תסריט מעודכן להצגת הפוסטר - Rootzone Prediction

מטרה: הצגה של כ-24 דקות, דמו של 3-4 דקות, והשאר לשאלות שופטים.

שפת דיבור: עברית. להשאיר מונחים מקצועיים באנגלית כאשר הם מופיעים בפוסטר או בקוד: Root-zone, pH, EC, Soft Sensor, Micro-climate, XGBoost, Walk-forward validation, Holdout, MAE, R2, Baseline, Feature engineering.

הערה חשובה: לא להזכיר בהצגה את שם הגרסה הטכנית הפנימית. מול השופטים אומרים פשוט "המודל הסופי".

0:00-3:00 - פתיחה, רקע חקלאי והבעיה המרכזית

מציג: Eden

להצביע בפוסטר על: Motivation.

שלום, אנחנו Eliyahu Eden ו-Chaushu Oren. הפרויקט שלנו עוסק בחיזוי משתני Root-zone בחממה, בעיקר pH ו-EC.

לפני שנכנסים למודל עצמו, חשוב להבין מה זה בכלל Root-zone. Root-zone הוא אזור בית השורשים - הסביבה שבה השורשים נמצאים וקולטים מים, דשן ומינרלים. בחממה מודרנית, במיוחד בגידול במצע מנותק, האזור הזה רגיש מאוד. אם תנאי בית השורשים יוצאים מאיזון, הצמח יכול להיפגע גם אם מעל פני הקרקע הכול נראה תקין.

בפרויקט שלנו התמקדנו בחיזוי שניים מהמשתנים החשובים ביותר בבית השורשים: EC ו-pH.

EC, כלומר Electrical Conductivity, הוא מדד שמייצג את כמות המלחים המומסים בתמיסה סביב השורשים. הוא נותן לנו אינדיקציה לריכוז הדשן והמלחים שהצמח פוגש בפועל. אם EC נמוך מדי, זה יכול להעיד על חוסר הזנה או על תמיסה מדוללת מדי. אם EC גבוה מדי, זה יכול להעיד על הצטברות מלחים ועל סטרס אוסמוטי - מצב שבו לצמח קשה יותר לקלוט מים גם אם יש מים בסביבה.

pH הוא מדד לחומציות או בסיסיות של התמיסה סביב השורשים. הוא חשוב כי הוא משפיע על זמינות יסודות ההזנה לצמח. גם אם נתנו דשן בכמות נכונה, pH מחוץ לטווח האופטימלי יכול לגרום לכך שחלק מהיסודות יהיו פחות זמינים, ולכן הצמח לא באמת יוכל לנצל את כל ההזנה שקיבל.

כלומר, pH ו-EC הם לא רק מספרים טכניים. הם מספרים לנו האם סביבת השורשים מאפשרת לצמח לקלוט מים ודשן בצורה טובה, או האם מתפתח סיכון ליבול, לקצב צימוח, לבריאות הצמח וליעילות השימוש במים ובדשן.

הבעיה המעשית היא שברוב החממות כן מבצעים מדידות מדויקות של pH ו-EC, אבל המדידות האלה נעשות בתדירות נמוכה יחסית. לכן רוב הזמן אנחנו לא יודעים מה המצב האמיתי של בית השורשים, אלא רק מה היה המצב בנקודות הדגימה האחרונות.

בין שתי מדידות יכולים להתרחש שינויים משמעותיים בגלל דישון, השקיה, הוספת חומצה, שינויי מזג אוויר, קרינה חזקה, אידוי או הצטברות מלחים. אם השינויים האלה יוצרים חריגה מהטווח האופטימלי, נגלה את זה רק בפעם הבאה שנדגום. ואז נצטרך לבצע פעולה מתקנת אחרי שהבעיה כבר הופיעה.

אנחנו לא רוצים להגיע למצב הזה. הרעיון שלנו הוא לבנות חיישן תוכנה שמאפשר לחזות, על סמך הנתונים שכבר זמינים בחממה, מה צפוי להיות מצב בית השורשים. כך אפשר להתאים מראש את תוכניות הדישון וההשקיה לפי מזג האוויר, התכנון המקורי, מצב הצמח והמדידה האחרונה, ולוודא שאנחנו נשארים בטווח האופטימלי.

במילים אחרות, המטרה היא לעבור מגישה ריאקטיבית - שבה מזהים חריגה ורק אז מתקנים - לגישה פרואקטיבית, שבה מנסים לצפות את הבעיה מראש ולשנות את התוכנית לפני שהצמח נפגע.

