



UNIVERSITATEA TEHNICĂ

DIN CLUJ-NAPOCA

CENTRUL UNIVERSITAR NORD DIN BAIA MARE

FACULTATEA DE INGINERIE

Programul de studii: CALCULATOARE

PREDICȚIE CULTURI

STUDENT,

Țințaș Gheorghe Mihăiță

COORDONATOR ȘTIINȚIFIC,

Prof. Rudolf Erdei

2024

Cuprins

1. INTRODUCERE	2
1.1 Importanța	2
Optimizarea resurselor:	3
1.2 Scopul proiectului	3
2. Contextul bazei de date și al proiectului	3
3. STAREA ACTUALĂ A DOMENIULUI	5
3.1 Agricultură de precizie: suprafețe, locații și detalii despre cultivare	6
3.2 Beneficiile agriculturii de precizie	7
4. ASPECTE TEORETICE RELEVANTE	7
5. ANALIZĂ	9
5.1 Corectitudinea datelor	9
5.2 Analiza datelor și Corelațiile	9
6. IMPLEMENTARE	11
7. Testare și Validare	12
7.1 Rezultatele testării	13
8. Concluzii	19
9. Bibliografie	19

1. INTRODUCERE

Agricultura de precizie este o paradigmă modernă în agricultură, care utilizează tehnologii de vârf pentru a aduce o abordare mai informată, eficientă și sustenabilă în gestionarea fermelor agricole. Această abordare revoluționară își propune să optimizeze utilizarea resurselor agricole, să maximizeze randamentul culturilor și să reducă impactul negativ asupra mediului înconjurător. Prin utilizarea datelor și a tehnologiilor avansate, agricultorii pot lua decizii mai precise și mai bine informate în ceea ce privește practicile agricole, contribuind astfel la creșterea durabilității și eficienței în agricultură.

Prin valorificarea unei multitudini de date privind factori precum compoziția solului, istoricul performanței culturilor și modelele meteorologice, agricultura de precizie oferă fermierilor informații și recomandări specifice pentru optimizarea randamentelor și profitabilității acestora.

În momentul actual, fermierii se confruntă cu provocări semnificative, cum ar fi schimbările meteorologice imprevizibile și cerințele fluctuante ale pieței, agricultura de precizie apare ca un instrument crucial pentru atenuarea riscurilor și maximizarea randamentelor. Oferind fermierilor informații concrete cu privire la cele mai potrivite culturi pentru terenurile și condițiile lor specifice, aceste sisteme permit luarea unor decizii în cunoștință de cauză, care pot duce la recolte mai reușite și la o stabilitate financiară sporită.

Agricultura de precizie reprezintă o schimbare de paradigmă în practicile agricole, îndepărtându-se de abordările tradiționale, universale, către strategii personalizate adaptate nevoilor unice ale fiecărei ferme. Prin adoptarea acestor tehnologii inovatoare și a sistemelor de sprijinire a deciziilor, fermierii pot naviga prin complexitatea agriculturii moderne cu mai multă încredere și rezistență, deschizând calea pentru un viitor mai durabil și mai prosper.

1.1 Importanța

Agricultura de precizie este crucială în contextul actual al schimbărilor climatice, creșterii demografice și nevoii de a produce alimente în mod sustenabil și eficient. Prin utilizarea datelor precise și a modelelor predictive, agricultorii pot adapta mai bine practicile lor la condițiile locale și pot gestiona resursele într-un mod mai eficient. Aceasta nu numai că contribuie la creșterea randamentului și a

profitabilității, dar poate și să reducă impactul negativ asupra mediului, să minimizeze riscurile asociate cu condițiile meteorologice extreme și să asigure securitatea alimentară pe termen lung.

Iată câteva motive pentru care acest subiect este esențial:

Optimizarea resurselor: Agricultura de precizie utilizează tehnologii avansate, cum ar fi senzori, imagini satelitare, drone și sisteme de informații geografice (GIS), pentru a monitoriza și gestiona variabilitatea în câmpurile agricole. Aceasta permite fermierilor să utilizeze resursele precum apă, îngrășăminte și pesticide în mod mai eficient, reducând astfel costurile și impactul asupra mediului.

Creșterea randamentului: Prin înțelegerea mai profundă a condițiilor locale, inclusiv a compoziției solului, a condițiilor meteorologice și a altor factori de mediu, agricultura de precizie poate ajuta fermierii să ia decizii mai bine informate cu privire la tipurile de culturi de plante sau de animale care ar prospera cel mai bine într-o anumită zonă. Acest lucru poate duce la o creștere semnificativă a randamentului și a productivității.

Reducerea impactului asupra mediului: Prin utilizarea precisă a resurselor și prin adoptarea practicilor agricole durabile, agricultura de precizie poate contribui la reducerea poluării solului și apei, la conservarea biodiversității și la protejarea habitatelor naturale.

În plus, reducerea utilizării de pesticide și îngrășăminte poate contribui la diminuarea efectelor negative asupra sănătății umane și a ecosistemelor.

Adaptarea la schimbările climatice: În contextul schimbărilor climatice și al fenomenelor meteorologice extreme tot mai frecvente, agricultura de precizie oferă instrumente și tehnici pentru a gestiona și a adapta culturile agricole la condițiile în continuă schimbare. Prin monitorizarea și anticiparea schimbărilor climatice, fermierii pot lua măsuri preventive pentru a minimiza impactul acestora asupra recoltelor și a veniturilor lor.

Sustenabilitatea și securitatea alimentară: Într-o lume în continuă creștere demografică, agricultura de precizie joacă un rol crucial în asigurarea unei producții alimentare suficiente, sigure și sustenabile. Prin creșterea eficienței și a productivității agricole, agricultura de precizie poate contribui la reducerea foametei și la asigurarea accesului la alimente nutritive pentru populația globală în creștere.

1.2 Scopul proiectului

Scopul proiectului este de a dezvolta și implementa un sistem de agricultură de precizie, utilizând date și tehnologii avansate pentru a oferi fermierilor recomandări personalizate și informații detaliate despre practicile agricole optime pentru fiecare fermă în parte. Prin colectarea și analizarea unor seturi de date complexe, inclusiv date meteorologice, de sol și despre îngrășăminte, proiectul va dezvolta modele predictive și sisteme de suport decizional care să ofere recomandări personalizate pentru fiecare fermier cu privire la cele mai potrivite culturi pe care aceștia să le cultive în fermele lor din India.

