Automatizarea climatului unei sere

PROIECT DE SEMESTRU

Student: **Lavinia-Maria ABRUDAN**

Disciplina: **Sisteme bazate pe cunoaștere**

Cuprins

[1 Introducere 2](#_Toc181667062)

[1.1 Context general 2](#_Toc181667063)

[1.2 Obiective 3](#_Toc181667064)

[1.3 Specificații 3](#_Toc181667065)

[2 Cunoașterea și analiza setului de date 5](#_Toc181667066)

[2.1 Mediul de implementare 5](#_Toc181667063)

[2.2 Citirea setului de date 5](#_Toc181667064)

[2.3 Analiza de date 5](#_Toc181667065)

2.4  [Statistici descriptive 7](#_Toc181667063)

[2.5 Analize vizuale 7](#_Toc181667064)

[2.6 Analiza de corelatie 9](#_Toc181667065)

[3 Pre-procesarea setului de date 10](#_Toc181667067)

[3.1 Eliminarea zgomotului 10](#_Toc181667063)

[3.2 Reducerea dimensiunii 11](#_Toc181667064)

[3.3 Eliminarea tenintelor 11](#_Toc181667065)

[3.3.1 Prin metoda medie mobila 12](#_Toc181667077)

[3.3.2 Folosind polinomul 12](#_Toc181667077)

[3.4 Interpolarea esantioanelor lipsa 13](#_Toc181667065)

[3.5 Eliminarea valorilor aberante 14](#_Toc181667065)

[4 Modelarea sistemului 16](#_Toc181667068)

[4.1 Tehnici de modelare 16](#_Toc181667063)

[4.2 Descriere procedurilor 16](#_Toc181667064)

[4.3 Implentare 17](#_Toc181667063)

[4.4 Evaluarea modelului 22](#_Toc181667063)

[5 Concluzii 24](#_Toc181667069)

[5.1 Rezultate obținute 24](#_Toc181667070)

[5.2 Direcții de dezvoltare 24](#_Toc181667071)

[6 Bibliografie 25](#_Toc181667072)

[7 Reguli de formatare 26](#_Toc181667073)

[7.1 Formatarea paginii 26](#_Toc181667074)

[7.2 Titluri și stiluri 26](#_Toc181667075)

[7.3 Figuri, tabele și ecuații 27](#_Toc181667076)

[7.3.1 Figuri 27](#_Toc181667077)

[7.4 Tabele 27](#_Toc181667078)

[7.5 Ecuații 27](#_Toc181667079)

[7.6 Referințe bibliografice 28](#_Toc181667080)

# Introducere

## Context general

Proiectul are in vedere realizarea unei sere inteligente in interiorul careia climatul sa fie automatizat. Variatiile de umiditate, iluminat si de temperatura, de asemenea si alti factori, pot afecta negativ sanatatea si dezvoltarea plantelor, dar si productivitatea. Prin urmare voi analiza setul de date specificat, iar pe baza acestuia voi crea o medie a datelor pentru optimizarea acestora ce urmeaza sa fie utilizate in tehnici de machine learning.

Controlul climatului intr-o sera este important datorita asigurarii unui mediu optim de crestere a plantelor. Sistemul de automatizare permite monitorizarea, controlul si reglarea conditiilor in timp real, utilizand senzorii specifici si actuatoare care vor fi integrati in partea fizica, actionand automat asupra factorilor de mediu.

Climatizarea corecta intr-o sera asigura un mediu optim pentru cresterea plantelor. De exemplu, trebuie analizata temperatura din interiorul acesteia, pentru a deschide un „geam” in momentul in care temperatura este prea ridicata, iar pornirea unui ventilator care sa mentina o temperatura constanta si sa reduca temperatura foarte ridicata.

Umiditatea solului este de asemenea importanta, deoarece pe baza acesteia se va porni pompa de apa atunci cand solul devine prea uscat. Iar aceasta se va opri atunci când umiditatea este suficient de potrivita.

Aceasta lucrare presupune dezvoltarea unui sistem automat de control al climatului intr-o sera, bazat pe senzori si actuatoare, pentru a monitoriza si regla conditiile de mediu potrivite precum temperatura, umiditatea si luminozitatea.

Agricultura moderna se confrunta cu diverse dificultati, printre care cresterea eficientie utilizarii resurselor naturale, reducerea impactului asupra mediului si asigurarea unei productivitati constante. Contexul general al acestei aplicatii este in legatura cu agricultura moderna care imbina cunostinte din electronica, informatica, biologie si agronomie. Cu cresterea utilizarii tehnologiilor IoT si a sistemelor autonome, aceasta lucrare contribuie la dezvoltarea unor solutii moderne pentru agricultura.

Senzorii sunt dispozitive care masoara conditiile de mediu si transmit aceste date. Actuatoarele sunt dispozitive care intervin asupra mediului pentru a modifica parametrii in functie de conditiile necesare. Tehnologia IoT se refera la interconectarea senzorilor si actuatoarelor prin intermediul internetului pentru monitorizare si controlul sistemului de la distanta.

