
ניתוח ביצועים בספורט (sport analytics) בעידן נתוני העתק (big-data) – ציוני דרך, פיתוח מטריקות ואתגרים

איליה מורגולב,^{1,2,3} רוני לידור,² איל אליקים^{4,2}

¹המכללה האקדמית לחינוך ע"ש קיי

²המכללה האקדמית בוינגייט

³המחלקה למנהל עסקים, אוניברסיטת בן גוריון בנגב

Zone7 Technologies Inc⁴

תקציר

מטרת מאמר זה היא לבחון את ההתפתחות של תחום ניתוח ביצועים בספורט (sport analytics) בעת הזו, עידן נתוני העתק (big data), ואת תרומתו של תחום זה לחיזוק ההבנה של חוקר הספורט את התופעות שבהן הוא צופה ומנסה להבין. ההתפתחות הטכנולוגית המואצת שחלה בשלושת העשורים האחרונים אפשרה לחוקר הספורט לפתח טכניקות צפייה ומעקב אחר ביצועים של ספורטאים במהלך אימון, משחק או תחרות. טכניקות אלו מאפשרות לחוקר הספורט ללמוד (כמעט) על כל תנועה ופעולה שהספורטאי והקבוצה עושים בזמן אמת. החוקר יכול גם לשמור נתונים אלו ולקבצם בקובצי עתק המאפשרים לו להגדיל משמעותית את היקף הנתונים שהוא מעוניין לנתח. תיעוד רחב היקף של ביצועים המתרחשים בזמן אמת וניתוחם מאפשרים לחוקר הספורט לחזק את הבנתו את התופעה הנצפית, להסביר אותה, לנבא את התרחשותה, ולעיתים גם לשלוט בה. אנו בוחנים במאמר זה את ציוני הדרך המרכזיים בהתפתחות תחום ניתוח ביצועים בספורט, דנים בצורך לפתח מטריקות (מדדים) מורכבות שיסייעו למאמן בתהליכי קבלת החלטות, מתארים כמה אתגרים העומדים כיום בפני חוקרי ביצועים בספורט, וטוענים שניתוח ביצועים בספורט יכול לשמש לא רק את חוקר הספורט אלא גם חוקרים בתחומי דעת אחרים כמו כלכלה ופסיכולוגיה. אנו מביאים במאמר זה דוגמאות של מטריקות מענפי ספורט שונים ובעיקר משני משחקי הכדור הפופולריים – כדורגל וכדורסל.

תאריכים: ניתוח ביצועים, ספורט, מטריקות, קבלת החלטות, כדורגל, כדורסל

"האסכולה האנגלית השרישה באירופה את משחק המסירות הארוכות והגבוהות, אולם אותם קרובי משפחה רחוקים מהקצה השני של העולם לא כיבדו את המסורת המשפחתית. הם העדיפו להמציא כדורגל של כדורים קצרים מרגל לרגל, עם שינויי קצב מסחררים והתקפות מהירות". (גלאנו, 2006, עמ' 58)

"... אני גם לא מת על סגנון המשחק שלהם. סגנון שהוא בלי ספק סולידי, אתלטי וקבוצתי, שבמקרה הטוב אפשר לקרוא לו גרמני, עם טיפוסים כמו גרטס או פיליפ אלבר, ובמקרה הרע הוא מזכיר יותר את סגנון המשחק של קבוצות כמו ורגס או בוורן, עם טיפוסים מלאים שליליות, עם נעליים מלאות בוץ...." (טוסן, 2020, עמ' 38)

דברים אלו שכתבו הסופר האורוגוואי אדוארדו גלאנו והסופר הבלגי ז'אן-פיליפ טוסן, מבטאים תובנות שהם גיבשו על משחק הכדורגל. לשניהם מוניטין לא רק של אנשי רוח אלא גם של אוהבי כדורגל. הם הרבו לכתוב על סגנון המשחק ההגנתי וההתקפי של שחקנים מנבחרות שונות שהשתתפו באליפויות אירופה ובאליפויות עולם בכדורגל. הם אף ערכו השוואות בין יעילות הביצוע של נבחרות מיבשות שונות. גלאנו וטוסן אינם היחידים המנסים לגבש תובנות המתבססות על תצפיות של ביצועי ספורטאים וקבוצות. אוהדים הצופים בעקיבות בספורטאים ובקבוצות, עיתונאים המסקרים ענפי ספורט שונים ואנשי מקצוע כמו מאמנים ומנהלים מגבשים אף הם תובנות על תופעות שונות ומגוונות בספורט. כך לדוגמה, בהמשך לדבריהם של גלאנו וטוסן, רבים מאוהדי הכדורגל תופסים את הכדורגל הברזילאי כיצירתי ומלהיב, את הכדורגל האיטלקי כהגנתי ואת הכדורגל הגרמני כמכונה משומנת (Anderson & Sally, 2013; Goldblatt, 2006, 2019). נראה שתובנות אלו עוברות מדור לדור ומוטמעות בניתוחים מקצועיים של משחק הכדורגל. השאלה הנשאלת בהקשרן היא – האם הן אכן משקפות הלכה למעשה את המתרחש על המגרש או שהן משקפות כיצד המתבונן תופס את האירוע? ובמילים אחרות, האם תובנות אלו בכלל נכונות?

חוקרים בתחומי דעת שונים כמו חינוך, כלכלה, ניהול, פסיכולוגיה ורפואה שמו להם למטרה לתאר תופעות, להסביר אותן, לנבא את התרחשותן ובסופו של דבר גם לשלוט בהן (Hatcher, 2013; Livingston & Lewis, 2019). חוקרים אלו אוספים נתונים באופן שיטתי, מנתחים אותם תוך שימוש בשיטות סטטיסטיות מתאימות, ועל בסיס הממצאים העולים מהניתוחים הסטטיסטיים, הם מציגים ראיות מבוססות נתונים (evidence-based data). אנשי מחקר, אם כך, מנסים להציג תובנות שהן תוצאה של תהליך מבוקר ושיטתי ולא תוצאה של התרשמות אישית מהתופעה הנצפית. חוקרי ספורט אינם שונים מחוקרים בתחומי דעת אחרים: גם הם אוספים נתונים ומנתחים אותם כדי לתאר תופעות המעניינות אותם, להסביר תופעות אלו, לנבא מתי הן יתרחשו או לא יתרחשו, וגם כיצד ניתן לשלוט בהן (Martin, 2016; Miller, 2016).

אחד התהליכים הראשונים שחוקרים נדרשים לו בעבודתם הוא תיעוד התופעה הנצפית. ככל שתהליך תיעוד התופעה הנלמדת יהיה רחב יותר בהיקפו ויכלול היבטים רבים ומגוונים שלה, כך יהיו הנתונים שיעמדו לרשות החוקר אותנטיים יותר ויסייעו לו לתאר אותה, להסביר אותה, ומכאן גם לנבא את התרחשותה ברמת דיוק גבוהה. אחד התחומים המתפתחים בעשורים האחרונים במדעי הספורט והתנועה, המסייע לחוקר הספורט לתעד נתונים, לתאר אותם ולנתחם במדויק, הוא התחום העוסק בניית ביצועים בספורט (sport analytics). תחום זה מתבסס על עבודת תיעוד מקיפה של התופעות הנצפות, המלווה בשימוש באמצעים טכנולוגיים מתקדמים, המאפשרים לחוקר לתעד כמעט כל פעולה שהספורטאי עושה בזמן אמת – במהלך אימון, במשחק או בתחרות. אמצעים טכנולוגיים אלו מסייעים לחוקר לתעד אירועים (ביצועים) בהיקף שבעבר לא יכול היה להגיע אליו (Alamar, 2013; Miller, 2016).

מטרת מאמר זה היא לבחון את ההתפתחות של תחום ניתוח ביצועים בספורט ואת תרומתו לחיזוק הבנת החוקר את התופעות שהוא חוקר. הטענה המרכזית המועלת במאמר היא שניתוח ביצועים בספורט הוא שלב הכרחי בטיפול ההמשגה של חוקר הספורט את התופעה הנצפית. היישום של שלב זה – תיעוד הביצועים וניתוחם – מאפשר לחוקר להסביר את התופעה הנצפית ולנבא את התרחשותה ובכך לסייע לבעלי עניין – קובעי מדיניות, מנהלים, מאמנים ומדריכים בספורט – לחזק את יכולתם לשלוט בתופעה הנחקרת. השימוש בנתוני עתק בספורט לא רק יחזק אפוא את הבנתו של החוקר את התופעה הנחקרת אלא גם את יכולתו להנגיש את ממצאיו לבעלי העניין. אנו בוחנים במאמר זה את ציוני הדרך המרכזיים בהתפתחות התחום, מציגים כמה מטריקות (מדדים) שפותחו בתחום ואת השימוש בהן, דנים בכמה אתגרים העומדים בפני חוקרי ספורט המנתחים ביצועים של ספורטאים וקבוצות וטוענים שניתוח ביצועים בספורט יכול לשמש לא רק את חוקר הספורט אלא גם חוקרים בתחומי דעת אחרים כמו כלכלה ופסיכולוגיה. אנו מביאים במאמר זה דוגמאות של מטריקות מענפי ספורט שונים ובעיקר משני משחקי כדור פופולריים – כדורגל וכדורסל.

