**CARBON FOOTPRINT**

DOCUMENTAȚIE PENTRU „ANALIZA DATELOR”

|  |  |  |
| --- | --- | --- |
|  | Student: | **Balog David Alexandru** |
|  |  |  |
|  |  |  |

**2024**

[Capitolul 1. Introducere 1](#_Toc129864352)

[Capitolul 2. Contextul bazei de date și al proiectului 2](#_Toc129864353)

[Capitolul 3. Aspecte teoretice relevante 3](#_Toc129864354)

[3.1. Definirea amprentei de carbon 3](#_Toc129864356)

[3.2. Metode de calcul și măsurare a amprentei de carbon 3](#_Toc129864356)

[3.3. Impactul amprentei de carbon asupra mediului și sănătății umane 3](#_Toc129864356)

[3.4. Strategii și soluții pentru reducerea amprentei de carbon 4](#_Toc129864356)

[3.5. State of the art în măsurarea și gestionarea amprentei de carbon 4](#_Toc129864356)

[Capitolul 4. Implementarea aspectelor teoretice în cadrul proiectului 5](#_Toc129864355)

[4.1. Preprocesarea datelor 5](#_Toc129864356)

[4.2. Explorarea datelor 6](#_Toc129864356)

[4.3. Pregătirea datelor pentru modelare 6](#_Toc129864356)

[4.4. Construirea modelelor predictive 7](#_Toc129864356)

[Capitolul 5. Testare și validare 9](#_Toc129864358)

[5.1. Metodologia de testare 9](#_Toc129864356)

[5.2. Rezultatele testării 9](#_Toc129864356)

[Capitolul 6. Rezultate 12](#_Toc129864359)

[6.1. Analiza performanței modelelor 12](#_Toc129864356)

[6.2. Interpretarea rezultatelor 12](#_Toc129864356)

[6.3. Recomandări 12](#_Toc129864356)

[Capitolul 7. Concluzii 13](#_Toc129864360)

[Bibliografie 14](#_Toc129864362)

# Introducere

Amprenta de carbon, denumită și amprenta CO2, reprezintă emisiile totale de gaze cu efect de seră pe care o organizație, un eveniment, produs sau o persoană le produce într-un anumit interval de timp.

Gazele cu efect de seră sunt rezultatele diverselor activitati pe care omul le intreprinde ca urmare a dezvoltării și progresului: transport, consumul de combustibili, alimente, produse fabricate, materiale, lemn, drumuri, clădiri, servicii etc.

Scopul acestui proiect este de a dezvolta un sistem pentru monitorizarea și analiza amprentei de carbon. Prin utilizarea unei baze de date adecvate acestui proiect, vom colecta, stoca și analiza date referitoare la emisiile de carbon, facilitând astfel luarea deciziilor informate în vederea reducerii impactului negativ asupra mediului în care trăim.

Am ales această bază de date deoarece selecția de date este foarte bună în contextul proiectului, oferind o imagine detaliată a diferiților factori care contribuie la emisiile de carbon. Fiecare coloană reprezintă un aspect important al vieții de zi cu zi care poate influența amprenta de carbon a unui individ. Analiza detaliată a acestor variabile ne va permite să identificăm zonele cu cel mai mare impact și să propunem măsuri eficiente pentru reducerea amprentei de carbon. Baza de date dispune de foarte multe intrări, analizele și predicțiile fiind mult mai precise și mai exacte.

# Contextul bazei de date și al proiectului

Emisiile de carbon sunt un mare factor care ne influențeaza viața de zi cu zi. Cunoașterea acestor date este esențială pentru a preveni și a gestiona impactul negativ al schimbărilor climatice asupra mediului înconjurător și asupra sănătății noastre.

Înțelegerea motivelor pentru care reducerea amprentei de carbon contează cuprinde mai multe aspecte semnificative. În primul rând, contribuie direct la eforturile globale împotriva schimbărilor climatice, acest lucru ajutand la atenuarea fenomenelor meteorologice extreme si în reducerea temperaturilor în creștere.

Reducerea emisiilor de carbon se traduce în conservarea resurselor naturale ale planetei, promovând sustenabilitatea în domenii legate de apă, energie și materii prime. Acest lucru nu numai că aduce beneficii mediului, ci și promovează eficiența resurselor pentru generațiile care vor urma.

Scăderea emisiilor are un impact pozitiv asupra sănătății fiecarui individ, prin îmbunătățirea calității aerului și a sănătății fizice generale. Adoptarea practicilor durabile, inclusiv măsurarea și gestionarea amprentei de carbon, încurajează alegerile inteligente cu mediul și contribuie la construirea unei societăți mai durabile.

