

DATA SCIENCE PROJECT

יובל הרונגי



מי אנחנו

קרן להלוואות בערבות המדינה לעסקים קטנים ובינוניים
עסקים עם מחזור הכנסות עד 100 מ' ש

מחלקה חיתום - אחראית על החלטה האם לאשר את הלוואה
מחלקה סיכון - מטפלת בתיקי הלוואות שהגיעו לגבייה



מסלול השקעה



מסלול עסקים בהקמה



מסלול כללי



מסלול חקלאים



מסלול יצואנים



מסלול חון חוזר



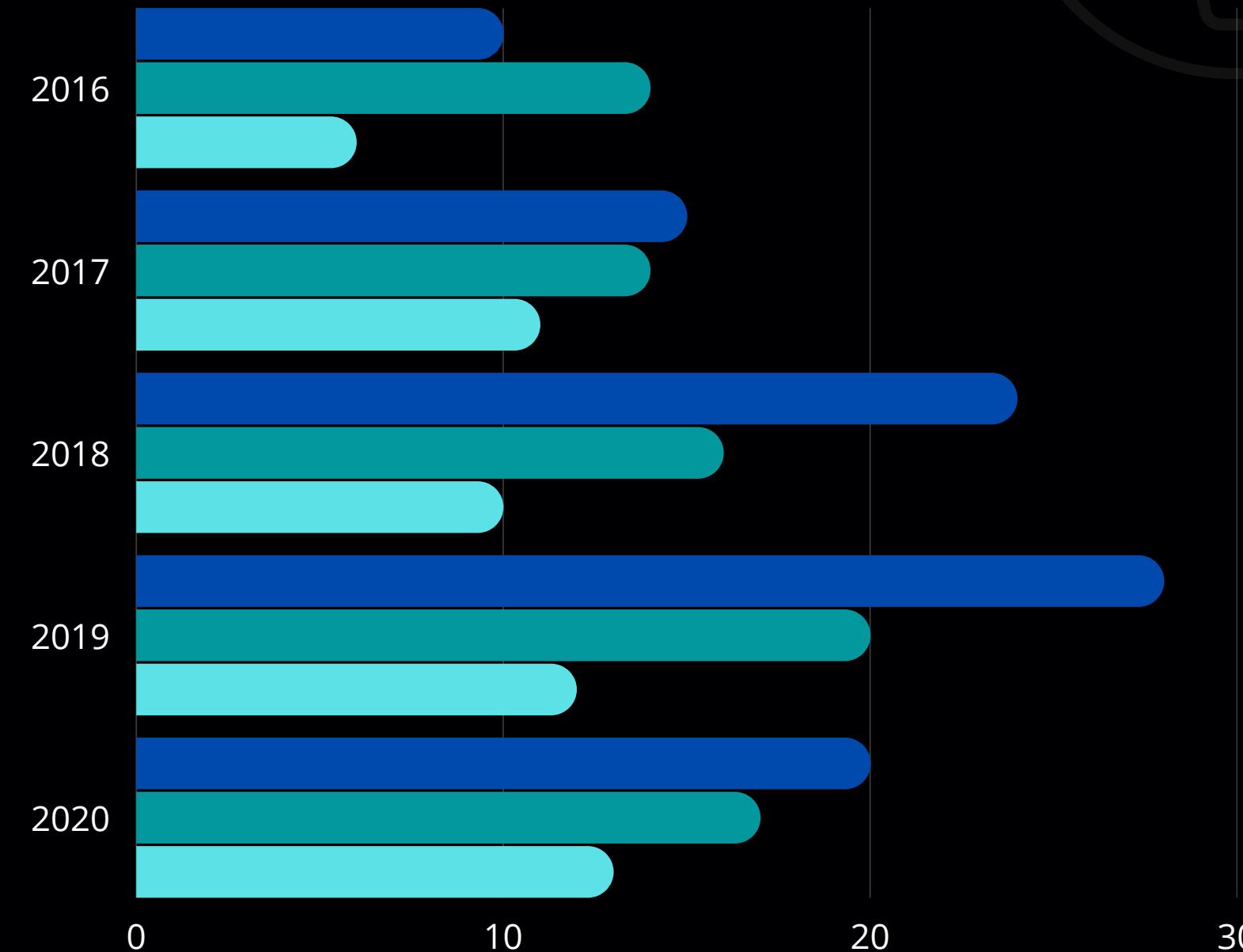


שאלה מהקרו

האם תיק הלוואה יגיע לגביה / לא יגיע לגביה



חקירת הנתונים A



מדד הנתונים

משתנים קטגוריים

בנק, אישור, מסלול, סקטור פעילות,
גודל עסק, וממד משפטית

משתנים מספריים

סכום מבוקש, סכום מאושר, סכום שהעמד, גובה הריבית,
שיעור הביטחונות, שנת הקמה, שנת העמדת הלואאה,
מכירות, שיעור רווח נקי ויחס אובליגו למכירות

