Sistem Inteligent De Irigare Pentru Gradini Pluviale Folosind Datele De Umiditate A Solului

PROIECT DE SEMESTRU

Student: **Alexandra Malutan**

Disciplina: **Sisteme bazate pe Cunoaștere**

Cuprins

[1 Introducere (1-5 pag) 2](#_Toc181667062)

[1.1 Context general 2](#_Toc181667063)

[1.2 Obiective 2](#_Toc181667064)

[1.3 Specificații 2](#_Toc181667065)

[2 Cunoașterea și analiza setului de date (1-5 pag) 3](#_Toc181667066)

[3 Pre-procesarea setului de date (1-5 pag) 4](#_Toc181667067)

[4 Modelarea sistemului (1-10 pag) 5](#_Toc181667068)

[5 Concluzii (1-3 pag) 6](#_Toc181667069)

[5.1 Rezultate obținute 6](#_Toc181667070)

[5.2 Direcții de dezvoltare 6](#_Toc181667071)

[6 Bibliografie (1-2 pag) 7](#_Toc181667072)

[7 Reguli de formatare 8](#_Toc181667073)

[7.1 Formatarea paginii 8](#_Toc181667074)

[7.2 Titluri și stiluri 8](#_Toc181667075)

[7.3 Figuri, tabele și ecuații 9](#_Toc181667076)

[7.3.1 Figuri 9](#_Toc181667077)

[7.4 Tabele 9](#_Toc181667078)

[7.5 Ecuații 9](#_Toc181667079)

[7.6 Referințe bibliografice 10](#_Toc181667080)

# Introducere (1-5 pag)

## Context general

Într-un context global al schimbarilor climatice si al cresterii cererii pentru utilizarea sustenabila a resurselor, gestionarea eficienta a apei devine esentiala, in special in mediul urban. Gradinile de ploaie reprezinta o solutie ecologica importanta, avand rolul de a reduce poluarea apei, de a gestiona apa pluviala. Mentinerea acestor spatii necesita un echilibru intre irigare excesiva si insuficienta, ceea ce duce la necesitatea dezvoltarii unor sisteme inteligente pentru gestionarea resurselor de apa. Acest proiect reprezinta o solutie pentru problema. Proiectul propune un sistem de irigare inteligent bazat pe datele colectate de senzori de umiditate a solului. Prin analizarea continua a acestor date, sistemul poate decide momentul optim și cantitatea necesară de apă pentru fiecare tip de sol și plantă.

Umiditatea solului joacă un rol cheie în ciclul hidrologic al Pământului și în procesele meteorologice și climatice. Informațiile privind conținutul de umiditate a solului sunt necesare pentru programarea irigațiilor, prognoza randamentului culturilor, studiile privind vremea și schimbările climatice, monitorizarea și prognoza evenimentelor meteorologice extreme, cum ar fi inundațiile și seceta, și estimarea scurgerii și eroziunii solului. Pentru aceste aplicații sunt necesare estimarea și prognozarea precisă și în timp util a umidității solului. Algoritmii de învățare automată (ML), cum ar fi rețelele neuronale artificiale, mașinile vectoriale de suport, arborii de decizie, pădurile aleatorii și așa mai departe, sunt utilizați pe scară largă pentru evaluarea umidității solului datorită capacității lor de a modela relații neliniare și complexe între variabile. Acești algoritmi sunt utilizați pentru a dezvolta funcții de pedotransfer care pot prezice proprietățile hidraulice ale solului, cum ar fi capacitatea disponibilă a apei, conductivitatea hidraulică, curba de retenție a apei din sol și multe altele. Acești algoritmi sunt utilizați și pentru recuperarea umidității solului prin teledetecție. Prin furnizarea de date meteorologice, de vegetație, topografice și istorice despre variația umidității solului, acești algoritmi ML pot prognoza cu precizie umiditatea solului după câteva zile. Aceste informații pot fi utilizate pentru programarea irigațiilor în sistemul automat de irigare inteligent. Acești algoritmi sunt, de asemenea, utilizați pe scară largă pentru reducerea la scară grosieră a produselor de umiditate a solului, derivate din satelit, la rezoluții spațiale mai fine, astfel încât aceste produse să poată fi aplicate la nivel regional sau bazin hidrografic. Algoritmii ML contribuie semnificativ la progresul cercetării umidității solului. În acest capitol, este prezentată o privire de ansamblu asupra aplicabilității algoritmilor ML pentru evaluarea umidității solului în diferitele domenii ale cercetării umidității solului.