În cele din urmă, recunoașterea și abordarea impactului propriu de carbon reflectă responsabilitatea socială și ambientală, subliniind importanța adoptării unor măsuri proactive pentru un viitor mai durabil.

Această bază de date a fost creată pentru a înregistra și analiza informații referitoare la diferite aspecte ale vieții individuale și amprenta de carbon asociată acestora. Datele colectate includ informații precum tipul de corp, sexul, dieta, frecvența dușului, sursa de energie pentru încălzire, mijloacele de transport, tipul de vehicul, activitatea socială, cheltuielile lunare cu alimentele, frecvența călătoriilor cu avionul, distanța lunară parcursă cu vehiculul, dimensiunea sacului de gunoi, numărul săptămânal de saci de gunoi, timpul petrecut zilnic în fața televizorului sau a calculatorului, numărul de haine noi achizitionate lunar, timpul petrecut zilnic pe internet, eficiența energetică, reciclarea, modul de gătit și emisiile de carbon. Baza de date a fost descărcată de pe site-ul Kaggle prin următorul link: <https://www.kaggle.com/datasets/dumanmesut/individual-carbon-footprint-calculation> .

Prin utilizarea acestei baze de date și a analizelor asociate, ne propunem să identificăm tendințe și modele în comportamentul populației în ceea ce privește emisiile de carbon și să dezvoltăm soluții personalizate pentru reducerea acestora. În final, dorim să contribuim la eforturile globale de combatere a schimbărilor climatice și protejare a mediului înconjurător pentru generațiile viitoare.

# Aspecte teoretice și relevante

## Definirea amprentei de carbon:

Amprenta de carbon reprezintă cantitatea totală de gaze cu efect de seră emise direct sau indirect în timpul unei activități specifice, cum ar fi producția de bunuri, consumul de energie sau deplasările cu mijloace de transport. Aceasta este exprimată în echivalent de dioxid de carbon (CO2e), pentru a permite o comparație uniformă a impactului diferitelor gaze cu efect de seră.

**Referințe:**

* IPCC (2006). "Guidelines for National Greenhouse Gas Inventories."
* Wiedmann, T., Minx, J. (2008). "A definition of 'carbon footprint'." Ecological Economics, 65(1), 1-5.

## Metode de calcul și măsurare a amprentei de carbon

Există mai multe metode și instrumente disponibile pentru calcularea și măsurarea amprentei de carbon, care variază în complexitate și precizie. Acestea includ analiza ciclului de viață, evaluările de la poartă la poartă, precum și diverse standarde și ghiduri internaționale.

**Referințe:**

* ISO 14040:2006, Environmental management - Life cycle assessment - Principles and framework.
* GHG Protocol, "Corporate Accounting and Reporting Standard."

## Impactul amprentei de carbon asupra mediului și sănătății umane

Emisiile de carbon și creșterea concentrației de gaze cu efect de seră în atmosferă sunt asociate cu schimbările climatice, care pot avea efecte devastatoare asupra ecosistemelor, biodiversității și sănătății umane. Aceste efecte includ creșterea temperaturilor globale, modificări ale regimurilor de precipitații, creșterea nivelului mărilor și intensificarea fenomenelor meteorologice extreme.

**Referințe:**

* IPCC (2018). "Global Warming of 1.5°C."
* Watts, N. et al. (2019). "The Lancet Countdown on health and climate change: ensuring that the health of a child born today is not defined by a changing climate." The Lancet, 394(10211), 1836-1878.

## Strategii și soluții pentru reducerea amprentei de carbon:

Există o varietate de strategii și soluții disponibile pentru reducerea amprentei de carbon, care includ eficientizarea energetică, adoptarea surselor de energie regenerabilă, promovarea transportului durabil, minimizarea deșeurilor și încurajarea consumului responsabil.

**Referințe:**

* Creutzig, F. et al. (2018). "Towards demand-side solutions for mitigating climate change." Nature Climate Change, 8(4), 268-271.
* Rogelj, J. et al. (2018). "Scenarios towards limiting global mean temperature increase below 1.5°C." Nature Climate Change, 8(4), 325-332.

## State of the art în măsurarea și gestionarea amprentei de carbon:

Cercetările recente și inovațiile tehnologice au condus la dezvoltarea unor instrumente și tehnici mai precise și mai eficiente pentru măsurarea și gestionarea amprentei de carbon. Acestea includ utilizarea senzorilor IoT, analiza big data, modelele de predicție avansate și tehnologiile blockchain.

**Referințe:**

* Marinakis, V. et al. (2020). "Internet of Things and Blockchain: A Systematic Review in the Era of Industry 4.0." Sensors, 20(8), 2297.
* Mastrucci, A. et al. (2020). "Big data and the environment: A literature review on opportunities, challenges, and changes." Journal of Big Data, 7(1), 1-29.

