

סיכום העבודה

Gender Classification

במחברת הראשונה הייתי צריכה לסווג בין זכר לנקבה- לצורך זה השתמשתי ב **Logistic** | **KNN** | **Regression**. בפרויקט הנוכחי ניסיתי ליצור שיפור באחוז דיוק באמצעות **Ensamble Learning**:

.Random Forest, Voting, Bagging, Decision Tree, Stacking, Adaboost, XGBoost

כמובן שגם השתמשתי ב GridSearch על מנת למצוא את הפרמטרים הטובים ביותר. לא הצלחתי לשפר את רמת הדיוק מ KNN שעמדה על כ-96.0%.

Fashion MNIST

מהלך העבודה במחברת זו היה ייבוא הדאטה מקאגל בצורה המוכרת והידועה, ניתן היה לראות כי

כמות הפיקסלים הייתה 784 לכל תמונה ולכן החלטתי באמצעות PCA להוריד את כמות המימדים ל-16. עשיתי split לדאטה ומשם עבדתי עם מודלים שונים על מנת להגיע לאחוזי דיוק ולשפרם.

ניתן היה לראות כי עבור המודל Random Forest היה את האחוז דיוק הגבוהה ביותר כאשר אימנתי על הדאטה המקורי (ללא הורדת מימדים) – 92% אחוז דיוק, אבל כמובן שזמן ההרצה לקח המון זמן ולכן זה פחות יעיל.

השתמשתי במודלים שונים כמו: KNN, Decision Tree כמו כן השתמשתי במודלים שונים מתוך Ensamble Learning כמו: Gradient Boosting, Adaboost, XGBoost, Stacking.

Cat VS Dog

כאן הדאטה היה שונה מאחר וקיבלנו תיקייה ובה 12,500 תמונות של חתולים ו-12,500 תמונות של כלבין שונים והמטרה הייתה כמובן לסווג בין חתול לכלב באמצעות למידת מכונה. אז על מנת להפוך את התמונות האלה לדאטה קונקרטי לעבודה השתמשתי בספריות cv2, osi, תחילה הפכתי את כל התמונות לגודל אחיד והשתמשתי ב gray scale לצורך הפיכת התמונות לגוונים של שחור לבן ובכך גם הורדתי משמעותית מימדים. יצרתי דאטה של כלבים וחתולים יחד מתוך תקיית "train" ולאחר נירמול הדאטה באמצעות MinMaxScaler הורדתי ל-16 מימדים וגם זה באמצעות PCA.

מכאן המשכתי לעבודת מודלים והשתמשתי במודלים הבאים: KNN, Logistic Regression, Random Forest, K-Means, Gradient Boosting, Adaboost, XGBoost, Stacking, bagging

קיבלתי את האחוז הדיוק הגבוהה ביותר באמצעות Bagging: אחוז דיוק של 64.80% וגם זה על הדאטה המקורי.