ציוני דרך מרכזיים

כבר במאה ה-19 החלו עיתונאים ופעילי ספורט לתעד ולנתח ביצועים של ספורטאים ושל קבוצות. אחד מענפי הספורט הראשונים שהחל בו תהליך של תיעוד מסודר ועקיב של ביצועי שחקנים וניתוח ביצועים אלו הוא הבייסבול, אחד מענפי הספורט הפופולריים ביותר בארצות הברית, יפן וקובה (Keri, 2006). כבר בשנת 1859 פיתח הנרי צ'קדוויק, עיתונאי ופעיל ספורט בולט שנחשב לאבי הבייסבול, שיטת ניקוד שתארה בשיטתיות ובאובייקטיביות את רצף ההתרחשויות במשחק, ביניהן חבטות, ריצות, שגיאות ופסילות (Schiff, 2008). קידוד וכימות מהימן זה של פעולות משמעותיות תוך שימוש בלוחות ובאיורים נקרא *Notational Analysis* (Hughes & Franks, 2004). משחק הבייסבול, נוסף על היותו ענף ספורט מקצועני המסוקר בהרחבה בעיתונות הכתובה, הדיגיטלית והמשודרת, מורכב מפעולות הגנה והתקפה בדידות עם תוצאה ברורה, המאפשרות כר נוח לקידוד ביצועים. לקראת סוף המאה ה-19, בעידן שלפני המחשב והטכנולוגיה המתקדמת, החלו כתבי ספורט

ואנשי צוות האימון לכמת בעזרת עיפרון ונייר את ביצועי הקבוצות והשחקנים בליגות הבייסבול השונות בצפון אמריקה ולשקף את המתרחש הלכה למעשה בשדה המשחק.

כתבי עת מקצועיים בראשית המאה ה-20 החלו לפרסם מידע גם על ביצועים מענפי ספורט נוספים. לדוגמה, החלו להופיע בעיתונות הכתובה לוחות עם שכיחויות של חבטות מוצלחות ואייסים (חבטות פתיחה שלא נענו מצד היריב) במשחק הטניס או מידע על מרחקי ריצה ושכיחויות של בעיטות והפלות במשחק הפוטבול האמריקני. ב-*Baseball Magazine*, עיתון שהופיעו בו מאמרים על בייסבול בלבד, נכתב כבר בשנת 1913 שאיסוף הנתונים מאפשר לאנשי המקצוע לא רק לתאר אלא גם להסביר את הגורמים המבדילים בין הניצחון להפסד במשחק (Eaves, 2015). מאמץ מוקדם נוסף להבנת גורמי ההצלחה במשחק הכדורסל עשה נט הולמן (שחקן, מאמן וחבר בהיכל התהילה של הכדורסל האמריקני) בספרו *Scientific Basketball* (Holman, 1922).

אחד הניסיונות המוקדמים הנוספים לתעד ביצועים בספורט ולהבין את משמעותם הוא מחקרו של צ'ארלס ריפ, רואה חשבון שפרש מהשירות בחיל האוויר הבריטי והחל להתעניין בקשר בין מספר המסירות בהתקפה לבין סיכויי ההבקעה של קבוצות כדורגל. במהלך שנות החמישים והשישים של המאה הקודמת תיעד ריפ את מספר המסירות בהתקפה ואת השערים שהובקעו במאות משחקים בליגות באנגליה. ניתוח הנתונים הראה כי 80% מהשערים הובקעו בהתקפות של שלוש מסירות או פחות. מכך הסיק ריפ שכדורגל של מסירות ארוכות וישירות, המקדמות את הכדור מהר מהגנה להתקפה (long-ball), יעיל יותר מכדורגל של הנעת כדור רוחבית ממושכת. בשפה מדעית, בשלב הראשון תיאר ריפ את ההתקפות בליגה האנגלית מבחינת מספר המסירות והשערים, ובשלב השני הוא ניתח את הנתונים וניסה להסביר מה הם הגורמים המעלים את סיכויי הבקעת השערים בכדורגל (Reep & Benjamin, 1968).

במהלך השנים ספג ריפ ביקורת על כך שהתעלם מהעובדה כי שערים רבים בכדורגל מובקעים כתוצאה מהתקפות מתפרצות או לאחר לחץ שמפעילים שחקני ההגנה על שחקני ההתקפה, המוביל לחטיפת כדור באזור שער היריב ולהבקעת שער. לכן, הממצא כי רוב השערים במשחק מובקעים בהתקפות קצרות אינו בהכרח מצביע על כך שיש לחתור להתקפות קצרות במקרים שבהם הקבוצה מתחילה לארגן התקפה מסודרת מאזור השער שלה (Hughes & Franks, 2005). בעוד הדיון סביב המסקנות של ריפ עדיין נמשך (Pollard, 2019), חשוב להדגיש שכדי לבצע ניתוח מתקדם יותר על מהות הקשר הסטטיסטי בין מספר המסירות וסוגיהן לבין סיכויי ההבקעה של הקבוצה, יש צורך להשתמש בכלי איסוף ובשיטות לניתוח נתונים מתקדמים יותר מאשר הכלים שבהם השתמש ריפ – עיפרון ופנקס.

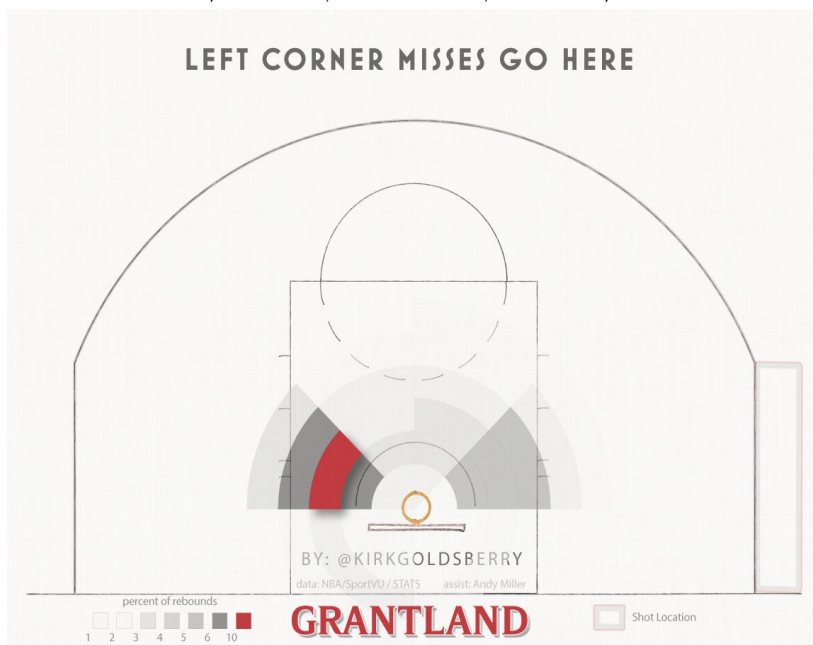
בשנת 1971, קבוצה של בעלי עניין בתהליכי איסוף ועיבוד נתונים בבייסבול ייסדה את החברה האמריקנית למחקר בבייסבול – *Society for American Baseball Research* (SABR). שנה לאחר מכן החלה החברה להוציא לאור כתב עת מקצועי – *The Baseball Research Journal*, המהווה קפיצת מדרגה מבחינת תיאור היקף ועומק הניתוחים הסטטיסטיים בספורט (Morgulev et al., 2018). אחד החברים הבולטים בארגון ה-SABR הוא ביל גיימס (Bill James), שעבודתו בתחום

הערכת תפוקה של שחקני בייסבול היוותה השראה למנכ"ל קבוצת אוקלנד אתלטיקס בילי בין (Billy Beane). בילי בין הצליח להשיג את רצף הניצחונות הארוך ביותר בהיסטוריה בליגת הבייסבול המקצוענית בארצות הברית, ה-*Major League Baseball*. רצף 20 הניצחונות של בילי בין בעונת 2002 מתואר בספר *מאניבול* (Lewis, 2004) ובסרט שיצא לאקרנים באותו שם. הן הספר והן הסרט נתפסים כציוני דרך מרכזיים בהפצת תחום ניתוח ביצועים בספורט לציבור הרחב. שיטות לקידוד ידני של מהלכי משחק פותחו גם בענפי המחבט, אולם המספר הגדול של חבטות ואירועי ניקוד במשחק הסקווש והטניס לא אפשר לשיטות ידניות אלו להיות מיושמות בפועל, שכן המחשב הפך לכלי מרכזי בעבודתו של חוקר הספורט (Hughes et al., 2007). בהקשר של מחשוב תהליכי הקידוד של ביצועי הספורטאים והניתוח של נתונים אלו יש לחזור שוב למשחק הבייסבול. ביל גיימס שהיה מתוסכל מסירוב הנהלת הליגה לפרסם באופן מרוכז את רצף האירועים מכל משחק (play-by-play), הקים בשנת 1984 את פרויקט ה-*Scoresheet*, רשת של חובבי בייסבול שקידדו וריכזו את הנתונים שאספו ממשחקים שונים ושיתפו בהם את הציבור הרחב. בשנת 1989, על בסיס התשתית של *Scoresheet*, הקים דיוויד סמיט, פרופסור לביולוגיה, את פרויקט ה-*Retrosheet* שאיחד חובבי סטטיסטיקה בבייסבול כדי למחשב את לוחות הניקוד של כל המשחקים בליגה המקצוענית מתחילת המאה ה-20 (Costa et al., 2019). יוזמות אלו הביאו, בין היתר, לצמיחתה בשנות השמונים של המאה הקודמת של חברת נתוני הספורט הגדולה בעולם, ה-*STATS, Inc.*, כיום ה-*Stats Perform* (Hutchins, 2016).

