



# ניתוח מעורבות קהילתית בספריות ציבוריות בקונטיקט

נושאים מתקדמים בלמידת מכונה

ברק וירצברגר 316597541 | רותם בן עטר 318299914

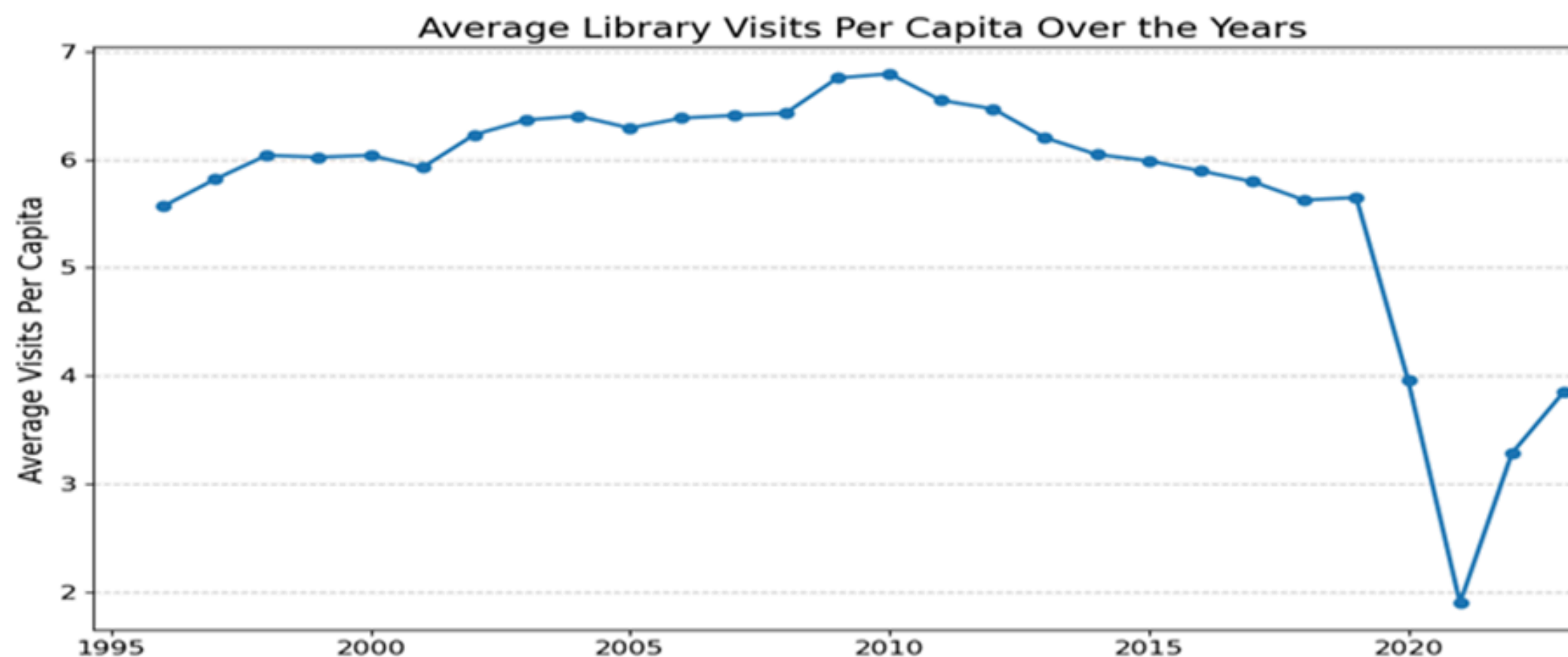
מרצה: חן חגג'

# הבעיה

לאחר הקורונה, נרשמה עלייה מסוימת במספר הביקורים, אך לא במידה שמחזירה את הספריות לרמת הרלוונטיות והפעילות שהיו להן בעבר.

בשנת 2020, מגפת הקורונה האיצה את הירידה בביקורים, אך המגמה השלילית החלה עוד לפני המשבר העולמי.

ניתוח נתונים במדינת קונטיקט מצביע על ירידה עקבית במספר הביקורים בספריות לנפש החל משנת 2010.

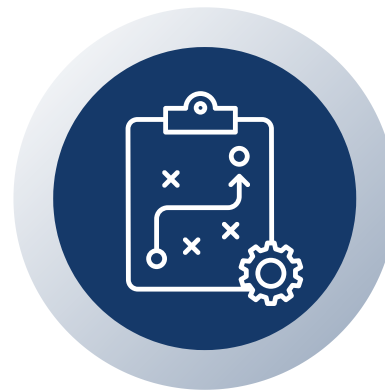


הירידה בשימוש בספריות עלולה להוביל לצמצום התרומה החברתית-קהילתית שלהן ולפגיעה במטרותיהן כמרכזים חינוכיים ותרבותיים.

# מטרה



לסווג ספריות לפי מדד מעורבות חברתית, במטרה לזהות ספריות מצליחות מבחינת חיבור לקהילה לעומת ספריות עם מעורבות נמוכה.



לגבש אסטרטגיות ממוקדות עבור ספריות בעלות מעורבות נמוכה, ובכך לשפר את יכולתן למשוך קהל ולהפוך לרלוונטיות יותר עבור הקהילה.



בעזרת שילוב של מודלים מפוקחים ולא מפוקחים, נזהה מאפיינים ותבניות ייחודיות שמסייעות להבנת הגורמים למעורבות קהילתית גבוהה.

