סיכום עבודה גדולה אלי

בגדול הדברים שעשינו בעבודה כי:

scaling - standard scaler

PCA - רצינו לצמצם את המידע אבל שמרנו על 95% מהמידע ועדיין לקחו המון זמן בגלל שכבה

grid search - מצאנו את הפרמטרים הכי טובים.

pipeline - scale, pCA, model.

models - knn, logistic regression.

ensemble models - random forest, xgboost, adaboost, bagging, voting, stacking

clustering - kmeans

מטלה 1 — שימוש באלגוריתם קרב

בקוביץ) נקודת עבורי של פאקץ שמהן ג'ני אב נוון את זווית
בכ'ס ש'אלמש – פאקג'ר נירצת אינו מיוני. להרי נרון ג 93%.
נתר, יקלה גרצתי לרכז עם קוביץ ואבר זמען ולקבל, שטלי
אחר נן בעזרת ensemble methodes כך ג'אגנו זמען לקבצר
ונטלס נ את עם Kmeans – ג אב ג'אגנו ורצתי, שטר את נירא בעבץ) 3%.

מטלה 2- fashion mnist

בעזרתי ב PCA ו'כי קיבצתי)-50 מימים ומ בזי feature importance.
אחר נן בעזרתי ג mnist ו knn ו ensemble אב בכרגית של נוורים
כך נבצתי ברצתי מטרי gridsearch. קיבל בעזרתי נאומ ברצולום
אל ג ה testing ונגרת) 86.61% בערצ מטרתי stacking.

מחברת 3 - dogs vs cats

בתחילה טעינו את התמונות (שאותם הורדנו df) אבל לקח יותר מדי וטענו גם בגרייסקייל וגם reashape-ו grayscale.
ואחרי כן טענו את התמונות בגרייסקייל, בגלל שלוקח יותר מדי זמן.
בהמשך ב PCA לניתוח) - 50 שימוש, ולאחר מכן רצנו על ה train
והרצנו כל המודלים על כל הנתונים, בקטגורי ואז את המודל עם אחוז הכי גבוה ואהרצנו
C בהרצת voting 64.x6%.

מחברת 4 - יער..

לאחר יותר בהתחלה בהרצת הקוד את הנתונים עם pandas, ואחרי הורדנו גם את
ה train התמונות לדייסט, כי רצנו ובגלל ראינו אם הזמן רץ יותר מאם,
לא ה אלא שינוי קוד (את רובי ובהרצה לאחר קוד לדייסט).
בהמשך ב PCA לניתוח) - 15 שימוש, ולאחר מכן רצנו על ה train
והרצנו כל המודלים על כל הנתונים, בקטגורי ואז את המודל עם אחוז הכי גבוה ואהרצנו) 88.4%
בהרצת Xgboost ולאחר כל הנתונים עם אחוז הכי גבוה על כל הנתונים לדייסט.