Această lucrare adresează probleme globale precum:

* Creșterea consumului necontrolat de apă în mediul urban.
* Lipsa soluțiilor sustenabile pentru gestionarea irigațiilor.
* Necesitatea maximizării sănătății plantelor și a productivității solului.

Soluția propusă are aplicații în agricultură, peisagistică și infrastructura verde.

A diagram of a soil monitoring system

Description automatically generated

Un alt studiu realizat in 2019 (vezi bibliografie) propune un model bazat pe rețele neuronale profunde de regresie (DNNR) pentru prognoza umidității solului în zona Beijing, folosind date meteorologice și date despre umiditatea solului colectate între 2012 și 2016. Modelul analizează serii temporale și corelațiile dintre parametrii meteorologici și umiditatea solului prin diagrame Taylor, ceea ce permite selectarea optimă a caracteristicilor relevante pentru prognoză.

Rezultatele testelor demonstrează că modelul bazat pe deep learning oferă o predicție precisă a tendințelor și valorilor umidității solului, datorită capacității sale de a procesa big data și de a identifica relații complexe între variabile. Aceasta poate contribui la optimizarea sistemelor de irigație și la prevenirea secetei, oferind o bază teoretică solidă pentru utilizarea eficientă a apei în agricultură.

## Obiective

* **Obiectivele generale:**

1. Dezvoltarea unui sistem inteligent de irigare pentru grădinile de ploaie.
2. Optimizarea consumului de apă pe baza datelor în timp real.

* **Obiective specifice:**

1. Analiza datelor de umiditate a solului colectate pe parcursul unui an.
2. Construirea unui model de predicție a nevoilor de irigare folosind algoritmi de învățare automată. (regresie)
3. Evaluarea eficienței sistemului prin reducerea risipei de apă și menținerea sănătății plantelor.

## Specificații

Tip de sistem: Automatizat, bazat pe analiza datelor în timp real.

Intrări: Date de la senzori de umiditate a solului, temperatură, umiditate atmosferică (dacă sunt disponibile), prognoza meteo (dacă este integrată).

Ieșiri: Controlul pompei de irigare, activarea sau dezactivarea udării în funcție de pragurile definite.

* Algoritmi și metode utilizate:

Preprocesarea datelor: Eliminarea valorilor aberante, interpolarea datelor lipsă, analiza tendințelor.

Modelul de predicție: Regresie liniară/multivariată, rețele neuronale artificiale, modele bazate pe arbori de decizie.

Optimizarea consumului de apă: Algoritmi de control bazat pe feedback pentru ajustarea debitului de apă.

* Arhitectura sistemului:

Hardware: Microcontroler (ESP32, Raspberry Pi, Arduino) pentru colectarea datelor și controlul sistemului.

Software: Python pentru prelucrarea datelor și antrenarea modelului, API pentru comunicare cu microcontrolerul.

Bază de date: Stocarea datelor istorice pentru analiză și îmbunătățirea modelului.

* Performanță și fiabilitate:

Timp de răspuns: Sistemul trebuie să reacționeze la modificările umidității solului în maxim 15 minute.

Precizia modelului: Se va evalua utilizând metrici precum RMSE (Root Mean Square Error) și R² (coeficientul de determinare).

Robustețe: Funcționare stabilă în diverse condiții meteorologice.

* Limitări:

Dependenta de senzorii existenți: Calitatea și precizia predicției sunt influențate de fiabilitatea senzorilor.

Necesar de calibrare: Sistemul trebuie ajustat periodic în funcție de schimbările sezoniere.

Consumul energetic: Necesitatea unui sistem de alimentare eficient pentru funcționare pe termen lung.

