

Deployment report

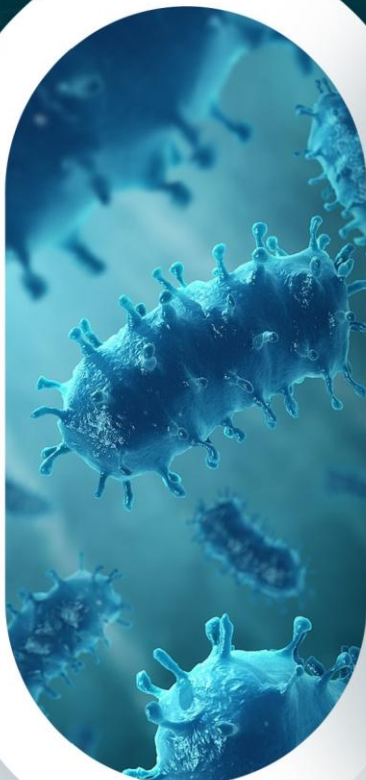
פרויקט גמר - מסמך מימוש ופריסה

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: 09.2025
שם המרצה: מר זכאי אבי
שם המנחה: גב' גוטפריד ג'ניה

תוכן עניינים

1.....	שער
2.....	תוכן עניינים
3.....	1. תוכנית מימוש/פריסה/הטמעה (Deployment Plan)
5.....	2. תכנון ניטור ותחזוקה (Planning Monitoring and Maintenance)
7-15.....	3. נספח תיעוד Dashboards

1. תוכנית מימוש/פריסה/הטמעה (Deployment Plan)

הפרויקט שלנו עוסק במחלת הכלבת ובניתוח מקרים מאומתים של המחלה לאורך זמן. אספנו נתונים אמיתיים ממקורות רשמיים, ביצענו עיבוד וניקוי מקיף לנתונים, ובנינו מודל חיזוי מסוג Gradient Boosting. המודל מתמקד בניבוי האזור והחודש שבהם עלול להתרחש האירוע הבא, ובכך מספק כלי משמעותי לאיתור מוקדם, הקצאת משאבים ותכנון צעדי מניעה יעילים.

שלב הפריסה (Deployment) מהווה נקודת מפתח קריטית בכל פרויקט ניתוח נתונים. לאחר תהליכי איסוף הנתונים, ניקוי, עיבוד והכנה, בחירת המודל המתאים ואימון המערכת, מגיע השלב שבו עלינו להפוך את התובנות והמודלים לכלים מעשיים שניתן ליישם, לשתף ולהפיק מהם ערך ממשי. מטרת שלב זה היא להנגיש את התוצרים שפותחו בצורה ברורה ושימושית, כך שקהל היעד, בין אם מדובר באנשי מקצוע, חוקרים או מקבלי החלטות, יוכל להיעזר בהם לקבלת החלטות ותכנון עתידי.

תוכנית הפריסה שלנו בפרויקט זה שמה דגש על מספר ערוצי יישום ושיתוף, שנועדו להבטיח את שימור העבודה, תחזוקתה לאורך זמן, והנגשה של התוצרים גם עבור מי שאינם בעלי רקע טכני. במסגרת התוכנית שילבנו כלים מגוונים כגון GitHub, Hugging Face ו-Streamlit אשר ביחד מרכיבים פתרון לפריסה, שיתוף והפעלה של המודל והדשבורדים שפיתחנו.

בדוח הזה נפרט את תוכנית הפריסה שבוצעה באמצעות הפעולות הבאות:

1. שמירה, ניהול ובקרת קוד באמצעות GitHub

כל קוד המקור של המערכת, החל משלב עיבוד הנתונים, דרך בניית המודל ועד לפיתוח הדשבורד, הועלה למאגר GitHub ייעודי. פלטפורמה זו מאפשרת ניהול גרסאות מסודר, כך שכל שינוי או שדרוג בקוד מתועד ונשמר. יתרה מזאת GitHub מאפשר שיתוף פעולה אפקטיבי בין חברי הצוות, מימוש עקרונות של עבודה בצוותים טכנולוגיים, ותיעוד מלא של תהליך הפיתוח. מעבר לכך, עצם השימוש בו מבטיח כי הפרויקט יוכל להמשיך להתפתח גם בעתיד, תוך אפשרות לאיתור ותיקון באגים, הוספת פיצ'רים חדשים ושיפור ביצועי המודל לאורך זמן.

מה שמאפשר לנו:

- שמירה וגיבוי בענן: כל שינוי נשמר ומגובה אוטומטית עם אפשרות לשחזור גרסאות קודמות.
- סטנדרט בתעשייה: שמירה ב-Git הוא חלק בלתי נפרד מתהליכי הפיתוח בתעשייה, השילוב שלו בפרויקטים אקדמיים מדמה סביבת עבודה אמיתית ונותן התנסות.