3:00-5:30 - מטרת הפרויקט, התוצר והערך היישומי

מציג: Eden

להצביע בפוסטר על: Project Goal.

המטרה שלנו הייתה לבנות Soft Sensor לבית השורשים. זה לא חיישן פיזי שמודד ישירות, אלא מודל שמעריך את pH ו-EC מתוך נתונים אחרים: אקלים, השקיה, דישון, מצב הצמח, והמדידה הידנית האחרונה.

התוצר שאליו כיוונו הוא מערכת תומכת החלטה: המגדל מזין או מעלה את תמונת המצב הנוכחית, כולל pH ו-EC אחרונים, נתוני אקלים, מצב הצמח ותכנון השקיה ודישון; המערכת מחזירה תחזית לערכי pH ו-EC בזמן יעד עתידי. הערך היישומי הוא מעבר מניהול תגובתי לניהול

פרואקטיבי: אם המודל מזהה שתרחיש השקיה או דישון צפוי להוביל ל-pH או EC בעייתיים, אפשר לשנות את התוכנית לפני שהבעיה מופיעה במדידה הידנית הבאה.

איך רצינו להשיג את זה?

1. לבנות ציר זמן אחיד שמחבר את כל מקורות הנתונים.
2. לחזות את תנאי המיקרו-אקלים בתוך החממה כאשר אנחנו מסתכלים קדימה.
3. לבנות מודל Root-zone שלומד את השינוי מ-t0 ל-t1.
4. לעטוף את המודל בדשבורד שמדגים איך מגדל או חוקר יכול להפעיל את התחזית.

5:30-8:30 - הדאטה: מקורות, סדר גודל והכנה

מציג: Eden

להצביע בפוסטר על: Data.

בנינו בסיס נתונים אחיד ברזולוציה של 10 דקות. בגרסה הסופית יש 16,693 רשומות, 20 עמודות, מטווח התאריכים 29 במאי עד 21 בספטמבר 2025.

הנתונים הגיעו מכמה קבוצות:

1. נתוני מזג אוויר חיצוניים וקרינה - תחנת מזג אוויר וקובצי radiation חיצוניים. אלו נותנים לנו את התנאים שמניעים את החממה מבחוץ.
2. נתוני מיקרו-אקלים בתוך החממה - טמפרטורה פנימית, לחות, קרינה פנימית, ET0 וטמפרטורת קרקע או הערכה שלה. אלו מייצגים את הסביבה שהצמח באמת חווה.
3. נתוני ממשק גידול - השקיה, דישון, סוגי דשן, חומצה, מלחים, גבס ו-Kortin. אלו הפעולות שהמגדל מבצע, והן משפיעות ישירות על בית השורשים.
4. נתוני מצב צמח - כיסוי נוף וגיל הצמח. אלו חשובים כי צמח צעיר וצמח מפותח לא צורכים מים ודשן באותו אופן.
5. מדידות ידניות של pH ו-EC - אלו תוויות האמת של מודל Root-zone.

האתגר המרכזי הוא שהנתונים צפופים מאוד בצד האקלימי והתפעולי, אבל תוויות pH ו-EC דלילות מאוד: יש רק 109 מדידות ידניות. כלומר, יש לנו הרבה מידע שיכול להסביר את התהליך, אבל מעט נקודות אמת שבהן אנחנו יודעים בוודאות מה היה מצב בית השורשים.

העבודה על הדאטה כללה:

- יישור timestamps מכל הקבצים.
- המרה לפורמט זמן אחיד של 10 דקות.
- איחוד עמודות שהגיעו בשמות ובפורמטים שונים.
- חישוב משתנים חסרים או נגזרים, למשל ET0 או טמפרטורת קרקע.
- סיכום אירועי השקיה ודישון לחלונות זמן.
- חיבור מדידות pH ו-EC לנקודות עוגן ולנקודות יעד.

8:30-13:30 - ארכיטקטורה ושיטה: למה שני שלבים ואיך מתכננים את החיזוי

מציג: Oren

להצביע בפוסטר על: Method, Micro-Climate Model, Architecture.

תכננו את המערכת כ-pipeline דו-שלבי.

הסיבה לפיצול היא שבזמן חיזוי עתידי יש לנו שתי בעיות שונות:

1. קודם צריך לדעת מה יהיו תנאי החממה הפנימיים.
2. אחר כך צריך להשתמש בתנאים האלה כדי לחזות את השינוי ב-pH וב-EC.