	יחס אובליגו למכירות	שיעור רווח נקי	מכירות הלואאה	שנת העמדת הלואאה	גודל העסק	ממד משפטית	סקטור פעילות	תקופה	ויק	האם תיק גבוהה	שיעור הביטחונות	גובה הריבית	סכום ההמואר שהעמד	סכום מבקש	סכום מאושר	מסלול הצפון	אישור לאומי	בנק
0	0.27	-0.03	1385.0	2017.0	קטן	חברה	שירותים	ותיק	לא	3.50	25.00	1	135	135.0	150.0	משולב	הצפון	לאומי
1	Nan	Nan	Nan	2019.0	קטן	חברה	שירותים	עיר	לא	2.80	16.00	0	500	500.0	500.0	הצפון	הקמה	מזרחי טפחות
2	Nan	Nan	Nan	2019.0	קטן	שותפות	שירותים	עיר	לא	3.00	10.00	1	100	100.0	300.0	הצפון	הקמה	מזרחי טפחות
3	Nan	Nan	Nan	2018.0	קטן	שותפות	שירותים	בוגר	לא	2.90	13.75	1	250	400.0	500.0	הצפון	הקמה	לאומי
4	9.25	4.84	4668.0	2018.0	קטן	חברה	שירותים	בוגר	לא	2.75	25.00	1	200	200.0	300.0	משולב	המרכז	מזרחי טפחות
...
3249	Nan	Nan	Nan	2016.0	קטן	חברה	שירותים	בוגר	לא	3.00	16.00	0	500	500.0	500.0	הדרום	אוצר החיל	הקמה