# Cunoașterea și analiza setului de date (1-5 pag)

Setul de date utilizat în cadrul acestui proiect conține informații despre nivelurile de umiditate a solului colectate din grădini de ploaie din Portland, Oregon, pe parcursul unui an. În această secțiune, analizăm detaliat setul de date pentru a identifica caracteristicile sale principale, potențialele probleme de calitate și pentru a efectua analize descriptive și vizuale.

In ceea ce priveste mediul de lucru, am utilizat limbajul de programare Python, versiunea 3.13 si am folosit mediul de dezvoltare Jupyter Notebook, deoarece este un mediu flexibil si interactiv care faciliteaza explorarea si manipularea datelor.

Setul de date este format din 4 coloane: Time, care prezinta timpul masuratorii, si wfv\_1, wfv\_2, wfv\_3, care sunt 3 senzori de la care s-au colectat informatii timp de 1 an, din 15 in 15 minute. Acesti senzori au fost plasati intr-o gradina cu sol mixt si diferite plante. Astfel, in setul de date sunt facute 14889 inregistrari.

Analiza descriptiva: Acest pas este necesar deoarece permite explorarea rapida a datasetului, evidentiind structura sa, numarul de inregistrari, tipuri de date, descrierea statistica a acestora si altele.

-citirea setului de date

Table 2.1 Citirea setului de date

A number on a white background

Description automatically generated

-numarul de inregistrari realizate



-analiza tipurilor de date a coloanelor cat si informatiilor legate despre faptul daca acestea sunt nule sau nu, dar si memoria pe care setul de date o solicita

Table 2.2 Analiza tipurilor de date

A screenshot of a computer code

Description automatically generated

-descrierea statistica a datelor

Table 2.3 Descrierea statistica a datelor

A table of numbers and a number

Description automatically generated

-verificarea coloanelor



-verificarea valorilor lipsa: Exista cate 2 valori lipsa pe fiecare coloana, care pot creea probleme in analiza si modelare. Pot distorsiona rezultatele si afecta performanta modelelor predictive. Aceasta problema o vom rezolva in partea de preprocesare a datelor.

Table 2.4 Verificare valori lipsa

A number of text on a white background

Description automatically generated

-verificarea valorilor duplicate: Exista valori duplicate la care va trebui sa renuntam, deoarece afecteaza performanta.

Table 2.5 Verificare valori duplicate

A white background with black text

Description automatically generated

Analiza vizuala: Acest pas este necesar pentru a observa anomalii in setul de date, de asemenea, pentru a observa pattern-uri, tendinte si perspective, lucru care se realizeaza prin EDA ( Exploratory Data Analysis).

-analiza univariata

A graph of a graph

Description automatically generated

Figure 2.1 Count plot senzor 1

A graph of a graph

Description automatically generated

Figure 2.2 Count plot senzor 2

A graph of a graph

Description automatically generated

Figure 2.3 Count plot senzor 3

Aceasta imiplica utilizarea graficelor pentru a intelege mai bine datele. In acest scop, se utilizeaza bibliotecile Matplotlib (creearea graficelor 2D )si Seaborn (personalizarea diagramelor). Se observa ca valorile predominate de umiditate se situeaza in intervalul de 10-30%, lucru ce sugereaza ca umiditatea solului este mai frecventa in acest interval. De asemenea, senzorii 2 si 3 au o distributie asemanatoare, de unde rezulta o posibila corelatie.

A screenshot of a graph

Description automatically generated

Figure 2.4 KDP

De aici putem deduce faptul ca, pentru senzorii 1 si 2 asimetria este una pozitiva, iar pentru senzorul 3 este negativa.

-analiza bivariata: Este utila pentru a arata tipare clare.

A group of graphs showing a number of data

Description automatically generated with medium confidence

Figure 2.5 Pair plot

Se poate observa ca perechile de senzori (1-2), (1-3), (2-3) prezinta tipare diagonal clare. Acest lucru sugereaza o corelatie puternica intre toti cei trei senzori de umiditate. (punctele sunt asezate de-a lungul unei directii diagonale, ceea ce inseamna ca pe masura ce creste valorea unui sensor, si ceilalti doi tind sa creasca).

