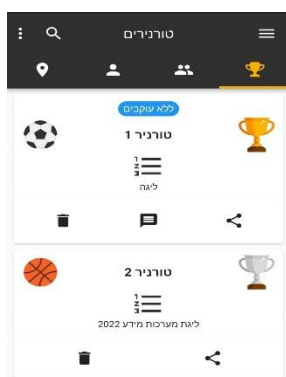


סדנת הכנה לפרויקט – עבודה 1

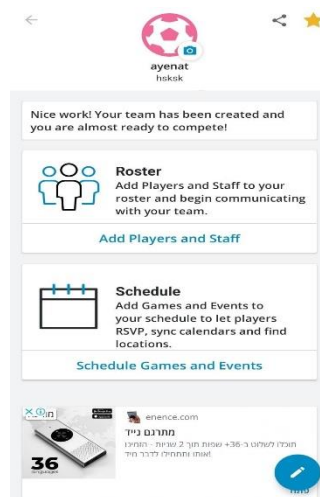
סקירה מעשית – מערכות לניהול ליגות:

ניהול ליגות היא משימה המורכבת ממספר תחומים. מערכת לניהול ליגות היא מערכת מורכבת הצריכה לתמוך בפעולות רבות, ביניהן הוספה ושיבוץ של שחקנים, מאמנים, בעלי קבוצות ושופטים, ניהול יומן משחקים, שליחת התראות למשתמש בהתאם לסוג המשתמש, ניהול המשחקים ושיבוץ, חישובי טבלאות הליגה. המערכת צריכה להיות מסוגלת לבצע אינטגרציה בין דרישות רבות ואילוצים. אנו נסקור ונערוך השוואה בין שלוש מערכות קיימות לניהול ליגות. האחת Monclubsportif , השנייה Winner והשלישית sports engine.

1. Winner – אפליקציה זו מאוד נוחה לשימוש, הממשק שלה מאוד פשוט, נקי ואינטואיטיבי. האפליקציה ניתנת להצגה גם בשפה העברית, ולכל פעולה יש איור המתאר באופן מדויק את ביצוע הפעולה. באפליקציה זו ניתן לנהל טורנירי ספורט, ליגות playoff. ניתן ליצור במהירות טורנירים בכל ענפי הספורט (כדורסל, כדורגל, טניס, אגרוף, פריסבי, בייסבול ועוד), הטורניר יכול להיות אישי, זוגי או קבוצתי. האפליקציה תומכת בכל סוגי פורמט הטורנירים (ליגה, גביע, בתים), בשיתוף טורנירים עם משתמשים אחרים, בעדכון online ו-offline ובהתאמה אישית של המשתמש. תהליך הניהול של הליגות נעשה באופן הדרגתי, שמתחיל בבניית הליגה, הרכבת קבוצות הליגה, וכלה ביצירת פרופיל אישי לשחקנים ושיבוץ בקבוצות שונות.
2. Monclubsportif – מערכת המאפשרת לנהל ליגה ולנהל קבוצות. למערכת זו קיים אתר אינטרנטי וגם אפליקציה שדרכם ניתן לעשות את התהליך של בנייה וניהול של ליגות. מערכת זו נוצרה בצרפת ולכן תומכת רק בשפה הצרפתית והאנגלית. המערכת בנויה כך שיש דגש על חיסכון בזמן כמו למשל אישורי הגעה לאימונים ומשחקים, ניהול מערכת מסרונים פנימית שהיא רק לחברי המועדון, סטטיסטיקות על המועדון והקבוצה. בנוסף, המערכת מציעה סקר יומי לכל השחקנים כדי לעקוב אחרי מצבם הבריאותי, אפשרות לשחקנים חלופיים במקרה של היעדרות, אפשרות לחלוק אסטרטגיות בין מאמנים ויצירה של אלבומי תמונות מהמשחק כדי לחלוק זיכרונות מרגעי המשחק. יחד עם זאת, חסרה במערכת האפשרות ליצור משחק או אירוע בין שתי קבוצות ששייכות לאותה ליגה – כלומר, המערכת לא מבצעת קישוריות בין קבוצות שרשומות למשחק ספציפי, אלא קישור בין קבוצה לשם של קבוצה יריבה בלבד.
3. Sports engine – המערכת מאפשרת ניהול ליגות וקבוצות. למערכת אפליקציה ואתר אינטרנטי לשימוש המשתמשים, והממשק שלה לא אינטואיטיבי, קצת מסובך ומציף. המערכת מתרכזת במספר פעולות, הראשונה שבהן היא ניהול ליגה והוספה של ליגות וקבוצות חדשות, שיבוץ קבוצות למשחקים, ניהול ושימור מידע על הליגות וגם סטטיסטיקות בזמן אמת. במערכת קיימת אפשרות ייחודית של בדיקת רקע והכשרה לצוות. הרישום לאפליקציה והתשלום מאובטחים, אם כי מסובך ולא ברור. בנוסף, ניתן לנהל חנות מזכרות ייעודית לקבוצה. יחד עם זאת, מאוד קשה להוציא ולשתף מידע על הליגה או הקבוצה לשימוש אישי או לשיתוף בין משתמשים.