3250 rows × 19 columns



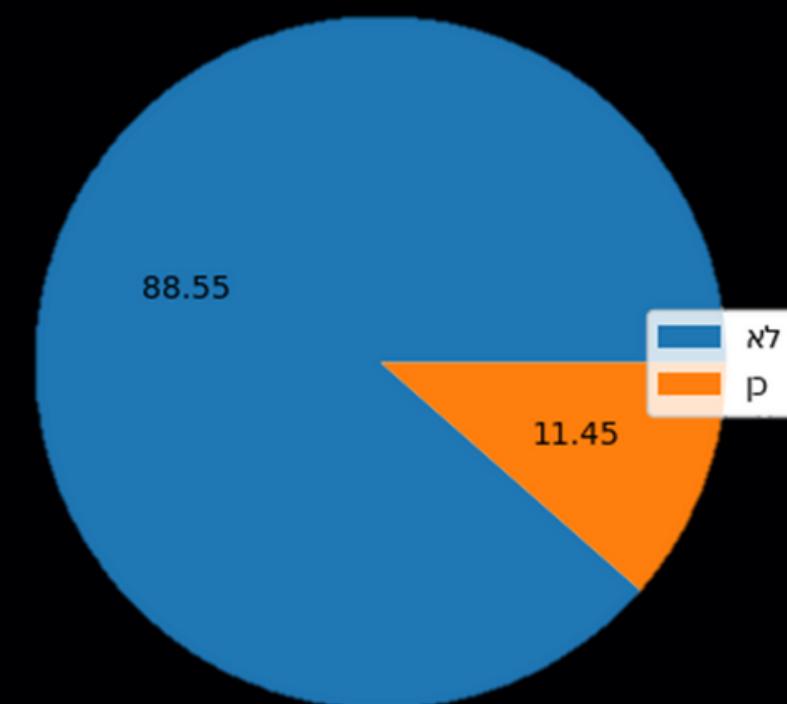
משתנה המטרה האם תיק גביה

88.55%

לא הגיע לגביה

11.45%

הגיע לגביה





בחירה המשתנים הרלוונטיים למודל

בנק ואזר

משתני דפוס של אופי העסק מבחינה סוציאו-קונומית

מסלול, סקטור פעילות, גודל העסק ומעמד משפטי

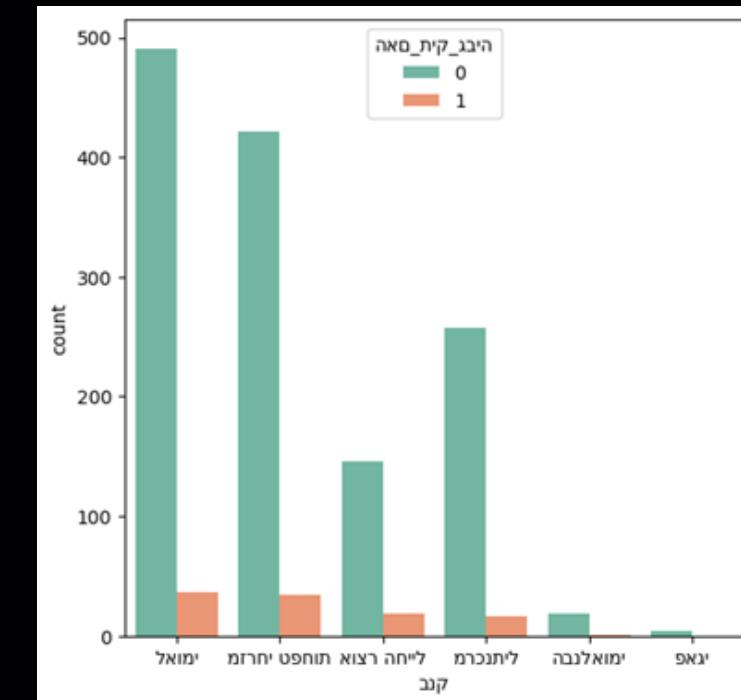
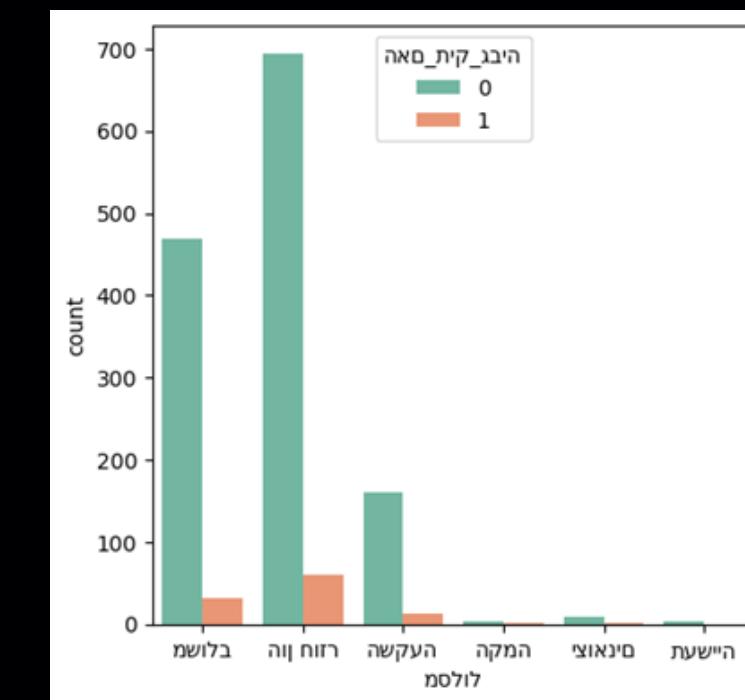
משתנים שיכולים להשפיע על התמודדות העסק עם סיכון צד נגדי, ועל החוסן וההצלחה שלו

מעמד_משפטי	
אגודה	0.000000
חברה	0.075968
עובד סורשה	0.038462
עמותה	0.000000
שותפות	0.166667

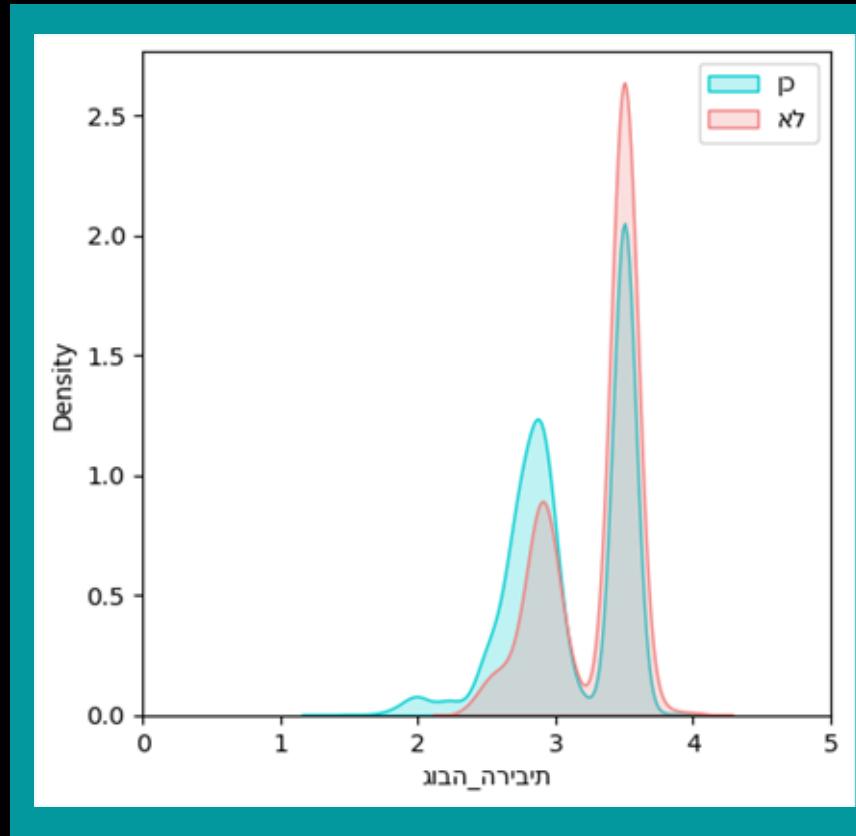
איזור	
הדרום	0.064378
המרכז	0.098870
הצפון	0.064646
הדרום	0.071942
יק"ש	0.000000