# Implementarea aspectelor teoretice în cadrul proiectului

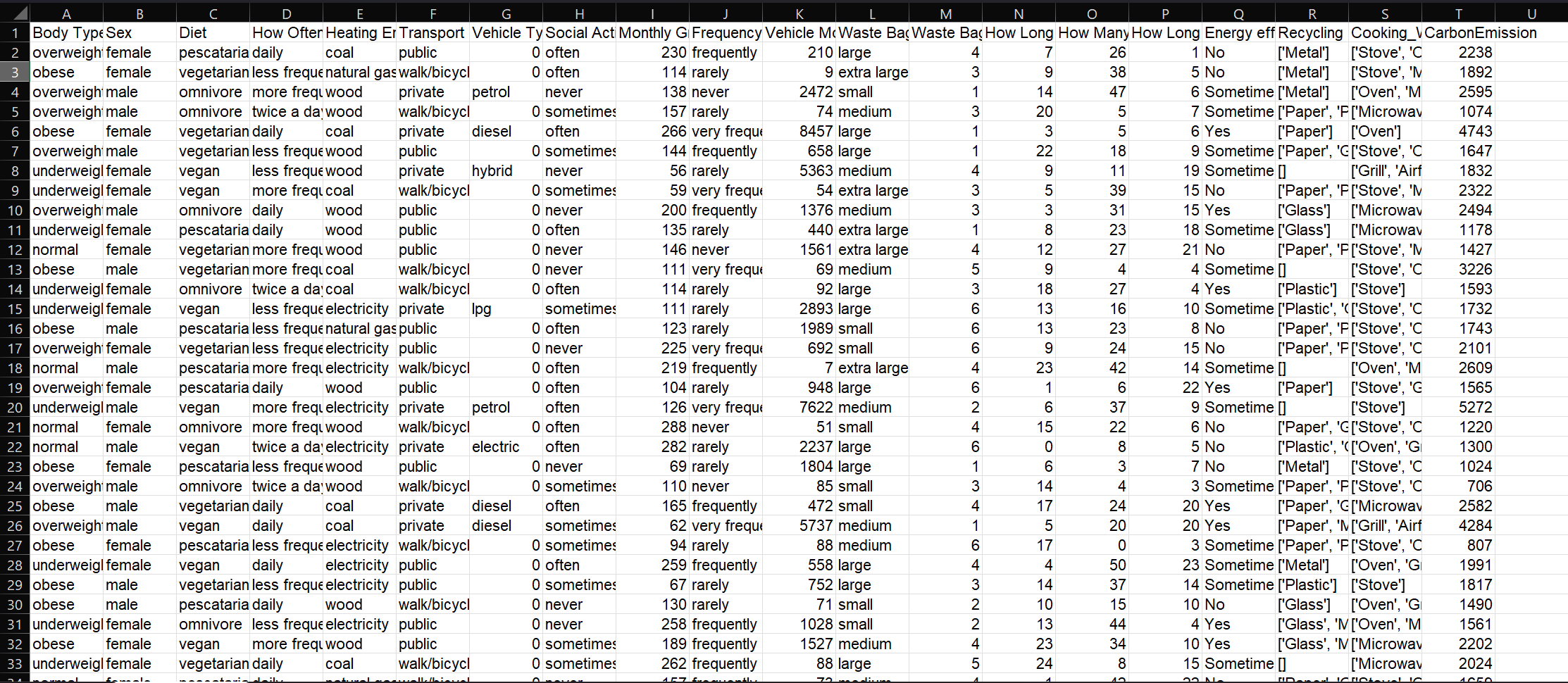
## Preprocesarea datelor

Pentru a pregăti datele pentru analiză în RapidMiner Ai Studio, am realizat mai mulți pași de preprocesare:

**Importul datelor:** Am încarcat baza de date Carbon Emission în aplicatie, aceasta fiind preluata de pe site-ul „Kaggle” din următorul link: <https://www.kaggle.com/datasets/dumanmesut/individual-carbon-footprint-calculation> .

**Conversia categoriilor în valori numerice:** Unele coloane, cum ar fi Body Type, Sex, Diet, etc., sunt categorice. Am utilizat operatorul „Nominal to Numerical” în RapidMiner pentru a le converti în valori numerice.

**Normalizarea datelor:** Am normalizat datele utilizând operatorul „Multiply”, pentru a asigura că toate variabilele au aceeași unitate de măsură și contribuie echitabil la model.