שמירת ביצועים מאלפי משחקים במאגר ממוחשב (*Retrosheet*) מאפשרת לאנשי ניתוח ביצועים בספורט לפתח בשלב ראשון מטריקות (מדדים; metrics) עבור כל שחקן בקבוצה (כמו למשל אחוז חבטות מוצלחות, אחוז עלייה לבסיס ראשון), ובשלב השני לבחון את מערכת היחסים בין מטריקות אלו ובין סיכוייה של הקבוצה להשגת ניצחון. רובד נוסף בניתוח הביצועים הוא מציאת קשרים בין משתני היכולת של השחקן (מהירות ריצה, עוצמת הגשה) לתפוקות שלו במגרש. שימוש ברגרסיה מרובה למציאת הקשר בין משתנים בלתי תלויים, כגון יכולות פיזיות של השחקן ומדדים שהוא משיג במשחק, לבין משתני הצלחה, כגון צבירת נקודות והשגת ניצחונות, הוא הבסיס הסטטיסטי שמתבסס עליו תהליך ניתוח הביצועים בספורט (Alamar, 2013).

בשנות התשעים של המאה הקודמת החלו ליישם מערכות ייחודיות לאיסוף נתונים במשחקי כדור. לדוגמה, בעונת 1995-1996 כל קבוצות ה-NBA השתמשו בפלטפורמת ה-*Advanced Scout* לקידוד נוח ויעיל של רצף האירועים במשחק. מערכת זו מבוססת על מתייג אנושי שצופה במשחק ומשתמש במקלדת ייחודית המאפשרת קידוד נוח של משתנים כמו מיקום הזריקה לסל, סוג הזריקה, תוצאת הזריקה, ריבאונד (כדור חוזר מהסל), חסימה וחטיפה. בסיום המשחק הנתונים עולים למאגר שיתופי הנגיש לכל קבוצות הליגה (Bhandari, 1997). נתונים מעונות משחקים שלמה, הזמינים במאגר מידע, מאפשרים פיתוח של מטריקות שהן מעבר לאחוז קליעות מוצלחות של שחקן מהשדה או ממוצע ריבאונדים למשחק. לדוגמה, ה-*Shot Chart* הוא מטריקה מרחבית המציגה את דפוס הזריקות של השחקן, של הקבוצה או של הליגה כולה. מאמן הצופה במטריקה זו יכול לראות מאיזה אזור

במגרש השחקן זורק יותר זריקות לסל, מאיזה אזור הוא זורק מעט, ומהו אחוז ההצלחה שלו בכל אזור. תיאור דומה של ביצועים ניתן לייצר גם עבור קליטת ריבאונדים בהגנה ובהתקפה (Reich et al., 2006), כפי שניתן לראות באיור 1.



איור 1.

התפלגות כיווני נפילה של כדורים חוזרים בזריקה מטווח שלוש נקודות מהפינה השמאלית של המגרש¹

בסוף שנות התשעים של המאה הקודמת התחילו להשתמש בכדורגל במערכות מעקב לניטור תנועת השחקנים והכדור במגרש בעזרת מערכות וידאו. מערכות אלו (כמו *Prozone* ו-*Amisco*) עושות שימוש במצלמות וידאו מיוחדות ובאלגוריתמים לזיהוי אובייקטים ותנועות במרחב. המערכת מצלמת את מיקומי השחקנים והכדור עשרות פעמים בשנייה, פעולה המאפשרת לדעת את המהירויות ואת כיווני התנועה של כל האובייקטים הנעים במגרש. מתייג אנושי מלווה את עבודת המערכת על ידי תיעוד הפעולות שהשחקנים מבצעים כמו מסירות, בעיטות, הרמות, נגיחות, תיקולים ומאבקים על כדור. שילוב זה של מתייג אנושי לצד מערכת חצי-אוטומטית מספק מאגר נתונים מגוון בסיום כל משחק. מאגר זה מאפשר לחשב את תנועת השחקנים במגרש, את מהירות התנועה, את מספר הפעולות בהגנה ובהתקפה ואת אחוזי ההצלחה בכל סוג פעולה. זאת לצד מטריקות מרחביות מתקדמות יותר, כגון

¹ איור זה נגיש לכול; הוא מבוסס על ניתוח של קירק גולדסברי של 95 אלף כדורים חוזרים בעונת 2013-2014 ב-NBA. מתוך: <https://grantland.com/features/how-rebounds-work>

עמדות השחקנים במהלך המשחק, אזורים "חמים" במגרש שבהם נבעטות רוב הבעיטות, ושחקנים מרכזיים שדרכם עברו רוב ההתקפות (Castellano et al., 2014). כאשר בוחנים את ציוני הדרך המאפיינים את התפתחות תחום ניתוח הביצועים בספורט, ניתן לציין גם את תרומתם של חוקרים ישראלים. לדוגמה, גל עוז ומיקי תמיר, בעלי רקע בטכנולוגיות ביטחוניות של זיהוי אופטי וניטור טילים, פיתחו בתחילת שנות ה-2000 מערכת מעקב וידאו נוספת ה-*SportVu*. לאחר ניסויים ראשוניים במגרשי הכדורגל בישראל נרכש סטארטאפ ישראלי זה על ידי ה-*STATS Inc*. בסוף שנות ה-2000 הותקנה המערכת באולמות ה-NBA, והחל בעונת 2013-2014 כל קבוצות הליגה משתמשות בה (Fried & Mumcu, 2016). פעילות מערכת זו מלווה ברכיבי בינה מלאכותית (Duarte et al., 2021), המסוגלים לזהות ולתעד תרגילי כדורסל מובנים שכל קבוצה מבצעת במהלך משחק (Wang & Zemel, 2016). בהקשר הישראלי יש להזכיר גם את חברת פלייסייט,² המתמחה בפיענוח וידאו במשחקי טניס. גם לחברה זו שורשים בתעשייה הביטחונית עם טכנולוגיה שסייעה לאסוף נתונים על אימוני טיסה בחיל האוויר.

שלב נוסף בהתפתחות מערכות זיהוי וקידוד נתונים הוא ניתוח או פיענוח אוטומטי של קטעי וידאו מצולמים (video recognition). טכנולוגיית למידת מכונה (machine learning) מקרבת את היום שבו מערכות אלו יוכלו לזהות את הדפוס של מנחי הגוף, התנועה והכדור הייחודיים לפעולה ספציפית כמו זריקת ניתור, צעד וחצי, חסימה ויציאה להתקפה מתפרצת בכדורסל.³ בסיום שלב הלמידה תהיה המערכת מסוגלת לקדד (ללא מעורבות אדם) שידור וידאו של משחק ספורט למאגר נתונים הכולל לא רק את מיקומי השחקנים והכדור עם מהירויות וכיווני התנועה, אלא גם את רצף הפעולות במגרש כמו כדור, הגנה, חטיפת כדור והטבעה (Chen & Wang, 2020; Pan & Li, 2020).

מקור מידע נוסף על אודות תנועה ותפקוד הספורטאים במהלך האירוע הוא השימוש ב"ביגוד חכם" (smart clothing).⁴ ביגוד זה שוזר בתוכו חיישנים המספקים מידע על תנועת הספורטאי, על התאוצה שלו ועל מדדים פיזיולוגיים כגון דופק, קצב נשימה וטמפרטורת גוף (Feng et al., 2021). נתונים המתקבלים מחיישנים אלו מספקים: (א) משוב על טכניקת הביצוע של הספורטאי (ביצועים כמו קפיצה, בעיטה, זריקה); (ב) מידע בזמן אמת על עצימות האימון; (ג) אפשרות לניטור עומסים והקטנת ההסתברות לפציעה (Hanуска et al., 2016).

ריבוי מקורות המידע, הכמות הגדולה של הנתונים הנאספים, מורכבות הניתוחים הסטטיסטיים וזרם בלתי פוסק של חידושים טכנולוגיים הביאו להקמת פלטפורמות ייעודיות שבהן מתאפשר שיח בין אנשי אקדמיה מדיסציפלינות מגוונות לאנשי טכנולוגיה ואנשי ספורט, כמו למשל ה-*Big Data Analytics in Sports*⁵, מרכז מחקר המתמקד ביצירת תובנות והפצת ידע בעיקר בתחום משחקי הכדור או ה-*S-TRAINING*, פורום רב-תחומי המתמקד בפציעות ספורט. אתגר נוסף בעידן נתוני עתק הוא יצירת ממשקים נוחים להנגשת המידע עבור איש המקצוע בשטח.

