המעלקה להערכה

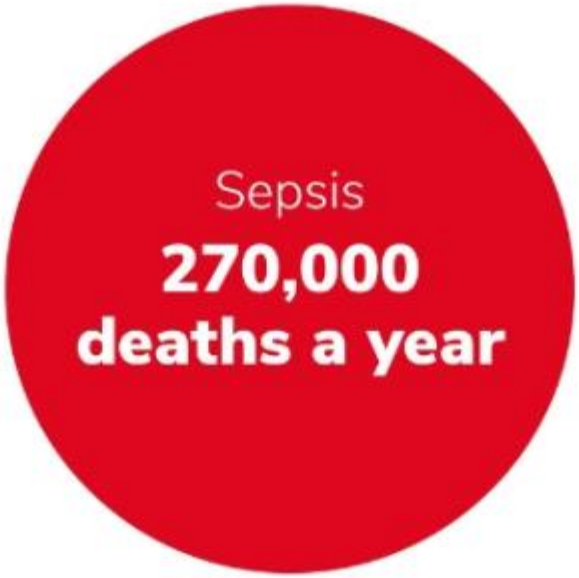
האוניברסיטה הפתוחה

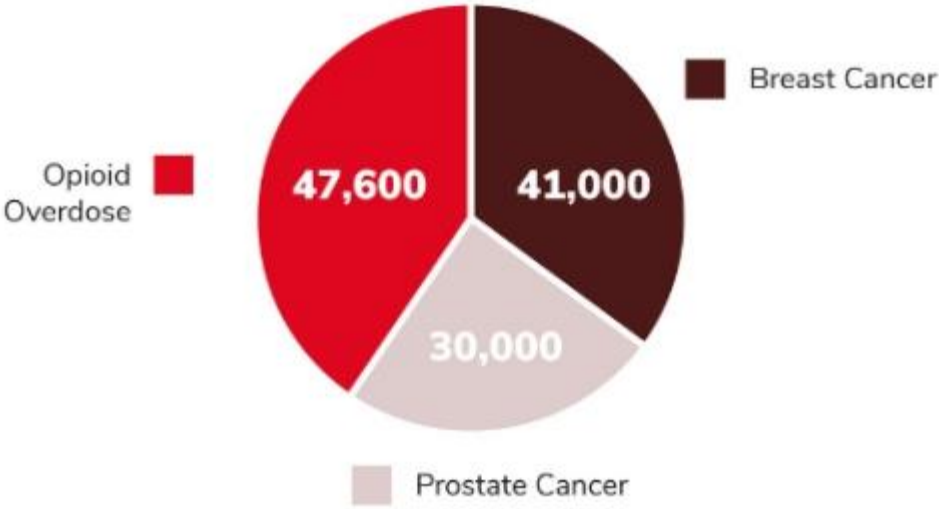
על מודלים של ניבוי במוסדות להשכלה גבוהה

ניצן וינטראוב
עינת נוטע-קורן

הכינוס ה- 16 של האגודה הישראלית לפסיכומטריקה (אפ"י)
פברואר, 2020

Each year in the United States, sepsis takes more lives than *opioids, breast cancer, and prostate cancer combined.*





= 118,600 deaths a year

*all numbers are approximate

Learn the signs at Sepsis.org. You could save a life.