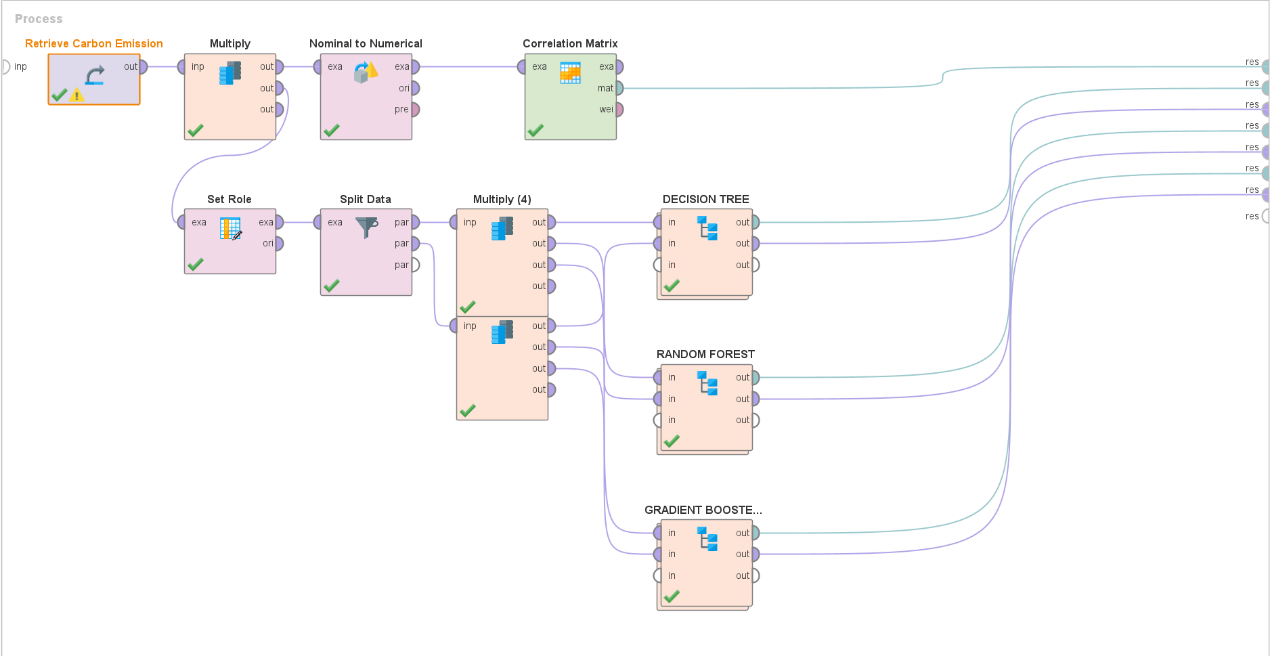
**Figura 4.1. Baza de date Carbon Footprint**

## Explorarea datelor

Explorarea inițială a datelor a fost realizată folosind:

**Matricea de corelație:** Am utilizat operatorul „Correlation Matrix” pentru a identifica relațiile dintre diferite variabile și amprenta de carbon. Aceasta ne-a permis să identificăm factorii cu cel mai mare impact asupra amprentei de carbon.

**Vizualizări grafice:** Am creat diverse grafice și diagrame pentru a înțelege distribuția și relațiile dintre variabile.