² ראו: PlaySight Interactive

³ https://www.youtube.com/watch?v=66ko_cWSHBU&t=200s

⁴ https://www.youtube.com/watch?v=k1VgJ_XPwHk

⁵ <https://bodai.unibs.it/bdsports>

הבקשר זה ניתן לראות יותר ויותר יישומים ייעודיים המציגים את הנתונים בצורה ידידותית וגרפית כגון InStat Basketball Scout⁶ ו-Stats Zone⁷.

פיתוח מטריקות והשימוש בהן

השילוב בין השימוש בטכנולוגיות מתקדמות לאיסוף נתונים, הכולל כל תנועה ופעולה שמבצע הספורטאי במשחק או בתחרות, מיקומו הפיזי של הספורטאי באירוע הספורט הנצפה והפעולה שהוא עושה במהלך האירוע, ובין השימוש בתוכנות סטטיסטיות ובתכנות ספציפי של שלב ניתוח הנתונים, מאפשר לחוקרי הספורט להצביע על תופעות המאפיינות את הביצועים שנצפו. ביתר פירוט, ניתוח קובצי עתק (big data files) של ביצועים אותנטיים של ספורטאים ושל קבוצות מאפשר לחוקרים לתאר את התופעות הנצפות, להבין אותן ואף לנבא את התרחשותן. החוקרים מנסים להבין, בעיקר, מה הם הגורמים המסייעים לספורטאי ולקבוצה להגיע להישגים ומה הם הגורמים המונעים מבעדם להגיע להישגים. כך למשל בכדורסל, החוקרים מנסים לאתר מה מאפיין את ההתקפות שהסתיימו בסל, מה ייחודי לשחקנים שנקלטו היטב בליגה, או מה מבדיל בין קבוצות שהגיעו או לא הגיעו לשלב הפלייאוף (Morgulev et al., 2018).

תהליך ניתוח ביצועים בספורט המתבסס על נתוני עתק, כלומר על היקף נתונים של מאות אלפי ואף מיליוני פעולות שהספורטאים מבצעים במהלך הפעילות, דורש מהחוקר לפתח אלגוריתם (רצף של פעולות המוביל לפתרון בעיה מסוימת) המנתח ביצועים שהתרחשו לאורך זמן, לדוגמה, במהלך כמה עונות של משחקים בליגה מסוימת. הפיתוח של האלגוריתם הוא חיוני, שכן רק כך יוכל החוקר, לדוגמה, לזהות אילו פרמטרים (משתנים) חשובים להשגת ניצחון, וגם מה משותף לכל הקבוצות שסיימו במקומות הראשונים בעונת המשחקים, מה משותף לכל הקבוצות שסיימו במקומות האחרונים, ומה הם הגורמים הבולטים המבדילים בין אלו שדורגו ראשונים לבין אלו שדורגו אחרונים. כאשר היקף הנתונים אינו גדול, אין צורך בפיתוח אלגוריתם מיוחד אלא ניתן לנתח את הנתונים בעזרת ניתוחים סטטיסטיים כמו מתאמים.

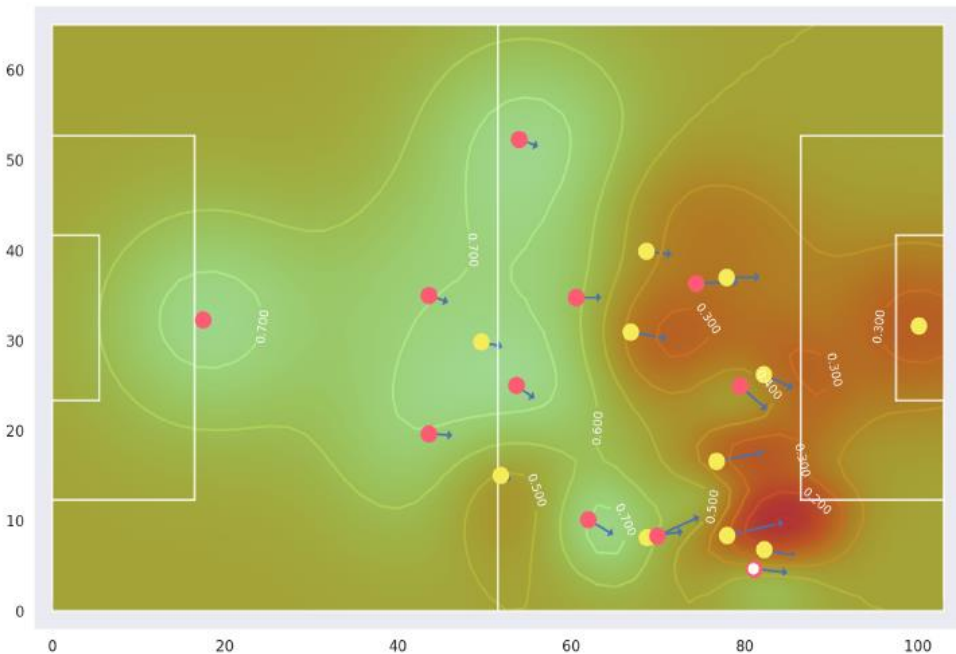
אם החוקר מעוניין לבחון, לדוגמה, את התרומה של גורמים מסוימים להשגת ניצחון במשחקי כדור, בשלב הראשון של הניתוח יבחן האלגוריתם עשרות ולפעמים גם מאות של משתנים שנאספו על כל ספורטאי או קבוצה בכל משחק או תחרות וינסה ללמוד (ומכאן השם *בינה מלאכותית*) איזה פרמטר אכן מצליח להסביר הצלחה (Elyakim et al., 2020). בשלב השני של הניתוח יזהה האלגוריתם מה דומה ומה שונה בין ספורטאים וקבוצות, למשל מה מבחין בין קבוצות צמרת לקבוצות תחתית. פרמטרים מובהקים מוגדרים בספרות כמציני ביצוע משמעותיים (Key Performance Indicators, להלן KPIs) (Mackenzie & Cushion, 2013). המונח *מציני ביצוע משמעותיים* לקוח מהעולם העסקי: אלו הם מדדים (כלכליים ולא כלכליים), שעל פיהם הארגון אומד את מצבו ובעזרתם הוא יכול לקבוע דרכי פעולה

⁶<https://play.google.com/store/apps/details?id=basketball.scout&hl=iw&gl=US>

⁷<https://apps.apple.com/us/app/stats-zone/id1325189497>

כמו העלאת מחויבות העובדים לארגון, פיתוח מנהיגות והגברת מידת שביעות הרצון של חברי הארגון (Marr, 2011).

חוקרים אחדים מדגישים את היכולת המוגבלת של השימוש בפרמטרים בודדים של ביצועי הספורטאי או הקבוצה כגון מספר ספרינטים שהשחקן מבצע במהלך משחק, שכיחות מסירות מדויקות או חטיפות כדור כדי לתאר את המהות הדינמית של משחקי הכדור (Lames & McGarry, 2007; Low et al., 2020). המגמה הנצפית בעשור האחרון היא לנסות לזקק מתוך ריבוי הנתונים מטריקות מורכבות, אך משמעותיות, של ביצועים כגון מידת הביזור והאקראיות (אנטרופיה) בהנעת כדור במשחק (Martínez et al., 2020), השטחים במגרש שבהם קבוצה שולטת במהלך משחק (Fonseca et al., 2012) (ראו איור 2), מידת התיאום (סינכרוניזציה) של תנועת השחקנים במהלך המשחק (Ribeiro et al., 2020) ומידת הריכוזיות של שחקן מסוים ברשת המסירות של הקבוצה (Grund, 2012).



איור 2.

אזורי שליטה של שחקני ברצלונה (נקודות אדומות) ושחקני היריבה (נקודות צהובות) במהלך התקפה מאגף ימין (כדור מסומן בנקודה לבנה). האזורים בגוונים ירוקים הם אזורים ששחקני ברצלונה יהיו הראשונים להגיע אליהם, אזורים בגוונים אדומים הם בשליטת שחקני היריבה.⁸ המחשה גרפית דומה ניתן לייצר על המיקום הממוצע של כל שחקן ושחקן במהלך המשחק כולו.

כאשר עולה בידי מנתח נתונים בספורט לתאר מאפיינים של תופעה מסוימת ולהיות מסוגל להסביר אותה, יעדיו הבאים עשויים להיות הנגשת הממצאים לבעלי

⁸ איור זה נגיש לכול ולקוח מתוך: Fernandez & Bornn, 2018.