תמונה 1 – מסך ראשי, אפליקציית Winner



תמונה 2 – מסך ראשי, אפליקציית SportEngine



תמונה 3 – מסך ראשי, אתר MonClubSportif

קריטריון/להשוואה/פיצ'ר מרכזי	Winner	Monclubsportif	Sports engine
חשוב טבלת ליגה	• קיים.	• קיים .	• קיים.
התראות על משחקים ואירועים	<ul style="list-style-type: none"> • אין אפשרות לשלוח הודעות מתוזמנות לעוקבי (ואוהדי) הליגה. • מאפשר שליחת הודעה ידנית לכל עוקבי (ואוהדי) הליגה שמקפיצה להם הודעת push. 	<ul style="list-style-type: none"> • מאפשר הגדרת התראות אוטומטיות לקבוצות ולעוקבים (ואוהדים) בהתאם לבחירת המשתמש (ניתן לקבל התראה באמצעות אימייל/הודעת push/שניהם ביחד). 	<ul style="list-style-type: none"> • מאפשר הגדרת התראות אוטומטיות לאוהדים על משחקים אך אין התראות לשופטים.
שיבוץ משחקים	<ul style="list-style-type: none"> • שיבוץ המשחקים נעשה באופן אוטומטי ורנדומלי או באופן ידני ע"י רשימה. (במידה והמשתמש בחר שיבוץ באופן אוטומטי רנדומלי המערכת משבצת משחקים בין כל שתי קבוצות בהתאם למספר המפגשים שאחראי הליגה בוחר) 	<ul style="list-style-type: none"> • שיבוץ המשחקים נעשה באופן ידני לבחירת המשתמש. עבור כל קבוצה מוצגות האפשרויות בלוח שנה. 	<ul style="list-style-type: none"> • שיבוץ המשחקים נעשה באופן ידני לבחירת המשתמש. עבור כל קבוצה מוצגות האפשרויות בלוח שנה.
בעלי תפקידים אפשריים במערכת	<ul style="list-style-type: none"> • שחקן. 	<ul style="list-style-type: none"> • שחקן. • שחקן מחליף. • מאמן. 	<ul style="list-style-type: none"> • שחקן. • עובד צוות.
נתונים אופציונליים לבעלי התפקיד.	<ul style="list-style-type: none"> • שם בעל התפקיד. • רמה של בעל התפקיד. 	<ul style="list-style-type: none"> • שם בעל התפקיד. • פרטים ליצירת קשר. • מין בעל התפקיד. • מספר חולצה. • תאריך יום הולדת. • תמונה. • הערות למאמן. 	<ul style="list-style-type: none"> • שם בעל התפקיד. • אימייל. • מספר חולצה.
עלות השימוש	<ul style="list-style-type: none"> • בגרסה הפשוטה – ללא עלות. • גרסת פרו – בעלות של 8 ₪ לחודש/22 ₪ 3ל חודשים/82 ₪ לשנה. 	<ul style="list-style-type: none"> • העלות החודשים נקבעת עפ"י מספר הקבוצות שהמשתמש מתחזק. עבור כל קבוצה 6\$. 	<ul style="list-style-type: none"> • עבור הגרסה הפשוטה – ללא עלות. • עבור גרסת הפרימיום 69\$ לחודש.
מספר קבוצות מקסימלי	<ul style="list-style-type: none"> • 40 קבוצות. 	<ul style="list-style-type: none"> • 1000 קבוצות. 	<ul style="list-style-type: none"> • 900 בגרסה הרגילה. • 2000 בקבוצת הפרימיום.
פלטפורמות שימוש	<ul style="list-style-type: none"> • Mobile 	<ul style="list-style-type: none"> • Mobile. • Web. 	<ul style="list-style-type: none"> • Mobile • web
ממשק משתמש	<ul style="list-style-type: none"> • ממשק שנתמך ב40 שפות. • חוויית משתמש טובה. על הפעולות נגישות ופשוטות לשימוש עם איורים שמסמנים את הפעולה הרצויה. • מתאפשר לנהל ליגה בלבד. • יצירת הליגה ויצירת הקבוצות נעשית באופן ברור וסידרתי (שלב אחר שלב) ולכן זה נוח, ברור והדרגתי למשתמש. 	<ul style="list-style-type: none"> • ממשק שנתמך ב-2 שפות (אנגלית או צרפתית). • חוויית המשתמש טובה אך השימוש מורכב (משתמש ממוצע לא יצליח להנות בריבוי הפונקציונליות של המערכת). • ממשק שמאפשר גם ניהול ליגה וגם ניהול קבוצה. • יצירת הליגה ויצירת הקבוצות לא נעשית באופן היררכי ולכן זה מקשה על הסדר, הבנה ותפעול המערכת. 	<ul style="list-style-type: none"> • ממשק שנתמך בשפה אחת אוניברסלית – אנגלית. • חוויית משתמש מסורבלת, מחסור באייקונים על המסך והמשתמש הממוצע יתקשה להתמצא באפליקציה. • ביצירת הקבוצה ניתנת אפשרות לבחירת צבע הקבוצה שייחודי למערכת הזאת.

