

## חלק ראשון

עבדתי עם דאטה סט על סיווג יינות - שיפרתי את המודל בחמישה אחוזים לאחר שהשתמשתי במודלים נוספים כגון Voting Classifier, Logistic Regression.

## חלק שני

בדאטה ה Fashion MNIST השתמשתי במודלים Decision Tree, Random Forest Classifier, Logistic Regression ולאחר מכן בAdaBoost וXGBoost, הדיוק המקסימלי שקיבלתי היה 89.6% עם XGBOOST, ניסיתי לשפר עוד באמצעות PCA – הוא הוריד לי את הדיוק בכמה אחוזים.

## חלק שלישי

בסיווג הכלבים והחתולים עשיתי ספליט לאימון וטסט והדיוק הכי גבוה שקיבלתי היה 65.4% עם XGBOOST, אחרי שהשתמשתי כבר ב K-means, AdaBoost, Random Forest Classifier, KNN, לשפר עוד באמצעות PCA, הופתעתי לגלות שהוא הוריד לי את הדיוק בהרבה, למרות שהexplained\_variance\_ratio היה יותר מ99% זהה לדאטה פריים המקורי.

## חלק רביעי

איחדתי את הדאטה סטס של קבצי האימון והולידציה (לפי ההוראות) והחלטתי להשתמש בכל הדאטה - השתמשתי במספר מודלים, הדיוק המקסימלי היה 73.7% עם Logistic Regression. אמנם לא כללתי זאת במחברת, אבל כשפיצלתי את הדאטה לאימון וטסט, הדיוק היה גבוה במיוחד, אבל בגלל שהדאטה כמעט זהה שם בגלל שאלו סרטונים של אותם אנשים בלי הבדל בתנועתם והתנהגותם, הבנתי מדוע יש קובץ נפרד של ולידציה. חוץ מXGBoost, השתמשתי ב Decision Tree, Random Forest Classifier, Logistic Regression, AdaBoost כשניסיתי לשפר עוד באמצעות PCA, הוא הוריד לי את הדיוק בכמה אחוזים, למרות שה explained\_variance\_ratio היה יותר מ99% זהה לדאטה פריים המקורי.