2. Contextul bazei de date și al proiectului

Proiectul de machine learning are ca scop dezvoltarea unui sistem inteligent de recomandare a culturilor agricole. Acest sistem va folosi algoritmi de învățare automată pentru a analiza datele și a sugera culturile care se potrivesc cel mai bine condițiilor specifice de sol și climat. Aceste recomandări

personalizate vor ajuta fermierii să optimizeze utilizarea resurselor și să îmbunătățească randamentul recoltei.

Pentru acest proiect am ales o baza de date ce conține resurse valoroase pentru dezvoltarea și implementarea agriculturii de precizie. Aceasta conține o serie de parametri esențiali care pot fi utilizați pentru a construi modele predictive pentru recomandarea celor mai potrivite culturi de cultivat într-o anumită fermă. Setul de date „Crop_recommendation.csv” conține 2200 de înregistrări și 8 coloane, fiecare reprezentând diferiți parametri relevanți pentru recomandarea culturilor agricole.

N (Azot), P (Fosfor), K (Potasiu): Acești parametri reprezintă rapoartele conținutului de azot, fosfor și potasiu din sol. Acești nutrienți sunt esențiali pentru creșterea plantelor și influențează în mod direct producția agricolă. Nivelele optime ale acestor nutrienți în sol pot varia în funcție de cultură și de alte condiții locale.

Temperatură și Umiditate: Acești factori climatici influențează în mod semnificativ dezvoltarea plantelor și absorbția nutrienților din sol. Temperatura și umiditatea pot afecta rata de transpirație a plantelor, rata de creștere și sensibilitatea la boli și dăunători.

pH-ul solului: Valoarea pH-ului indică nivelul de aciditate sau alcalinitate al solului. Un pH optim este crucial pentru absorbția adecvată a nutrienților de către plante. Anumite culturi preferă soluri mai acide, în timp ce altele prosperă în soluri mai alcaline.

Precipitații: Cantitatea și distribuția precipitațiilor sunt esențiale pentru asigurarea aportului adecvat de apă către culturi. Deficitul sau excesul de apă pot avea un impact semnificativ asupra randamentului culturilor și pot duce la probleme precum uscăciunea sau inundarea.

Label: Tipul de cultură agricolă recomandată, reprezentat sub formă de etichetă textuală. Exemple de culturi includ „rice” (orez), „wheat” (grâu), „maize” (porumb), etc.

N	P	K	temperature	humidity	ph	rainfall	label
90	42	43	20.88	82.00	6.50	202.94	rice
85	58	41	21.77	80.32	7.04	226.66	rice
60	55	44	23.00	82.32	7.84	263.96	rice
74	35	40	26.49	80.16	6.98	242.86	rice
78	42	42	20.13	81.60	7.63	262.72	rice

Figura 1: Înregistrări din setul de date

În figura 1 este prezentat un set de înregistrări din baza de date. Acest set de date poate fi utilizat pentru antrenarea și testarea modelelor de machine learning în vederea realizării de predicții privind culturile agricole optime pentru condițiile specifice de sol și climat. Modelele pot învăța să coreleze parametrii de mediu și sol cu tipurile de culturi care ar avea cele mai bune rezultate, ajutând astfel fermierii să ia decizii informate și să îmbunătățească productivitatea și sustenabilitatea agriculturii. Prin analiza acestui set de date, se poate dezvolta un sistem de recomandare a culturilor care să ofere sfaturi precise și personalizate bazate pe datele locale de mediu și sol, contribuind astfel la optimizarea proceselor agricole și la utilizarea eficientă a resurselor. Acest set de date a fost descărcat de pe

platforma Kaggle de la următorul link: <https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset>.

3. STAREA ACTUALĂ A DOMENIULUI

Decizia fermierului cu privire la ce cultură să crească este, în general, tulburată de intuiția sa și de alți factori irelevanți, cum ar fi obținerea de profituri instantanee, lipsa de conștientizare a cererii pieței, supraestimarea potențialului unui sol de a susține o anumită cultură și așa mai departe. O decizie foarte greșită din partea fermierului ar putea pune o presiune semnificativă asupra stării financiare a familiei sale. Poate că acesta ar putea fi unul dintre numeroasele motive care contribuie la nenumăratele cazuri de sinucidere ale fermierilor pe care le auzim zilnic din mass-media. Într-o țară precum India, în care agricultura și sectorul conexe contribuie la aproximativ 20,4% din valoarea sa adăugată brută, o astfel de judecată eronată ar avea implicații negative nu doar asupra familiei fermierului, ci și asupra întregii economii a unei regiuni.

Agricultura este coloana vertebrală pentru țările în curs de dezvoltare precum India, deoarece 70% din populație depinde de agricultură. Agricultura din India joacă un rol predominant în economie și ocuparea forței de muncă. Problema comună existentă în rândul fermierilor indieni este că aceștia nu aleg cultura potrivită pentru nevoile lor de sol și, de asemenea, ce îngrășământ să fie folosit pentru cultura lor. datorită acestui fapt, se confruntă cu o grea scădere a productivității. [1]

Agricultura de precizie (PF) este o abordare a managementului fermei care utilizează tehnologia informației pentru a se asigura că culturile și solul primesc exact ceea ce au nevoie pentru o sănătate și productivitate optime. În loc să aplice inputuri similare în întregul domeniu, abordarea își propune să le gestioneze și să le distribuie pe o bază specifică locației pentru a maximiza beneficiile costurilor pe termen lung, precum și pentru a preveni orice risipă. [2]

Ca și agricultura de conservare, PF este o combinație de tehnologii diferite, mai degrabă decât o abordare unică și permite managementului specific sitului să-și utilizeze eficient resursele și să obțină câștiguri economice. Înțelegerea și dobândirea mai multor cunoștințe despre variabilitatea naturală în cadrul unei ferme este unul dintre aspectele vitale ale PF.

În ultimele decenii, multe tehnologii au fost dezvoltate pentru PF; Ele pot fi împărțite în "moale" și "tare". Agricultura de precizie moale depinde de observarea vizuală a culturilor și de deciziile de gestionare a solului bazate pe experiență și intuiție, mai degrabă decât pe analize statistice și științifice. În schimb, agricultura de precizie utilizează toate tehnicile moderne, cum ar fi GPS-ul, teledetecția și tehnologia cu rată variabilă. [3]

Inovația și progresele tehnologice se remarcă tot mai mult în practicile agricole ale fermierilor. Integrarea tehnologiei în agricultură simplifică viețile oamenilor și îmbunătățește constant productivitatea activităților agricole. Transformările constante prin care a trecut agricultura în ultimii ani

au arătat că implementarea unui sistem de management agricol contribuie la o producție eficientă, durabilă și profitabilă.

