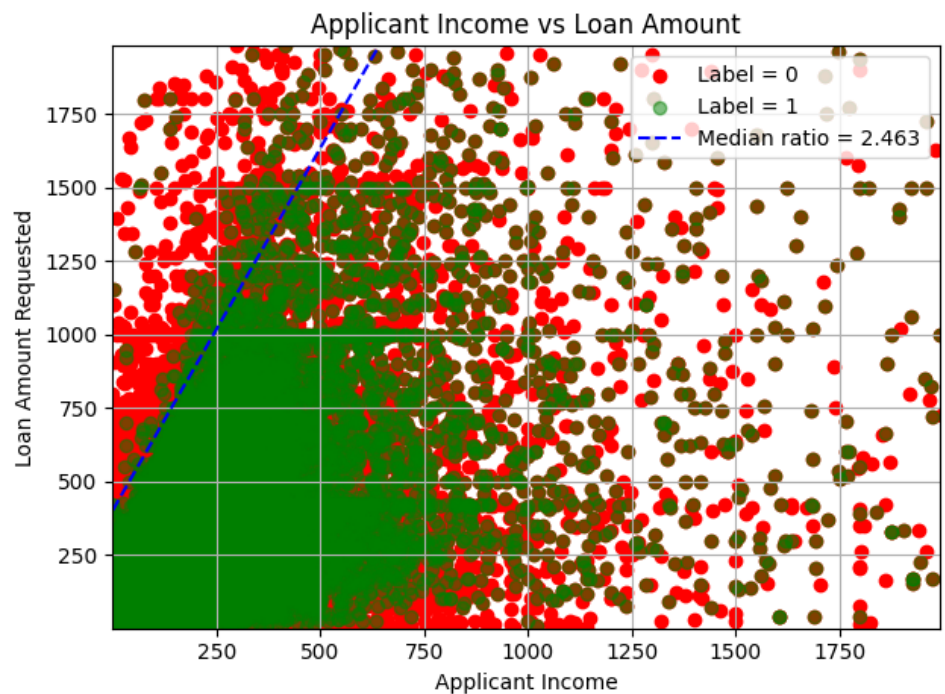
Number of outliers for applicant_income: 19548



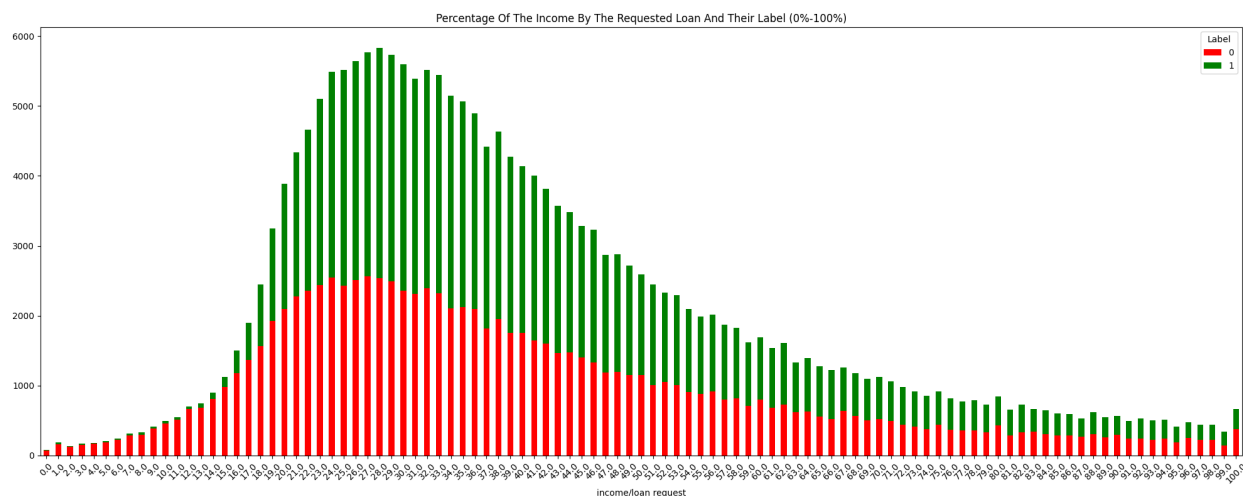
1.7 בגרף הגבוה התפלגות של גובה ההכנסה של מגיש הבקשה ביחס לlabel, בגרף הנמוך רק הערכים בהם יש יותר בקשות שאושרו מאשר שנדחו (תא 30):



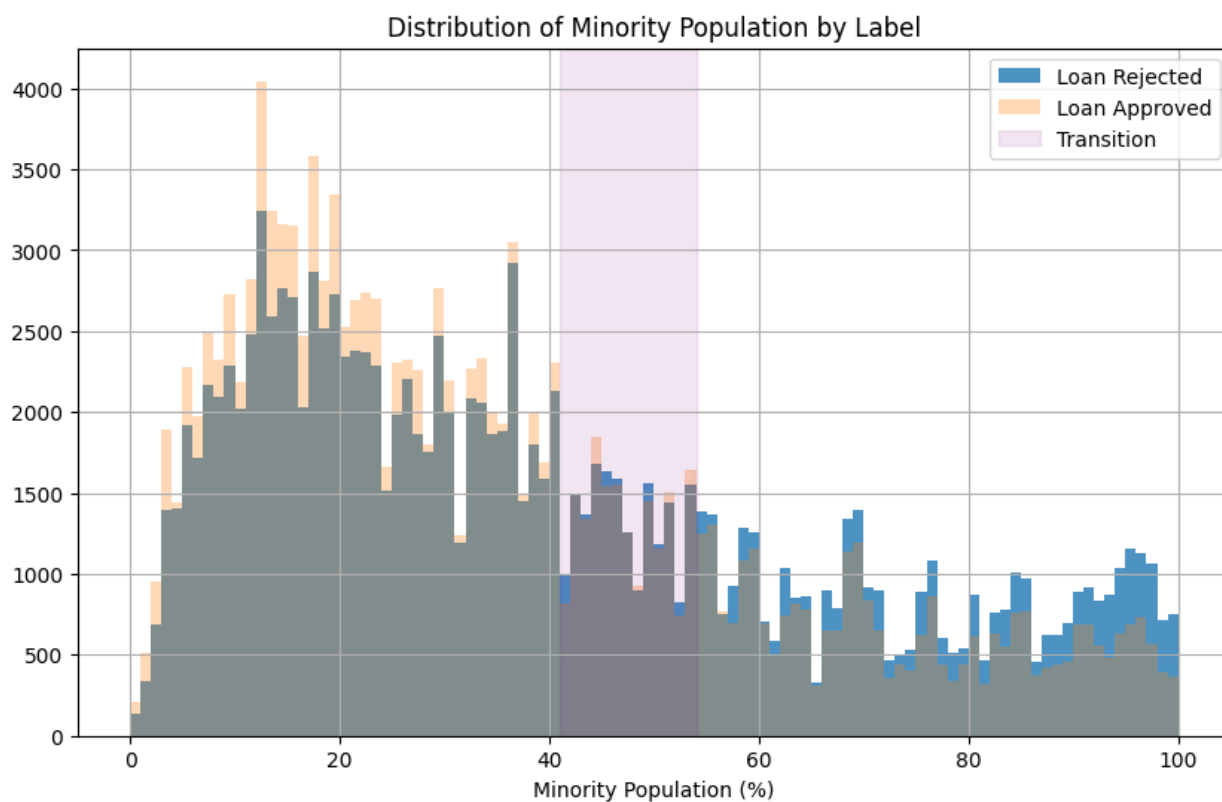
1.8 מיפוי גובה ההכנסה וגודל ההלוואה ביחס לאישור ההלוואה (תא 33):



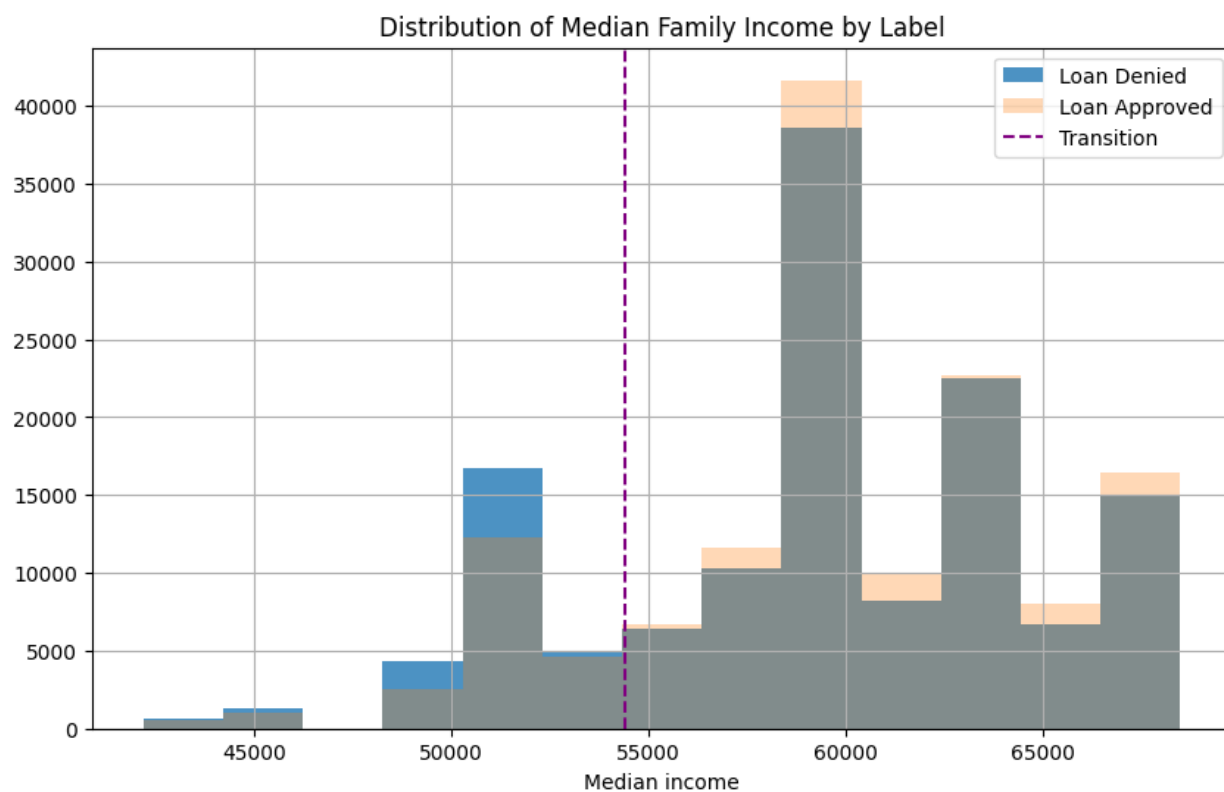
1.9 גרף עמודות של כמה אחוזים גובה ההכנסה של מגיש הבקשה מתוך גובה ההלוואה שביקש, מחולק לפי אישור ההלוואה (תא 36):



