



מוביליות חברתית-כלכלית בין-דורית: מודל ניבוי

עופר פיינשטיין אורן הלר
מיכל גרינשטיין-וייס

צוות ג'ונט-אשלים למחקר:
נועה בן-דוד, יואב בוקעי, רוני ליאור

עריכת לשון: רויטל אביב מתוק
עיצוב גרפי: ענת פרקו טולדנו

המחקר הוזמן על ידי ג'וינט-אשלים והמשרד לשוויון חברתי וקידום מעמד האישה ומומן בסיועם.

מכון מאירס ג'וינט ברוקדייל

ת"ד 3886 ירושלים 9103702

טלפון: 02-6557400

brook@jdc.org | brookdale.jdc.org.il

ירושלים | שבט תשפ"ו | כברואר 2026

תקציר

דוח זה מציג מודל ניבוי חדשני המבוסס על נתוני עתק (Big Data) ובינה מלאכותית, שנועד לחולל שינוי תפיסתי: מעבר ממדידה רטרוספקטיבית של מוביליות חברתית-כלכלית המביטה אל העבר, לחיזוי פרוספקטיבי הצופה פני עתיד. המודל פותח על בסיס נתונים מנהליים שהונגשו בחדר המחקר של הלשכה המרכזית לסטטיסטיקה, והוא מעריך את סיכויי ההצלחה של צעירים בעתיד ומגשר על פער הזמנים המובנה במדידה המסורתית. גישה זו מאפשרת להעריך את פוטנציאל המוביליות עוד בשלבי ההתפתחות הקריטיים של הצעירים, ובכך הופכת את תובנות המחקר לשימות עבור הדור הנוכחי. המודל נועד לשמש תשתית לכלי תומך החלטה, המאפשר למעצבי מדיניות לפעול בזמן אמת ולהשפיע על מסלולי החיים בעודם מתהווים.

הניתוח מתמקד בחיזוי מוביליות בהכנסה ובהשכלה ובזיהוי סיכונים ל"רצפה דביקה", תוך כדי בחינת משקלם של משתני רקע וסביבה. הממצאים חושפים תמונה מורכבת: בעוד מאפיינים דמוגרפיים, ובראשם מגדר והשתייכות לקבוצת אוכלוסייה, נותרים מנבאים חזקים המשקפים פערים מבניים, לגורמים סביבתיים הניתנים להשפעה, כגון איכות ההוראה והרכב הכיתה, יש כוח ניבוי רב שעשוי לשמש מנוף לשינוי. הדוח מדגים כיצד אימוץ גישה של התערבות מקדימה תאפשר למעצבי המדיניות לעבור מניהול תגובתי לניהול יוזם, המזהה סיכונים והזדמנויות בזמן אמת וממקד משאבים והתערבויות בצמתים הקריטיים ביותר לקידום שוויון הזדמנויות.

תמצית

רקע

מוביליות בין-דורית והמעבר ממדידה רטרוספקטיבית לחיזוי פרוספקטיבי צופה פני עתיד

מוביליות בין-דורית היא יכולתם של יחידים לשפר את מעמדם החברתי-כלכלי לעומת זה של הוריהם, והיא מדד ליבה להערכת שוויון הזדמנויות בחברה. בישראל, המאופיינת בשונות רבה ובפערים חברתיים עמוקים על רקע מוצא, קבוצה באוכלוסייה ומיקום גאוגרפי של מגורים. הבנת המנגנונים הקשורים למוביליות היא סוגיה ראשונה במעלה בתחום המדיניות ובעיצוב פרקטיקות חדשניות לקידום שוויון הזדמנויות.

עם זאת, המאמץ הממשלתי והחברתי לקידום מוביליות נתקל בחסם מתודולוגי מובנה: המדידה הרווחת היא רטרוספקטיבית במהותה (מביטה לאחור). כלומר, אפשר לאמוד את מידת המוביליות רק בדיעבד, בחלוף דור, כאשר כבר אפשר למדוד את תוצאותיהם הכלכליות של הפרטים בבגרותם, לרוב בעשור הרביעי לחייהם, בתקופה שבה מסלול הקריירה מתגבש וכן מתאפשרת צבירת הון. פער זמנים זה מקשה על קובעי המדיניות לזהות בזמן אמת ילדים ובני נוער המצויים בסיכון לסטגנציה (קיפאון) כלכלית, ומוביל לרוב למדיניות המגיבה לאתגרי העבר במקום למנוע אותם מראש.

דוח זה מציג היתכנות למעבר לגישה פרוספקטיבית (צופה פני עתיד). השימוש במודלים מתקדמים לניבוי, המבוססים על נתוני עתק (Big Data), מאפשר להעריך את סיכויי ההצלחה של הדור הצעיר בעתיד על בסיס נתונים קיימים, ובכך לגשר על פער הזמנים ולהפוך את המוביליות ליעד בר מדידה כבר בהווה.

מטרת המחקר

הנחת יסודות לשינוי התפיסה

מטרתו המרכזית של המחקר היא לבחון את ההיתכנות של פיתוח יכולות ניבוי בשדה המדיניות החברתית ועיצוב פרקטיקות. הדוח מציג **אב-טיפוס** של מודל ניבוי למוביליות, אשר פותח על בסיס נתונים מנהליים רחבי היקף של הלשכה המרכזית לסטטיסטיקה.

המודל המוצג הוא מתווה ראשוני שמדגים וממחיש כי פיתוח יכולות ניבוי מתקדמות בתחום המוביליות הבין-דורית הוא בר השגה. המחקר מציג תשתית מתודולוגית חדשנית ומדגים את הערך של שימוש בנתונים קיימים כדי לסמן את הכיוון לפיתוח כלי עבודה יישומיים שיאפשרו להתקדם בשלושה מישורים:

- 1. זיהוי מוקדם:** איתור קבוצות באוכלוסייה ומרחבים גאוגרפיים המצויים בסיכון למוביליות נמוכה ("רצפה דביקה").
- 2. מיקוד משאבים על בסיס מנבאי מוביליות:** שימוש בממצאים על אודות משתנים בעלי כוח ניבוי להגדרת התחומים והקבוצות באוכלוסייה שהשקעה נוספת בהם עשויה להניב את התשואה החברתית הגבוהה ביותר.
- 3. תכנון התערבויות:** העמקת בסיס הידע המאפשר זיהוי מענים מובחנים ומדויקים המותאמים לשילוב המשתנים המשפיעים על מוביליות בקרב קבוצות שונות באוכלוסייה.

שיטה

המחקר מתבסס על מסד נתונים מקיף הכולל את ילידי השנים 1970–1987: קובצי רשומות פרט שהוכנו לצורך ביצוע המחקר על ידי הלשכה המרכזית לסטטיסטיקה ומהם הושמטו פרטי זיהוי. לצורך אימון המודל, התמקד הניתוח בשלישון התחתון של התפלגות הכנסות ההורים, קבוצה הכוללת כ-430,000 פרטים. עבור אוכלוסייה זו נבחנו משתנים מסבירים מגוונים, לרבות נתונים דמוגרפיים, נתוני רקע משפחתיים (השכלה והכנסה), מאפייני בית הספר ומשתנים ייחודיים כגון מאפייני איכות צוות ההוראה וההרכב החברתי-כלכלי של הכיתה.

המודל מתמקד בחיזוי שלושה משתני תוצאה:

1. מוביליות כלפי מעלה בהכנסה: הסיכוי של ילד שגדל עם הורים שהכנסתם בשלישון התחתון, להגיע בבגרותו לשלישון העליון של התפלגות ההכנסות.

2. מוביליות כלפי מעלה בהשכלה: הסיכוי של ילד שגדל עם הורים שהכנסתם בשלישון התחתון, לרכוש השכלה גבוהה.

3. היעדר מוביליות ("רצפה דביקה"): הסיכוי של ילד שגדל עם הורים שהכנסתם בשלישון התחתון, להישאר בשלישון התחתון גם בבגרותו.

הניתוח בוצע באמצעות שילוב של שתי גישות סטטיסטיות:

1. רגרסיות לוגיסטיות: שימשו לזיהוי המגמות הכלליות ולאמידת הקשר הממוצע והישיר בין כל משתנה ובין סיכויי המוביליות.

2. עצי החלטה: שימשו לזיהוי הקשרים ייחודיים (אינטראקציות). שלא כמו רגרסיה, מודל זה מזהה כי השפעתו של גורם מסוים אינה אחידה לכל הילדים, אלא **תלויה בהקשר** הסביבתי והמשפחתי של הילד.

באמצעות שתי הגישות הללו יישמנו מודלים לניבוי כל אחד ממשתני התוצאה שלעיל. עבור כל פרט המודלים הניבו ערך מספרי כנגד ההסתברות שנאמדה למוביליות (בניבוי מוביליות כלפי מעלה בהכנסה או בהשכלה) או היעדרה (בניבוי היעדר מוביליות). בשלב הבא, ההסתברות שנאמדה הומרה לערך בינארי (1 או 0) על בסיס ערך גבולי מוגדר, כאשר ערכי ההסתברות הגבוהים מן הערך הגבולי הומרו ל-1, וערכים נמוכים ממנו הומרו ל-0. **בשתי השיטות, ניתוח הממצאים כלל התבוננות בתרומתו של כל משתנה מנבא לתוצאת הניבוי.**

ממצאים מרכזיים

ממצאי המודלים חושפים מציאות מורכבת שבה ישנם פערים מבניים עמוקים ובצידם גורמים סביבתיים בעלי כוח ניבוי רב.

1. מבנים חברתיים וריבוד.

הממצאים מראים שפערים בין המינים ובין קבוצות שונות באוכלוסייה הם מנבאים חשובים:

- **מגדר:** בשני המודלים נמצא כי **למשתנה "מגדר" ההשפעה הרבה ביותר** על תוצאות הניבוי עבור משתנה התוצאה של מוביליות כלפי מעלה בהכנסה. לנשים סיכוי נמוך יותר למוביליות כלפי מעלה בהכנסה לעומת לגברים, אף על פי שיש להן סיכוי גבוה יותר למוביליות כלפי מעלה בהשכלה.
- **קבוצה באוכלוסייה:** השתייכות לקבוצה באוכלוסייה נותרה מנבא חזק: ליהודים (שאינם חרדים) סיכויי המוביליות הגבוהים ביותר, לעומת ערבים וחרדים.

2. מנפי שינוי: כוחן של סביבת החינוך והמשפחה.

- לצד מאפיינים דמוגרפיים הקשורים במוביליות אך אינם ניתנים להשפעה ישירה באמצעות מדיניות, זיהינו במודלים גורמים נוספים שלהם כושר ניבוי גבוה ואשר עיצוב מדיניות מתאימה עשוי להשפיע עליהם:
- **איכות ההוראה:** למשתנים הנוגעים לאיכות המורים, כמו שיעור המורים בעלי השכלה גבוהה וותק המורים, יש קשר חיובי לסיכויי המוביליות כלפי מעלה.
 - **הרכב הכיתה:** הרקע החברתי-כלכלי של עמיתים לספסל הלימודים התגלה בתור גורם מנבא חשוב ביותר. למידה בכיתה שבה ממוצע הכנסות ההורים גבוה, קשורה בסיכויים גבוהים למוביליות כלפי מעלה.
 - **הון משפחתי:** השכלת ההורים (ובפרט השכלת האם) זוהתה בתור גורם מנבא חשוב להעלאת הסיכויים למוביליות כלפי מעלה.

3. אינטראקציות כמפתח להתערבות מדויקת.

- ערך מוסף עיקרי של מודל עצי ההחלטה הוא שימוש באינטראקציות מורכבות בין המשתנים. הממצאים מראים כי דפוסי הניבוי אינם אחידים, וגורמים מסוימים קשורים בעוצמה שונה בקבוצות שונות:
- **הקשר בין השכלת מורים לסיכויי המוביליות:** ניתוח הנתונים מצביע כי כושר הניבוי של משתנים הקשורים למערכת החינוך אינו אחיד, אלא תלוי הקשר חברתי-כלכלי. המודל מזהה כי הימצאותם של מורים בעלי השכלה גבוהה קשורה בסיכויי המוביליות הצפויים של התלמידים, בייחוד באשכולות חברתיים-כלכליים נמוכים. **ממצא זה מדגיש כי למערכת החינוך יש פוטנציאל גבוה יותר לשמש מנוף לשינוי מסלול החיים עבור אוכלוסיות מרקע חברתי-כלכלי נמוך.**
 - **גורמי חוסן באוכלוסיות מיעוט:** השכלת האם זוהתה בתור גורם מגן משמעותי בקרב פרטים מכלל הקבוצות באוכלוסייה; היא משנה את משוואת הסיכויים ומצמצמת את הפערים אל מול האוכלוסייה הכללית.

סיכום וכיווני פעולה

הדוח מדגים את ההיתכנות והנחיצות של מעבר למדיניות חברתית מבוססת נתונים הצופה פני עתיד. אב-הטיפוס שפותח מראה שאפשר להשתמש במידע הקיים כדי לאתר צומתי התערבות יעילים במסלול חייהם של ילדים וצעירים, ולנהל סיכונים חברתיים באופן מדויק יותר.

לנוכח הממצאים, מומלצים כיווני הפעולה האלה:

1. **שימוש במודל בתור כלי תומך החלטה:** המחקר הנוכחי מראה כי אפשר לפתח מודל ניבוי המבוסס על נתוני עתק מנהליים הקיימים ממילא במערכת. יש לבחון את שילובו של אב-הטיפוס בתהליכי העבודה השוטפים של גופי המטה, לא ככלי אוטומטי המקבל החלטות, אלא כתשתית תומכת החלטה המאפשרת מיפוי דינמי של מוקדי סיכון והזדמנות. הטמעת כלי כזה תאפשר למקבלי ההחלטות לעבור לניהול פרואקטיבי המזהה מגמות בזמן אמת וקובע סדרי עדיפויות מבוססי נתונים למיקוד המאמץ המשלתי.
2. **ממנבאים להתערבות – קידום מחקר סיבתי ומדיניות מותאמת:** המודל הנוכחי מאפשר לזהות את "מקור הסיכון" למוביליות חברתית-כלכלית (למשל, האם הוא מושפע מסביבת בית הספר או נעוץ ברקע המשפחתי); אך כדי לגזור מכך התערבויות בנות ביצוע, מומלץ לקדם מחקרי המשך סיבתיים. מחקרים אלו יבחנו האם שינוי ספציפי במנבאים שזוהו, למשל, חיזוק השכלת הורים או שיפור איכות ההוראה, אכן מוביל לשיפור בתוצאות המוביליות. תשתית זו תאפשר לעבור למדיניות מכלילה המציעה מענים מובחנים: במקום שבו החסם הוא סביבתי, יש לחזק את מערכת החינוך; ובמקום שבו החסם משפחתי, יש למקד מאמץ בתוכניות חוסן ותמיכה בתא המשפחתי.