Conceptul de agricultură de precizie înglobează diferite tehnologii digitale care permit monitorizarea precisă a culturilor. Practicarea acestui concept reprezintă, în momentul de față, cel mai important pas către progres pentru fermele din India.

Tehnologiile din domeniu, de la sisteme GPS automatizate, senzori pentru sol, soluții inteligente de pulverizare până la aplicații inovatoare, îi ajută pe cei care lucrează în sectorul agricol să obțină cele mai bune rezultate într-un mod mai simplu și eficient, fără să risipească resursele. Tocmai din acest motiv, analizăm pe larg cât de inteligentă este agricultura de precizie, care sunt avantajele și caracteristicile acesteia și cum îi poate ajuta pe fermieri să-și atingă obiectivele. [4]

Ca abordare modernă a agriculturii, agricultura de precizie se concentrează pe toate aspectele resurselor (sol, apă, nutrienți) necesare agriculturii. Cu toate acestea, în India, practica este dezvoltată până acum pentru eficiența utilizării nutrienților și eficiența utilizării apei. PF în peisajul indian nu a devenit încă o parte integrantă a sistemelor agricole principale. Literatura de specialitate și consultările părților interesate confirmă, de asemenea, că practica se află într-un stadiu incipient în India.

3.1 Agricultură de precizie: suprafețe, locații și detalii despre cultivare

Cât de multă suprafață din India se află sub agricultura de precizie? Întrucât practica se află într-un stadiu incipient, nu există estimări agregate ale suprafeței disponibile pentru întreaga țară și pentru diferite tehnici. Cu toate acestea, tehnicile de irigare de precizie sau micro-irigarea, cum ar fi irigarea prin picurare și prin aspersiune, au o acoperire semnificativă a suprafeței de aproximativ 9,2 milioane de hectare în aproximativ 29 de state (în martie 2017). Din această suprafață, aproximativ 4,2 milioane de hectare sunt acoperite de irigare prin picurare, în timp ce 4,9 milioane de hectare sunt sub tehnologii de aspersoare.

La ce dimensiune a fermei se practică PF? PF este practică de un număr limitat de fermieri, adesea pe un singur câmp sau pe bază experimentală sau ferme comerciale care cultivă culturi de mare valoare. Potrivit părților interesate consultate, în Karnataka, pachetul complet de tehnologie PF este utilizat numai de marii proprietari progresivi de terenuri, deoarece costul ridicat limitează absorbția sa.

Câți fermieri din India practică agricultura de precizie? PF este practică de un număr limitat de fermieri, adesea pe un singur câmp sau pe bază experimentală sau ferme comerciale care cultivă culturi de mare valoare. Potrivit părților interesate consultate, în Karnataka, pachetul complet de tehnologie PF este utilizat numai de marii proprietari progresivi de terenuri, deoarece costul ridicat limitează absorbția sa. Deoarece suprafața sub micro-irigare este de aproximativ 9,2 milioane ha, atunci gama de fermieri care adoptă PF se calculează prin împărțirea suprafeței la dimensiunea terenului implementatorilor, care sunt în mare parte semi-medii spre medii. Prin astfel de metode, estimăm în linii mari că aproximativ 3 milioane dintre aceștia practică PF, în special activități de micro-irigare.

Unde în India este răspândită agricultura de precizie? PF este practicat de un număr minim de fermieri din țară. Deși nu există cifre agregate oficiale disponibile la nivel de țară, se estimează o estimare aproximativă de aproximativ 3 milioane de adoptatori. Cele mai multe dintre ele sunt situate în zonele irigate; aproximativ 1.000 de fermieri au adoptat tehnica LLL în vestul Uttar Pradesh și Haryana. Pentru CLCC, aproximativ 48000 de exemplare sunt înregistrate pentru a fi vândute conform celor mai recente date disponibile, ceea ce oferă un sentiment generic al ratei sale de adoptare, care este scăzută.

Care sunt principalele culturi cultivate în cadrul agriculturii de precizie din India? În cea mai mare parte, culturile comerciale și horticole din fermele cooperative găsesc mult domeniu, deoarece este o provocare pentru micii fermieri să utilizeze utilajele agricole în parcele fragmentate. Cu toate acestea, unele aspecte ale PF, cum ar fi CLCC, pot fi adoptate de ferme pentru multe culturi, inclusiv cereale. În același timp, sistemele de fertilizare prin picurare, aplicatoarele cu rată variabilă și alți senzori sunt mai profitabile pentru culturile cu valoare ridicată, având în vedere costul ridicat asociat acestora. Prin urmare, vedem mai ales fermieri care practică PF implicați în cultivarea fructelor, legumelor, condimentelor, florilor, culturilor medicinale și aromatice. [5]

3.2 Beneficiile agriculturii de precizie

Marea promisiune din spatele tehnologiilor de agricultură de precizie emergente și pregătite pentru câmp nu ne ajută doar să creștem mai multe alimente, ci să facem acest lucru cu o mai mare inteligență. Creșterea culturilor se poate face cu resurse mai puține și cu o atenție sporită pentru mediul înconjurător, iar asta face din agricultura de precizie un element cheie în viitorul agriculturii. Însă avantajele agriculturii de precizie nu sunt doar declarative, ci diferă în funcție de nivelul fermei,

- Ferme de dimensiuni reduse
 - optimizarea resurselor
 - reducerea pierderilor
 - reducerea materiei prime utilizate
 - reducerea poluării
 - utilizarea unor soluții diferite în funcție de nevoile ariilor lucrate
- Ferme de dimensiuni mari
 - creșterea productivității
 - creșterea profitabilității
 - reducerea pierderilor
 - reducerea materiei prime utilizate [6]

4. ASPECTE TEORETICE RELEVANTE

Machine Learning (ML) este, de fapt, folosirea inteligenței artificiale într-un mod în care se oferă sistemelor abilitatea de a învăța singure în mod automatizat, dar și oportunitatea de a deveni mai bune pe parcursul acestei experiențe fără a fi asistate de programatori umani. Altfel spus, ML se concentrează pe dezvoltarea de programe computerizate care pot accesa singure date și le și pot folosi pentru propriile nevoi de învățare. Exact ca și în cazul ființelor biologice, procesele de învățare ale sistemelor cibernetice încep cu etapa de observare, care, în acest ultim caz, se poate traduce în acumularea de date. Ulterior, experiențele directe din lumea biologică pot fi transpuse în seturi de instrucțiuni pe care mașinile cibernetice le înțeleg cu scopul de a căuta modele în seturile de date puse la dispoziție și de a lua decizii autonome pe baza acestora. Ținta finală este de a permite computerelor să poată învăța singure, în mod automat, fără asistarea sau intervenția umană. [7]