# סקירת הדאטה

**29**

פיצ'רים

**data.ct.gov**

מקור המידע

**1996–2023**

שנות המדידה

**208**

ספריות

נתונים  
כלכליים

נתונים  
חברתיים

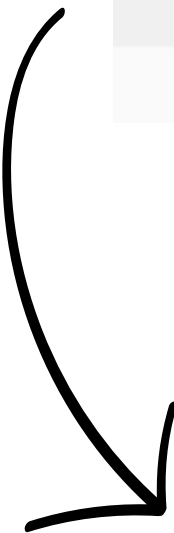
נתונים  
דמוגרפיים

# למידה לא מפוקחת

1 קיבוץ הדאטה לפי כל ספריה כשורה מייצגת לכל השנים

2 יצירת פיצ'רים סטטיסטיים לשימור המידע

	Fiscal Year	Library	County	Principal Public?	AENGLC Rank	Population of Service Area	Total Library Visits	Library Visits Per Capita Served	Total Registered Borrowers	Percent of Residents with Library Cards	...
0	1996	Andover	Tolland	P	109.0	2815.0	9000.000000	5.3	1313.0	0.47	...
1	1997	Andover	Tolland	P	109.0	2782.0	9193.111111	5.4	1676.0	0.60	...



	Library	Library Visits Per Capita Served_mean	Library Visits Per Capita Served_min	Library Visits Per Capita Served_max	Library Visits Per Capita Served_std	Library Visits Per Capita Served_median	Percent of Residents with Library Cards_mean	Percent of Residents with Library Cards_min	Percent of Residents with Library Cards_max	Percent of Residents with Library Cards_std	...
0	Andover	3.050000	1.2	5.400000	0.882127	3.000000	0.757857	0.44	1.000000	0.186813	...
1	Ashford-Babcock	2.807143	0.8	3.600000	0.804123	3.100000	0.510357	0.18	0.840000	0.206908	...

# למידה לא מפוקחת

הפחתת מימדים בעזרת PCA

3

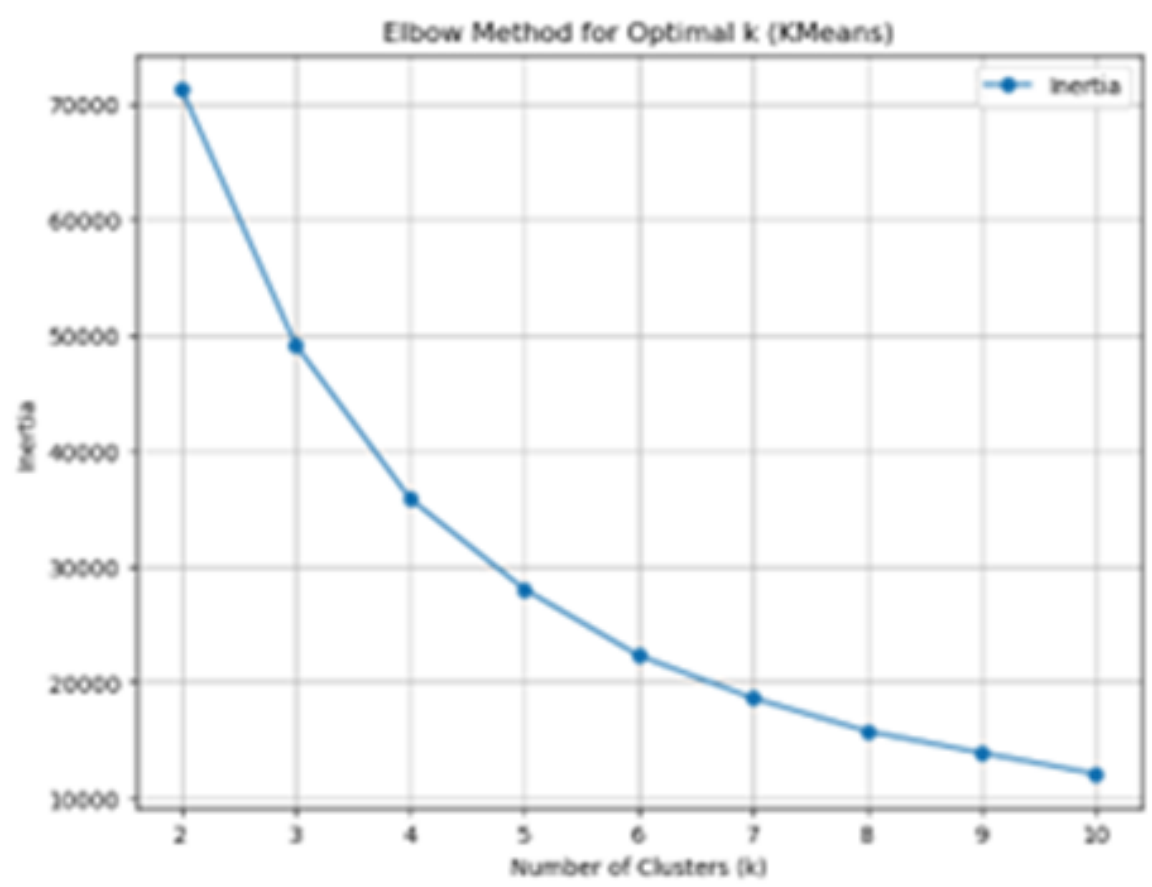
K-Means, Gaussian Mixture ,Agglomerative Clustering - בחירת מודלים ללמידה