אם היינו בונים מודל אחד שמקבל רק נתוני עבר, הוא היה נשאר בעיקר תרגיל offline. אבל המטרה היא לאפשר תחזית. לכן השלב הראשון הוא Micro-climate prediction.

שלב 1 - Micro-climate prediction

השלב הזה מקבל נתוני מזג אוויר וקרינה חיצוניים, ומייצר את משתני האקלים הפנימיים שהמודל השני צריך: טמפרטורה פנימית, ET0, קרינה פנימית וטמפרטורת קרקע.

המטרה שלו היא לגשר בין מה שזמין לתחזית מבחוץ לבין מה שהצמח חווה בתוך החממה. זה חשוב במיוחד כי נתונים פנימיים עתידיים לא תמיד זמינים בזמן אמת.

שלב 2 - Root-zone prediction

בשלב השני הפכנו את הבעיה ל-Interval prediction. כל דוגמה במודל היא לא שורה אחת רגילה, אלא מרווח זמן:

- t0 - זמן שבו יש מדידת pH ו-EC ידועה.
- t1 - זמן עתידי שאליו רוצים לחזות.

המודל לא שואל "מה הערך עכשיו?" אלא "מה קרה בין t0 ל-t1 שיכול להסביר שינוי בבית השורשים?"

חשוב להבין שגם במודל בית השורשים אנחנו לא חוזים ערך מנותק מהקשר. אנחנו חוזים שינוי מ-t0 ל-t1, כי לבית השורשים יש זיכרון. הערך האחרון שנמדד הוא נקודת התחלה משמעותית, והפעולות מאז הן אלו שמסבירות את הדינמיקה.

איך הארכיטקטורה עובדת בפועל

מבחינת העבודה ההנדסית, בנינו את המערכת כשרשרת מודולרית: קודם שכבת נתונים שמייצרת master dataset אחיד, אחר כך מודל Micro-climate שמייצר את תנאי החממה הפנימיים, ורק לאחר מכן מודל Root-zone שמקבל את תמונת המצב המלאה ומחזיר תחזית pH ו-EC.

היתרון בפיצול הזה הוא שכל שלב פותר בעיה מוגדרת וניתן לבדוק אותו בנפרד. אם תחזית המיקרו-אקלים לא טובה, אפשר לזהות את זה לפני שמאשימים את מודל בית השורשים. ואם מודל בית השורשים טועה, אפשר לבדוק האם מקור הבעיה הוא הקלט האקלימי, אירועי ההשקיה והדישון, או דלילות מדידות pH/EC.

מבחינת זרימת החיזוי, המשתמש מתחיל מנתונים זמינים: weather, radiation, מדידת pH/EC אחרונה ותוכנית ממשק. השלב הראשון מתרגם weather/radiation לתחזית תנאים פנימיים. השלב השני מתרגם את המצב ההתחלתי והתרחיש המתוכנן לשינוי צפוי בבית השורשים.

לכן הארכיטקטורה לא בנויה כמודל אחד גדול שמנסה ללמוד הכול יחד, אלא כמערכת שבה כל שלב מייצג חלק אמיתי בתהליך החקלאי: קודם הסביבה, אחר כך תגובת בית השורשים.

מכאן עברנו לשאלה איזה אלגוריתם מתאים לדאטה כזה.

13:30-16:30 - אלגוריתמים ושיטות

מציג: Oren

להצביע בפוסטר על: Algorithms & Model Choice.

המודל הסופי הוא Unified model שחוזה גם pH וגם EC מאותו סט features. במקום שני pipelines נפרדים לגמרי, רצינו להשתמש באותו תיאור של המצב החקלאי והאקלימי כדי ללמוד את שתי התגובות.

האלגוריתם המרכזי הוא XGBoost. בחרנו בו כי זו בעיית tabular data, עם מספר תוויות לא גדול, והרבה אינטראקציות לא ליניאריות:

- אקלים משפיע אחרת לפי מצב הצמח.
- דישון משפיע אחרת לפי כמות השקיה.
- שינוי EC תלוי גם ב-EC ההתחלתי.
- אותו אירוע יכול להשפיע אחרת לפי הזמן שעבר מאז שהוא התרחש.

Deep Learning לא היה מתאים כאן כבחירה ראשית, כי יש רק 109 תוויות pH ו-EC. מודל עמוק היה עלול ללמוד רעש או לזכור את הדאטה במקום ללמוד תהליך.

בנוסף השתמשנו במודל היברידי: XGBoost כמודל המרכזי, ורכיב Huber / robust linear למצבים חריגים ומוגדרים מראש.