יצירת דפים אישיים לקבוצות ולשחקנים	<ul style="list-style-type: none"> • אין דף אישי לכל שחקן. • יש אפשרות לבחור שם לשחקן , מספר חולצה, צבע החולצה נבחר אוטומטית. 	<ul style="list-style-type: none"> • יש דף אישי לכל שחקן. • לבחור צבע חולצה, שם, מספר חולצה. • יש אפשרות לשים תמונה לשחקן. • יש אפשרות לשמור תמונות למזכרת ממשחקים שבהם השחקן השתתף. 	<ul style="list-style-type: none"> • יש דף אישי לכל שחקן. • יש אפשרות לבחור שם, צבע קבוצה, מספר חולצה. • אפשר לשמור נתונים אישיים ומידע רקע על שחקן.
ניהול שחקנים	• לא קיים.	<ul style="list-style-type: none"> • אפשרות בחירה אם השחקן נמצא או לא. 	• לא קיים.
שיבוץ שופטים	• השיבוץ מתבצע אוטומטי.	• השיבוץ מתבצע אוטומטי.	• השיבוץ מתבצע אוטומטי.
התראות לאוהדים	<ul style="list-style-type: none"> • יש התראות לאוהדים על משחקים, תוצאות משחק, עדכונים חשובים. 	<ul style="list-style-type: none"> • יש התראות לאוהדים על משחקים, תוצאות משחק, עדכונים חשובים. 	<ul style="list-style-type: none"> • יש התראות לאוהדים על משחקים, תוצאות משחק, עדכונים חשובים.
מערכת המלצה לאוהדים	• לא קיים.	• לא קיים.	• לא קיים.
בקרה תקציבית	• לא קיים.	• לא קיים.	• קיים.
שיבוץ מגרשים	• לא קיים.	• לא קיים.	<ul style="list-style-type: none"> • אפשר לשבץ מגרשים לפי מיקום.
שיתוף פעולה עם חברות ביטוח ורפואה	• לא קיים.	• לא קיים.	• קיים.

טבלה מספר 1. ריכוז והשוואת פיצ'רים בין מערכות קיימות.

לסיכום, ניתן לראות כי מערכת Monclubsportif פחות רלוונטית מכיוון שהיא אינה תומכת בניהול ליגה אלא מתאימה יותר לניהול רק של קבוצות ושחקנים. המערכת sports engine היא המערכת הרלוונטית ביותר והייחודית ביותר מכיוון שהיא מספקת אפשרויות יוצאות דופן כגון חנות מזכרות אינטרנטית, שיתוף פעולה עם חברות ביטוח ורפואה, בדיקת רקע לחברי הצוות והיא יותר מעמיקה. המערכות שסקרנו לעיל עדיין אינן מאפשרות כיסוי מלא לניהול ליגה כיוון שאינן תומכות במערכות המלצה לעובדים, בבקרה תקציבית, בשיבוץ שופטים וצוותים הדרושים לניהול ליגה. ניתן לגשר על פערים אלו ע"י הוצאה של מערכת מתאימה שתוסיף את הפיצ'רים החסרים למערכות הקיימות.

מערכות קיימות לחיזוי תוצאות משחקים:

בחלק זה נתאר ונשווה בין מערכות לחיזוי תוצאות משחקי כדורגל. אנו בחרנו לסקור שתי מערכות – האחת Forebet והשנייה SoccerVista. מערכות חיזוי לתוצאות משחקים עוסקות בניית נתונים ממשחקי עבר ובעזרתם לחזות תוצאות משחקים עתידיים.

1. Forebet – אתר זה הופיע ראשון כמערכת החיזוי לתוצאות משחקים בחיפוש באתר גוגל. האתר מאוד נוח לשימוש למשתמשים חדשים, הוא מכיל הרבה פיצ'רים ואיורים ומציג מידע אודות שיטות החיזוי והכלים בהם הוא נעזר כדי לבצע חיזוי. למשל: התפלגות פואסונית, משולש פסקל לחישוב הסתברויות, קריטריון קלי, שימוש ב-databases המכיל נתוני עבר של הקבוצות על פני יותר מ-400 ליגות ב-15 שנים האחרונות.
2. SoccerVista – אתר זה הינו אתר פשוט יחסית והוא אינו מכיל פונקציונליות רבה ומסובכת. האתר נוח לשימוש ומתמקד בחיזוי של תוצאות משחקים ומציג יחס הימורים ותוצאות חמשת המשחקים האחרונים עבור כל קבוצה. התחזיות נוצרות על ידי אלגוריתם מתמטי מורכב וייחודי שהם אינם מפרסמים, המבוסס על נתוני כדורגל סטטיסטיים אשר מוצגים לאחר מכן למשתמשים.

