

Evaluation report

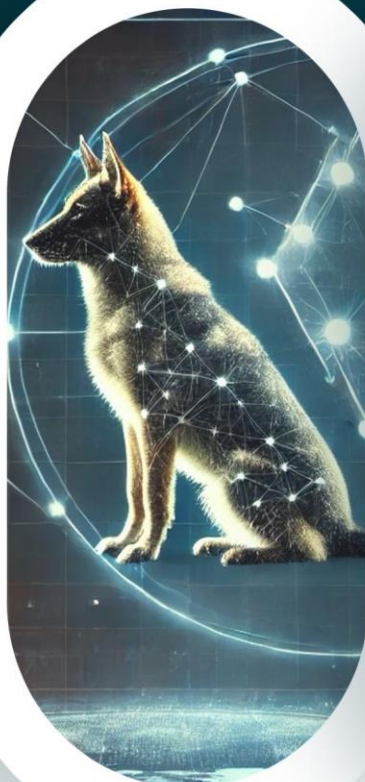
פרויקט גמר מסמך הערכה

מערכת לחיזוי אזורים ומועדים בסיכון לחשיפה
ולהתפרצות מחלת הכלבת והצגת מפת סיכונים.

סטודנטיות מגישות:

אל קוזלי אימאן, ת.ז. 212175582

רם דניאל, ת.ז. 208220509



קורס פרויקט גמר
שנת לימודים תשפ"ה
תאריך הגשה 7.6.25
שם המרצה מר זכאי אבי
שם המנחה גב' גוטפריד ג'ניה

תוכן עניינים

1.....	שער
2.....	תוכן עניינים
3.....	1.הערכת התוצאות (Evaluating the Results)
10.....	2.תהליך הביקורת (Review Process)

1. הערכת התוצאות (Evaluating the Results)

הערכת המודל הסופי מבוססת על תהליך שיטתי שנערך לאחר שלב השוואת המודלים (כמתואר במסמך הקודם), בו נבדקו למעלה מעשרה מודלים שונים לצורך חיזוי אזור התרחשות מקרה הכלבת (Region) וחודש הדיווח (Month). תהליך ההשוואה כלל בחינה אמפירית של הביצועים לפי קריטריונים אובייקטיביים – מדדי דיוק (Accuracy) לכל יעד, ואופן היציבות בין ריצות שונות – ולצידם שיקולים סובייקטיביים כגון יכולת הסבר, נוחות השימוש של המודל, מידת ההתאמה לדאטה ומורכבות הביצוע. במהלך ההשוואה נמדד דיוק ממוצע של כל מודל בכל יעד, שני הקריטריונים המרכזיים שנמדדו עבור כל מודל היו:

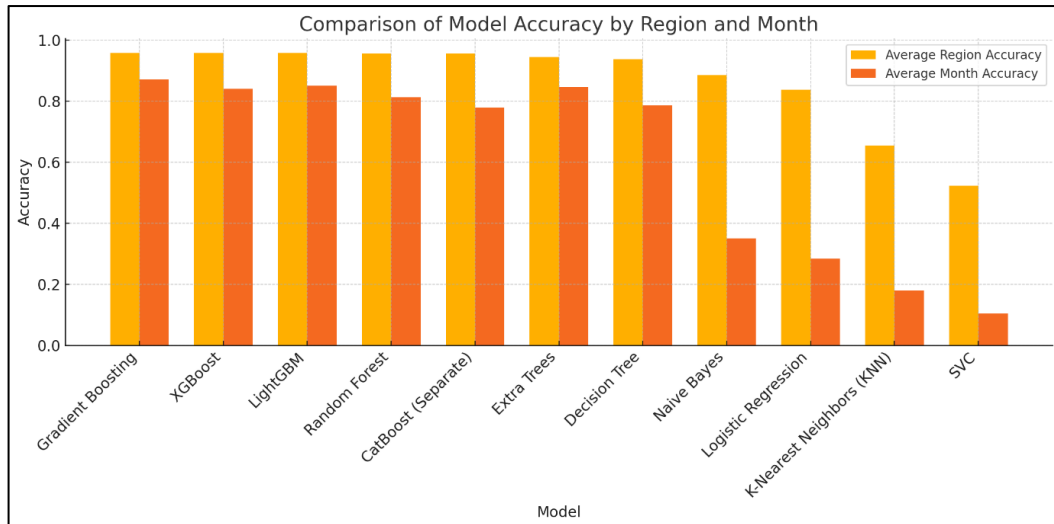
- דיוק ממוצע בזיהוי האזור (Average Region Accuracy)
- דיוק ממוצע בזיהוי החודש (Average Month Accuracy)