3. הרחבת בסיס הנתונים: ניתוח משלים שבוצע באמצעות נתוני הסקר החברתי הראה כי למשתנים "רכים", כגון אוריינות דיגיטלית, שליטה בשפה וחוסן משפחתי, עשוי להיות כוח ניבוי משמעותי למוביליות כלפי מעלה. מלבד תרומתם לשיפור דיוק המודל, חשיבותם של משתנים אלו טמונה בכך שהם ניתנים להשפעה ולשינוי (כגון תוכניות להקניית מיומנויות דיגיטליות או חיזוק השפה). **לכן, מומלץ להעשיר את מסדי הנתונים המנהליים ולאסוף משתנים אלו בשיטתיות.** צעד זה יאפשר לקובעי המדיניות לא רק לחזות סיכונים היטב, אלא גם לטפל ישירות בגורמים המשפיעים ביותר על סיכויי ההצלחה של הפרט.

אימוץ כיווני הפעולה ומעבר ממדידה רטרוספקטיבית לשימוש במודלי ניבוי מבוססי נתונים עשויים לאפשר, לראשונה, לגבש מדיניות המבוססת על זיהוי מוקדם של פוטנציאל מוביליות כלפי מעלה. הממצאים מדגישים כי זיהוי מוקדם של מנבאים הניתנים להשפעה עשוי לספק למקבלי החלטות מצפן מבוסס נתונים. הטמעת המודל בתור כלי תומך החלטה, לצד השלמת מחקרי המשך סיבתיים, תהווה תשתית לבחינת התערבויות ופרקטיקות בשדה החברתי.

דברי תודה

מחקר זה הוא פרי שותפות בין ג'וינט-אשלים למכון מאירס-ג'וינט-ברוקדייל. נבקש להודות למשרד לשוויון חברתי וקידום מעמד האישה ולעמיתינו בג'וינט-אשלים על ההשקעה במחקר. תודה מיוחדת ליפעת קלופשטוק ולכל צוות הנגשת מידע למחקר בלשכה המרכזית לסטטיסטיקה על ההכנה וההנגשה של הנתונים שעליהם מבוסס דוח זה.

תוכן עניינים

1	1. מבוא
2	2. שיטה
4	3. ממצאי ליבה: זיהוי מנופי התערבות
9	4. ניתוח דפוסים מורכבים כבסיס למדיניות מותאמת
13	5. יישום מדיניות: המודל ככלי עבודה תומך החלטה
15	6. סיכום: כלי דינמי למדיניות עתידית
17	7. מונחון מושגים
19	מקורות
20	נספחים

רשימת לוחות

22	לוח ב-1: רגרסיה לוגיסטית לעומת עצי החלטה
26	לוח ג-1: תוצאות רגרסיה לוגיסטית – תרומת המשתנים המנבאים מוביליות כלפי מעלה בהכנסה
26	לוח ג-2: תוצאות רגרסיה לוגיסטית – תרומת המשתנים המנבאים מוביליות כלפי מעלה בהשכלה
27	לוח ג-3: תוצאות רגרסיה לוגיסטית – תרומת המשתנים המנבאים היעדר מוביליות
28	לוח ג-4: תוצאות מבחנים סטטיסטיים

רשימת תרשימים

4	תרשים א-1: תרומת המשתנים המנבאים לניבוי מוביליות כלפי מעלה בהכנסה
4	תרשים ב-1: תרומת משתנים מנבאים עבור חיזוי מוביליות כלפי מעלה בהכנסה, גרף SHAP
8	תרשים 2: גרף SHAP – תרומת המשתנים המנבאים לניבוי מוביליות כלפי מעלה בהכנסה כאשר מובאים בחשבון נתוני הסקר החברתי
10	תרשים א-4: התפלגות סיכויי מוביליות חזויים לפי שיעור המורים בעלי השכלה גבוהה (בחלוקה לאשכול יישוב)
11	תרשים ב-4: התפלגות הסתברויות חזויות בין קבוצות של אינטראקציית משתנים, שיוך לקבוצה באוכלוסייה והשכלת אם

רשימת תרשימים בנספחים

23	תרשים ב-1: השוואת ביצועי המודלים עבור משתנה מנובא של מוביליות כלפי מעלה בהכנסה
24	תרשים ב-2: השוואת ביצועי המודלים עבור משתנה מנובא של מוביליות כלפי מעלה בהשכלה
25	תרשים ב-3: השוואת ביצועי המודלים עבור משתנה מנובא של "Sticky Floor"
27	תרשים ג-1: גרף SHAP תרומת המשתנים המנבאים מוביליות כלפי מעלה בהשכלה
28	תרשים ג-2: גרף SHAP המציג את תרומת המשתנים המנבאים היעדר מוביליות

1. מבוא

דוח זה הוא חלק ממהלך מחקרי רחב הכולל סקירת ספרות מקיפה בנושא מוביליות חברתית-כלכלית בישראל (הלר ואח', 2025). סקירת הספרות פורשת את היריעה התאורטית וממפה את המנגנונים המרכזיים המעצבים את המוביליות, החל מן המשפחה, דרך מערכת החינוך ועד לשוק העבודה, ואילו הדוח הנוכחי מתמקד בהצעת פתרון לאתגר מתודולוגי ויישומי מרכזי: פער הזמנים במדידה.

מעבר ממדידה רטרואקטיבית לחיזוי פרוספקטיבי

היכולת לאמוד את רמת שוויון ההזדמנויות מתאפשרת לרוב רק בדיעבד, בחלוף דור. דפוס זה, המאפיין את מרבית המחקר הקלאסי בתחום (כגון Corak, 2013; Chetty et al., 2014), יוצר מצב שבו קובעי המדיניות נאלצים לקבל החלטות על בסיס נתונים היסטוריים ללא תמונת מצב עדכנית על סיכויי ההצלחה של הדור הצעיר בעתיד.

דוח זה מציע מעבר לגישה פרוספקטיבית (צופה פני עתיד), גישת מחקר הנפוצה במחקרים בעולם הרפואה. ברפואה המודרנית, השימוש במערכות תומכות החלטה ובמודלים לחיזוי סיכונים מאפשר אבחון מוקדם והתערבות מונעת עוד בטרם התפרצות המחלה (Rajkomar et al., 2018). בדומה לאיתור "סמנים מקדימים" רפואיים, הגישה המוצגת כאן מאפשרת לזהות בנתונים המנהליים הקיימים "סמנים חברתיים" המנבאים את פוטנציאל המוביליות העתידית. היתכנות המעבר למדידה כזו נתמכת במחקרם של דויטשה ומצומדר (Deutscher & Mazumder, 2020) המראים כי אפשר לאמוד מוביליות גם מתוך נתוני חיים חלקיים וראשוניים. אימוץ תפיסה של "רפואה מונעת חברתית" מאפשר לעבור מניהול מדיניות המגיב לנזקים שכבר התקבעו, לניהול פרואקטיבי המזהה הזדמנויות וסיכונים בזמן אמת.

המודל בתור כלי עבודה לעיצוב מדיניות

המחקר הנוכחי מציג אב-טיפוס של מודל ניבוי המיישם כלי למידת מכונה על נתונים מנהליים רחבי היקף של הלשכה המרכזית לסטטיסטיקה. המודל משלב שתי גישות משלימות:

- **רגרסיה לוגיסטית** לאמידת המגמות הכלליות והקשרים הישירים של כל משתנה.
- **עצי החלטה** לזיהוי אינטראקציות מורכבות ודפוסי חיים שאינם ליניאריים.

אתגרים אתיים והסתגלות לעולם משתנה

השימוש בטכנולוגיות ניבוי במגזר הציבורי מחייב משנה זהירות, כפי שמדגישים ברוקס וסלבסט (Barocas & Selbst, 2016). שילוב אלגוריתמים במדיניות ציבורית מחייב התייחסות מעמיקה לסוגיית ההוגנות האלגוריתמית (Algorithmic Fairness) כדי שלא להנציח באופן סמוי פערים היסטוריים. חשיבות היישום האחראי עומדת גם במוקד מסמך מדיניות הנכתב בימים אלו (פיינשטיין ואח', 2026), העוסק בפיתוח סטנדרטים להוגנות בבניה מלאכותית בשירותים במימון ציבורי. עבודה זו משלימה את המודל המוצג כאן, שכן היא מבקשת להבטיח שכלי הניבוי ישמשו מצפן מבוסס נתונים לקידום שוויון הזדמנויות תוך שמירה על עקרונות אתיים, שקיפות אלגוריתמית ואחריות מקצועית.

2. שיטה

תהליך פיתוח מודל הניבוי כלל ארבעה שלבים: הגדרת משתני התוצאה, אפיון המשתנים המסבירים, בחירת המודל המתאים ובניית בסיס הנתונים.

2.1 הגדרת משתני התוצאה

כדי לפתח כלי לחיזוי סיכונים לצד חוסן, התמקדנו בשלוש תוצאות מרכזיות המשקפות מוביליות חברתית-כלכלית. המודל שבנינו נועד לחזות את ההסתברות של כל פרט לחוות אחת מן התוצאות האלה:

- **מוביליות כלפי מעלה בהכנסה:** הוגדרה כמעבר מן השלישון התחתון של התפלגות הכנסות ההורים מעבודה (כשהפרט היה בן 15–19) אל השלישון העליון של התפלגות ההכנסות מעבודה (כשהפרט היה בן 30–34). זוהי "קפיצה" לא מבוטלת המשקפת שינוי חברתי-כלכלי מהותי. זהו משתנה התוצאה המרכזי במחקר הנוכחי.
- **מוביליות כלפי מעלה בהשכלה:** הוגדרה כהשגת השכלה גבוהה עבור מי שהוריו היו בשלישון התחתון של התפלגות ההכנסות מעבודה (כשהפרט היה בן 15–19).
- **היעדר מוביליות ("רצפה דביקה"):** הוגדרה כמצב שבו פרט שהוריו היו בשלישון התחתון של התפלגות ההכנסות מעבודה (כשהפרט היה בן 15–19), נשאר בעצמו בשלישון התחתון של התפלגות ההכנסות מעבודה (כשהיה בן 30–34). הגדרה זו מאפשרת לזהות קבוצות באוכלוסייה השוהות בעקביות בתחתית הסולם הכלכלי.

2.2 אפיון המשתנים המסבירים

המודל בוחן קשת רחבה של גורמים שנמצאו בספרות המחקר (הלר ואח', 2025) קשורים למוביליות:

- **נתוני הורים:** השכלה, הכנסות, מצב משפחתי
 - **נתונים דמוגרפיים:** מגדר, שיוך לקבוצה באוכלוסייה (ערבים, חרדים, יהודים שאינם חרדים), מוצא, אשכול יישובי
 - **נתונים על מאפייני מערכת החינוך:** סוג הפיקוח, מאפייני כיתת הלימוד (גודל, הרכב חברתי-כלכלי) ומאפייני צוות ההוראה (ותק, השכלה)
- ראו נספח א: פירוט המשתנים המסבירים.

2.3 מודל הניבוי

במחקר זה השתמשנו בשתי גישות חיזוי סטטיסטיות מקובלות המתקפות ומשלימות זו את זו. הגישה הראשונה היא **מודל רגרסיה לוגיסטית**, המספק מידע על הקשר הממוצע בין כל גורם ובין ההסתברות להיארעות משתנה התוצאה. הגישה השנייה היא **מודל עצי החלטה**, אשר תורם מידע על כושר הניבוי של כל משתנה, כמו גם על דפוסים מורכבים ואינטראקציות בין משתנים.

מבחינה יישומית, למודל עצי ההחלטה יתרון הנובע מיכולתו הייחודית והמובנית לזהות שילובים מורכבים של משתנים (אינטראקציות) ולכמת את הקשר בינם ובין משתנה התוצאה. שלא כמו מודל רגרסיה לוגיסטית אשר אומד את ההשפעה הממוצעת של כל גורם, עצי החלטה מזהים באיזה אופן המתאם של הגורמים השונים עם המשתנה המנובא הוא תלוי הקשר. כך לדוגמה, הממצאים המובאים בהמשך מראים כיצד ניבוי מוביליות כלפי מעלה באמצעות ותק המורים תלוי באשכול החברתי-כלכלי של יישוב המגורים; וכיצד עוצמת הניבוי של השכלת האם תלויה במוצא.

פירוט נוסף על אודות השיטות מופיע בנספח ב אשר מציג גם את ההשוואה הטכנית ביניהן; הגדרות והסברים למונחים המקצועיים השזורים בפרק זה ובדו"ח כולו ניתן למצוא במונחון המושגים המופיע בסופו.