4

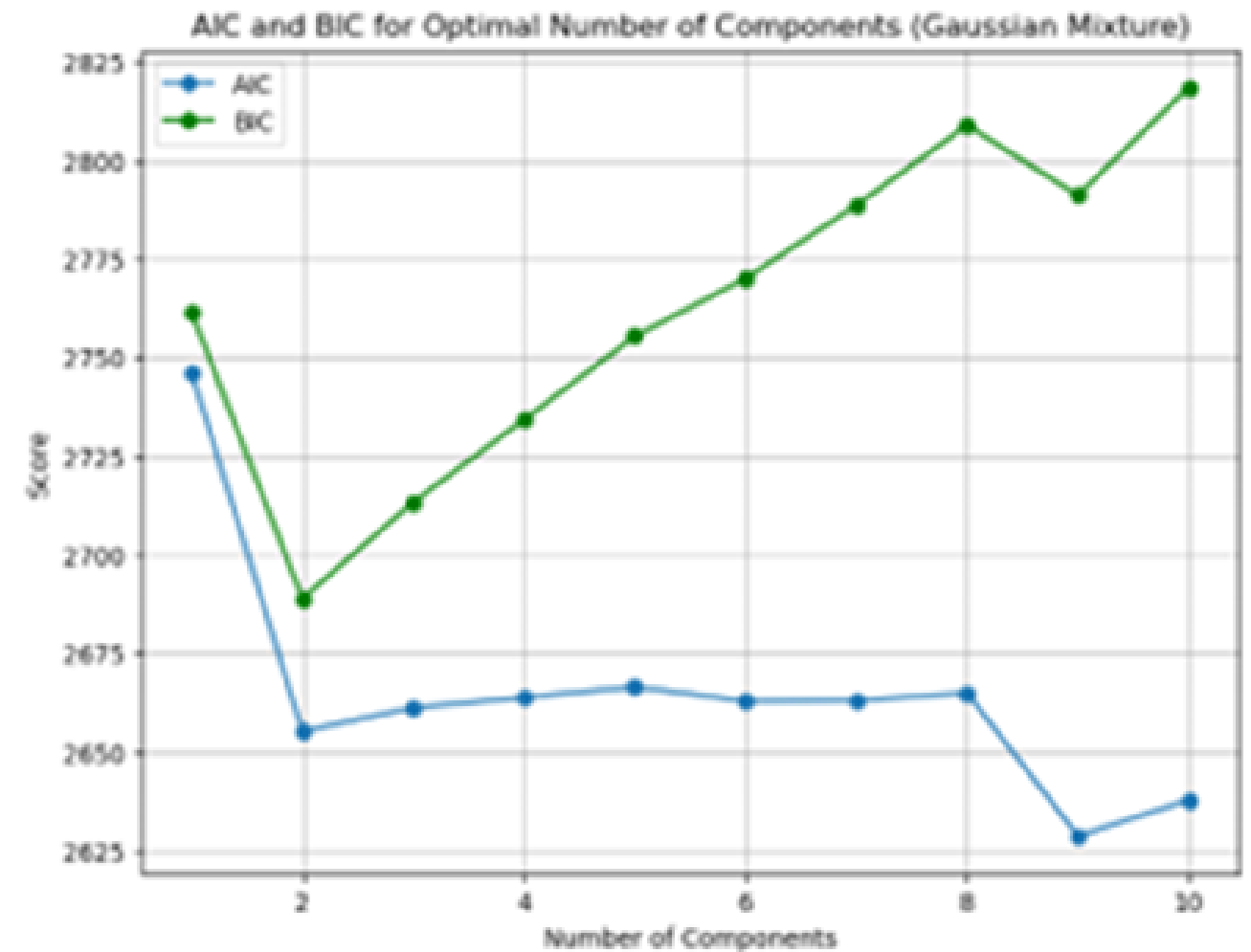
בחינת פרמטרים אופטימליים עבור כל מודל

5

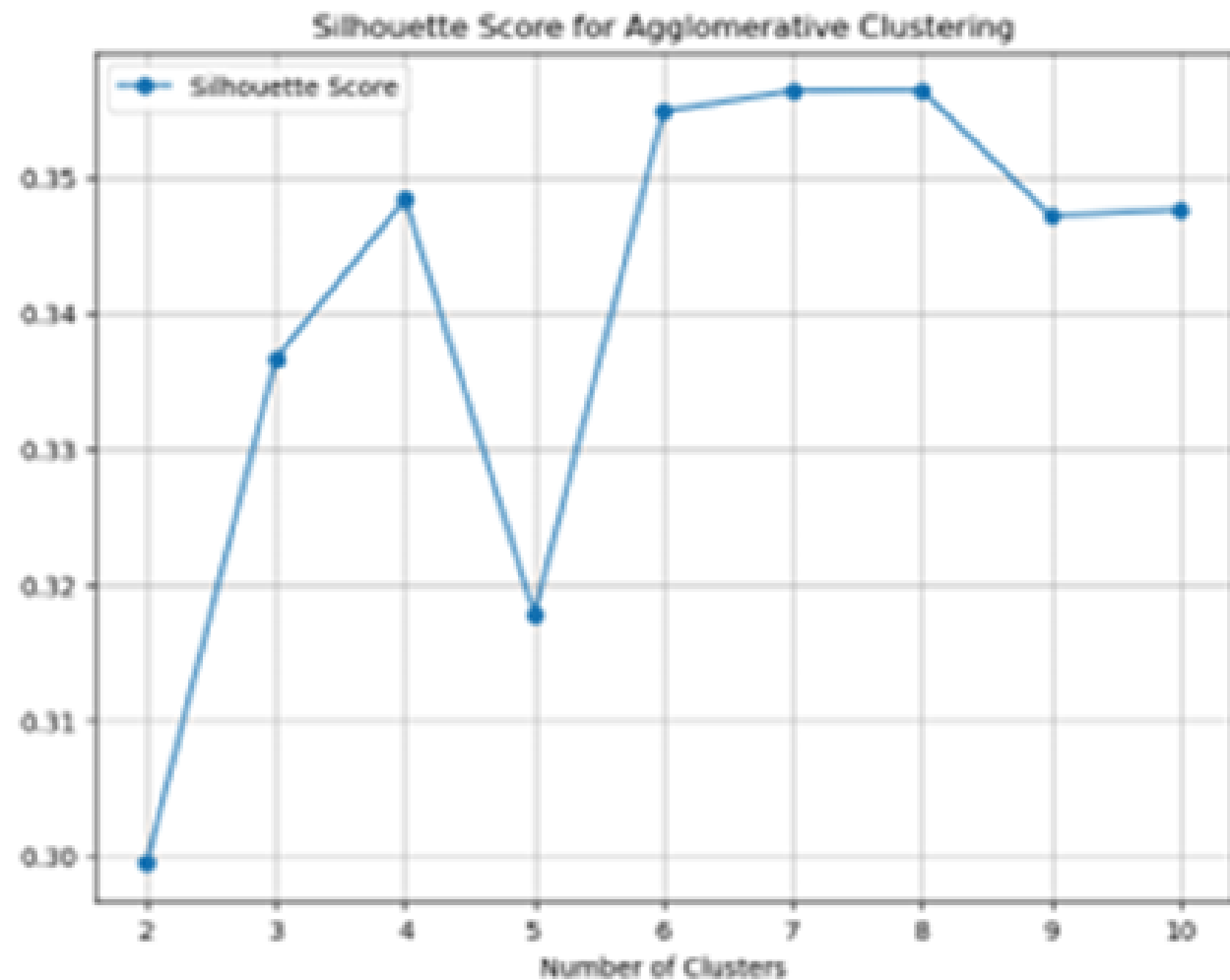
K-Means



Gaussian Mixture



Agglomerative Clustering

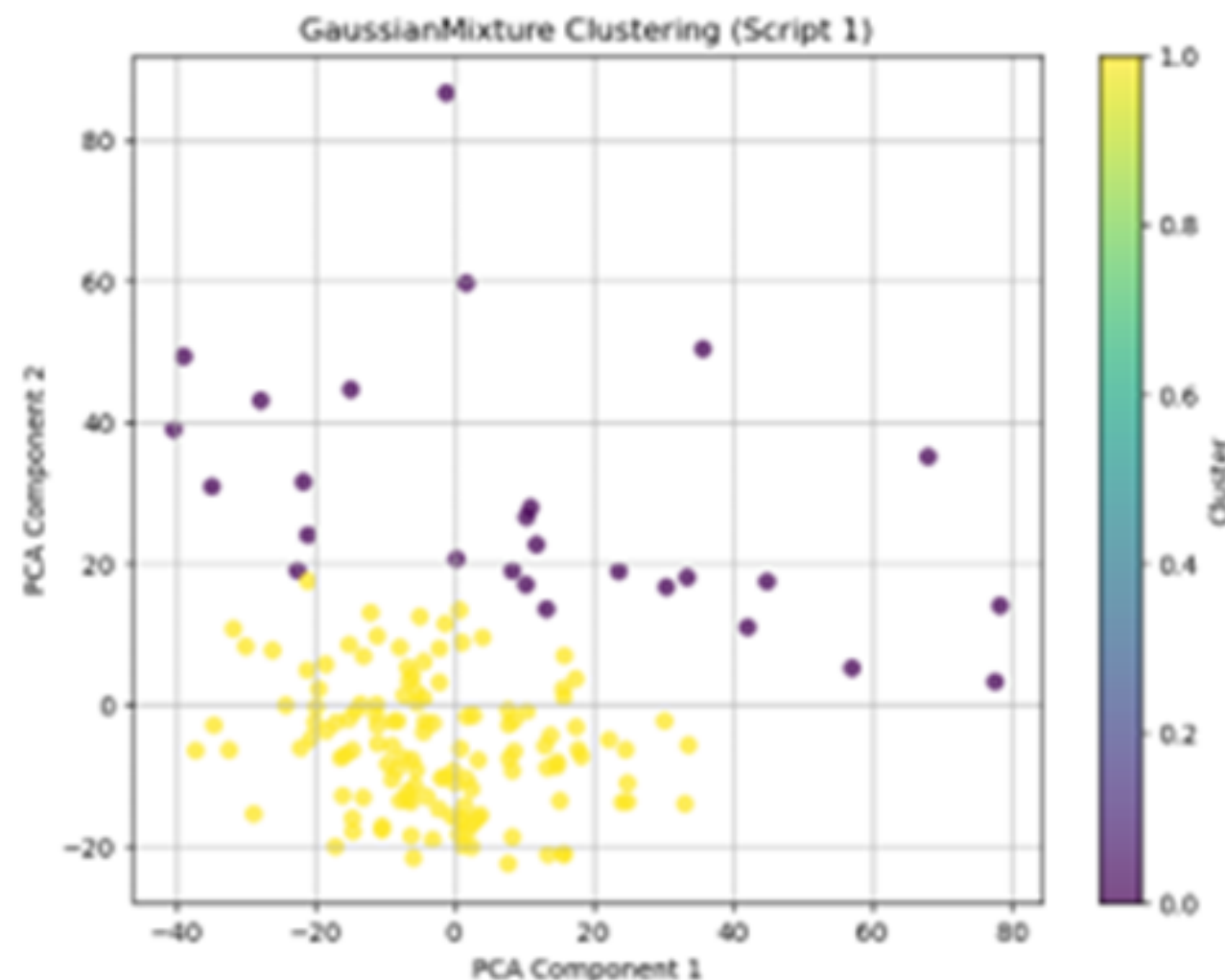


# למידה לא מפוקחת

	Silhouette Score
KMeans	0.430529
GaussianMixture	0.495432
AgglomerativeClustering	0.336652

מציאת החלוקה המועדפת על פי מדד נבחר

6



החלוקה ל-2 קלסטרים באמצעות Gaussian Mixture הראתה שספריות עם מעורבות חברתית גבוהה מתאפיינות במשאבים ותשתיות טובים יותר ונמצאות באזורים סוציו-אקונומיים גבוהים, בעוד הספריות בקלסטר השני מציגות מעורבות ומשאבים נמוכים יותר.