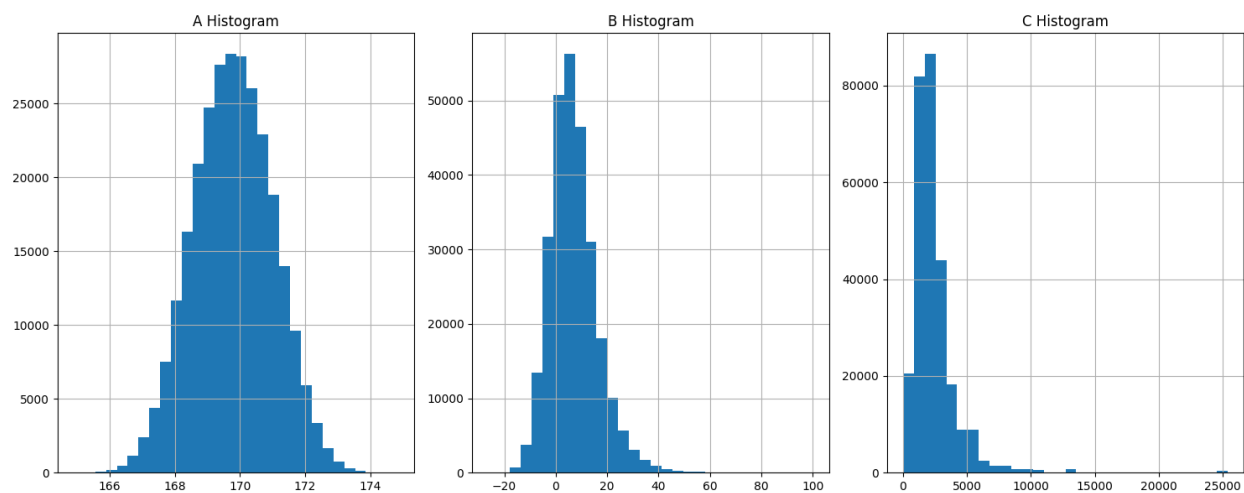
1.10 היסטוגרמה של אחוז אוכלוסיית המיעוטים באזור ביחס לאישור ההלוואה (תא 40):



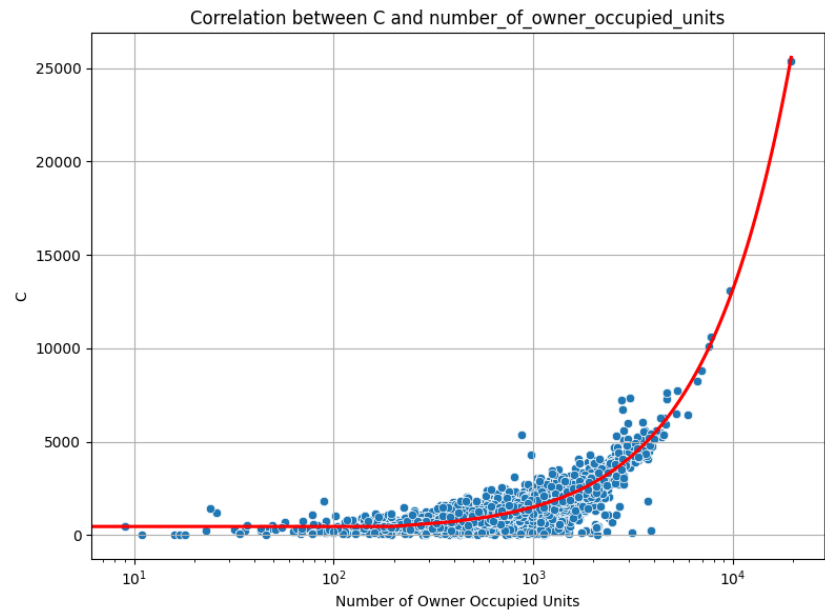
1.11 היסטוגרמה של median family income באזור ביחס לאישור ההלוואה (תא 43):



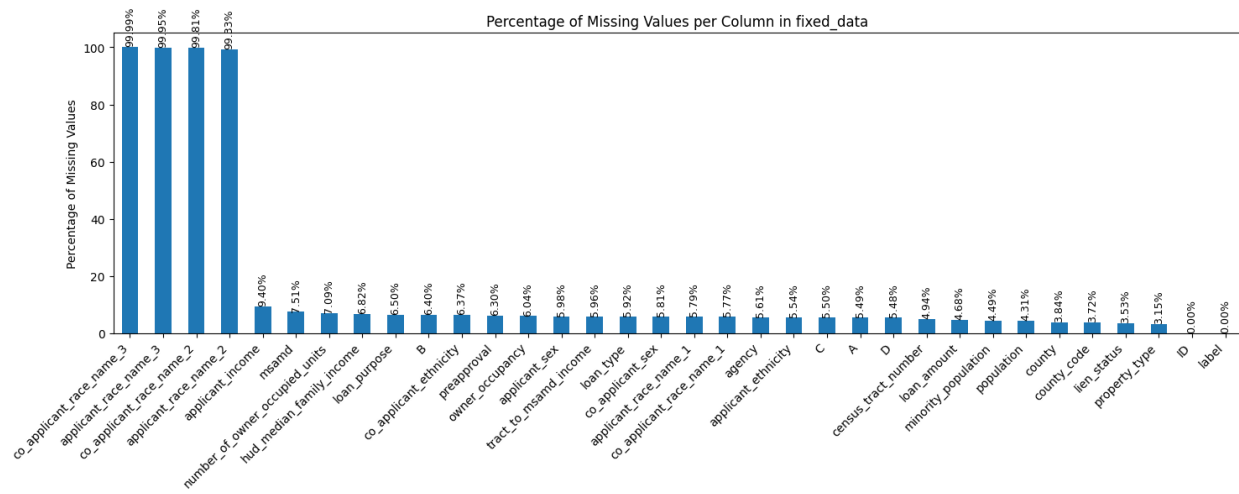
1.12 היסטוגרמה של עמודות A, B ו-C (תא 46):



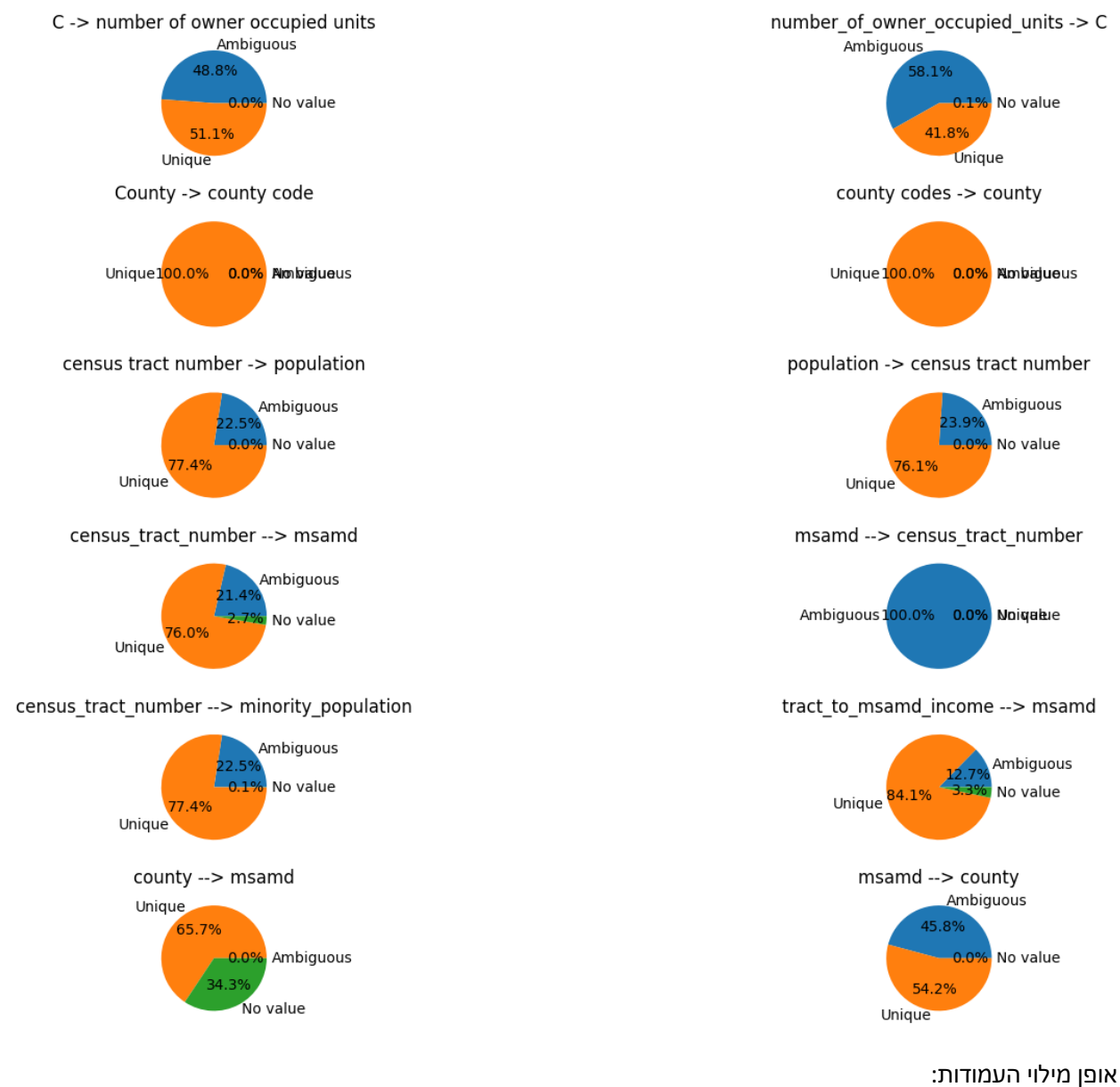
1.13 קורלציה בין עמודה C לבין number_of_owner_occupied_units (תא 48):



2.1 כמה ערכים חסרים בכל עמודה (תא 53):

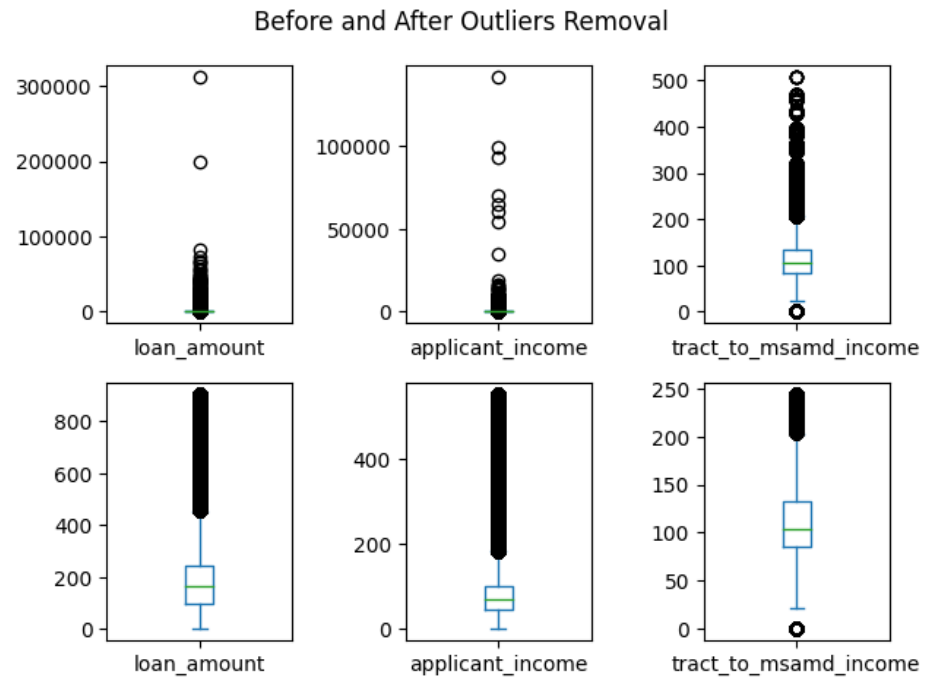


2.2 אחוזי חד חד ערכיות של יחסים בין עמודות כפי שמחושב בתא 61 ואופן מילוי העמודות בשלב 2 בסעיף 1.1 (תא 61):