2.4 אוכלוסיית המחקר

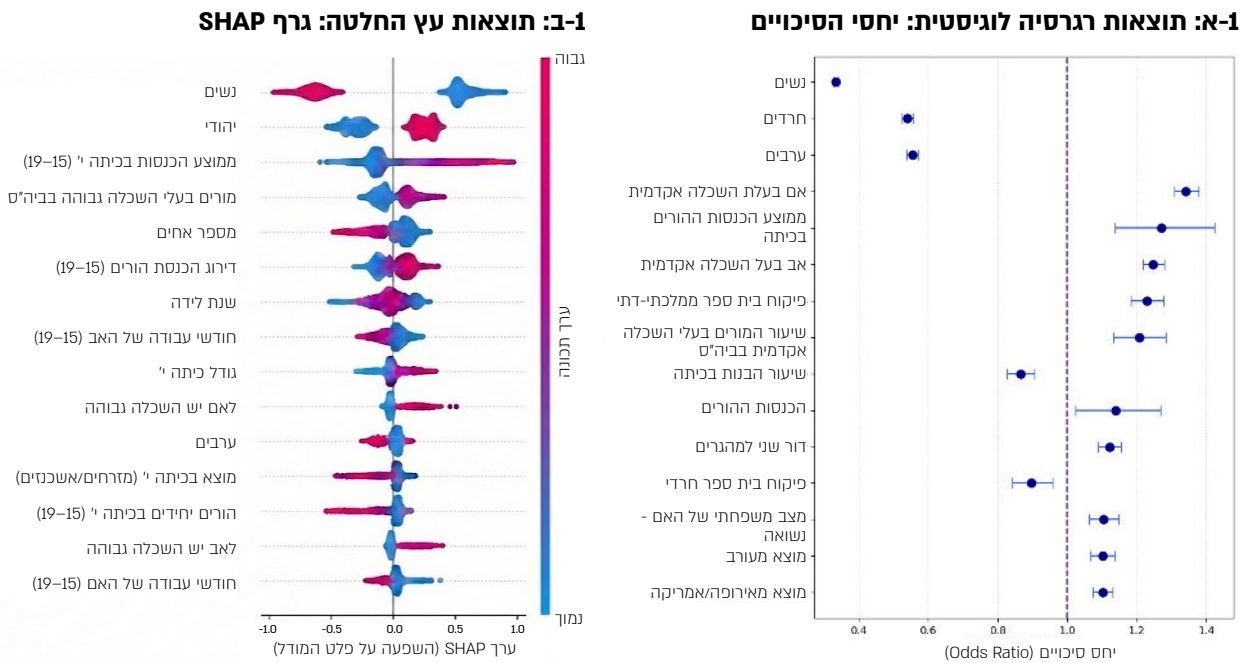
במחקר זה נעשה שימוש בנתונים מנהליים אשר הונגשו לחוקרים בחדר המחקר של הלשכה המרכזית לסטטיסטיקה. אוכלוסיית המחקר מבוססת על ילידי השנים 1970–1987. הנתונים על ילידי שנתונים אלו, כוללים 1,395,381 רשומות. תהליך האימון של מודל הניבוי התבצע על השלישון התחתון של הכנסות ההורים מעבודה, שכלל 430,212 רשומות. על פי הגדרת משתני התוצאה, נמצא כי 27% (שהם 114,472 פרטים) חוו מוביליות בהכנסה כלפי מעלה; 25% (שהם 109,700 פרטים) חוו מוביליות בהשכלה; ו-40% (שהם 170,395 פרטים) מצויים בתופעה של "רצפה דביקה" ולא חוו מוביליות.

3. ממצאי ליבה: זיהוי מנופי התערבות

בפרק זה, אנו מציגים לראשונה ניתוח מקיף לחיזוי מוביליות בין דורית בישראל, המבוסס על מאגר נתונים רחב היקף הכולל עשרות משתנים. הממצאים המובאים כאן מציגים את התרומה של כל גורם מנבא, כפי שהיא עולה משילוב שני המודלים: רגרסיות לוגיסטיות ועצי החלטה.

תרשימים 1א ו-1ב להלן מסכמים באופן ויזואלי את ממצאי שני המודלים עבור משתנה היעד המרכזי: 'מוביליות כלפי מעלה בהכנסה'. בחינתם זה לצד זה מדגישה את התרומה הייחודית של כל גישה. נספח ג מציג את התרשימים המקבילים בניבוי שני משתני התוצאה הנוספים, ואת ערכי האומדים (תוצאות החישובים) ברגרסיות הלוגיסטיות.

תרשים 1: תרומת המשתנים לניבוי מוביליות כלפי מעלה בהכנסה



הערות: הערות: תרשים 1א מציג את יחסי הסיכויים של 15 המשתנים עבורם נאמדו הערכים (המוחלטים) הגבוהים ביותר; תרשים 1ב מציג את גרף SHAP של מודל עץ ההחלטה. כל נקודה מייצגת אדם/תצפית ומראה כיצד ערך המשתנה השפיע על תחזית המודל. הצבעים מציינים את ערכי המשתנה: אדום = ערכים גבוהים, כחול = ערכים נמוכים. המיקום על הציר האופקי מצביע על כיוון ההשפעה: נקודות שמופיעות ימינה תרמו להגדלת הסיכוי לתוצאה (לדוגמה מוביליות כלפי מעלה), ונקודות שמופיעות שמאלה תרמו להקטנת הסיכוי. המשתנים מסודרים מלמעלה למטה לפי מידת חשיבותם הכוללת לחיזוי.

שני התרשימים מציגים את המשתנים השונים על פי מדרג של כוח ניבוי, כאשר ההבדלים ביניהם משקפים הבדלים מהותיים בין שתי הגישות:

מודל הרגרסיה (1א) מתמקד בבידוד הקשר הממוצע בין כל משתנה ובין סיכויי המוביליות. גישה זו חושפת ומחדדת את עוצמתם של גורמי רקע קבועים כמו מגדר ודת, ומעניקה להם בולטות המשקפת את משקלם המרכזי בניבוי.

לעומתו, **מודל עצי החלטה (1ב) מציג קשר כוללני יותר**, המבליט את ערכם המוסף של הגורמים הסביבתיים (כגון הרכב הכיתה ואיכות ההוראה). במודל זה, כושר הניבוי של גורמי הסביבה עולה משמעותית, שכן תרומתם למוביליות הופכת למכרעת בתוך הקשרים ספציפיים ואינטראקציות (למשל, רמת ההשכלה של המורה בקרב אוכלוסייה מוחלשת) – דינמיקה שמודל הרגרסיה אינו מביא בחשבון באופן מובנה.

שילוב הממצאים מאפשר לא רק דירוג מדויק יותר, אלא מצביע על כוונתניאל לשינוי מדיניות: הוא מסיט את המיקוד מנתוני הרקע של הפרט, שאינם ניתנים לשינוי, אל סביבת החינוך, שבה התערבות מושכלת יכולה לשנות את המציאות החברתית. הסעיפים הבאים מעמיקים בניתוח הממצאים ומתמקדים בכמה היבטים מרכזיים שבהם בלטו במיוחד כושר הניבוי והקשר למוביליות. הניתוח המפורט להלן מסכם את המגמות העיקריות שנצפו עבור כלל משתני התוצאה – מוביליות כלפי מעלה, רכישת השכלה והיעדר מוביליות. הנתונים הכמותיים המלאים זמינים לעיון בנספח ג.

3.1 פערים מבניים כחסם למוביליות

תרשים 1, על שני חלקיו, מראה כימות של הקשר המשמעותי בין מוביליות ובין פערים מוכרים במבנה החברה הישראלית:

■ **פערים בין מינים:** 'מגדר' זוהה כאחד המנבאים החזקים ביותר בשני המודלים; הוא מדורג בתור המשתנה החשוב ביותר הן במודל הרגרסיה הלוגיסטית (תרשים 1-א) הן במודל עצי ההחלטה (תרשים 1-ב). ממצאי הרגרסיה מראים כי לנשים סיכוי נמוך בכ-66% מלגברים (יחס סיכויים של כ-0.34) לחוות מוביליות כלפי מעלה בהכנסה. בד בבד יש להן סיכוי גבוה פי 2.4 להישאר במצב של היעדר מוביליות (לוח ג-3 בנספח). זאת, אף על פי שבהיבט של רכישת השכלה, לנשים סיכוי גבוה פי 4.9 לחוות מוביליות (לוח ג-2 בנספח).

ממצא זה מעגן בנתונים אמפיריים רחבים את הפער בין הישגי נשים בהשכלה ובין מעמדן הכלכלי. הניתוח הנוכחי מראה כיצד פער השכר המוכר מתפקד בפועל כחסם למוביליות: אף שנשים מרקע מוחלש רוכשות השכלה בשיעורים גבוהים משמעותית מן הגברים, יכולתן לתרגם את ההון האנושי למוביליות בהכנסה מוגבלת. נתוני העתק מאששים כי התשואה להשכלה עבור נשים נמוכה יותר מזו של גברים, וכי ההשכלה לבדה איננה ערובה מספקת להיחלצות ממעגל העוני.

■ **פערים בין קבוצות באוכלוסייה:** השתייכות לקבוצות שונות באוכלוסייה נמצאה כגורם מנבא חיוני לסיכויי המוביליות. בחלוקה לקבוצות העיקריות באוכלוסייה (יהודים שאינם חרדים, חרדים, ערבים), השתייכות לקבוצת היהודים שאינם חרדים זוהתה בתור גורם מנבא חיוני בשני המודלים. גם במודל הרגרסיה הלוגיסטית (תרשים 1-א) וגם במודל עצי ההחלטה (תרשים 1-ב) המשתנה מדורג שני בחשיבותו. לפי מודל הרגרסיה הלוגיסטית, השתייכות לקבוצת החרדים נמצאה קשורה לסיכוי נמוך יותר למוביליות כלפי מעלה בהכנסה, עם יחס סיכויים של כ-0.5. נוסף על כך נמצא כי השתייכות לקבוצת הערבים קשורה לסיכוי נמוך למוביליות כלפי מעלה בהכנסה, עם יחס סיכויים של כ-0.55.

3.2 סביבת החינוך

למערכת החינוך ולסביבת הלימודים בבית הספר יש קשר ישיר לסיכויי המוביליות של הפרט. ממצאי המודלים מצביעים על כמה גורמי מפתח:

■ **איכות המורים וניסיונם:** שיעור המורים בעלי השכלה גבוהה בבית הספר זוהה כגורם מנבא חיוני. משתנה זה הוא משתנה רציף הנע בין 0 ל-1 ומשקף את שיעור המורים האקדמאים בבית הספר. לפי תוצאות הרגרסיה הלוגיסטית, מעבר תאורטי מבית ספר שבו אין כלל מורים אקדמאים לבית ספר שבו כל המורים הם אקדמאים, קשור בעלייה של כ-12% בסיכוי למוביליות בהכנסה. יחס זה משקף השפעה ממוצעת מתונה יחסית במודל הרגרסיה לעומת משתנים אחרים, אך במודל עצי ההחלטה (תרשים 1-ב) משתנה זה מדורג במקום הרביעי בחשיבותו. יתרה מזאת, בניבוי ההסתברות למוביליות בהשכלה, משתנה זה הוא שלישי בחשיבותו בשני המודלים עם יחס סיכויים של 2.56, שמשמעותו סיכוי של פי 2.56 למוביליות בהשכלה בקרב ילדים שלמדו בבית ספר שבו כל המורים בעלי השכלה גבוהה לעומת ילדים שלמדו בבית ספר בלי מורים בעלי השכלה גבוהה.

ממצאים אלו עולים בקנה אחד עם תובנות בין-לאומיות, כגון אלו של ה-OECD (Schleicher, 2018), המדגישות כי חיזוק ההון האנושי בבתי ספר, ובפרט הבטחת שיעור גבוה של מורים בעלי השכלה גבוהה וניסיון בפריפריה החברתית-כלכלית, הוא מנוף יעיל לצמצום פערים ולמוביליות.

השערנו לפער שנמצא בנתוני ההכנסה היא שלאיכות המורים תפקיד מרכזי באינטראקציות מורכבות עם משתנים אחרים, ולכן כוח הניבוי שלה בפועל גבוה בהרבה מן האומדן הממוצע שעולה מן הרגרסיה הלוגיסטית, ובהתאם לכך בפרק הבא נרחיב ונציג כיצד אינטראקציה בין אשכול יישוב חברתי-כלכלי ובין השכלת המורים בבית הספר עשויה לשנות את הסיכוי למוביליות בין-דורית.

■ **ההרכב החברתי-כלכלי של הכיתה:** נמצא קשר בין ממוצע הכנסות ההורים בכיתה ובין סיכויי המוביליות. בדומה למשתנה איכות המורים, גם כאן הפער בין המודלים מלמד על חשיבות ההקשר: בעוד במודל הרגרסיה (תרשים 1-א) הקשר הממוצע בין משתנה זה ובין מוביליות בהכנסה הוא מתון יחסית, עם יחס סיכויים של כ-1.27, במודל עצי ההחלטה (תרשים 1-ב) משתנה זה מדורג במקום השלישי בחשיבותו. הדירוג הגבוה מעיד כי הרכב הכיתה משמש גורם מפתח באינטראקציות, וכי עבור אוכלוסיות מוחלשות עשויה להיות לו השפעת ניכרת.

■ **מאפייני מסגרת החינוך:** בעוד במודל הרגרסיה הלוגיסטית פיקוח ממלכתי-דתי הוא משתנה בעל חשיבות מכרעת (לדוגמה, יחס סיכויים של 2.56 לסיכוי נמוך למוביליות), במודל עצי ההחלטה (תרשים 1-ב) משתנה זה כלל לא עלה בין המשתנים החשובים ביותר. היעדרותו של משתנה הפיקוח מצמרת הדירוג בעץ, בניגוד לדומיננטיות של מאפייני הכיתה (כגון הכנסות ההורים), היא שהובילה אותנו למסקנה כי השפעת המסגרת אינה עומדת בפני עצמה, אלא שלובה במאפייני התלמידים הלומדים בה.

3.3 התשתית המשפחתית-קהילתית

המודל מאשש את החשיבות של הרקע המשפחתי ושל הסביבה הקהילתית כגורמים מנבאים.

■ **השכלת האם:** השכלה גבוהה של האם, זוהתה בתור גורם מנבא חיובי עקבי וחזק בשני המודלים (מדורג גבוה בתרשים 1-ב, ועם יחס סיכויים של 1.34 בתרשים 1-א). אף שמחקרים מסוימים גורסים כי במדינות עם אי-שוויון גבוה השפעת האב דומיננטית יותר, הממצאים האמפיריים בדוח זה עולים בקנה אחד עם ספרות המחקר המצביעה על כך שבמדינות מפותחות, להשכלת ההורים תפקיד מכריע במוביליות של ילדיהם (כגון OECD, 2018).