אירגון מרכאז שלוקים באלה דם

Harborview Medical Center

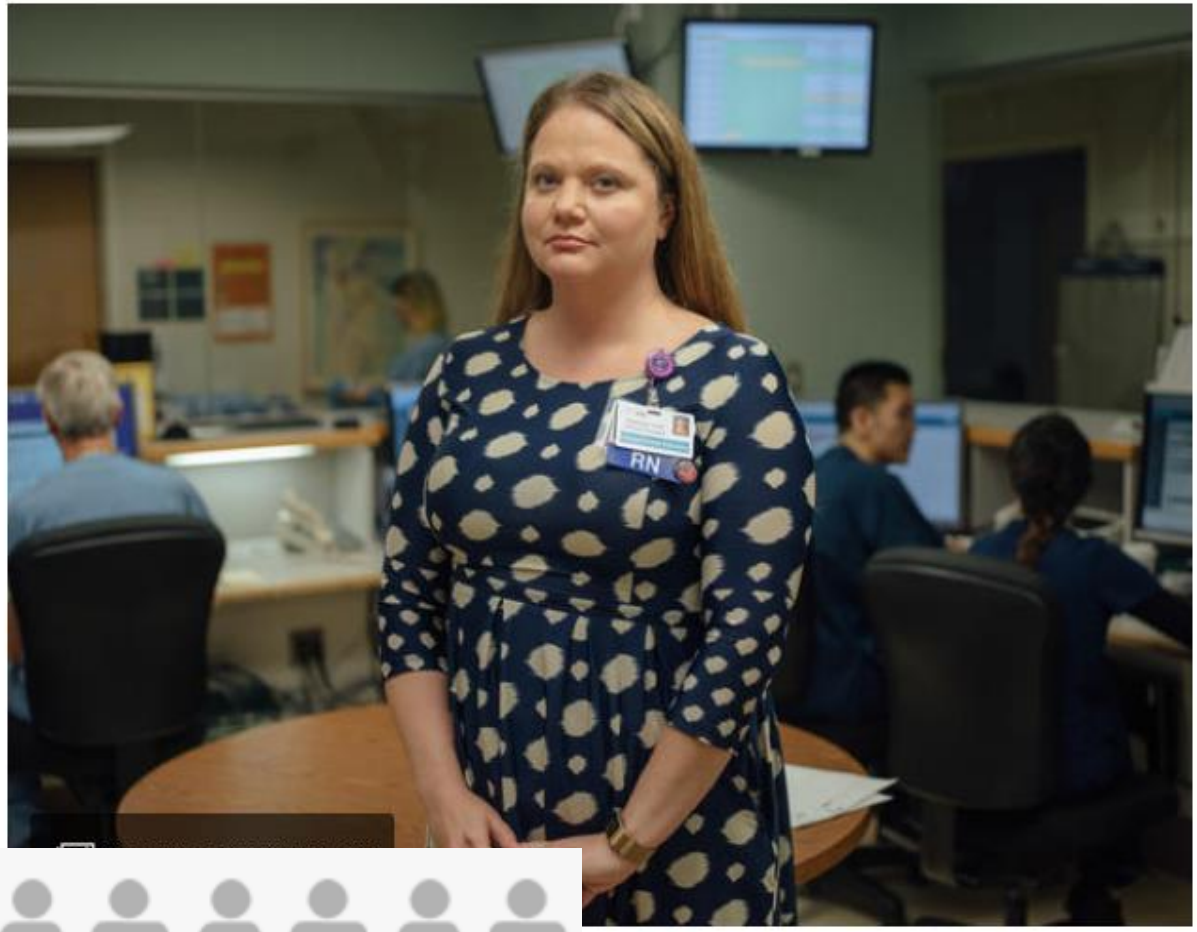
Synergy Between Nurses And Automation Could Be Key To Finding Sepsis Early

By RICHARD HARRIS • FEB 22, 2018

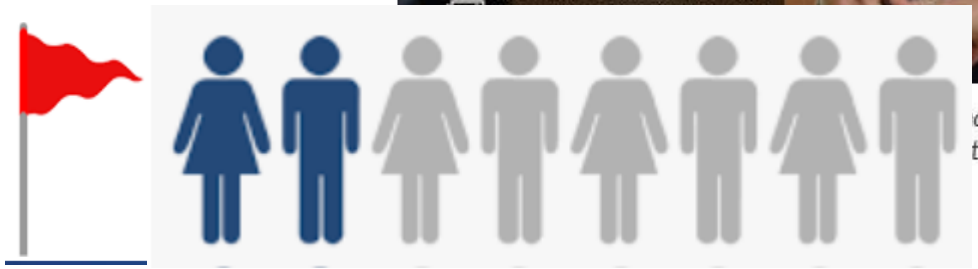
Share

Tweet

Email



ate sepsis care at Harborview Medical Center in Seattle.
t be developing sepsis antibiotics within three hours.



מיומת ה-

הסמכה

הסמכת הגורם האנושי

השכלה גבוהה

בסיית נשירה

NASPA Student Affairs Administrators in Higher Education

April 2017

PREDICTIVE ANALYSIS OF STUDENT DATA

A Focus on Engagement and Behavior

December 12, 2017

Big data could solve the college-dropout problem

Share this page:

Colleges have access to extraordinary amounts of data — not just on students' academic performance, but also on their eating habits, social life and daily routines. Administrators could use such data to identify which students are at risk of dropping out and could then intervene to give them additional support.