Proiectul presupune o importanta in contextul agriculturii moderne, avand ca scop principal optimizarea utilizarii resurselor naturale si cresterea productivitati. Mini-serele inteligente pot fi utilizate in diverse cazuri sau situatii, de la productii agricole pana la cercetari siintifice.

## Obiective

Scopul principal al proiectului este dezvoltarea unei solutii eficiente pentru automatizarea conditiilor de mediu dintr-o mini-sera. Obiectivele specifice includ:

1. Crearea unui sistem capabil sa monitorizeze si sa regleze automat temperatura, umiditatea si lumina din mini-sera pe baza datelor colectate de la senzori.
2. Analiza conditiilor de mediu si identificarea parametrilor cei mai buni pentru cresterea plantelor, utilizand datele furnizate pentru o reglare precisa.
3. Dezvoltarea unui sistem bazat pe implementarea unui algoritm bazat pe conditii logice, pentru controlul echipamentelor din mini-sera precum ventilatorul, pompa de apa si sursa de lumina.
4. Asigurarea unui sistem eficient care raspunde in timp real la schimbarile de mediu.
5. Activarea proceselor doar atunci cand este necesar.
6. Analiza eficientei sistemului in vederea mentinerii parametrilor in intervalele optime.

Aceste obiective contribuie la crearea unui sistem inovativ si scalabil, care poate fi adaptat in functie de nevoile specifice din domeniul agriculturii.

## Specificații.

Proiectul presupune implementarea unui sistem care include urmatoarele functionalitati:

1. Monitorizarea conditiilor de mediu: valorile furnizate de senzori le voi procesa si utiliza pentru functionearea corecta in timp real a sistemului.
2. Reglarea automata a climatului din mini-sera: pornirea si oprirea automata a ventilatorului pentru mentinerea temperaturii optime. Activarea sursei de lumina in conditii de luminozitate scazuta. Controlul irigarii in functie de nivelul de umiditate a solului.
3. Utilizarea senzorilor necesari:

* Pentru temperatura si umiditate
* Pentru masurarea intensitatii luminii si ajustarea sursei de lumina artificiala (de tip bec sau leduri)
* Pentru determinarea nivelului apei
* Pentru analiza nivelului de nutrienti din sol, senzor de NPK – nutrienti.

1. Actuatori:

* Ventilator pentru mentinearea si controlul temperaturii
* Pompa de irigare pentru gestionarea umiditatii solului
* Sursa de lumina pentru a suplimenta lumina naturala

1. Platforma de control:

* Utilizarea unui microcontroller precum, Raspberry Pi, pentru colectarea datelor de la senzori si controlul actuatorilor.

1. Interfata utilizator:

* Dezvoltarea unei aplicatii mobile pentru monitorizare si controlul de la distanta.

# Cunoașterea și analiza setului de date

**2.1 Detalii despre mediul de implementare**

Mediul de implementare pentru analiza datelor este platforma Jupyter Notebook prin intermediul caruia pot sa citesc datele din dataset-ul furnizat, pot sa observ valorile lipsa, forma datelor, informatii si descrieri despre acestea.

Dupa ce datele sunt analizate prin intermediul principiului EDA, setul de date este pre-procesat, aceste procese realizandu-se tot prin intermediul platformei Jupyter. Pentru partea de modelare a sistemului avem nevoie de un sistem bazat pe IoT cum este Rasperry Pi, prin intermediul caruia sa putem crea sistemul propriu zis si implementa partea de cod disponibila pentru functionarea fizica a sistemului creat in urma analizei datelor si pre-procesarea lor.

**2.2 Citirea setului de date**

*Tabelul 2.2.1. Setul de datele*

O imagine care conține text, captură de ecran, număr, linie

Descriere generată automat

**2.3 Analiza de date**

*Tabelul 2.3.1. Numărul de caracteristici si de observații*

O imagine care conține text, captură de ecran, Font, linie

Descriere generată automat

*Tabelul 2.3.2. Tipul de date si cantitatea maxima, validitatea datelor, memoria utilizata*

O imagine care conține text, captură de ecran, număr, Font

Descriere generată automat

O imagine care conține text, captură de ecran, număr, Font

Descriere generată automat**O imagine care conține text, captură de ecran, Font, număr

Descriere generată automat***Tabel 2.3.6. Lista coloanelor Tabelul 2.3.7. Lista cu suma valorilor lipsa din fiecare coloana*

O imagine care conține text, captură de ecran, Font, număr

Descriere generată automat*Tabelul 2.3.8. Valorile unice din fiecare coloana*

**2.4 Statistici descriptive**

Setul de date include urmatoarele statistici descriptive pentru principalele variabile:

1. Temperatura

* Valoarea medie: 18,76
* Valoarea minima: 3
* Valoarea maxima: 41
* Abaterea standard: 7,14

1. Umiditatea

* Valoarea medie: 50,32
* Valoarea minima: 0
* Valoarea maxima: 100
* Abatere standard: 22,99

1. Nivelul apei

* Valoarea medie: 79.40
* Valoarea minima: 0
* Valoarea maxima: 100
* Abatere standard: 36.82

1. Nutrientii solului (NPK)

* Azot: valoarea medie: 231,96 , intervalul:[0, 255]
* Fosfor: valoarea medie: 234,12 , intervalul:[0, 255]
* Potasiu: valoarea medie: 235,17, intervalul: [0,255]

