

הצעת פרויקט בלמידת מכונה

שמות המגישים: איתי מייזליש ואלי קוזינץ

04.06.2025

שם הפרויקט

סיווג תמונות לפי מדינות באמצעות למידת מכונה <https://huggingface.co/datasets/nateraw/country211>

מאגר הנתונים

הפרויקט יתבסס על מאגר הנתונים **Country211** הזמין דרך HuggingFace: המאגר מכיל מעל 30,000 תמונות גאוטגיות, אשר סווגו ל-211 מדינות שונות לפי קוד ISO.

שאלות המחקר

- האם ניתן לאמן מודל למידה חישובית שיסווג תמונות חדשות לפי המדינה בה צולמו?
- אילו תכונות בתמונה משפיעות יותר על זיהוי המדינה (נוף, סגנון אדריכלי, צבעוניות)?
- כיצד משפיעה חלוקת הדוגמאות הלא-מאוזנת על ביצועי המודל?

כלים בהם נשתמש

לצורך הפרויקט נשתמש בארבעה כלים לפחות, מתוכם שלושה שאינם רשתות נוירונים:

1. SVM – למודל סיווג מבוסס גבולות החלטה.
2. k-NN – לבדיקת דמיון חזותי בין תמונות לפי קרבה וקטורית.
3. PCA – לצמצום ממדים והדמיה תלת-ממדית של ההתפלגות לפי מדינות.

4. Forest Random או XGBoost – כשיטות מבוססות עצים להשוואה.

5. (אופציונלי) ResNet50 – לשם השוואה למודל רשת נוירונים עמוק.

פלט צפוי

-- אחוזי הצלחה של כל שיטה והשוואה ביניהם.

-- טבלת בלבול שתמחיש באילו מדינות נרשמות טעויות סיווג.

-- הדמיה של קבוצות המדינות לאחר צמצום ממדים.

-- ממשק בסיסי לבחירת תמונה וחיזוי מדינה.