predictive models מודלים חיזוי

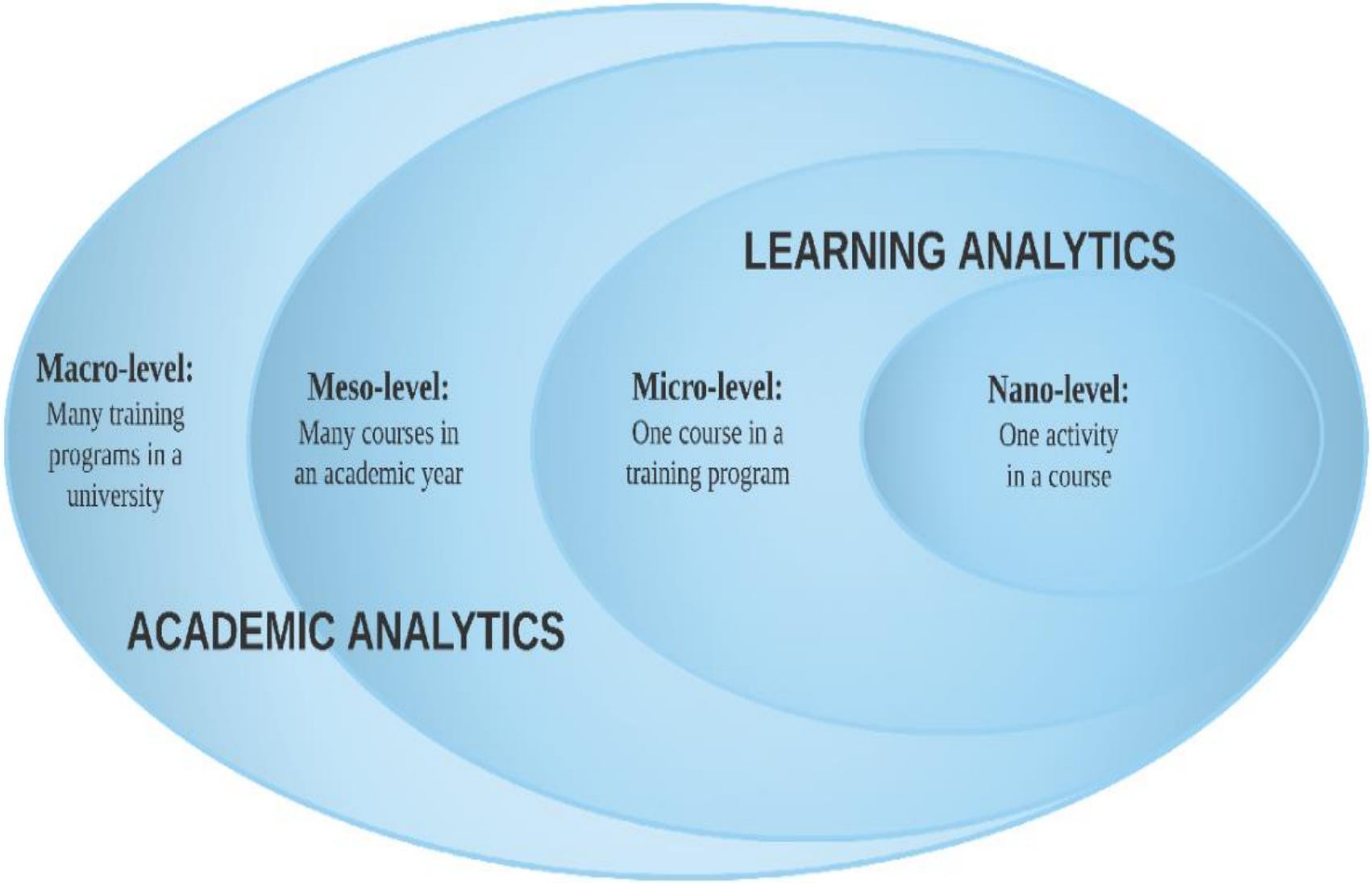
אלגוריתמים סטטיסטים מורכבים שבוחנים לעומק מידע עשיר ומגוון שהצטבר על סטודנטים מהעבר, ומזהים בו תבניות שמנבאות התנהגות/התנהגויות עתידיות של סטודנטים

Predictive analytics refers to the use of highly sophisticate statistical algorithms that sift through large amounts of assorted historical data to identify notable patterns and thus predictive future related outcomes *

* taken from: CCRC Working Paper No. 103, by S. Klempin, M. Grant and M. Ramos (2018)

שימושים במוסדות להשכלה גבוהה

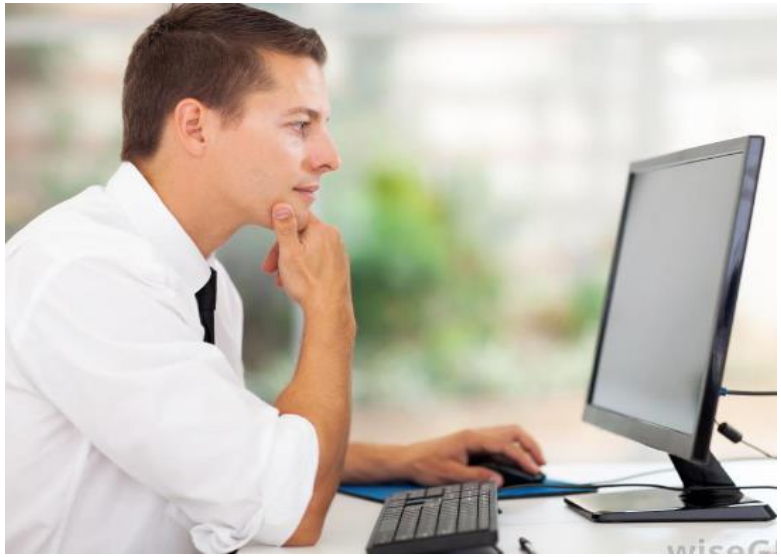
- איתור והפניית משאבי ייעוץ למי שזקוק
- המלצות לקורסים (באמצעות יועץ או אוטומטי)
- למידה מותאמת *
- תכנון השווק, צפי כספי, תכנון מלגות

