**פרויקט גמר: Kepler vision**

**תהליך העבודה – בניית מודל למידה עמוקה**

מגישים: דולב הינדי 312126642

ערד זקלר 305600579

נאור דהן 308399393

בשלב הראשון (Data Extraction) היינו צריכים להתמודד עם מסד נתונים גולמי שמגיע מNASA ולברור מתוכו את הפיצ'רים שיעזרו לנו בסיווג או להוריד את הפיצ'רים שעלולים לפגוע לנו בסיווג. מבין יותר מ-120 פיצ'רים נבחרו בסביבות ה-70 (אחרי קריאה מעמיקה על כל פיצ'ר והתועלת/רמת הדיוק שלו).

אחר כך היינו צריכים לנרמל את כל הדאטה (Data Augmentation): העברת נתונים טקסטואליים למספריים (למשל, בעזרת ONE-HOT-ENCONDING), העברת כל הנתונים לאותו הטווח וטיפול בנתונים חסרים. אחרי שהדאטה הוכן כראוי הוא הוטען לעיבוד מקביל בTensorflow- שם המודל יעבוד עליו (ה-Dataset הראשוני מכיל כ-7000 תצפיות).

במהלך בניית המודל למדנו המון על הגדרת גלקסיות ועל מושגים כמו Goldilock zone שזהו השטח שבו יכול להימצא כוכב לכת שיש בו חיים. במהלך הפרויקט ייעדנו לאפשר חקר מעמיק יותר של מודלים באמצעות מושגים אלו.

למידה עמוקה ממשיכה להתמודד עם בעיות באמצעות רשתות עצביות עמוקות, לכן נורמליזציה של כמויות גדולות של נתונים והזנתם לרשת עצבית דורשת כמויות גדולות של זיכרון ועוצמת עיבוד ולכן חזינו שבעיה זו תהווה עבורנו גורם מגביל במהלך פיתוח הפרויקט, אך הצלחנו להגיע לסיווג אידאלי.

בניית המודל עצמו קרתה אחרי ניסוי, טעיה וקריאה על ארכיטקטורות שונות. אחרי הרבה ניסויים הוחלט על מודל DNN (Deep Neural Network) הכולל מאות נוירונים שנבנה בעזרת Sci-kit וTensorflow-. המודל בתחילתו עמד על 21% דיוק וכרגע הוא עומד על קרוב ל-97% בסיווג פלנטות.

דבר נוסף שעבדנו עליו זו מערכת GUI למשתמש על מנת להריץ את המודל ולקבל אחוזי דיוק של סיווג פלנטות.

את ה GUI-בנינו בשרת על ידי Flask python.

מטרתו העיקרית של הפרויקט היא להוות חלק משמעותי בעבודתם של אסטרונאוטים וחוקרי חלל. מודל זה יהיה נגיש לכל אדם שמתעניין בסיווג פלנטות ולא רק לאנשי מקצוע בתחום.

קישור לגיטהב שלנו:

https://github.com/naordahan/Final-Project.git