על מנת להעלות את כל המסמכים של הפרויקט בצורה מסודרת יצרנו repositories עם הקבצים מסודרים בתיקיות בהתאם למבנה המקצועי שמוסבר גם בקבצי readme.

יצירת ממשק אינטראקטיבי והנגשה באמצעות Streamlit

Streamlit היא פלטפורמה וספריית Python שמאפשרת לבנות אפליקציות אינטראקטיביות, באופן יעיל יחסית מבלי לכתוב קוד ותשתית מאפס בשפות כגון JavaScript, CSS, HTML.

כדי להבטיח שהתוצרים לא יישארו ברמה תאורטית בלבד, פיתחנו 2 דשבורדים אינטראקטיביים אפשריים לפרויקט שלנו. דשבורד אחד בעזרת Streamlit ודשבורד נוסף בעזרת Hugging Face Streamlit. הדשבורדים מאפשרים הצגה חזותית של תחזיות המודל, הצגת גרפים אינטואיטיביים

וביצוע ניתוחים אינטראקטיביים של הנתונים. בכך, גם משתמשים שאינם בעלי רקע טכני יכולים להבין ולהפיק ערך מהתוצאות שהופקו. הכלים שפותחו מהווים דוגמה ליישום פרקטי של מודל חיזוי, ומאפשרים קבלת החלטות מושכלות על בסיס נתונים בזמן אמת. בנוסף, הדשבורדים מהווים שכבת ממשק ידידותית בין המודל לבין המשתמש, ומדגימים כיצד ניתן לשלב תובנות מדע הנתונים במערכות מידע קיימות או בתהליכי קבלת החלטות.

2. אחסון, שיתוף ונגישות המודל באמצעות Hugging Face

Hugging Face היא פלטפורמה שמאפשרת שימוש במודלים בצורה פשוטה ויעילה, ללא צורך בתשתית מורכבת או משאבים מיוחדים. הפלטפורמה תומכת בשיתוף פעולה עם קהילה רחבה של מפתחים ובספריות ושפות תכנות שונות, מה שמעניק גמישות ומקל על פיתוח וניתוח נתונים.

היתרון המרכזי של השימוש ב־Hugging Face הוא שקיפות מלאה ויכולת לשחזר תוצאות בקלות, והכל במסגרת חנימית שמאפשרת לפרוס מודלים ולהתנסות בהם במהירות וללא עלות.

כדי להעלות את המודל יוצרים קודם Space חדש באתר Hugging Face ובחרים את סוג ה Space למשל Streamlit, לאחר מכן מעלים את כל קבצי הפרויקט וקובץ requirements.txt שמכיל את כל הספריות הדרושות. עם ההעלאה Hugging Face מריץ את הכל באופן אוטומטי וה Space הופך לזמין באינטרנט, כך שכל משתמש יכול לגשת ולהשתמש בו דרך דפדפן.

לסיכום

באמצעות השילוב בין Streamlit, GitHub, cursor, Hugging Face יצרנו תוכנית פריסה מלאה המאפשרת שמירה ותחזוקה של הקוד, שיתוף המודל עם הקהילה, והנגשה נוחה ואינטראקטיבית של התוצרים למשתמשי קצה. בדרך זו הצלחנו להפוך את המודל שלנו למחלת הכלבת מכלי מחקרי בלבד לכלי עם פוטנציאל מעשי ויישומי, שניתן בעתיד לשלבו בתהליכי עבודה שוטפים, להציגו לקהל יעד מקצועי או ציבורי, ולהמשיך לפתחו בעתיד שכן מדובר בפרויקט לימודי ולא במוצר שלם מוכן לתעשייה.

הרחבה יישומית והשלמות (Addendum)

החלק הבא מרחיב את היישום בפועל ותכנון לעתיד של המימוש/פריסה והנגשת המודל בשני הדשבורדים שנבנו – Streamlit (פיתוח מקומי בסביבת Cursor) ו-Hugging Face Spaces (פריסה אונליין) – ומוסיף מפרטים תפעוליים, מפת דרכים (Roadmap), ושיקולי איכות ותפעול נוספים.

א. חיבור למקורות נתונים (תכנון עתידי) - אינטגרציה מתוכננת ל-APIs: מאגרי נתוני כלבת בישראל, Visual Crossing ו-Meteostat לצורך נתוני מזג אוויר ונתוני סביבה/חיות בר.

ב. קלטים/פלטטים בדשבורד (מפרט תפעולי) - קלטי משתמש: אזור/יישוב, תאריך, סוג בעל-חיים, מאפייני מזג אוויר, ועוד שדות קונפיגורציה לפי הצורך.

פלטטי מודל: הסתברות להתפרצות וסימון בינארי ("קיים סיכון"/"אין סיכון") עם קוד צבעים, לצד טבלת תחזיות לשורות הקלט.

ג. שקיפות המודל – Feature Importance - הצגה בדשבורד של Feature Importance (לדוגמה: אזור, סוג בעל-חיים, פרמטרי מזג אוויר) לחיזוק אמון המשתמשים, שיפור פרשנות והסבר החלטות.