7



# למידה מפוקחת

1 יצירת עמודת המטרה

1

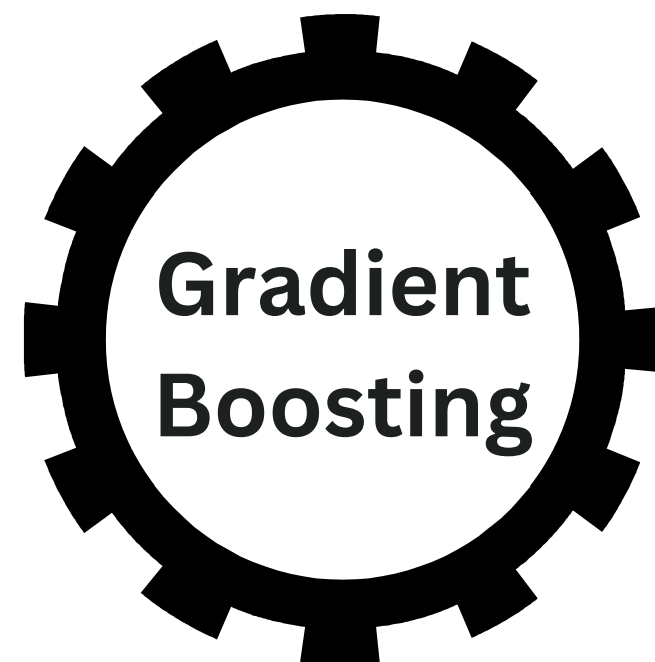
$$\text{Performance Score} = (\text{Library Visits Per Capita Served} \times 0.7) + (\text{Total Registered Borrowers per Capita} \times 0.3)$$

2 בניית פיצ'רים נוספים

2

3 בחירת מודלים

3

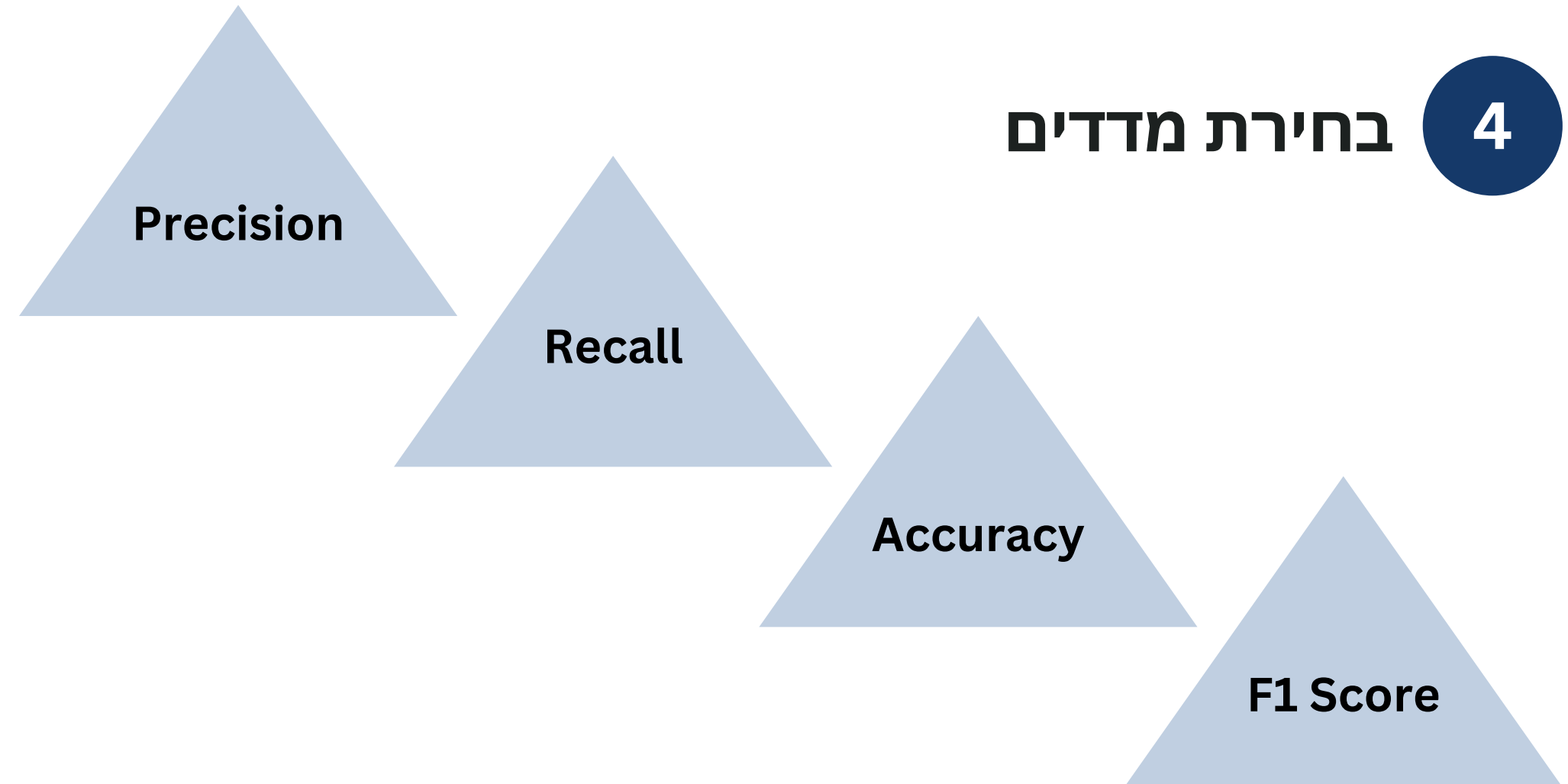




# למידה מפוקחת

## 4 בחירת מדדים

$$\text{Precision} = \frac{TP}{TP + FP} \quad \text{Recall} = \frac{TP}{TP + FN}$$
$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$
$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$



