

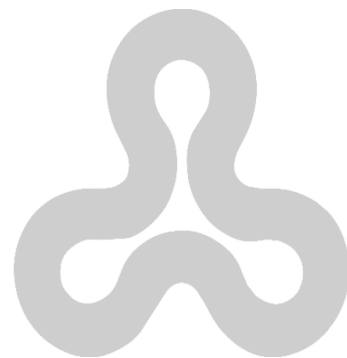


המגזק אהערכה

האוניברסיטה
הפתוחה

על מודלים של ניבוי במוסדות אהשכלה גבוהה

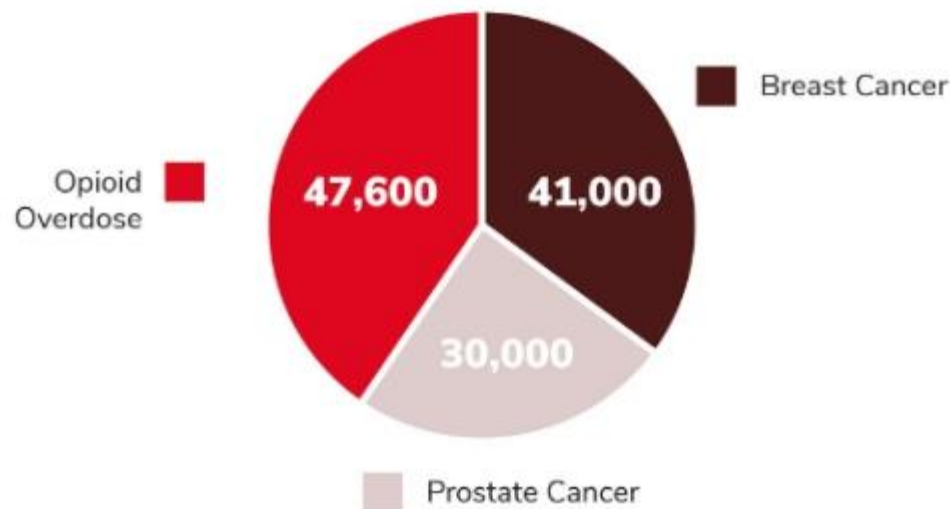
ניצן וינטראוב
עינת נוטע-קורן



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.

**Learn the signs
at [Sepsis.org](https://www.sepsis.org).
You could save a life.**



= 118,600 deaths a year

*all numbers are approximate

איגוד מראש של אוקייס באא 38



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.



מיילת ה-



העצמה

העצמת הגורם האנושי



השכחה גבוהה



בעיית נשירה

PREDICTIVE ANALYSIS OF STUDENT DATA

A Focus on Engagement and Behavior



NEW
AMEI

NEW
AMEI

MANUELA

IRIS PAL

THE
PE
AN
HI
CH
A
A

A Gui

A Land



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 אגז'קטיביים*

אלגוריתמים סטטיסטיים 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 sophisticated statistical algorithms that sift through large amounts of assorted historical data to identify notable patterns and thus predictive future related outcomes *