הסיבה היא ההבדל בין אינטרפולציה לאקסטרפולציה. מודלים מבוססי עצים, כמו XGBoost, חזקים מאוד באינטרפולציה - כלומר כאשר הדוגמה החדשה דומה לאזורים שכבר ראינו בדאטה. הם יודעים לחלק את מרחב הפיצ'רים לאזורים וללמוד דפוסים לא ליניאריים מורכבים. אבל כאשר מגיע תרחיש קיצוני או נדיר, למשל קפיצת מלחים אחרי דישון או שינוי חד ב-EC, מודל עצים לא באמת ממשיך קו מעבר למה שראה; הוא נוטה להחזיר ערך לפי העלים והדוגמאות הדומות ביותר.

לכן הוספנו רכיב robust linear שמבוסס על Huber. הוא פשוט יותר, אבל מתאים יותר למצבים שבהם רוצים תגובה יציבה למגמה חריגה בלי להיות רגישים מדי ל-outliers. הוא לא מחליף את XGBoost בכל התחזיות.

בפועל, ברירת המחדל היא XGBoost. הרכיב ה-Robust מופעל רק כאשר תנאי gate מזהים תרחיש כימי רגיש: למשל כמות מלחים/דישון גבוהה, מעט השקיה אחרי אירוע מלח, ריכוז מלח גבוה, EC התחלתי נמוך יחסית, ופער זמן קצר יחסית. כלומר, רוב התחזיות הן אינטרפולציה לא ליניארית של XGBoost, ובמצבי shock מוגדרים אנחנו עוברים לרכיב היציב יותר כדי להתמודד טוב יותר עם תרחישים שנמצאים בקצה ההתפלגות.

עבור EC השתמשנו ב-log-change logic, כי EC הוא ערך חיובי, והיחס בין שינוי קטן לשינוי גדול חשוב יותר מערך מוחלט בלבד. זה עוזר למודל להתייחס לשינויי מליחות בצורה יציבה יותר.

16:30-19:30 - ולידציה: מיקרו-אקלים ו-Root-zone

מציג: Eden

להצביע בפוסטר על: Validation Strategy.

בפרויקט כזה קל מאוד לקבל תוצאה טובה בצורה לא הוגנת אם משתמשים בטעות במידע מהעתיד. לכן שמנו דגש על validation שמדמה deployment.

במודל המיקרו-אקלים

במודל המיקרו-אקלים השתמשנו גם כן ב-Walk-forward. התחלנו מכמות התחלתית של דאטה היסטורי, אימנו את המודל על התקופה הזו, ואז בכל איטרציה ביקשנו ממנו לחזות 3 ימים קדימה מתוך נתוני האקלים החיצוני והקרינה בלבד.

לאחר שהתקופה של 3 הימים הסתיימה, השווינו בין התחזית לבין הערכים האמיתיים שנמדדו בתוך החממה. רק לאחר מכן הוספנו את אותם 3 ימים לסט האימון והמשכנו לחלון הבא.

בצורה הזו בדקנו האם המודל באמת מסוגל לחזות קדימה, ולא רק להתאים את עצמו לדאטה שכבר ראה. זה חשוב כי המיקרו-אקלים הוא השלב שמזין את מודל בית השורשים. אם השלב הראשון לא אמין, גם השלב השני יקבל קלט בעייתי.

במודל Root-zone

במודל Root-zone השתמשנו גם ב-Walk-forward, אבל כאן יחידת הבדיקה היא מדידת pH/EC ולא רצף של 3 ימים.

Walk-forward עובד כמו סימולציה של שימוש אמיתי לאורך זמן. מתחילים מאוסף מדידות עבר ראשוני. בכל צעד מאמנים את המודל רק על מדידות שכבר קרו, ואז מבקשים ממנו לחזות את המדידה הבאה או אינטרוול עתידי מ-t0 ל-t1. רק אחרי שהגענו בזמן למדידה הזו, והיא כבר לא עתידית, מותר להוסיף אותה לסט האימון.

כך אנחנו מונעים Data leakage: אין מצב שבו המודל רואה מדידת pH/EC עתידית בזמן שהוא אמור לחזות אותה.

בנוסף ביצענו מבחני Holdout משניים מול המכון הוולקני. כאן יצרנו אינטרוולים חדשים מתוך המדידות הקיימות: בחרנו זוגות של נקודת עוגן t0 ונקודת יעד t1, בעיקר בפערים ארוכים או מאתגרים יותר, והוצאנו אותם מהאימון. לאחר מכן אימנו את המודל בלי האינטרוולים האלה ובדקנו האם הוא מצליח לחזות אותם כאילו היו תרחישים חדשים.

