רעיון לפרויקט גמר – מבוא למדעי הנתונים

שמות המגישים:

שפיר חפיף - 208666909

יוני ירמיהו – 204797682

רקע בקצרה:

שנינו ספורטאים, שבאים מענפים שונים שפחות מוכרים במדינת ישראל.

שפיר שחקן כדורגל אולמות בנבחרת ישראל.

יוני שחקן טניס שולחן בליגת העל ונבחרת לשעבר.

בחרנו לקחת את ענף הטניס שולחן, ספורט לא מוכר בארץ בהתפתחות תמידית, המונה מעל 1000 שחקנים פעילים במסגרות הליגה השונות, ומעל 70 מועדונים בכל רחבי הארץ, שלכל מועדון מספר קבוצות בליגות השונות.

לכן - שאלת המחקר שלנו עוסקת בתחום הטניס שולחן:

\*האם ניתן להגיע לרמת חיזוי של תוצאה מדויקת (מערכות), והאם ניתן להגיע לרמת חיזוי גבוהה של טפסי ליגה שונים של הדרגים הבכירים, להבין ולחשב את ההשפעות השונות של כל מיני משתנים בין משחק למשחק.

\*האם ניתן לזהות דפוסים באמצעותם מאמנים מרכיבים את טפסי הליגה השונים והאם ניתן לחזות טופס של משחק.

טפסי הליגה מורכבים משתי קבוצות, לכל קבוצה יש שלושה מקומות לרשום כל שחקן, כך שהמטרה של כל קבוצה היא לא ליפול על השחקנים החזקים של הקבוצה השניה ובעצם "לגנוב נקודות". לטופס המשחק חשיבות עצומה כחלק מטקטיקת המאמן לאותו משחק.

מצורף טופס לדוגמה:



דרכי התמודדות עם שאלת המחקר:

כל הרכשת הנתונים תעשה באמצעות מה שנלמד במסגרת הקורס.

Crawling with BeautifulSoup

את התוכן נסרוק מהאתר של ניהול הטניס שולחן בישראל:

<http://www.tttm.co.il/>

את רשימת השחקנים המובילים נסרוק משני העמודים הראשונים ומכל קישור של שחקן בנפרד,

<http://www.tttm.co.il/rk/MS-1/%D7%98%D7%A0%D7%99%D7%A1-%D7%A9%D7%95%D7%9C%D7%97%D7%9F-%D7%93%D7%99%D7%A8%D7%95%D7%92-%D7%92%D7%91%D7%A8%D7%99%D7%9D>

את רשימת המשחקים נקח מכל שחקן בנפרד בערך כ150 משחקים אחרונים (עד 2017).

נטפל בנתונים, נסיר כפילויות וניצור טבלה מספיק גדולה שבאמצעותה נוכל להתמודד עם השאלות השונות.

נרכז את כל 200 השחקנים המדורגים בארץ, ניקח את הניקוד של כל אחד, דירוג, סימן זיהוי, כתובת לפרופיל ושם השחקן.

בטבלה השניה, טבלת המשחקים, נסרוק באמצעות הרכשת נתונים את כל המשחקים ששוחקו במסגרות השונות עבור כל שחקן.

בטבלה זו נמלא את העמודות הבאות:

תאריך המשחק, מסגרת המשחק (ליגה, תחרות, גביע וכו..), עמודות המייצגות את שחקן מספר 1 ו2 ששם יש את שמות השחקנים, סימן זיהוי, דירוג נוכחי למשחק, יתרון ביתיות, מספר מערכות שכל שחקן השיג.

ננסה לבדוק את הקשרים השונים בין דירוג השחקן לבין התוצאה, נבדוק על אילו שחקנים יתרון הביתיות משפיע על המשחק עצמו.

בהמשך בעזרת לומדת מכונה ובדיקות עצמאיות ננסה להסיק מהנתונים השונים תוצאות עבור משחקים עתידיים.

בשאלת המחקר השניה נסרוק לטבלה את כל טפסי משחקי הליגה שהתקיימו בשלוש שנים האחרונות וננסה לעלות על דפוסים קבועים של כל קבוצה עבור רשימת טופס המשחק ובאמצעות כך לנסות ולהרכיב טפסים דומים ולהתחיל לנבא משחקים עתידיים להתקיים.

בטבלה זו עמודותיה יהיו מאמני שתי הקבוצות, קבוצה 1, קבוצה 2, סדר רשימת השחקנים לכל קבוצה ואיפה משוחק המשחק.

כספורטאים שבאים מתחום הספורט הקטן בישראל, לא הצלחנו למצוא נתונים, סטטיסטיקות וחיזויים ואנו מצאנו עניין אדיר בנושא שקשור אלינו באופן אישי.

שאלות המחקר שלנו מיוחדות עבורנו ועבור הענף, מכיוון שלא קיים שום דבר שדומה לכך ונצטרך לטפל, לחשב ולהרכיש הכל בעצמנו, בתקווה שבעתיד העבודה הזאת תוכל לעזור לענפים האלה לצמוח ולהתפתח, שכן הם גם ככה במגמת צמיחה ובעצם לבדוק אילו שחקנים כדאי לרכוש לכל קבוצה, לבנות אסטרטגיות מול קבוצות אחרות וכו.