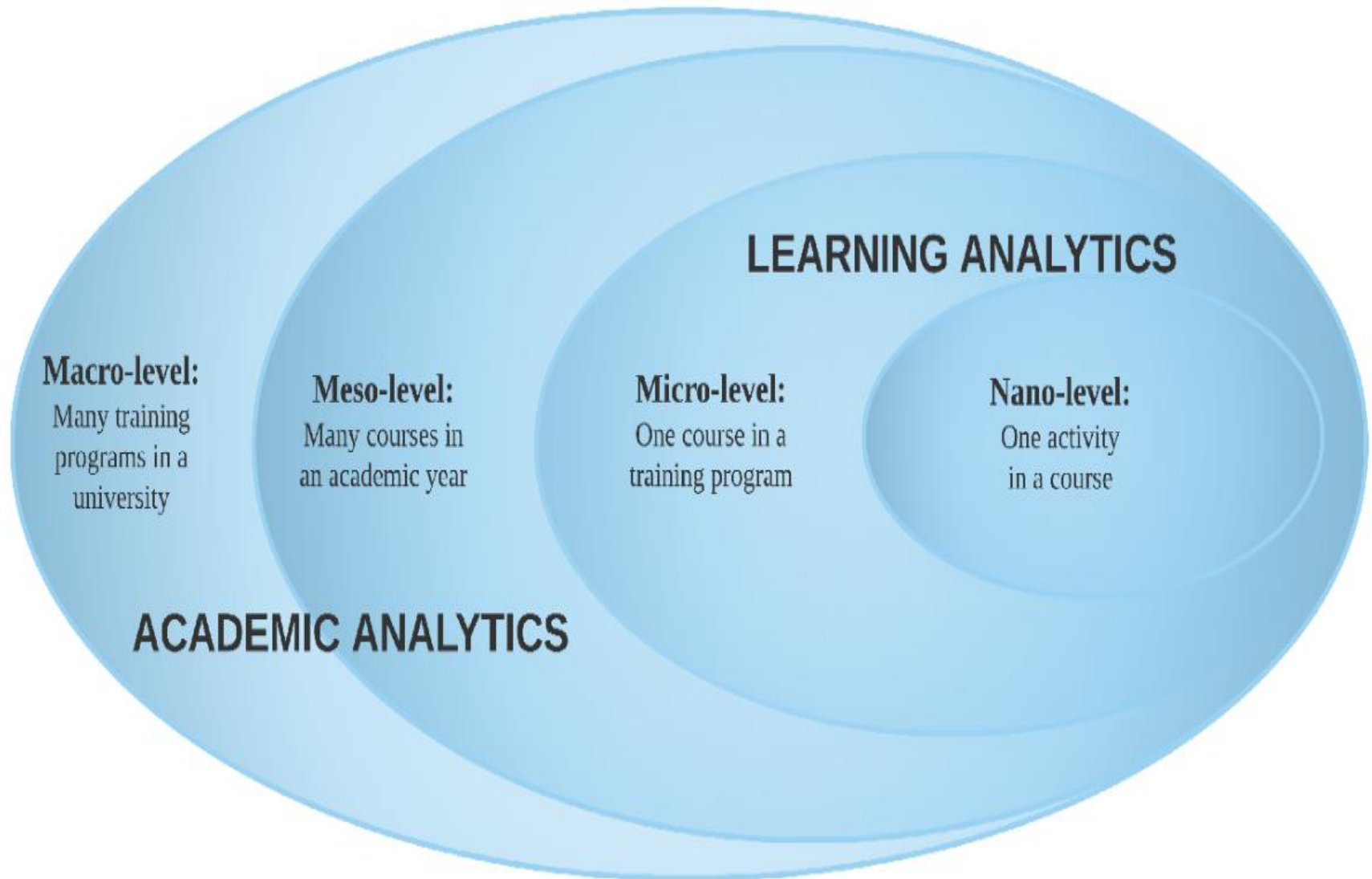
* taken from: CCRC Working Paper No. 103, by S. Klempin, M. Grant and M. Ramos (2018)