מניתוח התוצאות עלה כי המודל המוביל באופן מובהק הוא Gradient Boosting, שהציג דיוק ממוצע של 95.84% עבור יעד האזור, ו-87.08% עבור יעד החודש. מעבר לכך, המודל הפגין עקביות גבוהה, פער נמוך בין הריצות, ותוצאה גבוהה גם כאשר נמדד בשיטות הערכה שונות. מודלים נוספים כמו XGBoost ו-LightGBM הציגו ביצועים קרובים, אך נותרו מעט מאחור, בעיקר בחיזוי חודש. לעומת זאת, מודלים פשוטים יותר כמו Logistic Regression ו-SVC הראו ירידה חדה בביצועים, במיוחד בתחזיות החודשיות, מה שמעיד על קושי בזיהוי משתנים אקלימיים ומורכבות הקשר בין זמן להתפרצות. טבלת ההשוואה הבאה מסכמת את ביצועי כלל המודלים שנבחנו לפי שני מדדי הדיוק:

Average Month Accuracy	Average Region Accuracy	Model
0.870836	0.958406	Gradient Boosting
0.840485	0.957847	XGBoost
0.850862	0.957847	LightGBM
0.813507	0.956440	Random Forest
0.779531	0.956437	CatBoost (Separate)
0.845825	0.944933	Extra Trees
0.786544	0.937351	Decision Tree
0.350346	0.886193	Naive Bayes
0.284864	0.837928	Logistic Regression
0.180327	0.655627	K-Nearest Neighbors (KNN)
0.105049	0.523940	SVC

הגרפים הבאים מציגים השוואה ויזואלית בין ביצועי המודלים שנבחנו, לצורך חיזוי שני היעדים – האזור הגיאוגרפי של ההתפרצות וחדש הדיווח. גרפים אלו מאפשרים להמחיש בצורה ברורה את ההבדלים בין המודלים, ולזהות אילו מהם מצליחים לשמור על ביצועים עקביים בשני סוגי התחזיות.

• גרף 1 – השוואת ביצועי דיוק (Accuracy) בין מודלים שונים בשני יעדי החיזוי: אזור מול חודש



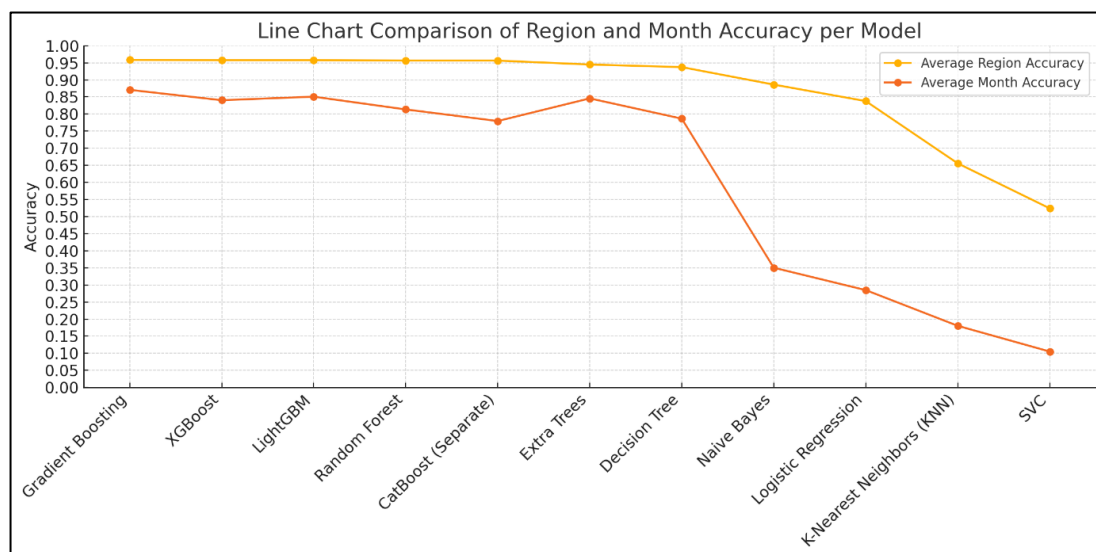
הגרף ה-1 מציג השוואה בין המודלים באמצעות עמודות צבעוניות, כאשר כל זוג עמודות מייצג מודל אחד.

• עמודה כתומה בהירה – מציינת את דיוק המודל בחיזוי האזור (Average Region Accuracy)

• עמודה כתומה כהה – מציינת את דיוק המודל בחיזוי החודש (Average Month Accuracy)

הגרף מדגיש את יתרונותיו של מודל Gradient Boosting, אשר מציג את העמודות הגבוהות ביותר בשני המדדים. גם המודלים XGBoost, LightGBM ו-CatBoost הראו רמות דיוק טובות, אך ירדו מעט בחיזוי החודשי. מנגד, ניתן לראות בבירור את הירידה החדה של מודלים פשוטים יותר כמו Naive Bayes, Logistic Regression, KNN ו-SVC – במיוחד במדד הדיוק החודשי – מה שמעיד על קושי שלהם בזיהוי דפוסיים עונתיים ואקלימיים.