מסלול

בנק

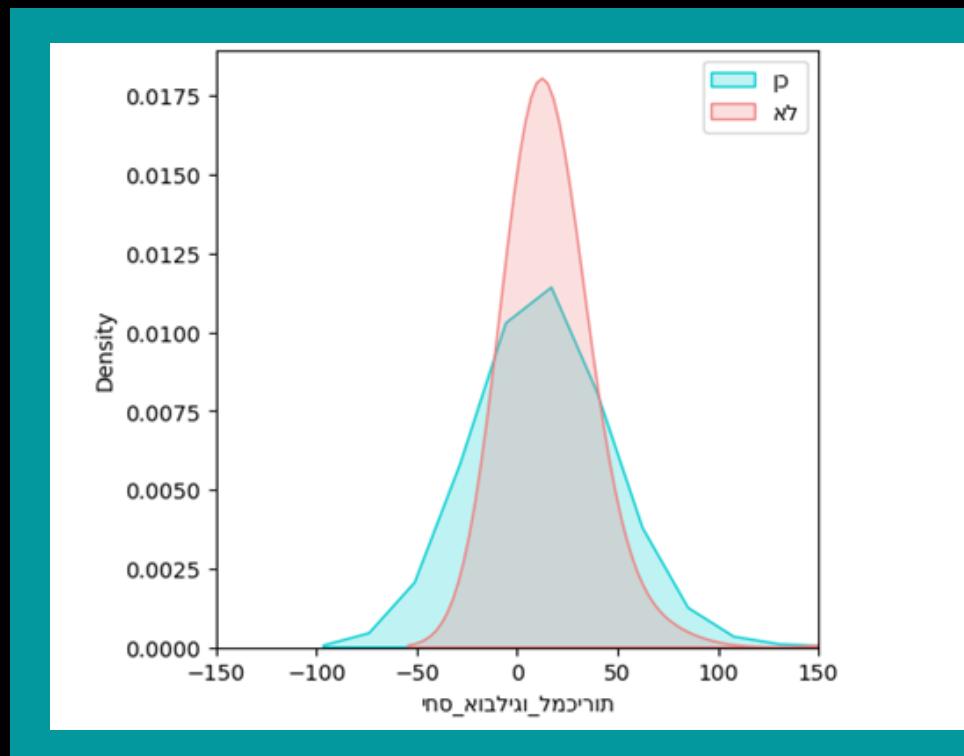


בחירה המשתנים הרלוונטיים למודל



יחס אובליגו למכירות

גובה הריבית
מתאים גבוהה עם מצב האינפלציה בשוק, لكن אינדיקציה של ריבית הלוואה יכולה להעיד על רמת הקושי בהישרות העסק



שיעור רווח נקי ויחס אובליגו למכירות
משתנים המותאמים ורלוונטיים להסברת יציבות ואיתנות העסק, המשלבים בתוכם את משתנה המכירות

שיעור בטעונות

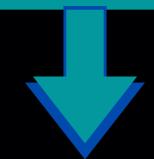
שיעור בטעונות נמור לעול להיעד על מעורבות נמוכה יותר של בעל העסק בהחזר הלוואה, דבר המunik לו אפשרות לקחת סיכון גבוהים יותר בהתקנותו