ד. מפות אינטראקטיביות ומיקוד אזורי - שילוב שכבות מפה (Heatmap ונקודות) להצגת מוקדי סיכון.

אפשרויות סינון מהירות לפי אזור (North / Center / South) ותצוגת מיקומים מדויקים.

ה. דוחות ומגמות (תכנון עתידי) - הפקת דוחות שבועיים/חודשיים על מגמות התפרצות, לרבות יצוא נתונים/איורים לצורכי ניטור מתמשך, תיעוד והקצאת משאבים.

ו. אבטחה והרשאות - גישה מאובטחת והרשאות לפי תפקיד (וטרינרים, גורמי אכיפה, חוקרים).

הקפדה על פרטיות משתמשים ואזרחים, כולל נתונים על חיות בית ומשק, ועל תאימות לנהלי הארגון.

ז. מפת דרכים תפעולית (Roadmap) לדשבורד – Cursor + Streamlit:

- שמירת וטעינת מודל - לאחר אימון, המודל נשמר כקובץ pickle; נטען באפליקציית Streamlit להפעלה מיידי.
- טעינת נתונים - העלאת CSV/Excel או שימוש בדאטה היסטורי שמור; מאחורי הקלעים מתבצעים ניקוי והשלמת ערכים חסרים.
- סינון ופרמטריזציה - פילטרים מהירים לפי אזור/חודש/סוג חיה, בחירת מאפיינים (Feature toggles) לפי תרחיש שימוש.
- הצגת תוצאות וניתוח - תוצאות בלייב (באותו Tab), טבלת תחזיות לכל רשומה, גרפים של מקרים לפי אזור/חודש, Feature Importance, ו-מדדי ביצועים (Accuracy / Precision / Recall / F1) כחלק מה-UI.
- פריסה לענן - Hugging Face Spaces - יצירת Space מסוג Streamlit, העלאת קוד וקובץ requirements.txt. הרצה אוטומטית מספקת נגישות אונליין אחידה לקהל היעד, ללא התקנות.
- ח. Best Practices למניעת כשלים (איכות ותפעול):
 - מניעת Data Leakage בכל שלבי אימון/עדכון/פריסה, הימנעות מ-Features בעייתיים, להתמקד בתחזית סיכון עתידי, להימנע מ"דשבורד ריק": לצרף שכבת הסבר/תיעוד קצרה לקריאת התוצאות ולפרשנותן.
 - הערה: כל הסעיפים לעיל חלים הן על הדשבורד המקומי ב-Streamlit (פיתוח ב-Cursor) והן על פריסת ה-Hugging Face Spaces, כאשר רכיב ה-Roadmap מגדיר את הזרימה מקצה לקצה – מאימון ושמירת מודל, דרך העלאת נתונים וסינונם, ועד להצגה, שקיפות ואיכות.

2. תכנון ניטור ותחזוקה (Planning Monitoring and Maintenance)

לאחר פיתוח מודל החיזוי ובדיקת ביצועיו, נדרש שלב נוסף וקריטי של ניטור ותחזוקה מתמשכים. מטרתו של שלב זה היא להבטיח שהמודל ישמור על רמת דיוק גבוהה לאורך זמן, יתמודד בהצלחה עם שינויים בנתונים, וימשיך לייצר תחזיות אמינות גם בסביבות משתנות.

1. מעקב אחרי גורמים משפיעים

בפרויקט החיזוי של מחלת הכלבת קיימים מספר גורמים חיצוניים שעשויים להשפיע על ביצועי המודל לאורך זמן, ולכן יש צורך במעקב קבוע אחריהם שמאפשר לזהות מוקדם תופעות חדשות ולהיערך לעדכון המודל בהתאם.

- מיקום גיאוגרפי: ישנם אזורים בארץ שבהם נמצאו התפרצויות בתדירות גבוהה יותר מאחרים. שינוי בפיזור המקרים, למשל מעבר התפרצות לאזור חדש עלול להשפיע על אמינות התחזית.
- סוגי בעלי חיים: חלק מהחיות כגון תנים או שועלים נוטות להדבקה גבוהה יותר מבעלי חיים אחרים. שינוי ביחס המינים הנדבקים עשוי לשנות את דפוסי המחלה.
- שינויים במבנה הנתונים: לעיתים נתונים חדשים נאספים בפורמט שונה, או שחלק מהעמודות שנכללו במודל אינן מתעדכנות עוד. מצב זה עלול להוביל להטיות במודל ולפגוע בדיוק.