■ **רקע כלכלי:** להכנסת ההורים עצמם יש קשר מובהק למוביליות (ראו תרשים 1-א). משתנה זה מייצג את דירוג הכנסות ההורים בהתפלגות הכללית. אף שאוכלוסיית המחקר מוגבלת לשלישון התחתון (כלומר, ערכי המשתנה במדגם נעים בין 0 ל-0.33), המודל מזהה שגם בתוך טווח מצומצם זה, למיקום היחסי יש משמעות: ככל שהכנסת ההורים גבוהה יותר (קרובה יותר לרף העליון של השלישון), כך גדל הסיכוי למוביליות בהכנסה, ובד בבד קטן הסיכוי להישארות במצב של היעדר מוביליות.

■ **מבנה המשפחה:** למספר האחים במשפחה יש קשר שלילי לסיכויי המוביליות. ממצא זה עולה באופן עקבי בשני המודלים; ככל שמספר האחים גדול יותר, כך יורד הסיכוי למוביליות בהכנסה.

■ **מאפייני יישוב המגורים:** האשכול החברתי-כלכלי של היישוב שבו גדל הפרט זוהה גם הוא כגורם בעל חשיבות. אף שברגרסיה הלוגיסטית נאמד קשר מתון, במודל עצי ההחלטה הוא בולט יותר, בעיקר בשל תפקידו ביצירת אינטראקציות עם משתנים אחרים (כגון שיעור המורים האקדמאים בבית הספר), נושא שיידון בהרחבה בפרק הבא.

לצד משתני הרקע הקשיחים (הכנסה והשכלה), ניתוח משלים הבוחן את השפעתם של משתני מיומנות וחוסן, כגון אוריינות דיגיטלית וחוסן משפחתי, מוצג בתיבה 1 להלן.

תיבה 1: מבט משלים על מוביליות - מעבר למידע המנהלי

מחקרים המבוססים על נתונים מנהליים מאופיינים לרוב בהטיה מבנית, הנובעת מהתמקדות במשתנים הזמינים המתועדים במערכות המידע של הממשלה, כגון הכנסה, השכלה ודמוגרפיה. הטיה זו עלולה להותיר מחוץ לניתוחים נתונים על אודות משאבים רכים שאינם מתועדים רשמית כמו הון תרבותי, מיומנויות דיגיטליות, חוסן משפחתי ושליטה בעברית. למשאבים אלו עשויה להיות השפעה רבה על סיכויי המוביליות של הפרט.

במחקר זה ביקשנו לאתגר מגבלה זו באמצעות מהלך חדשני של שילוב נתונים: הצלבת הרשומות המנהליות שעליהן מבוסס מחקר זה עם נתוני הסקר החברתי של הלשכה המרכזית לסטטיסטיקה, תוך כדי קישור נתוני הסקר של ההורים לרשומות המנהליות של ילדיהם (אוכלוסיית המחקר). בדרך זו היה באפשרותנו להתבונן בסוגי הון נסתרים מן הנתונים המנהליים ולבחון את משקלם בניבוי מוביליות.

המחיר המתודולוגי: עומק מול רוחב

מהלך זה גבה מחיר טכני שחשוב להכיר בו. המודל הראשי הוא בעל חוזק סטטיסטי רב אשר נובע מן השימוש בנתוני עתק של כ-430,000 רשומות, המאפשר זיהוי דפוסים עדינים גם בקרב תתי-אוכלוסיות, ואילו החיתוך עם הסקר צמצם את המדגם לכ-5,000 הורים בלבד. משמעות הדבר היא ויתור כפוי על היתרון היחסי של עצי ההחלטה: במדגם קטן זה, המודל מאבד את הרזולוציה הנדרשת לזיהוי אינטראקציות מורכבות וספציפיות. עם זאת, הויתור על רוחב היריעה הכמותי אפשר לנו להפיק תועלת מעושר רעיוני, ובכך להשיג הבנה מעמיקה יותר של המנגנונים הקשורים למוביליות.

הממצאים: חשיבותם של משתני המיומנות

ניתוח חשיבות המשתנים במודל המצומצם חושף כי למשתני הסקר החברתי עשוי להיות כוח ניבוי מפתיע בעוצמתו, לעיתים אף מעבר למשתנים דמוגרפיים מסורתיים:

1. נגישות דיגיטלית: משתנים העוסקים בנגישות דיגיטלית כמו היעדר מחשב מחובר לאינטרנט בבית ושימוש באינטרנט דורגו גבוה מאוד בחשיבותם לניבוי (ראו **תרשים 2** להלן). יש לזכור כי משתנה "המחשב בבית" נבחן בהקשר לתקופת ילדותם של הנחקרים, עת היה מדובר במשאב נדיר ומבדל יותר מאשר כיום. עם זאת, המודל מסמן נגישות זו לא רק בתור סממן סטטוס של אותה תקופה, אלא בתור כלי עבודה יום-יומי ומפתח המרחיב את הגישה להזדמנויות ולמוביליות לאורך מעגל החיים.

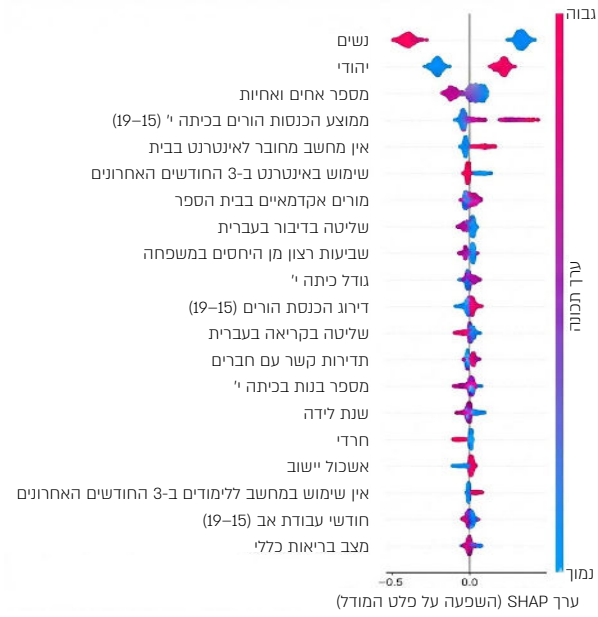
2. שליטה בשפה העברית: רמת שליטה גבוהה בשפה העברית (דיבור וקריאה) נמצאה במתאם חיובי עם הסיכוי למוביליות. ממצא זה מדגיש כי שליטה בשפה הרשמית של המדינה היא משאב הון אנושי חיוני להשתלבות בשוק העבודה.

3. חוסן משפחתי: גם משתנים סובייקטיביים, כגון שביעות רצון מן היחסים במשפחה, הראו תרומה לניבוי, עדות לחשיבותה של רשת ביטחון משפחתית. מבט כולל על מדרג החשיבות חושף כי רכיבי החוסן המשפחתי תופסים את המקומות הראשונים בניבוי. דומיננטיות זו של רכיבי החוסן המשפחתי, הכוללים את הכנסות ההורים, השכלת האם ומספר האחים, מלמדת כי המשאבים העומדים לרשות התא המשפחתי הם בסיס החוסן החשוב ביותר לסיכויי המוביליות של הילד.

סיכום

הממצאים המובאים בתיבה זו שופכים אור על תפקידם החשוב של משאבים רכים: אוריינות דיגיטלית, שליטה בשפה וחוסן משפחתי אינם רק מאפיינים נלווים להצלחה, אלא מנבאים חשובים שלה. במבט רחב יותר, הניתוח מדגים הלכה למעשה כיצד התחשבות במשתנים רלוונטיים, גם אם איסופם מורכב יותר, מובילה לתמונה עשירה, מדויקת ומלאה יותר של מנועי המוביליות, מעבר למה שמתאפשר בשימוש בנתונים מנהליים סטנדרטיים. תובנה זו טומנת בחובה קריאה לפעולה למחקר עתידי: העמקת איסוף הנתונים בממדים נוספים ושילובם במודלים מסוג זה עשויה לחשוף מנופי התערבות חדשים ובלתי צפויים, ולהרחיב את ארגז הכלים של קובעי המדיניות ומובילי התערבויות בשדה החברתי.

תרשים 2: גרף SHAP – תרומת המשתנים המנבאים לניבוי מוביליות כלפי מעלה בהכנסה כאשר מובאים בחשבון נתוני הסקר החברתי



4. ניתוח דפוסים מורכבים כבסיס למדיניות מותאמת

פרק זה מתמקד בניתוח מעמיק של יחסי הגומלין בין המשתנים הנבחרים, בהמשך לכרך הקודם אשר הציג את המגמות הכלליות המאפיינות את המוביליות החברתית-כלכלית ברמת האוכלוסייה כולה. בחרנו לבצע עיבודים להשוואת הפערים בין שני המודלים, וממצאיהם מוצגים להלן: כאשר משתנה מסוים מציג השפעה מתונה במודל הרגרסיה, ולצד זה מדורג גבוה בחשיבותו במודל עצי ההחלטה, הדבר סימן לנו כי כוחו של המשתנה נובע מאינטראקציות מסוימות ולא מהשפעה אחידה. מהלך זה מאפשר מעבר מפתרונות אחידים להתערבויות המותאמות למאפיינים הייחודיים של קבוצות שונות.

4.1 מאומדן ממוצע להקשר ייחודי

ניתוח סטטיסטי מסורתי כדוגמת הרגרסיות הלוגיסטיות אומד את תרומתו השולית של גורם מסוים למוביליות, בהנחה שיתר משתני הרגרסיה נותרים קבועים. האומדנים המתקבלים משקפים את עוצמת הקשרים בין משתנה התוצאה ובין המשתנים המסבירים, בממוצע באוכלוסייה. כך למשל, ממצאי מודל הרגרסיה מצביעים על מתאם חיובי מובהק בין ותק מורים למוביליות.

לעומת זאת, גישת עצי ההחלטה מתמקדת בזיהוי התנאים הייחודיים שבהם גורם מסוים הופך למנבא דומיננטי. המודל מאתר באופן אלגוריתמי שילובים של משתנים מנבאים, שבהם עוצמת הקשר בין גורם אחד ובין התוצאה תלויה ברמתו של גורם אחר, ובכך הוא מייצר פרופיל סיכון/סיכוי דינמי, שבו האלגוריתם מעדכן את התחזית בהתאם להקשר הספציפי של הפרט; בעוד ממצאי רגרסיה לוגיסטית מתארים קשר ממוצע וקבוע של משתנים, עצי ההחלטה יוצרים "מסלול" גמיש שיכול לזהות כיצד גורם סיכון עשוי להיות קריטי עבור קבוצה אחת אך זניח עבור אחרת. יכולת זו הופכת את המודל לכלי אקטיבי לניהול סיכונים: הוא מאפשר למקבלי ההחלטות לא רק לקבל תחזית, אלא לדרג את האוכלוסייה לפי רמת דחיפות או פוטנציאל הצלחה, ובכך לקבוע ספי החלטה המייעלים את הקצאת המשאבים הציבוריים בזמן אמת.

4.2 מקרי בוחן לאינטראקציות

כדי להמחיש את הערך המוסף של גישה זו, אנו מביאים להלן שתי דוגמאות בולטות לאינטראקציות, שנחשפו בעקבות הפערים בדירוג המשתנים בין המודלים באוכלוסייה הכללית.

« מקרה בוחן 1: חשיבות ההקשר - מורים בעלי השכלה גבוהה באשכולות יישוב חברתיים-כלכליים נמוכים לעומת גבוהים

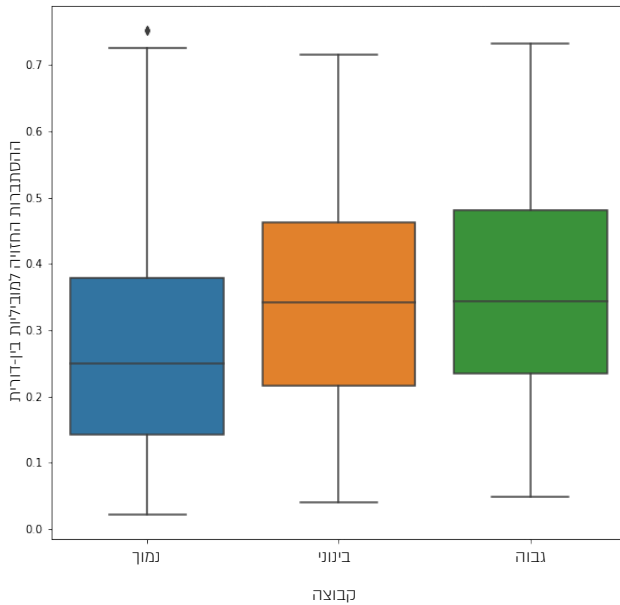
הממצאים בפרק הקודם הצביעו על כך שלמשתני איכות ההוראה (ובפרט שיעור המורים בעלי השכלה גבוהה בבית הספר) יש תרומה רבה לניבוי בשני המודלים. עם זאת, הדירוג הגבוה במיוחד שקיבל משתנה זה במודל עצי ההחלטה מרמז כי השפעתו אינה אחידה, אלא תלויה בהקשר הסביבתי.

ראשית, חשוב לעמוד על הפערים הקיימים בתשתיות ההון האנושי בין יישובים באשכולות חברתיים-כלכליים נמוכים ובין יישובים באשכולות חברתיים-כלכליים גבוהים. בדיקת הנתונים מעלה תמונת מצב של אי-שוויון בנקודת המוצא: ביישובים באשכולות חברתיים-כלכליים גבוהים (6-10), שיעור המורים בעלי השכלה גבוהה עומד בממוצע על 71% (סטיית תקן של 0.18). לעומת זאת, ביישובים באשכולות הנמוכים (1-5), הממוצע עומד על 59% בלבד (סטיית תקן של 0.24).