עניין רלוונטיים כמו קובעי מדיניות, מנהלים, מאמנים ומדריכים (Alamar, 2013). על חוקר הספורט להגיש את ממצאיו לבעלי העניין באופן שיעלה בידיהם להבין אותם ולהפעיל שיקול דעת אם וכיצד ליישם בעת עבודתם בשדה. על החוקר לתכלל את הממצאים, לאתר את המשותף ביניהם, כמו גם להצביע על תופעות ייחודיות. ככל שהממצאים יוגשו באופן קוהרנטי לבעלי העניין, כך יהיה גדול יותר הסיכוי ליישומם.

הדוגמאות הבאות ממחישות יעדים אלו של החוקר: הדוגמה הראשונה היא הניסיון לנבא את סיכויי ההשתלבות של שחקן בקבוצת הספורט. ניתוח רטרוספקטיבי (מבט לאחור) של מאגרי נתונים על ביצועי הספורטאי מאפשר לחוקר להתמקד בשחקנים בוגרים שהצלחו במיוחד לאורך זמן ולזהות את הפרמטרים שבהם שחקנים אלו היו שונים בגילים צעירים (לדוגמה: גיל 16-18 שבו השחקן אמור להיות בשלב ההתמחות שלו) משחקנים אחרים שלא זכו להצלחה בספורט. מדובר למעשה בזיהוי KPIs ייחודיים לגיל הצעיר, שעשויים לנבא קריירה בוגרת מוצלחת (Morgulev et al., 2018). נוסף על כך, מטריקות של יכולות גופניות כגון זריקת כדור כוח, לחיצת חזה, ניתור אנכי וספרינט למרחק 40 מ' מנותחות על ידי חוקרי ביצועים כדי לזהות גורמי הצלחה משמעותיים עבור תפקידי משחק ספציפיים (Robbins, 2010). השימוש בנתוני עתק מאפשר לחוקר לתעד את ביצועי הספורטאי הצעיר הן במבדקים גופניים שנערכים בתנאי מעבדה ובתנאי שדה, הן הלכה למעשה בזירת הספורט. החוקר יכול לפתח מטריקות המתבססות על נתוני עתק של הספורטאי הצעיר שנאספו הן במבדקים השונים הן במשחק או בתחרות. לדוגמה: החוקר יכול לפתח מטריקה המשלבת בין הישגי כדורסלן צעיר במבדקי כוח מתפרץ אנכי לבין ביצועיו במשחק עצמו הדורשים כוח מתפרץ כמו קליטת כדורים חוזרים בהגנה ובהתקפה וחסימות בהגנה. זאת ועוד, החוקר יכול גם לתעד מה עשה השחקן הצעיר לאחר שקפץ אנכית במשחק: מה הוא עשה לאחר שקלט כדור חוזר בהגנה – עצר את המשחק וחיכה ששחקן אחר יגיע אליו כדי לקבל את הכדור, כדרך, או מסר את הכדור לחבר בקבוצה ובכך יצר התקפה מתפרצת. תכלול הממצאים העולים מניתוחים מסוג זה יסייע לחוקר להבין טוב יותר את התופעה (ניתור אנכי בכדורסל), ובכך לספק מידע רחב ומגוון יותר למאמן על יכולת השחקן בהיבט זה של המשחק. הדוגמה השנייה היא בחירת שחקני כדורסל לליגת ה-NBA במסגרת אירוע הבחירה – הדראפט (*Draft*). הדראפט היא סיטואציה ייחודית שבה לאחר סיום העונה, כל קבוצה על פי סדר שנקבע מראש מקבלת הזדמנות להחתים שחקן ממאגר השחקנים שהגישו בקשה להשתלב בליגה בעונה הקרובה. לקראת הדראפט אנשי מקצוע מנסים לנבא מי מהשחקנים הזמינים ישפר בצורה משמעותית את הישגי הקבוצה בעונות הקרובות. חוקרי ביצועים בספורט מיישמים בניסיונותיהם להעריך את סיכויי השחקן להשתלב בליגת ה-NBA את טכניקת הנדסת תכונה (*feature engineering*): בשלב הראשון של יישום הטכניקה יבקש החוקר ממאמן הקבוצה לספק דוגמאות של כמה שחקנים המשחקים בליגה שלדעתו יתרמו משמעותית לקבוצה שלו. בשלב השני, על בסיס הנתונים הרבים הקיימים על אודות מאות משחקים של השחקנים שעליהם הצביע המאמן, יזהה החוקר מה מבדיל שחקנים אלו משחקנים אחרים המשחקים באותה עמדה, כלומר הוא ינסה לזהות תכונות (*features*) ייחודיות של שחקנים שהמאמן מגדיר ככאלו העשויים לתרום לקבוצתו.

בשלב השלישי, במידת האפשר, ילמד החוקר את הביצועים שהציגו השחקנים הללו בעונות שקדמו לכניסתם לליגת ה-NBA. בשלב האחרון של יישום הטכניקה יציג החוקר למאמן ולצוות המקצועי המלווה אותו בעבודתו מי מהשחקנים הזמינים לבחירה דומים בביצועיהם לשחקנים שעליהם הוא הצביע (Alamar, 2013).

הדוגמה השלישית היא הניסיון לבקרת עומסים ולמניעת פציעות בספורט, בעיקר במשחקי כדור. מועדונים בליגות בכירות בבייסבול, כדורגל, כדורסל ופוטבול אמריקני משתמשים בחיישני תאוצה, בדיקות שתן ודם, מבדקי פלטות כוח ומבדקים איזוקינטיים וביומכניים במטרה לאסוף כמה שיותר מידע על המצב הגופני של השחקנים ועל והעומסים המצטברים שהם נחשפים אליהם במהלך עונת המשחקים. מידע היקפי זה, המצטבר מעונה לעונה, מאפשר לחוקרים לפתח אלגוריתם המנסה לזהות דפוסי עומס ומצב גופני כדי להעריך את סיכויי השחקן לחוות פציעה (Eliakim et al., 2020).

כמה שימושים במטריקות שפותחו הפכו זה מכבר לכלי עבודה בידי מאמנים. בכדורגל לדוגמה, ניתוח רשת המסירות שבוצעו במהלך המשחק מסייע למאמן לזהות את השחקן המשמעותי בהנעת הכדור של הקבוצה היריבה ולהתמקד בו בהגנה. בכדורסל, ניתוח אלפי זריקות הביא את המאמנים להבנה כי זריקות מטווח שלוש נקודות הן יעילות יותר מזריקות שתי נקודות, ולכן שיעור הזריקות מהשלוש בליגת ה-NBA עלה משמעותית מתחילת שנות ה-2000 (Goldsberry, 2019). זאת ועוד, הכנה טקטית ברמת הקבוצה לקראת משחק ספציפי, עבודה על טכניקה ברמת השחקן על סמך נתוני הצלחה בפעולה זו או אחרת, בניית תוכניות אימון אישיות להכנה ולתחזוקת הגוף בהתאם לנתונים הספציפיים של השחקן, הכנה לדו-קרב פנדלים על סמך התפלגות בעיטות קודמות של שחקני היריבה, עבודה על שיפור תהליכי קבלת החלטות על סמך נתונים על אודות דפוסי תנועה, זריקה, בעיטה ומסירה – כל אלה הן דוגמאות לשימושים של מטריקות שפיתחו חוקרי ביצועים בספורט, הנגישות כיום למאמנים וספורטאים (Goes et al., 2020; Hutchins, 2016).

אמידת הלא-נמדד

אחת ההתפתחויות המרתקות בעידן ניתוח נתוני עתק של ביצועים בספורט היא היכולת של החוקר לאמוד את מה שלא היה ניתן למדוד בעבר. כך למשל בכדורגל ניתן לתאר כיצד נוכחותו של קשר קדמי מסוים משפיעה על נתוני אחזקת הכדור של קבוצתו, על דיוק במסירות ועל הגעת ההתקפה לשליש העליון של המגרש. בכדורסל ניתן לתאר כיצד אחוזי הקליעה לסל או מספר איבודי הכדור עולים או יורדים כאשר הרכז המחליף עולה מהספסל ותופס את מקומו של הרכז הפותח במשחק. מטריקות מסוג זה הן מטריקות הוליסטיות בעלות תוקף סביבתי (ecological validity) גבוה משום שהן מצליחות לאמוד בצורה מקיפה את התרומה של השחקן להצלחת קבוצתו (Alamar, 2013).

עם זאת, אנדרסון וסאלי (Anderson & Sally, 2013) מזהירים חוקרי ספורט כי: *Not everything that counts can be counted, and not everything that can be counted counts* (p. 5) (לא כל מה שנחשב יכול להימדד, ולא כל מה שיכול להימדד אכן נחשב). בספרם *The number games – why everything you know about soccer is wrong*, אנדרסון וסאלי (Anderson & Sally, 2013) מציינים את

פאולו מלדיני, קפטן קבוצת מילאן בכדורגל, שנחשב לאחד משחקני ההגנה האיכותיים ביותר בתולדות המשחק. למרות מעמדו המקצועי הרם, הסטטיסטיקה מגלה ששחקני הגנה רבים תיקלו יותר ממנו וחטפו כדורים משחקנים יריבים יותר ממנו (שתי פעולות הגנה משמעותיות שעל מגן בכדורגל המודרני לבצע היטב). מלדיני השכיל להתמקם נכון במהלך המשחק, לסגור קווי מסירה של היריב, לצפות ביעילות את תנועת היריב מבעוד מועד וכתוצאה מכך למנוע את המהלך ההתקפי מלהתרחש, כל זאת ללא הצורך לבצע תיקול של היריב או חטיפת כדור, כפי שנוהגים לעשות שחקני הגנה רבים.