שימושים במוסדות אהשכ"ה גבוהה

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





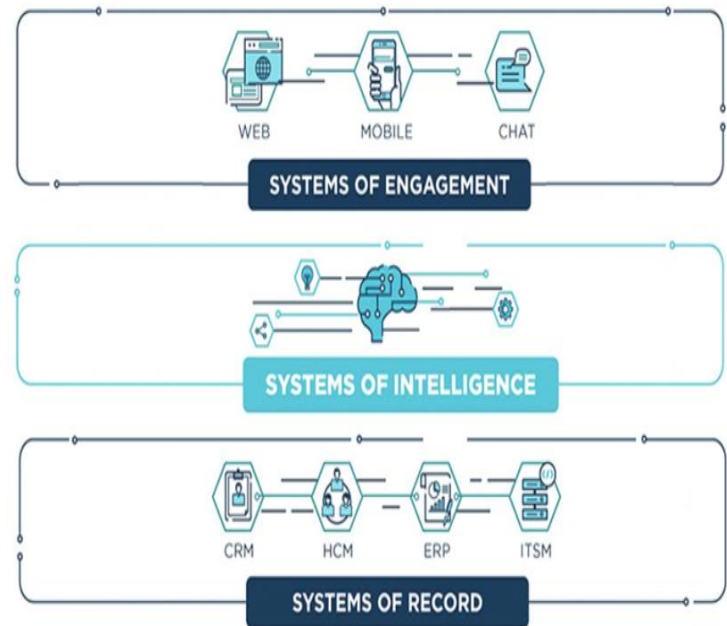
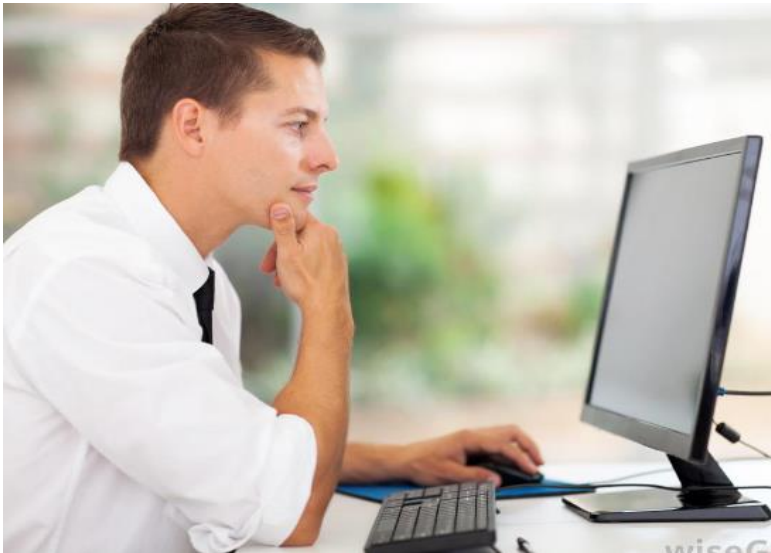
Hervatis, V, Loe, A, Barman, L., O'Donoghue, Zary, N (2015)



Learning Analytics in Higher Education

Academic Analytics in Higher Education





Tornborough, Civitas Learning, 2018



אסמ"ס, ג'פיסוט שליילוט וביקורוט

- 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 ולהטמעה

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

צוות ההוראה

קבוצות הלימוד

הקורסים של האוניברסיטה הפתוחה

גמישות מרבית בארגון הלימודים



ככה לומדים באוניברסיטה הפתוחה

האוניברסיטה הפתוחה מאפשרת לכל אחד ואחת מכם גמישות מרבית 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 ולתנאים האישיים, ללמוד מכל מקום ובכל זמן את חומרי הלימוד האיכותיים והוידיוותיים, ולהיעזר במפגשי הנחיה במגוון ימים ושעות באחד מ-60 מרכזי הלימוד הפרושים ברחבי הארץ או במפגשי הנחיה מקוונים.

יכולות מגוונות, רקע מוקדם

הרבה סטודנטים פעילים

יחסית הרבה נשירה

אין תנאי קבלה

גמישות

הרשמה לקורס

מטרות מגוונות



האוניברסיטה הפתוחה



מאגר נתונים מסודר - משאבים

טכנולוגיה מתקדמת מפותחת

מנגנוני תמיכה וייעוץ מפותחים



הצגת המוצר



המטרה

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



מוצאים לביצוע - אפשרויות

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



מוציא לאוריו הצלחה בקורס - אוכלוסייה

כלל הסטודנטים לקורסים אקדמיים באוניברסיטה הפתוחה

הוצאו מהמודל:

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



קבוצות המשגים

שלב
בתואר

משתנים
דמוגרפים

הישגים
לימודיים

קצב
לימודים

משתנים
לימודיים
שונים

מדדי
הקורס

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



בניית האלגוריתם

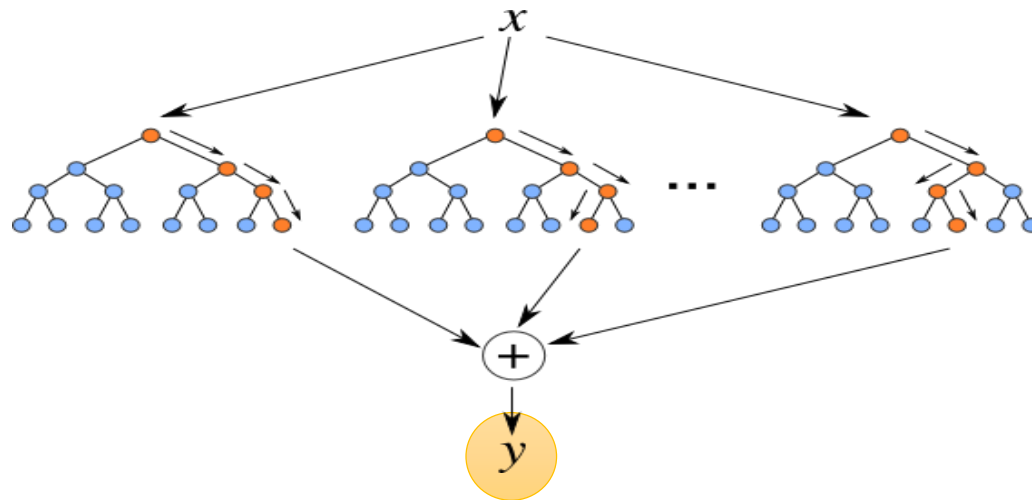
לצורך בחירת המודל הנכון בדקנו מודלים ממגוון משפחות:

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



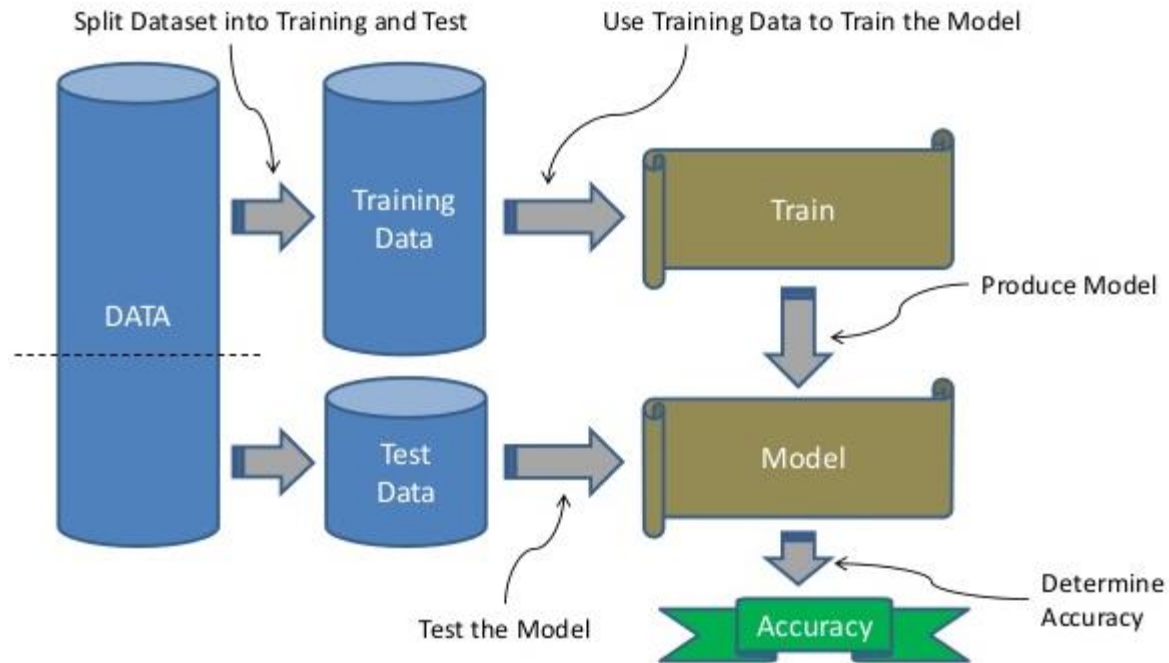
קצת על האובדן (Random Forest)

- 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.