**2.5 Analize vizuale**

sns.set\_style("darkgrid")

numerical\_columns = df.select\_dtypes(include=["int64", "float64"]).columns

plt.figure(figsize=(14, len(numerical\_columns)\*3))

for idx, feature in enumerate(numerical\_columns, 1):

plt.subplot(len(numerical\_columns), 2, idx)

sns.histplot(df[feature], kde=True)

plt.title(f"{feature} | Skewness: {round(df[feature].skew(), 2)}")

plt.tight\_layout()

plt.show()

In urma analizarii graficului, distributia temperaturii este aproape normala cu o usoara asimetrie pozitiva, majoritatea valoriilor se concentreaza intre 15 si 30 grade Celsius cu cateva valori extreme 35-40. Distributia umiditatii indica o asimetrie negativa pronuntata de -1.15, indicand o acumulare mare de valori aproape de maxim 60-80%, indicand faptul ca umiditatea este mentinuta ridicata constant. Nivelul apei arata ca sistemul mentine un nivel optim de apa, iar valorile scazute sunt rare.

*O imagine care conține diagramă, Interval, captură de ecran, text

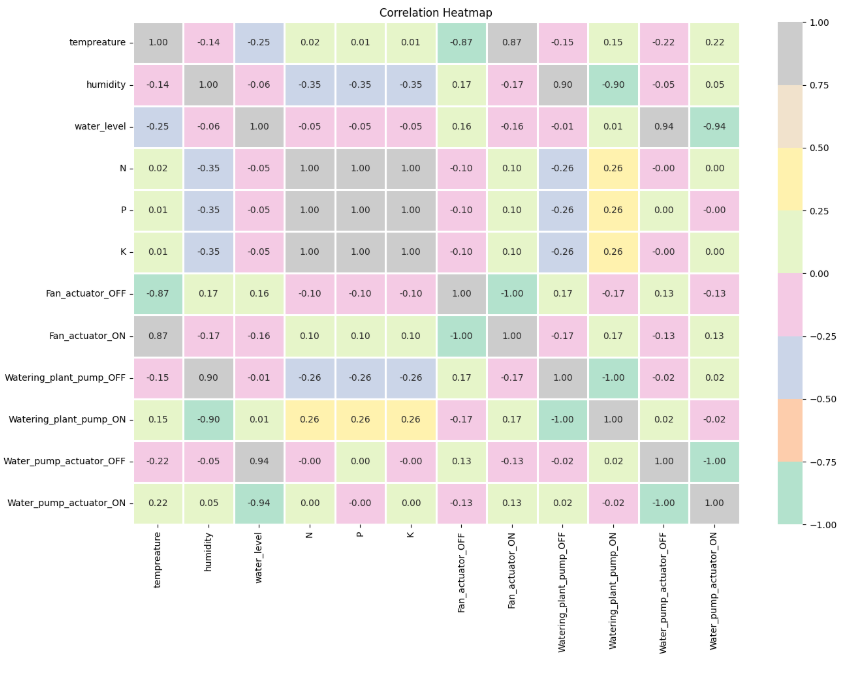
Descriere generată automatO imagine care conține Interval, diagramă, linie, captură de ecran

Descriere generată automatO imagine care conține text, captură de ecran, diagramă, Interval

Descriere generată automatFigure 2.5.1 Kernel Density Plots*

Nutrientii N,P,K au o asimetrie usor negativa, indicat de skewness, care prezinta o acumulare moderata de valori in partea superioara a intervalului [-0.84, -0.87]. Actuactorii (ventilator si pompa de apa) indica o distributie puternic asimetric pozitiva ceea ce inseamna ca ventilatorul este activ doar la anumite temperaturii ridicate si activarea pompelor in momente de necesitate.

**2.6 Analiza de corelație**

*Figure 2.6.1 Analiza de corelatie*

Actuatorii (ventilatoare si pompa de apa) sunt sincronizati in mod logic, deoarece autocorelatie ridicata intre ele. Nutrientii din sol au relatii de corelatii pozitive intre ei. Temperatura este in stransa corelatie cu actuatorul ventilatorului pornit, umiditatea este in relatie de corelatie pozitiva cu pompa de apa oprita, nivelul de apa se afla in corelatie cu actuatorul pompei de apa oprit. Valorile ridicate ale nutrientilor din sol pot influenta direct decizile actuatorilor, evidentiind o posibila interdependenta intre nevoile solului si compozitia acestuia si activarea sistemului de irigare.

Prin urmare, analiza de corelatie dintre parametrii senzorilor si activarea actuatorilor indica un model de functionare a sistemului, permitand anumite ajustari, pentru a optimiza modelul ce urmeaza a fi creat.

# Pre-procesarea setului de date

Pre-procesarea datelor este o etapa importanta pentru asigurarea calitati si relevantei setului de date furnizat si utilizat. In capitolul 2 am analizat valorile lipsa, valorile duplicate, iar in pre-procesare vom reduce zgomotul, eliminam valorile aberante, reducerea dimensionalitatii si filtrarea zgomotului.