2. בדיקה וניטור של ביצועי המודל

כדי להבטיח שהמודל ימשיך לספק תחזיות אמינות גם לאחר פריסתו בפועל, נדרש מערך שיטתי של בדיקות תקופתיות:

- בדיקות חוזרות: יש להריץ את המודל על נתונים חדשים המתקבלים באופן שוטף, למשל אחת לחודש או לרבעון, ולבחון האם הביצועים נשארים ברמה מספקת.
- מדדי ביצועים: יש למדוד באופן עקבי את מדדי הדיוק השונים ולהשוות אותם לערכים שהושגו בעת האימון הראשוני. ירידה עקבית במדדים אלו היא סימן לכך שהמודל מאבד מיכולתו לנבא.
- בקרת דריפט: יש לעקוב אחר תופעות של Data Drift (שינוי בהתפלגות הקלטים) ו Concept Drift (שינוי בקשר בין משתני הקלט לתוצאה). למשל, אם בעבר מרבית המקרים הגיעו מצפון הארץ ופתאום מתחילה עלייה חדה בדרום, זהו סימן ל Data Drift המחייב בחינה מחודשת.

3. זיהוי מודל שפג תוקפו

מודל חיזוי נחשב ככזה ש"פג תוקפו" כאשר הוא אינו מסוגל עוד לספק תחזיות אמינות או כאשר תופעות חדשות משנות באופן מהותי את התמונה. ניתן לזהות זאת במספר דרכים:

- ירידה מתחת לסף ביצועים מוגדר מראש: לדוגמה, אם הוחלט שהמודל חייב לשמור על דיוק של לפחות 85% כדי להיחשב אמין, ירידה עקבית מתחת לרמה זו מצביעה על צורך באימון מחדש.
- שינוי בנתונים: הופעת משתנים חדשים, היעלמות של נתונים חשובים, או שינוי בקצבי האיסוף עלולים לשבש את יכולת המודל לעבוד כפי שעבד בעבר.
- הופעת דפוסים חדשים: אם מתגלים מקרי כלבת בסוגי בעלי חיים שלא נכללו במערך האימון המקורי, המודל אינו מסוגל להתמודד עם התופעה החדשה.

במצב כזה נדרש תהליך של עדכון ואימון מחדש של המודל על בסיס נתונים עדכניים, כדי להשיב אותו לרמת ביצועים מיטבית.

הרחבה יישומית והשלמות לניטור ותחזוקה (Addendum)

- החלק הבא מוסיף קריטריונים תפעוליים, מטריקות, תדירויות ובקורות איכות שמיושמים בפועל במערכת ומדייקים את עבודת הניטור באופן שוטף.
- א. גורמים לניטור – השלמות תחום
- מזג האוויר: שינויים עונתיים ותנאי קיצון (גלי חום/קור, גשם חריג) המשליכים על דפוסי התפשטות.
 - דפוסי אוכלוסיית בעלי חיים: תנודות במספר השועלים/תנים/כלבים משוטטים.


- שינויים סביבתיים ותשתיתיים : פיתוח/סגירת שטחים פתוחים, שינוי שימושי קרקע, אזורי מגורים/חקלאות/ פריצה וסגירת של גדרות גבולות בזמני מלחמה ושלוש ואכיפת הרשויות.
 - מדיניות ואכיפה : קמפיינים, חיסונים, עיקור/סירוס והגברת מודעות.
 - מחלות/תופעות נוספות : הופעת נגיפים/מחלות חדשות בבעלי חיים.
- ב. מטריקות ביצוע ומדיניות מדידה
- מדידה חודשית של Precision, Recall ו-F1 בקטגוריה "התפרצות קיימת", לצד Accuracy למעקב כללי.
 - השוואת תחזיות ל-Ground Truth מהשטח בכל מחזור מדידה ; ניתוח פערים על פני אזורי (North/Center/South) וחודשים.
- ג. ספי התרעה והכרזה על "פקיעת תוקף"
- סף Recall קריטי : אם Recall בקטגוריית "התפרצות קיימת" $0.9 >$ במשך שני חודשים רצופים – המודל יוכרז כלא תקף ויוזנק תהליך עדכון.
 - סטיית התפלגות מאפיינים : זיהוי שינוי מובהק סטטיסטית במאפיינים מרכזיים (לדוגמה מעבר עומק מוקדי סיכון לאזור חדש) ידרוש בחינה מחודשת.
 - הופעת מודלים עדיפים : אם קיימת אלטרנטיבה עם ביצועים משופרים באופן מובהק – יש לשקול החלפה/Ensemble.
- ד. תדירות ניטור ורה"אימון (Retraining)
- איסוף נתונים חודשי/רבעוני והטמעתם לצורך עדכון תכופ של התמונה האפידמיולוגית.
 - אימון מחדש לפחות פעמיים בשנה, או מוקדם יותר בעת ירידה במדדים/חריגות בהתפלגות.
 - בחינה ממוקדת אזורית/עונתית בעת זיהוי Drift נקודתי (למשל, גל חריג בדרום).
- ה. לולאת משוב משתמשים מהשטח
- איסוף פידבק שוטף מוטרנירים/רשויות מקומיות/משרד החקלאות על תחזיות שגויות/חסרות.
 - הטמעה למחזור השיפור : תיעדוף בקשות, עידון פיצ'רים רגשיים (למשל משקלי מזג אוויר/עונתיות), ושיפור חוויית השימוש בדשבורד.