**Figura 4.2. Procesul din RapidMiner Ai Studio**

## Pregătirea datelor pentru modelare

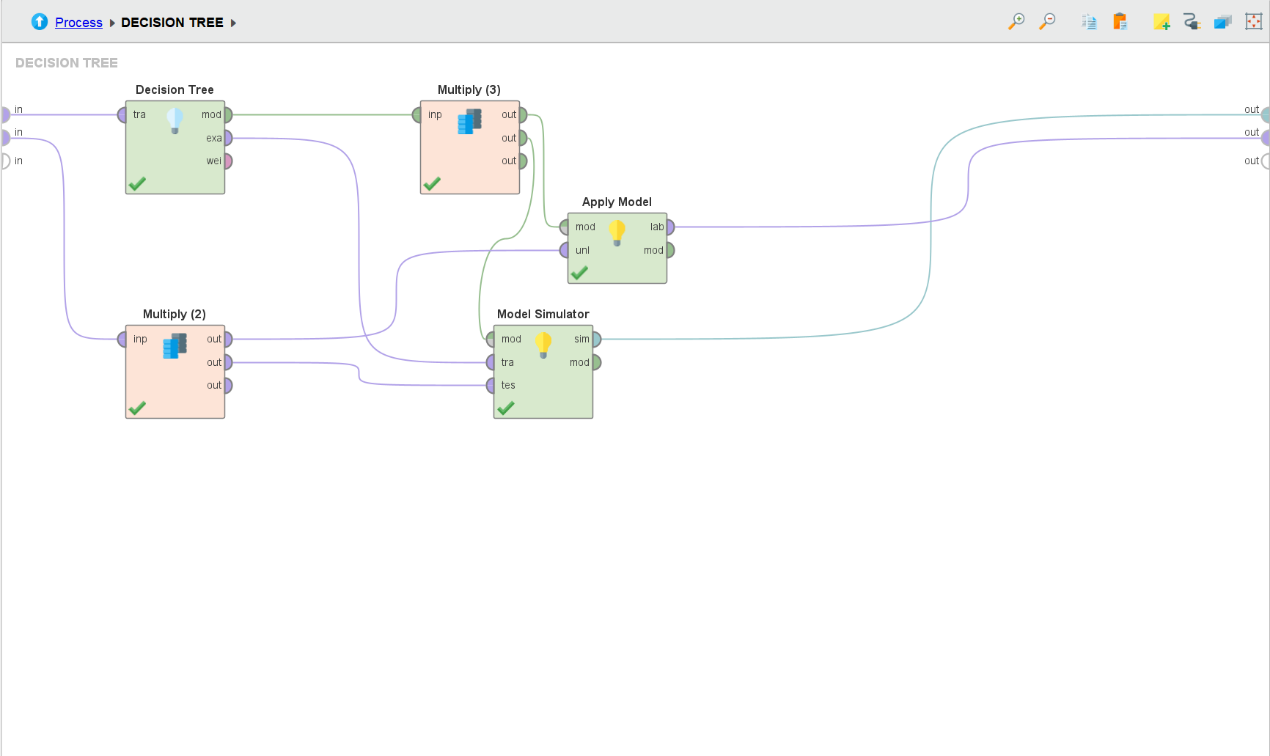
**Setarea rolului variabilelor:** Am utilizat operatorul „Set Role” pentru a specifica variabila țintă (CarbonEmission).

**Împărțirea datelor:** Datele au fost împărțite în seturi de antrenament și testare folosind operatorul „Split Data”, asigurându-ne că modelul este testat pe date neutilizate în antrenament, acestea având rația de 0,3 si 0,7.

## Construirea modelelor predictive

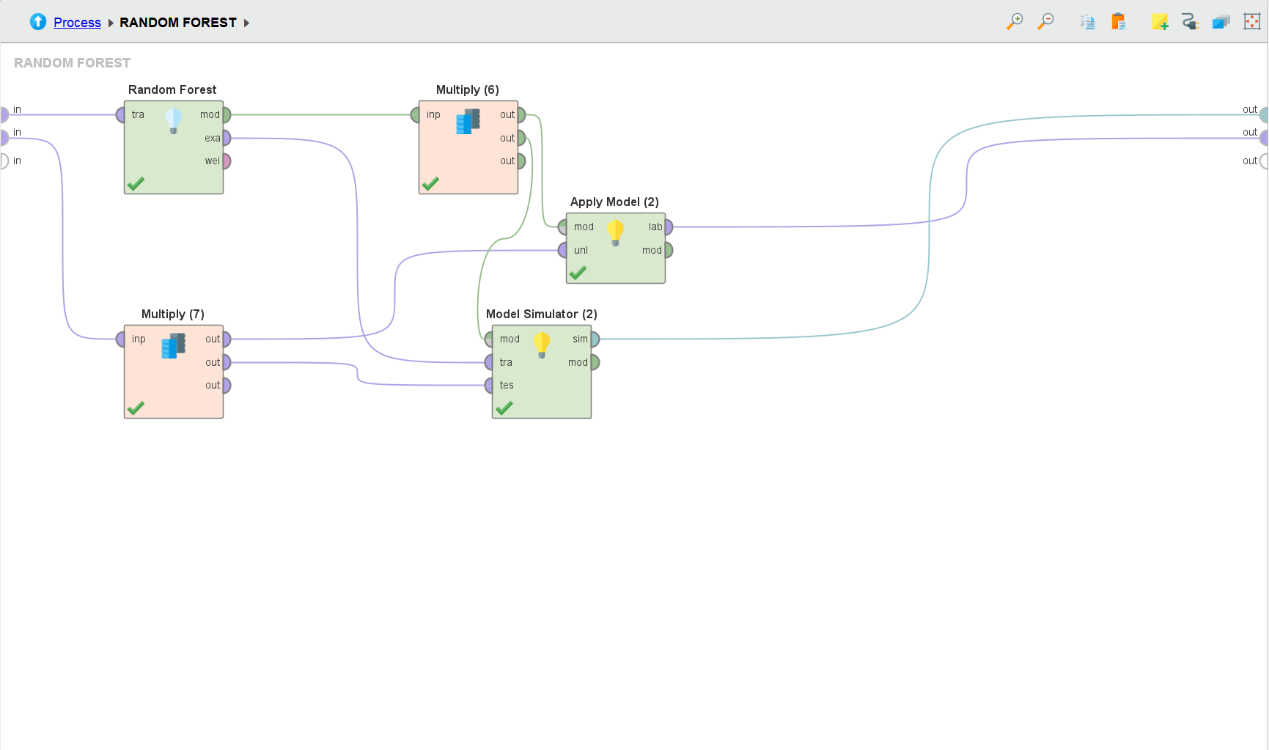
Am utilizat mai multi algoritmi de învățare automată pentru a construi modele predictive și pentru a compara performanța acestora:

1. **Arbore de decizie:** Am folosit operatorul „Decision Tree” pentru a crea un model explicabil și ușor de interpretat, care ne permite să vizualizăm criteriile de decizie utilizate pentru predicția amprentei de carbon.



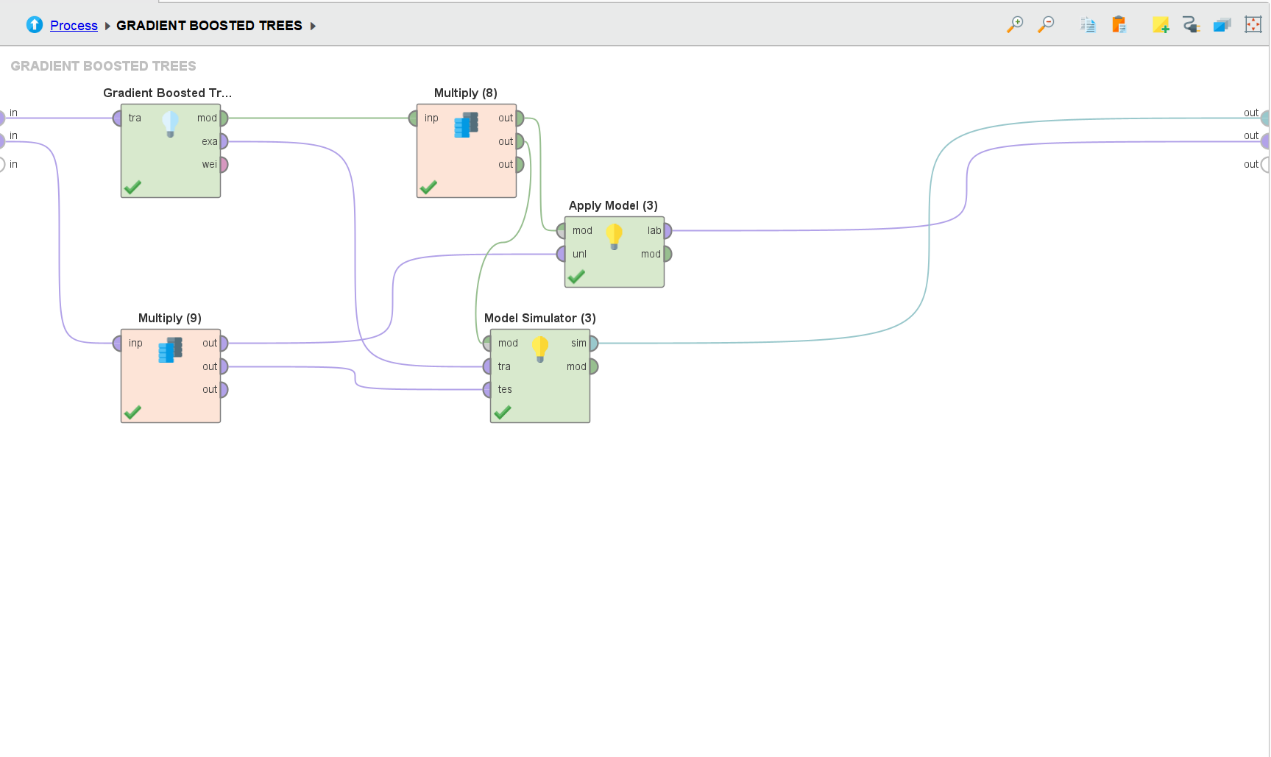
**Figura 4.3. Decision Trees**

1. **Pădure aleatorie:** Am folosit operatorul ”Random Forest” pentru a construi un model robust, care combină predicțiile mai multor arbori de decizie pentru a reduce riscul de overfitting.



**Figura 4.4. Random Forest**

1. **Gradient Boosted Trees:** Am utilizat operatorul „Gradient Boosted Trees” pentru a construi un model performant care corectează greșelile modelelor anterioare și oferă predicții precise.



**Figura 4.5. Gradient Boosted Trees**

# Testare și validare

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

## Metodologia de testare

Testarea a fost făcută cu operatorul “Split Data” pentru a împărți datele în două seturi, cel de antrenament (procentaj de 70%) și cel pentru testare (procentaj de 30%). Metoda de sampling a acestui operator a fost ajustată automat.