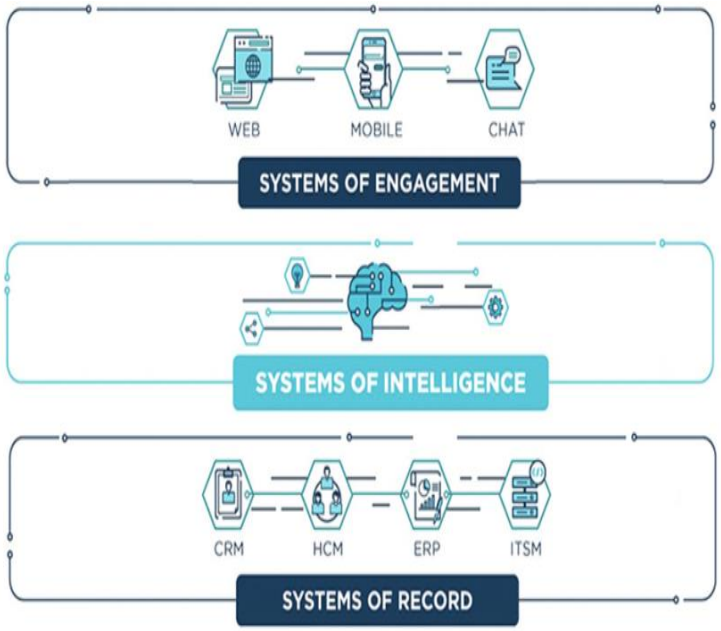


Hervatis, V, Loe, A, Barman, L., O'Donoghue, Zary, N (2015)

Learning Analytics in Higher Education

Academic Analytics in Higher Education





Tornbourgh, Civitas Learning, 2018

אמון, תביעות שוויון וביקורת

- התנגדות ואי-אמון של המשתמשים/ במודל ובתוצאותיו
- אי-הבנה של האלגוריתמים ("קופסה שחורה")
- העדר שקיפות של המשתנים שנלקחים בחשבון
- העדר הכשרה מספקת של היועצים, של מה לעשות עם הממצאים
- סוגיות של פרטיות ואתיקה:
  - הטיה/פגיעה בקבוצות חלשות, הנצחת אי-שוויון והעמקתו
  - אי שמירה על פרטיות
  - חשש לפגיעה ביצירת הקשר יועץ-סטודנט
  - תיוג של הסטודנט

אבני דרך עיקריות ליישום והטמעה

1. יעד ארגוני והגדרת צרכים: לשם מה
2. בניית תשתית בארגון (תקציב, כ"א, מיחשוב)
3. איסוף נתונים: אלו נתונים, כיצד אוספים, בדיקת איכות
4. מדיניות של פרטיות ואבטחת המידע
5. פיתוח מודל: להיות מודע להטיות אפשריות, שקיפות
6. הכשרת המשתמשים (יועצים), רתימה שלהם, העצמה, ליווי

ללמוד מאחרים

האוניברסיטה הפתוחה

איך לומדים באוניברסיטה הפתוחה?

אין תנאי קבלה

גמישות

הרשמה לקורס

מטרות לימוד מגוונות

גמישות מרבית בארגון הלימודים | הקורסים של האוניברסיטה הפתוחה | קבוצות הלימוד | צוות ההוראה

האוניברסיטה הפתוחה מאפשרת לכל אחד ואחת מכם גמישות מירבית בארגון הלימודים: אפשר להחליט אם ללמוד קורס אחד או מספר קורסים במהלך הסמסטר (ההרשמה באוניברסיטה הפתוחה היא לסמסטר), להחליט אם למקד את הלימודים בתחומים מסוימים או ללמוד תואר רב-תחומי, לבחור את קבוצת הלימוד המתאימה לסגנון הלמידה שלכם ולתנאיכם האישיים, ללמוד מכל מקום ובכל זמן את חומרי הלימוד האיכותיים והידידותיים, ולהיעזר במפגשי הנחיה במגוון ימים ושעות באחד מ-60 מרכזי הלימוד הפרושים ברחבי הארץ או במפגשי הנחיה מקוונים.