הציטוט של אנדרסון וסאלי, המופיע בפסקה הקודמת, רלוונטי במיוחד למקרה של מלדיני: לעיתים במהלך משחק או תחרות ספורטאים מבצעים פעולות משמעותיות ביותר התורמות להצלחתם או להצלחת הקבוצה אך קשה למדוד אותן כמו שקשה למדוד את דפוסי משחק ההגנה של מלדיני. אחת השאלות המעניינות חוקרי ביצועים בספורט היא כיצד ניתן בכל זאת למדוד את "מה שלא קרה". בשנים האחרונות מתאפשר לחוקרי ספורט, לאור ריבוי הנתונים הנאספים ושכלול ביכולות הסינון, הקטלוג והניתוח שלהם, להציג מדדים קבוצתיים עונתיים עבור דקות המשחק שבהן היה שחקן מסוים בהרכב בהשוואה לדקות שבהן הוא לא שיחק. במקרה של מלדיני, ניתן למשל לתאר כמה פעמים מצליחה קבוצת היריב שנגדה שיחק להגיע עם כדור לרחבת ה-16 בדקות שבהן הוא היה על המגרש, ואם מספר זה נמוך באופן מובהק בהשוואה לדקות שבהן הוא לא שותף במשחק. ניתן לתאר פרמטרים נוספים כמו מספר בעיטות למסגרת או שליטה בכדור על ידי הקבוצה היריבה באזור המסכן את השער, כאשר שחקן מסוים שיחק או לא שיחק. ברוח זו, גולדסברי ווייס (Goldsberry & Weiss, 2013) הראו כיצד משפיעה נוכחותו באזור הסל של שחקן NBA דווייט הווארד על דפוס הזריקות ועל אחוזי הקליעה של שחקנים מהקבוצות היריבות. החוקרים מצאו כי אף שהווארד הוא לא בין החוסמים המובילים בליגה, הוא מצליח בכל זאת להוריד משמעותית את אחוזי הקליעה של קלעי הקבוצות היריבות באזור שבו הוא נמצא. יתרה מזאת, הווארד הוא השחקן המוביל בליגה בהרחקת זריקות היריב מאזורים הקרובים לסל קבוצתו לאזורים רחוקים יותר. נוכחותו במשחק היא כה דומיננטית, עד ששחקני קבוצת היריב אינם זורקים בקרבתו, וכתוצאה מכך לא עולה בידו לצבור חסימות רבות במהלך המשחק.

אתגרים העומדים בפני חוקרי ביצועים בספורט ובעיית ההסקה הסיבתית

אשר לטענתם של אנדרסון וסאלי (Anderson & Sally, 2013) שלא כל מה שנחשב יכול להימדד, ולא כל מה שיכול להימדד אכן נחשב, יש לציין כי רוב המחקר בתחום ניתוח ביצועים בספורט מתמקד בהיבטים נוחים למדידה (Martin, 2016). חוקרים יבחנו בדרך כלל משתני ביצוע קלאסיים. לדוגמה, אם החוקר מעוניין להעריך את משחק ההתקפה של קבוצה בכדורגל הוא ינתח את מספר הבעיטות שנבעטו לשער היריב, דיוק הבעיטות לשער, מספר המסירות שבוצעו במשחק, דיוק המסירות, מספר הרמות הכדור לרחבה, מספר איבודי הכדור, מספר קרנות ומספר נבדלים. החוקרים ינסו להסביר בעזרת חישובי רגרסיה את התרומה של ביצועי

התקפה אלו להצלחת הקבוצה, כלומר הם יצביעו על קשר בין מספר הבעיטות למסגרת לבין ניצחונות (Brito Souza et al., 2019). עם זאת, הבעיה העיקרית בדרך זו של ניתוח ביצועים היא שחוקר המסיק על סיבתיות ממתאם בין מספר הבעיטות לשער לבין מספר הניצחונות שהשיגה הקבוצה בליגה, מסיק למעשה ממתאם מדומה, משום שמשתנה הבעיטות לשער אינו באמת משתנה בלתי-תלוי. כלומר, שני המדדים – בעיטות לשער ומספר ניצחונות – הם תוצר של איכות השחקנים בקבוצה בהשוואה לאיכות השחקנים בקבוצה היריבה. משתנים בלתי-תלויים אחרים כמו רמת התיאום בין השחקנים (סינרגיה) ורמת הכושר הגופני שלהם גם הם עשויים להיות תלויים באיכות שחקני הקבוצה, כך שגם הם אינם משתנים בלתי-תלויים טהורים.

ההנחה המקובלת בקרב חוקרים ואנשי שדה בספורט היא שככל שהספורטאי הוא איכותי יותר כך יהיו ביצועיו טובים יותר. לדוגמה, אנו מניחים שחלוצים טובים בכדורגל יבעטו יותר בעיטות לשער בהשוואה לחלוצים טובים פחות, או שככל שבקבוצת כדורעף יהיו שחקנים טובים יותר, היא תשיג מספר ניצחונות גבוה יותר בליגה מקבוצה שבה יהיו שחקנים טובים פחות. כיוון שקיים קושי בקרב חוקרי ביצועים להעריך את איכות השחקנים, הם מודדים בפועל את ביצועי השחקן ומנסים להצביע על ביצועים אלו כבעלי תרומה פוטנציאלית להצלחה. כך, חוקר שמצא מתאם בין מספר הבעיטות לשער ובין מספר הניצחונות שהשיגה הקבוצה, וטוען כי ככל שהקבוצה תבעט יותר לשער כך תשפר את סיכוייה להשגת ניצחון, מסיק למעשה ממתאם לסיבתיות. ייתכן שקשר זה בין שני המשתנים הוא תוצר של הפער בין איכות שחקני שתי הקבוצות.

כדי להמחיש את הבעייתיות בפרשנות של מתאמים מסוג זה, נביא טענה של חוקר שמצא מתאם גבוה בין מכירה של מטקות לבין מכירה של בגדי ים. לטענתו, עלייה ברכישת המטקות מסבירה את העלייה ברכישת בגדי הים. עם זאת, נראה שדווקא העלייה בטמפרטורה בתקופת הקיץ, שבה רבים מעדיפים לבלות בחופי הים, היא המשתנה הבלתי-תלוי, כלומר המשתנה האמיתי המסביר את העלייה במכירה של שני המוצרים (בייט-מרום, 2001).

ניתן להצביע על בעייתיות דומה בפרשנות של מתאמים גם בתחומי דעת אחרים כמו סוגיית האנדוגניות (endogeneity) בכלכלה וסוגיית הערפול (confounding) בפסיכולוגיה. לדוגמה, חוקרת בפסיכולוגיה של הספורט מצאה קשר בין הובלה (יתרון בנקודות) בסיום המחצית הראשונה בכדורסל לבין ההסתברות לניצחון בסיום המשחק. החוקרת עלולה להסיק סיבתיות ולטעות למשל כי הובלה במחצית גורמת להופעת מומנטום – יתרון פסיכולוגי – המסייע לקבוצה לנצח בסיום המשחק. אולם, ייתכן מאוד, כי חוקרת זו התעלמה ממשתנים מתערבים (sconfounder) - יחסי הכוחות בין הקבוצות עוד לפני תחילת המשחק; על הקבוצה המובילה לקלוע פחות סלים מיריבתה על מנת לנצח. יתרה מכך, לרוב, הקבוצה הטובה יותר היא זו שתוביל במחצית, לכן קבוצה המובילה במחצית היא גם בעלת הסיכוי הגבוה יותר לנצח בסיום המשחק (Morgulev & Avugos, 2020).

ספרות מתודולוגיה במדעי הספורט והתנועה מזהירה מפני הסקה לא מבוקרת ממתאם לסיבתיות בעידן נתוני עתק (Shiffirin, 2016). כך למשל, ברגר ופופ (Berger & Pope, 2011) דיווחו על תופעה מפתיעה בכדורסל, הפוכה לדוגמה הנדונה בפסקה

הקודמת: מדגם של אלפי משחקים הראה שקבוצות שסיימו את המחצית הראשונה בפיגור של נקודה אחת היו בעלות סיכויי ניצחון גבוהים יותר מאשר קבוצות שסיימו את המחצית הראשונה בשוויון. החוקרים פירשו את הממצא במונחי תאוריית הצבת מטרות וטענו כי מצב זה של "כמעט" גורם לעלייה ייחודית במוטיבציה ובמאמץ המגדילה את סיכויי הניצחון של הקבוצה בפיגור נקודה.