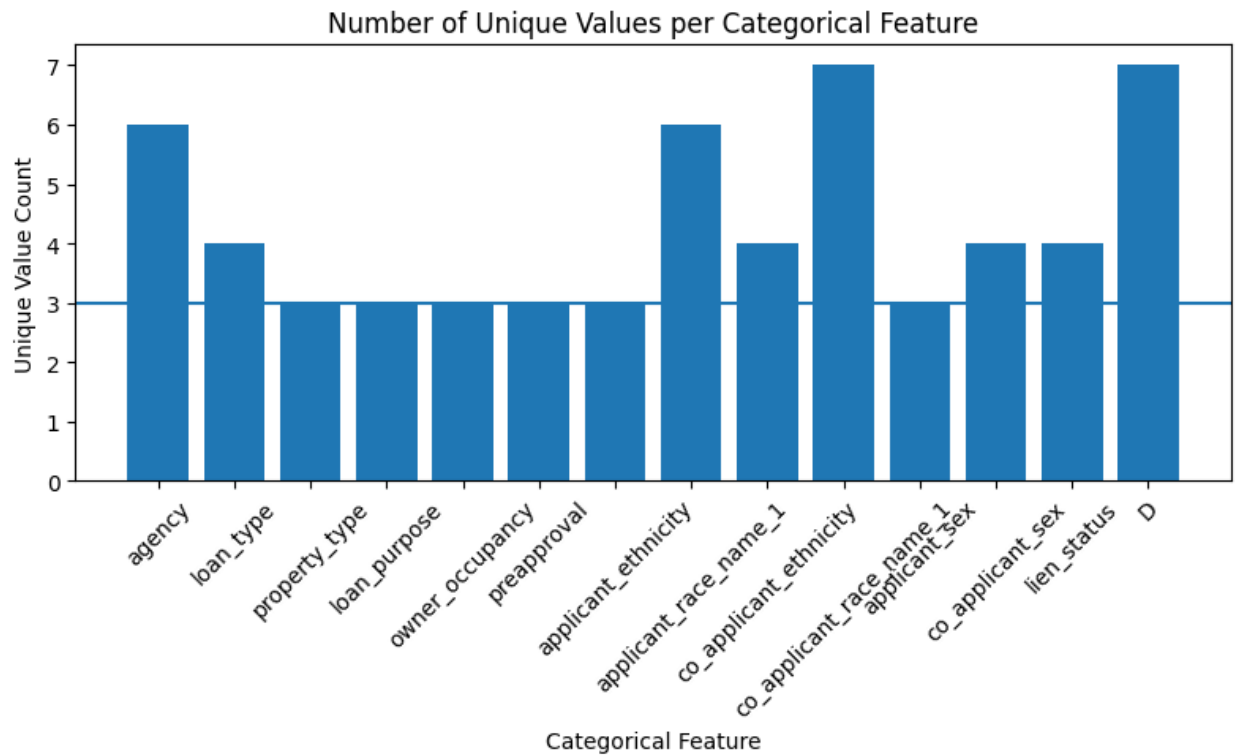


- county <-> county_code ○
- census_tract_number <-> population ○
- tract_to_msamd_income -> msamd ○
- census_tract_number -> minority_population ○

2.3 נתונים חריגים בloan_amount, applicant_income, ו-tract_to_msamd_income לפני ואחרי מחיקתם (תא 83):



2.4 מספר הקטגוריות בפיצ'רים הקטגוריאליים:



2.5 השוואה של Accuracy על דאטה עם הורדת מימדים ע"י RF, ע"י PCA ובלי הורדת מימדים (תא 166):

Random Forest:

RF Selection:

Train: 1.000

Val: 0.897

Difference = 0.103

PCA results:

Train: 1.000

Val: 0.828

Difference = 0.172

RAW results:

Train: 1.000

Val: 0.897

Difference = 0.103

MLP:

RF Selection:

Train: 0.887

Val: 0.882

Difference = 0.005

PCA results:

Train: 0.884

Val: 0.877

Difference = 0.007

RAW results:

Train: 0.886

Val: 0.879

Difference = 0.007

KNN:

RF Selection:

Train: 0.837

Val: 0.828

Difference = 0.009

PCA results:

Train: 0.818

Val: 0.811

Difference = 0.008

RAW results:

Train: 0.822

Val: 0.813

Difference = 0.010

Logistic Regression:

RF Selection:

Train: 0.814

Val: 0.817

Difference = -0.002

PCA results:

Train: 0.813

Val: 0.816

Difference = -0.003

RAW results:

Train: 0.815

Val: 0.818
 Difference = -0.003

נספח 3.1 ההיפרפרמטרים המלאים של כל המודלים (תא 156):

```
Random Forest{
  'bootstrap': True,
  'ccp_alpha': 0.0,
  'class_weight': None,
  'criterion': 'gini',
  'max_depth': 31,
  'max_features': 0.5,
  'max_leaf_nodes': None,
  'max_samples': None,
  'min_impurity_decrease': 0.0,
  'min_samples_leaf': 1,
  'min_samples_split': 2,
  'min_weight_fraction_leaf': 0.0,
  'n_estimators': 291,
  'n_jobs': -1,
  'oob_score': False,
  'random_state': 42,
  'verbose': 0,
  'warm_start': False}

Multi-layerd Perceptron{
  'activation': 'tanh',
  'alpha': 0.0008637846941310894,
  'batch_size': 'auto',
  'beta_1': 0.9,
  'beta_2': 0.999,
  'early_stopping': True,
  'epsilon': 1e-08,
  'hidden_layer_sizes': (100,),
  'learning_rate': 'constant',
  'learning_rate_init': 0.0034791409848015362,
  'max_fun': 15000,
  'max_iter': 200,
  'momentum': 0.9,
  'n_iter_no_change': 10,
  'nesterovs_momentum': True,
  'power_t': 0.5,
  'random_state': 42,
  'shuffle': True,
  'solver': 'adam',
  'tol': 0.0001,
  'validation_fraction': 0.1,
  'verbose': False,
  'warm_start': False}
```

```

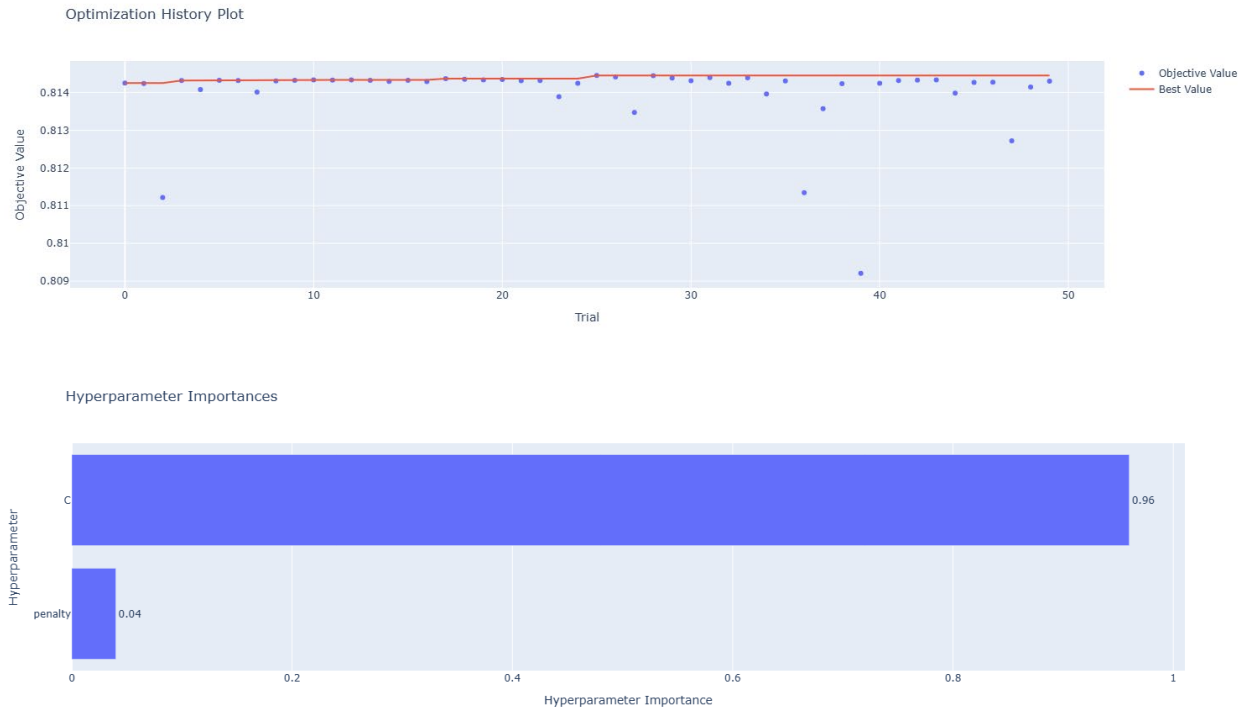
KNN-Classfier{
    'algorithm': 'auto',
    'leaf_size': 30,
    'metric': 'minkowski',
    'metric_params': None,
    'n_jobs': -1,
    'n_neighbors': 40,
    'p': 1,
    'weights': 'uniform'}

Linear Regression{
    'C': 0.0003540266337546,
    'class_weight': None,
    'dual': False,
    'fit_intercept': True,
    'intercept_scaling': 1,
    'l1_ratio': None,
    'max_iter': 100,
    'multi_class': 'auto',
    'n_jobs': None,
    'penalty': 'l2',
    'random_state': None,
    'solver': 'saga',
    'tol': 0.0001,
    'verbose': 0,
    'warm_start': False}

```

נספח 3.2 גרפים של תהליך טיוב ההיפרפרמטרים של המודלים וחשיבות ההיפרפרמטרים והלוגים של תהליך הטיוב לפי מודל:

רגרסיה לינארית (תאים 120 ו-123):



[I 2025-07-11 17:52:27,244] Trial 0 finished with value: 0.8142518479860833 and parameters: {'C': 0.00869041022360225, 'penalty': 'l1'}. Best is trial 0 with value: 0.8142518479860833.

[I 2025-07-11 17:52:31,694] Trial 1 finished with value: 0.8142390782248778 and parameters: {'C': 0.0031874359674166188, 'penalty': 'l2'}. Best is trial 0 with value: 0.8142518479860833.

[I 2025-07-11 17:52:36,052] Trial 2 finished with value: 0.8112169176219932 and parameters: {'C': 0.0003501457946313366, 'penalty': 'l1'}. Best is trial 0 with value: 0.8142518479860833.

[I 2025-07-11 17:52:40,562] Trial 3 finished with value: 0.8143199530168163 and parameters: {'C': 0.3392769618255998, 'penalty': 'l2'}. Best is trial 3 with value: 0.8143199530168163.

[I 2025-07-11 17:52:45,756] Trial 4 finished with value: 0.8140815835068457 and parameters: {'C': 0.00021255546585945728, 'penalty': 'l2'}. Best is trial 3 with value: 0.8143199530168163.

[I 2025-07-11 17:52:51,853] Trial 5 finished with value: 0.8143242096944754 and parameters: {'C': 0.9359138192280809, 'penalty': 'l1'}. Best is trial 5 with value: 0.8143242096944754.

[I 2025-07-11 17:52:57,365] Trial 6 finished with value: 0.8143156966109292 and parameters: {'C': 0.5066833536885004, 'penalty': 'l1'}. Best is trial 5 with value: 0.8143242096944754.