• גרף 2 – השוואת דיוק (Accuracy) של מודלים שונים לפי יעד החיזוי (אזור / חודש): גרף line chart



הגרף 2 מציג השוואה בין ביצועי כלל המודלים שנבחנו, באמצעות גרף קווים כפול. על ציר ה- X מופיעים שמות המודלים, ועל ציר ה- Y ערכי הדיוק (Accuracy) שנעים בין 0 ל-1.

שני הקווים בגרף מתארים :

קו כתום בהיר - (Average Region Accuracy) : מתאר את ממוצע הדיוק לפי אזור.

קו כתום כהה - (Average Month Accuracy) : מתאר את ממוצע הדיוק לפי חודש.

הגרף מדגיש את המודלים ששומרים על רמת דיוק גבוהה בשני המדדים, כמו Gradient Boosting, XGBoost ו-LightGBM. מודלים אלו מראים עקביות גבוהה ולכן מתאימים לשימושים שבהם דרושה תחזית מדויקת גם בממד הזמן וגם בממד המרחב.

לעומת זאת, ניתן לראות כי מודלים פשוטים יותר כמו Naive Bayes, Logistic Regression, KNN ו-SVC מציגים ירידה משמעותית בדיוק לפי חודש, למרות דיוק סביר לפי אזור – דבר שמעיד על מגבלותיהם בהבחנה בין חודשים או בזיהוי משתנים זמן-תלויים.

שני הגרפים יחד תומכים בהחלטה להעמיק את הבדיקה במודל Gradient Boosting בלבד, ולהתמקד בו כחלק משלב ההערכה המעמיקה שמוצג בהמשך סעיף זה.

לסיכום שלב זה, הבחירה הסופית במודל Gradient Boosting התקבלה על סמך תוצאותיו הגבוהות והעקביות בשני יעדי התחזית, בשילוב עם יכולת הסבריות גבוהה וקלות יחסית להטמעה. על כן, בשלב הבא של הערכת התוצאות (שלב זה), אנו מתמקדות בהעמקת הניתוח של מודל זה – תוך שימוש בגרסה חדשה של הקוד, חישוב מדדים נוספים, ביצוע Feature Importance, ניתוח Cross Validation וגרפים נלווים לצורך הסקת מסקנות ודיוק נוסף.

1. הערכת התוצאות שלב 5 evaluation:

בשלב זה ביצענו הערכה מורחבת למודל הסופי שנבחר – Gradient Boosting Classifier – אשר נמצא בעל הביצועים הגבוהים והעקביים ביותר מבין כלל האלגוריתמים שנבדקו בשלב המידול. הקוד עודכן בקובץ נפרד והורכב מחדש בצורה מודולרית, תוך הפרדה ברורה בין שלבי ההרצה, ההערכה, ניתוח הפיצ'רים והצגת התובנות. כחלק מהעבודה על הערכת המודל הוספנו 2 עמודות-פיצ'רים חדשים שעזרו לחיזוי מדויק יותר. לשם הבטחת יציבות התוצאות והעמקת הבדיקות של המודל, הופעלו שתי גרסאות שונות אך כמעט זהות של הקוד, כאשר ההבדל היחיד ביניהן הוא שבגרסה השנייה הופעלה בהרצה נוספת פונקציית Cross Validation עם 5 קיפולים, ובגרסה ה-1 פונקציה זו אינה מופעלת כחלק מהרצת המודל.

שתי הגרסאות כללו חישוב של ארבעה מדדים מרכזיים – Accuracy, Precision, Recall ו-F1 (ממוצע מאקרו) – בנפרד עבור כל אחד משני יעדי התחזית : אזור ההתרחשות (Region) וחודש ההתרחשות (Month). חלוקת הנתונים בוצעה באופן מאוזן תוך שימוש ב-stratify=y כדי לשמר את התפלגות הקטגוריות בכל חלוקה אקראית.

התוצאות המתקבלות מגרסת הריצות הרנדומליות הצביעו על ביצועים מרשימים במיוחד. במודל חיזוי אזור השיג דיוק ממוצע של 95.1% עם סטיית תקן נמוכה, וחיזוי החודש הגיע לדיוק של 89.1% עם F1 של 87.7%. המדדים הנוספים תמכו בעקביות הביצועים והראו רמות דיוק גבוהות בכל ההיבטים.

פירוט מעמיק יותר לגבי הקוד והמודל:

בשלב זה בעצם פותח קוד הערכה מעמיק ומודולרי שתכליתו לבחון את המודל הסופי באופן יסודי, לאחר שלב המידול הראשוני ובחינת האלגוריתמים השונים. הקוד הורכב כך שיאפשר הפרדה ברורה בין שלבי ההכנה, האימון, ההערכה וההמחשה הוויזואלית של התוצאות, תוך הקפדה על שקיפות, אמינות ויכולת הסבר (explainability) של המודל.