מנגד, מריט וקלאוזט (Merritt & Clauset, 2014) ניתחו מדגם של קרוב ל-40 אלף משחקים בענפי ספורט שונים (ביניהם כדורסל) וטענו שיייתכן שקבוצות חזקות, המובילות ומרגישות בטוח במשחקן, מאפשרות לשחקנים כוכבים לנוח בדקות האחרונות של המחצית הראשונה. יכול להיווצר אפוא מצב שדווקא הקבוצות החזקות יסיימו במקרים רבים את המחצית בפיגור קטן. דוגמאות אלו מחדדות שבמחקר לא ניסויי – ומחקרים מבוססי ניתוח נתוני עתק בספורט הם לרוב מחקרים לא ניסויים – על החוקרים להיות מודעים היטב להנחות שביסוד המחקר, ולחשוב ביסודיות על הסברים חלופיים לקשר הסיבתי שהם מבקשים להצביע עליו (Titunik, 2015).

כדי להתמודד עם האתגרים העומדים בפניהם חוקרי ביצועים בספורט מדגישים, בעיקר בעת ניתוחי ביצועים במשחקי כדור, את חשיבות הבקרה על מאזן הכוחות בין הקבוצות משום שאותם שחקנים באותה קבוצה יציגו מדדים אישיים וקבוצתיים שונים כתלות באיכות הקבוצה היריבה (Bilek & Ulas, 2019). כך לדוגמה, לבד (Lebed, 2017) בוחן את משחק הכדורגל (וגם משחקי כדור אחרים) דרך הפרספקטיבה של מערכות מורכבות ומציין כי רוב הזמן במהלך המשחק המערכות נמצאות בשיווי משקל: קבוצת ההגנה מצליחה לסגור את זוויות המסירה של הקבוצה היריבה ולמנוע את ניסיונות החדירה לשער עד שפעולה ייחודית מצליחה ליצור פרטורבציה (הפרעה), כמו פריצה של שחקן המוציאה את המערכות מאיזון ומייצרת מהלך עם פוטנציאל לסיכון השער. בהקשר זה לבד ובר-אלי (Lebed & Bar-Eli, 2013) מציעים להתמקד בניתוח בפעולות פרואקטיביות, קרי, פעולות של יצירתיות, יוזמה ותעוזה, כמו חטיפת כדור בכדוריד או כדור בכדורגל, משום שיכולת השחקן לייצר פעולות ייחודיות אלה היא משתנה מסביר משמעותי להצלחת הקבוצה.

ניתוח נתונים בספורט: לא רק לתועלת הספורט

מנקודת המבט של חוקרים בפסיכולוגיה, סוציולוגיה וכלכלה הספורט הוא כר פורה לניתוח נתונים של התנהגות יחידים וקבוצות בסביבת ביצוע תחרותית עם חוקים מוגדרים ותוצרים מדידים (Morgulev et al., 2014). בניסוי מעבדה, למשל, בניגוד למתרחש בזירת הספורט האותנטית – אולם ההתעמלות, אצטדיון האתלטיקה, בריכת השחייה, מגרש הכדורגל – המשתתפים (המבצעים) במקרים רבים הם סטודנטים מתנדבים, לא מומחים, המתמודדים עם משימות בעלות תוקף סביבתי נמוך תוך העדר תמריצים משמעותיים להצליח במשימה שהם מתבקשים לבצע בניסוי. בעשרות השנים האחרונות הפך הספורט למעבדה עבור חוקרים ממגוון דיסציפלינות המבקשים לבחון קבלת החלטות, ביצועים והתנהגות של מקצוענים בעלי מוטיבציה גבוהה להגיע להישגים (Morgulev et al., 2019).

לדוגמה, כלכלנים בחנו התפלגות של מכות פתיחה בטניס ופנדלים (בעיטות עונשין מ-11 מ') בכדורגל כדי לראות אם המגשים, הבוועטים והשוערים מתנהגים בהתאם לשיווי משקל נאש (Walker & Wooders, 2001; Chiappori et al., 2002). דוגמה בולטת נוספת בכלכלה היא הבחינה של תאוריית הפרוספקט של כהנמן וטברסקי בעזרת ניתוח של מיליוני חבטות גולף (Pope & Schweitzer, 2011). בפסיכולוגיה, חוקרים ניתחו את אחוזי הקליעות לסל של שחקני NBA במטרה לחקור הטיות קוגניטיביות (Gilovich et al., 1985), או בחנו את ההשפעה של קהל ולחץ על ביצועים בבייסבול (Baumeister & Steinhilber, 1984). היריעה כאן קצרה מלהתייחס לכל קשת המחקרים שעשו שימוש בנתוני ספורט על מנת לבחון תאוריות מדעיות בתחומי דעת שאינם מתמקדים בחקר ביצועים בספורט (Bar-Eli et al., 2020).

סיכום

מחבריהם של שני ספרי כדורגל שיצאו בשנים האחרונות בישראל (מונטל, 2018; קבסה ואלוני, 2021) מבליטים את חשיבות השימוש בנתוני עתק בתהליכי ניתוח ביצועים של שחקנים וקבוצות. היקף הנתונים שחוקר ביצועים בספורט יכול לנתח כיום הוא עצום. הטכנולוגיה שבה החוקר משתמש מאפשרת לו לנתח כמעט כל תנועה ופעולה שהספורטאי או הקבוצה מבצעים בזמן אמת. היקף הנתונים הנחקרים והטכנולוגיה המתקדמת מאפשר לחוקר ללמוד לעומק על היבטים רבים ומגוונים של התופעה הנחקרת. עם זאת, על החוקר להפעיל שיקול דעת אילו פרמטרים לנתח ואילו מטריקות לפתח, שכן אפשרויות הניתוח העומדות לרשותו הן רבות ומגוונות ועליו להימנע מלהגיע למצב "שמרוב עצים (נתונים) לא רואים (מתקשים להבין) את היער (התופעה)". על חוקר הספורט להביא בחשבון שהמטריקות שהוא מפתח אמורות לסייע לאנשי מקצוע בספורט – מאמנים, מנהלים וקובעי מדיניות – לקבל החלטות מבוססות ראיות ומכאן גם החלטות טובות יותר. ממצאי מחקרים בתחום ניתוח ביצועים בספורט אמורים להיות נגישים למקבלי ההחלטות, שעשויים ליהנות מהתובנות העולות ממחקרים אלו ולשפר את עבודתם בשדה עם הספורטאים (Alamar, 2013). זאת ועוד, על חוקרי ביצועים בספורט לסייע לאנשי שדה למדוד את אשר הם מתקשים למדוד ולא את מה שנח או קל למדידה. מאמץ משותף של חוקרים ואנשי שדה עשוי לחזק עוד יותר את השימוש בניתוח ביצועים בספורט ואת הפיכתו של ניתוח זה לכלי עבודה חיוני בעבודתם על אנשי השדה בספורט. יש להדגיש שהתובנות העולות מהניתוח, ולא הניתוח עצמו, הן החשובות והן שעשויות לסייע לחוקרים ולאנשי שדה להבין טוב יותר את התופעות המעסיקות אותם.

מקורות

- בייט-מרומ, ר' (2001). שיטות מחקר במדעי החברה עקרונית המחקר וסגנונותיו (מהדורה שנייה). האוניברסיטה הפתוחה.
- גלאנו, א' (2006). כדורגל באור ובצל. גלורי.
- טוסן, ז"י"פ (2020). כדורגל. לוקוס.
- מונטל, ט' (2018). הטיות או לא להיות. אייפאבליש.
- קבסה, ר' ואלוני, א' (2021). פשוט להבין כדורגל. ספרי ניב.
- Alamar, B. (2013). *Sports analytics*. Columbia University Press.
- Anderson, C., & Sally, D. (2013). *The numbers game: Why everything you know about soccer is wrong*. Penguin.
- Bar-Eli, M., Krumer, A., & Morgulev, E. (2020). Ask not what economics can do for sports - Ask what sports can do for economics. *Journal of Behavioral and Experimental Economics*, 89, 101597.
- Baumeister, R. F., & Steinhilber, A. (1984). Paradoxical effects of supportive audiences on performance under pressure: The home field disadvantage in sports championships. *Journal of Personality and Social Psychology*, 47(1), 85.
- Berger, J., & Pope, D. (2011). Can losing lead to winning? *Management Science*, 57, 817-827.
- Bhandari, I., Colet, E., Parker, J., Pines, Z., Pratap, R., & Ramanujam, K. (1997). Advanced scout: Data mining and knowledge discovery in NBA data. *Data Mining and Knowledge Discovery*, 1(1), 121-125.
- Bilek, G., & Ulas, E. (2019). Predicting match outcome according to the quality of opponent in the English premier league using situational variables and team performance indicators. *International Journal of Performance Analysis in Sport*, 19(6), 930-941.
- Brito Souza, D., López-Del Campo, R., Blanco-Pita, H., Resta, R., & Del Coso, J. (2019). A new paradigm to understand success in professional football: Analysis of match statistics in LaLiga for 8