סכום שהועמד

סכום הלוואה זה הינו הסכום הסופי שהעסקלקח

הטאמת הנתונים PREPROCESSING

סינון נתונים לא
רלוונטיים בקובץ אקסל



סינון מתקדם
בקובץ פירטו



בנייה משתנים
חדשים



יצירת משתנים חדשים

סכום מבוקש גובה מהماושר

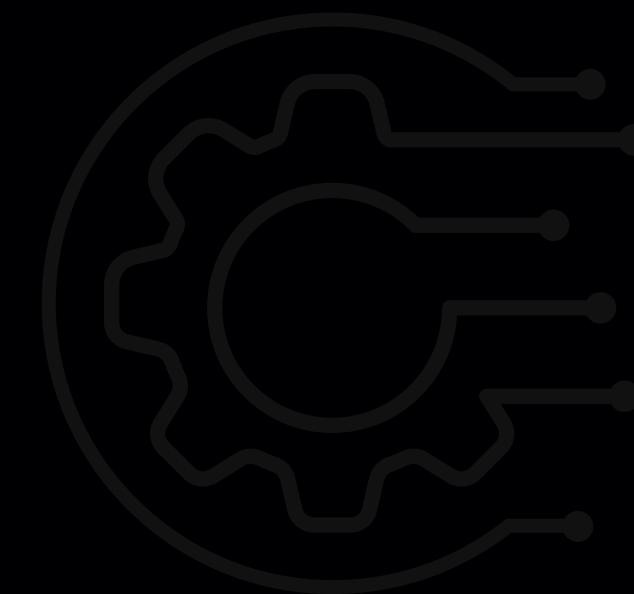
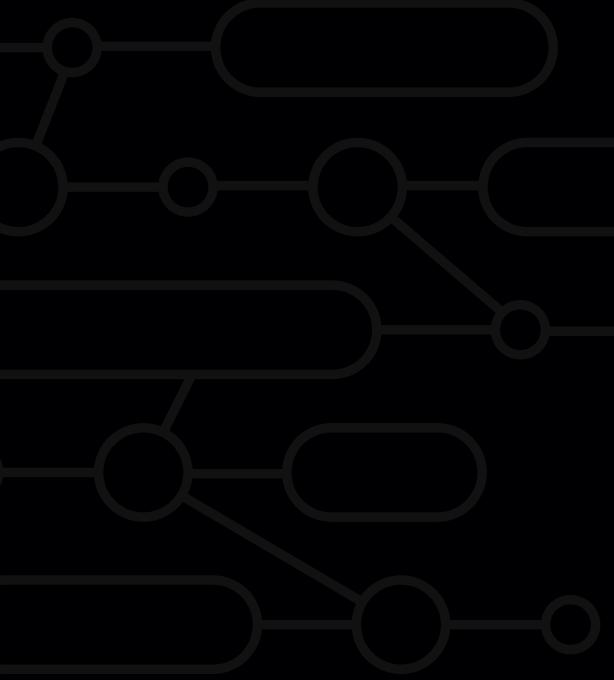
במידה והסכום המאושר גובה
מהסכום המבוקש יקרא "1",
במידה ושווה יקרא "0"

ותק

עסק בעל פחות מ-5 שנים פעילות יקרא "עיר",
עסק בין 5-10 שנים פעילות יקרא "בוגר", עסק
בעל מעל 10 שנים פעילות יקרא "ותיק"

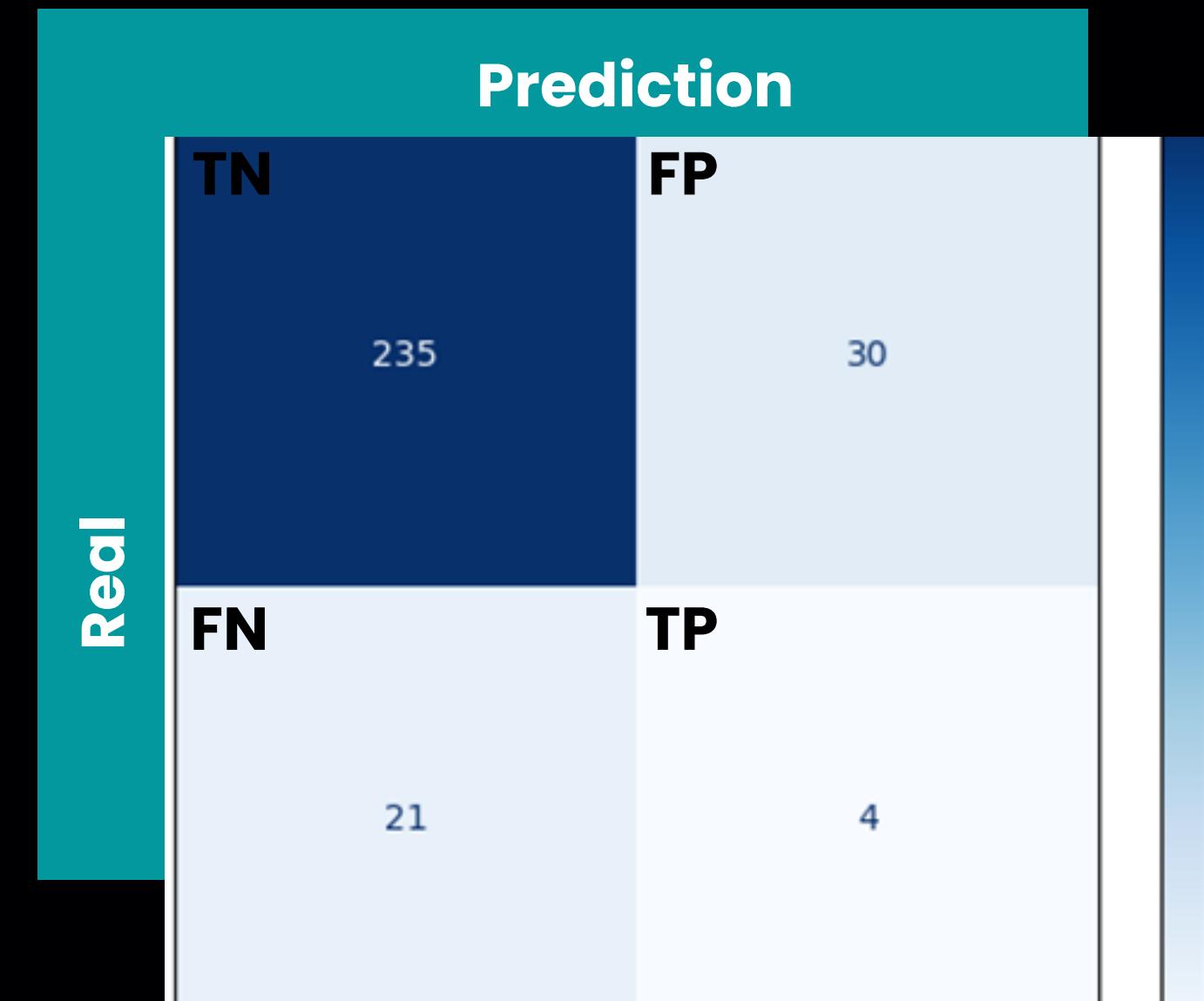
האם_תיק_גביה
סכום_մերժական_գիւղական_մասնաշխատ
0 0.047619
1 0.084713