Random Forest este un algoritm de învățare. Acesta face parte din familia algoritmilor bazați pe arbori de decizie și este conceput pentru a îmbunătăți acuratețea și a reduce supra învățarea care poate apărea atunci când se utilizează un singur arbore de decizie. Algoritmul Random Forest funcționează

prin construirea unui set de arbori de decizie independenți și combinând rezultatele acestora. În timpul antrenării, se selectează un set aleatoriu de caracteristici pentru a fi utilizate în construirea fiecărui arbore de decizie, astfel încât fiecare arbore să fie diferit și să aibă o abordare unică asupra setului de date. Pentru a returna un rezultat, algoritmul colectează rezultatul fiecărui arbore de decizie și le combină pentru a obține un rezultat final mai precis și mai stabil decât cel care ar fi fost obținut prin utilizarea unui singur arbore. Acest proces de îmbinare a rezultatelor din mai mulți arbori de decizie este numit ensemble learning. [8]

Decision Tree. Un arbore decizional este un algoritm de învățare supravegheat care este utilizat pentru modelarea clasificării și regresiei. Regresia este o metodă utilizată pentru modelarea predictivă, astfel încât acești arbori sunt utilizați fie pentru a clasifica datele, fie pentru a prezice ce va urma.

Arborii decizionali arată ca niște scheme logice, începând de la **nodul rădăcină** cu o întrebare specifică de date, ceea ce duce la ramuri care dețin răspunsuri potențiale. Ramurile conduc apoi la **noduri de decizie (interne)**, care pun mai multe întrebări care duc la mai multe rezultate. Acest lucru continuă până când datele ajung la ceea ce se numește **un nod terminal (sau "frunză")** și se termină.

În învățarea automată, există patru metode principale de algoritmi de instruire: învățare supravegheată, nesupravegheată, învățare de consolidare și învățare semi-supravegheată. Un arbore decizional ne ajută să vizualizăm modul în care un algoritm de învățare supravegheată duce la rezultate specifice. [9]

Gradient Boosted Trees (GBT) este un algoritm de învățare automată pentru clasificare și regresie care creează un model puternic prin combinarea unei secvențe de modele slabe, de obicei arbori de decizie. Fiecare arbore este construit pentru a corecta erorile arborelui anterior.

Principiu de funcționare:

- **Adăugarea secvențială:** Arborii sunt adăugați secvențial, fiecare nou arbore fiind antrenat pentru a corecta erorile reziduale ale modelului actual.
- **Optimizare gradientală:** Utilizarea metodelor de optimizare gradientală pentru a minimiza funcția de pierdere a modelului combinat, ajustând ponderile arborilor noi pentru a corecta erorile din predicțiile anterioare.
- **Ponderarea arborilor:** Fiecare arbore este ponderat în funcție de performanța sa, astfel încât arborii care fac mai puține erori să aibă o influență mai mare.

Avantaje:

- **Performanță ridicată:** Este capabil să modeleze relații complexe și non-liniare între caracteristici și variabila țintă.
- **Reducerea erorilor:** Corectarea iterativă a erorilor reduce treptat eroarea generală a modelului.

Limitări:

- **Timp de antrenare:** Antrenarea unui model GBT poate fi computațional costisitoare și poate necesita mult timp, mai ales pentru seturi de date mari.

- **Sensibilitate la hiperparametri:** Performanța modelului poate fi sensibilă la setarea hiperparametrilor, cum ar fi numărul de arbori, rata de învățare și adâncimea maximă a arborilor. [10]

Algoritmul K-NN se bazează pe compararea unei înregistrări necunoscute cu K înregistrări care sunt cele mai apropiate. Aproximarea este definită în termeni de distanță în spațiul n-dimensional, definită de n atribute din setul de date de antrenament. În al doilea pas, algoritmul clasifică înregistrarea necunoscută printr-un vot majoritar al vecinilor găsiți. Poate evalua distanța vecinilor, astfel încât cei mai apropiați vecini să aibă o pondere mai mare în contribuție la medie decât cei mai îndepărtați. [11]

5. ANALIZĂ

5.1 Corectitudinea datelor

Corectitudinea datelor este esențială pentru acest proiect, care utilizează tehnici de machine learning pentru a estima agricultura de precizie.

Pentru citirea și curățarea datelor am folosit un subprocess care conține următorii operatori.

1. Retrieve - Citește datele primite
2. Remove Duplicates - Pentru a elimina duplicatele
3. Nominal to Numeric – Pentru a transforma o coloană din formă nominală în valori numerice

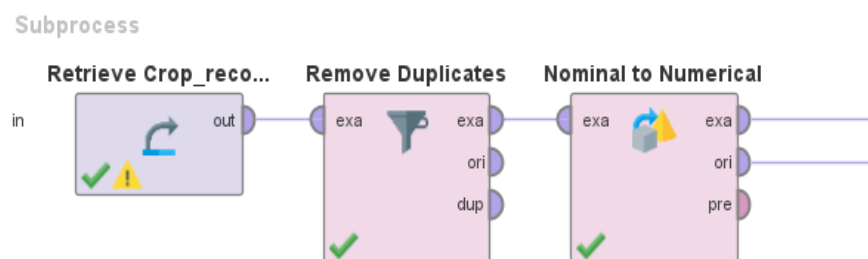


Figura 2 Subprocess- Citire și curățare date

Pentru a pregăti datele noastre pentru analiză în RapidMiner am aplicat operația "Remove Duplicates" pentru a curăța setul de date și pentru a elimina eventualele duplicate. Din fericire, în timpul prelucrării datelor am observat că baza de date nu conține rânduri care să aibă valori nule, de asemenea nu am identificat nici rânduri duplicate, deci putem afirma faptul că folosind această operație, ne-am asigurat că informațiile pe care le analizăm sunt relevante și non-redundante, deci baza noastră de date este curată. Am folosit operația "Nominal to Numeric" pentru a transforma valorile coloanei "Label" din formă nominală în valori numerice, deoarece operațiile de modelare și analiză de date necesită date numerice pentru a fi aplicate, astfel lucrul în RapidMiner devine mult mai eficient.