אין תנאי קבלה

גמישות

הרשמה לקורס

מטרות מגוונות

יכולות מגוונות, רקע מוקדם

הרבה סטודנטים פעילים

יחסית הרבה נשירה

האוניברסיטה הפתוחה

מאגר נתונים מסודר - משאבים

טכנולוגיה מתקדמת מפותחת

מנגנוני תמיכה וייעוץ מפותחים

הכנת המודל

המטרה

- לסייע לאוניברסיטה להקצות משאבי תמיכה מותאמים לקבוצות שונות של סטודנטים
- לאתר סטודנטים ברמות שונות של סיכון לנשירה מהלימודים

מודלים ראיכוי - אפשרויות

- התמדה לאורך התואר
- התמדה לסמסטרים הבאים
- התמדה בתוך קורס (מבוסס על נתוני אתר הקורס)
- עמידה בדרישות הקורס מבחינת הגשת המטלות
- הצלחה בקורס

מודל ליצוי הבלמה בקורס - אוכלוסייה

כלל הסטודנטים לקורסים אקדמיים באוניברסיטה הפתוחה

הוצאו מהמודל:

- סטודנטים חדשים, בסמסטר הראשון
- סטודנטים בקורסים לא אקדמיים
- עבודות סמינריוניות

קבוצות המשתנים

הישגים לימודיים

משתנים דמוגרפים

שלב בתואר

מדדי הקורס

משתנים לימודיים שונים

קצב לימודים

הכרות הבעיה ובסיס הנתונים קריטי לבחירת המשתנים למודל ובנייה שלהם!
השפעה ישירה על ביצועי המודל

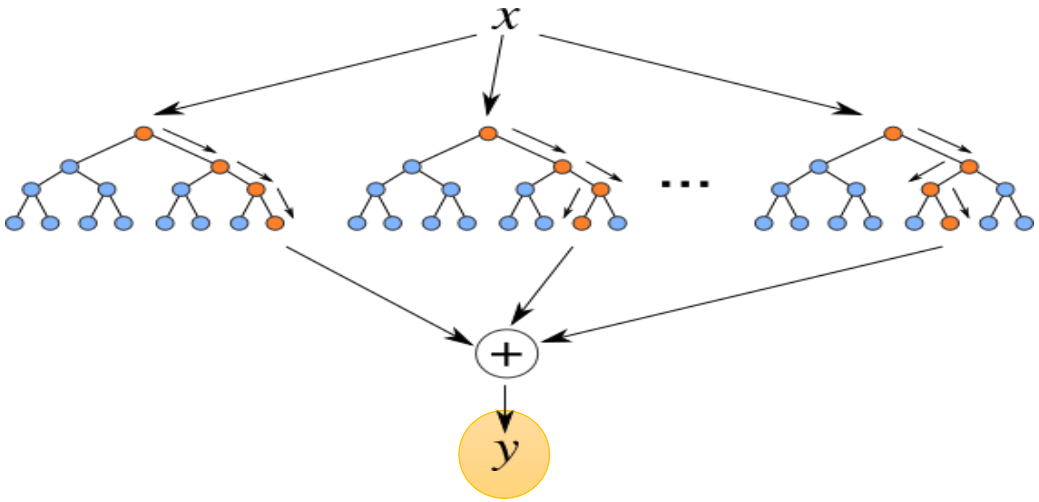
בחירת האלגוריתם

לצורך בחירת המודל הנכון בדקנו מודלים ממגוון משפחות:

- מודלים מבוססי עצים – Random Forest
- Gradient boosting (XGBoost)
- מודלי רגרסיה מקומית (שכנים קרובים) – KNN
- רשת עצבית מלאכותית – ANN
- ועוד...