ההבדל הוא ש-Walk-forward בודק שימוש רציף בזמן, ו-Holdout בודק יכולת להתמודד עם אינטרוולים שנשמרו בצד ולא השתתפו בלמידה. שני המבחנים חשובים כי הם בודקים את המודל מזוויות שונות.

הבסיס להשוואה היה Last-value baseline: להניח שהערך האחרון שנמדד נשאר נכון. זה baseline חשוב כי הוא מייצג את מה שאפשר לעשות בלי מודל. אם המודל לא מנצח אותו, אין הצדקה מעשית לכל המערכת.

19:30-22:30 - תוצאות: Micro-climate ו-Root-zone

מציג: Oren

להצביע בפוסטר על: Micro-Climate Model ו-Root-zone Model Results.

בשלב המיקרו-אקלים קיבלנו תוצאות טובות עבור המשתנים הפנימיים המרכזיים:

- טמפרטורה פנימית: MAE בערך 0.43 מעלות, R2 בערך 0.97.
- ET0: MAE בערך 0.0035, R2 בערך 0.97.
- קרינה פנימית: MAE בערך 22.5 W/m2, R2 בערך 0.97.
- טמפרטורת קרקע: MAE בערך 0.86 מעלות, R2 בערך 0.92.

המשמעות היא שהשלב הראשון הצליח לייצר קלט פנימי סביר למודל בית השורשים, במקום להסתמך רק על מדידות פנימיות עתידיות שלא תמיד קיימות.

במודל הסופי של Root-zone:

- עבור pH קיבלנו Walk-forward MAE של 0.330, R2 של 0.932, ושיפור של 62.4% לעומת Last-value baseline.
- עבור EC קיבלנו Walk-forward MAE של 0.127 mS/cm, R2 של 0.982, ושיפור של 35.3% לעומת Last-value baseline.
- במבחני Holdout משניים קיבלנו pH MAE של 0.288 ו-EC MAE של 0.131.

איך מפרשים את זה?

MAE אומר בכמה יחידות בממוצע טעינו. עבור pH זו טעות ביחידות pH. עבור EC זו טעות ב-mS/cm.

R2 אומר כמה מהשונות המודל מצליח להסביר. אבל בגלל שהדאטה קטן, אנחנו לא מסתמכים רק על R2. לכן ההשוואה ל-baseline חשובה מאוד.

22:30-24:30 - מסקנות, הישגים וצעד הבא

מציג: Eden

להצביע בפוסטר על: Conclusions.

המסקנה המרכזית היא שאפשר לבנות Soft Sensor שימושי לבית השורשים גם כאשר מדידות pH ו-EC דלילות, בתנאי שמשלבים אותן עם דאטה תפעולי ואקלימי צפוף.

ההישגים המרכזיים:

1. בנינו בסיס נתונים אחיד ממקורות שונים.
2. בנינו שכבת Micro-climate שמאפשרת להשתמש במודל גם כתחזית.
3. בנינו מודל Root-zone שמנבא pH ו-EC קדימה ומשפר את הדיוק מול baseline.
4. בנינו דשבורד שמתרגם את המודל ל-workflow שימושי.
5. ביצענו validation שמדמה שימוש אמיתי ומצמצם סיכון ל-Data leakage.

חשוב להדגיש: לא ניסינו להחליף את המגדל ולא לבנות מערכת בקרה אוטונומית. המטרה היא Decision Support - כלי שנותן תחזית ותובנה, והמגדל נשאר מקבל ההחלטות.

הצעד הבא בתהליך הוא לא רק לחזות מה יקרה, אלא להמליץ מה לעשות. כיום צוותים נוספים מהמכון הוולקני מתמקדים בכיוון של מערכת אופטימיזציה: מערכת שתשתמש במודל בית השורשים שלנו כדי להמליץ כמה מים ודשן לתת ומתי, כך שהתחזית תהפוך לכלי תכנון פעולות.

כלומר, המודל שלנו יכול להיות שכבת החיזוי שעליה בנויה שכבת החלטה.

24:30-26:30 - מגבלות ואתגרים

מציג: Eden או Oren

להצביע בפוסטר על: Challenges & Limitations.