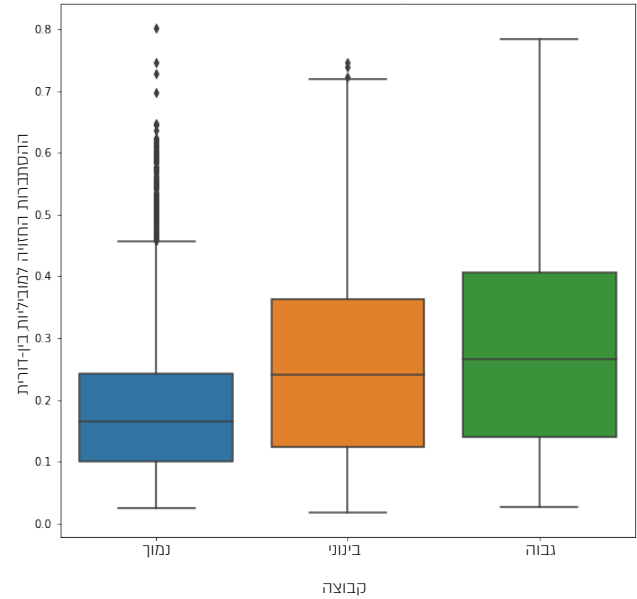
פער זה מחדד את חשיבות הממצא המוצג ב**תרשים 4-א** להלן: דווקא באשכולות הנמוכים, שבהם מורים אקדמאים הם משאב נדיר יותר ושונות האיכות בין בתי הספר גבוהה יותר, הימצאותם בבית הספר היא קריטית ביותר לסיכויי המוביליות של התלמיד.

תרשים 4-א: התפלגות סיכויי מוביליות חזויים לפי שיעור המורים בעלי השכלה גבוהה [בחלוקה לאשכול יישוב]

התפלגות הסתברויות חזויות בין קבוצות של אינטראקציית משתנים: אשכול יישוב נמוך ושיעור המורים בעלי השכלה גבוהה



התפלגות הסתברויות חזויות בין קבוצות של אינטראקציית משתנים: אשכול יישוב גבוה ושיעור המורים בעלי השכלה גבוהה



הערה: התרשים מציג את שיעור התלמידים שצפויים, לפי המודל, להשיג מוביליות (בציר האנכי) לפי שיעור המורים האקדמאים בבית הספר בו למדו (בחלוקה לשלוש קבוצות שוות גודל: נמוך, בינוני וגבוה; בציר האופקי). התרשים הימני מתייחס ליישובים מאשכולות חברתיים-כלכליים גבוהים (6-10) והשמאלי ליישובים מאשכולות נמוכים (1-5). כיצד לקרוא את התרשים: הקופסה הצבעונית מייצגת את הטווח שבו נמצאים 50% מהתלמידים (החציון מסומן בקו בתוך הקופסה), והקווים האנכיים בקצותיה מציגים את טווח הפיזור של שאר התלמידים, כאשר חריגים מוצגים כנקודות.

מן התרשים עולה כי גם ביישובים מאשכולות גבוהים (התרשים הימני) וגם ביישובים מאשכולות נמוכים (התרשים השמאלי), יש קשר בין שיעור המורים בעלי השכלה גבוהה ובין סיכויי המוביליות, כאשר ביישובים מאשכולות נמוכים קשר זה חזק יותר. להלן ניתוח הממוצעים וגודל האפקט שחושף הבדלים משמעותיים בהשכלת המורים בתור גורם מקדם מוביליות בהתאם להקשר היישובי (ראו ניתוח סטטיסטי מפורט בלוח ג-4 בנספח):

■ **אשכולות חברתיים-כלכליים נמוכים:** בקבוצה זו, בבתי ספר עם מעט מורים אקדמאים שיעור התלמידים שחוו מוביליות הוא 19%, לעומת שיעור של 29% בבתי ספר עם מורים אקדמאים רבים. זהו פער של יותר מ-50% בשיעור התלמידים שחוו מוביליות. ממצא זה ממחיש כי הוראה איכותית עשויה לתרום ניכרת לסיכויי המוביליות.

■ **אשכולות חברתיים-כלכליים גבוהים:** בקבוצה זו שיעור גבוה יותר של תלמידים שחוו מוביליות – ממוצע של 27% בבתי ספר עם מעט מורים אקדמאים ו-36% בבתי ספר עם הרבה מורים אקדמאים, אך הפער בין שני סוגי בתי הספר נמוך יותר ועומד על כ-33%. יש לציין כי גם ההבדלים בסטיות התקן של הממוצעים, הגדולים יותר באשכולות הגבוהים, בפרט בבתי ספר עם מעט מורים אקדמאים (ראו לוח ג-4 בנספח), מעידים על כך שבאשכולות הגבוהים סיכויי המוביליות ככל הנראה מושפעים ממכלול רחב יותר של משאבים חיצוניים (כמו רקע משפחתי וכלכלי), ולכן איכות ההוראה לבדה משפיעה פחות מן האשכולות הנמוכים.

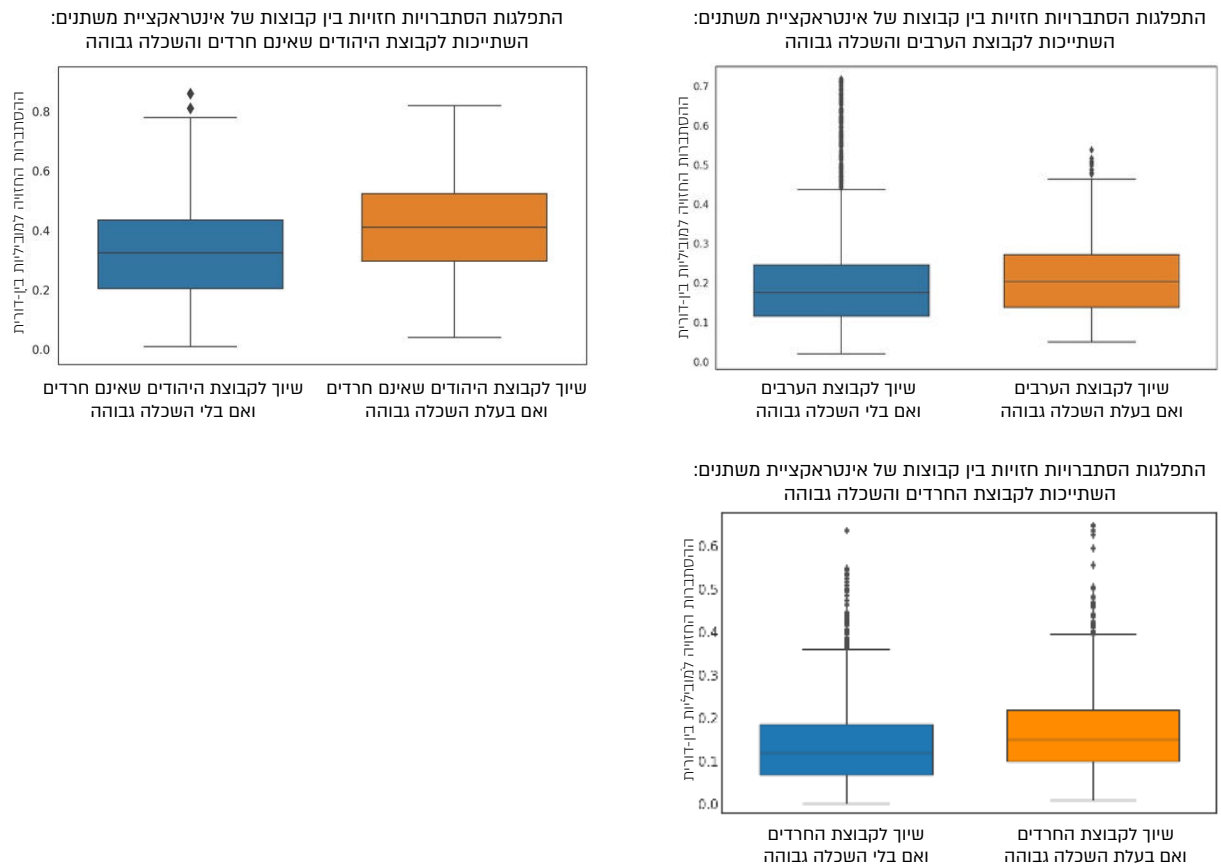
הסבר אפשרי לעוצמת הקשר שנמצאה טמון בתפקידו הייחודי של בית הספר עבור קבוצות שונות באוכלוסייה. בעוד בקרב אוכלוסיות מבוססות יש משאבים משפחתיים וקהילתיים משלימים, עבור אוכלוסיות מוחלשות ומודרות בית הספר משמש לעיתים קרובות בתור המכשיר הבלעדי והמרכזי למוביליות חברתית. משום כך, איכות התשומות הבית-ספריות – ובפרט איכות ההוראה – הופכת להיות גורם מכריע בעיצוב עתידם הכלכלי של תלמידים אלו, באופן משמעותי הרבה יותר מאשר בקרב עמיתיהם מרקע חזק.

מקרה בוחרן 2: גורם חוסן - השכלת האם בתור גורם מצמצם פערים

השכלת האם זוהתה בתור אחד הגורמים המנבאים העוצמתיים לסיכויי המוביליות של הדור הבא. כדי להבין לעומק את מגנוני ההשפעה של משאב זה, בחנו את האינטראקציות שלו עם קבוצות שונות באוכלוסייה.

המטרה בבחינת האינטראקציות היא לבדוק אם התשואה על השכלת האם היא אחידה באוכלוסייה או משתנה על פי ההקשר החברתי שאליו משתייך הפרט. הממצאים המפורטים להלן (תרשים 4-ב) חושפים כי השכלת האם אינה פועלת בחלל ריק – השתייכות לקבוצה באוכלוסייה משפיעה על המידה שבה השכלת האם קשורה לסיכויי ההצלחה של הילד.

תרשים 4-ב: התפלגות הסתברויות חזויות בין קבוצות של אינטראקציית משתנים, שיוך לקבוצה באוכלוסייה והשכלת אם



הערה: התרשים מציג את שיעור התלמידים שצפויים, לפי המודל, להשיג מוביליות (בציר האנכי) בחלוקה לפי קבוצות אוכלוסייה (בכותרות) והשכלת האם (בציר האופקי). הקופסה הכחולה מייצגת תלמידים לאימהות ללא השכלה גבוהה, והקופסה הכתומה מייצגת תלמידים לאימהות בעלות השכלה גבוהה. כיצד לקרוא את התרשים: הקופסה הצבעונית מייצגת את הטווח שבו נמצאים 50% מהתלמידים (החציון מסומן בקו בתוך הקופסה), הקווים האנכיים בקצותיה מציגים את טווח הפיזור של שאר התלמידים, כאשר חריגים מוצגים כנקודות.

משלושת חלקי התרשים עולה קשר חיובי עקבי בין השכלת האם לסיכויי המוביליות בכל הקבוצות באוכלוסייה, עם סיכויים גבוהים יותר למוביליות בקרב תלמידים לאימהות בעלות השכלה גבוהה לעומת אימהות בלי השכלה גבוהה. עם זאת, קשר זה שונה בעוצמתו בקבוצות השונות באוכלוסייה: בקרב יהודים שאינם חרדים הקשר בין השכלת האם לסיכויי המוביליות כמעט כפול (הבדל של 7.3 נקודות אחוז) מסיכויי המוביליות בקבוצת הערבים (הבדל של 3.8 נקודות אחוז) והחרדים (הבדל של 3.6 נקודות אחוז; ראו לוח ג- 4 בנספח). ממצא זה מעיד על כך שבקבוצות אלו באוכלוסייה יש חסמים מבניים המגבילים את ההעברה הבין-דורית של הון אנושי.

הממצאים הללו מרמזים על כך שנוסף על התועלות החברתיות המיידיות בהשכלה, בטווח הארוך השכלת הנשים עשויה להוות גורם בעל חשיבות אסטרטגית בצמצום פערים חברתיים ובקידום מוביליות חברתית-כלכלית. הממצאים מדגישים כי השקעה בהשכלה גבוהה עבור נשים בכלל הקבוצות באוכלוסייה היא צעד חיוני, אך באוכלוסייה הערבית והחרדית עליה להיות מלווה במדיניות משלימה להסרת חסמים מבניים בשוק התעסוקה. טיפול משלים במשתני הרקע הממתנים כיום את השפעת ההשכלה, עשוי לשפר את המידה שבה השכלת האם היא מנוף למוביליות עבור הדור הבא כפי שמתקיים בקרב יהודים שאינם חרדים.

יחד עם זאת, תמונת המוביליות באוכלוסייה הערבית מחייבת התייחסות גם למגמות בקרב גברים: בעוד ניכרת עלייה מרשימה בהשכלת הנשים, מסתמן קיפאון בהשכלת הגברים. פער זה מוכר כחסם המקשה על תרגום ההישגים האקדמיים של הנשים למוביליות כלכלית ברמת משק הבית. משמעות הדבר היא כי ייתכן שכדי למצות את פוטנציאל המוביליות של הנשים, יש לתת את הדעת גם לסגירת הפערים בקרב הגברים.

5. יישום מדיניות: המודל ככלי עבודה תומך החלטה

הטמעתו של מודל הניבוי בעבודת מטה לקידום שוויון בהזדמנויות היא בסיס לשינוי תפיסתי עמוק בדרכי גיבוש המדיניות החברתית. באופן מסורתי, ההתמודדות עם פערים במוביליות מתבססת על נתונים רטרוספקטיביים, כלומר, המדידה מתבצעת בדיעבד, דור לאחר שהפערים כבר התקבעו והפכו לעובדה מוגמרת. המודל הנוכחי מציע מעבר לגישה צופה פני עתיד (פרוספקטיבית), המאפשרת לזהות את פוטנציאל המוביליות, או את הסיכון להיעדרה, עוד בטרם התממשו התוצאות החברתיות-כלכליות. יכולת זו מאפשרת למעצבי המדיניות לעבור מגישה מתקנת (ריאקטיבית) לגישה מונעת (פרואקטיבית), ולהקצות משאבים באופן ממוקד על בסיס זיהוי מוקדם.