[I 2025-07-11 17:53:02,531] Trial 7 finished with value: 0.8140134809220626 and parameters: {'C': 0.0036959736224322914, 'penalty': 'l1'}. Best is trial 5 with value: 0.8143242096944754.

[I 2025-07-11 17:53:07,981] Trial 8 finished with value: 0.8143071837085645 and parameters: {'C': 0.09906581391797013, 'penalty': 'l1'}. Best is trial 5 with value: 0.8143242096944754.

[I 2025-07-11 17:53:13,262] Trial 9 finished with value: 0.8143242096944754 and parameters: {'C': 0.8457865873169674, 'penalty': 'l1'}. Best is trial 5 with value: 0.8143242096944754.

[I 2025-07-11 17:53:18,070] Trial 10 finished with value: 0.8143327231403845 and parameters: {'C': 0.05536999248097811, 'penalty': 'l2'}. Best is trial 10 with value: 0.8143327231403845.

[I 2025-07-11 17:53:22,598] Trial 11 finished with value: 0.8143284667344975 and parameters: {'C': 0.05373585976521838, 'penalty': 'l2'}. Best is trial 10 with value: 0.8143327231403845.

[I 2025-07-11 17:53:26,961] Trial 12 finished with value: 0.8143369797274531 and parameters: {'C': 0.058506918847369055, 'penalty': 'l2'}. Best is trial 12 with value: 0.8143369797274531.

[I 2025-07-11 17:53:31,230] Trial 13 finished with value: 0.8143242100568383 and parameters: {'C': 0.044849079843517256, 'penalty': 'l2'}. Best is trial 12 with value: 0.8143369797274531.

[I 2025-07-11 17:53:35,182] Trial 14 finished with value: 0.8142944143097219 and parameters: {'C': 0.024384477856027455, 'penalty': 'l2'}. Best is trial 12 with value: 0.8143369797274531.

[I 2025-07-11 17:53:39,894] Trial 15 finished with value: 0.8143242096944754 and parameters: {'C': 0.1595054437489305, 'penalty': 'l2'}. Best is trial 12 with value: 0.8143369797274531.

[I 2025-07-11 17:53:44,117] Trial 16 finished with value: 0.8142901578132442 and parameters: {'C': 0.010711883339934426, 'penalty': 'l2'}. Best is trial 12 with value: 0.8143369797274531.

[I 2025-07-11 17:53:48,298] Trial 17 finished with value: 0.8143710317898659 and parameters: {'C': 0.0007884268748623374, 'penalty': 'l2'}. Best is trial 17 with value: 0.8143710317898659.

[I 2025-07-11 17:53:52,745] Trial 18 finished with value: 0.8143540055321827 and parameters: {'C': 0.000746404673447678, 'penalty': 'l2'}. Best is trial 17 with value: 0.8143710317898659.

[I 2025-07-11 17:53:57,175] Trial 19 finished with value: 0.8143369793650901 and parameters: {'C': 0.0007398672682435838, 'penalty': 'l2'}. Best is trial 17 with value: 0.8143710317898659.

[I 2025-07-11 17:54:01,350] Trial 20 finished with value: 0.814345492176864 and parameters: {'C': 0.0010155252439513352, 'penalty': 'l2'}. Best is trial 17 with value: 0.8143710317898659.

[I 2025-07-11 17:54:05,718] Trial 21 finished with value: 0.8143114396614978 and parameters: {'C': 0.0010385249936069524, 'penalty': 'l2'}. Best is trial 17 with value: 0.8143710317898659.

[I 2025-07-11 17:54:09,923] Trial 22 finished with value: 0.8143156959767939 and parameters: {'C': 0.0011018889110740412, 'penalty': 'l2'}. Best is trial 17 with value: 0.8143710317898659.

[I 2025-07-11 17:54:14,305] Trial 23 finished with value: 0.8138900367264007 and parameters: {'C': 0.0001453164152472144, 'penalty': 'l2'}. Best is trial 17 with value: 0.8143710317898659.

[I 2025-07-11 17:54:18,364] Trial 24 finished with value: 0.8142433357178536 and parameters: {'C': 0.002199076606905365, 'penalty': 'l2'}. Best is trial 17 with value: 0.8143710317898659.

[I 2025-07-11 17:54:22,242] Trial 25 finished with value: 0.8144519066723948 and parameters: {'C': 0.0003540266337546, 'penalty': 'l2'}. Best is trial 25 with value: 0.8144519066723948.

[I 2025-07-11 17:54:26,356] Trial 26 finished with value: 0.8144135974793691 and parameters: {'C': 0.0003777910646389006, 'penalty': 'l2'}. Best is trial 25 with value: 0.8144519066723948.

[I 2025-07-11 17:54:30,208] Trial 27 finished with value: 0.8134728930055036 and parameters: {'C': 0.0001009467573430944, 'penalty': 'l2'}. Best is trial 25 with value: 0.8144519066723948.

[I 2025-07-11 17:54:33,910] Trial 28 finished with value: 0.8144476497229635 and parameters: {'C': 0.00040212688862217166, 'penalty': 'l2'}. Best is trial 25 with value: 0.8144519066723948.

[I 2025-07-11 17:54:38,138] Trial 29 finished with value: 0.8143838011887083 and parameters: {'C': 0.0004310311160935933, 'penalty': 'l2'}. Best is trial 25 with value: 0.8144519066723948.

[I 2025-07-11 17:54:42,026] Trial 30 finished with value: 0.8143114396614977 and parameters: {'C': 0.0002801700210659719, 'penalty': 'l2'}. Best is trial 25 with value: 0.8144519066723948.

[I 2025-07-11 17:54:46,357] Trial 31 finished with value: 0.8143965710405047 and parameters: {'C': 0.0004434768671093744, 'penalty': 'l2'}. Best is trial 25 with value: 0.8144519066723948.

[I 2025-07-11 17:54:50,336] Trial 32 finished with value: 0.8142475922143314 and parameters: {'C': 0.002189526733515378, 'penalty': 'l2'}. Best is trial 25 with value: 0.8144519066723948.

[I 2025-07-11 17:54:54,806] Trial 33 finished with value: 0.8143923144534361 and parameters: {'C': 0.0004411170699309814, 'penalty': 'l2'}. Best is trial 25 with value: 0.8144519066723948.

[I 2025-07-11 17:54:58,706] Trial 34 finished with value: 0.8139623988877466 and parameters: {'C': 0.0001881024471406837, 'penalty': 'l2'}. Best is trial 25 with value: 0.8144519066723948.

[I 2025-07-11 17:55:02,865] Trial 35 finished with value: 0.8143071837085645 and parameters: {'C': 0.005349103363996304, 'penalty': 'l2'}. Best is trial 25 with value: 0.8144519066723948.

[I 2025-07-11 17:55:07,274] Trial 36 finished with value: 0.8113446155058203 and parameters: {'C': 0.00037041292770648344, 'penalty': 'l1'}. Best is trial 25 with value: 0.8144519066723948.

[I 2025-07-11 17:55:11,328] Trial 37 finished with value: 0.8135750493739234 and parameters: {'C': 0.00011136375312555013, 'penalty': 'l2'}. Best is trial 25 with value: 0.8144519066723948.

[I 2025-07-11 17:55:15,269] Trial 38 finished with value: 0.8142348223625353 and parameters: {'C': 0.0019473750666550546, 'penalty': 'l2'}. Best is trial 25 with value: 0.8144519066723948.

[I 2025-07-11 17:55:19,744] Trial 39 finished with value: 0.8092078111527983 and parameters: {'C': 0.00022497728215016053, 'penalty': 'l1'}. Best is trial 25 with value: 0.8144519066723948.

[I 2025-07-11 17:55:24,555] Trial 40 finished with value: 0.8142433349025371 and parameters: {'C': 0.009600433952193357, 'penalty': 'l1'}. Best is trial 25 with value: 0.8144519066723948.

[I 2025-07-11 17:55:28,934] Trial 41 finished with value: 0.8143156961579754 and parameters: {'C': 0.0005056306216190977, 'penalty': 'l2'}. Best is trial 25 with value: 0.8144519066723948.

[I 2025-07-11 17:55:33,185] Trial 42 finished with value: 0.8143284658285902 and parameters: {'C': 0.0005174892081477141, 'penalty': 'l2'}. Best is trial 25 with value: 0.8144519066723948.

[I 2025-07-11 17:55:36,908] Trial 43 finished with value: 0.8143369789121364 and parameters: {'C': 0.00031973881756631783, 'penalty': 'l2'}. Best is trial 25 with value: 0.8144519066723948.

[I 2025-07-11 17:55:41,049] Trial 44 finished with value: 0.8139836819136796 and parameters: {'C': 0.00019384446451458224, 'penalty': 'l2'}. Best is trial 25 with value: 0.8144519066723948.

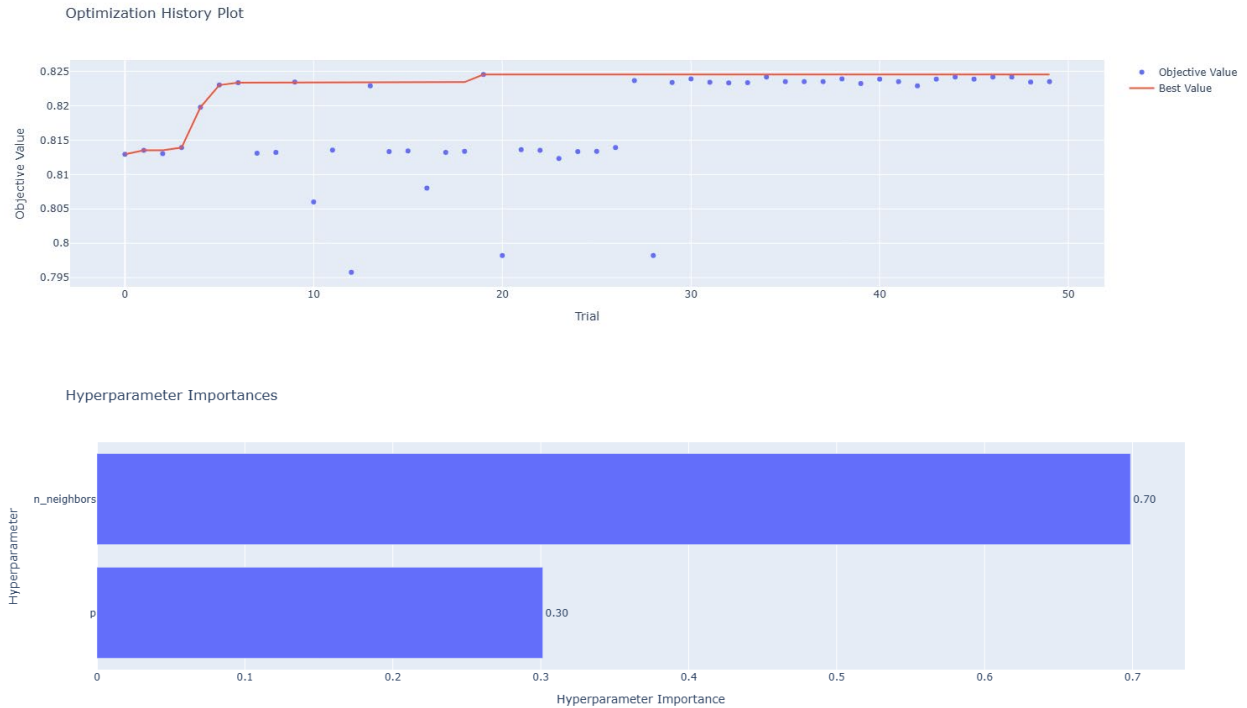
[I 2025-07-11 17:55:45,451] Trial 45 finished with value: 0.8142688746061296 and parameters: {'C': 0.0015351941119072763, 'penalty': 'l2'}. Best is trial 25 with value: 0.8144519066723948.