5.2 Analiza datelor și Corelațiile

Matricea de corelație oferă o perspectivă asupra relațiilor dintre variabile, ceea ce poate fi util pentru înțelegerea tiparelor și pentru a face predicții în date. Elementele de pe diagonala principală sunt toate 1, deoarece reprezintă corelația dintre o variabilă și ea însăși. În general, valoarea unei celule din matricea de corelație indică gradul de relație între două variabile, unde valorile apropiate de 1 indică o corelație pozitivă puternică, valorile apropiate de -1 indică o corelație negativă, iar valorile apropiate de zero indică o lipsă de corelație între variabile.

Pentru a genera matricea de corelații am folosit operatorul Correlation Matrix.

Attributes	label	N	P	K	temper...	humidity	ph	rainfall
label	1	0.283	-0.168	0.144	0.181	0.524	0.052	0.121
N	0.283	1	-0.231	-0.141	0.027	0.191	0.097	0.059
P	-0.168	-0.231	1	0.736	-0.128	-0.119	-0.138	-0.064
K	0.144	-0.141	0.736	1	-0.160	0.191	-0.170	-0.053
temperat...	0.181	0.027	-0.128	-0.160	1	0.205	-0.018	-0.030
humidity	0.524	0.191	-0.119	0.191	0.205	1	-0.008	0.094
ph	0.052	0.097	-0.138	-0.170	-0.018	-0.008	1	-0.109
rainfall	0.121	0.059	-0.064	-0.053	-0.030	0.094	-0.109	1

Figura 3 Matricea de corelații

În urma analizei matricei de mai sus, nu s-au descoperit corelații așa de puternice. Cea mai mare corelație de 0.736 este între attributele P (fosfor) și K (potasiu) deoarece ele influențează solul cel mai mult, astfel un sol bogat în fosfor este și un sol bogat în potasiu, acest sol fiind potrivit pentru un anumit tip de cultură. Următoarea corelație de 0.524 este observată între attributele Humidity si Label. Această corelație puternică, pozitivă, influențează cel mai mult tipul de cultură care va fi plantat în fermele din India. Sunt anumite timpuri de plante care necesita un sol mai umed cum ar fi orezul, acesta este cultivat într-un sol foarte umed , iar alte tipuri de plante au nevoie de un tip de sol mai uscat. Umiditatea din sol influențează dezvoltarea plantelor.

O altă corelație este observată între attributele N (azot) și Label (tipul de cultură), aceasta are un coeficient de 0.283 și ne specifică că solul care conține azot este fertil , dar totodată aici apare și două corelații negative , una între N (azot) și P (fosfor) ce are un coeficient negativ de -0.168 și una între N (azot) și K (potasiu) ce are un coeficient negativ de -0.144, aceste corelații ne sugerează faptul că un sol care conține azot și potasiu sau azot si fosfor nu este la fel de optim pentru anumite culturi ca și un sol ce conține doar azot sau doar fosfor și potasiu.

Entropia este o măsură a incertitudinii sau a dezordinii într-un set de date. Entropia poate fi folosită pentru a evalua cât de bine sunt separate diferitele tipuri de culturi la un anumit nod al arborelui de decizie. Cu cât entropia este mai mare, cu atât mai mixt este setul de date în acel nod.

```
Entropy label': 4.459431618637295
Entropy K': 5.812815935039871
Entropy P': 6.560747386240556
Entropy N': 6.755148352882815
Entropy ph': 11.103287808412068
Entropy humidity': 11.103287808412068
Entropy temperature': 11.103287808412068
Entropy rainfall': 11.103287808412068
```

Figura 4 Entropia pentru fiecare atribut din setul de date

6. IMPLEMENTARE

Pentru a putea estima agricultura de precizie folosind datele primite am folosit softul RapidMiner. O imagine de ansamblu a tuturor operatorilor folosiți poate fi văzută în figura următoare.