במהלך עיבוד הנתונים (Data Preparation) בתוך הקוד החדש בוצעו מספר שדרוגים משמעותיים:

ראשית, נוספה עמודה חדשה בשם Rabies_Cases_Per_Month, המייצגת את מספר מקרי הכלבת בכל חודש ושנה. פיצ'ר זה נבנה מתוך קובץ הנתונים באמצעות חישוב ספירה מצטברת, ומטרתו הייתה להוסיף את כמות אירועי הכלבת בכל חודש לאורך השנים הנתונות בקובץ, ובכך להעניק למודל הקשר עונתי וחוזק סטטיסטי נוסף בתחזית החודש.

בנוסף, שולבה מחדש בהתאם עמודת Event Per Year, אשר הייתה קיימת בנתונים אך לא שולבה בשלב המידול הקודם. פיצ'ר זה מאפשר למודל ללמוד מגמות בין-שנתיות ולשקלל את העומס הכללי של אירועים במהלך כל שנה קלנדרית.

הקוד כלל גם קידוד עמודות קטגוריות, נירמול של משתנים מספריים, ושימוש באיזון קטגוריות (stratify=y) בכל חלוקה. המודל אומן על בסיס Gradient Boosting Classifier ב-5 ריצות נפרדות, תוך אפשרות להפעלת Cross Validation עם 5 קיפולים.

בתהליך ההערכה חושבו מדדי דיוק מרכזיים – Accuracy, Precision, Recall ו-F1 – עבור כל אחד משני יעדי התחזית (Month ו-Region), ונותחו ממוצעים סטטיסטיים על פני הריצות.

בנוסף, בוצע ניתוח חשיבות פיצ'רים (Feature Importance) בנפרד עבור כל יעד, כולל גרף השוואתי המדגיש אילו משתנים תורמים יותר לתחזית מרחבית לעומת תחזית עונתית. כמו כן, הופקה טבלת ציר (Pivot Table) להצגת סך מקרי הכלבת לפי שנה וחודש, לצורך ניתוח מגמות כלליות לאורך זמן.

• פלט הטבלה והחישובים עבור עמודת Rabies_Cases_Per_Month:

Pivot Table:												
Month	0	1	2	3	4	5	6	7	8	9	10	11
Year												
2006	0	1	0	0	1	0	2	1	0	1	0	2
2007	0	6	1	2	0	0	3	1	0	0	1	1
2008	0	2	1	0	0	0	0	1	1	1	0	1
2009	2	15	23	10	4	2	2	0	4	3	10	10
2010	4	2	4	10	6	7	2	3	1	5	5	5
2011	5	1	2	1	6	2	6	4	1	0	1	2
2012	2	1	0	2	2	1	1	3	9	2	0	1
2013	1	3	2	0	2	4	3	6	2	0	1	4
2014	1	0	3	5	0	1	0	0	0	0	0	3
2015	1	4	4	4	3	0	0	4	2	2	1	1
2016	3	0	2	4	4	2	3	0	1	1	3	6
2017	3	14	20	21	3	2	2	2	1	1	2	3
2018	24	1	1	1	11	8	3	6	2	2	1	1
2019	4	3	2	2	1	1	1	0	0	1	2	0
2020	2	9	10	5	5	2	3	4	1	2	1	2
2021	5	3	2	5	6	2	8	3	3	2	1	0
2022	5	1	5	2	3	2	2	0	0	0	5	4
2023	8	1	3	6	4	4	4	3	4	3	4	1
2024	5	6	2	8	7	7	9	3	2	1	3	2
2025	7	0	0	0	9	9	0	0	0	0	0	0

Rabies Case Count per Year-Month:			
Year	Month	Rabies Case Count	
0	2006	1	1
1	2006	4	1
2	2006	6	2
3	2006	7	1
4	2006	9	1
...
185	2024	10	3
186	2024	11	2
187	2025	0	7
188	2025	4	9
189	2025	5	9

[190 rows x 3 columns]

שלב זה של הערכה מעמיקה סיפק תובנות קריטיות על אופן קבלת ההחלטות של המודל, אישר את עמידתו ביעדי הדיוק שהוגדרו מראש, ותרם לחיזוק יכולת ההסבר והאמינות של המערכת כולה.