Am ales acesta metoda deoarece ne ajută la evaluarea performanței modelelor predictive, asigurându-se că modelele nu sunt supraînvățate (overfitting) și că pot generaliza bine pe date noi.

## Rezultatele testării

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

* 1. Regression Tree Model

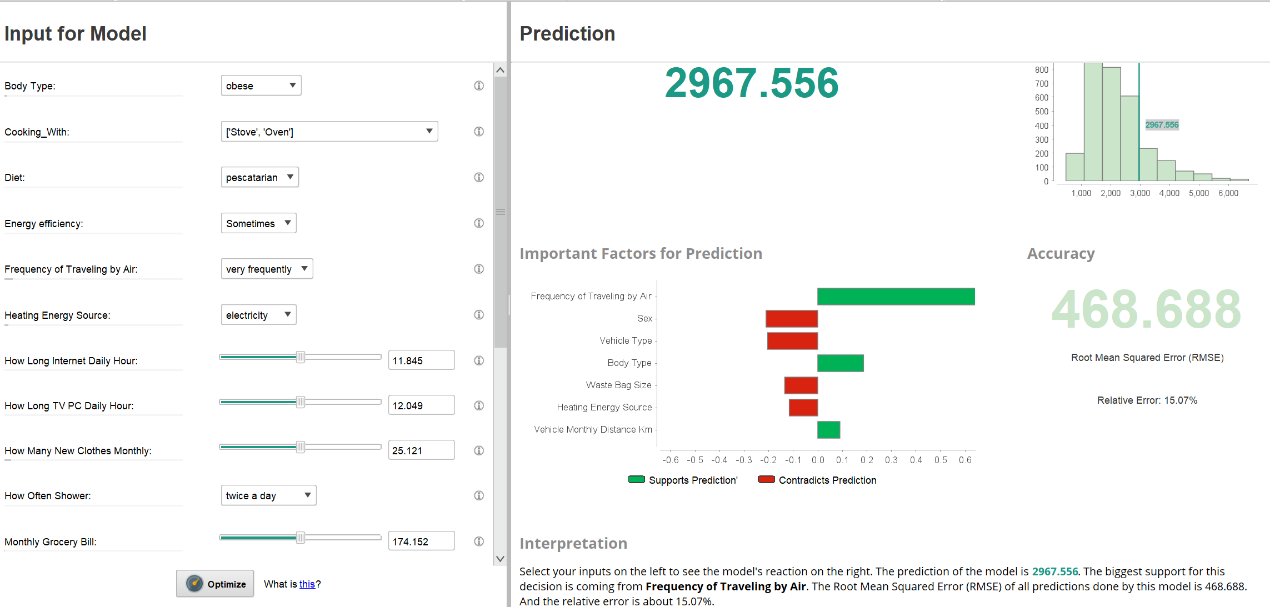
Valoare predicție: 2967.556

Root Mean Squared Error (RMSE): 468.688

Eroarea Relativă: 15.07%

Factorii importanți care suportă predicția sunt: frecvența călătoriilor aeriene, tipul corpului, distanța parcursă pe lună în km.

Factorii importanți care nu suportă predicția sunt: sex, tipul vehicului, marimea sacului de gunoi și sursa de energie.