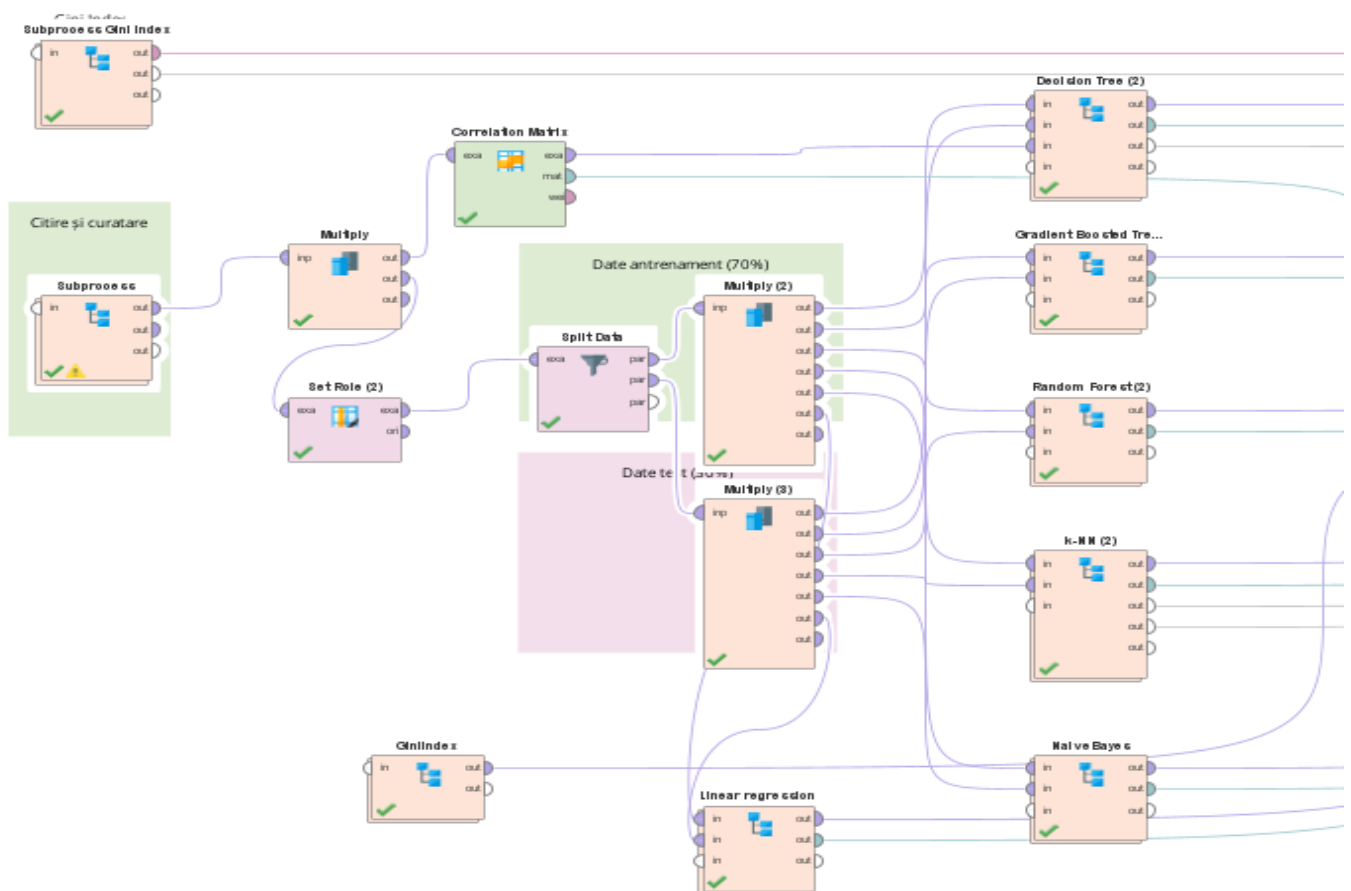


Figura 5 Vedere de ansamblu a tuturor operatorilor

7. Testare și Validare

Datele curățate provenite de la subprocesul de citire și curățare a datelor sunt folosite și de algoritmi de machine learning. Aceste date sunt împărțite în 2 părți, astfel:

- 70% din date sunt folosite pentru antrenament
- 30% din date sunt folosite pentru testare

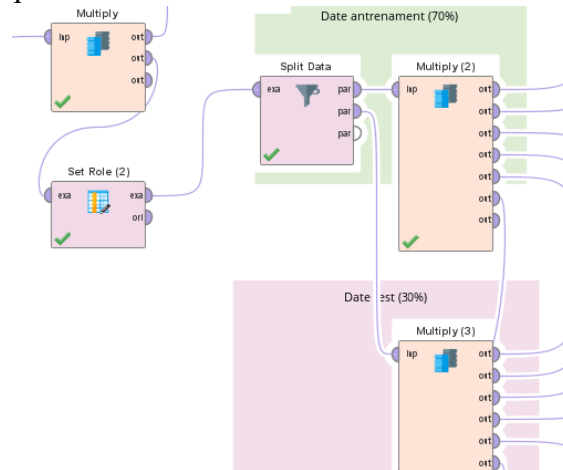


Figura 6 Împărțirea datelor pentru antrenament și testare

Pentru fiecare algoritm de machine learning am creat un subproces separat pentru a putea vedea mai clar flow-ul datelor. Algoritmii implementați sunt:

- Decision Tree
- Gradient Boosted Trees
- Random Forest

Fiecare subproces în care este implementat un algoritm arată ca în figura următoare.

Într-un astfel de subproces se găsesc următorii operatori:

- Algoritm de machine learning (în figura este implementat algoritmul Decision Tree) primește la intrare cele 70% din date necesare pentru antrenament, la ieșire se oferă modelul creat de acest algoritm și la o altă ieșire se oferă setul de date primit la intrare
- Apply Model - acest operator aplică modelul pe un set de date. Operatorul primește la intrare modelul de la algoritm și restul de 30% din date necesare pentru testare.
- Model Simulator – Simulează algoritmul, putând să modificăm input-urile pentru un set de date. Acest operator primește la intrare modelul creat de algoritmul de machine learning, datele de test și de antrenament.

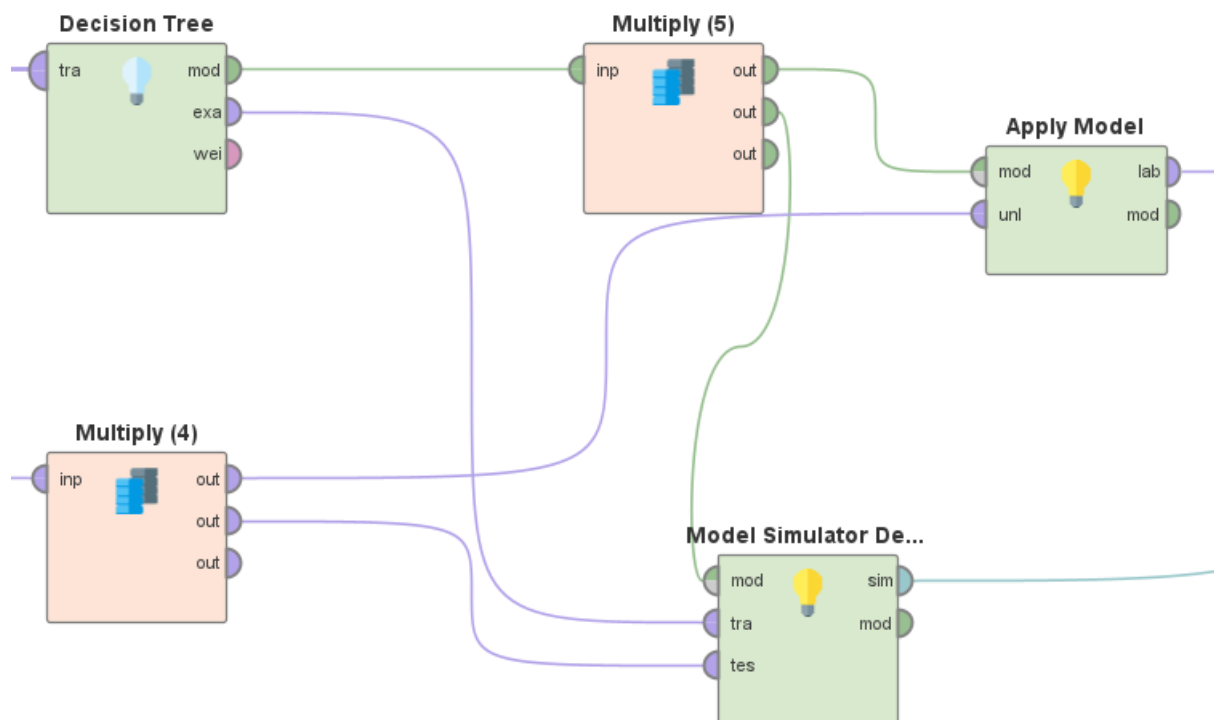


Figura 7 Implementarea algoritmului DecisionTree

7.1 Rezultatele testării

Vom prezenta rezultatele obținute pentru fiecare model, subliniind punctele forte și limitările fiecăruia.