## Eliminarea zgomotului

Am realizat eliminearea zgomotului din datele care furnizeaza temperatura pentru a asigura acuratetea modelului final. Zgomotul poate influenta negativ analiza datelor si performanta algoritmilor de predictie. Am aplicat mai multe tehnici de filtrare, evaluand eficienta fiecareia si am identificat ca filtrul median este cel mai potrivit pentru acest set de date.

*O imagine care conține text, captură de ecran, diagramă, afișaj

Descriere generată automatO imagine care conține text, captură de ecran, diagramă, linie

Descriere generată automat*Filtru median are performante superioare in netezirea semnalului fara a afecta caracteristicile esentiale ale datelor. Acest proces l-am efectuat pentru toate tipurile de date pentru a crea un set de date curat si relevant pentru prelucare.

O imagine care conține text, diagramă, linie, Paralel

Descriere generată automat *Figura 3.1.1.Semnalul original* *Figura 3.1.2 Aplicarea filtrului*

*Figura 3.1.3 Toate tipurile di filtre aplicate*

**3.2 Dimensionality reduction**

PCA este metoda pe care am utilizat-o pentru reducerea dimensionalitatii datelor, astfel incat sa se pastreze cat mai multa variabilitate a acestora. PCA reduce datele numerice din seul de date pentru a simplifica procesarea datelor, pentru vizualizarea relatiilor dintre punctele de date si pentru a inlatura colinearitatea si valorile redudante. Prima data am eliminat valorile lipsa pentru a nu influenta procesarea urmatoare.

**O imagine care conține captură de ecran, text, Interval, diagramă

Descriere generată automat**

*Figura 3.2.1 Reducerea dimensiunii*

Componenta 1 reprezinta directia cu cea mai mare variabilitate a datelor, alaturi de componenta 2 care de asemenea captureaza a doua cea mai mare variabilitate. Distributia datelor arata ca acestea sunt in grupuri disticte, sugerand o posibila relatie intre variabilele numerice initiale. Punctele dispersate pot indica valori extreme.

## Eliminarea tendintelor

Eliminarea tendintelor este importanta pentru a extrage semnalul relevant pentru analiza, fara a fi influentat de variatii lente si nesemnificative care pot controla si distorsiona rezultatele modelelor.

**O imagine care conține text, captură de ecran, Interval, linie

Descriere generată automat** *Figure 3.3.1 Datele din coloana umiditate*

**3.3.1 Eliminarea tendintei folosind media mobila**

Aceasta metoda am utilizat-o pentru a identifica si elimina tendintele generale din semnal, cum este pentru coloana indicata de umiditate prin cresterea sau scaderea lenta a valorilor care nu reflecta variatii rapide.

*O imagine care conține text, captură de ecran, Interval, Color

Descriere generată automat*

*Figura 3.3.1.1 Scaderea mediei mobile*

**3.3.2 Eliminarea tendintei folosind polinoame**

Si aceasta metoda am folosit-o pentru aceeasi coloana, ce indica umiditatea, si reprezinta o metoda mai avansata pentru eliminarea tendintelor, care modeleaza si elimina variatiile sistematice prin utiliarea unei functii polinomiale.

Aceasta metoda este mai precisa decat media mobila atunci cand tendinta generala a semnalului are o forma curbata sau care nu este liniara.

polynomial\_coeffs = np.polyfit(time, signal, 3)

polynomial\_trend = np.polyval(polynomial\_coeffs, time)

detrended\_polynomial = signal - polynomial\_trend

plt.figure(figsize=(10, 6))

plt.plot(time, signal, label='Original')

plt,plot(time, detrended\_polynomial, label=’Polynomial Detrended (Cubic)’)

plt.title("Polynomial Detrended Signal")

O imagine care conține text, captură de ecran, Interval

Descriere generată automatplt.legend() plt.show()

*Figura 3.3.2.1 Semnal de tendinta polinomial*

**3.4 Interpolarea esantioanelor lipsa**

Exista valori lipsa care trebuie tratate inainte de a realiza modelarea ulterioara. Am realizat interpolarea prin metoda cubica spline utilizata pentru a completa valorile lipsa din coloana ce indica nivelul apei. Prin intermediul metodei creez o functie continua care ajusteaza curbele in functie de punctele cunoscute, reducand astfel consecintele valorilor lipsa.

**O imagine care conține text, captură de ecran, Dreptunghi, linie

Descriere generată automat***Figura 3.4.1 Valori lipsa din coloana water\_level*

Graficul arata distributia valorilor lipsa, indicand golurile ce trebuie completate.

Metoda spline cubic ajusteaza o functie polinomiala de ordin 3 intre punctele de date cunoscute, astfel se genereaza o interpolare care respecta tendintele normale ale semnalului.

**O imagine care conține text, captură de ecran, Interval, linie

Descriere generată automat**

*Figura 3.4.2 Interpolare cubic spline*

Interpretarea rezultatelor: interpolarea reface semnalul pastrand o continuitate si caracteristicile firesti ale datelor originale. Aceasta metoda este potrivita pentru coloana water\_level din setul de date, deoarece nu introduce schimbari bruste care ar putea fi nereale.

