**עבודת גמר מדעי הנתונים** - שירה עוזרי, ת"ז: 318901428

**המודלים בהם נעשה שימוש:**

Knn

## Decision tree

## AdaBoost Classifier

## Pipeline

## Voting

### Random forest

### GradientBoost

## Stacking

**פירוט המחברות:**

מחברת 1- שימוש במודלים שונים על מנת לשפר את אחוז הדיוק מסמסטר הקודם.

ניסיתי להפעיל PCA על מנת להפוך את המודל ליעיל יותר אך ראיתי שלמרות שהורדתי מעט פיצ'רים הדיוק ירד משמעותית ולכן העדפתי להישאר ללא PCA במקרה זה.

*במחברות הבאות הדאטה חולקה לtrain וולידציה שבהם השתמשתי לצורך האימונים ולטסט שיועד להרצה האחרונה עבור המודל הנבחר.*

מחברת 2 FMnist- השתמשתי במודלים שונים והראיתי עבור כל מודל את אחוז הדיוק שלו לפני PCA ואחריו. בהתאם לתוצאות ראיתי איזה מודל עדיף (לפני או אחרי PCA , שיקולי אחוז דיוק, וקומפקטיות) . לאחר מכן אספתי את דיוקי המודלים, השוויתי ביניהם ובחרתי את המודל בעל שילוב היעילות ודיוק המוצלח ביותר.

ניסיתי להוריד עוד ממדים ולראות איך זה משפיע על הדיוק. אחוז הדיוק ירד מעט, אך העדפתי להשאיר את המצב לטובת היעילות והקומפקטיות.

לבסוף השתמשתי בtest כדי לבחון את דיוק המודל.

מחברת 3 dogs VS cats- המחשתי בצורה ויזואלית את משמעות הפחתת ממדים. השתמשתי במודלים שונים והראיתי את אחוז הדיוק שלהם לאחר PCA , בחרתי את המודל המוצלח ביותר ובעזרת הtest בחנתי את דיוקו.

מחברת 4 hands – במחברת זו ייבאתי את הקבצים, אותם הכנסתי לרשימות. נעשתה חלוקה לtrain, validation וtest. ניקיתי את הדאטה מNaN . איחדתי את קובץ righthand עם הalone איחדתי כל 3 שורות והתחלתי כל הקלטה משניה 7 כמבוקש. ניקיתי עמודות מיותרות, והוספתי עמודות רלוונטיות. איחדתי את 3 סוגי הדאטות של התנועות עבור הtrain, validation ו test ולמעשה יצרתי 3 דאטה פריים של train, validation, ו test . חילקתי את הtrain וvalidation לtrain ו test כהכנה לPCA