SoccerVista	Forebet	קריטריון להשוואה
• Web.	• Web. • Mobile.	פלטפורמות בהן נתמך
• עולמי.	• עולמי. • מעל 400 ליגות.	מדינות בהן ניתן להשתמש
• 1999	• 2009	שנת הקמת האתר
• חינמי	• חינמי	עלות שימוש
• האם יהיו מעל או מתחת 2.5 שיערים במשחק. • האם הקבוצה תנצח במשחק או שיהיה תיקו.	• האם יהיו מעל או מתחת 2.5 שיערים במשחק. • האם שתי הקבוצות יבקייעו במשחק העתידי. • האם הקבוצה תנצח במשחק או שיהיה תיקו.	נתונים נגזרים
• חיזוי תוצאת משחק סופית.	• חיזוי של תוצאת משחק סופית. • חיזוי של תוצאת מחצית המשחק. • למי מהשחקנים המשתתפים במשחק יש סיכוי להבקיע ומה הסיכוי.	פיצ'רים נחזים מרכזיים
• הצגת יחסי ההימורים.	• המלצות למשתמש בזמן אמת. • תנאי מזג האוויר במיקום בו מתקיים המשחק. • פירוט אודות אילו שחקנים משתתפים, אילו מושעים ואילו שחקנים פצועים. • הצגת יחסי ההימורים.	האינפורמציה שמוצגת עבור כל משחק
• האתר נתמך בשפה האנגלית בלבד. • ניתן לחפש באתר רק עפ"י תאריך וליגה. • מבחינת חוויית משתמש קל לתפעל את האתר והוא מאוד פשוט להבנה, יש הסבר כללי על האתר וגם שאלות ותשובות נפוצות.	• תמיכה ב-10 שפות שונות. • סרגל חיפוש באופן חופשי. • אופציה לסימון משחקים המועדפים על המשתמש. • מציג בדף הראשי באופן אוטומטי את טבלת הליגה הבכירה המקומית שמתבססת על פי מיקום המשתמש. • האתר נגיש וקל להבנה עבור משתמשים חדשים ממוצעים.	ממשק משתמש ונראות האתר
• האתר מציג את הליגות הבכירות בעמוד הראשי. • האתר מציג את התוצאות של חמשת המשחקים עבור כל קבוצה שמשחקת ביום הנוכחי במסך הראשי. • האתר מציג קישורים לאתרי הימורים חיצוניים שבהם אפשר להמר באמצעות הידע שהמשתמש קיבל באתר.	• האתר מציג כתבות הקשורות אליו, לחיזויים שלו. • האתר מציג קישורים לפרופיל האתר ברשתות חברתיות כגון: טוויטר, יוטיוב ופייסבוק. • האתר מאפשר השוואה בין שתי קבוצות שהמשתמש בוחר ומציג עבורו סטטיסטיקות.	פונקציונליות נוספת

טבלה מספר 2. ריכוז והשוואות פיצ'רים בין מערכות חיזוי קיימות.

בחינת תוצאות חיזוי: אנו בחרנו לבדוק את איכות יכולות החיזוי של המערכות שסקרנו. נבחן את מחזורי המשחקים עבור הליגה האנגלית, ונבצע השוואה בין תוצאות הניבוי של שני האתרים אל תוצאות המשחק שקרה בפועל (קבוצה מנצחת בלבד).

SoccerVista	Forebet	קריטריון
3 מתוך 7 חיזויים מוצלחים עבור קבוצה מנצחת	3 מתוך 7 חיזויים מוצלחים עבור קבוצה מנצחת	בחינת תוצאות חיזוי

טבלה מספר 3. בחינת תוצאות החיזוי ובידיקת אופן הדיוק של האתרים.

FOREBET
Football is mathematics

HOME PREDICTIONS TRENDS MATCH PREVIEWS ARTICLES LIVESCORE TEAM COMPARISON INJURED PLAYERS TENNIS

Ads by Google

Free football predictions and tips for England

Home team	Away team	Probability %	Forebet	Correct score	Avg. goals	Weather conditions	Odds	Score	Kelly Criterion
1	X	2							
Premier League - Round 31									
Liverpool	Watford	71 15 14	1	3-0	3.21	8°	1.13	FT 2-0 (1-0)	
Brighton	Norwich City	42 26 33	1	2-0	2.45	8°	1.50	FT 0-0 (0-0)	
Burnley	Manchester City	14 21 66	2	0-3	2.66	8°	1.22	FT 0-2 (0-0)	
Chelsea	Brentford	66 22 13	1	2-1	2.50	8°	1.36	FT 1-4 (0-0)	
Leeds United	Southampton	34 26 40	2	1-2	2.90	8°	2.88	FT 1-1 (1-0)	0.08

FEATURED MATCH: Sheffield United vs QPR

Acrobat Pro DC

תמונה 4 – בחינת תוצאות חיזוי- מסך חיזוי, אתר Forebet

SoccerVista - football betting

Choose soccer league for 02-04-2022 and you will find statistics, picks, tables and information for all your betting needs. All football information on this site is free. You can choose a football game by date or select league from the country list. For detailed information about match click into score link.

Main page Bet of the day Picks Telegram Livescore Bookmakers FAQ

Soccer games by date: 02-04-2022

England Premier League	Odds	Tips
1 X 2 1X2 Goals Score		
12:30 Liverpool vs Watford	1.14 8.5 15	1 O 3:0
15:00 Brighton vs Norwich	1.5 4.2 7	1 U 1:0
15:00 Burnley vs Man City	15 6.5 1.2	2 O 0:3
15:00 Chelsea vs Brentford	1.36 4.75 8.5	1 U 1:0
15:00 Leeds vs Southampton	2.4 3.75 2.7	X2 O 2:2

Top Leagues: England, Spain, Germany, Italy, France, International, Netherlands, Portugal, Russia

1XBET

תמונה 5 – בחינת תוצאות חיזוי- מסך חיזוי, אתר SoccerVista

סקירת ספרות

משחר ההיסטוריה ספורט תמיד היווה חלק בלתי נפרד מחייהם של בני האדם. הספורט היה למטרות כושר, תרבות ובידור. בשלב מסוים, רצונם של אנשים להיות חלק ממשי מאירועי הספורט הלך וגבר ולכן החלו בניסיון לחזות את תוצאותיהם של אירועי ספורט שונים [11].