In continuare voi utiliza diferite metode de interpolare pentru toate coloane din setul de date, astfel incat sa fie potrivit si pregatit pentru modelarea sistemului.

**3.5 Eliminarea valorilor aberante**

Aceste valori sunt niste puncte de date care difera destul de mult fata de celelalte valori din setul de date. Acestea afecteaza negativ analiza si modelarea ce urmeaza, de aceea este important sa fie detectate si eliminate. Voi lucra pe toate coloanele de date, utilizand algoritmul DBSCAN.

**O imagine care conține text, captură de ecran

Descriere generată automatO imagine care conține text, captură de ecran, linie, Paralel

Descriere generată automat**

**O imagine care conține text, Interval, captură de ecran, linie

Descriere generată automatO imagine care conține text, captură de ecran, Interval, linie

Descriere generată automat** *Figura 3.5.1 Pentru nivelul apei Figure 3.5.2 Pentru umiditate*

**O imagine care conține text, captură de ecran, Interval, linie

Descriere generată automatO imagine care conține text, captură de ecran, Interval, linie

Descriere generată automat***Figura 3.5.3 Pentru temperatura Figura 3.5.4 Pentru nutrientul N*

*Figura 3.5.5 Pentru nutrientul P Figura 3.5.6 Pentru nutrientul K*

Din graficul initial se observa multe valori care trebuie eliminate cele colorate cu gri pentru ca nu urmaresc trendul normal. Dupa aplicarea metodei DBSCAN se pastreaza doar punctele relevante, adica datele nuantate cu albastru. Astfel se curata de valori aberante setul de date, aplicand aceasta metoda pentru toate coloanele.

# Modelarea sistemului

O imagine care conține text, desen animat, clădire, proiectare

Descriere generată automat

*Figura 4.1 IoT Smart Greenhouse*

**4.1 Tehnici de modelare utilizate:**

* Algoritmi de clasificare pe care ii voi utiliza sunt Random Forest si Decision Tree, pentru a da detalii legate de comportarea actuatorilor in etapa de predictie.
* Algoritmul de regresie pe care il utilizez este Linear Regression pentru a analiza relatia dintre datele importante ce stabilesc starea actuatorilor, precum temperatura, umiditatea si nivelul apei din pompa de apa.

**4.2 Descrierea procedurilor:**

* Preprocesarea datelor efectuata in etapa de analiza a datelor.
* Vizualizarea initiala a datelor si procesarea lor pentru a intelege variabilele relevante.
* Impartirea setului de date in seturi de identificare si testare.
* Construirea unui model pentru a testa setul de date pentru a evalua performanta.
  1. **Implementare**
* Detalii tehnice: mediul de implementare Python, platforma Rasberry Pi si PyCharm, modul de utilizare fiind o aplicatie mobila sau o interfata web pentru a seta de la distanta sau a monitoriza starea si conditiile mediului din interiorul mini-serei.
* Controlabilitatea aplicatiei: senzorii colecteaza date despre temperatura, umiditate, nivelul apei, apoi datele sunt procesate si transmise catre software, iar mai apoi deciziile sunt transmise actuatorilor pentru control.

Vizualizarea datelor prin analiza caracteristicilor realizata prin decompozitia ETS:

*decomposed = seasonal\_decompose(df['tempreature'], model='additive', period=48)*

*decomposed.plot()*

*plt.show()*

*O imagine care conține text, captură de ecran, linie, Interval

Descriere generată automatO imagine care conține text, captură de ecran, Font, Interval

Descriere generată automat*

*O imagine care conține captură de ecran, text, linie, Font

Descriere generată automat Figura 4.3.1 ETS decomposition pentru temperatura Figura 4.3.2 Pentru umiditate*

*Figura 4.3.3 Pentru nivelul de apa Figura 4.3.4 Pentru nutrientul N*

*Figura 4.3.5 Pentru nutrientul O imagine care conține captură de ecran, linie, Paralel

Descriere generată automatP Figura 4.3.6 Pentru nutrientul K*

Impartirea setului de date in date de testare si date de validare:

*X = data[['tempreature', 'humidity', 'water\_level', 'N', 'P', 'K']]*

*y=data[['Fan\_actuator\_OFF','Fan\_actuator\_ON','Watering\_plant\_pump\_ON','Water\_pump\_actuator\_ON']*

*X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)*

*models = {}*

*for column in y\_train.columns:*

*model = RandomForestClassifier(n\_estimators=100, random\_state=42)*

*model.fit(X\_train, y\_train[column])*

*models[column] = model*

Incerc realizarea modelului prin metoda regresiei liniare:

*regressor = LinearRegression()*

*regressor.fit(X\_train, y\_train['Water\_pump\_actuator\_ON'])*

*y\_pred\_reg = regressor.predict(X\_test)*

*for column, model in models.items():*

*predictions = model.predict(X\_test)*

*print(f"Evaluare pentru {column}:")*

*print("Acuratete:", accuracy\_score(y\_test[column], predictions))*

*print("Raport clasificare:\n", classification\_report(y\_test[column], predictions))*

*print("Matrice de confuzie:\n", confusion\_matrix(y\_test[column], predictions))*