-analiza multivariata: Aceasta parte este foarte utila deoarece sugereaza daca exista variabile puternic corelate.

A red and blue squares with numbers

Description automatically generated

Figure 2.6 Analiza de corelatie

Din acest heatmap de corelatie se observa doar valori >0.9 ceea ce inseamna ca toti senzorii sunt puternic corelati intre ei. Acest lucru usureaza partea de modelare si putem face predictia doar pe 1 senzor, ceilalti doi fiind redundanti. Cei mai corelati sunt senzorul 2 si 3 (0.93 si 0.96), de aceea vom alege unul dintre acestia pentru modelare.

A graph of a graph showing the same number of data

Description automatically generated with medium confidence

Figure 2.7 Setul de date din studiul realizat in 2019, cu informatii referitoare la terenurile din China

A graph of a number of delay

Description automatically generated

Figure 2.7 Autocorelatia pe setul de date de la studiul din 2019

# Pre-procesarea setului de date (1-5 pag)

Preprocesarea datelor este un pas esențial în construirea unui sistem inteligent de irigație, deoarece calitatea datelor influențează direct acuratețea modelului de predicție. Această etapă include:

* Eliminarea zgomotului din date (denoising)
* Reducerea dimensionalității (dimensionality reduction)
* Eliminarea tendințelor (detrending)
* Interpolarea eșantioanelor lipsă (data interpolation)
* Identificarea și eliminarea valorilor aberante (outlier removal)

A graph of different colored lines

Description automatically generated

Figure 3.1 Evolutia umiditatii solului in timp

Reprezentare grafică a seriilor temporale. In acest grafic se poate observa evolutia umiditatii solului in timp (trendurile). Se pot observa fluctuatii periodice. Toti cei 3 senzori urmeaza aceeasi tendinta, confirmand corelatia puternica dintre ei. Exista perioade in care umiditatea scade semnificativ, ceea ce ar putea indica nevoia de irigare.

A graph with lines and dots

Description automatically generated

Figure 3.2 Sezonalitatea umiditatii

Aceasta este o analiza sezoniera a datelor. Am extras lunile, am calculate media umiditatii lunare si am vizualizat sezonalitatea. Umiditatea solului este mai ridicata in lunile de primavera si toamna. In lunile de vara(iunie-august), umiditatea este scazuta, lucru care ne sugereaza o nevoie mai mare de irigare. Exista variatii mici intre senzori, posibil cauzate de pozitionarea acestora in pamant, sau de tipul solului si a plantelor. Deci, sistemul de irigare trebuie sa fie active in lunile de vara, iar toamna si primavara putem reduce udarea.

-denoising: Elimină zgomotul din date, permițându-ne să ne concentrăm pe tiparele semnificative. Zgomotul poate apărea din cauza erorilor echipamentelor, interferențelor de mediu sau a altor surse imprevizibile, ceea ce face analiza datelor mai dificilă. Tehnicile de filtrare ajută la izolarea semnalelor esențiale, eliminând în același timp zgomotul.

A graph showing a line graph

Description automatically generated with medium confidence

Figure 3.3 Filtru median

Am ales filtrul median, deoarece trebuie sa mentinem tendintele sezoniere si variatiile naturale, iar acest filtru este potrivit, deoarece elimina zgomotul (valorile anormale) pastrand forma si tendintele originale ale datelor.

-In ceea ce priveste reducerea dimensionalitatii, aceasta nu este necesara in cazul nostru deoarece lucram doar cu senzorul 2 (in urma analizei setului de date in care am determinat ca cei 3 senzori sunt puternic corelati).

-interpolarea: Este esentiala pentru gestionarea datelor lipsă în seriile temporale, asigurând continuitatea seturilor de date. Valorile lipsă pot perturba analiza, iar interpolarea oferă o metodă de a completa aceste goluri cu precizie, fără a introduce erori sau prejudecăți.