במרוצת השנים, ענף הספורט שהפך למאוד פופולרי הוא ענף הכדורגל. 4.7 מיליארד אנשים ברחבי העולם צופים על בסיס שנתי במשחקי ליגה אנגלית. תעשיית ההימורים סביב ענפי הספורט הפכה פופולרית, גדלה ורווחיהן של מועצות ההימורים במדינות רבות בעולם גדלו באופן משמעותי. למשל, מחזור הכנסת מועצות ההימורים בישראל, נכון לשנת 2018, היו כ-2.93 מיליארד שקלים [12].

בשנים האחרונות, ענף תחומי מדעי המחשב ככלל וענף תחומי הבינה המלכותית ולמידת המכונה בפרט התפתחו בצורה דרמטית יחד עם התפתחות הטכנולוגיה. ענפים אלו עוסקים בפיתוח היכולת של מחשבים לדמות את הבינה האנושית ולהעניק למחשבים תכונות, יכולות למידה ודרכי חשיבה עצמאית ע"י תכנותם באמצעות אלגוריתמים. אחת המשימות הנפוצות בלמידת מכונה היא משימת הסיווג (Classification), אשר מטרתה לחזות משתנה מטרות שלא נראה בעבר בעזרת בניית מודל סיווג המבוסס סט אימון וסט בדיקה. אנשים מסביב לעולם הבינו כי ביכולתם להשתמש באלגוריתמים אלו על מנת לחזות תוצאות אפשריות של משחקי ספורט, ולתעל זאת לרווחתם האישית. יותר ויותר אתרים החלו להציע תוכנות ועמודי אינטרנט שיכולים לחזות עפ"י נתוני עבר את תוצאותיהם של משחקים עתידיים [3].

במשחקי ספורט ככלל ובענף הכדורגל בפרט, יש מספר גדול של פרמטרים העשויים להשפיע על תוצאותיו של משחק. אדם ממוצע לרוב לא יצליח לפתור בעיות המושפעות מגורמים רבים, בניגוד לאלגוריתמי למידת מכונה אשר מסוגלים להתמודד עם כמויות עצומות של מידע ולתעל אותו לביצוע חיזוי [12], [11]. אלגוריתמי למידת מכונה מאופיינים באחוזי דיוק גבוהים לעומת ניסיונות החיזוי של מומחים רבים בתחום. ניתן לראות את בעיית חיזוי התוצאות של משחקי ספורט כניצחון, תיקו או הפסד כבעיית סיווג. יש הסבורים כי בעיה זו הינה בעיית רגרסיה – ניסיון לחזות את הפרש התוצאות בין ניקוד הקבוצה המנצחת למפסידה (הפרש מספר הנקודות). ברוב עבודות המחקר שסקרנו התייחסנו לבעיה זו כבעיית סיווג עם שלוש מחלקות, ואני נפרט את המודלים והאלגוריתמים העיקריים בהם השתמשו בעבודות המחקר שסקרנו.

Classification הינו תהליך המתקשר לקטגוריזציה שבו אובייקטים מזוהים ומשווים לקבוצות מסוימות עפ"י קריטריונים שנבדקים, הן ע"י אלגוריתמים שלמדו סט נתונים שלפיו יש לקטגוריה, והן באופן ידני [3]. שיטה אפקטיבית ומשתמשים בה הרבה בלמידת מכונה היא שיטה הנקראת XGboost. שיטה זו היא open source (חינמית) וסקלאבילית (יכולה להתמודד עם מספר רב של מידע בלי לצרוך יותר מדי משאבים כמו זיכרון) ומשתמשים בה בשביל לשפר תוצאות סיווג ולמנוע מצב של overfitting (מצב בו המודל מותאם יותר מדי training set ובגלל זה מצליח פחות בביצוע תחזית). השיטה מבצעת feature selections (בחירת תת קבוצה של פיצ'רים מתוך המודל) Leaf pruning שזה חיתוך של העץ בכדי לא לחפש מידע בעלים לא רלוונטיים והיא ניתנת להרצה באופן מקבילי. השיטה יכולה לטפל במצבים בהם המידע חסר, ובכל איטרציה שלה ניתן לבצע cross validation (שיטה בשביל לאבחן overfitting וחוסר זיהוי תבנית) כך שמספר האיטרציות יהיה אופטימלי [6].

אחד מבין אלגוריתמי למידת המכונה הפופולריים הוא אלגוריתם ANN, שהוא מודל מתמטי שפותח כרשת נוירונים מלאכותית המחקק את פעולת המוח האנושית ותהליכים קוגניטיביים, ובכך ניתן ללמד את המודל סיווגים חדשים עפ"י נתונים אחרים שלמד בעבר. מודל זה מורכב מיחידות רבות של מידע המקושרות ביניהן ומכיל שלוש סוגים של שכבות: Input layer – מידע התחלתי לרשת, Hidden layer – שכבה מתווכת ו- Output layer – הפלט. הבעיה שעולה מאלגוריתם זה היא בעיית זיהוי תווי קול שבוצעו בתיאום עם מערכות הטלפון הקווי. שימוש במודל השכבות החבויות הוא פתרון המתמודד עם הבעיה הנוצרת. מודל שכבות מורכב משכבת קלט אחת, שכבת פלט אחת ושתי שכבות נסתרות. בשכבת הקלט ישנם 19 נוירונים אשר כל אחד מהם משתמש בפונקציית העברה לינארית [10].