האם_תיק_גביה
ותק
בוגר 0.092593
ותיק 0.068345
עיר 0.090909



הערכת המודל באמצעות מדדים

הבעיה שנרצה לפתור היא בעיית קלאסיפיקציה,
כלומר, ייצורת תחזית של משתנה בינהי
 $\text{לא} = 0, \text{כן} = 1$



מדדִי הערכת הביצועים

01

תחית - Precision

02

עולם אמיתי - Recall

03

F1

04

דיוק - Accuracy

05

בינארי 0/1 - Roc AUC

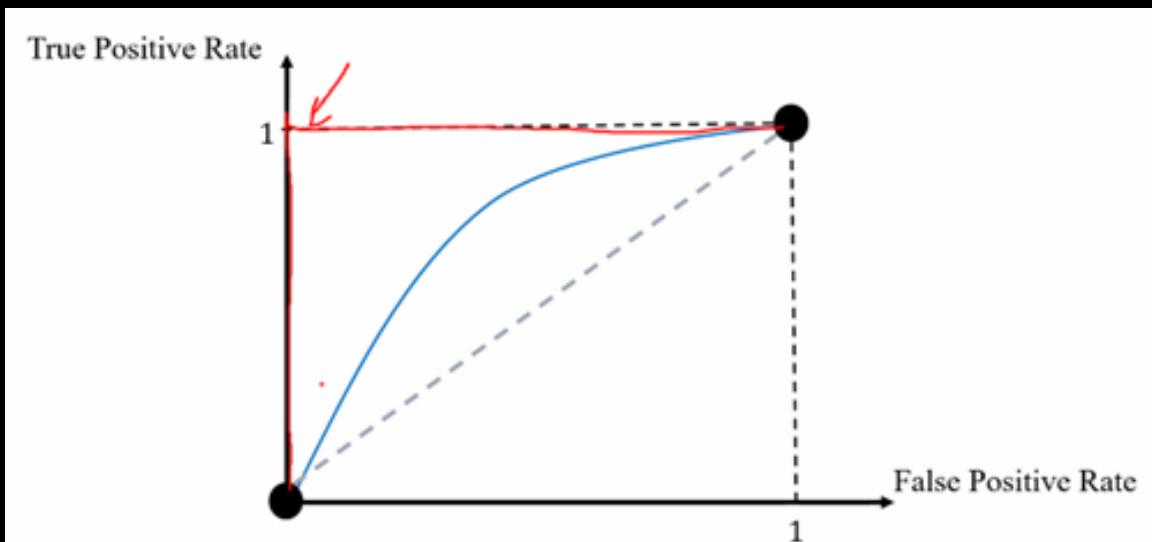
```
recall: 0.1910384068278805
precision: 0.18069085347850047
F1: 0.18572061154643618
auc: 0.5434556404469755
accuracy: 0.8061538461538461
```