האתגר הראשון היה דלילות התוויות. יש הרבה נתוני חיישנים ותפעול, אבל רק 109 מדידות pH ו-EC. התמודדנו עם זה באמצעות ניסוח של בעיית interval prediction, feature engineering עשיר, והשוואה ל-baseline.

האתגר השני היה סנכרון מקורות מידע. כל מקור הגיע בפורמט ובתדירות אחרת. התמודדנו עם זה באמצעות master dataset אחיד של 10 דקות.

האתגר השלישי היה חיזוי EC. EC רגיש מאוד לאירועי השקיה, דישון ומלחים, ובערכים נמוכים גם שינוי קטן יכול להיות משמעותי יחסית. לכן השתמשנו ב-log-change ובפיצ'רים שמייצגים הצטברות מלחים והיסטוריה.

האתגר הרביעי היה להימנע מ-Data leakage. בגלל שהדאטה הוא סדרת זמן, split רגיל יכול להטעות. לכן השתמשנו ב-Walk-forward validation.

המגבלות:

- המודל נבדק על עונה אחת וחממה אחת.
- מספר תוויות pH ו-EC עדיין קטן.
- צריך לבדוק Generalization על עוד חממות, עונות וגידולים.
- הדשבורד הוא prototype תומך החלטה ולא מערכת Production מלאה.

26:30-30:00 - דמו של הדשבורד

מציג: Oren

מטרת הדמו: להראות איך המודל הופך מכלי מחקרי ל-workflow שימושי.

זרימת הדמו:

1. מעלים קובצי Weather/Radiation.
2. המערכת יוצרת Micro-climate forecast.
3. מזינים נתוני מצב נוכחי: pH, EC, זמן עוגן, מצב צמח ואירועי השקיה/דישון.
4. בוחרים זמן יעד.
5. מריצים חיזוי ומקבלים pH ו-EC צפויים.

שאלות שופטים מעודכנות ותשובות קצרות

שאלות רקע חקלאי

מה זה Root-zone?

Root-zone הוא אזור בית השורשים - הסביבה שבה השורשים קולטים מים, דשן ומינרלים. בחממה במצע מנותק האזור הזה רגיש מאוד, ולכן שינוי ב-pH או EC יכול להשפיע מהר על הצמח.

מה זה pH ולמה הוא חשוב?

pH מודד חומציות או בסיסיות. הוא חשוב כי הוא משפיע על זמינות יסודות ההזנה. גם אם הדשן קיים, pH לא מתאים יכול לגרום לכך שהצמח לא יקלוט חלק מהיסודות בצורה טובה.

מה זה EC ולמה הוא חשוב?

EC הוא Electrical Conductivity, מדד לכמות המלחים המומסים. EC נמוך מדי יכול להעיד על חוסר הזנה, ו-EC גבוה מדי יכול להעיד על הצטברות מלחים וסטרס אוסמוטי שמקשה על קליטת מים.

מה הבעיה המרכזית בחממה?

המדידות האמינות של pH ו-EC נעשות ידנית ובדלילות. בין מדידות יכולים להתרחש השקיה, דישון, שינוי אקלים או הצטברות מלחים, והמגדל לא תמיד רואה את זה בזמן.

שאלות מטרת הפרויקט

מה התוצר של הפרויקט?

Soft Sensor ודשבורד תומך החלטה שמקבל מדידה אחרונה, אקלים, השקיה, דישון ומצב צמח, ומחזיר תחזית pH ו-EC לזמן יעד עתידי.

למה זה Decision Support ולא בקרה אוטונומית?

כי המודל נותן תחזית ותובנה, אבל ההחלטה החקלאית נשארת אצל המגדל. לפני בקרה אוטונומית צריך validation על עוד עונות, חממות וגידולים.

מה הערך היישומי?

מעבר מניהול תגובתי לניהול פרואקטיבי: לזהות מראש סיכון לסטייה ב-pH או EC ולשנות השקיה/דישון לפני שהבעיה מופיעה במדידה הידנית הבאה.

שאלות דאטה

אילו סוגי נתונים היו לכם?

מזג אוויר וקרינה חיצוניים, מיקרו-אקלים בתוך החממה, השקיה, דישון, חומצה ומלחים, כיסוי נוף, גיל הצמח, ומדידות ידניות של pH ו-EC.

מה סדר הגודל של הדאטה?

16,693 שורות ברזולוציה של 10 דקות, 20 עמודות, מטווח 29.05.2025 עד 21.09.2025. יש 109 מדידות ידניות של pH ו-EC.

למה 109 תוויות זה אתגר?