גאוקר הנאונים והכרזג

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



איכות המוצא – 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 המוצר - ציוק

מציאות			מודל
עבר - מציאות	נכשל - מציאות		
10%	22%	נכשל - חיזוי	
65%	3%	עבר - חיזוי	



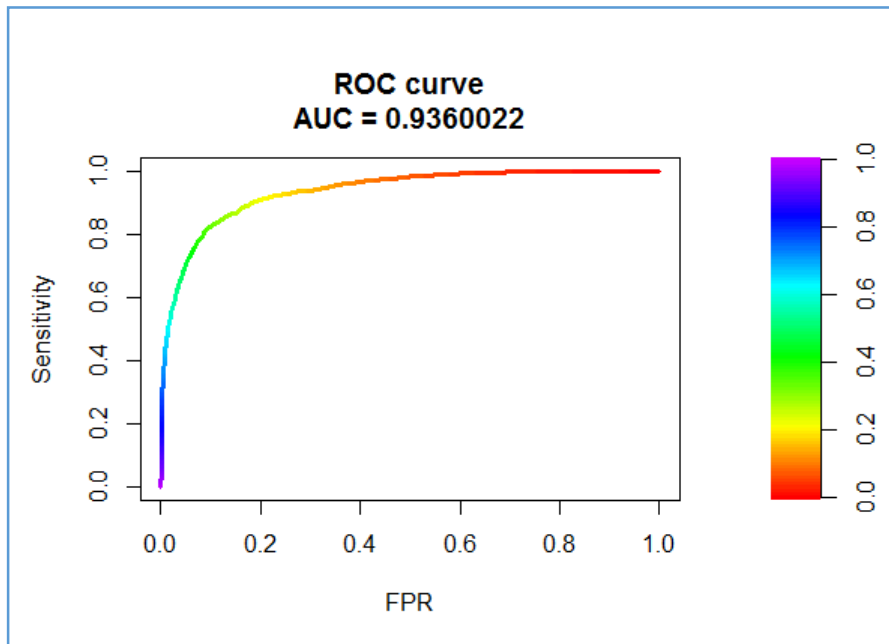
הדגמה:
עבור 50,000
הרשמות לקורסים

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

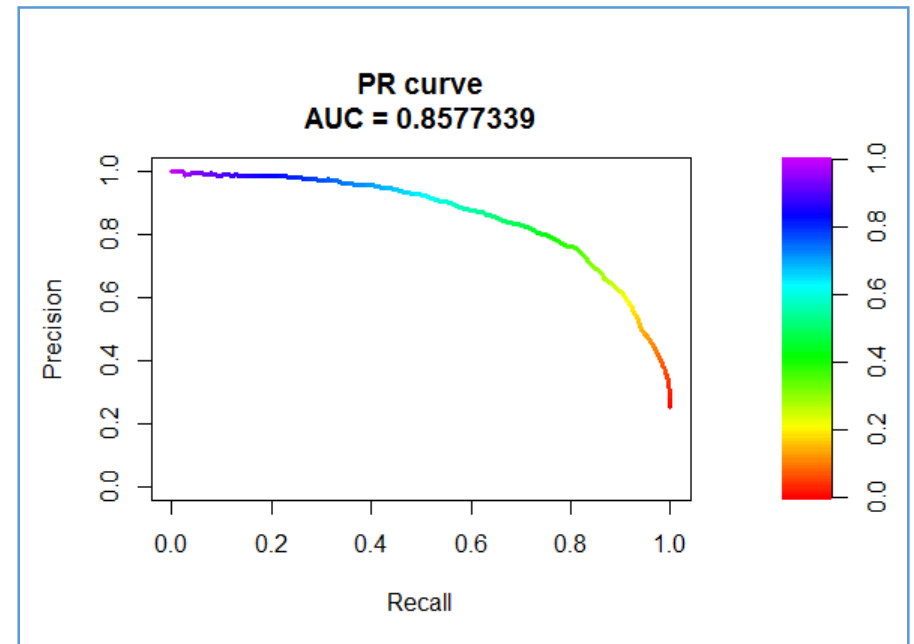