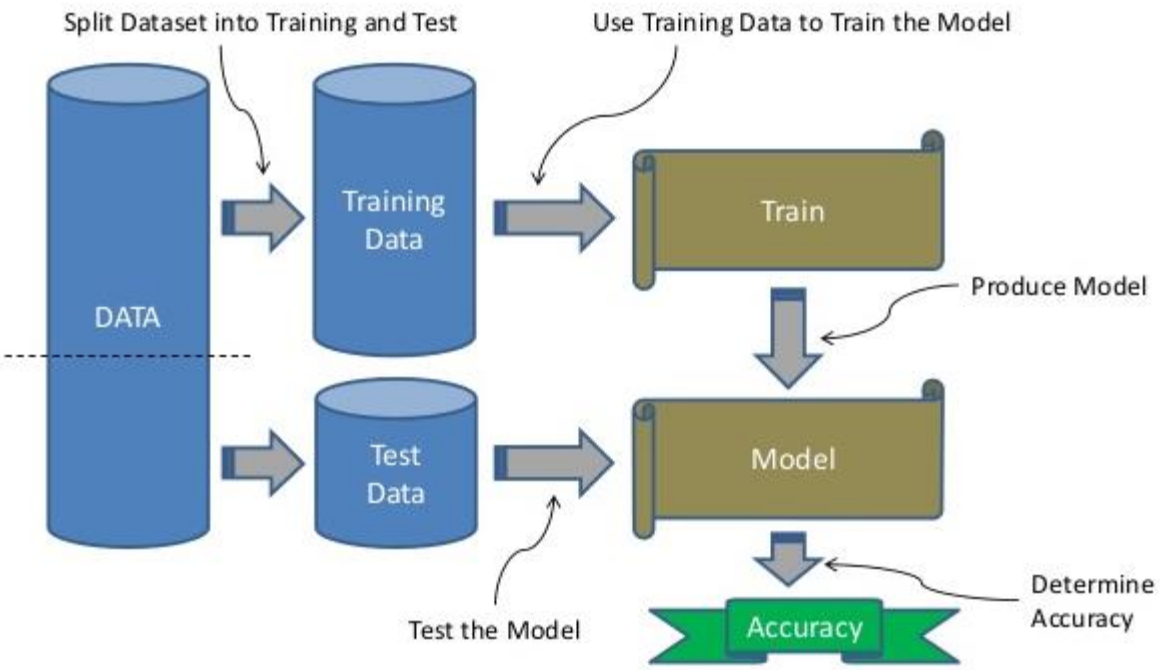
קבלת עץ המודל (Random Forest)

- אלגוריתם Random Forest מייצר אלפי עצי החלטה מגוונים השונים אחד מהשני.
- אקראיות בתצפיות – לכל עץ נבחר סט נתונים מעט שונה באופן אקראי.
- אקראיות במשתנים – לכל עץ נבחרים חלק מהמשתנים האפשריים לקבלת העץ.



חלוקת הנתונים והפרדתם

- אימון על סט נתוני נפרד ושימוש ב Cross-validation
- בחירת המודל לפי מדדי דיוק על סט נתוני "מבחן":
  AUC , Specificity , Sensitivity ,Accuracy



איכות המודל – דיוק

ברמת הרשומה:

	Accuracy	Balanced Accuracy	Specificity	Recall	Precision	F1 - Score
Random Forest	86.0	85.9	85.8	85.9	70.0	0.77

*כל המדדים נבדקו על נתונים חדשים (שהמודל לא ראה)

בפשטות:

- נוכל לחזות בדיוק של **86%** האם סטודנט יצליח בקורס אליו נרשם!
- Specificity – אחוז המצליחים בקורס שזיהינו מכלל המצלחים.
- Recall – אחוז הנכשלים שזיהינו מכלל הנכשלים.
- Precision – אחוז הנכשלים שזוהו מקרב המסווגים כנכשלים.

איכות המודל – דיוק

מודל

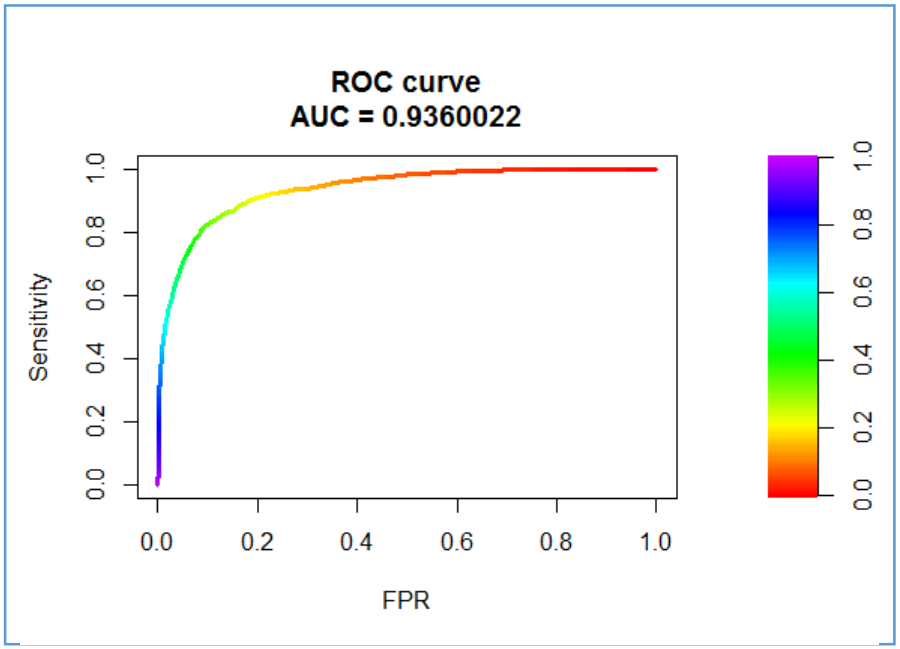
	מציאות	
	נכשל – מציאות	עבר – מציאות
נכשל – חיזוי	22%	10%
עבר – חיזוי	3%	65%