*print("-" \* 50)*

*mse = mean\_squared\_error(y\_test['Water\_pump\_actuator\_ON'], y\_pred\_reg)*

*r2 = r2\_score(y\_test['Water\_pump\_actuator\_ON'], y\_pred\_reg)*

*print("Eroarea medie patratica (MSE):", mse)*

*print("Scorul:", r2)*

**Eroarea medie patratica (MSE): 0.01819009520888483**

**Scorul: 0.8954828067878883**

O imagine care conține text, captură de ecran, diagramă, Dreptunghi

Descriere generată automatO imagine care conține text, captură de ecran, diagramă, Dreptunghi

Descriere generată automatPentru Water\_pump\_actuator\_ON, MSE are valoarea 0.0181, ceea ce indica o performanta a modelului de regresie liniara. Scorul de 0.895 demonstreaza ca modelul se potriveste pe datele disponibile.

*Figura 4.3.7 Matricele de confuzie*

Insa folosind regresia liniara apar cateva probleme pentru celelalte categorii de actuatoare:

*mse = mean\_squared\_error(y\_test['Fan\_actuator\_ON'], y\_pred\_reg)*

*r2 = r2\_score(y\_test['Fan\_actuator\_ON'], y\_pred\_reg)*

*print("Eroarea medie patratica (MSE):", mse)*

*print("Scorul:", r2)*

**Eroarea medie patratica (MSE): 0.3275880679014678**

**Scorul: -0.5363270379784844**

Pentru Fan\_actuator\_ON, MSE are valoarea 0.3275, ceea ce indica o performanta slaba. Scorul negativ demonstreaza ca modelul are o problema la aceasta parte, insa nu este foarte mare deoarece valoarea este subunitara.

*mse = mean\_squared\_error(y\_test['Watering\_plant\_pump\_ON'], y\_pred\_reg)*

*r2 = r2\_score(y\_test['Watering\_plant\_pump\_ON'], y\_pred\_reg)*

*print("Eroarea medie patratica (MSE):", mse)*

*print("Scorul:", r2)*

**Eroarea medie patratica (MSE): 0.3370932122002202**

**Scorul: -0.8939559470008789**

Pentru Watering\_plant\_pump\_ON, MSE are valoarea 0.337, ceea ce indica o performanta slaba prin regresie liniara. Scorul negativ demonstreaza ca modelul are o problema.

Pentru cele doua variabile nu functioneaza bine metoda de regresie liniara deoarece scorul este valoare negativa, prin urmare nu are sens sa prezicem o valoare binara cu regresie, deci pentru cele doua folosim modelarea prin RandomForestClassifier.

*# Initializez clasificatorul pentru Fan\_actuator\_ON*

*model\_fan = RandomForestClassifier(n\_estimators=100, random\_state=42)*

*model\_fan.fit(X\_train, y\_train['Fan\_actuator\_ON'])*

*y\_pred\_fan = model\_fan.predict(X\_test)*

*print("Evaluare pentru Fan\_actuator\_ON:")*

*print("Acuratete:", accuracy\_score(y\_test['Fan\_actuator\_ON'], y\_pred\_fan))*

*print("Raport clasificare:\n", classification\_report(y\_test['Fan\_actuator\_ON'], y\_pred\_fan))*

*print("Matrice de confuzie:\n", confusion\_matrix(y\_test['Fan\_actuator\_ON'], y\_pred\_fan))*

*print("-" \* 50)*

*O imagine care conține text, captură de ecran, Font, număr

Descriere generată automat*

*Tabelul 4.3.1. RandomForestClassifier pentru actuatorul de ventilatie*

Acuratetea sugereaza faptul ca modelul a prezis corect toate valorile deoarece are valoarea 1. Raportul de clasificare in momentul in care ventilatorul este pornit are precizia corect cu valoare 1, analog si pentru cazul in care ventilatorul este oprit. Recall sugereaza procentul de situatii identificate corect, iar valoare 1 indica faptul ca toate cazurile reale din setul de predictie au fost identificate corect. Iar F1-score indica o medie armonica intre precizie si recall, valoarea 1 sugerand ca modelul a prezis destul de bine ambele proprietati. Support indica numarul de exemple din fiecare proprietate care au fost efectuate in predictie, si anume 5247 de exemple pentru momentul cand ventilatorul este oprit, si 2338 de exemple efectuate cand ventilatorul este pornit.

*model\_pump = RandomForestClassifier(n\_estimators=100, random\_state=42)*

*model\_pump.fit(X\_train, y\_train['Watering\_plant\_pump\_ON'])*

*y\_pred\_pump = model\_pump.predict(X\_test)*

*print("Evaluare pentru Watering\_plant\_pump\_ON:")*

*print("Acuratete:", accuracy\_score(y\_test['Watering\_plant\_pump\_ON'], y\_pred\_pump))*

*print("Raport clasificare:\n", classification\_report(y\_test['Watering\_plant\_pump\_ON' ] , y\_pred\_pump))*

*print("Matrice de confuzie:\n", confusion\_matrix(y\_test['Watering\_plant\_pump\_ON'], y\_pred\_pump))*

