



## חיזוי תוצאות משחקי כדורגל (ניצחון/הפסד/תיקו)

### פרויקט בבינה מלאכותית - 236502

שם מנחה: סימון דרוד

<u>מייל</u>	<u>תעודת זהות</u>	<u>שם פרטי</u>
firasbathish@campus.technion.ac.il	318452208	פיראס בטחיש
mahmoud-nas@campus.technion.ac.il	318489200	מחמוד נסאר



## תיאור הבעיה :

### חיזוי תוצאות משחקי כדורגל בליגה הספרדית (ניצחון/הפסד/תיקו).

הערה : בכל פעם שמדברים על תוצאה אנחנו מדברים על (ניצחון/הפסד/תיקו) ולא מתעניינים בתוצאה המספרית.

## תיאור הדרך פתירת הבעיה:

הבעיה אנו מנסים לפתור היא בעיית חיזוי, נשתמש באלגוריתמי למידה וחיזוי שלמדנו בקורס מבוא לבינה מלאכותית.

תחילה נזכיר את תהליך הלמידה והחלקים השונים שמרכיבים אותו:

### 1. איסוף המידע:

תהליך זה כולל איסוף מידע רב עבור משחקי כדורגל של הליגה הספרדית, מידע זה יכלול תוצאת המשחקים, ומתבסס על קריטריונים שונים פר קבוצה וגם פר שחקנים, כולל שחקני מפתח, שחקני הרכב ראשי, מיקום בטבלת הליגה, מספר אוהדים לכל קבוצה, וכו' ..

### 2. בחירת התכונות:

בחירת התכונות הינו אחד האתגרים בפרויקט זה, בחרנו תכונות רבות שיכסו כמעט את כל האספקטים והקריטריונים שמשפיעים על תוצאת המשחק, מבניהם: לצורך נוחות אנו נסמן עבור קבוצת כדורגל ב  $A$  ועבור משחק מסוים ב  $G$ :

- ערך מסחרי של קבוצה  $A$  בשוק.
- מספר הצופים של קבוצה  $A$  במשחק  $G$ .
- מספר השחקנים המכריעים (חשובים) שפצועים בקבוצה  $A$  לקראת משחק  $G$ .
- תוצאות של 5 המשחקים האחרונים שנפגשים בהם שתי הקבוצות שמשחקות במשחק  $G$ .
- תוצאות של חמשת המשחקים האחרונים של קבוצה  $A$ .
- מצבם הגופני של שחקני הקבוצה  $A$  (עייפים/ לא כל כך בכושר).
- האם המשחק  $G$  הוא משחק בית או חוץ של קבוצה  $A$ .
- האם תוצאות הקבוצה  $A$  במשחקי בית יותר טובות ממשחקי חוץ או להפך.



- מיקום של קבוצה  $A$  בטבלת הליגה.
- מספר תארי הליגה שיש לקבוצה  $A$ .
- מספר תארי הליגות האירופיות של קבוצה  $A$  (ליגת האלופות / הליגה האירופית).
- המוטיבציה של שחקני קבוצה  $A$  לקראת המשחק. (תלוי בהיסטוריה של משחקי עבר וגם חשיבות המשחק עצמו).

### 3. שלב הלימוד:

אנו נשתמש במספר של אלגוריתמי למידה, ועבור כל אלגוריתם למידה אנו ניקח את הגרסה הטובה ביותר שלו (שנקבל בעזרת ניסויים, למשל כיוון פרמטרים וכו') כהתחלה נשתמש באלגוריתמים הבאים:

- KNN
- Decision Tree
- Radom forest

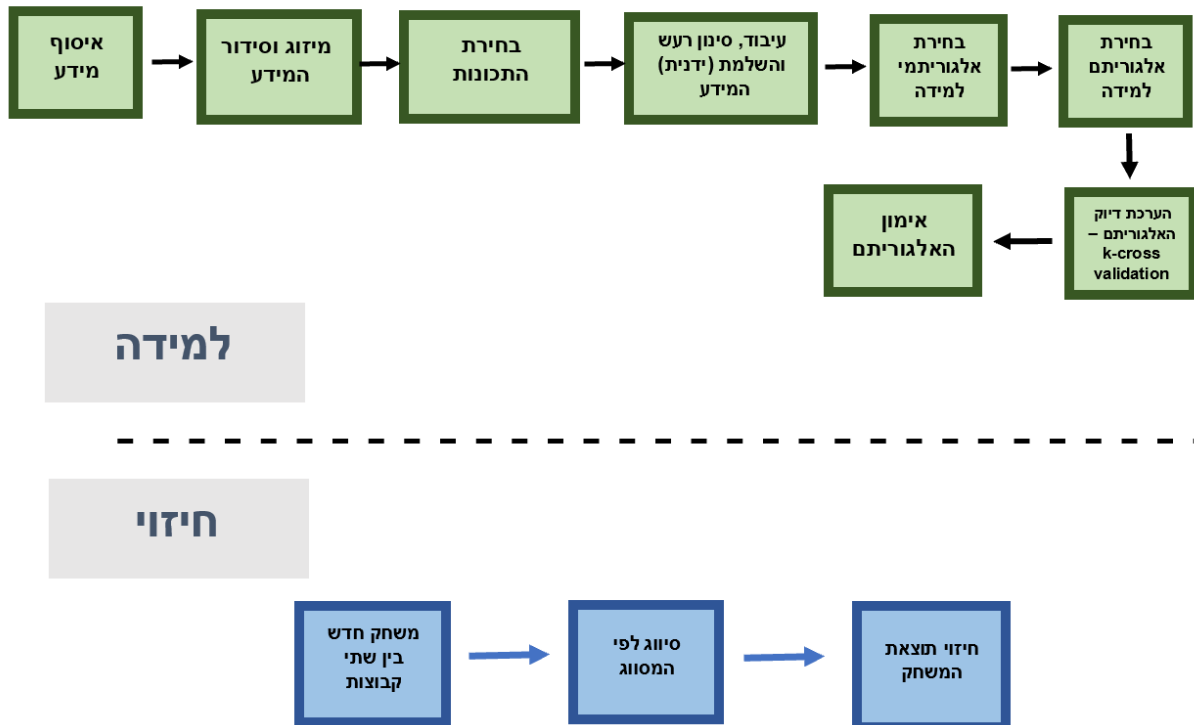
וייתכן שינויים ולהיעזר באלגוריתמים נוספים.

כדי לאמן את המסווגים הנ"ל, אנו נשתמש בשיטת cross-validation על מאגר המידע שאספנו כלומר משחקי עבר בליגה הספרדית – עונות 2017, 2018. שיטה זו מניבה הערכה על טיב המסווג, ולבחון את יכולת המודל לחיזוי נתונים שלא השתמשנו בהם קודם, שיטה זו עוזרת להתמודד עם בעיות כמו התאמת יותר, selection bias וכו' ..

בשיטה זו אנו נחלק את dataset למספר קבוצות נסמן את מספר הקבוצות ב- $X$ , ערכו של  $X$  נקבע בזמן הלמידה לפי הניסויים השונים, ב- $X$  איטרציות נבחר בכל פעם  $X-1$  קבוצות מתוך  $X$  קבוצות לאימון המסווג, וכאשר הקבוצה הנותרת נשתמש בה בקבוצת המבחן, מודדים את הצלחת האלגוריתם כממוצע של  $X$  התוצאות מכל איטרציה.



## תיאור כללי של דרך הפתרון:



## תיאור הניסויים מתכננים לבצע לשם הערכת הפתרון המוצע:

הניסויים שזכנו עולים לנו בראש חלק למדנו אותם בקורס מבוא לבינה מלאכותית, כגון כוונן פרמטרים, ובדיקת התלות שלהם בדיוק המסווג עבור כל אלגוריתם מוצע. בדיקת רלוונטיות של תכונה (במהלך הלמידה ובניית המסווג) כפונקציה של הדיוק, עבור כל אלגוריתם שהצענו מקודם.

## פרטים טכניים:

שפת תכנות: Python

סביבת עבודה: Pycharm (JetBrains), Git

פריית רלוונטיות: pandas, sklearn, random