

Machine Learning Final Project

בפרויקט זה, world-cup 2022, אחוז הדיוק של המודל הוא כ- 70%.

בנינו מודל הרצנו אותו על כ- 80% מהמשחקים ולאחר מכן על כ-20% מהמשחקים בדקנו את הסטטיסטיקות שקיבלנו (testing), למדנו ובדקנו על data-set של כ- 20,000 נתונים ולבסוף קיבלנו דיוק של כ- 70%.

Accuracy of model: 0.6817136886102403

שאלה 1:

מה אחוז הדיוק שקבוצת הבית תנצח?

Percentage of games won by the host team: 49.57%

כלומר המודל יחזה שקבוצת הבית במאגר הנתונים תנצח, עם דיוק של כמעט 50%.

שאלה 2:

מה אחוז הדיוק שקבוצת החוץ תנצח?

Percentage of games won by the away team: 50.43%

כלומר המודל יחזה שקבוצת החוץ במאגר הנתונים תנצח, עם דיוק של יותר מ- 50%.

שאלה 3:

בכמה אחוזים FIFA תדייק? כלומר, קבוצות שמדורגות במקום הגבוה ביותר (דירוג נמוך יותר), בכמה אחוזים של דיוק הם ינצחו?

Percentage of times FIFA was correct in the test data: 67.94%

כלומר דירוג FIFA מקבל כ- 68% דיוק, זאת אומרת שבכמעט 70% דיוק, קבוצות עם דירוג נמוך יותר (קבוצות שממוקמות במקום גבוה יותר) – ינצחו.

שאלה 4:

בכמה אחוזים FIFA תטעה? כלומר, "קבוצות שהן פחות טובות- ינצחו"

Percentage of times FIFA was incorrect in the test data: 32.06%

כלומר כ- 32% מהמקרים, קבוצות שממוקמות במקום נמוך יותר ינצחו קבוצות שממוקמות מעליהן בטבלת הדירוג של FIFA. זאת אומרת FIFA תטעה בזהות המנצחת ביחס לדירוג ב- 32% מהמקרים (!)

שימוש במודלים:

בפרויקט זה השתמשנו במודל של רגרסיה לוגיסטית ובמודל של קלסיפיקציה בינארית.

binary classification - הסיווג מתבצע באמצעות רגרסיה לוגיסטית, כאשר אנו חוזים אם הקבוצה המארחת מנצחת או לא בהתבסס הדירוג FIFA של הקבוצות.

במהלך הפרויקט השתמשנו ב:

- מודל של Random Forest, אך קיבלנו דיוק נמוך יותר ולכן שינינו מודל.

ניתן לראות שאחוז הדיוק כאן הוא: 60.8% ולכן לא השתמשנו במודל זה על מנת לקבל אחוז דיוק גבוה יותר.

```
/Users/yoellchemla/Downloads/pythonProject1/venv/bin/python /Users/  
Accuracy of model: 0.6087774294670847  
Percentage of games won by the host team: 49.57%  
Percentage of games won by the away team: 50.43%  
The team with the lower FIFA rank won 3251 times in the test data.  
Percentage of times FIFA was correct in the test data: 67.94%  
Percentage of times FIFA was incorrect in the test data: 32.06%
```

- באמצעות שימוש במודל linear regression קיבלנו את אותה תוצאה של דיוק המודל, כלומר 68%.

```
Accuracy of model in linear regression: 0.6895110739657334
```