1. Decision Tree

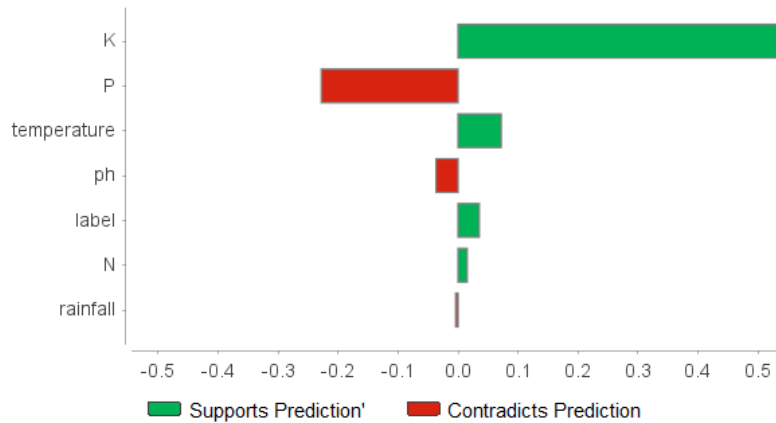
- Valoare predicție: 80067
- Root Mean Squared Error (RMSE): 5200
- Eroarea Relativă: 5.41%

Factorii importanți care suportă predicția sunt: K, temperature, N, rainfall, label.

Factorii importanți care nu suportă predicția sunt: P, ph-ul.

Prediction

Important Factors for Prediction



Accuracy

5.200

Root Mean Squared Error (RMSE)

Relative Error: 5.41%

Interpretation

Select your inputs on the left to see the model's reaction on the right. The prediction of the model is **80.067**. The biggest support for this decision is coming from **K**. The Root Mean Squared Error (RMSE) of all predictions done by this model is 5.200. And the relative error is about 5.41%.

Figura 8 Predicție folosind Decision Tree

2. Gradient Boosted

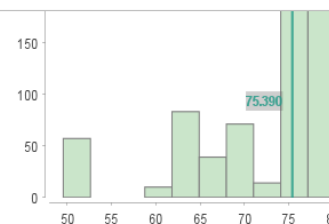
- Valoare predicție: 75390
- Root Mean Squared Error (RMSE): 13641
- Eroarea Relativă: 16.00%

Factorii importanți care suportă predicția sunt: K, temperature, N.

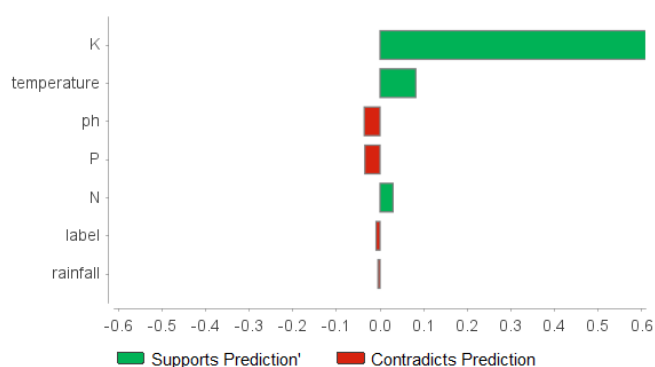
Factorii importanți care nu suportă predicția sunt: P, ph-ul, rainfall, label.

Prediction

75.390



Important Factors for Prediction



Accuracy

13.641

Root Mean Squared Error (RMSE)

Relative Error: 16.00%

Figura 9 Predicție folosind GradientBoosted

3. Random Forest

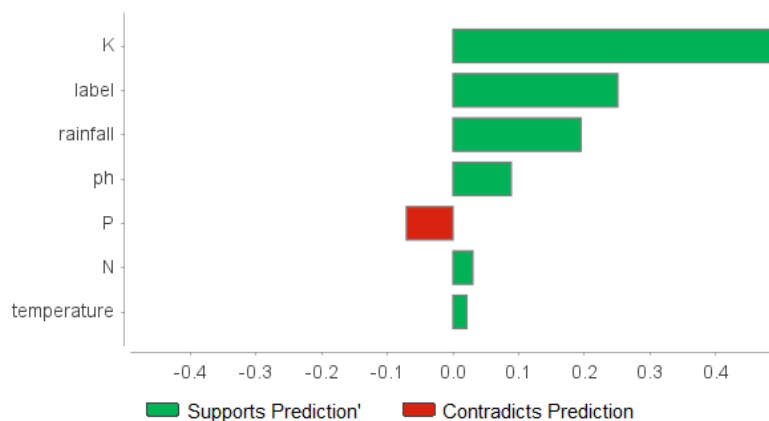
- Valoare predicție: 75390
- Root Mean Squared Error (RMSE): 13641
- Eroarea Relativă: 16.00%

Factorii importanți care suportă predicția sunt: K, temperature, N, ph-ul, rainfall, label.

Factorii importanți care nu suportă predicția sunt: P,

Prediction

Important Factors for Prediction



Accuracy

4.193

Root Mean Squared Error (RMSE)

Relative Error: 4.66%

Interpretation

Select your inputs on the left to see the model's reaction on the right. The prediction of the model is **85.851**. The biggest support for this decision is coming from **K**. The Root Mean Squared Error (RMSE) of all predictions done by this model is 4.193. And the relative error is about 4.66%.

Figura 10 Predicție folosind RandomForest

În tabelele de mai jos apare acuratețe și eroare pentru fiecare algoritm implementat setând în operatorul SetRole fiecare atribut din setul nostru de date.

