

דו"ח הפרויקט

## תיאור מסד הנתונים

מסד הנתונים מכיל מידע על עצמים אסטרונומיים. כל עצם מאופיין על ידי תכונות, הכוללות:

- קואורדינטות (אלפא, דלתא)
- בהירות בטווחי ספקטרום שונים (u, g, r, i, z)
- הסטה אדומה (Redshift)
- סוג העצם (גלקסיה, קואזר, או כוכב)

## שאלות על מסד הנתונים

- האם נוכל לסווג במדויק עצמים אסטרונומיים על בסיס המאפיינים הנצפים שלהם?
- אילו תכונות הן החשובות ביותר לסיווג?
- האם קיימות קבוצות טבעיות בנתונים מעבר לסוגים הידועים?
- האם ניתן לגלות סוגים חדשים של עצמים שלא היו ידועים קודם לכן?

## תקציר האלגוריתמים שנעשה בהם שימוש

- Random Forest
- CatBoost
- מכונת וקטורים תומכים (SVM)
- פרספטרון רב-שכבתי (MLP)
- רשת עצבית עמוקה
- K-means לניתוח אשכולות
- שיטה ניסיונית של רשתות עצביות KAN

## תוצאות

- Random Forest הראה את הדיוק הטוב ביותר: 97.81%
- CatBoost: 97.77%
- רשת עצבית: 96.62%
- SVM: 95.93%
- התכונות החשובות ביותר: הסטה אדומה, בהירות בטווחי g-z
- ניתוח האשכולות אישר את קיומן של שלוש קבוצות עיקריות התואמות לסוגים הידועים

## בעיות ודרכי התמודדות

- בחירת מספר אופטימלי של אשכולות: נפתר באמצעות שיטת המרפק וניתוח סילואט

- פרשנות תוצאות האשכול: התגברנו על ידי ניתוח ערכים ממוצעים של תכונות בכל אשכול
- יעילות מוגבלת של שיטת KAN הניסיונית: נדרש מחקר נוסף עם ארכיטקטורות מורכבות יותר

### גישות מוצלחות ופחות מוצלחות

- מוצלחות: שיטות אנסמבל (Random Forest, CatBoost) בזכות היכולת לתפוס תלויות לא ליניאריות ועמידות לרעש
- פחות מוצלחות SVM כנראה בגלל הקושי בכיוון היפר-פרמטרים למשימת סיווג מרובת מחלקות

### ניתוח מסד הנתונים, כלים ותוצאות

- מסד הנתונים התאים היטב למשימת הסיווג, עם הפרדה ברורה של הסוגים לפי מאפיינים מרכזיים
- הכלים שנעשה בהם שימוש (Python, scikit-learn, CatBoost, Tensorflow) הראו יעילות גבוהה לניתוח ומידול
- התוצאות אישרו את ההשערה לגבי התפקיד המרכזי של ההסטה האדומה בסיווג עצמים אסטרונומיים
- הדיוק הגבוה בסיווג (עד 97.81%) מדגים את הפוטנציאל של שיטות למידת מכונה באסטרונומיה

### תצפיות נוספות

- חשיבות ההסטה האדומה מוסברת בהבדלים במרחקים לסוגים שונים של עצמים: קוואזרים הם הרחוקים ביותר, אחריהם גלקסיות, כוכבים בעיקר מהגלקסיה שלנו
- הניסיון לגלות אשכולות נוספים לא נתן תוצאות ברורות, אך נשאר כיוון מבטיח למחקר נוסף
- שיטת KAN הניסיונית הראתה פוטנציאל, אך דורשת פיתוח ובדיקה נוספים עם ארכיטקטורות מורכבות יותר