- לינק לקוד ולמודל המעודכן בסביבת google colab :
https://colab.research.google.com/drive/1bDx-6s84lh_cTIEJOV7X-esS7BOAjkdt?usp=sharing
- קובץ הנתונים עליו עובד המודל :
content/Rabies__Weather__War_Combined_1.4.25.xlsx/
- צילום הפלט המסכם את מדדי הביצוע עבור Region ו-Month (Accuracy, Precision, Recall, F1)

```
Performance summary for 'Region' (based on 5 random splits):
Accuracy: 0.9510 ± 0.0147
Precision (avg): 0.8892
Recall (avg): 0.8859
F1 Score (avg): 0.8758
```

```
Performance summary for 'Month' (based on 5 random splits):
Accuracy: 0.8937 ± 0.0304
Precision (avg): 0.8909
Recall (avg): 0.8775
F1 Score (avg): 0.8770
```

בנוסף, הרצת הקוד בגרסת Cross Validation אפשרה לבחון את מידת היציבות של המודל. גם כאן נרשמה רמת ביצועים גבוהה מאוד בתחזית האזור ($\text{Accuracy} \approx 90.6\% \pm 6.4\%$), בעוד שבתחזית החודש זוהתה שונות גבוהה יותר ודיוק נמוך יותר ($\text{Accuracy} \approx 57.4\% \pm 6.5\%$), כנראה עקב חפיפה עונתית או רגישות גבוהה למאפיינים זמנים דומים.

- צילום של הפלט עבור Cross Validation של Region ושל Month

```
Cross Validation results for 'Region':
Accuracy: 0.9061 ± 0.0647
Precision (avg): 0.8458
Recall (avg): 0.8318
F1 Score (avg): 0.8153

Cross Validation results for 'Month':
Accuracy: 0.5745 ± 0.0654
Precision (avg): 0.5706
Recall (avg): 0.5474
F1 Score (avg): 0.5138
```

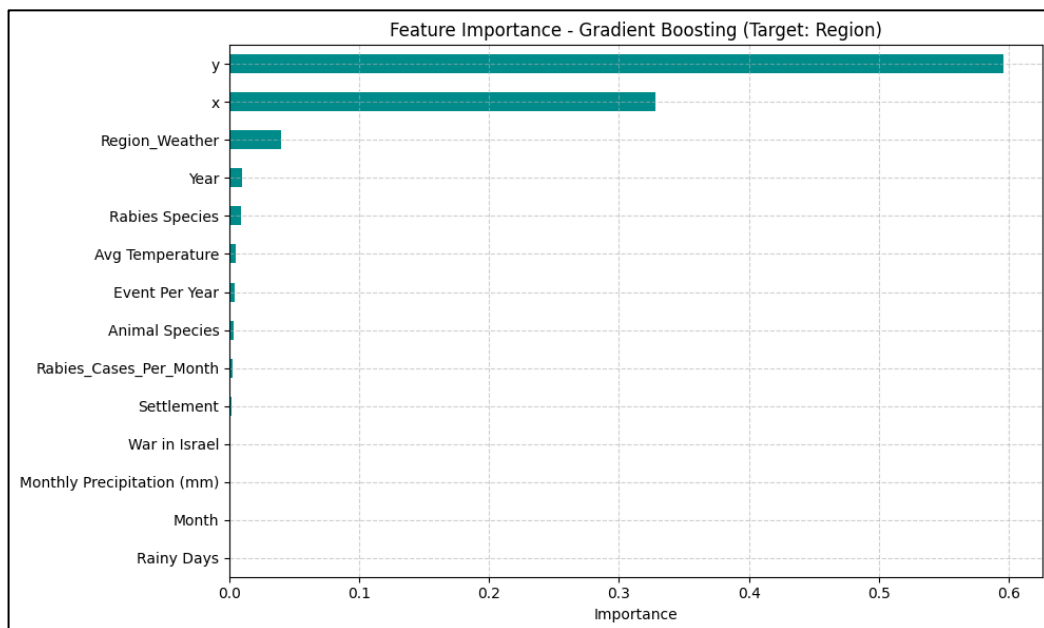
אחד מהשיפורים המרכזיים שנוספו בגרסה זו הוא ניתוח חשיבות משתנים (Feature Importance) עבור כל אחד מהיעדים. הממצאים חיזקו את ההבנה התיאורטית והמעשית שלנו בנוגע לתרומת הפיצ'רים לתחזית.

עבור חיזוי האזור, שני הפיצ'רים הדומיננטיים ביותר היו קווי הרוחב (y) והאורך (x), עם משקל של 59.6% ו-32% בהתאמה. משתנים נוספים כמו תנאי מזג האוויר וסוגי חיות השפיעו בצורה שולית בלבד. המסקנה היא שחיזוי מיקום ההתפרצות אכן מבוסס במידה רבה על משתנים גיאוגרפיים כגון Y, X.

לעומת זאת, בחיזוי החודש, תרמו בעיקר משתנים אקלימיים וזמניים – הטמפרטורה הממוצעת (47%), כמות מקרי הכלבת החודשיים (13.7%), ומספר האירועים בשנה (15.4%). פרמטרים כמו משקעים וימי גשם השפיעו גם הם, אך בעוצמה נמוכה יותר.