בברכה,

דניאל ואימאן.

3.נספח תיעוד Dashboards

Huggingface dashboard:



Rabies / Weather / War Prediction

הזן פרטי רשומה חדשה

Animal Species

Badger

Rabies Species

I

Settlement

Abbirim

Region Weather

Center

War in Israel

Yes

x

0.00

-

+

y

0.00

-

+

Avg Temperature

20.00

-

+

Monthly Precipitation (mm)

50.00

-

+

Rainy Days

10.00

-


+

Year

2025

-

+

 Make Prediction >>

✓ Model Prediction: Region = Negev (65.38%), Month = October (95.07%)

Most Similar Record to Your Input (Similarity Based)

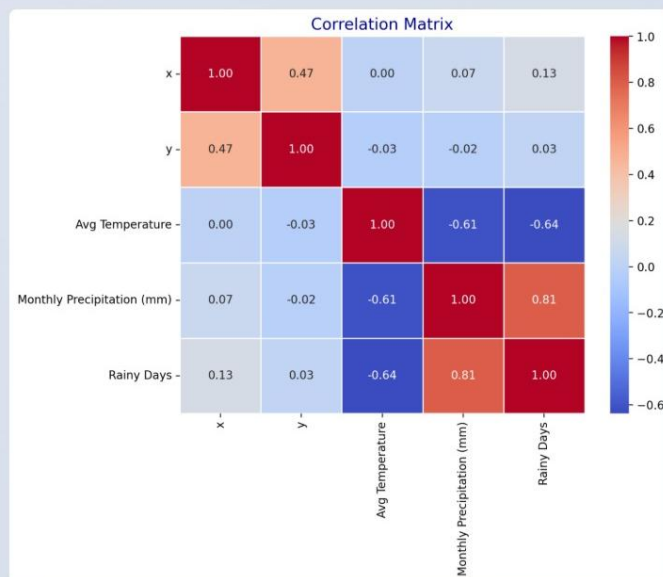
The record from the existing dataset that is most similar to your input:

	Index Event ID	Event Per Year	Date	Year	Month	Animal Species	Rabies Species
673	39	9	2009-04-05 00:00:00	2009	4	Sheep	I
58	654	21	2024-04-10 00:00:00	2024	4	Badger	Na
704	8	9	2006-10-24 00:00:00	2006	10	Fox	I
536	176	6	2011-02-13 00:00:00	2011	2	Dog	VII
485	227	8	2013-02-28 00:00:00	2013	2	Dog	VII
682	30	11	2008-11-05 00:00:00	2008	11	Jackal	I
290	422	26	2018-02-06 00:00:00	2018	2	Jackal	VII B
293	419	24	2018-02-04 00:00:00	2018	2	Jackal	VII B
34	678	46	2024-11-19 00:00:00	2024	11	Badger	Na
547	165	50	2010-12-19 00:00:00	2010	12	Cattle	III

Correlation Matrix (Numeric Features)

The correlation matrix shows the pairwise **Pearson correlation coefficients** between numeric features.

- Values close to 1 indicate a strong positive correlation.
- Values close to -1 indicate a strong negative correlation.
- Values around 0 indicate little or no linear correlation.



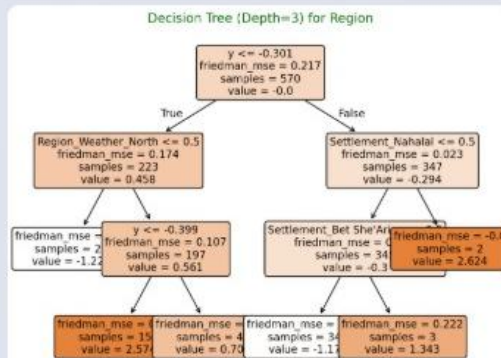
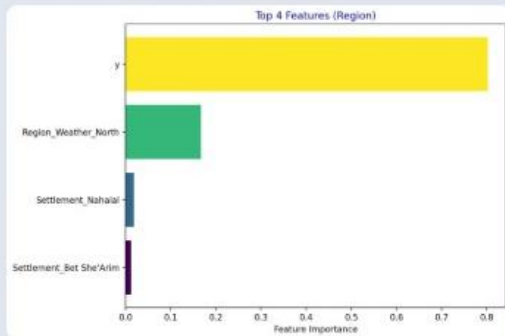
Pearson p-values (Numeric Features)

The Pearson p-values indicate the statistical significance of the correlation between numeric features.