A graph showing a line graph

Description automatically generated with medium confidence

Figure 3.4 Interpolarea

Am aplicat Forward and Backward Filling Interpolation, deoarece este utila daca valorile lipsa sunt putine si apropriate. In cazul nostru avem doar 2 valori lipsa. De asemenea, e o metoda foarte rapida, deoarece nu face calcule complexe.

-detrending: Acesta elimină tendințele pe termen lung, izolând tiparele pe termen scurt din date. Aceasta este utilă pentru analizarea componentelor periodice și a fluctuațiilor în jurul unei valori de bază.

A graph showing a moving average signal

Description automatically generated

Figure 3.5 Detrending

Am ales metoda Mediei Mobile (Substraction of moving average), deoarece elimina tendintele pe termen lung fara sa afecteze sezonalitatea, si pastreaza variatiile locale, care sunt esentiale pentru predictia ML.

-eliminarea valorilor aberante(outliers): valorile aberante pot distorsiona analiza, astfel că identificarea și eliminarea acestora este esențială. Valorile aberante pot fi eliminate pe baza scorului Z sau a intervalului intercuartil (IQR), pentru a păstra doar punctele de date centrale. Eliminarea valorilor aberante ajută la asigurarea faptului că datele reprezintă mai bine condițiile tipice, reducând zgomotul și crescând calitatea informațiilor obținute din date.

A graph showing a red line

Description automatically generated

Figure 3.6 Eliminarea valorilor aberante

Am ales metoda IQR (Interquartile Range), deoarece e cea mai sigura pentru acest set de date. Elimina doar valorile extreme fara a afecta trendul general si putem ajusta pragul pentru a controla cat de agresiv eliminam outlierii.

A graph showing a signal

Description automatically generated

Figure 3.7 Semnalul dupa ce am eliminat valorile duplicate

# Modelarea sistemului (1-10 pag)

Acest capitol prezinta procesul de modelare a sistemului inteligent de irigatie. Modelul a fost dezvoltat pentru a prezice valorile viitoare ale umiditatii si pentru a permite ajustarea consumului de apa in functie de nevoile sololui.

Tehnici de modelare aplicate:

Regresia liniara: Pentru a intelege relatiile dintre variabilele temporale si umiditatea solului

Regresia random forest: Pentru a captura relatiile non-liniare din date si a imbunatatii precizia predictiilor.

Pasi:

* + - am creat niste caracteristici temporale: hour, day, month, year pentru a usura procesul.
    - am separat datele in seturi de antrenare(80%) si testare(20%)
    - am creat lag features pentru ultimele 5 valori anterioare ale umiditatii
    - am antrenat un model de regresie liniara si un model random forest cu arbori de decizie
    - am evaluat modelele folosind metrici statistice

Metrici utilizate:

* RMSE (Root mean squared error), care masoara deviatia medie a predictiilor fata de valorile reale;

(4.1)

MAE (Mean absolute error), care masoara eroarea medie absoluta.

(4.2)

Score, care indica cat de bine explica variabilele independente variabila dependenta

(4.3)

Rezultatele obtinute:

Table 4.1

A number with black text

Description automatically generated with medium confidence

A graph of a graph showing different values

Description automatically generated with medium confidence

Figure 4.1 Modelarea

Se observa foarte clar ca regresia liniara a oferit o performanta superioara, cu un de 0.8562, ceea ce indica o corelatie puternica intre variabile.

Regresia liniara a fost eficienta datorita corelatiei puternice dintre factorii temporali si umiditatea solului, iar Random Forest nu a estimat bine.

Studiul realizat in anul 2019 a comparat modelul DNNR cu metodele existente precum Regresia Liniară (LR), Mașini cu Vectori de Suport (SVM) și Rețele Neuronale Artificiale (ANN) arată că DNNR depășește aceste modele în toate măsurile de performanță:

R² al DNNR este cu 9% mai mare decât SVM și cu 24% mai mare decât ANN;

RMSE al DNNR este mai mic cu 80.74% față de SVM și 87.02% față de ANN;

MAE al DNNR este mai mic cu 91.73% față de LR, 84.38% față de SVM, 88.51% față de ANN și 54.76% față de AGNN.