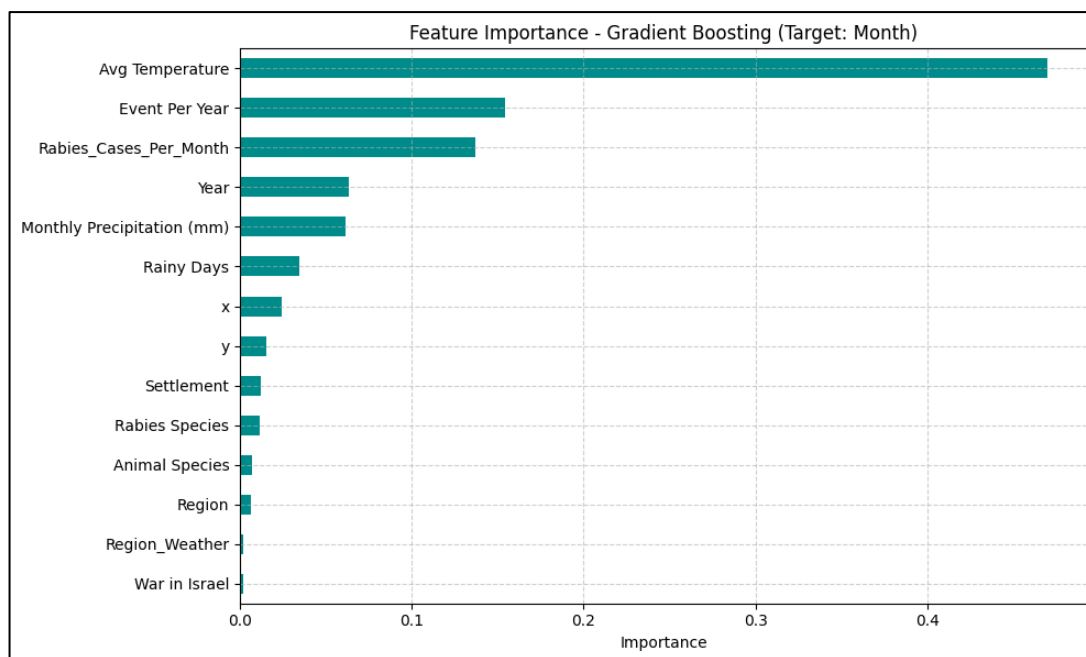
- תוצאות וגרף עמודות של feature importance עבור Region אזור :

```
Feature Importances:
y                0.596275
x                0.328146
Region_Weather  0.039703
Year            0.009896
Rabies Species  0.008795
Avg Temperature 0.004838
Event Per Year   0.004056
Animal Species   0.003301
Rabies_Cases_Per_Month 0.002208
Settlement       0.001456
War in Israel    0.000836
Monthly Precipitation (mm) 0.000195
Month            0.000159
Rainy Days       0.000137
dtype: float64
```



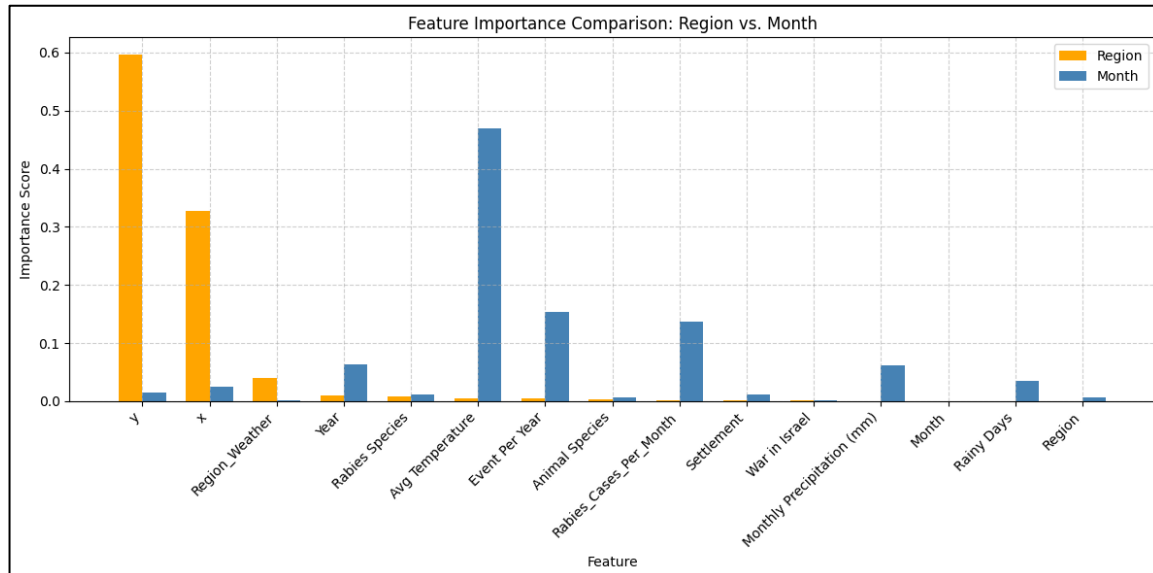
- תוצאות וגרף feature importance עבור Month חודש :

```
Feature Importances:
Avg Temperature  0.470151
Event Per Year    0.154299
Rabies_Cases_Per_Month 0.137267
Year            0.063162
Monthly Precipitation (mm) 0.061300
Rainy Days       0.034320
x                0.024183
y                0.015059
Settlement       0.012182
Rabies Species   0.011242
Animal Species   0.006662
Region           0.006527
Region_Weather   0.002080
War in Israel    0.001566
dtype: float64
```

