

# עבודת סוף בהדמיית נתונים

## מחברת קלסיפיקציה(סמסטר קודם):

במחברת הזאת השתמשתי בדברים שלמדתי הסמסטר כלומר מסווגים חדשים כגון:

Adaboost (with Decision Tree Classifier), Voting classifier, Bagging and XGBoost

(העברית אנגלית בוורד מסתבך לי , מתבייש להיות שנה ג' במדמח) בעקרון כל המסווגים חוץ ממסווג ההצבעה לא שיפרו לי את המודל , וגם המודל ההצבעה שיפר לי ממש באחוזים מעטים , זה מקובל כי המודל שלי כבר היה 94% בעזרת מסווג השכנים

## FMNIST:

במודל זה השתמשתי רבות בגריד סרץ' ובראנדום סרץ' כדי למצוא את הערכים המתאימים ולאחר בדיקה רבה המסווגים שהשתמשתי בהם הם:

KNeighbors, Random Forest, XGboost, AdaBoost, Voting classifier, Stacking Classifier

בעוד שניסיתי להשתמש ב PIPELINE KMEANS ב מודל , הוא הוריד לי את הדיוק בערך כ-20%

לכן החלטתי לוותר עליו , בסופו של דבר המודל שדייק ברמה הגבוהה ביותר היה ה VOTING

עם אחוז דיוק של 68%

## Dogs Vs Cats:

במודל הזה השתמשתי יותר לרוב ב random search כדי למצוא את הערכים המתאימים לכל

חשוב לי גם להוסיף כי בהרצה הראשונה שלי המודל לא הצליח לסווג בכלל כלבים ולא הבנתי למה עד שהגעתי להבנה שהדאטה שהזנתי לו היה חצי ראשון כולו חתולים וחצי שני כולו כלבים לכן הוא התחיל לחשוב שכל תמונה היא בעצם חתול ,

השתמשתי לראשונה בצמצום של כל התמונות לגודל 90 על 90 של שסיכם לי את כל הדאטה בגודל של 8100 וב GRAYSCALE אך לאחר בדיקות רבות החלטתי להשאיר את שלושת הצבעים אך לשנות את הגודל ל 34 על 34 ולכן הגודל היה 3468 והעלה לי את הדיוק בהרבה וכך סיימתי את המודל עם דיוק של 87.3%

בין היתר המסווגים שהשתמשתי בהם הם:

KNeighbors, AdaBoost, Xgboost, Voting(winner), Stacking, RandomForest, PipeLine(with kmean) and dummy classifier

## Hand movements:

מה שעשיתי במודל הזה זה להתחיל מלהעלות כל קובץ , להעיף את כל האוטלינארים כגון למשל , שבקובץ HandRight היו ידיים שמאליות , או בקובץ של לבד היו ידיים ימניות , או נגיד מספר הידיים היה 1 בחלק מהקבצים כאשר לא היה צריך להיות ככה וכו'.

לאחר מכן צמצמתי כל קובץ ככה שבכל שורה יהיה את יד ימין ויד שמאל ביחד ובעזרת groupby צימצמתי כל 5 פריימים לשורה אחת ארוכה ככה שלכל דאטהפריים היו 200 פ'צרים וכן היה לי יותר קל לעבוד , כמו בשאר העבודה (למרות שלא ציינתי) השתמשתי בפונקציות והדמיות של הנתונים כדי לשפר את המודל עוד יותר ולהבין יותר טוב איך הנתונים עובדים (אך חלק מהפונקציות לא השארתי לתוצר הסופי)

במודל זה השתמשתי במסווגים :

XGboost, LogisticRegression, KNeighbors, Adaboost, Random  
forest, ExtraTrees, Stacking, Voting, Bagging

בעוד שבכל העבודה לא השתמשתי ב pasting מכיוון שלפי הספר אומרים שלרוב הוא פחות טוב מ bagging

המנצח במחברת זו היה ה stacking classifier עם 89.8% דיוק

נק' צד: חשוב לי להוסיף (וזה לא התחנפות אני מבטיח) שנהנתי מאוד מהעבודה והעבודה עם העיבוד תמונה והבנה של הדאטה , ותודה רועי על 2 הקורסים האחרונים!