בחירה מדד

ROC AUC

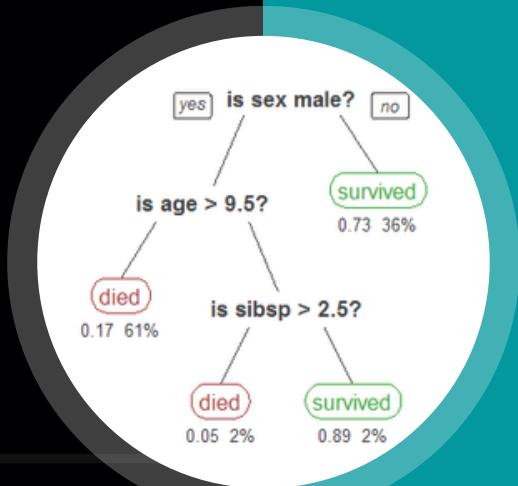


- ✓ מטפל במשתנים לא מאוזנים
- ✓ מדד זה מתאים לשימוש עם משתנים בינאריים, כלומר תוצאות של 0 או 1
- ✗ מדד **Accuracy** אינם מתאים למודל מהסיבה שהוא בודק אחוז דיווק, שכן כאשר המשתנים אינם מאוזנים, המחשב לומד את המקרים הנפוצים יותר של משתנה המטרה, וכתוצאה לכך מציג תוצאות מוטעות
- ✗ המדדים **Precision** ו/**Recall** כאשר הם נבדדים בנפרד הם אינם אפקטיביים, لكن נستخدم במדד **F1** המשלב בין שנייהם. למרות זאת, מדד זה הביא אותנו לתוצאות נוחות מאוד





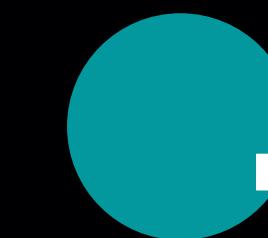
סוגי מודלים



Decision Tree



Logistic Regression



בחירה מודל

Logistic Regression



- ✓ מודל זה נוקשה יותר מאשר המודלים, ונדרשת בו בחירה של רמת סף - האם אני שמרן \ רגש וועזרת בההתאמת התוצאות למקורה
- ✓ השתמשתי ברמת סף יחסית נמוכה כי אני מאמין שעדייף שיהיו יותר מייד התראות שווא, ככלمر מצב שבו נחשב שעסק יגיע לגביה, אך הוא לא הגיע, מאשר מצב שבו נחשב שעסק לא יגיע לגביה, אך בסוף הוא כן יגיע.

Decision Tree

```
recall: 0.1910384068278805
precision: 0.18069085347850047
F1: 0.18572061154643618
auc: 0.5434556404469755
accuracy: 0.8061538461538461
```

Logistic Regression

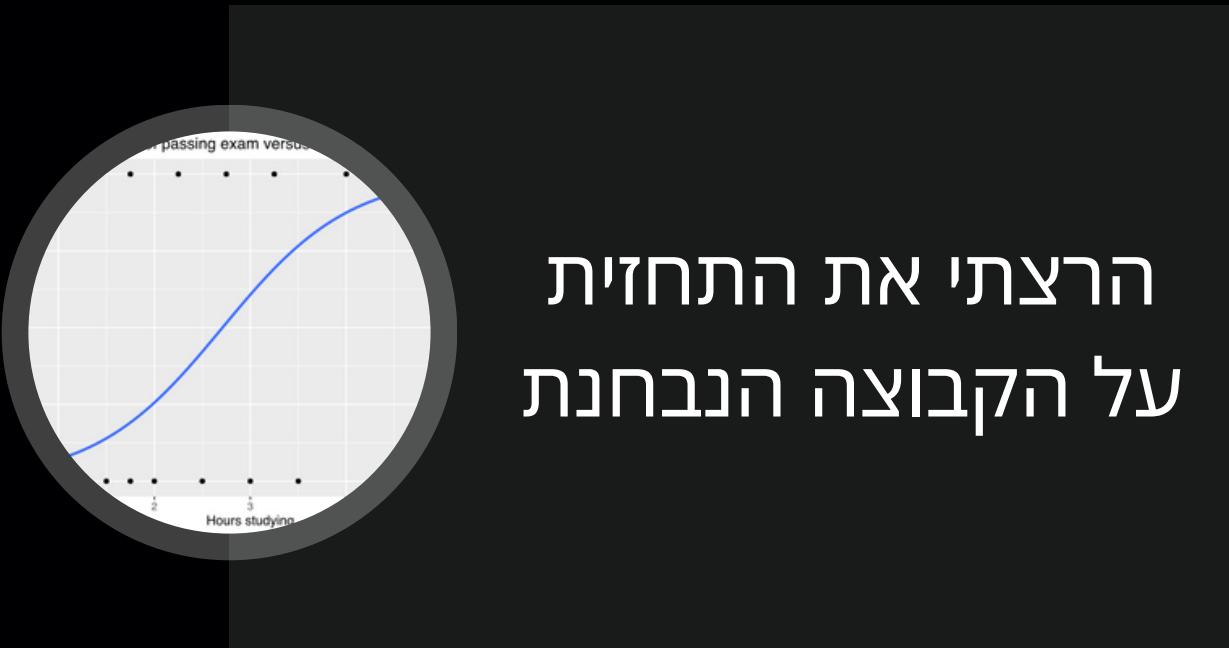
```
recall: 0.005405405405405406
precision: 0.2
F1: 0.010526315789473684
auc: 0.7033261266492974
accuracy: 0.8855384615384615
```