לצורך הבנת ההבדלים במבנה ההחלטות של המודל, יצרנו גרף השוואתי בין תרומת הפיצ'רים לשני יעדי התחזית. הגרף מדגים בצורה ברורה כיצד משתנים מרחביים שולטים בתחזית אזור ההתפרצות, בעוד שמשתנים עונתיים ואקלימיים הם המרכזיים בתחזית חודש ההתפרצות/אירוע.

• גרף ההשוואה בין השפעת/חשיבות הפיצ'רים על 2 יעדי החיזוי האזור והחודש – Comparison Plot :



מעבר לביצועים הגבוהים, עלו מהניתוח גם תובנות חדשות. לדוגמה, הפיצ'ר Event Per Year, שלא הוערך כמשפיע משמעותית בתחילת הדרך, התברר כמשתנה בעל חשיבות גבוהה בחיזוי החודש. נתון זה עשוי להעיד על השפעה של מגמות בין-שנתיות או תדירות כללית של התפרצויות לאורך זמן. תובנה זו מעלה את האפשרות להוספת משתנה "עונה" (Season) או יצירת פיצ'רים סדרתיים מבוססי זמן (כגון time-lags) לשיפור המודל.

המסקנה הכוללת היא ששני המודלים שנבנו עומדים היטב ביעדים העסקיים שהוגדרו בתחילת הפרויקט – השגת דיוק של מעל 80% – ואף עולים עליהם. Gradient Boosting הוכיח את יכולתו לספק תוצאות מדויקות, פרשנות טובה ועמידות גבוהה, תוך שימוש במגוון מאפיינים מורכבים ממקורות נתונים שונים. לכן, הוא נבחר כמודל הסופי למימוש בשלב הבא.

2. תהליך הביקורת (Review Process)

אחת מהעקרונות המרכזיים של מתודולוגיית CRISP-DM היא ביצוע תהליך ביקורת בסיום שלב העבודה, שבו נערכת רפלקציה כנה על הדרך שעברנו, על ההחלטות שקיבלנו, על נקודות החוזק והחולשה בתהליך, ועל הלקחים שנוכל ליישם בפרויקטים עתידיים. במסגרת פרויקט זה, בו עברנו דרך חמישה שלבים עיקריים – החל מהבנה עסקית, דרך הבנת הנתונים והכנת נתונים ומידול הנתונים, ועד להערכת המודל והפרויקט – התבצע לאורך כל הדרך תהליך ביקורתי שיטתי, שנשען על תיעוד, בדיקה עצמית ויכולת למידה מתמשכת.

במהלך הפרויקט ראינו כיצד תכנון יסודי בשלב ההתחלתי תרם ליצירת מטרות ברורות, מדידות ושימויות. הגדרנו שני יעדי חיזוי: חיזוי האזור הגיאוגרפי של מקרי הכלבת וחיזוי החודש שבו התרחשו, תוך התאמה ליכולת של אלגוריתמים לפתרון בעיות מסוג סיווג מרובה יעדים (Multi-Output Classification). הגדרה מדויקת זו אפשרה לנו לבנות שלבי עבודה מדורגים וברורים, ולבחור בגישות טכניות מתאימות בהמשך הדרך.

עבודת הצוות שהתנהלה לאורך הפרויקט התבססה על חלוקת תפקידים ברורה, אחריות עצמאית של כל אחת מאיתנו על חלק מהמשימות, ותיאום הדדי קבוע. כך הצלחנו לקדם את העבודה באופן שיטתי ולשמור על גמישות מחשבתית והתאמה תוך כדי תנועה, בעיקר כאשר התברר לנו שיש להכניס שינויים או לנסות כיוונים חדשים.

בשלבם הטכניים נתקלנו בקשיים שונים, בעיקר באינטגרציה של מקורות נתונים מגוונים (משרד הבריאות, מזג אוויר, נתוני בעלי חיים), שכל אחד מהם הצריך טיוב, התאמה מבנית והמרות ערכים. בשלב המידול, בנינו מערך בדיקות השוואה בין למעלה מעשרה מודלים שונים תוך שימוש בקוד מודולרי ומבוקר. עם זאת, גילינו שמודלים פשוטים כמו Logistic Regression ו-Naive Bayes כשלו בעיקר במשימת חיזוי החודש, ולעומתם מודלים מבוססי עצים כמו Gradient Boosting הראו ביצועים טובים ויכולת הכללה גבוהה יותר.