הדגמה:
עבור 50,000
הרשמות לקורסים

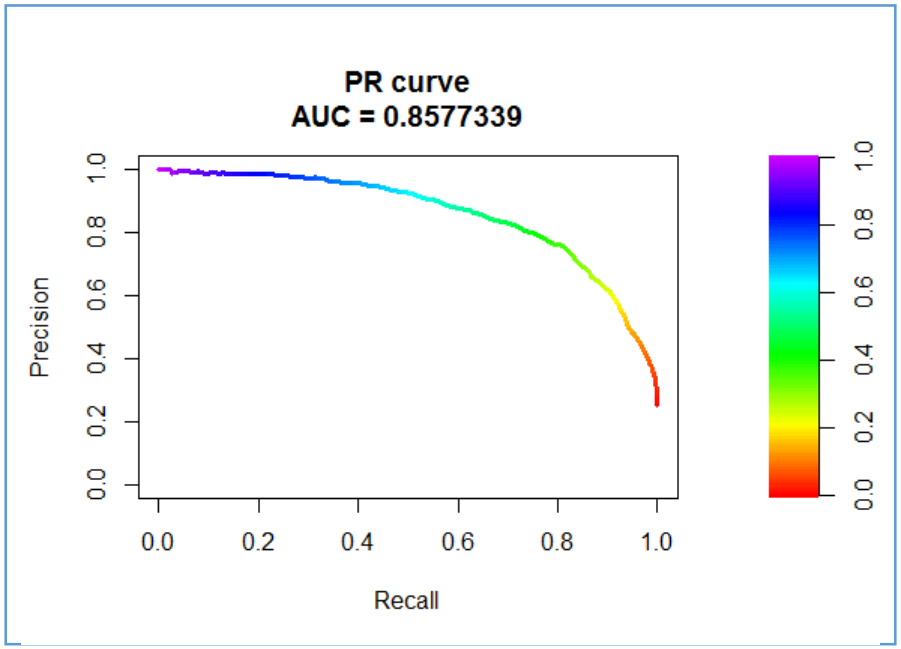
	נכשל – מציאות	עבר – מציאות
נכשל – חיזוי	11,000	5,000
עבר – חיזוי	1,600	32,400