*print("-" \* 50)*

*O imagine care conține text, captură de ecran, Font, număr

Descriere generată automat*

*Tabelul 4.3.2. RandomForestClassifier pentru pompa de apa*

Acuratetea sugereaza faptul ca modelul a prezis corect toate valorile deoarece are valoarea 1. Raportul de clasificare in momentul in care pompa de apa este pornita are precizia corect cu valoare 1, analog si pentru cazul in care pompa de apa este oprita. Recall sugereaza procentul de situatii identificate corect, iar valoare 1 indica faptul ca toate cazurile reale din setul de predictie au fost identificate corect. Iar F1-score indica o medie armonica intre precizie si recall, valoarea 1 sugerand ca modelul a prezis destul de bine ambele proprietati. Support indica numarul de exemple din fiecare proprietate care au fost efectuate in predictie, si anume 5828 de exemple pentru momentul cand pompa de apa este oprita, si 1757 de exemple efectuate cand pompa de apa este pornita.

*O imagine care conține text, captură de ecran, linie, Interval

Descriere generată automat*

*Figura 4.3.8. Importanta caracteristicilor*

Temperatura este cea mai importanta caracteristica, avand cea mai mare influenta asupra activitatii actuatorilor, in conditiile de pornire si oprire. Umiditatea are de asemenea o influenta destul de semnificativa, aceasta determina activarea pompei de apa. Nivelul apei este de asemenea important legandu-se de cele doua caracteristici principale, controland pompa de apa si influenteaza indirect anumiti actuatori. Nutrientii au o importanta redusa avand o influenta redusa asupra actuatorilor.

**4.4 Evaluarea modelului:**

* Metodologia de evaluare: impartirea datelor in seturi de identificare pentru antrenare si in seturi de testare pentru validare.
* Metrici precum acuratetea si precizia
* Eroarea medie patratica pentru predictii
* Matricea de confuzie arata numarul de predictii corecte si incorecte facute de model in functie de valorile reale.
* Valorile de pe diagonala principala reprezinta predictiile corecte.
* Valorile inafara diagonalei principale sunt erori.
* Pentru oprirea ventilatorului, pe diagonala principala 2338 instante corecte in care ventilatorul trebuia sa fie oprit si a fost prezis corect ca fiind oprit, iar 5247 instante corecte in care ventilatorul trebuia sa fie pornit si a fost prezis corect ca fiind oprit.
* Pe diagonala secundara sunt 0 erori, ceea ce inseamna ca modelul a avut o precizie perfecta.
* Evaluare modelului s-a realizat pe baza punctului anterior.
* O imagine care conține text, captură de ecran, număr, linie

  Descriere generată automatUn punct de vedere esential ar fi faptul ca intitial incercand modelarea datelor pe baza regresiei liniare aceasta nu a functionat corect pentru anumiti actuatori, chiar daca pentru actuatorul sipompa de apa functiona corespunzator. Prin urmare pe setul de date cel mai bine a functionat metoda de Random Forest Classifier.

*Figura 4.4.1 Semnalul pentru predictie pentru ventilator*

Semnalul cu albastru este cel actual iar cu rosu este cel prezis ceea ce din imagine reiese ca predictia functioneaza destul de bine, existand doar cateva diferente mici.

O imagine care conține text, linie, număr, Interval

Descriere generată automat

*Figura 4.4.2 Semnalul de predictie pentru actuatorul pompei de apa*

O imagine care conține text, număr, linie, captură de ecran

Descriere generată automat

*Figura 4.4.3 Semnal de predictie pentru pompa de apa*

***O imagine care conține text, captură de ecran, diagramă, Dreptunghi

Descriere generată automat***O imagine care conține text, captură de ecran, diagramă, Dreptunghi

Descriere generată automatCaracteristicile importante pentru fiecare actuator:

**O imagine care conține text, captură de ecran, diagramă, Dreptunghi

Descriere generată automat***Figura 4.4.4 Caracteristici importante pentru pompa de apa*

*Figura 4.4.5 Carcateristici importante pentru ventilator*

# Concluzii

## Rezultate obținute

In urma analizei setului de date si a pre-procesarii acestuia, am realizat modelarea datelor prin metoda Random Forest, de unde am obtinut urmatoarele concluzii: faptul ca randul care indica temperatura este cel mai important factor dintr-un mediu de crestere a plantelor. Umiditatea este de asemenea esentiala pentru activarea pompei de udare, dar si dezactivarea acesteia cand solul este prea umed. Nivelul apei este strans legat de umiditate si temperatura avand un impact asupra activarii pompei de apa. Nutrientii din sol nu sunt atat de semnificativi in clasificare, deoarece nu se pune atat de mult accent ca si pe proprietatiile esentiale, astfel sistemul nu este influentat foarte tare de acesti factori.

Performanta modelului prin metoda Random Forest este una buna, avand o acuratete de 98% si o capacitate buna de predictie, astfel validarea modelului nu difera mult sau semnificativ de predictia realizata.

Matricea de confuzie arata ca nu sunt erori in clasificare, deoarece pe diagonala secundara sunt valori de 0, insa avem o eroare mica la pornirea pompei de apa pe care o pot remedia printr-un model mai precis sau imbunatatit.