Zona de testare este situată în Beijing, China (E 115°7' ~ E 117°4', N 39°4' ~ N 41°6'), în regiunile Shunyi, Yanqing și Daxing. Aceasta reprezintă un climat musonic continental semi-umed tipic din zona temperată nordică. Vara este caldă și ploioasă, iar iarna este rece și uscată. Primăvara și toamna sunt scurte. Textura solului este predominant nisipoasă sau asemănătoare solului nisipos.

Table 4.2 Analiza comparativa dintre mai multe metode de modelare, realizat de studiul din 2019

A white sheet with black text

Description automatically generated

A collage of graphs and charts

Description automatically generated

Figure 4.2 The comparison of daily soil moisture predictions in Yanqing, Shunyi and Daxing(MLP).

Acest model are mai multe aplicatii, printre care:

Aplicarea modelului in amenajarile urbane si infrastructura verde pentru a asigura utilizarea sustenabila a apei in gradini si spatii similare.

Extinderea utilizarii modelului in agricultura, grdinarit si spatii verzi urbane unde nivelurile de umiditate trebuie monitorizate atent pentru a evita supra- sau sub- irigarea.

Utilizarea modelului pentru a ajuta cercetatorii sa inteleaga relatia dintre diferite tipuri de sol, specii de plante si retentia apei, contribuind la cele mai bune practici in proiectarea gradinilor si a peisajelor.

# Concluzii (1-3 pag)

## Rezultate obținute

Rezultatele obtinute prin aceasta lucrare au demonstrat ca regresia liniara este cea mai eficienta metoda pentru predictia nivelului de umiditate al solului, cu un de 0.8563, indicand o corelatie puternica intre factorii temporali si nivelul de umiditate.

Pe de alta parte, Random Forest nu a oferit rezultate satisfacatoare, avand cu negativ (-0.8605), ceea ce sugereaza ca modelul nu a reusit sa generalizeze datele corespunzator. Acest rezultat poate fi explicat prin lipsa unor date aditionale relevante, precum precipitatiile si temperatura.

Contributii ale acestei lucrari:

* + - demonstrarea eficientei regresiei liniare pentru analiza datelor de umiditate a solului
    - dezvoltarea unui pipeline complet pentru procesarea si modelarea datelor
    - crearea unui sistem replicabil de analiza si predictie pentru irigarea inteligenta
    - identificarea variabilelor esentiale care afecteaza nivelul de umiditate a solului

Totodata studiul din 2019 a confirmat ca metoda de modelare DNNR este cea mai performanta pentru analiza umiditatii solului.

## Direcții de dezvoltare

Pentru imbunatarirea acestui model, urmatoarele directii de dezvoltare pot fi explorate:

* + - Integrarea unor variabile suplimentare: adaugarea temperaturii aerului, precipitatiilor si umiditatii relative pentru a imbunatatii precizia predictiilor
    - Implementarea unui model de tip LSTM, deoarece datele analizate sunt de tip serie temporala, utilizarea unui model bazat pe retele neuronale recurente poate imbunatati predictiile
    - Crearea unui sistem de irigare automatizat: implementarea unui sistem hardware care să controleze automat irigarea pe baza predicțiilor oferite de model.
    - Testarea și validarea în condiții reale: Aplicarea modelului într-un mediu real, într-un sistem de irigare inteligent, pentru a analiza eficiența acestuia în practică.
    - Optimizare modelului Random Forest: Ajustarea hiperparametrilor și testarea unor tehnici de selecție a caracteristicilor pentru a îmbunătăți performanța acestui model.

# Bibliografie

* <https://www.sciencedirect.com/science/article/abs/pii/B978032385214200001X>
* <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0214508>
* <https://www.sciencedirect.com/science/article/abs/pii/S0048969722021593>
* Research on soil moisture prediction model based on deep learning by Yu Cai, Wengang Zheng, Xin Zhang, Lili Zhangzhong, Xuzhang Xue.

Published: April 3, 2019