[I 2025-07-11 17:55:49,677] Trial 46 finished with value: 0.8142731303778813 and parameters: {'C': 0.003851291912448635, 'penalty': 'l2'}. Best is trial 25 with value: 0.8144519066723948.

[I 2025-07-11 17:55:54,808] Trial 47 finished with value: 0.8127194837980959 and parameters: {'C': 0.0006340627151308605, 'penalty': 'l1'}. Best is trial 25 with value: 0.8144519066723948.

[I 2025-07-11 17:55:58,941] Trial 48 finished with value: 0.8141454328564173 and parameters: {'C': 0.00024009914120037203, 'penalty': 'l2'}. Best is trial 25 with value: 0.8144519066723948.

[I 2025-07-11 17:56:03,241] Trial 49 finished with value: 0.8143029264873606 and parameters: {'C': 0.001427166293604604, 'penalty': 'l2'}. Best is trial 25 with value: 0.8144519066723948.



```
[I 2025-07-11 18:12:42,937] Trial 10 finished with value: 0.8060111275812021 and parameters:
{'n_neighbors': 10, 'p': 2}. Best is trial 10 with value: 0.8060111275812021.
[I 2025-07-11 18:13:16,708] Trial 12 finished with value: 0.7957783358427013 and parameters:
{'n_neighbors': 5, 'p': 2}. Best is trial 10 with value: 0.8060111275812021.
[I 2025-07-11 18:13:34,168] Trial 16 finished with value: 0.808033000187973 and parameters:
{'n_neighbors': 12, 'p': 2}. Best is trial 16 with value: 0.808033000187973.
[I 2025-07-11 18:13:39,266] Trial 0 finished with value: 0.8129493358761646 and parameters:
{'n_neighbors': 19, 'p': 2}. Best is trial 0 with value: 0.8129493358761646.
[I 2025-07-11 18:14:02,787] Trial 1 finished with value: 0.8135282299056644 and parameters:
{'n_neighbors': 39, 'p': 2}. Best is trial 1 with value: 0.8135282299056644.
[I 2025-07-11 18:14:06,552] Trial 17 finished with value: 0.8132260134014813 and parameters:
{'n_neighbors': 35, 'p': 2}. Best is trial 1 with value: 0.8135282299056644.
[I 2025-07-11 18:14:10,099] Trial 3 finished with value: 0.8139113204770598 and parameters:
{'n_neighbors': 46, 'p': 2}. Best is trial 3 with value: 0.8139113204770598.
[I 2025-07-11 18:14:18,354] Trial 11 finished with value: 0.8135537662573993 and parameters:
{'n_neighbors': 24, 'p': 2}. Best is trial 3 with value: 0.8139113204770598.
[I 2025-07-11 18:14:30,712] Trial 2 finished with value: 0.8130387227551511 and parameters:
{'n_neighbors': 55, 'p': 2}. Best is trial 3 with value: 0.8139113204770598.
[I 2025-07-11 18:14:36,160] Trial 18 finished with value: 0.8133664775134747 and parameters:
{'n_neighbors': 21, 'p': 2}. Best is trial 3 with value: 0.8139113204770598.
[I 2025-07-11 18:15:06,024] Trial 14 finished with value: 0.8133281654215454 and parameters:
{'n_neighbors': 23, 'p': 2}. Best is trial 3 with value: 0.8139113204770598.
[I 2025-07-11 18:15:11,748] Trial 15 finished with value: 0.8134175527534854 and parameters:
{'n_neighbors': 29, 'p': 2}. Best is trial 3 with value: 0.8139113204770598.
[I 2025-07-11 18:18:03,589] Trial 4 finished with value: 0.8197768733603003 and parameters:
{'n_neighbors': 17, 'p': 1}. Best is trial 4 with value: 0.8197768733603003.
[I 2025-07-11 18:18:04,998] Trial 13 finished with value: 0.8228884187585841 and parameters:
{'n_neighbors': 29, 'p': 1}. Best is trial 13 with value: 0.8228884187585841.
[I 2025-07-11 18:18:10,674] Trial 19 finished with value: 0.824544228410504 and parameters:
{'n_neighbors': 40, 'p': 1}. Best is trial 19 with value: 0.824544228410504.
[I 2025-07-11 18:18:12,158] Trial 5 finished with value: 0.8230033479682944 and parameters:
{'n_neighbors': 24, 'p': 1}. Best is trial 19 with value: 0.824544228410504.
```


[I 2025-07-11 18:19:22,306] Trial 23 finished with value: 0.8123278766101153 and parameters: {'n_neighbors': 17, 'p': 2}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:20:44,553] Trial 24 finished with value: 0.8133281654215454 and parameters: {'n_neighbors': 23, 'p': 2}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:20:48,289] Trial 8 finished with value: 0.8132260134014813 and parameters: {'n_neighbors': 35, 'p': 2}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:20:58,899] Trial 22 finished with value: 0.813511201564394 and parameters: {'n_neighbors': 33, 'p': 2}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:21:21,749] Trial 21 finished with value: 0.8136133571174972 and parameters: {'n_neighbors': 36, 'p': 2}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:21:26,390] Trial 25 finished with value: 0.8133664775134747 and parameters: {'n_neighbors': 21, 'p': 2}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:21:35,985] Trial 26 finished with value: 0.8139198347382857 and parameters: {'n_neighbors': 48, 'p': 2}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:22:44,739] Trial 7 finished with value: 0.8130940566658174 and parameters: {'n_neighbors': 20, 'p': 2}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:27:44,313] Trial 28 finished with value: 0.7982343757103305 and parameters: {'n_neighbors': 5, 'p': 1}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:28:04,603] Trial 20 finished with value: 0.7982343757103305 and parameters: {'n_neighbors': 5, 'p': 1}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:28:27,817] Trial 27 finished with value: 0.8236503453073029 and parameters: {'n_neighbors': 26, 'p': 1}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:29:30,318] Trial 9 finished with value: 0.8234332641681192 and parameters: {'n_neighbors': 41, 'p': 1}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:29:40,555] Trial 29 finished with value: 0.8233821867539305 and parameters: {'n_neighbors': 49, 'p': 1}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:30:19,791] Trial 30 finished with value: 0.8238972307997233 and parameters: {'n_neighbors': 44, 'p': 1}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:30:31,794] Trial 31 finished with value: 0.8234034692363194 and parameters: {'n_neighbors': 47, 'p': 1}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:30:44,656] Trial 6 finished with value: 0.8233481307961161 and parameters: {'n_neighbors': 35, 'p': 1}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:35:37,372] Trial 34 finished with value: 0.8241696514661992 and parameters: {'n_neighbors': 42, 'p': 1}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:35:46,924] Trial 33 finished with value: 0.8233481307961161 and parameters: {'n_neighbors': 35, 'p': 1}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:36:05,566] Trial 35 finished with value: 0.8235098850907111 and parameters: {'n_neighbors': 43, 'p': 1}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:36:20,224] Trial 42 finished with value: 0.8228884187585841 and parameters: {'n_neighbors': 29, 'p': 1}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:36:20,361] Trial 40 finished with value: 0.8238674306136605 and parameters: {'n_neighbors': 28, 'p': 1}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:36:20,644] Trial 37 finished with value: 0.8235098850907111 and parameters: {'n_neighbors': 43, 'p': 1}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:36:20,733] Trial 36 finished with value: 0.8235098850907111 and parameters: {'n_neighbors': 43, 'p': 1}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:36:21,975] Trial 39 finished with value: 0.823224690495856 and parameters: {'n_neighbors': 45, 'p': 1}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:36:22,527] Trial 41 finished with value: 0.8235098850907111 and parameters: {'n_neighbors': 43, 'p': 1}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:36:23,926] Trial 32 finished with value: 0.8233268549268514 and parameters: {'n_neighbors': 51, 'p': 1}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:38:28,596] Trial 43 finished with value: 0.8238674306136605 and parameters: {'n_neighbors': 28, 'p': 1}. Best is trial 19 with value: 0.824544228410504.

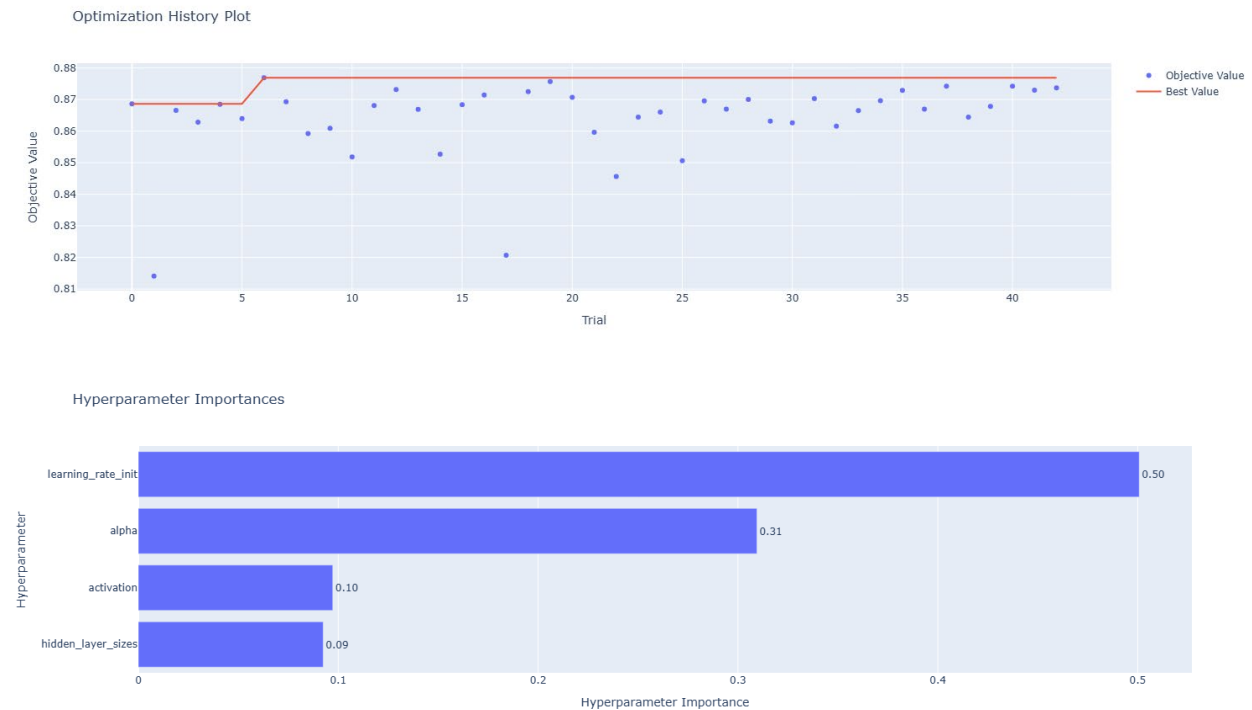
[I 2025-07-11 18:38:28,955] Trial 38 finished with value: 0.8238972307997233 and parameters: {'n_neighbors': 44, 'p': 1}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:38:37,287] Trial 45 finished with value: 0.8238674306136605 and parameters: {'n_neighbors': 28, 'p': 1}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:39:04,230] Trial 48 finished with value: 0.8234332641681192 and parameters: {'n_neighbors': 41, 'p': 1}. Best is trial 19 with value: 0.824544228410504.