איכות המודל – דיוק

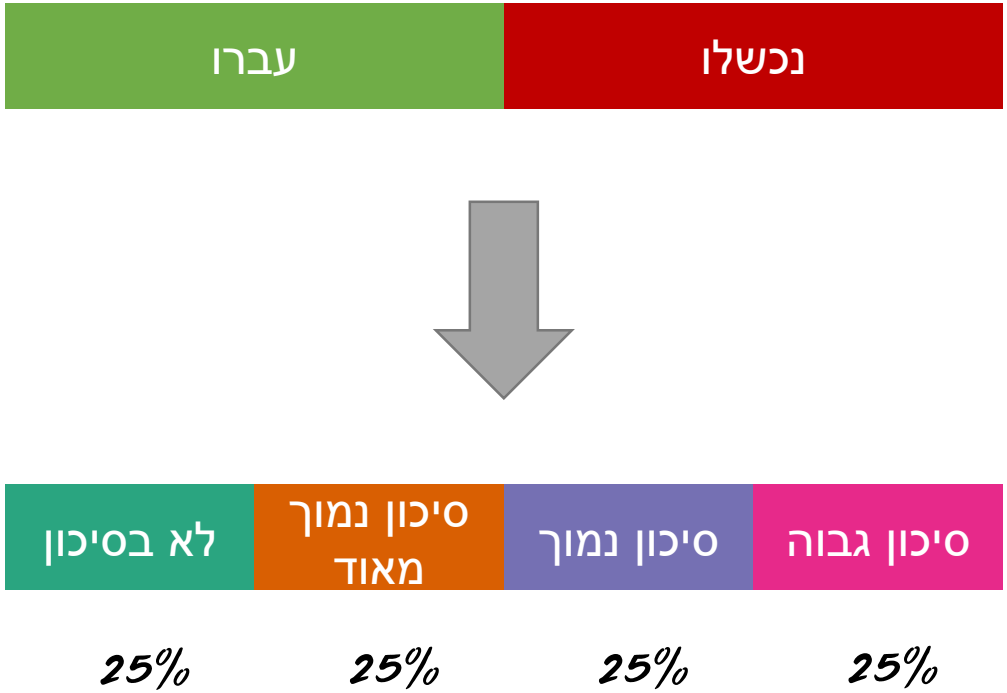
Sensitivity VS FPR



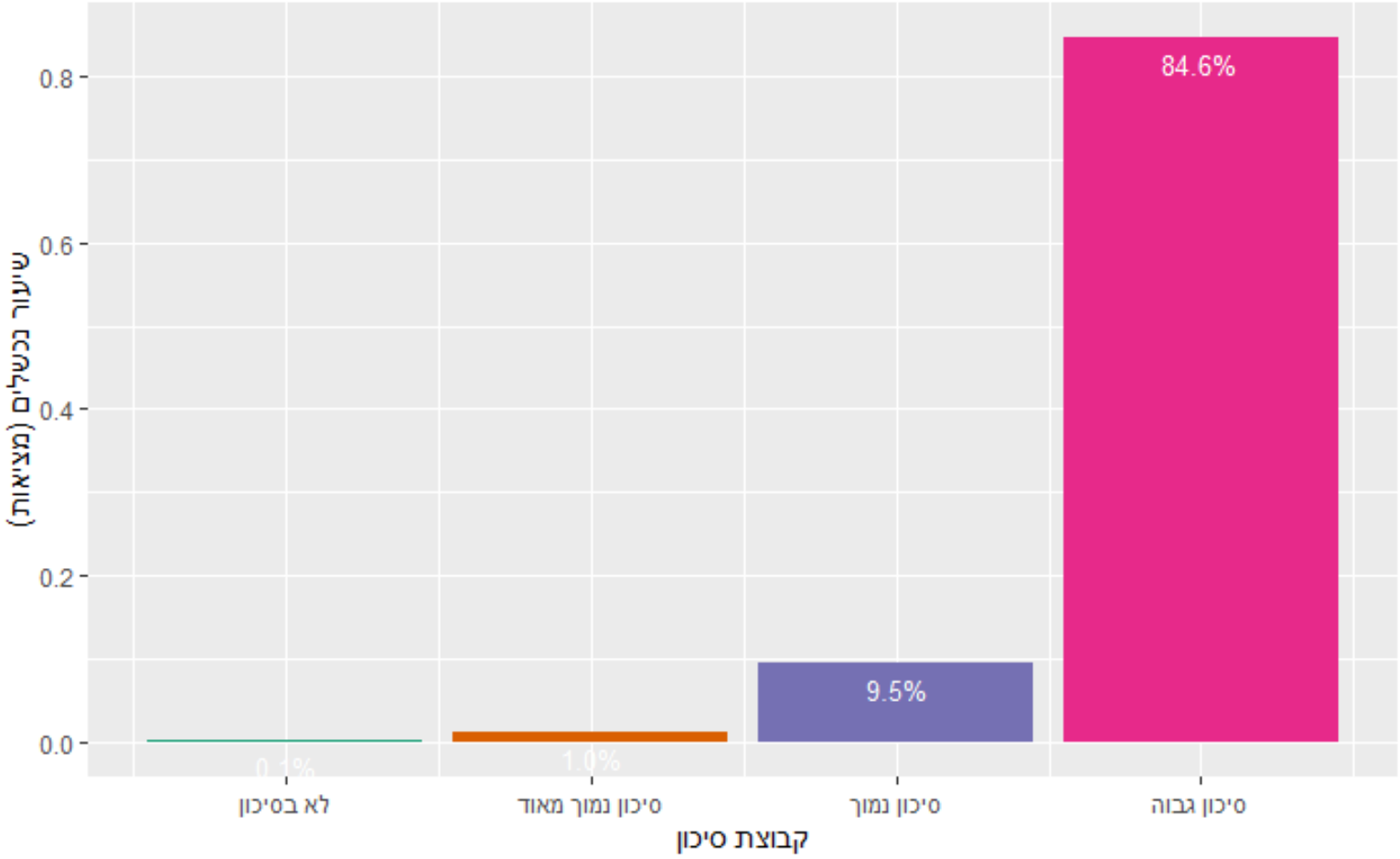
Precision VS Recall



חלוקה למספר קבוצות סיכון



דוגמה – תלוקה למספר קבוצות סיכון



לסיום – הרבה אתגרים בפתח

- נדרשות עוד בדיקות וטיוב המודל/ים לפני יישום (התמודדות עם הטיות וכו')
- הטמעה במערכות
- הכשרה של הגורם האנושי, יצירת אמון
- מדיניות פרטיות

מבחינת ההיתכנות הסטטיסטית, הבשורה היא ש:

אפשר להגיע למודלים ברמת דיוק גבוהה תוך שימוש בנתונים מובנים "קונבנציונליים"

המחלקה להערכה

תודה רבה!