איכות המוצר - ציון

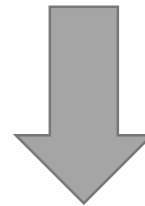
Sensitivity VS FPR



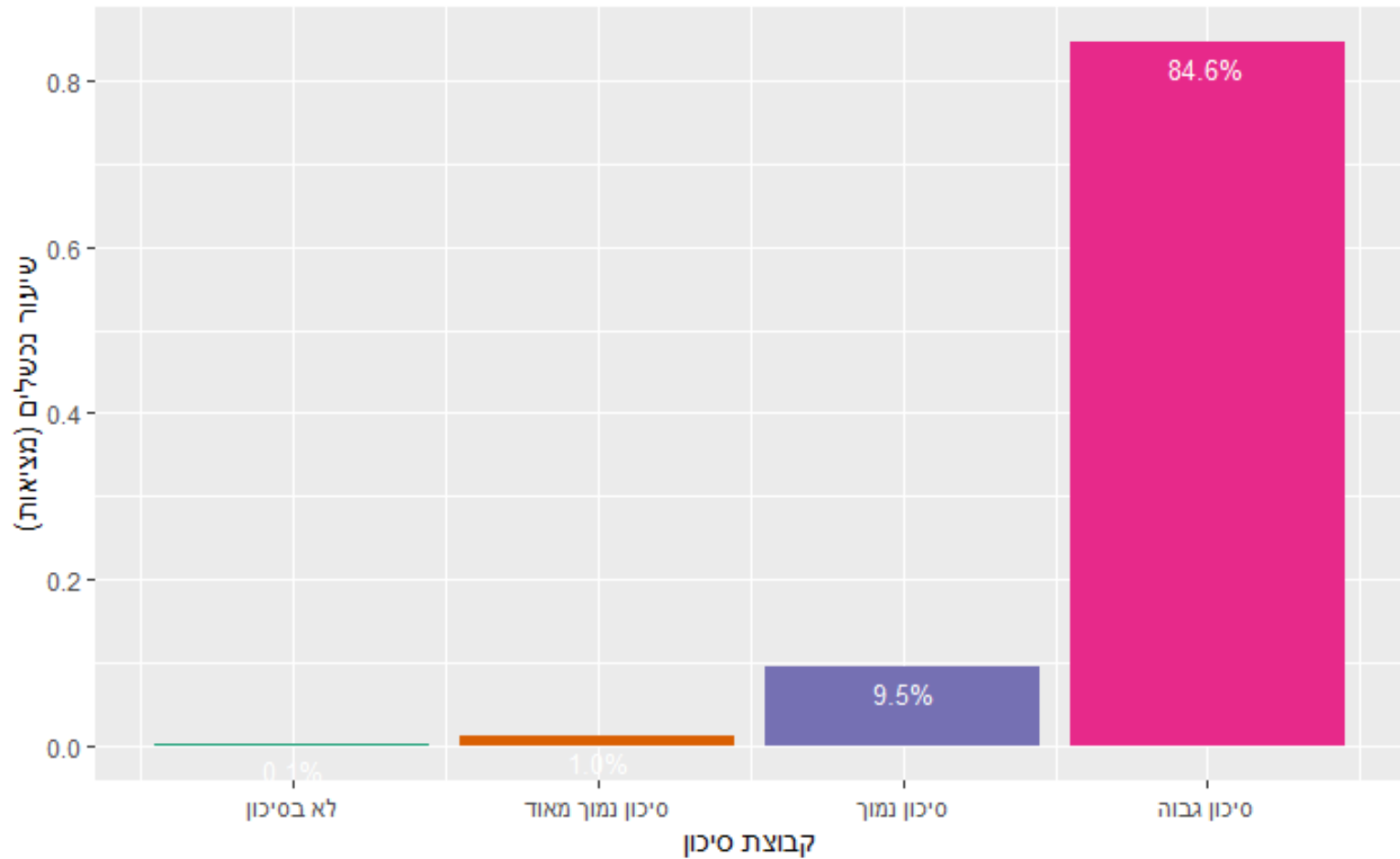
Precision VS Recall



גאוקה אמספר קבוצות סיכון



צבחה - גאוקה אמספר קבוצה סיכון



אסיוס – הרבה אנג'לים בפתח

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

מבין ההיבטים הסטטיסטיים, הבטחה היא ש:

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



תודה רבה!