[I 2025-07-11 18:39:05,046] Trial 46 finished with value: 0.8241696514661992 and parameters: {'n_neighbors': 42, 'p': 1}. Best is trial 19 with value: 0.824544228410504.
 [I 2025-07-11 18:39:06,714] Trial 44 finished with value: 0.8241696514661992 and parameters: {'n_neighbors': 42, 'p': 1}. Best is trial 19 with value: 0.824544228410504.
 [I 2025-07-11 18:39:28,598] Trial 47 finished with value: 0.8241696514661992 and parameters: {'n_neighbors': 42, 'p': 1}. Best is trial 19 with value: 0.824544228410504.
 [I 2025-07-11 18:39:29,080] Trial 49 finished with value: 0.8235098850907111 and parameters: {'n_neighbors': 43, 'p': 1}. Best is trial 19 with value: 0.824544228410504.

:(144 ו 141 תאים) MLP



[I 2025-07-11 22:11:31,703] Trial 4 finished with value: 0.8684847922757655 and parameters: {'hidden_layer_sizes': (20,), 'activation': 'tanh', 'alpha': 1.102764743599042e-05, 'learning_rate_init': 0.0030350083399687594}. Best is trial 4 with value: 0.8684847922757655.
 [I 2025-07-11 22:14:08,637] Trial 1 finished with value: 0.8140943558045919 and parameters: {'hidden_layer_sizes': (100,), 'activation': 'tanh', 'alpha': 0.0011017971080137707, 'learning_rate_init': 1.029563567019831e-05}. Best is trial 4 with value: 0.8684847922757655.
 [I 2025-07-11 22:14:41,630] Trial 8 finished with value: 0.8592480299533133 and parameters: {'hidden_layer_sizes': (20,), 'activation': 'relu', 'alpha': 6.597188453878661e-05, 'learning_rate_init': 0.0052593059522298684}. Best is trial 4 with value: 0.8684847922757655.
 [I 2025-07-11 22:15:59,706] Trial 15 finished with value: 0.868357088684722 and parameters: {'hidden_layer_sizes': (20,), 'activation': 'tanh', 'alpha': 0.00036662386434865857, 'learning_rate_init': 0.0015586455946380558}. Best is trial 4 with value: 0.8684847922757655.
 [I 2025-07-11 22:18:21,291] Trial 18 finished with value: 0.8724689376607266 and parameters: {'hidden_layer_sizes': (40,), 'activation': 'tanh', 'alpha': 0.002134039754574151, 'learning_rate_init': 0.005639128473640169}. Best is trial 18 with value: 0.8724689376607266.
 [I 2025-07-11 22:19:52,927] Trial 6 finished with value: 0.8768957743515713 and parameters: {'hidden_layer_sizes': (100,), 'activation': 'tanh', 'alpha': 0.0008637846941310894, 'learning_rate_init': 0.0034791409848015362}. Best is trial 6 with value: 0.8768957743515713.
 [I 2025-07-11 22:22:29,260] Trial 17 finished with value: 0.8206707478573811 and parameters: {'hidden_layer_sizes': (20,), 'activation': 'relu', 'alpha': 0.015790544718240304, 'learning_rate_init': 3.437326573879386e-05}. Best is trial 6 with value: 0.8768957743515713.

[I 2025-07-11 22:25:58,978] Trial 19 finished with value: 0.8756954182477095 and parameters: {'hidden_layer_sizes': (100,), 'activation': 'tanh', 'alpha': 0.0011119783431795649, 'learning_rate_init': 0.0014735594593995307}. Best is trial 6 with value: 0.8768957743515713.

[I 2025-07-11 22:26:07,137] Trial 13 finished with value: 0.8669098498967525 and parameters: {'hidden_layer_sizes': (60,), 'activation': 'relu', 'alpha': 0.004135824695773162, 'learning_rate_init': 0.006128903331742478}. Best is trial 6 with value: 0.8768957743515713.

[I 2025-07-11 22:26:19,122] Trial 3 finished with value: 0.8628405783870475 and parameters: {'hidden_layer_sizes': (20,), 'activation': 'tanh', 'alpha': 0.004599431354459305, 'learning_rate_init': 0.000321195743606703}. Best is trial 6 with value: 0.8768957743515713.

[I 2025-07-11 22:26:27,845] Trial 7 finished with value: 0.8693233347645855 and parameters: {'hidden_layer_sizes': (60,), 'activation': 'relu', 'alpha': 0.000265237161663165, 'learning_rate_init': 0.005422176201988498}. Best is trial 6 with value: 0.8768957743515713.

[I 2025-07-11 22:29:37,983] Trial 0 finished with value: 0.868633780613966 and parameters: {'hidden_layer_sizes': (80,), 'activation': 'relu', 'alpha': 0.0001654468298235879, 'learning_rate_init': 0.0020748533536024856}. Best is trial 6 with value: 0.8768957743515713.

[I 2025-07-11 22:30:48,242] Trial 2 finished with value: 0.8665693329868456 and parameters: {'hidden_layer_sizes': (80,), 'activation': 'relu', 'alpha': 0.015646070705683395, 'learning_rate_init': 0.0033438145170782644}. Best is trial 6 with value: 0.8768957743515713.

[I 2025-07-11 22:32:01,672] Trial 11 finished with value: 0.868080423479746 and parameters: {'hidden_layer_sizes': (80,), 'activation': 'relu', 'alpha': 4.8786210544784735e-05, 'learning_rate_init': 0.001222293142167707}. Best is trial 6 with value: 0.8768957743515713.

[I 2025-07-11 22:32:22,621] Trial 14 finished with value: 0.8527227170359367 and parameters: {'hidden_layer_sizes': (20,), 'activation': 'relu', 'alpha': 1.0821663272162994e-05, 'learning_rate_init': 0.00020160759244217433}. Best is trial 6 with value: 0.8768957743515713.

[I 2025-07-11 22:39:30,878] Trial 12 finished with value: 0.8731542548824678 and parameters: {'hidden_layer_sizes': (100,), 'activation': 'tanh', 'alpha': 4.8704183032592405e-05, 'learning_rate_init': 0.00044456471118881037}. Best is trial 6 with value: 0.8768957743515713.

[I 2025-07-11 22:41:44,268] Trial 28 finished with value: 0.8700554738099451 and parameters: {'hidden_layer_sizes': (80,), 'activation': 'tanh', 'alpha': 0.018118097691089435, 'learning_rate_init': 0.0066489654803751925}. Best is trial 6 with value: 0.8768957743515713.

[I 2025-07-11 22:47:58,755] Trial 5 finished with value: 0.8639302775404655 and parameters: {'hidden_layer_sizes': (80,), 'activation': 'relu', 'alpha': 0.01896428415859577, 'learning_rate_init': 0.0003445806513845424}. Best is trial 6 with value: 0.8768957743515713.

[I 2025-07-11 22:49:09,589] Trial 9 finished with value: 0.860916604021585 and parameters: {'hidden_layer_sizes': (60,), 'activation': 'relu', 'alpha': 0.01174936927742976, 'learning_rate_init': 0.00038494525819940774}. Best is trial 6 with value: 0.8768957743515713.

[I 2025-07-11 22:56:20,463] Trial 22 finished with value: 0.8456355069953438 and parameters: {'hidden_layer_sizes': (20,), 'activation': 'tanh', 'alpha': 1.4986587898012585e-05, 'learning_rate_init': 3.6047589058246255e-05}. Best is trial 6 with value: 0.8768957743515713.

[I 2025-07-11 22:56:26,284] Trial 16 finished with value: 0.8714516263964243 and parameters: {'hidden_layer_sizes': (40,), 'activation': 'tanh', 'alpha': 0.0013444305761346866, 'learning_rate_init': 0.00039634476333702464}. Best is trial 6 with value: 0.8768957743515713.

[I 2025-07-11 22:58:45,134] Trial 23 finished with value: 0.8644580785741696 and parameters: {'hidden_layer_sizes': (60,), 'activation': 'tanh', 'alpha': 0.06567598294521834, 'learning_rate_init': 0.0004123367563655606}. Best is trial 6 with value: 0.8768957743515713.

[I 2025-07-11 22:59:15,630] Trial 30 finished with value: 0.862640523505547 and parameters: {'hidden_layer_sizes': (100,), 'activation': 'tanh', 'alpha': 0.09527821343451813, 'learning_rate_init': 0.0007607041314228242}. Best is trial 6 with value: 0.8768957743515713.

[I 2025-07-11 23:02:14,102] Trial 32 finished with value: 0.8615636042594972 and parameters: {'hidden_layer_sizes': (100,), 'activation': 'tanh', 'alpha': 0.09426335063541845, 'learning_rate_init': 0.0005278713901553464}. Best is trial 6 with value: 0.8768957743515713.

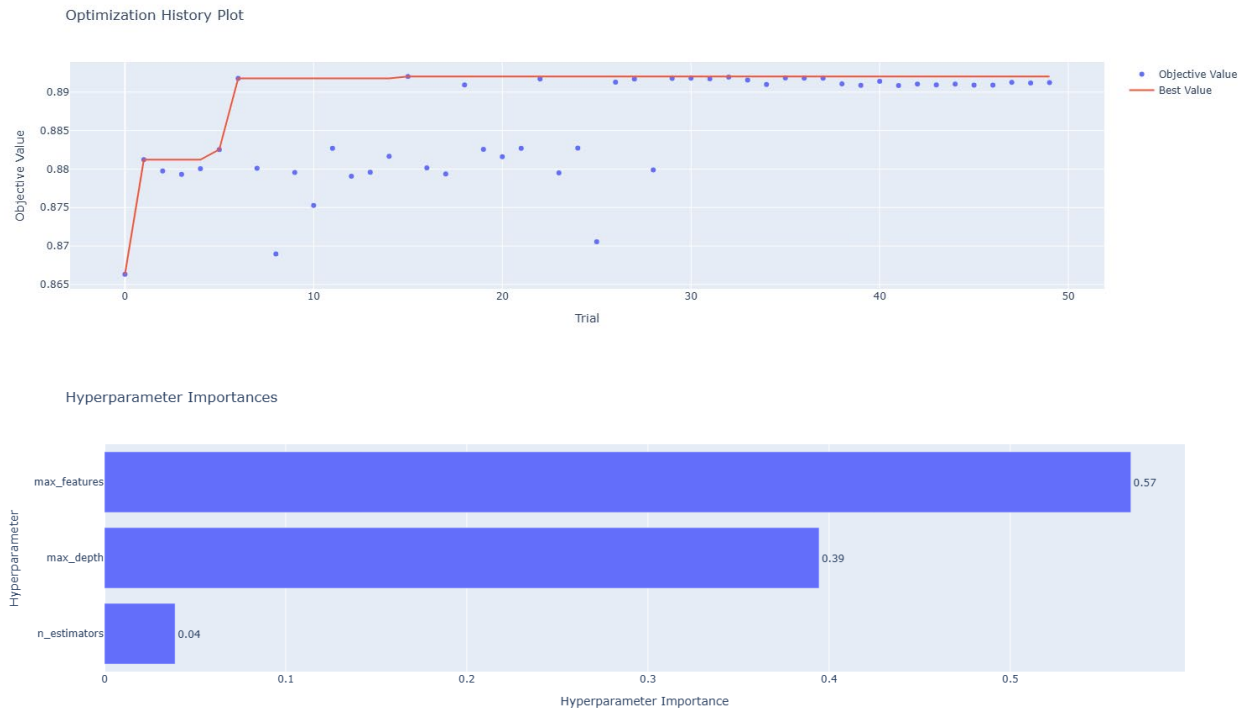
[I 2025-07-11 23:04:56,702] Trial 20 finished with value: 0.8707237439381332 and parameters: {'hidden_layer_sizes': (60,), 'activation': 'tanh', 'alpha': 0.026895066550211344, 'learning_rate_init': 0.00039581441000293463}. Best is trial 6 with value: 0.8768957743515713.