כדי לאפשר מעבר זה הלכה למעשה, נבחר השימוש במודל עצי החלטה. בהקשר המחקרי-יישומי, מודל זה מייצג שלב אבולוציוני ביכולת הניתוח: המעבר מתיאור של קשרים ממוצעים (המאפיין מודלים סטטיסטיים קלאסיים) למיפוי של מסלולי חיים ופרופילים מורכבים. המציאות החברתית אינה ליניארית, וגורמי סיכון נוטים להופיע במקבצים ובשילובים ייחודיים. עץ החלטה יכול לזהות דפוסים אלו של אינטראקציות ושל חוסר לינאריות באופן מובנה, והוא מספק את הרזולוציה הנדרשת למדיניות מותאמת אישית, ומאפשר לאתר במדויק את הצמתים שבהם להתערבות תהיה התועלת המרבית.

פרק זה מתווה את העקרונות המנחים לשימוש במודל ככלי עבודה יישומי בסביבת קבלת החלטות.

« מאפייני הכלי: הערכה הסתברותית וניהול סיכונים

נבחר כי מודל הניבוי אינו מספק סיווג דיכוטומי מוחלט בנוגע לעתידו של הפרט, אלא מפיק ערך רציף הנע בין 0 ל-1. ערך זה משקף את ההסתברות הסטטיסטית למוביליות, בהתבסס על שקלול מורכב של נתוני הרקע והסביבה. משמעות הדבר היא שהמודל משמש כלי תומך החלטה הממפה סיכונים והזדמנויות; הוא אינו פועל אוטומטית, והשימוש בו מחייב שיקול דעת והבנה של עקרונותיו. הגמישות בקביעת "ספי החלטה" (כפי שיפורט להלן) מאפשרת למקבלי החלטות לדרג קבוצות באוכלוסייה על פי רמת הדחיפות להתערבות או על פי פוטנציאל ההצלחה, ובכך להמליץ על תהליכי התעדוף והקצאת המשאבים.

אופן השימוש בציון הניבוי נגזר מאופי כלי המדיניות:

- **במדיניות של הקצאת משאבים רציפה** (למשל, קביעת גובה סל תקציב לתלמיד או לבית הספר), אפשר להסתייע בנתון הרציף (0 עד 1) כפי שהוא, כדי להתאים את היקף ההשקעה לעוצמת הצורך באופן יחסי.
- **במדיניות של התערבות** (למשל, החלטה על קבלה לתוכנית במקום מסוים, או הפניה לטיפול ייעודי), אי אפשר להישאר עם ערך הסתברותי; יש לתרגם אותו להכרעה בינארית של "התערבות" או "אי-התערבות". לשם כך, צריך לקבוע סף החלטה.

« קביעת סף [threshold] כהחלטה אסטרטגית

כאשר מדיניות דורשת בחירה בין שתי אפשרויות, קביעת 'סף החלטה' הופכת לנושא של סדרי עדיפויות. מדובר בהחלטת מדיניות המגדירה כיצד המודל יפרש את הנתונים לכדי פעולה. הבחירה נעה על רצף החלטה המייצג את המתח בין שני הקטבים האלה:

1. **בקצה האחד: גישת הכיסוי המרבי (סף נמוך)**. ככל שמורידים את הרף, כך גדל הסיכוי לאתר ולכלול בתוכנית את מרב הפרטים העשויים להזדקק לסיוע. גישה זו הולמת מצבים שבהם העלות החברתית של אי הכללת זכאים נתפסת

גבוהה במיוחד. המחיר של התקרבות לקוטב זה הוא השקעת משאבים גם בפרטים שהיו מצליחים ללא התערבות (חיוביים-כוזבים), אך היא מבטיחה רשת ביטחון רחבה.

2. בקצה השני: גישת מיקוד המשאבים (סף גבוה). ככל שמעלים את הרף, כך מתמקדת ההתערבות רק במקרים המובהקים ביותר שבהם רמת הסיכון (או הסיכוי) גבוהה במיוחד. גישה זו הולמת מצבים תקציביים המחייבים יעילות כלכלית מקסימלית. המחיר של התקרבות לקוטב זה הוא הסיכון להותיר ללא מענה אוכלוסיות גבוליות, אשר היו יכולות להפיק תועלת רבה מן ההתערבות. סיכון נוסף הוא תת-מימוש של התערבויות ייחודיות בגלל חסמי נטל בירוקרטי וחסמי נגישות אחרים.

לפיכך, סף ההחלטה הוא פרמטר גמיש שיש להתאימו התאמה דינמית בהתאם לאיזון הרצוי בין שוויון הזדמנויות, מגבלות תקציב ויעדי התוכנית הספציפית.

« רמות הפעולה: מתכנון אסטרטגי להתערבות פרטנית

המודל עשוי לשרת צרכים מגוונים של מקבלי החלטות בדרגים שונים ומטרות שאפשר לחלקן לשתי רמות:

- 1. ברמה האסטרטגית (הקצאת משאבים):** שימוש בנתונים המצרכיים כדי לזהות בתי ספר או רשויות שבהם ריכוז גבוה של בעלי סיכויי מוביליות נמוכים, ותקצובם על פי הפוטנציאל העתידי ולא רק על פי המצב הנוכחי.
- 2. ברמה הטקטית (ניהול תוכניות):** סינון ומיון מועמדים לתוכניות התערבות מותאמות, כדי לוודא שהסיוע מגיע למי שנמצא בסיכון הגבוה ביותר.

« מאבחון לטיפול באמצעות התאמת המענה

אחד היתרונות המובהקים של מודל עצי ההחלטה הוא היכולת לא רק לתת "ציון סיכון", אלא גם להסביר אילו משתנים תורמים לו. יכולת זו רלוונטית להתאמת המענה. לדוגמה, אם גורם הסיכון העיקרי הוא סביבתי-חינוכי (למשל, ותק מורים נמוך), ההתערבות הנדרשת שונה מאשר במקרה שבו גורם הסיכון הוא משפחתי. כלומר, ביכולתו של כלי זה גם להצביע על קבוצות באוכלוסייה שנמצאות בסיכון למוביליות נמוכה, וגם לסייע בעיצוב דרכי הפעולה להקטנת סיכון זה.

« מודל דינמי ללמידה מתמשכת

המודל המוצג בדוח זה תוכנן כאב-טיפוס בעל יכולת התפתחות. יתרונו המרכזי טמון באופיו הדינמי:

- 1. עדכון עיתי:** הפעלה מחזורית של המודל על נתונים עדכניים תאפשר לקבל תמונת מצב בזמן אמת ולהתאים את המדיניות לשינויים דמוגרפיים וחברתיים.
- 2. הרחבת בסיס הידע:** המבנה הגמיש של המודל יאפשר קליטה של משתנים חדשים ממקורות מידע נוספים (כגון נתוני רווחה, בריאות או חינוך בלתי פורמלי), תהליך הצפוי לשפר את דיוק התחזיות ואת יכולת האיתור של אוכלוסיות היעד שיפור מתמיד.

6. סיכום: כלי דינמי למדיניות עתידית

דוח זה מציג מהלך חדשני של פיתוח ויישום מודל ניבוי למוביליות חברתית-כלכלית בישראל. המודל מתבסס על נתונים מנהליים רחבי היקף ומציג מיפוי של תמונת המצב הקיימת, כמו גם הערכה כמותית של סיכויי המוביליות העתידיים ברמת הפרט.

המודל המוצג כאן הוא **אב־טיפוס** המדגים את הפוטנציאל הטמון בשילוב כלי למידת מכונה בתהליכי קבלת החלטות במגזר הציבורי, והוא אינו בגדר מוצר סופי וסטטי. חשיבותו של המהלך טמונה במעבר מתפיסה מחקרית רטרופקטיבית, לתפיסה ניהולית צופה פני עתיד.

« תובנות מתודולוגיות כבסיס למדיניות

אחת התרומות המרכזיות של דוח זה היא ההשוואה האמפירית בין גישות ניבוי שונות. הניתוח הראה כי בעוד מודל סטטיסטי מסורתי (רגרסיה לוגיסטית) ומודל למידת מכונה (עצי החלטה) הם בעלי כושר ניבוי כללי דומה, שני המודלים משלימים זה את זה בכך שהם חושפים תופעות שונות:

- מודל הרגרסיה מדגיש את כובד המשקל של **הפערים המבניים** הקבועים (כגון מגדר וקבוצה באוכלוסייה)
- מודל עצי ההחלטה חושף את כוחם של **הגורמים הסביבתיים והמשתנים**

הערך המוסף הייחודי של מודל עצי ההחלטה, כפי שהודגם בפרק 3, טמון ביכולתו לזהות **אינטראקציות מורכבות**. המודל הראה כי גורמים כמו ותק מורים או השכלת אם קשורים למוביליות באופן שאינו אחיד לכל חלקי האוכלוסייה, אלא משמשים מנופים למוביליות בהקשרים ספציפיים – בפריפריה החברתית או בקרב קבוצות מיעוט. תובנה זו היא המפתח למעבר ממדיניות המבוססת על ממוצעים למדיניות המבוססת על הקשרים.

« מדיניות מבוססת נתונים בזמן אמת

כוחו של הכלי הוא באופיו הדינמי. המציאות החברתית-כלכלית בישראל משתנה תדיר, ומודל המבוסס על נתוני עבר בלבד עלול להפוך ללא רלוונטי. המבנה הנוכחי מאפשר:

1. **עדכון שוטף:** הזנת נתונים של שנתונים חדשים ככל שאלו מתבגרים, מה שמאפשר למודל ללמוד שינויים במגמות ולעדכן את התחזיות.
2. **בחינת תרחישים:** היכולת לבדוק כיצד שינויים בפרמטרים מסוימים (למשל, עלייה בשיעור המורים האקדמאים ביישוב מסוים) עשויים להשפיע על תחזית המוביליות של התלמידים באותם יישובים.

« הזמנה להרחבת היריעה המחקרית

חשוב להכיר בכך שהמודל הנוכחי, על אף עוצמתו, מוגבל לנתונים המנהליים הקיימים הזמינים לו. הניתוח המצומצם שערכנו באמצעות נתוני הסקר החברתי (תיבה 1) חשף טפח מן הפוטנציאל הטמון במשתנים רכים שאינם מנהליים – מיומנויות דיגיטליות, שליטה בשפה וחוסן משפחתי. הממצאים מצביעים על כך שהתחשבות במשתנים הללו מעשירה את הידע על מנועי המוביליות, בעוד התעלמות מהם עלולה להותיר נקודות עיוורון חשובות. **לפיכך, דוח זה מהווה גם קריאה לפעולה: העמקת המחקר באמצעות איסוף שיטתי של נתונים נוספים (כגון נתוני רווחה, בריאות, חינוך בלתי פורמלי ומיומנויות רכות) ושילובם במודל, עשויה להוליד תובנות חדשות ופורצות דרך שאי אפשר לראותן כיום.**

« | ממדיניות אחידה למדיניות מותאמת

המודל המובא במחקר זה סולל את הדרך לשינוי פרדיגמה במדיניות החברתית: מעבר מפתרונות רוחביים ואחידים לפתרונות ממוקדים ומותאמים.

היכולת לזהות מבעוד מועד את הצמתים בחייו של צעיר שבהם התערבות היא יעילה – הנקודה שבה שילוב של נסיבות משפחתיות וסביבתיות מציב אותו בסיכון גבוה לקיבעון כלכלי – היא המפתח לקידום שוויון הזדמנויות. המודל מספק למקבלי ההחלטות את המצפן הנדרש כדי לנווט את המשאבים הציבוריים אל אותם הצמתים, ולהפוך את המוביליות החברתית משאיפה ערכית ליעד מדיד ובר-השגה.

7. מונחון מושגים

שלב	מושג	הסבר נגיש	דוגמה מעולם המוביליות
התשתית	למידת מכונה (ML)	לימוד המחשב לזהות דפוסים מנתוני עבר כדי לחזות את העתיד	המחשב לומד מנתוני עבר אילו גורמים (כמו השכלה) מובילים לעלייה בשכר
	מודל (Model)	"המוח" המתמטי שנוצר לאחר תהליך הלמידה כדי להפיק תחזיות	בנינו תוכנה שמקבלת נתונים על צעירים כדי לנבא מי מהם ישכר את מעמדו הכלכלי
שלב הקלט	קלט (Input)	הנתונים המוזנים למודל לצורך החיזוי	נתוני רקע על האדם: מקום מגורים, מספר אחים והכנסת ההורים
	אימון מודל (Training)	השלב שבו המודל נחשף לדוגמאות מן העבר כדי ללמוד את הקשרים בנתונים	המודל לומד על נתוני שנת 2010 כדי להבין מה הוביל למוביליות באותה שנה
	בחינת מודל (Testing)	בדיקה האם המודל יודע להתמודד עם נתונים חדשים שלא נחשף אליהם באימון	בדיקת המודל על נתוני 2024 כדי לראות אם התחזיות שלו אכן התממשו
סוגי המודלים	רגרסיה לוגיסטית	שיטה סטטיסטית לבדיקת השפעה של גורמים על תוצאה של "כן/לא"	בדיקת השפעת תואר אקדמי על הסיכוי של אדם לצאת ממעגל העוני
	מקדמים (Coefficients)	מספרים המראים את עוצמת ההשפעה של כל משתנה וכיוונה (חיובי/שלילי)	אם המקדם של "תואר" הוא חיובי גבוה, התואר מגדיל מאוד את סיכויי המוביליות
	יחס סיכויים (Odds Ratio)	מדד המראה כי כמה משתנה מסוים מעלה או מוריד את הסיכוי לתוצאה	האם תואר אקדמי מעלה את הסיכוי למוביליות פי 3 לעומת מי שאין לו תואר
	מודל עצי החלטה	קבלת החלטות לפי סדרת שאלות של "כן/לא" עד לקבלת תחזית	האם יש בגרות? אם כן, האם סיים תואר? אם כן, חיזוי למוביליות
פלט והחלטה	פלט (Output)	התשובה שהמודל מחזיר, לרוב בצורת הסתברות	המודל קובע שיש לאדם מסוים 75% סיכוי לעבור עשירון
	סף החלטה (Threshold)	הגבול שמעליו המודל יחליט לסווג מקרה "כן" (למשל מעל 0.5)	אם הסיכוי הוא מעל 0.5, המודל יחליט שהתחזית היא שהאדם יחווה מוביליות
	תחזית חיובית/שלילית	האם המודל חוזה שהאירוע יקרה (חיובי) או לא (שלילי)	חיובית: חיזוי שתתרחש מוביליות. שלילית: חיזוי שלא יהיה שינוי במעמד
מדדי איכות	גרף SHAP	הצגה גרפית של תרומת כל משתנה לתחזית ספציפית	גרף שמראה ש"שנות לימוד" תרמו 40% מן ההחלטה לחזות מוביליות חיובית