סוג נוסף של רשת נוירונים הוא רשת Bayesian. רשת זו הינה מודל גרפי הסתברותי שמייצג סט פיצ'רים בצורת גרף חסר מעגלים, על ידי ביטוי הקשרים ביניהם ביחס לאם הפיצ'רים תלויים אחד בשני או לא. נשתמש במודל זה כאשר נקבל סט של פיצ'רים, ועל ידי המידע המתקבל מסט זה, מנסים לבצע חיזוי. בעת ביצוע משימת חיזוי ברשת הביסיאנית החלק המרכזי הינו לבחור את הפיצ'רים שלפיהם נאמן את המודלים השונים. אחת הבעיות העיקריות של רשת זו היא שהרשת לא מציגה מודל מובן וברור למתבונן, ובשימוש בה לחיזוי תוצאות כדורגל היא מתמקדת בחיזוי תוצאות סופיות של המשחק בלבד. לכן, עיקרון חדשני נוסף המוצג הוא השאיפה לבצע חיזויים בזמן אמת, כלומר לתת גישה למאמן הקבוצה לתוצאות החיזוי של המודל במהלך מחצית המשחק, ובעקבות כך לאפשר למאמן להחליט האם ברצונו לשנות את טקטיקת המשחק של הקבוצה שלו [1]. בתהליך זה החלו להשתמש בפיצ'רים נוספים חדשים מלבד חיזוי תוצאות המשחק, כגון: מספר בעיטות לשער, מספר עבירות, מספר הקרנות הממוצע ומספר בעיטות ממוצע [1]. חידוש נוסף לרשת בייסיאנית הוא Hybrid Bayesian networks.

זהו מודל בייסיאני היברידי, והוא רשת גרפית בדומה למודל הביסיאני הרגיל, אך בעל שינויים בתכונות המשתנים - קודקודי הגרף. משתני הקודקודים בגרף במודל זה הם מתמשכים ובדידים. דבר שנותן לרשת טווח יותר גדול של שימוש [2].

לעומת אלגוריתמי רשתות נוירונים [10], [1], אלגוריתם נוסף המשמש לסיווג או חיזוי הוא אלגוריתם ID3 שהוא אלגוריתם היוצר מודלים של עצי החלטה. עיקר השימוש בו הוא לסיווג או חיזוי ערכו של משתנה מטרה נבחר, והפלט שלו מספק מודל הנבנה בעזרת סט האימון, כאשר הפיצ'רים (צמתי הפיצול) בעץ מסודרים לפי סדר החשיבות של המידע החדש והתועלת שנוויה מכל אחד מהם, ולפיהם ניתן לחזות מה התוצאה הטובה ביותר עבור המשתמש. המודל מוצג באופן מובן, קריא, היררכי וברור למשתמש, כך שניתן לראות את פיצול המידע לקטגוריות [7]. בנוסף, צריכת הזיכרון נמוכה. לצד היתרונות של אלגוריתם זה קיימים מספר חסרונות, הוא אינו יכול לחול על משתנים רציפים בגדלים אינסופיים, לרוב אינו יציב ונוטה לשגיאות (מטופל ע"י תחזוקה), ועל כל שינוי קטן בערכים הנצפים או במדגם הנתונים יש נטייה ליצור הפרעות בחילוף מידע. חסרון נוסף חוסר היכולת להתמודד עם ערכים חסרים, אלגוריתם המתמודד עם חיסרון זה הוא id3+ שמתמודד עם ערכי תכונות חסרים מטופלת בדרך כלל על ידי מערכות למידה מצטברות [9]. שימוש נוסף למודל הוא בתחום חיזוי תוצאות הכדורגל [8]. המחקר עוסק ביצירת מודל חדשני שכותביו יצרו, המתבסס על 2 חלקים עיקריים. החלק הראשון במודל הוא מערכת דירוג שנוצרה כחלק מהמודל, אשר מדרגת כל קבוצה ביחס לקבוצות האחרות בליגה בעזרת הנתונים מה- Database ופרמטרים נוספים, והחלק השני של המודל הוא Hybrid Bayesian Networks שמקבל את הדירוג מהחלק הראשון ומייצר את התפלגות ההסתברות של שלושת התוצאות האפשריות למשחק. יתרון משמעותי של מודל חדשני זה, הוא היכולת לחזות תוצאות משחק של קבוצות ממדינות מסוימות על ידי שימוש בנתונים על משחקים קודמים בהם השתתפו קבוצות ממדינות אחרות [8]. קושי הנובע משימוש בעצי ההחלטה הוא שעצי ההחלטה אינם מודל יציב, אלא כל שינוי, ולו הקטן ביותר, יכול לגרום לפגיעה באופטימליות של האלגוריתם [7]. מחקר חדשני נוסף הפותר בעיה זו ובעיות נוספות שנוצרות מעצי החלטה כגון overfitting, שונות, חוסר דיוק וחוסר במידע ונתונים הוא מחקר המציע אלגוריתם חדש ששמו Random Forest, והוא נפוץ בפתרון בעיות שונות בלמידת מכונה כגון סיווג ורגרסיה [4]. אופן הפעולה של האלגוריתם הוא להריץ מספר עצי החלטה שונים, כך שכל עץ מקבל מידע אחר ורץ בצורה שונה, ולבסוף הפלט הוא או ממוצע של הפלטים או פלט לפי רוב הפלטים.