- A **small p-value** (typically < 0.05) suggests that the correlation is statistically significant.
- A **large p-value** suggests that the correlation could be due to random chance.
- Diagonal cells are **NaN** because a feature's correlation with itself is not tested.

	x	y	Avg Temperature	Monthly Precipitation (mm)	Rainy Days
x	None	0.000	0.944	0.057	0.000
y	0.000	None	0.362	0.640	0.505
Avg Temperature	0.944	0.362	None	0.000	0.000
Monthly Precipitation (mm)	0.057	0.640	0.000	None	0.000
Rainy Days	0.000	0.505	0.000	0.000	None

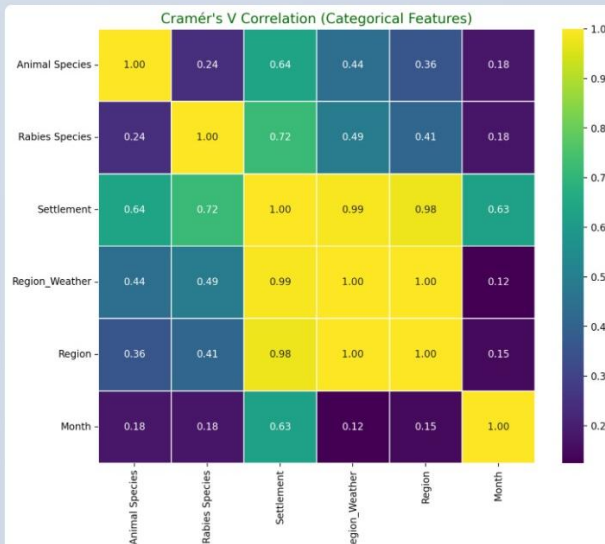
Target Name : Region



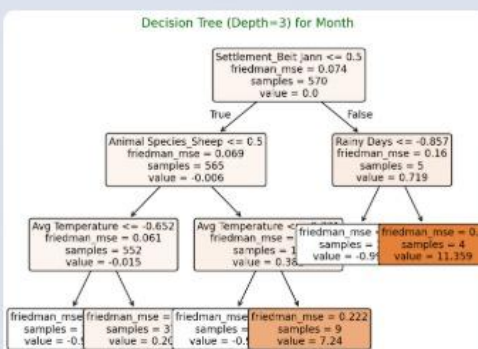
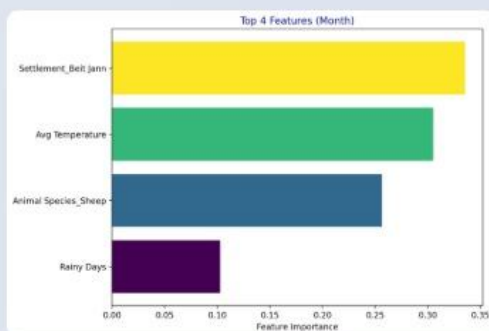
Cramér's V (Categorical Features + Targets)

Cramér's V measures the strength of association between categorical variables.

- Values range from 0 to 1:
 - 0 → no association
 - 1 → perfect association
- Higher values indicate stronger relationships between the categories.
- This includes both the original categorical features and the target variables (e.g., Region, Month).




Target Name : Month



Streamlit dashboard:



Prediction:




Rabies Prediction System

This dashboard provides machine learning-based predictions for rabies outbreaks and comprehensive data analysis tools.

Models loaded successfully!

 Prediction  Data Visualization



Rabies Outbreak Prediction

Make predictions for rabies outbreaks using our trained machine learning models.

Models loaded successfully!


Choose prediction mode:

☒ Single Observation ☐ Upload File ☐ Use Existing Data

Single Observation Prediction

Enter the details of a new observation to get predictions.

Feature Selection (Optional)

☐  Use Feature Selection ?

main

prediction

visualization

Navigation

System Status

☒ Models Loaded

Region Classes: 6


Month Classes: 12

☒ Preprocessor Loaded

Single Observation Prediction

Enter the details of a new observation to get predictions.

Feature Selection (Optional)



☐  Use Feature Selection ?


Input New Observation

Date	Region Weather	Rainy Days
2025/09/27	Center	10
Animal Species	X Coordinate	<input type="checkbox"/> War in Israel
Badger	0.00	Events Per Year
Rabies Species	Y Coordinate	10
I	0.00	
Region	Average Temperature (°C)	
Amakim	20.00	
Settlement	Monthly Precipitation (mm)	
Abbirim	50.00	

Make Prediction

Rabies Prediction System | Built with Streamlit

 Prediction  Data Visualization



Rabies Outbreak Prediction

Make predictions for rabies outbreaks using our trained machine learning models.

Models loaded successfully!


Choose prediction mode:

☐ Single Observation ☒ Upload File ☐ Use Existing Data

File Upload Prediction

Upload an Excel or CSV file to make batch predictions.