- complete seasons. *International Journal of Performance Analysis in Sport*, 19(4), 543-555.
- Castellano, J., Alvarez-Pastor, D., & Bradley, P. S. (2014). Evaluation of research using computerised tracking systems (Amisco® and Prozone®) to analyse physical performance in elite soccer: A systematic review. *Sports Medicine*, 44(5), 701-712.
- Chen, L., & Wang, W. (2020). Analysis of technical features in basketball video based on deep learning algorithm. *Signal Processing: Image Communication*, 83, 115786.
- Chiappori, P. A., Levitt, S., & Groseclose, T. (2002). Testing mixed-strategy equilibria when players are heterogeneous: The case of penalty kicks in soccer. *American Economic Review*, 92(4), 1138-1151.
- Costa, G. B., Huber, M. R., & Saccoman, J. T. (2019). *Understanding sabermetrics: An introduction to the science of baseball statistics*. McFarland.
- Duarte, A., Couceiro, M., Seifert, L., Sarmiento, H., & Davids, K. (2021). *Artificial intelligence in sport performance analysis*. Routledge.
- Eaves, J. S. (2015). A history of sports notational analysis: A journey into the nineteenth century. *International Journal of Performance Analysis in Sport*, 15(3), 1160-1176.
- Eliakim, E., Morgulev, E., Lidor, R., & Meckel, Y. (2020). Estimation of injury costs: financial damage of English Premier League teams' underachievement due to injuries. *BMJ Open Sport & Exercise Medicine*, 4(1), e000675.
- Elyakim, E., Morgulev, E., Lidor, R., Meckel, Y., Arnon, M., & Ben-Sira, D. (2020). Comparative analysis of game parameters between Italian league and Israeli league football matches. *International Journal of Performance Analysis in Sport*, 20(2), 165-179.
- Feng, Q., Liu, Y., & Wang, L. (2021). Wearable device-based smart football athlete health prediction algorithm based on recurrent neural networks. *Journal of Healthcare Engineering*, 2021.

- Fernandez, J., & Bornn, L. (2018). Wide open spaces: A statistical technique for measuring space creation in professional soccer. In *MIT Sloan Sports Analytics Conference*.
- Fonseca, S., Milho, J., Travassos, B., & Araújo, D. (2012). Spatial dynamics of team sports exposed by Voronoi diagrams. *Human Movement Science, 31*(6), 1652-1659.
- Fried, G., & Mumcu, C. (Eds.) (2016). *Sport analytics: A data-driven approach to sport business and management*. Taylor & Francis.
- Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology, 17*(3), 295-314.
- Goes, F. R., Meerhoff, L. A., Bueno, M. J. O., Rodrigues, D. M., Moura, F. A., Brink, M. S., & Lemmink, K. A. P. M. (2020). Unlocking the potential of big data to support tactical performance analysis in professional soccer: A systematic review. *European Journal of Sport Science, 1-16*.
- Goldblatt, D. (2006). *The ball is round – A global history of football*. Viking.
- Goldblatt, D. (2019). *The age of football – The global game in the twenty-first century*. Macmillan.
- Goldsberry, K. (2019). *Sprawlball: A visual tour of the new era of the NBA*. Houghton Mifflin Harcourt.
- Goldsberry, K., & Weiss, E. (2013). *The Dwight effect: A new ensemble of interior defense analytics for the NBA*. 7th Annual Sports Analytics Conference.
- Grund, T. U. (2012). Network structure and team performance: The case of English Premier League soccer teams. *Social Networks, 34*(4), 682-690.
- Hanuska, A., Chandramohan, B., Bellamy, L., Burke, P., Ramanathan, R., & Balakrishnan, V. (2016). *Smart clothing market analysis*. Technical Report, University of Berkeley.

- Hatcher, L. (2013). *Advanced statistics in research: Reading, understanding, and writing up data analysis results*. Shadow Finch Media.
- Holman, N. (1922). *Scientific basketball*. Inca.
- Hughes, M., Hughes, M. T., & Behan, H. (2007). The evolution of computerised notational analysis through the example of racket sports. *International Journal of Sports Science and Engineering*, 1(1), 3-28.
- Hughes, M., & Franks, I. M. (Eds.). (2004). *Notational analysis of sport: Systems for better coaching and performance in sport*. Psychology Press.
- Hughes, M., & Franks, I. M. (2005). Analysis of passing sequences, shots and goals in soccer. *Journal of Sports Sciences*, 23(5), 509-514.
- Hutchins, B. (2016). Tales of the digital sublime: Tracing the relationship between big data and professional sport. *Convergence*, 22(5), 494-509.
- Keri, J. (2006). *Baseball between the numbers*. Basic Books.
- Lames, M., & McGarry, T. (2007). On the search for reliable performance indicators in game sports. *International Journal of Performance Analysis in Sport*, 7(1), 62-79.
- Lebed, F. (2017). *Complex sport analytics*. Routledge.
- Lebed, F., & Bar-Eli, M. (2013). *Complexity and control in team sports*. Routledge.
- Lewis, M. (2004). *Moneyball: The art of winning an unfair game*. WW Norton & Company.
- Livingston, E., & Lewis, R. J. (2019). *JAMA Guide to statistics and methods*. McGraw-Hill Education.
- Low, B., Coutinho, D., Gonçalves, B., Rein, R., Memmert, D., & Sampaio, J. (2020). A systematic review of collective tactical behaviours in football using positional data. *Sports Medicine*, 50(2), 343-385.

- Mackenzie, R., & Cushion, C. (2013). Performance analysis in football: A critical review and implications for future research. *Journal of Sports Sciences*, 31(6), 639-676.
- Martin, L. (2016). *Sports performance measurement and analytics*. Pearson.
- Marr, B. (2011). Key Performance Indicators (KPI): The 75 measures every manager needs to know. FT.
- Martínez, J. H., Garrido, D., Herrera-Diestra, J. L., Busquets, J., Sevilla-Escoboza, R., & Buldú, J. M. (2020). Spatial and temporal entropies in the Spanish football league: a network science perspective. *Entropy*, 22(2), 172.
- Merritt, S., & Clauset, A. (2014). Scoring dynamics across professional team sports: Tempo, balance and predictability. *EPJ Data Science*, 3, 1-21.
- Miller, T. W. (2016). *Sports analytics and data science – Winning the game with methods and models*. Pearson.
- Morgulev, E., Azar, O. H., & Bar-Eli, M. (2019). Does a “comeback” create momentum in overtime? Analysis of NBA tied games. *Journal of Economic Psychology*, 75, 102126.
- Morgulev, E., & Avugos, S. (2020). Beyond heuristics, biases and misperceptions: the biological foundations of momentum (hot hand). *International Review of Sport and Exercise Psychology*, 1-21.
- Morgulev, E., Azar, O. H., & Lidor, R. (2018). Sports analytics and the big-data era. *International Journal of Data Science and Analytics*, 5(4), 213-222.
- Morgulev, E., Azar, O. H., Lidor, R., Sabag, E., & Bar-Eli, M. (2014). Deception and decision making in professional basketball: Is it beneficial to flop? *Journal of Economic Behavior & Organization*, 102, 108-118.
- Pan, Z., & Li, C. (2020). Robust basketball sports recognition by leveraging motion block estimation. *Signal Processing: Image Communication*, 83, 115784.

- Pollard, R. (2019). Invalid interpretation of passing sequence data to assess team performance in football: Repairing the tarnished legacy of Charles Reep. *The Open Sports Sciences Journal*, 19, 17-21.
- Pope, D. G., & Schweitzer, M. E. (2011). Is Tiger Woods loss averse? Persistent bias in the face of experience, competition, and high stakes. *American Economic Review*, 101(1), 129-57.
- Reep, C., & Benjamin, B. (1968). Skill and chance in association football. *Journal of the Royal Statistical Society. Series A (General)*, 131(4), 581-585.
- Reich, B. J., Hodges, J. S., Carlin, B. P., & Reich, A. M. (2006). A spatial analysis of basketball shot chart data. *The American Statistician*, 60(1), 3-12.
- Ribeiro, J., Lopes, R., Silva, P., Araújo, D., Barreira, D., Davids, K., & Garganta, J. (2020). A multilevel hypernetworks approach to capture meso-level synchronisation processes in football. *Journal of Sports Sciences*, 38(5), 494-502.
- Robbins, D. W. (2010). The National Football League (NFL) combine: Does normalized data better predict performance in the NFL draft? *The Journal of Strength & Conditioning Research*, 24(11), 2888-2899.
- Schiff, A. J. (2008). *"The father of baseball": A biography of Henry Chadwick*. McFarland.
- Shiffrin, R. M. (2016). Drawing causal inference from big data. *Proceedings of the National Academy of Sciences*, 113, 7308-7309.
- Titunik, R. (2015). Can big data solve the fundamental problem of causal inference? *Political Science & Politics*, 48, 75-79.
- Walker, M., & Wooders, J. (2001). Minimax play at Wimbledon. *American Economic Review*, 91(5), 1521-1538.
- Wang, K. C., & Zemel, R. (2016). Classifying NBA offensive plays using neural networks. *MIT Sloan Sports Analytics Conference*.