מאמר נוסף שמדבר על classification הוא מאמר שמציג אלגוריתם נוסף בתחום זה והוא אלגוריתם KNN [3]. אלגוריתם זה הוא אלגוריתם סיווג או רגרסיה פשוט יחסית לשאר האלגוריתמים שהצגנו עד כה בתחום למידת מכונה, והוא מתבסס על k התצפיות הקרובות ביותר במרחב הפיצ'רים. זהו אלגוריתם לימוד מבוסס מופעים אשר לרוב עושה שימוש במטריקת המרחק האוקלידי. אחת הבעיות העיקריות של אלגוריתם KNN היא בעיית הממדים הרבים שנוצרים ומאטים את זמן הריצה, וכמו כן, גם דיוק הסיווג עלול להיפגע עקב רגישות האלגוריתם למבנה המקומי של הנתונים, או מהיווצרות של רעש והתחשבות בפיצ'רים לא רלוונטיים [3]. הפתרון לכך שהוצע במאמר על ידי כותביו הוא Feature selection. בשיטה זו נבחר את המידע הכי חשוב ורלוונטי לפי מטרת המחקר וציפיות הפלט מהאלגוריתם, גודל dataset יקטן ולא יכלול פיצ'רים שעלולים לפגוע בדיוק או ברלוונטיות התוצאות. החוקרים השתמשו בשתי דרכים לבחירת המידע הרלוונטי. האחת נקראת פעולת Filter והיא מעבר על המידע על מנת למצוא את המחלקות הכי רלוונטיות, ולסוף מחלקות אלו יישמשו אותנו כדי לאמן את המודל. הדרך השנייה היא שיטת Wrapper בה מנסים להשתמש subset של features ולאמן את המודל באמצעותם, ולבסוף בעזרתם נחליט האם להוסיף או להסיר features מהsubset עד שנגיע למצב רצוי של features. בבעיה המתוארת במאמר השתמשו בסט נתונים של 1000 אימיילים, מתוכם 500 ספאם ו-500 תקינים. חידוש נוסף הוא אלגוריתם Weighted KNN using genetic algorithm. אלגוריתם זה פותר בעיה נוספת שנוצרת בבחירת K התצפיות ב-KNN. בעת בחירת הקלט, אם K קטן מידי ביחס למספר התצפיות שנבדקו אז האלגוריתם יהיה רגיש לנתונים חדשים ולא יצליח לבצע סיווג לקלסטר מסוים, ואם K יהיה גדול מידי ביחס למספר התצפיות, אז יהיו מספר גדול של נקודות ממחלקות אחרות. אלגוריתם זה הוא שיטת חיפוש סטוכסטית (שיטת חיפוש לבעיות שקיימות אצלן אקראיות במדידות שניתנות לאלגוריתם ויש אקראיות גם באלגוריתם עצמו) שהשראתו נלקחה מהגנטיקה של האבולוציה. האלגוריתם מדמה natural selection ככה שהמידע הכי רלוונטי נשאר ומשוכפל [5].

מקריאת המאמרים שבחרנו, בלט בעיקר הגורם היצירתי והגישות המגוונות בהיבט בחירת הפיצ'רים במודל. לאחר דיון מעמיק בנושא, לדעתנו יש עוד פוטנציאל מחקרי בתחום חיזוי תוצאות משחקי כדורגל. הסיבה העיקרית לכך היא שאחוזי הדיוק של המודלים הקיימים כיום לא עולים בהרבה על יכולות חיזוי של מומחים אנושיים בתחום. בנוסף, ברוב עבודות המחקר שהתייחסו לחיזוי תוצאות במשחקי ספורט והתמקדו בחיזוי תוצאות בכדורגל, ולא במחקרים שהתעסקו בשיטות סיווג ובאלגוריתמי חיזוי באופן כללי, שמנו לב כי החוקרים וכותבי המאמר התייחסו לקבוצה כישות אחת ובהתאם לכך גם על כלל הנתונים שקשורים אליה. אנו מציעים להסתכל על קבוצת כדורגל גם כסט של שחקנים ולא כישות אחת בלתי ניתנת לפירוק, כך שניתן יהיה לבחון את מרכיבי הקבוצה ולחקור את תרומתם של פיצ'רים שונים שרלוונטיים לשחקנים עצמם. דוגמאות לפיצ'רים כאלה הן - עמדה במגרש, כושר גופני של השחקן (פציעה עלולה לפגוע ביכולת), היבטים פסיכולוגיים, כמות השערים במשחק, כמות בישול שערים לשחקן ועוד.