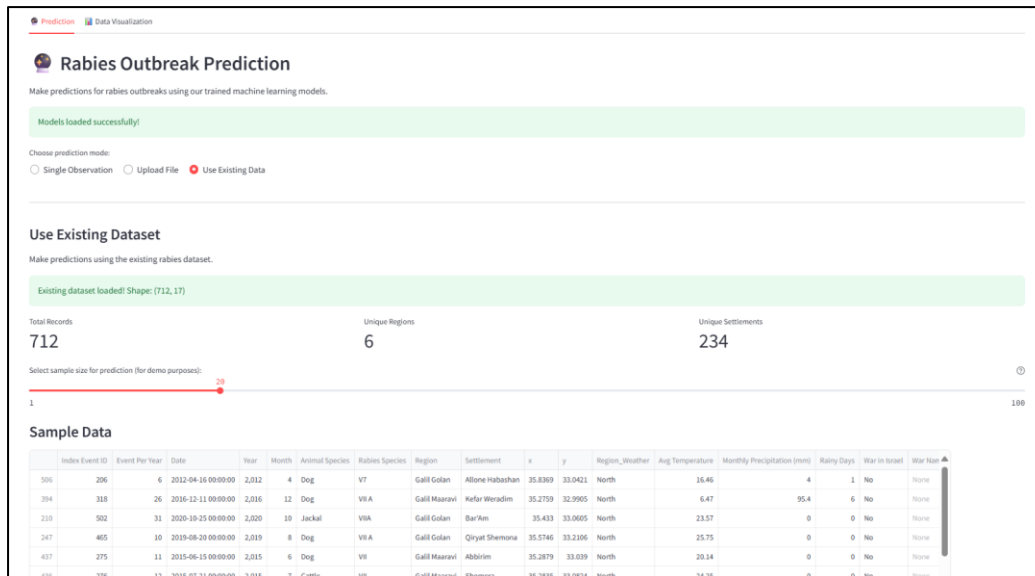
Choose a file

 Drag and drop file here

Limit 200MB per file • XLSX, CSV

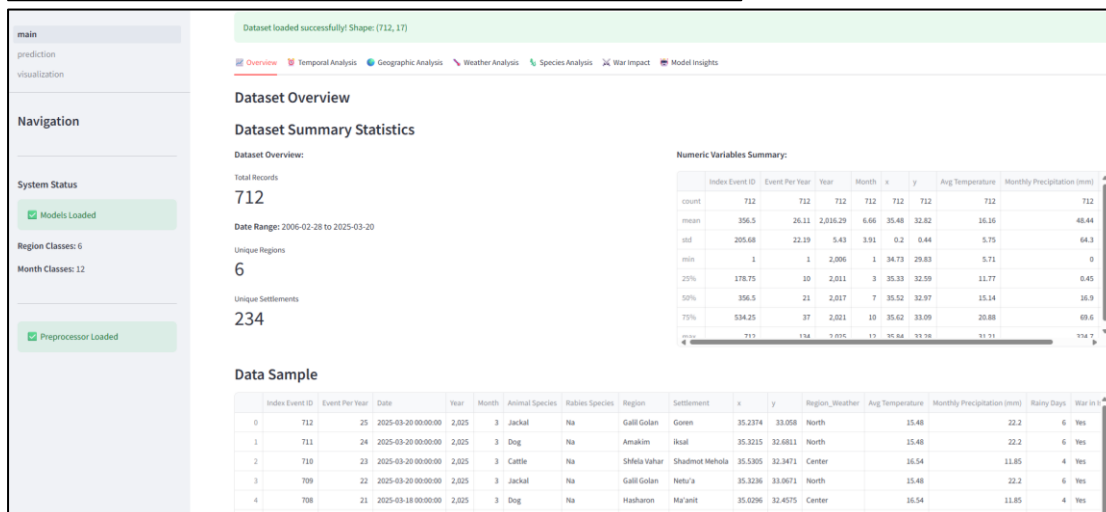
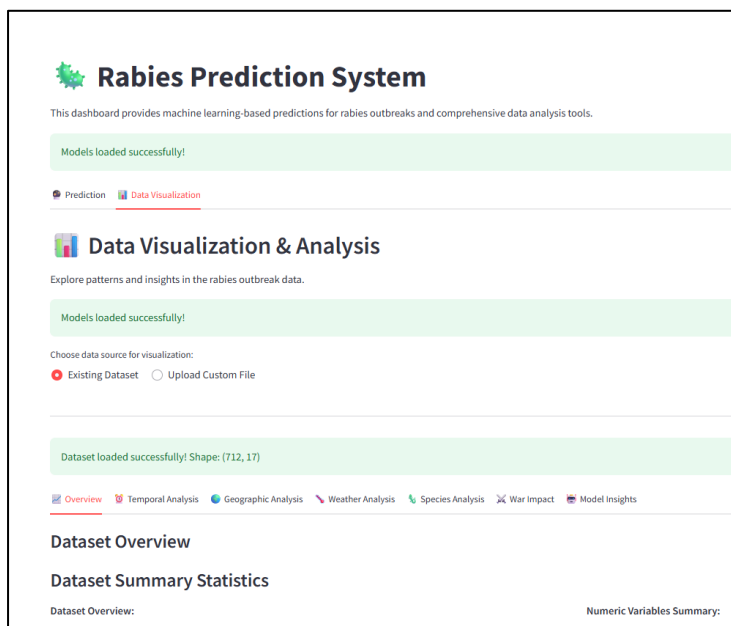
Browse files