כי יש מעט נקודות אמת ללמידה ולבדיקה. לכן השתמשנו ב-feature engineering, ניסוח של interval prediction, ו-validation שמרני.

שאלות ארכיטקטורה

למה פיצלתם לשני שלבים?

כי בזמן תחזית עתידי לא תמיד יש מדידות פנימיות של החממה. השלב הראשון חוזה Micro-climate פנימי מנתוני weather/radiation, והשלב השני משתמש בזה לחיזוי pH ו-EC.

למה לחזות שינוי מ-t0 ל-t1 ולא ערך מוחלט?

כי לבית השורשים יש זיכרון. הערך האחרון שנמדד הוא נקודת התחלה חשובה, והפעולות מאז מסבירות את השינוי הצפוי.

איזה Feature engineering היה חשוב?

ערכי עוגן pH/EC, gap hours, סכומי השקיה ודישון, ET0, קרינה, טמפרטורה, טמפרטורת קרקע, הצטברות מלחים, זמן מאז דישון, היסטוריה לפני t0, canopy cover וגיל הצמח.

שאלות מודל

למה XGBoost?

כי זו בעיית tabular data עם מעט תוויות יחסית והרבה אינטראקציות לא ליניאריות. XGBoost מתאים יותר מ-Deep Learning במקרה כזה.

למה לא Deep Learning?

Deep Learning דורש הרבה יותר תוויות. עם 109 מדידות pH/EC היה סיכון גבוה ל-overfitting.

למה EC קשה יותר?

EC רגיש מאוד לאירועי השקיה, דישון ומלחים. בנוסף, בערכים נמוכים גם שינוי אבסולוטי קטן יכול להיות משמעותי יחסית.

שאלות Validation

מה היה ה-validation המרכזי?

Walk-forward validation. בכל נקודה מאמנים רק על העבר ומנבאים עתיד, כך שאין שימוש במידע עתידי.

למה לא split רגיל?

כי זה דאטה בזמן. split אקראי יכול לערבב עבר ועתיד ולתת תוצאה אופטימית מדי.

מה זה Holdout אצלכם?

מבחן משני שבו יצרנו אינטרוולים שלא שימשו באימון ובדקנו עליהם את המודל. חלק מהמבחנים האלה נעשו לפי בקשות בדיקה נוספות של המכון הוולקני.

למה Baseline של last value?

כי הוא מייצג את הפתרון הפשוט ביותר: להניח שהמדידה האחרונה עדיין נכונה. המודל צריך לנצח את זה כדי להיות שימושי.

שאלות תוצאות

מה תוצאות המיקרו-אקלים?

בערך: טמפרטורה פנימית MAE 0.43°C ו-R2 0.97, ET0 MAE 0.0035 ו-R2 0.97, קרינה פנימית MAE 22.5W/m2 ו-R2 0.97, טמפרטורת קרקע MAE 0.86°C ו-R2 0.92.

מה תוצאות Root-zone?

במודל הסופי: pH MAE 0.330, R2 0.932, שיפור 62.4% מול baseline. EC MAE 0.127mS/cm, R2 0.982, שיפור 35.3%. במבחני Holdout: pH MAE 0.288, EC MAE 0.131.

למה לא להסתמך רק על R2?

כי בדאטה קטן R2 יכול להיות מטעה. MAE וההשוואה ל-baseline נותנים משמעות מעשית יותר.

שאלות המשך, מגבלות ו-XAI

מה הצעד הבא?

להשתמש במודל החיזוי כשכבה בתוך מערכת אופטימיזציה שממליצה כמה מים ודשן לתת ומתי. זה הכיוון שצוותים נוספים במכון הוולקני מתמקדים בו.

מה המגבלה המרכזית?

Generalization. המודל נבדק על עונה וחממה מוגבלות, עם 109 תוויות. צריך לבדוק על עוד עונות, חממות וגידולים.

האם יש XAI?

כרגע אפשר להסביר לפי Feature importance ולפי הפיצ'רים שהוזנו בתרחיש. XAI פר-תחזית מלא, למשל SHAP לכל חיזוי, הוא תוספת טובה לשלב הבא.

איך תסבירו למה המודל נתן תחזית מסוימת בדמו?

נבדוק אילו פיצ'רים בתרחיש יכולים לדחוף את התחזית: pH/EC התחלתי, השקיה, דישון, זמן מאז אירועים, ET0, קרינה, טמפרטורת קרקע והצטברות מלחים. חשוב לא לטעון להסבר SHAP מלא אם לא מציגים אותו.