השוואה בין מאמרים

מאמרים	הבעיה שהאלגוריתם פותר	איך האלגוריתם פותר את הבעיה	סוג הנתונים
Decision tree	גוזר את המסקנה של אירוע באמצעות סדרה של רגרסיה וסיווג	צומת השורש מתפצל באופן רקורסיבי לצמתי החלטה בצורה של ענפים או עלים בהתבסס על כמה נהלי למידה מוגדרים על ידי משתמש או אוטומטיים	משתנים קטגוריים
Decision Trees for the Prediction of Outcome of Soccer Games	ניבוי תוצאות משחקי כדורגל וניתוח השונות באחוז הניצחון של קבוצות ביתיות במדינות שונות	בחינה של מידע על ביצועים קודמים והיחסים ביניהם מוערכים באמצעות למידת מכונה כדי להעריך את המודלים	
Improved Decision Tree Algorithm: ID3+	פתרון בעיית חוסר היכולת להתמודד עם ערכי תכונות לא ידועים, ופתרון הבעיה ש ID3 מחמיץ הזדמנויות ביישומים מורכבים בעולם האמיתי	התמודדות עם ערכי תכונות חסרים מטופלת בדרך כלל על ידי מערכות למידה מצטברות, משנים רק חלקים 'פגומים' מבסיס הידע הנלמד כדי להתאים לתצפיות החדשות מסתמך על ייצוגים הסתברותיים ואסטרטגיית תיקון וכך מתמודד עם רעש (דרישות של יישומים בעולם האמיתי)	משתנים קטגוריים
Neural Network Solution for Secure Interactive Voice Response	משמש לפתירת משימות זיהוי, חיזוי, אופטימיזציה ועוד משימות מגוונות	באמצעות מודל של נירונים ושכבות אשר מדמה את חיבור הנירונים שבמוח	תווי קול בתיאום עם מערכת הטלפון הקווי.
Bayesian networks	חיזוי, דיאגנוזה גילוי אנומליות	מייצג קשרים בין משתנים תלויים ובלתי תלויים, משתמש במשפט בייס	נתונים של משחקי כדורגל כגון שערים.
Hybrid Bayesian networks	חיזוי, דיאגנוזה גילוי אנומליות ועוד. רק שהמשתנים יכולים להיות גם מתמשכים ובדידים	מייצג קשרים בין משתנים תלויים ובלתי תלויים (עובד גם אם המשתנים מתמשכים ובדידים), משתמש במשפט בייס	נתונים של משחקי כדורגל כגון שערים
KNN	classification and regression	מחשב את המרחק בין הדגימה לבין כל שאר הנתונים ומוצא את ה"דומה" ביותר	dataset of 1000 emails, 500 spam and 500 legitimate where word features only were used.
weighted KNN using genetic algo	לתת פתרונות איכותיים לבעיות חיפוש ואופטימיזציה שימשו לclassification	משמט מידע לא רלוונטי ומשאיר את המידע הרלוונטי, מריץ איטרציות רק על המידע הרלוונטי	breast diabetes glass waveform
Random forest	classification and regression	בונה עצי החלטה על מדגמים שונים ובסוף בוחר את הממוצע או את התוצאה השכיחה ביותר	Ecology, medicine, Astronomy, autopsy, traffic and transport, agriculture, bioinformatics
XGboost	classification, regression and ranking problems	כמו random forest אבל משפר את זמן הריצה ומבצע gradient boosting	Allstate, Higgs boson, Yahoo LTRC, Criteo

- [1] Razali, N., Mustapha, A., Yatim, F. A., & Ab Aziz, R. (2017, August). In IOP conference series: Materials science and engineering (Vol. 226, No. 1, p. 012099). IOP Publishing.
- [2] Anthony C. Constantinou (2020, March). Asian Handicap football betting with Rating-based Hybrid Bayesian Networks.
- [3] k-Nearest Neighbour Classifiers P´adraig Cunningham and Sarah Jane Delany
- [4] Random forests, article: Khaled Fawagreh, Mohamed Medhat Gaber & Eyad Elyan (2014)
- [5] Weighted K-Nearest Neighbor Classification Algorithm Based on Genetic Algorithm Xuesong Yan , Wei Li , Wei Chen , Wenjing Luo , Can Zhang , Qinghua Wu, Hammin Liu
- [6] XGBoost: A Scalable Tree Boosting System Tianqi Chen University of Washington
- [7] Decision Tree by Rajendra Mohan Panda and B S Daya Sagar
- [8] Decision Trees for the Prediction of Outcome of Soccer Games - Historical Data Analysis by Ercilia de Stefano, Leonardo de Oliveira Farroco, Gilson Brito Alves Lima, Luiz Octávio Gavião, Vitor Ayres Principe and Annibal Parracho Sant'Anna
- [9] Improved Decision Tree Algorithm: ID3+ by Min Xu, Jian-Li Wang & Tao Chen
- [10] Neural Network Solution for Secure Interactive Voice Response Syed Ayaz Ali Shah, Azzam ul Asar and S.F.
- [11] Fialho, G., Manhães, A., & Teixeira, J. P. (2019). Predicting sports results with artificial intelligence—a proposal framework for soccer games. Procedia Computer Science, 164, 131-136.
- [12] Ulmer, B., Fernandez, M., & Peterson, M. (2013). Predicting soccer match results in the English premier league (Doctoral dissertation, Doctoral dissertation, Ph. D. dissertation, Stanford).