[I 2025-07-11 23:05:37,793] Trial 31 finished with value: 0.8702810540817012 and parameters: {'hidden_layer_sizes': (100,), 'activation': 'tanh', 'alpha': 0.037553152210125965, 'learning_rate_init': 0.0006173612531910717}. Best is trial 6 with value: 0.8768957743515713.

[I 2025-07-11 23:07:16,025] Trial 29 finished with value: 0.8631300274853843 and parameters: {'hidden_layer_sizes': (80,), 'activation': 'relu', 'alpha': 0.05532213075229182, 'learning_rate_init': 0.0003685552814278312}. Best is trial 6 with value: 0.8768957743515713.

```
[I 2025-07-11 23:08:06,802] Trial 35 finished with value: 0.8729243753554051 and parameters:
{'hidden_layer_sizes': (100,), 'activation': 'tanh', 'alpha': 0.0008336238608207406,
'learning_rate_init': 0.0006935995135914501}. Best is trial 6 with value: 0.8768957743515713.
[I 2025-07-11 23:09:15,107] Trial 27 finished with value: 0.8669311430688484 and parameters:
{'hidden_layer_sizes': (80,), 'activation': 'tanh', 'alpha': 0.009641744745622464,
'learning_rate_init': 0.0002603140912455823}. Best is trial 6 with value: 0.8768957743515713.
[I 2025-07-11 23:12:13,031] Trial 34 finished with value: 0.8696000079415466 and parameters:
{'hidden_layer_sizes': (100,), 'activation': 'tanh', 'alpha': 0.058934904079283076,
'learning_rate_init': 0.000606330248890843}. Best is trial 6 with value: 0.8768957743515713.
[I 2025-07-11 23:14:13,600] Trial 37 finished with value: 0.8742609509421696 and parameters:
{'hidden_layer_sizes': (100,), 'activation': 'tanh', 'alpha': 0.00060801927247444048,
'learning_rate_init': 0.0008218877882210362}. Best is trial 6 with value: 0.8768957743515713.
[I 2025-07-11 23:14:17,953] Trial 36 finished with value: 0.8669524134120781 and parameters:
{'hidden_layer_sizes': (100,), 'activation': 'tanh', 'alpha': 0.06299551422819517,
'learning_rate_init': 0.0006045549053252608}. Best is trial 6 with value: 0.8768957743515713.
[I 2025-07-11 23:15:00,905] Trial 33 finished with value: 0.8664629059897926 and parameters:
{'hidden_layer_sizes': (100,), 'activation': 'tanh', 'alpha': 0.06273278949369246,
'learning_rate_init': 0.00044657555411409695}. Best is trial 6 with value: 0.8768957743515713.
[I 2025-07-11 23:16:26,347] Trial 38 finished with value: 0.8644367830467143 and parameters:
{'hidden_layer_sizes': (100,), 'activation': 'tanh', 'alpha': 0.07684564287294818,
'learning_rate_init': 0.0007054677561043651}. Best is trial 6 with value: 0.8768957743515713.
[I 2025-07-11 23:16:54,186] Trial 40 finished with value: 0.8742524403951641 and parameters:
{'hidden_layer_sizes': (100,), 'activation': 'tanh', 'alpha': 0.0006085980827035198,
'learning_rate_init': 0.0009344453861158368}. Best is trial 6 with value: 0.8768957743515713.
[I 2025-07-11 23:17:08,884] Trial 41 finished with value: 0.8729371712067435 and parameters:
{'hidden_layer_sizes': (100,), 'activation': 'tanh', 'alpha': 0.0005351954556402016,
'learning_rate_init': 0.0009833869630812458}. Best is trial 6 with value: 0.8768957743515713.
[I 2025-07-11 23:17:18,970] Trial 42 finished with value: 0.8736650449683235 and parameters:
{'hidden_layer_sizes': (100,), 'activation': 'tanh', 'alpha': 0.00037819132541755813,
'learning_rate_init': 0.0009158649618040673}. Best is trial 6 with value: 0.8768957743515713.
[I 2025-07-11 23:17:49,972] Trial 24 finished with value: 0.8660585802243744 and parameters:
{'hidden_layer_sizes': (80,), 'activation': 'tanh', 'alpha': 0.019073177239430018,
'learning_rate_init': 0.00016205786012626997}. Best is trial 6 with value: 0.8768957743515713.
[I 2025-07-11 23:18:32,514] Trial 39 finished with value: 0.8678250233637368 and parameters:
{'hidden_layer_sizes': (100,), 'activation': 'tanh', 'alpha': 0.061662270255128125,
'learning_rate_init': 0.0009781240145700359}. Best is trial 6 with value: 0.8768957743515713.
[I 2025-07-11 23:19:04,692] Trial 10 finished with value: 0.8518501177739853 and parameters:
{'hidden_layer_sizes': (40,), 'activation': 'relu', 'alpha': 0.00012069622607046634,
'learning_rate_init': 7.777749046617649e-05}. Best is trial 6 with value: 0.8768957743515713.
[I 2025-07-11 23:20:02,906] Trial 26 finished with value: 0.8695787337935059 and parameters:
{'hidden_layer_sizes': (100,), 'activation': 'tanh', 'alpha': 0.0014092803912401177,
'learning_rate_init': 0.0001506863047790653}. Best is trial 6 with value: 0.8768957743515713.
[I 2025-07-11 23:20:10,337] Trial 21 finished with value: 0.859660943177275 and parameters:
{'hidden_layer_sizes': (60,), 'activation': 'tanh', 'alpha': 2.2832763798564483e-05,
'learning_rate_init': 7.292012970285639e-05}. Best is trial 6 with value: 0.8768957743515713.
[I 2025-07-11 23:20:21,475] Trial 25 finished with value: 0.8505986697614167 and parameters:
{'hidden_layer_sizes': (60,), 'activation': 'tanh', 'alpha': 6.354202404255563e-05,
'learning_rate_init': 5.8505802192418866e-05}. Best is trial 6 with value: 0.8768957743515713.
```

RF (תאים 151 ו-154):



[I 2025-07-12 10:57:14,288] Trial 12 finished with value: 0.87905810545371 and parameters: {'n_estimators': 106, 'max_depth': 42, 'max_features': 'log2'}. Best is trial 12 with value: 0.87905810545371.

[I 2025-07-12 10:57:33,942] Trial 1 finished with value: 0.8812076715053617 and parameters: {'n_estimators': 109, 'max_depth': 37, 'max_features': 'sqrt'}. Best is trial 1 with value: 0.8812076715053617.

[I 2025-07-12 10:57:44,498] Trial 0 finished with value: 0.866322447856788 and parameters: {'n_estimators': 120, 'max_depth': 10, 'max_features': 0.5}. Best is trial 1 with value: 0.8812076715053617.

[I 2025-07-12 11:00:23,331] Trial 3 finished with value: 0.879309244815477 and parameters: {'n_estimators': 124, 'max_depth': 45, 'max_features': 'log2'}. Best is trial 1 with value: 0.8812076715053617.

[I 2025-07-12 11:00:31,753] Trial 14 finished with value: 0.881667381368716 and parameters: {'n_estimators': 128, 'max_depth': 58, 'max_features': 'sqrt'}. Best is trial 14 with value: 0.881667381368716.

[I 2025-07-12 11:00:32,750] Trial 10 finished with value: 0.8752697589067253 and parameters: {'n_estimators': 142, 'max_depth': 16, 'max_features': 'sqrt'}. Best is trial 14 with value: 0.881667381368716.

[I 2025-07-12 11:00:35,585] Trial 13 finished with value: 0.8795688965360637 and parameters: {'n_estimators': 144, 'max_depth': 52, 'max_features': 'log2'}. Best is trial 14 with value: 0.881667381368716.

[I 2025-07-12 11:00:50,580] Trial 9 finished with value: 0.8795433563795176 and parameters: {'n_estimators': 147, 'max_depth': 54, 'max_features': 'log2'}. Best is trial 14 with value: 0.881667381368716.

[I 2025-07-12 11:00:57,957] Trial 18 finished with value: 0.8909254230934716 and parameters: {'n_estimators': 132, 'max_depth': 43, 'max_features': 0.5}. Best is trial 18 with value: 0.8909254230934716.

[I 2025-07-12 11:03:58,604] Trial 16 finished with value: 0.8801477861266171 and parameters: {'n_estimators': 176, 'max_depth': 31, 'max_features': 'log2'}. Best is trial 18 with value: 0.8909254230934716.

[I 2025-07-12 11:04:21,052] Trial 2 finished with value: 0.8797434133446054 and parameters: {'n_estimators': 194, 'max_depth': 37, 'max_features': 'log2'}. Best is trial 18 with value: 0.8909254230934716.

[I 2025-07-12 11:04:37,050] Trial 17 finished with value: 0.879360319240171 and parameters: {'n_estimators': 198, 'max_depth': 13, 'max_features': 0.5}. Best is trial 18 with value: 0.8909254230934716.

[I 2025-07-12 11:07:48,738] Trial 4 finished with value: 0.8800584004253105 and parameters: {'n_estimators': 205, 'max_depth': 40, 'max_features': 'log2'}. Best is trial 18 with value: 0.8909254230934716.

[I 2025-07-12 11:08:23,467] Trial 5 finished with value: 0.8825357212352859 and parameters: {'n_estimators': 225, 'max_depth': 26, 'max_features': 'sqrt'}. Best is trial 18 with value: 0.8909254230934716.

[I 2025-07-12 11:08:27,908] Trial 7 finished with value: 0.8801052236983805 and parameters: {'n_estimators': 213, 'max_depth': 30, 'max_features': 'log2'}. Best is trial 18 with value: 0.8909254230934716.

[I 2025-07-12 11:13:07,219] Trial 19 finished with value: 0.8825527456811543 and parameters: {'n_estimators': 250, 'max_depth': 51, 'max_features': 'sqrt'}. Best is trial 18 with value: 0.8909254230934716.

[I 2025-07-12 11:13:08,633] Trial 20 finished with value: 0.8815992764285738 and parameters: {'n_estimators': 147, 'max_depth': 40, 'max_features': 'sqrt'}. Best is trial 18 with value: 0.8909254230934716.

[I 2025-07-12 11:13:28,667] Trial 8 finished with value: 0.868961522055451 and parameters: {'n_estimators': 274, 'max_depth': 14, 'max_features': 'log2'}. Best is trial 18 with value: 0.8909254230934716.

[I 2025-07-12 11:13:48,182] Trial 6 finished with value: 0.8917809917493841 and parameters: {'n_estimators': 249, 'max_depth': 33, 'max_features': 0.5}. Best is trial 6 with value: 0.8917809917493841.

[I 2025-07-12 11:18:00,211] Trial 23 finished with value: 0.8795007915959218 and parameters: {'n_estimators': 146, 'max_depth': 46, 'max_features': 'log2'}. Best is trial 6 with value: 0.8917809917493841.

[I 2025-07-12 11:18:02,808] Trial 11 finished with value: 0.882688958550933 and parameters: {'n_estimators': 292, 'max_depth': 41, 'max_features': 'sqrt'}. Best is trial 6 with value: 0.8917809917493841.