## 5 ביצוע ניסויים וכיוונון פרמטרים למודלים

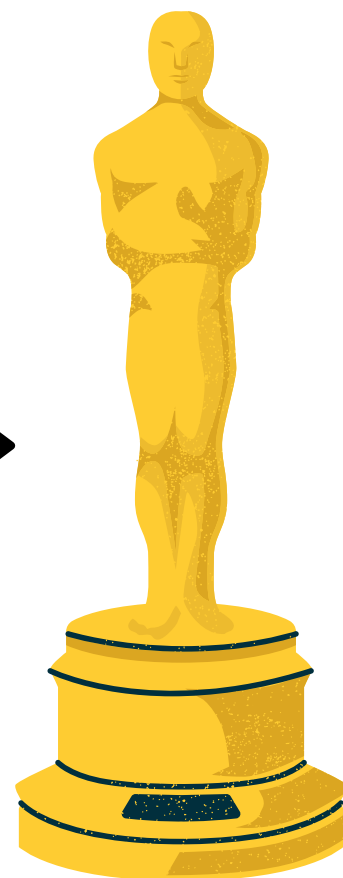
# למידה מפוקחת

השוואה בין תוצאות

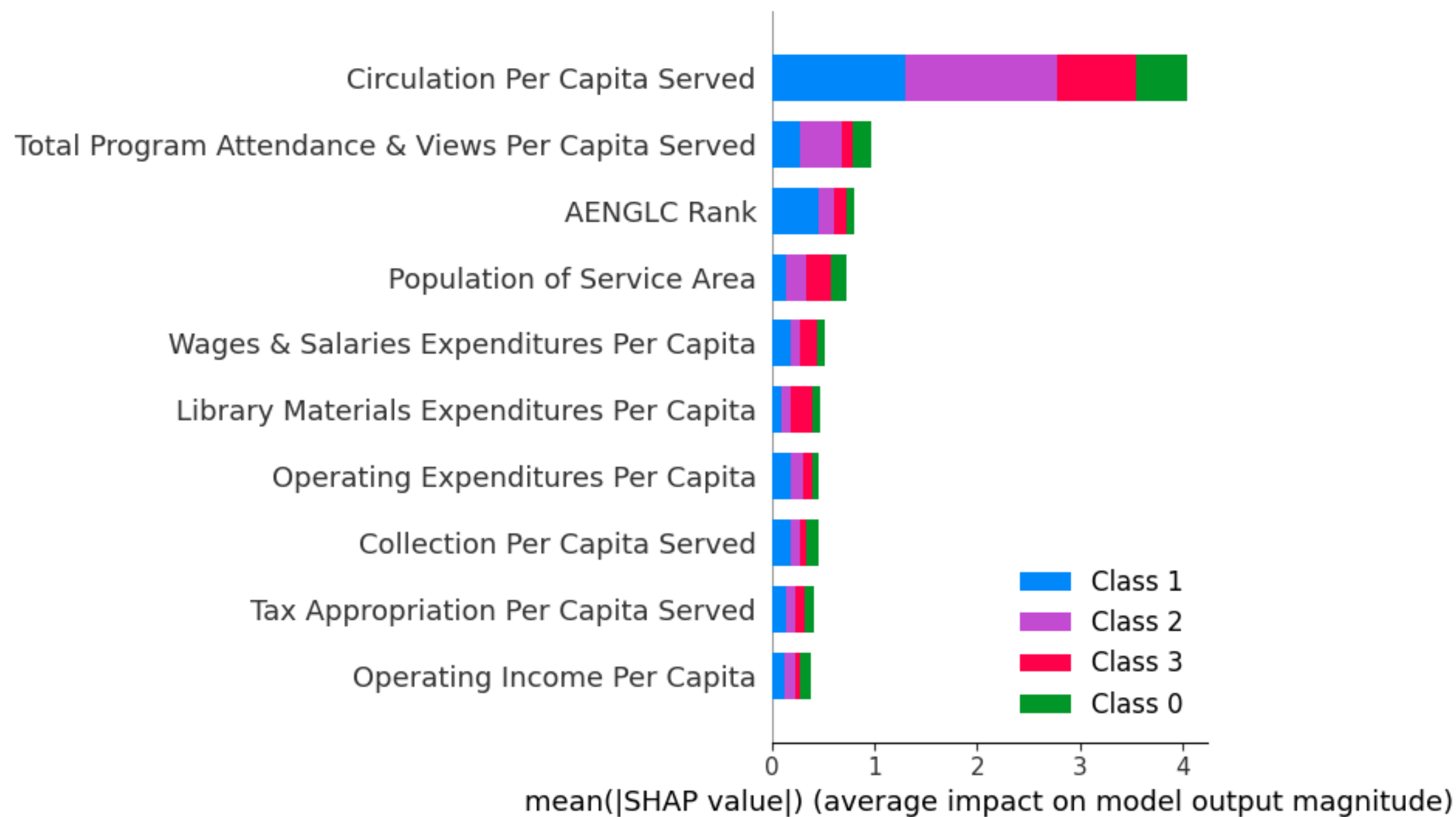
6

Sorted Results by Test F1 Score:

	Model	Mean CV Accuracy	Test Accuracy	Test F1 Score
3	Gradient Boosting	0.753697	0.750000	0.751272
2	XGBoost	0.736387	0.748884	0.750387
4	KNN	0.616029	0.631696	0.632773
1	Random Forest	0.640322	0.620536	0.629219
0	Decision Tree	0.602901	0.584821	0.598673