שלב	מושג	הסבר נגיש	דוגמה מעולם המוביליות
	AUC	ציון המודד כמה טוב המודל מבחין בין שתי קבוצות	ציון 0.85 אומר שהמודל מצליח מאוד להבחין בין מי שיחווה מוביליות למי שלא
	Precision (דיוק)	מתוך כל הפרטים שחזינו שיחוו מוביליות, כמה באמת חוו אותה?	מונע מצב של "חיובי שגוי" – זיהוי יתר של הצלחה חברתית שלא קרתה
	Recall (רגישות)	מתוך אלו שחוו מוביליות בפועל, כמה המודל הצליח לזהות?	בודק כמה מן ה"מטפסים" החברתיים המודל הצליח לאתר מראש
	Specificity (ספציפיות)	כמה המודל טוב בזיהוי אלו שלא יחוו מוביליות (מניעת אזעקות שווא)	היכולת לזהות נכון אנשים שיישארו באותו סטטוס כלכלי

מקורות

- הלר, א', ניסנוב, ז', בן-דוד, ג', אבנור, י' וגרינשטיין-וייס, מ. (2025). *מוביליות חברתית-כלכלית בישראל: סקירת ספרות*. דמ-063-25. מכון מאירס-ג'וינט-ברוקדייל. <https://brookdale.jdc.org.il/publication/socioeconomic-mobility-in-israel-a-literature-review/>
- פיינשטיין, ע', לבני-נבון ע' ופורזיקי ו'. (2026). *הוגנות בבינה מלאכותית בשירותים במימון ציבורי*. [דוח בהכנה]. מכון מאירס-ג'וינט-ברוקדייל.
- Barocas, S., & Selbst, A. D. (2016). Big Data's disparate impact. *California Law Review*, 104(3), 671-732. <http://dx.doi.org/10.2139/ssrn.2477899>
- Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States. *The Quarterly Journal of Economics*, 129(4), 1553-1623. <https://doi.org/10.1093/qje/qju022>
- Corak, M. (2013). Income Inequality, Equality of Opportunity, and Intergenerational Mobility. *Journal of Economic Perspectives*, 27(3), 79-102. <https://doi.org/10.1257/jep.27.3.79>
- OECD (2018). *A Broken Social Elevator? How to Promote Social Mobility*. OECD Publishing, Paris. <https://doi.org/10.1787/9789264301085-en>.
- Rajkomar, A., Dean, J., & Kohane, I. (2018). Machine learning in medicine. *New England Journal of Medicine*, 378(14), 1347-1358. DOI: 10.1056/NEJMra1814259
- Schleicher, A. (2018). *World Class: How to Build a 21st-Century School System, Strong Performers and Successful Reformers in Education*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264300002-en>

נספחים

נספח א: אוכלוסייה ומשתנים

רשימת המשתנים המלאה והגדרותיהם:

א. נתוני הורים:

- **השכלת הורים:** רמת ההשכלה הפורמלית של ההורים (לדוגמה: יסודית, תיכונית, על-תיכונית, גבוהה)
- **הכנסות מעבודה בגילי 15–19:** רמת ההכנסה של ההורים בתקופה שבה הפרטים היו בני 15–19, כמדד לרקע הכלכלי במשק הבית בשנות ההתבגרות
- **מצב משפחתי:** מבנה המשפחה בתקופת ההתבגרות (לדוגמה: הורים נשואים, הורה יחיד, צורות משפחה אחרות)
- **מספר אחים ואחיות:** מספר האחים והאחיות במשפחה הגרעינית

ב. דמוגרפיה:

- מגדר
- שיוך לקבוצה באוכלוסייה: יהודים שאינם חרדים, חרדים, ערבים
- מוצא: סיווג מוצא משפחתי לארץ מוצא
- שנת עלייה: שנת העלייה לישראל עבור מי שנולד/ה בחו"ל
- אשכול יישובי: דירוג האשכול החברתי-כלכלי של היישוב לפי סיווג האשכולות המקובל

ג. מאפייני מסגרת החינוך

- סוג הפיקוח על בית הספר
- סיווג סוג הפיקוח המוסדי (למשל: ממלכתי, ממלכתי-דתי, ערבי)
- שיוך מגזרי של מוסד הלימודים: שיוך המוסד למגזר (למשל: יהודי, ערבי, חרדי), בהתאם לסיווג הרשמי של מערכת החינוך
- השתייכות לחינוך מיוחד: אינדיקציה לכך שמסגרת החינוך מוגדרת כחינוך מיוחד (כן/לא).

ד. מאפייני כיתה הלימוד

- גודל הכיתה: מספר התלמידים הלומדים בכיתה
- יחס בנות-בנים: שיעור הבנות מתוך כלל התלמידים בכיתה
- יחס בין תלמידים ממוצא מזרחי ואשכנזי: התפלגות המוצא (מזרחי/אשכנזי) בקרב תלמידי הכיתה, כמדד להרכב החברתי-אתני
- המצב המשפחתי של הורי התלמידים בכיתה: התפלגות סוגי מבני המשפחה בקרב תלמידי הכיתה (לדוגמה: שיעור תלמידים ממשפחות דו-הוריות, חד-הוריות וכדומה)
- ממוצע הכנסות ההורים בכיתה: ממוצע ההכנסות של ההורים של כלל התלמידים בכיתה
- שיעור התלמידים שהכנסת הוריהם נמוכה: שיעור התלמידים בכיתה המשתייכים למשקי בית ברמות הכנסה נמוכות

ה. מאפייני צוות ההוראה

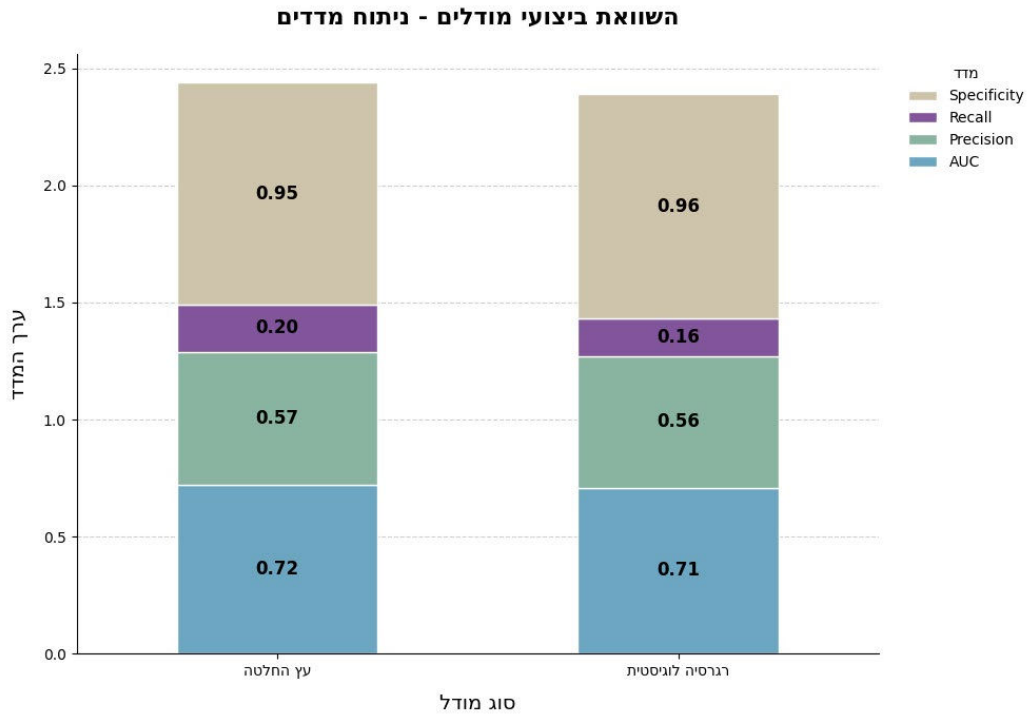
- גיל המורים: גיל המורים המלמדים בכיתה בבית הספר
- ותק המורים: שנות הוותק בהוראה של המורים המלמדים בבית הספר
- שיעור המורים בעלי תואר אקדמי: שיעור המורים בצוות ההוראה המחזיקים בתואר אקדמי (תואר ראשון ומעלה)

נספח ב: מתודולוגיה והערכת מודלים

לוח ב-1: רגרסיה לוגיסטית לעומת עצי החלטה

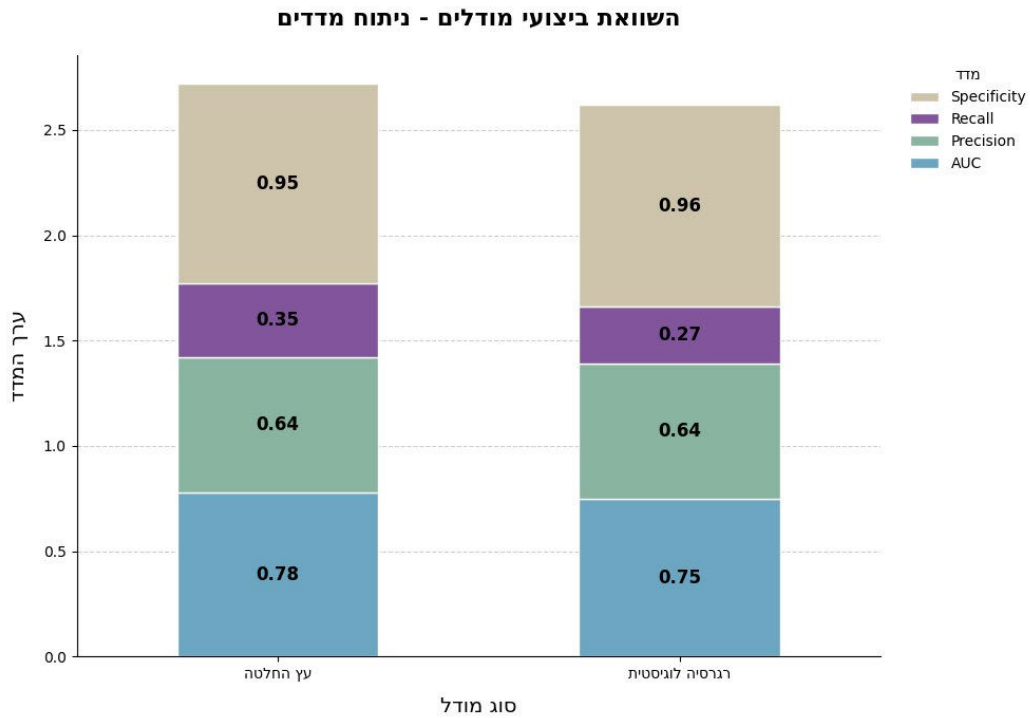
קריטריון השוואה	רגרסיה לוגיסטית	מודל עצי החלטה
מה השיטה עושה?	בוחנת כיצד כל גורם משפיע על הסיכוי שתתרחש תוצאה מסוימת. במחקר זה, הסיכוי למוביליות בין-דורית כלפי מעלה	מקבלת החלטות באמצעות סדרת שאלות ("כן/לא") ובונה תהליך שמוביל לחיזוי התוצאה. במחקר זה, חיזוי מוביליות בין-דורית כלפי מעלה
מתי השיטה עובדת טוב?	כאשר הקשרים בין המשתנים לתוצאה פשוטים יחסית וברורים	כאשר הקשרים מורכבים, יש תלות בין משתנים, או כאשר יש הרבה משתנים קטגוריאליים (משתנים שאינם מספריים אלא שייכים לקבוצות, כמו שיוך לקבוצה באוכלוסייה)
מה מקבלים מן המודל?	תובנות ברורות: אפשר להבין אילו גורמים מעלים או מורידים את הסיכוי לתוצאה, ובאיזו עוצמה	תחזיות מדויקות יחסית, אך פעמים רבות קשה יותר להבין מה בדיוק הוביל לתחזית
רמת דיוק	טובה, אך עשויה לרדת כאשר הנתונים מורכבים מאוד	לרוב מדויק יותר, במיוחד במצבי נתונים מורכבים ורבים
פרשנות המודל	קלה יחסית להסבר: המודל מספק מקדמים שמראים כיוון ועוצמת השפעה	מורכב יותר לפרש, אך כלים משלימים (כגון גרף SHAP) יכולים לסייע בהבנת התרומה של כל משתנה
יתרון מרכזי	מאפשר להבין מה משפיע על התוצאה ואת מידת ההשפעה, ולכן מהווה יתרון משמעותי ליצירת התערבויות וקבלת החלטות מדיניות	מספק יכולת חיזוי גבוהה, יעיל ליישום מעשי ולזיהוי קבוצות בסיכון (למשל: קבוצות עם סיכוי נמוך למוביליות)