תהליך הביקורת הדגיש במיוחד את חשיבות הנדסת הפיצ'רים. אחת התובנות המרכזיות הייתה שהוספת עמודות חדשות (כמו מספר מקרי כלבת חודשי או שנתי) שיפרה משמעותית את הביצועים, לעיתים יותר מהחלפת מודל. בנוסף, שילוב ניתוח חשיבות משתנים (Feature Importance) סיפק עומק נוסף להבנת תחזיות המודל, והוביל לתובנות חדשות – למשל, שהתחזית האזורית מתבססת בעיקר על קואורדינטות מיקום, בעוד התחזית החודשית דורשת משתנים עונתיים ואקלימיים בנוסף להיסטורית אירועי הכלבת.

בחלק מהשלבים נדרשה גמישות והיערכות מחדש. לדוגמה, כאשר גילינו שמודל CatBoost לא תומך במבנה נתונים מרובה יעדים, נדרשנו להתאים את שיטת העבודה או להמיר מבנה הנתונים. כמו כן, לאורך העבודה למדנו להפעיל שיפוט ביקורתי על תוצאות ראשוניות, לחפש הסברים ולהפעיל מנגנוני בקרה עצמיים – בין אם על ידי גרפים, השוואות בין ריצות, או קריאה מחדש של הנתונים.

סקירה רפלקטיבית של חמשת שלבי העבודה בפרויקט מאפשרת זיקוק של תובנות משמעותיות שיכולות לתרום לשיפור תהליכים עתידיים בפרויקטים דומים.

בשלב ההבנה העסקית, הצלחנו להגדיר מטרות חיזוי ישימות, מדידות ורלוונטיות – חיזוי אזור גיאוגרפי וחודש התרחשות של מקרי כלבת – באופן שתאם לתוצאות בשלב ההערכה ומידול הנתונים מבחינת ביצועי המודל ומגבלות הנתונים והיישום. יחד עם זאת, הבנו בדיעבד כי הגדרה מוקדמת של מאפיינים עונתיים או סביבתיים הייתה יכולה להעשיר את המודל ולשפר את החיזוי הזמני, בנוסף לשימוש במאגר נתונים גדול יותר מהקיים שמכיל תיעוד רק מ-20 השנים האחרונות. בשלב הבנת הנתונים, זיהינו אתגרים מבניים באינטגרציה של מקורות נתונים הטרוגניים, מה שהוביל אותנו להכיר בצורך בתכנון מוקדם של מיפוי משתנים, טיפול בערכים חסרים, אחידות פורמטים ותיעוד מלא של השדות כבר בשלב הראשוני.

שלב הכנת הנתונים בוצע ברמת פירוט גבוהה, תוך יישום שיטות קידוד, נירמול, טיוב ערכים והנדסת פיצ'רים, ומתוכו התחזקה ההבנה שלפיה איכות המבנה הנתוני והפיצ'רים משפיעה במידה רבה כמו הבחירה באלגוריתם מסוים. שלב המידול אופיין בגישה שיטתית של השוואת ביצועים על פני מספר רב של מודלים, ובמהלכו למדנו לאזן בין מדדים כמותיים לבין שיקולים פרשניים ואופרטיביים – דוגמת יציבות בין ריצות, גמישות יישומית ויכולת הסבר ויעילות המודל מבחינת זמן ריצה ואורך הפלט. לבסוף, שלב הערכת המודל סימן את השלב האנליטי המעמיק ביותר, שבו שולבו ניתוחי Feature Importance, חישובי ביצועים על פני ריצות חוזרות, Cross Validation וגרפים תומכי החלטה, אשר תרמו להבנה מעמיקה של פעולת המודל, חוזקותיו והיבטי השיפור האפשריים בו.

לסיכום, תהליך הביקורת הוא רכיב מרכזי בהתנהלות הפרויקט. הוא מחדד את הדרך בה אנו חושבות, מתכננות ובונות תהליך אנליטי מלא. הוא מאפשר לנו לשפר את תוצרי העבודה, להבין לעומק את משמעות המדדים, את תפקוד המודלים ואת המגבלות של כל שלב. הניסיון הזה חיזק את יכולתנו לנהל פרויקטים עתידיים במתודולוגיה מקצועית, ביקורתית ורפלקטיבית יותר, ולבצע בחירות מושכלות לאורך שלבי העבודה, גם כשהן דורשות שינוי כיוון או חשיבה מחודשת. לאחר מסמך זה אנחנו מוכנות להמשיך עבודה בתהליך של פרויקט הגמר ובכך להתקדם לשלב המימוש deployment עם המודל העדכני.

בברכה,

דניאל ואימאן.