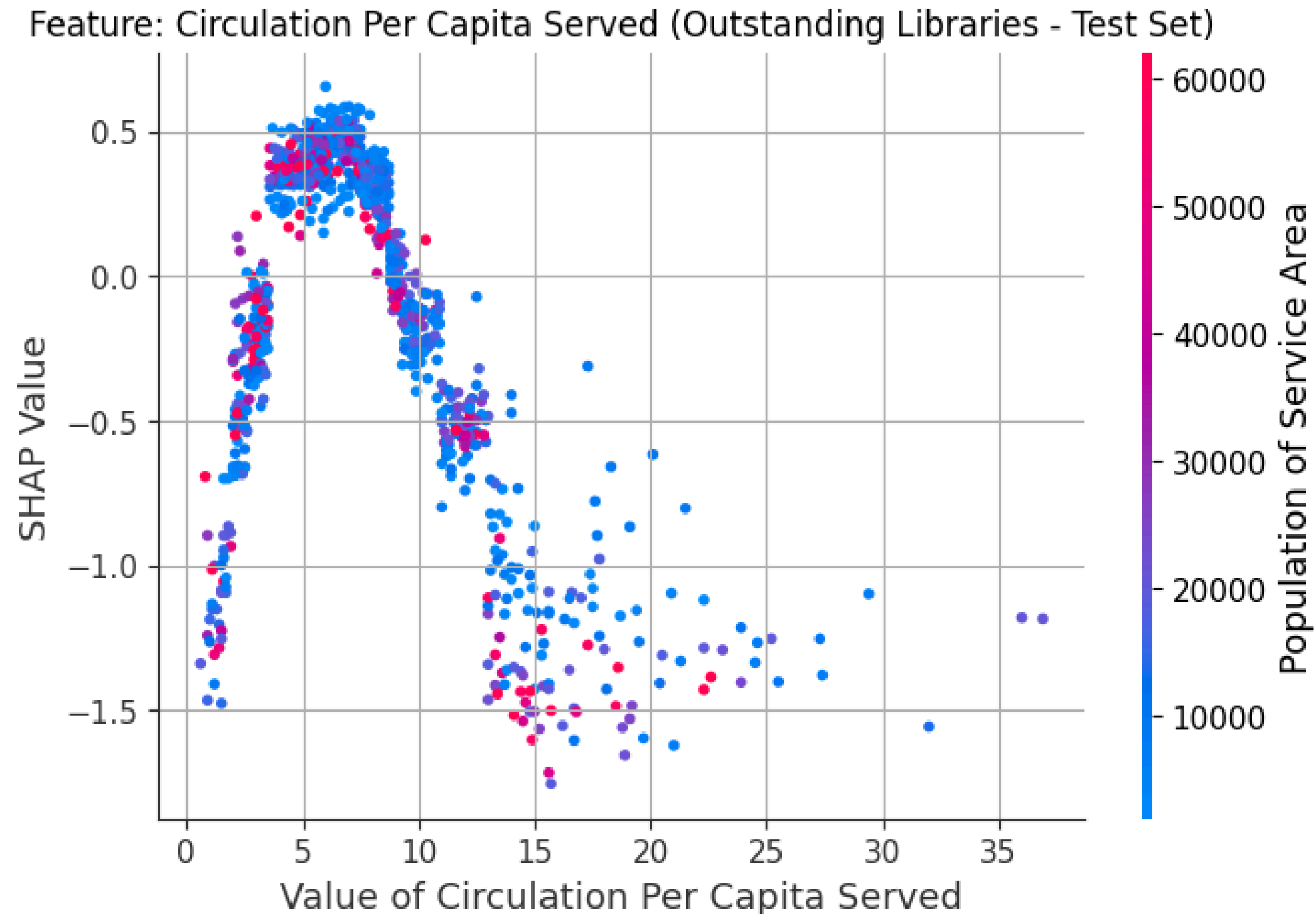


# ניתוח מעמיק



# ניתוח מעמיק

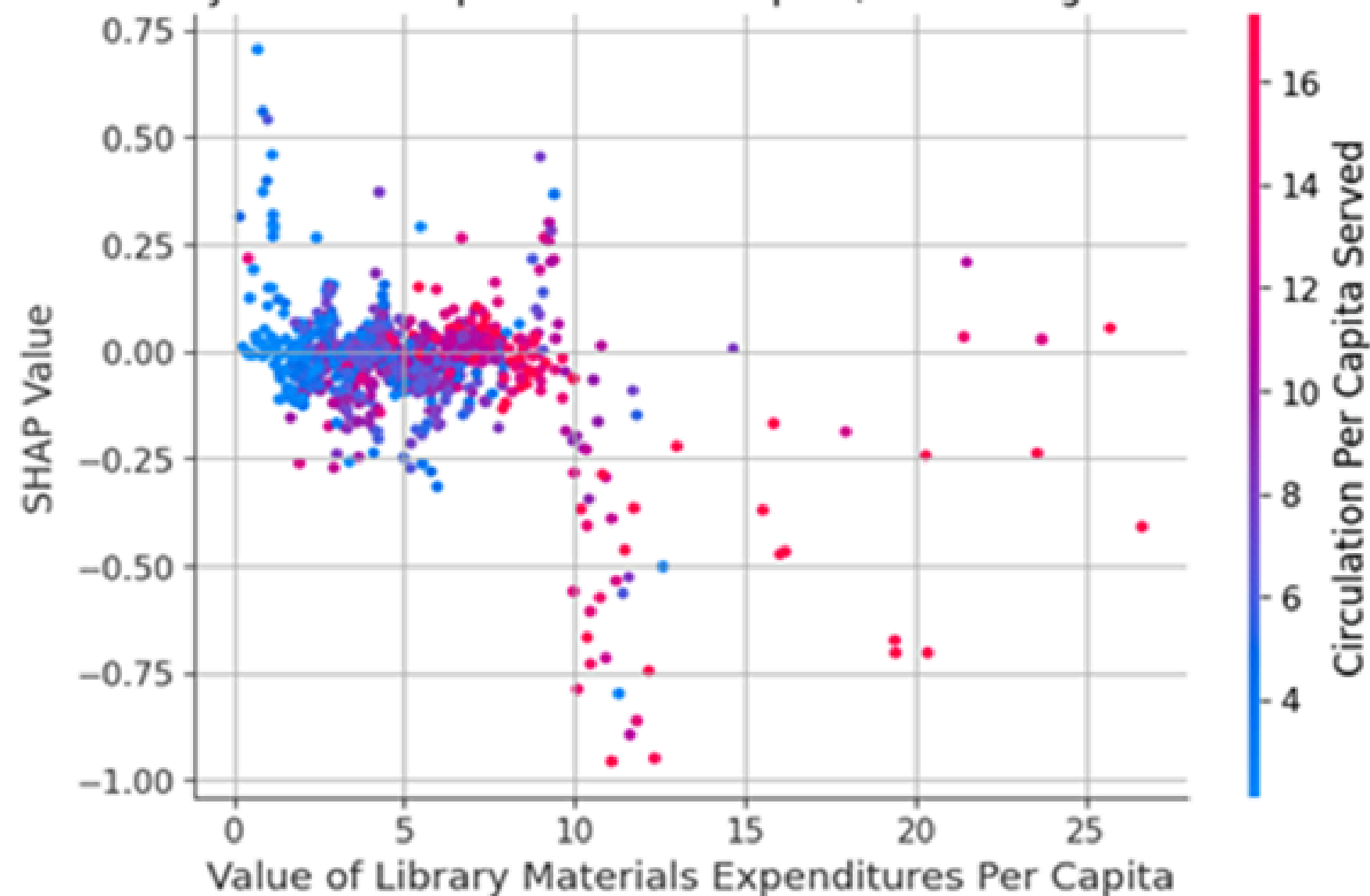
ההשפעה החיובית הגבוהה ביותר על  
הסיווג כספרייה מצטיינת נמצאת  
בטווח של 5–10 השאלות לנפש.  
מחוץ לטווח זה, ההשפעה פוחתת,  
מה שמעיד על חשיבות עידוד  
השאלות בטווח האופטימלי לשימור  
מעורבות קהילתית.