תרשים ב-1: השוואת ביצועי המודלים עבור משתנה מנובא של מוביליות כלפי מעלה בהכנסה



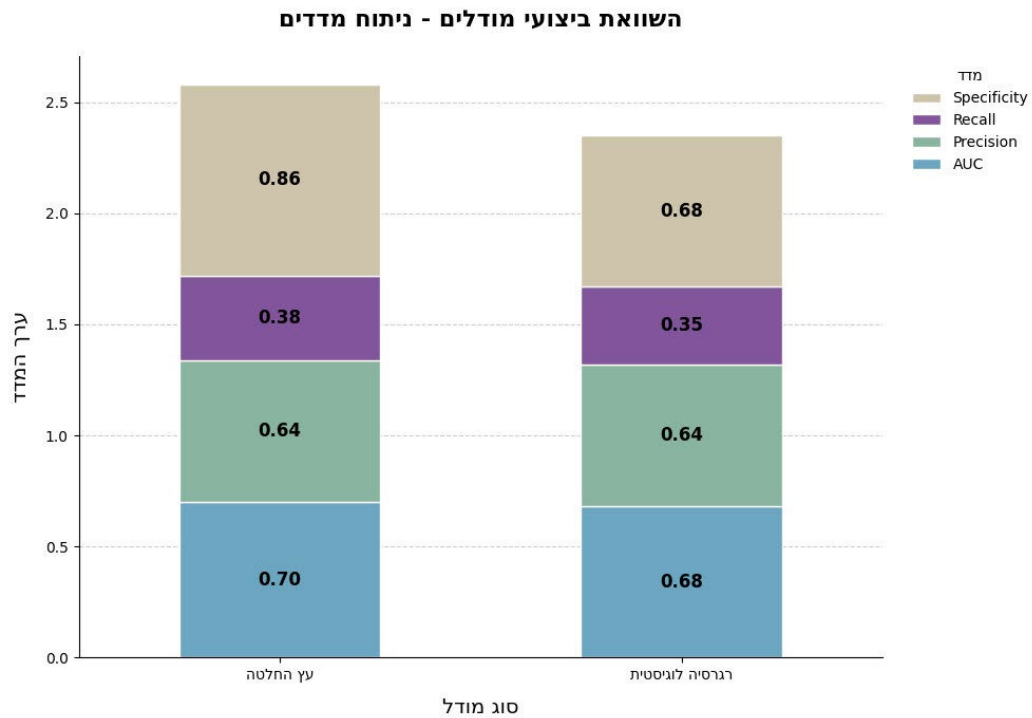
מבחינת ביצועי המודלים, מודל עצי ההחלטה ומודל הרגרסיה הלוגיסטית היו די דומים. שני המודלים הציגו תוצאות כמעט זהות בממד ה-Precision (0.57) כמו גם בממד ה-AUC: 0.72 עבור עצי ההחלטה ו-0.71 עבור הרגרסיה הלוגיסטית. הבדל קטן נצפה בממד Recall שבו עצי ההחלטה הצליחו לזהות יותר מקרים חיוביים (0.20 לעומת 0.16), וב-Specificity שם הרגרסיה הלוגיסטית הובילה במעט (0.96 לעומת 0.95). אף שהיו הבדלים קטנים במדדים השונים, הפער בביצועים בין המודלים מינורי, מה שמעיד על כך שבאופן כללי תפקודן של שני המודלים דומה מבחינת ביצועיהם. הדבר מהווה תיקוף לתוצאות המחקר, שכן הוא מצביע על כך שהממצאים במקרה הזה אינם תלויים דווקא במודל ספציפי אלא משקפים דפוסים עקביים בנתונים.

תרשים ב-2: השוואת ביצועי המודלים עבור משתנה מנובא של מוביליות כלפי מעלה בהשכלה



במדד AUC השיג עץ ההחלטה תוצאה של 0.78 והרגרסיה הלוגיסטית השיגה 0.75. במדד Specificity נצפו תוצאות דומות יחסית (0.95 לעומת 0.96). במדד Recall, זיהה עץ ההחלטה יותר מקרים חיוביים (0.35 לעומת 0.27), אך בשני המודלים מדד ה-Precision זהה, עם תוצאה של 0.64. אף על פי שהיו הבדלים קלים בין המדדים, הפער בביצועים בין המודלים מזערי, מה שמעיד על כך שפעולת שני המודלים דומה באופן כללי, גם עבור משתנה תוצאה זה. הדמיון בביצועים בין שני המודלים, המבוססים על אלגוריתמים שונים במהותם, מעיד על יציבות הממצאים.

תרשים ב-3: השוואת ביצועי המודלים עבור משתנה מנובא של "Sticky Floor"



במדד AUC השיג עץ ההחלטה תוצאה של 0.70, והרגרסיה הלוגיסטית השיגה 0.68. במדד Specificity עץ ההחלטה הוביל משמעותית עם 0.86 לעומת 0.68 ברגרסיה הלוגיסטית. במדד Precision שני המודלים השיגו תוצאה זהה של 0.64. במדד Recall, זיהה עץ ההחלטה מעט יותר מקרים חיוביים (0.38 לעומת 0.35). העליונות של עץ ההחלטה במדדי AUC ו-Specificity מראה כי זיהוי האינטראקציות עשוי לאפשר דיוק גבוה יותר מאשר הסתמכות על משתנה בודד.

נספח ג: ממצאים סטטיסטיים מפורטים

1. מוביליות בהכנסה

לוח ג-1: תוצאות רגרסיה לוגיסטית – תרומת המשתנים המנבאים מוביליות כלפי מעלה בהכנסה

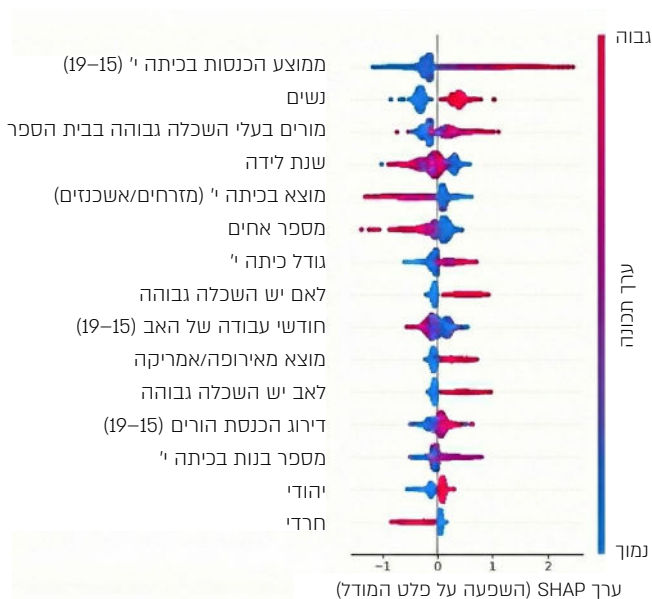
משתנה	מקדם	יחס סיכויים (OR)	גבול תחתון	גבול עליון	מובהקות
נשים	-1.094	0.335	0.329	0.341	כן
חרדים (לעומת יהודים לא חרדים)	-0.614	0.541	0.525	0.558	כן
ערבים (לעומת יהודים לא חרדים)	-0.589	0.555	0.538	0.572	כן
אם בעלת השכלה גבוהה	0.294	1.341	1.307	1.377	כן
ממוצע דירוג הכנסות ההורים בכיתה	0.240	1.271	1.136	1.423	כן
אב בעל השכלה גבוהה	0.221	1.247	1.217	1.279	כן
פיקוח בית ספר ממלכתי-דתי	0.206	1.228	1.182	1.276	כן
שיעור המורים בעלי השכלה גבוהה בביה"ס	0.187	1.206	1.132	1.284	כן
שיעור הבנות בכיתה	-0.145	0.865	0.826	0.906	כן
דירוג הכנסת ההורים	0.131	1.139	1.024	1.268	כן
דור שני למהגרים	0.114	1.121	1.089	1.155	כן
פיקוח בית ספר חרדי	-0.109	0.897	0.841	0.957	כן
מצב משפחתי של האם – נשואה	0.099	1.104	1.062	1.147	כן
מוצא מאירופה/אמריקה	0.097	1.102	1.074	1.130	כן
מוצא מעורב	0.097	1.102	1.066	1.138	כן

2. מוביליות בהשכלה

לוח ג-2: תוצאות רגרסיה לוגיסטית – תרומת המשתנים המנבאים מוביליות כלפי מעלה בהשכלה

משתנה	מקדם	יחס סיכויים (OR)	גבול תחתון	גבול עליון	מובהקות
ממוצע דירוג הכנסות ההורים בכיתה	1.578	4.847	4.321	5.438	כן
יחס בין מוצא מזרחי-אשכנזי בכיתה	-1.297	0.273	0.256	0.292	כן
שיעור המורים בעלי השכלה גבוהה בביה"ס	0.941	2.562	2.395	2.741	כן
פיקוח בית ספר ממלכתי-דתי (לעומת ממלכתי)	0.938	2.555	2.458	2.657	כן
חרדים (לעומת יהודים שאינם חרדים)	-0.747	0.474	0.459	0.490	כן
אם בעלת השכלה גבוהה	0.702	2.018	1.966	2.072	כן
נשים	0.608	1.836	1.804	1.869	כן
דור שלישי למהגרים	0.586	1.797	1.738	1.857	כן
דור שני למהגרים	0.541	1.718	1.666	1.772	כן
שיעור משפחות חד-הוריות בכיתה	-0.516	0.597	0.503	0.708	כן
אב בעל השכלה גבוהה	0.511	1.666	1.626	1.708	כן
שיעור תלמידים מרקע של הכנסה נמוכה בכיתה	-0.443	0.642	0.564	0.731	כן
מוצא מאירופה/אמריקה (לעומת אפריקה/אסיה)	0.424	1.529	1.489	1.570	כן
שיעור התלמידים המהגרים בכיתה	-0.375	0.687	0.632	0.747	כן

תרשים ג-1: גרף SHAP תרומת המשתנים המנבאים מוביליות כלפי מעלה בהשכלה

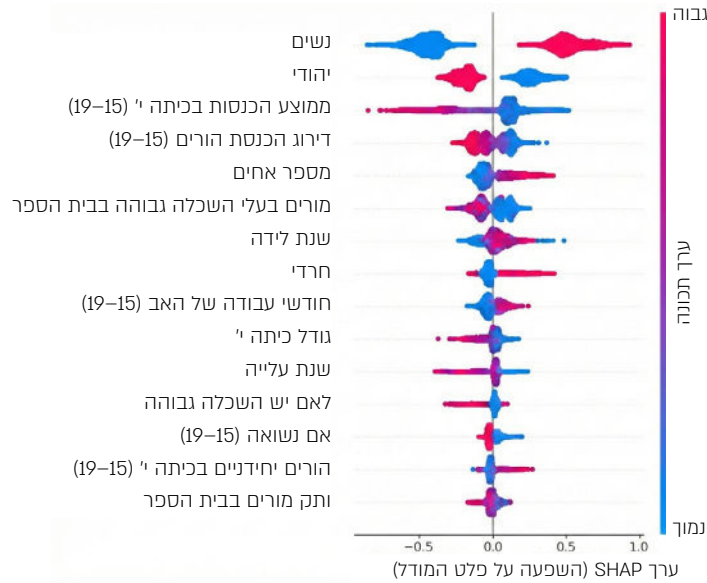


3. היעדר מוביליות ("רצפה דביקה")

לוח ג-3: תוצאות רגרסיה לוגיסטית – תרומת המשתנים המנבאים היעדר מוביליות

משתנה	מקדם	יחס סיכויים (OR)	גבול תחתון	גבול עליון	מובהקות
נשים	0.876	2.401	2.364	2.439	כן
ערבים (לעומת יהודים שאינם חרדים)	0.565	1.760	1.715	1.806	כן
חרדים (לעומת יהודים שאינם חרדים)	0.562	1.754	1.711	1.798	כן
פיקוח בית ספר ממלכתי-דתי	-0.237	0.789	0.761	0.818	כן
ממוצע דירוג הכנסות ההורים בכיתה	-0.206	0.814	0.736	0.900	כן
אם בעלת השכלה גבוהה	-0.186	0.830	0.809	0.852	כן
שיעור המורים בעלי השכלה גבוהה בביה"ס	-0.153	0.858	0.813	0.906	כן
דירוג הכנסת ההורים	-0.135	0.874	0.793	0.963	כן
אב בעל השכלה גבוהה	-0.130	0.878	0.857	0.899	כן
מצב משפחתי של האם – נשואה	-0.117	0.889	0.859	0.921	כן
שיעור תלמידים מרקע של הכנסה נמוכה	0.106	1.112	1.000	1.236	לא
דור שלישי למהגרים	0.099	1.104	1.071	1.137	כן
מוצא מעורב	-0.094	0.910	0.882	0.940	כן
מוצא מאירופה/אמריקה (לעומת אפריקה/אסיה)	-0.090	0.914	0.892	0.936	כן

תרשים ג-2: גרף SHAP המציג את תרומת המשתנים המנבאים היעדר מוביליות



לוח ג-4: תוצאות מבחנים סטטיסטיים

גודל האפקט (Cohen's d)	סטיית תקן	הסתברות ממוצעת למוביליות כלפי מעלה של הפרט	השכלת מורים (בחלוקה לשלוש קבוצות שוות בגודל התלמידים, לפי שיעור המורים האקדמאים)	אשכול חברתי-כלכלי
0.70	0.11	0.19	נמוך	נמוך (5-1)
	0.15	0.26	בינוני	
	0.16	0.29	גבוה	
0.58	0.16	0.27	נמוך	גבוה (10-6)
	0.15	0.34	בינוני	
	0.15	0.36	גבוה	
גודל האפקט (Cohen's d)	סטיית תקן	הסתברות ממוצעת למוביליות כלפי מעלה של הפרט	קבוצת השכלת אם	קבוצה באוכלוסייה
0.5	0.14	0.4	בעלת השכלה גבוהה	יהודי (שאינו חרדי)
	0.14	0.33	ללא השכלה גבוהה	
0.3	0.10	0.21	בעלת השכלה גבוהה	ערבי
	0.10	0.18	ללא השכלה גבוהה	
0.42	0.10	0.20	בעלת השכלה גבוהה	חרדי
	0.09	0.16	ללא השכלה גבוהה	