[I 2025-07-12 11:18:36,169] Trial 26 finished with value: 0.8912914868636396 and parameters: {'n_estimators': 159, 'max_depth': 49, 'max_features': 0.5}. Best is trial 6 with value: 0.8917809917493841.

[I 2025-07-12 11:18:37,239] Trial 15 finished with value: 0.8920236173028787 and parameters: {'n_estimators': 291, 'max_depth': 31, 'max_features': 0.5}. Best is trial 15 with value: 0.8920236173028787.

[I 2025-07-12 11:19:03,560] Trial 27 finished with value: 0.891695860460968 and parameters: {'n_estimators': 159, 'max_depth': 35, 'max_features': 0.5}. Best is trial 15 with value: 0.8920236173028787.

[I 2025-07-12 11:23:49,785] Trial 25 finished with value: 0.8705492243212788 and parameters: {'n_estimators': 169, 'max_depth': 15, 'max_features': 'log2'}. Best is trial 15 with value: 0.8920236173028787.

[I 2025-07-12 11:24:35,876] Trial 28 finished with value: 0.8798753654601414 and parameters: {'n_estimators': 191, 'max_depth': 45, 'max_features': 'log2'}. Best is trial 15 with value: 0.8920236173028787.

[I 2025-07-12 11:26:27,589] Trial 22 finished with value: 0.8916916049609884 and parameters: {'n_estimators': 203, 'max_depth': 33, 'max_features': 0.5}. Best is trial 15 with value: 0.8920236173028787.

[I 2025-07-12 11:32:31,670] Trial 24 finished with value: 0.8827230082579863 and parameters: {'n_estimators': 240, 'max_depth': 43, 'max_features': 'sqrt'}. Best is trial 15 with value: 0.8920236173028787.

[I 2025-07-12 11:33:37,296] Trial 21 finished with value: 0.8826889570108902 and parameters: {'n_estimators': 280, 'max_depth': 41, 'max_features': 'sqrt'}. Best is trial 15 with value: 0.8920236173028787.

[I 2025-07-12 11:45:06,240] Trial 29 finished with value: 0.8917809922929285 and parameters: {'n_estimators': 254, 'max_depth': 31, 'max_features': 0.5}. Best is trial 15 with value: 0.8920236173028787.

[I 2025-07-12 11:45:37,792] Trial 30 finished with value: 0.8918150447177042 and parameters: {'n_estimators': 270, 'max_depth': 58, 'max_features': 0.5}. Best is trial 15 with value: 0.8920236173028787.

[I 2025-07-12 11:45:53,202] Trial 31 finished with value: 0.8917256569328105 and parameters: {'n_estimators': 248, 'max_depth': 59, 'max_features': 0.5}. Best is trial 15 with value: 0.8920236173028787.

[I 2025-07-12 11:50:55,318] Trial 33 finished with value: 0.8915511364100486 and parameters: {'n_estimators': 253, 'max_depth': 26, 'max_features': 0.5}. Best is trial 15 with value: 0.8920236173028787.

[I 2025-07-12 11:51:12,344] Trial 34 finished with value: 0.8909935240476212 and parameters: {'n_estimators': 263, 'max_depth': 23, 'max_features': 0.5}. Best is trial 15 with value: 0.8920236173028787.

[I 2025-07-12 11:51:13,470] Trial 41 finished with value: 0.8908402893591054 and parameters: {'n_estimators': 168, 'max_depth': 23, 'max_features': 0.5}. Best is trial 15 with value: 0.8920236173028787.

[I 2025-07-12 11:55:03,450] Trial 32 finished with value: 0.8919299704849664 and parameters: {'n_estimators': 292, 'max_depth': 57, 'max_features': 0.5}. Best is trial 15 with value: 0.8920236173028787.

[I 2025-07-12 11:55:49,652] Trial 35 finished with value: 0.8918278138447745 and parameters: {'n_estimators': 263, 'max_depth': 48, 'max_features': 0.5}. Best is trial 15 with value: 0.8920236173028787.

[I 2025-07-12 11:55:53,271] Trial 36 finished with value: 0.8918150430870708 and parameters: {'n_estimators': 278, 'max_depth': 50, 'max_features': 0.5}. Best is trial 15 with value: 0.8920236173028787.

[I 2025-07-12 11:56:08,799] Trial 37 finished with value: 0.8918065322682931 and parameters: {'n_estimators': 256, 'max_depth': 48, 'max_features': 0.5}. Best is trial 15 with value: 0.8920236173028787.

[I 2025-07-12 12:04:00,779] Trial 42 finished with value: 0.8910403482265986 and parameters: {'n_estimators': 252, 'max_depth': 22, 'max_features': 0.5}. Best is trial 15 with value: 0.8920236173028787.

[I 2025-07-12 12:04:05,804] Trial 43 finished with value: 0.8909381902275454 and parameters: {'n_estimators': 250, 'max_depth': 22, 'max_features': 0.5}. Best is trial 15 with value: 0.8920236173028787.

[I 2025-07-12 12:04:21,760] Trial 38 finished with value: 0.8910786575102149 and parameters: {'n_estimators': 285, 'max_depth': 22, 'max_features': 0.5}. Best is trial 15 with value: 0.8920236173028787.

[I 2025-07-12 12:05:59,255] Trial 39 finished with value: 0.8908871097332718 and parameters: {'n_estimators': 282, 'max_depth': 21, 'max_features': 0.5}. Best is trial 15 with value: 0.8920236173028787.

[I 2025-07-12 12:06:05,932] Trial 44 finished with value: 0.8910403482265986 and parameters: {'n_estimators': 252, 'max_depth': 22, 'max_features': 0.5}. Best is trial 15 with value: 0.8920236173028787.

[I 2025-07-12 12:06:06,615] Trial 40 finished with value: 0.8914021578556479 and parameters: {'n_estimators': 294, 'max_depth': 24, 'max_features': 0.5}. Best is trial 15 with value: 0.8920236173028787.

[I 2025-07-12 12:06:22,393] Trial 45 finished with value: 0.8909169058427512 and parameters: {'n_estimators': 254, 'max_depth': 23, 'max_features': 0.5}. Best is trial 15 with value: 0.8920236173028787.

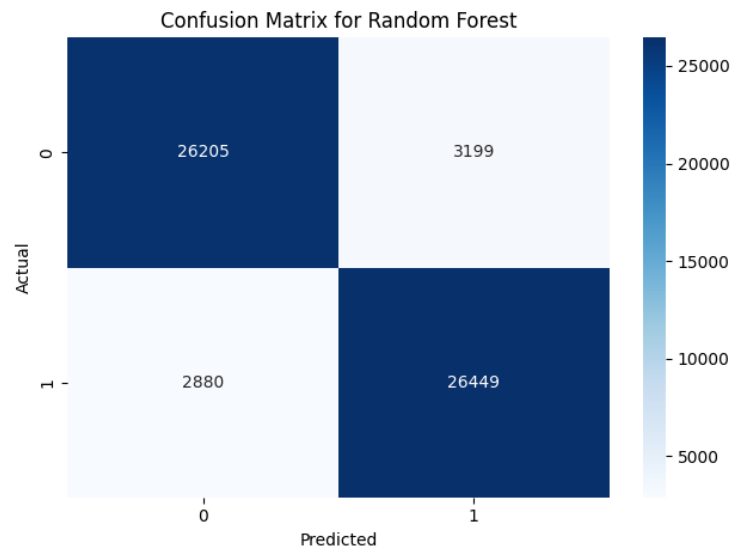
[I 2025-07-12 12:06:35,374] Trial 46 finished with value: 0.8909169058427512 and parameters: {'n_estimators': 254, 'max_depth': 23, 'max_features': 0.5}. Best is trial 15 with value: 0.8920236173028787.

[I 2025-07-12 12:07:18,164] Trial 47 finished with value: 0.8912574328082303 and parameters: {'n_estimators': 268, 'max_depth': 24, 'max_features': 0.5}. Best is trial 15 with value: 0.8920236173028787.

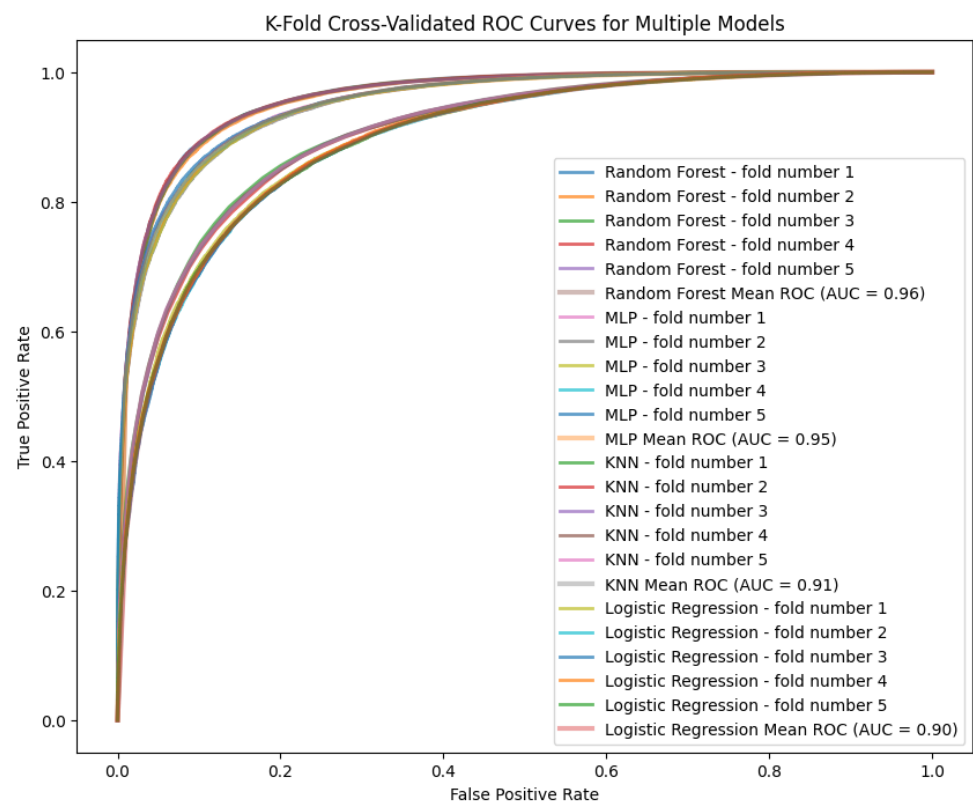
[I 2025-07-12 12:07:20,368] Trial 48 finished with value: 0.8911808149657234 and parameters: {'n_estimators': 263, 'max_depth': 24, 'max_features': 0.5}. Best is trial 15 with value: 0.8920236173028787.

[I 2025-07-12 12:07:54,100] Trial 49 finished with value: 0.891244663047025 and parameters: {'n_estimators': 266, 'max_depth': 24, 'max_features': 0.5}. Best is trial 15 with value: 0.8920236173028787.

נספח 4.1 מטריצת בלבול של מודל Random Forest (תא 162):



נספח 4.2: ROC של כל Fold ב K-Fold Cross Validation (תא 164):



4.3 פערים בין ה-train ל-validation (תא 164):

Random Forest:
 Train ROC AUC: 1.000 ± 0.000
 Val ROC AUC: 0.961 ± 0.001

MLP:

Train ROC AUC: 0.955 ± 0.001

Val ROC AUC: 0.951 ± 0.001

KNN:

Train ROC AUC: 0.918 ± 0.000

Val ROC AUC: 0.907 ± 0.001

Logistic Regression:

Train ROC AUC: 0.897 ± 0.000

Val ROC AUC: 0.897 ± 0.001