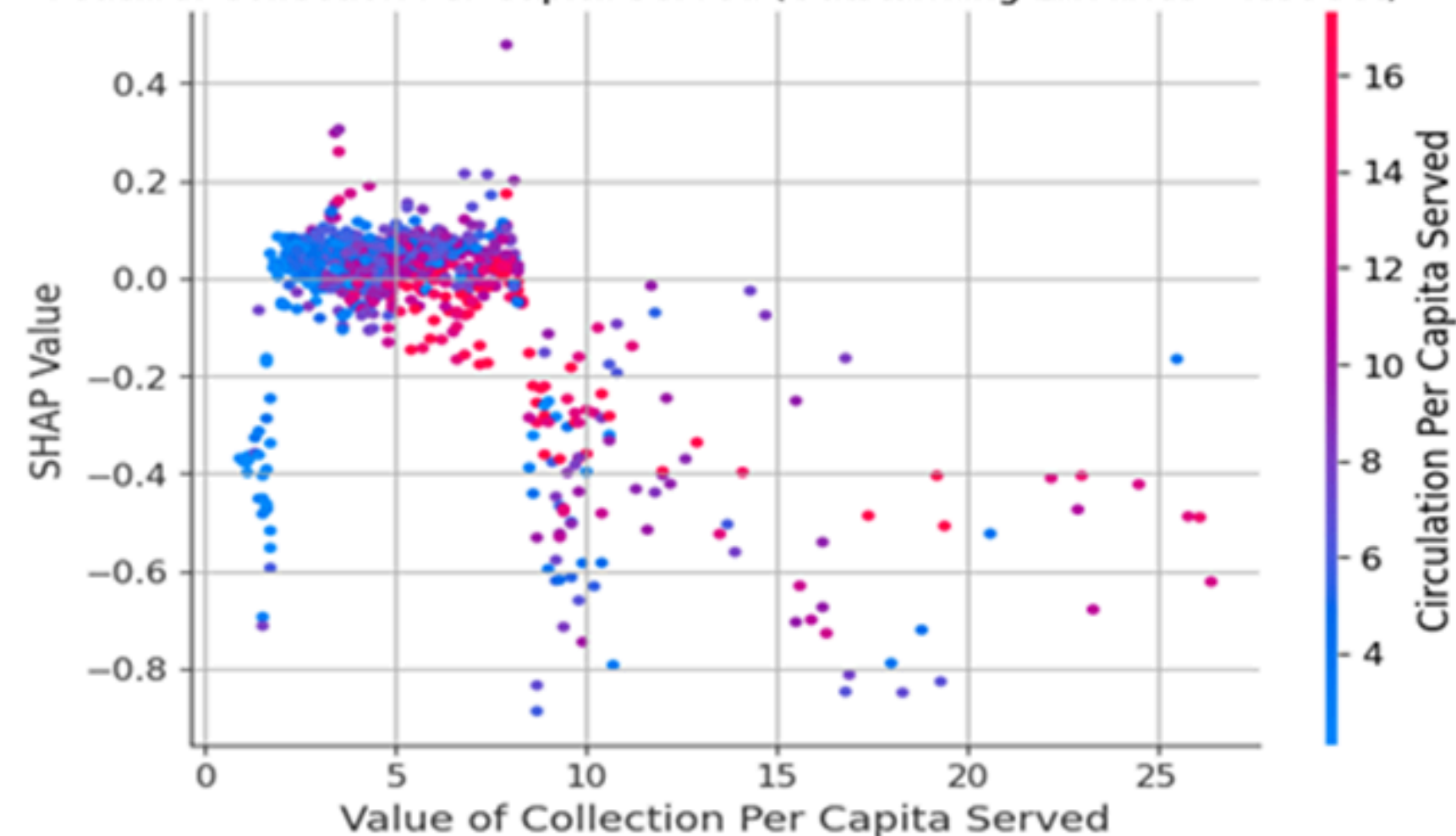
# ניתוח מעמיק

**השקעה בחומרים ובמשאבי הספרייה מעבר  
לכמות מסויימת לרוב אינה רלוונטית ואינה  
תורמת עוד לספרייה להיחשב כמצטיינת!**

Feature: Library Materials Expenditures Per Capita (Outstanding Libraries - Test Set)



Feature: Collection Per Capita Served (Outstanding Libraries - Test Set)



The background features several overlapping geometric shapes in various shades of blue. A large, dark blue triangle points towards the top right. To its left is a dark blue rounded rectangle. Below these are lighter blue shapes, including a rounded rectangle and a triangle, creating a layered, abstract composition.

# תודה על ההקשבה