Runbook מעודכן לדמו

מטרת הדמו: להראות איך המשתמש מעלה נתונים, מזין מצב חממה, ומקבל תחזית pH ו-EC דרך הדשבורד.

זרימת דמו מומלצת

1. לפתוח את הדשבורד המקומי.
2. להראות העלאת קובצי Weather/Radiation.
3. להסביר שהמערכת יוצרת Micro-climate forecast.
4. להראות הזנת Anchor: pH, EC וזמן מדידה אחרון.
5. להראות הזנת אירועי השקיה/דישון ומצב צמח.
6. לבחור Target time.
7. להריץ Prediction.
8. להציג predicted pH ו-predicted EC.

אם הדמו לא עובד

אם הדמו לא מגיב תוך 20 שניות:

כדי לשמור על זמן השיפוט, אעבור לצילומי המסך של אותה זרימה. המערכת עצמה ארוזה להרצה, והדמו נועד להמחיש את workflow המשתמש.

דף סיכום מעודכן להדפסה - Rootzone Prediction

One-liner

בנינו Soft Sensor שמחזֶה pH ו-EC עתידיים בבית השורשים מתוך מדידה ידנית אחרונה, אקלים, השקיה, דישון ומצב הצמח, כדי לתמוך בהחלטות השקיה ודישון לפני שמתרחשת חריגה.

רקע

Root-zone הוא אזור בית השורשים, שבו הצמח קולט מים ודשן.

pH מודד חומציות/בסיסיות ומשפיע על זמינות יסודות ההזנה.

EC מודד מוליכות חשמלית ומייצג את ריכוז המלחים בתמיסה.

כאשר pH או EC יוצאים מאיזון, הצמח עלול לסבול מחוסר זמינות דשן, הצטברות מלחים, סטרס מים ופגיעה ביבול.

הבעיה

המדידות האמינות של pH ו-EC נעשות ידנית ובדלילות. בין מדידות יכולים להתרחש השקיה, דישון ושינויי אקלים, ולכן המגדל עלול לזהות חריגה מאוחר מדי.

התוצר

מערכת תומכת החלטה:

- מקבלת מדידת pH/EC אחרונה.
- מקבלת נתוני אקלים, השקיה, דישון ומצב צמח.
- חוזה pH ו-EC בזמן יעד עתידי.
- מאפשרת לבדוק תרחיש לפני ביצוע פעולה בחממה.

הדאטה

- 16,693 שורות.
- רזולוציה של 10 דקות.
- 20 עמודות.
- 109 מדידות pH/EC.
- מקורות: Weather, radiation, micro-climate, irrigation, fertilization, crop state, manual pH/EC.

השיטה

Pipeline דו-שלבי:

1. Micro-climate model - חיזוי משתני חממה פנימיים מתוך מזג אוויר וקרינה חיצוניים.
2. Root-zone model - חיזוי שינוי pH ו-EC מ-t0 ל-t1.

ארכיטקטורה

המערכת בנויה כשרשרת מודולרית: master dataset אחיד, מודל Micro-climate לתנאי חממה פנימיים, ואז מודל Root-zone שחוזה שינוי pH/EC מ-t0 ל-t1. הפיצול מאפשר לבדוק כל שלב בנפרד ומקרב את השימוש לתרחיש אמיתי של תחזית.

תוצאות

Micro-climate:

- טמפרטורה פנימית: MAE 0.43°C, R2 0.97.
- ET0: MAE 0.0035, R2 0.97.
- קרינה פנימית: MAE 22.5W/m2, R2 0.97.
- טמפרטורת קרקע: MAE 0.86°C, R2 0.92.

Root-zone:

- pH: MAE 0.330, R2 0.932, שיפור 62.4% מול Last-value baseline.
- EC: MAE 0.127mS/cm, R2 0.982, שיפור 35.3%.
- Holdout: pH MAE 0.288, EC MAE 0.131.

מסקנות

הפרויקט מראה שאפשר לשלב מדידות ידניות דלילות עם דאטה תפעולי ואקלימי צפוף כדי לבנות Soft Sensor שימושי לבית השורשים.

הצעד הבא הוא להשתמש במודל כשכבת חיזוי בתוך מערכת אופטימיזציה שממליצה כמה מים ודשן לתת ומתי.

מגבלות

- עונה אחת וחממה אחת.
- 109 תוויות pH/EC בלבד.
- EC רגיש לאירועי השקיה, דישון ומלחים.
- נדרש validation על עוד עונות, גידולים וחממות.