Rabies Prediction System | Built with Streamlit



Streamlit dashboard:

Data Visualization



main

prediction

visualization

Navigation

System Status

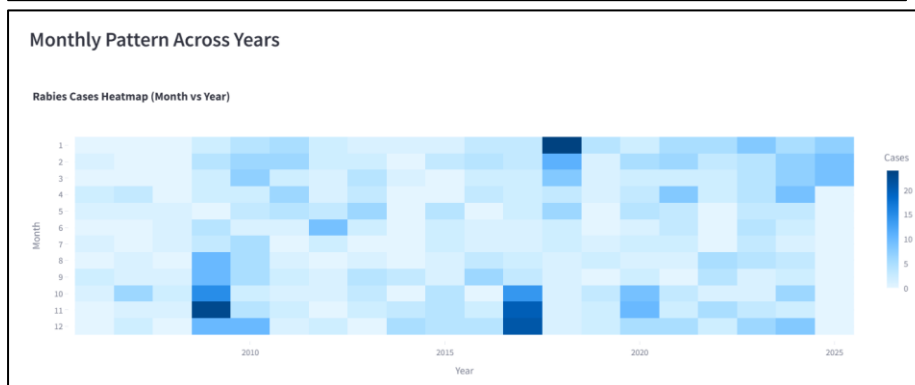
Models Loaded

Region Classes: 6

Month Classes: 12

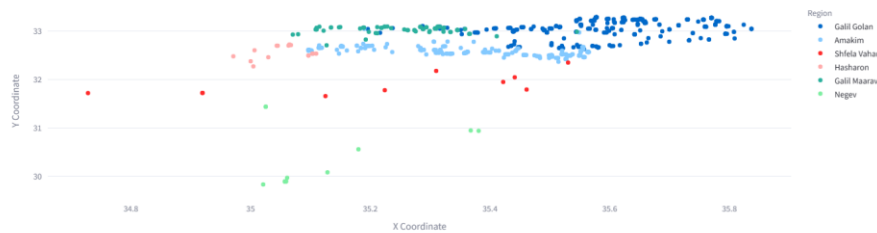
Column Information

	Column	Data Type	Non-Null Count	Null Count
	Index Event ID	int64	712	
	Event Per Year	int64	712	
	Date	datetime64[ns]	712	
	Year	int64	712	
	Month	int64	712	
	Animal Species	object	712	
	Rabies Species	object	712	
	Region	object	712	
	Settlement	object	712	
	x	float64	712	



Geographical Distribution of Cases

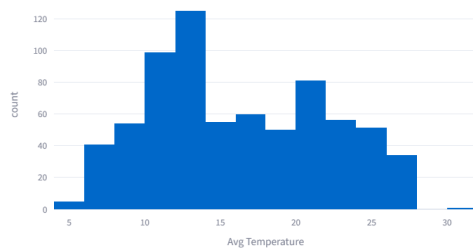
Geographical Distribution of Rabies Cases



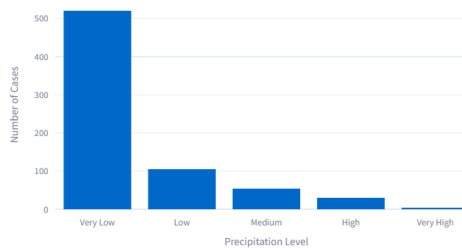
Overview Temporal Analysis Geographic Analysis **Weather Analysis** Species Analysis War Impact Model Insights

Weather Impact Analysis

Distribution of Average Temperature

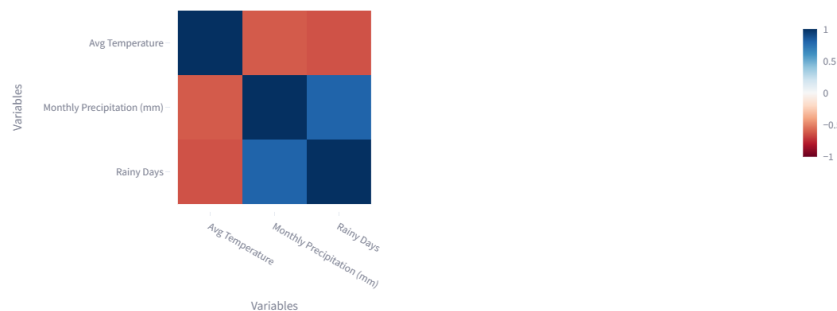


Cases by Precipitation Level



Weather Variables Correlation

Weather Variables Correlation Matrix



Overview Temporal Analysis Geographic Analysis **Weather Analysis** **Species Analysis** War Impact Model Insights

Species Analysis

Distribution by Animal Species