שיטות להערכת טיב התחזית

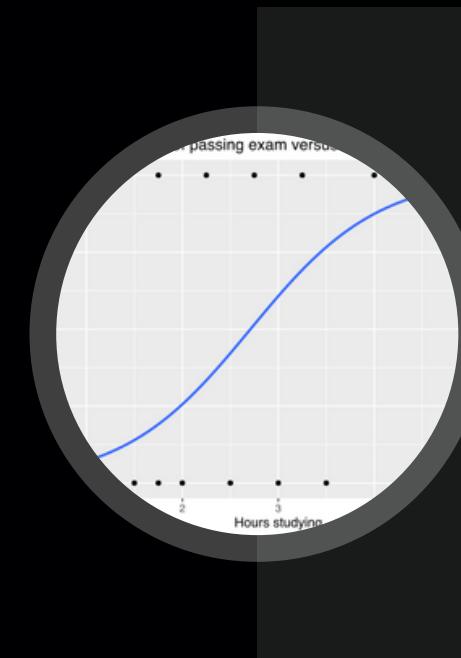
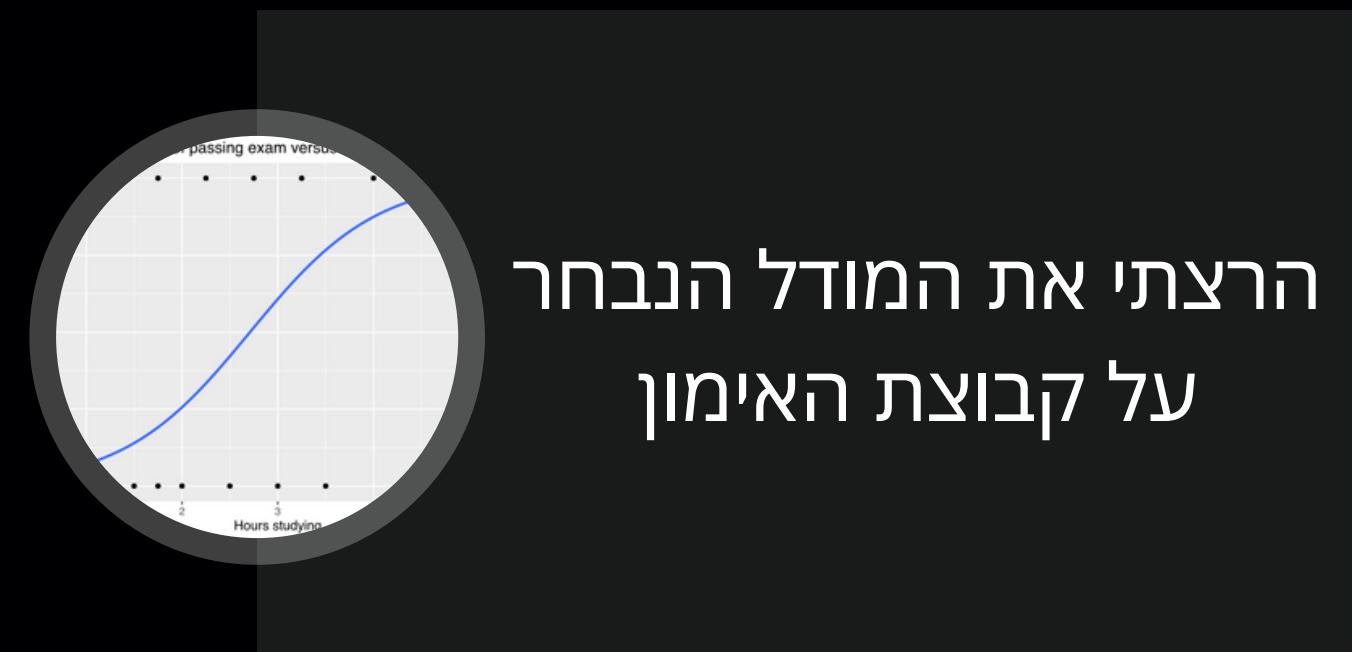
התמודדות עם OverFitting



הרצתי את התחזית
על הקבוצה הנבחנת



הרצתי את המודל הנבחר
על קבוצת האימון



פיצלי את המודל
לקבוצת אימון
וקבוצת מבחן



תובנות ומסקנות

למדינה

ኖכל למצוות את הוצאותיה בהתרבותות והגנה על עסקים שהגיעו לגביה, ותשולם הלוואה במקומות לבנים

לבנים

ኖכל לחסוך לבנים רדיפה מתמשכת תוך הוצאה משאבים אדרה אחר עסקים שmagיעים לגביה

לעסקים

ኖכל לחסוך את רישום ה"כטם הבנקאי" בדו"ח נתוני האשראי בכר שמראש לא נאשר להם הלוואה

תודה על ההקשבה



Ln: 123 Col: 2