## Direcții de dezvoltare

Imbunatatirea sistemului se poate realiza printr-o filtrare mai buna a datelor. De asemenea si utilizarea unor metode mai avansate cum sunt Retelele Neuronale care ar imbunatati clasificarea.

Se poate utiliza si un sistem de control inteligent care sa ajusteze mai bine preluarea datelor de la senzori si pornirea actuatorilor si a pompei de apa, dar si oprirea acestora pe baza informatiilor primite de la senzori. Incercarea crearii unui model mai performant, dar si dezvoltarea unei aplicatii prin care caracteristicile sa poata fi controlate si de la distanta de catre utilizator.

# Bibliografie

|  |  |  |
| --- | --- | --- |
| [1] |  | "https://www.insongreen.com/automated-greenhouse-system/" |
| [2] |  | "https://vs.inf.ethz.ch/publ/papers/Internet-of-things.pdf" |
| [3]  [4] |  | "https://www.instructables.com/Automated-Greenhouse/"  "https://opensource.com/article/21/5/monitor-greenhouse-open-source" |
| [5] |  | "https://users.utcluj.ro/~civan/thesis\_files/2020\_CorneaIulia.pdf" |

# Reguli de formatare

## Formatarea paginii

* + Dimensiunea paginii: A4
  + Margini: 2.5 cm (sus, jos, stânga, dreapta)
  + Antet și subsol: 1.27 cm de la marginea paginii
  + În antetul paginii (header): titlul capitolului, centrat, stil: Header\_style
  + În subsolul paginii: numărul paginii, centrat

## Titluri și stiluri

Titlurile capitolelor și subcapitolelor se marchează cu stilurile Heading 1 – 4, conform documentului model anexat în format Word. Descrierea stilurilor utilizate în document este prezentată în Tabelul 5.1.

Tabelul 5.1. Stiluri utilizate în acest document

| Nr. | Stil | Utilizat pentru | Format |
| --- | --- | --- | --- |
| 1 | Normal | Text normal | Font: (Default) Cambria, 12 pt, Justified, Line spacing: Multiple 1.1 li, Space After: 6 pt |
| 2 | Titlu | Titlul proiectului, prima pagină | Font: 24 pt, Small caps, Centered Line spacing: single, Space Before: 126pt, After: 0 pt, |
| 3 | Titlu2 | Titlul proiectului, pagina de prezentare | Font:14pt, Bold, Centered |
| 4 | Heading 1 | Titlurile capitolelor (nivel 1) | Font: 24 pt, Indent: Left: 0 cm Hanging: 0.76 cm, Space Before: 24pt, After: 12pt |
| 5 | Heading 2 | Titlurile subcapitolelor (nivel 2) | Font: 14 pt, Bold, Indent: Left: 0 cm  Hanging: 1.02 cm, Space Before: 18pt, After: 12pt |
| 6 | Heading 3 | Titlurile secțiunilor (nivel 3) | Font: Bold, Indent: Left: 0 cm Hanging: 1.27 cm, Space Before: 6 pt, After: 6pt |
| 7 | Heading 4 | Titlurile secțiunilor (nivel 4) | Font: Italic, Indent: Left: 0 cm Hanging: 1.52 cm, Space Before: 2 pt, After: 0 pt |
| 8 | Caption | Legenda figurilor și tabelelor | Font: Italic, Font color: Text 1, Line spacing: single, Space After: 10 pt, |
| 9 | Header\_style | Antetul paginii | Font: 10 pt, Italic, Centered, Border: Bottom: (Single solid line, Background 1, 0.5 pt Line width) |

## Figuri, tabele și ecuații

### Figuri

Figurile se inserează în text centrate, cu etichetă de numerotare și legendă (Caption) în partea de jos a figurii. Numărul figurii include și numărul capitolului, după exemplul prezentat în Figura 5.1.



Figura 5.1. Figură exemplu, stil: Caption

## Tabele

Tabelele se inserează în text centrate, cu etichetă și legendă (Caption) în partea de sus a tabelului, aliniată la stânga. Numărul tabelului include și numărul capitolului, după cum este prezentat, de exemplu, în Tabelul 5.1.

## Ecuații

Ecuațiile se inserează în text centrate, cu numerotare în partea dreaptă. Numărul ecuației include și numărul capitolului, conform exemplului din relația (5.1).

|  |  |
| --- | --- |
|  | (5.1) |

## Referințe bibliografice

Se recomandă ca citarea referințelor bibliografice să fie făcută în formatul IEEE.

În secțiunea Bibliografie sunt prezentate exemple pentru: o citare a unui capitol dintr-o carte [1], un articol publicat într-o revistă [2] și un articol publicat la o conferință [3].

Detalii cu privire la formatul citării diverselor tipuri de referințe pot fi găsite în [4] sau [5].

Referințele bibliografice se pot insera în text utilizând facilitățile Word de a adăuga surse și bibliografie unui document (References -> Citations & Bibliography). Dacă formatul IEEE pentru bibliografie nu este instalat implicit în Word, se poate descărca gratuit de la:

<https://bibword.codeplex.com/wikipage?title=Styles&referringTitle=Home>

Instrucțiunile de instalare pentru diferite versiuni de Word se pot obține de la aceeași adresă.