Humidity

	Eroare(%)	Acuratețe(%)
Decision Tree	5.41%	94.59%
Gradient Bossted Trees	16.00%	84.00%
Random Forest	4.66%	95.34%

K

	Eroare(%)	Acuratețe(%)
Decision Tree	10.64%	89.36%
Gradient Bossted Trees	32.18%	67.82%
Random Forest	9.50%	90.5%

N

	Eroare(%)	Acuratețe(%)
Decision Tree	28.33%	71.67%
Gradient Bossted Trees	34.46%	65.54%
Random Forest	26.37%	73.63%

P

	Eroare(%)	Acuratețe(%)
Decision Tree	19.69%	80.31%
Gradient Bossted Trees	28.75%	71.25%
Random Forest	17.07%	82.93%

Ph

	Eroare(%)	Acuratețe(%)
Decision Tree	7.45%	92.55%
Gradient Bossted Trees	7.32%	92.68%
Random Forest	6.50%	93.5%

Rainfall

	Eroare(%)	Acuratețe(%)
Decision Tree	11.65%	88.35%
Gradient Bossted Trees	23.49%	76.51%
Random Forest	11.27%	88.73%

Temperature

	Eroare(%)	Acuratețe(%)
Decision Tree	9.39%	90.61%
Gradient Bossted Trees	10.74%	89.26%
Random Forest	8.56%	91.44%

	Eroare(%)	Acuratețe(%)
Decision Tree	5.41%	94.59%
Gradient Bossted Trees	16.00%	84.00%
Random Forest	4.66%	95.34%

În tabelul de mai sus sunt prezentate rezultatele celor 3 algoritmi implementați (folosind atributul Humidity în operatorul SetRole, deoarece așa cum se observă și în matricea de corelații acesta influențează cel mai mult tipul de cultură), aici se poate vedea eroarea și acuratețea predicției pentru fiecare algoritm în parte. Cele mai bune rezultate le-am obținut cu algoritmul Random Forest, acesta având o eroare de 4,66% și o acuratețe de 95,34%, iar cel mai slab algoritm este Gradient Boosted Trees cu o eroare de 16,00% și o acuratețe de 84,00%.

Folosind o partiționare 70/30 s-a obținut o acuratețe de 95.34%, iar cu o partiționare 60/40 acuratețea devine 95.08%. Capacitatea modelului de a generaliza în ciuda scăderii cantității de date de antrenament, este remarcabilă.

Tot pentru testare am decis să folosesc în operatorul SetRole și atributul Label.

Pentru algoritmi implementați am găsit următoarele rezultate.

	Eroare(%)	Acuratețe(%)
Decision Tree	1.44%	98.56%
Gradient Bossted Trees	33.54%	66.46%
Random Forest	9.07%	90.93%

Decision Tree a obținut cea mai mică eroare (1.44%) comparativ cu alți algoritmi evaluați, acesta poate fi considerat cel mai bun algoritm.

- **Interpretabilitatea ușoară** oferă claritate în luarea deciziilor agricole.
- **Manipularea eficientă a datelor categoricale și numerice și gestionarea datelor lipsă** asigură flexibilitate și robustețe.
- **Detectarea automată a interacțiunilor și non-parametricitatea** oferă flexibilitate în capturarea relațiilor complexe.
- **Eficiența și scalabilitatea** îl fac potrivit pentru aplicații practice în agricultură.

Aceste avantaje, combinate cu performanța de top a modelului în datele tale, sugerează că Decision Tree este o alegere excelentă pentru a dezvolta un sistem de recomandare a culturilor agricole în cadrul proiectului tău de agricultură de precizie.

8. Concluzii

Proiectul de recomandare a culturilor agricole folosind algoritmi de învățare automată reprezintă o inițiativă inovativă în domeniul agriculturii de precizie. În cadrul acestui proiect, am explorat diverse metode și algoritmi pentru a dezvolta un model care să optimizeze selecția culturilor în funcție de caracteristicile solului și condițiile meteorologice. Iată concluziile principale ale acestui proiect:

1. **Importanța datelor în agricultură de precizie:** Setul de date utilizat, care conține informații despre diferite culturi, condiții de sol și climatice, este esențial pentru construirea unui model precis și fiabil. Colectarea și prelucrarea corectă a datelor sunt pași cruciali în orice proiect de învățare automată.
2. **Performanța algoritmilor de învățare automată:** Am implementat și evaluat mai mulți algoritmi de învățare automată, inclusiv Random Forest, Decision Tree, Gradient Boosted Trees,
3. **Direcții viitoare:**
 - **Extinderea setului de date:** Încorporarea de date suplimentare, cum ar fi date despre fertilitate, istoricul recoltelor și predicții meteorologice pe termen lung, poate îmbunătăți acuratețea modelului.
 - **Îmbunătățirea algoritmilor:** Investigarea altor algoritmi de învățare automată sau combinarea mai multor modele pentru a obține performanțe și mai bune.
 - **Implementarea practică:** Dezvoltarea unei aplicații user-friendly care să permită fermierilor să utilizeze modelul de recomandare în timp real, bazată pe date actualizate continuu.

Proiectul de recomandare a culturilor agricole bazat pe învățare automată demonstrează potențialul tehnologiilor avansate în a sprijini deciziile agricole. Alegerea algoritmilor adecvați și ajustarea corectă a parametrilor lor sunt esențiale pentru dezvoltarea unui sistem eficient și precis. Acest proiect subliniază importanța continuării cercetării și dezvoltării în domeniul agriculturii de precizie pentru a maximiza beneficiile aduse fermierilor și pentru a promova practici agricole sustenabile.

9. Bibliografie

- [1] „Agricultura de precizie India,” *UNIVERSITATEA SAVITRIBAI PHULE PUNE*, 2021-2022.
- [2] 2023. [Interactiv]. Available: <https://github.com/atharval1/precision-agriculture-using-machine-learning/tree/main/Project-docs>.
- [3] „CEEW THE COUNCIL,” 2023. [Interactiv]. Available: <https://www.ceew.in/publications/sustainable-agriculture-india/precision-farming>.
- [4] „BASF AGRO,” 18 Aprilie 2022. [Interactiv]. Available: <https://www.agro.basf.ro/ro/stiri/basf-in-camp/agricultura-de-precizie-ce-este-cum-ajuta-fermierii.html>.

- [5] 2024. [Interactiv]. Available: <https://www.ceew.in/publications/sustainable-agriculture-india/natural-farming>.
- [6] „CMEVO,” 2023. [Interactiv]. Available: <https://cmevo.com/ro/blog/ce-este-agricultura-de-precizie/>.
- [7] „THC Machine Learning,” 2023. [Interactiv]. Available: <https://www.thc.ro/blog/ce-este-machine-learning/>.
- [8] G. B. a. E. Scornet., *A random forest guided tour.*, 2016.
- [9] „Coursera-DecisionTree,” 2023. [Interactiv]. Available: <https://www.coursera.org/articles/decision-tree-machine-learning>.
- [10] GradientBoosted, „Friedman, J. H. "Greedy Function Approximation: A Gradient Boosting Machine".,” 2018.
- [11] „RapidMiner docs,” 2024.
- [12] [Interactiv]. Available: <https://www.ceew.in/publications/sustainable-agriculture-india/precision-farming>.