* 1. Random Forest Model

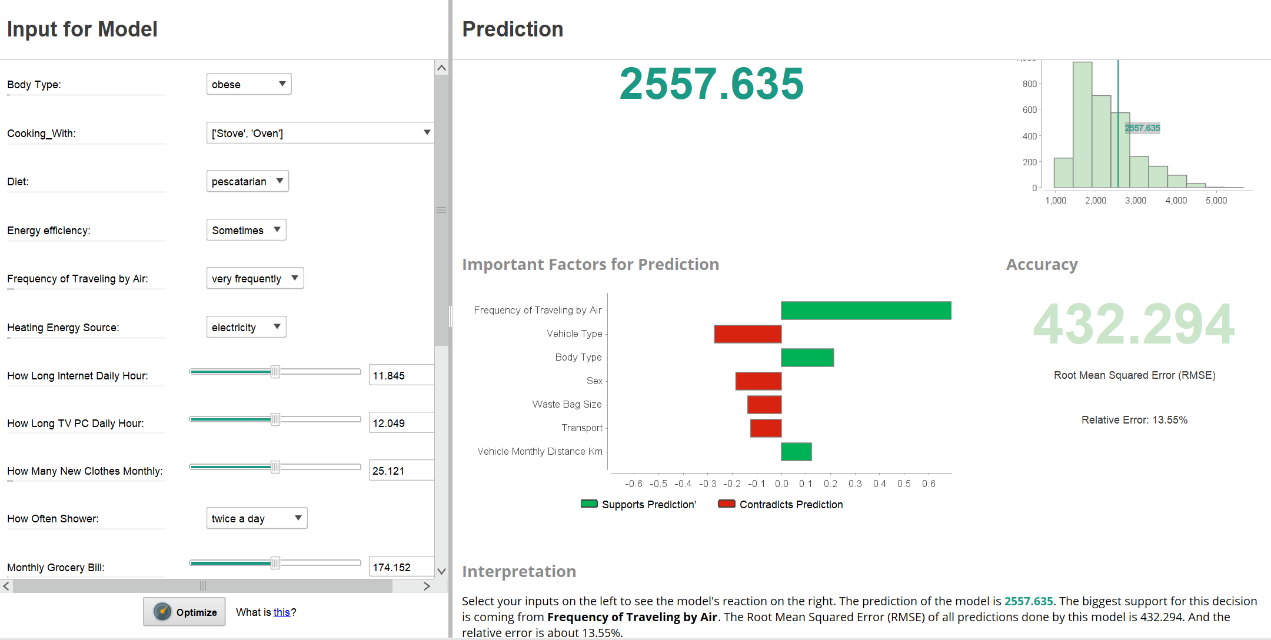
Valoare predicție: 2557.635

Root Mean Squared Error (RMSE): 432.294

Eroarea Relativă: 13.55%

Factorii care suportă predicția sunt: frecvența călătoriilor aeriene, tipul corpului, distanța parcursă pe lună în km.

Factorii importanți care nu suportă predicția sunt: tipul vehicului, sex, marimea sacului de gunoi și transportul.



* 1. Gradient Boosted Model

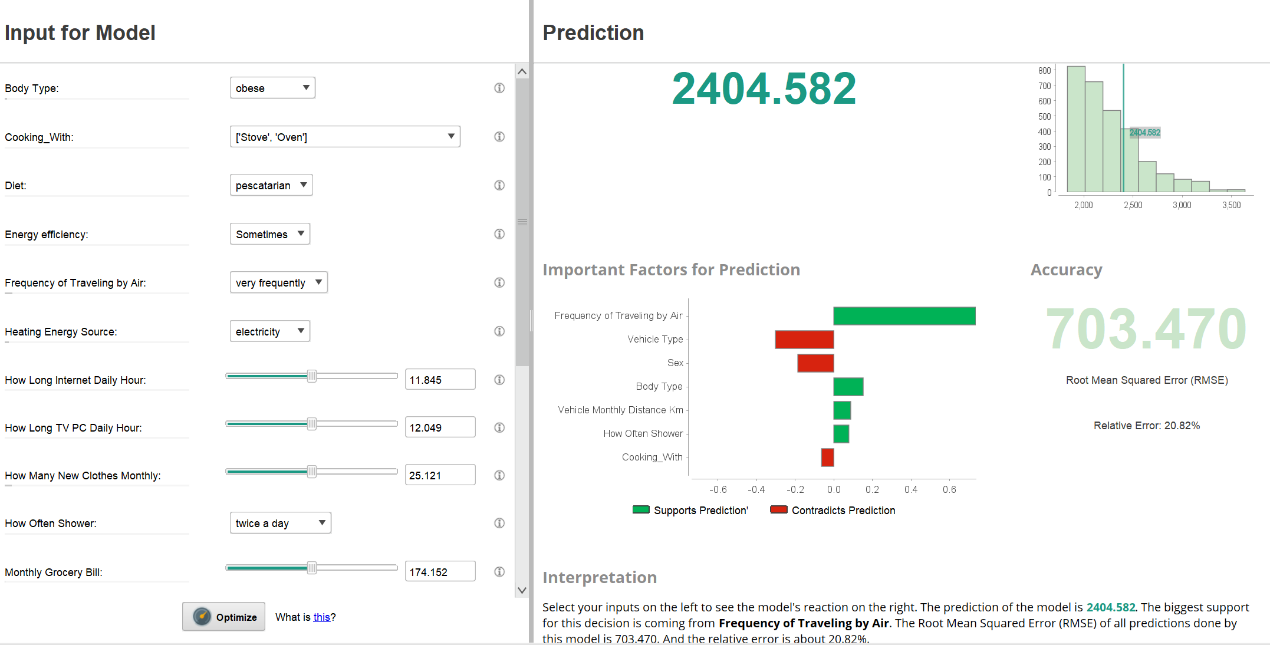
Valoare predicție: 2404.582

Root Mean Squared Error (RMSE): 703.470

Eroarea Relativă: 20.82%

Factorii importanți: Frecvența călătoriilor aeriene, tipul corpului, distanța parcursă pe lună în km, cât de des se face duș.

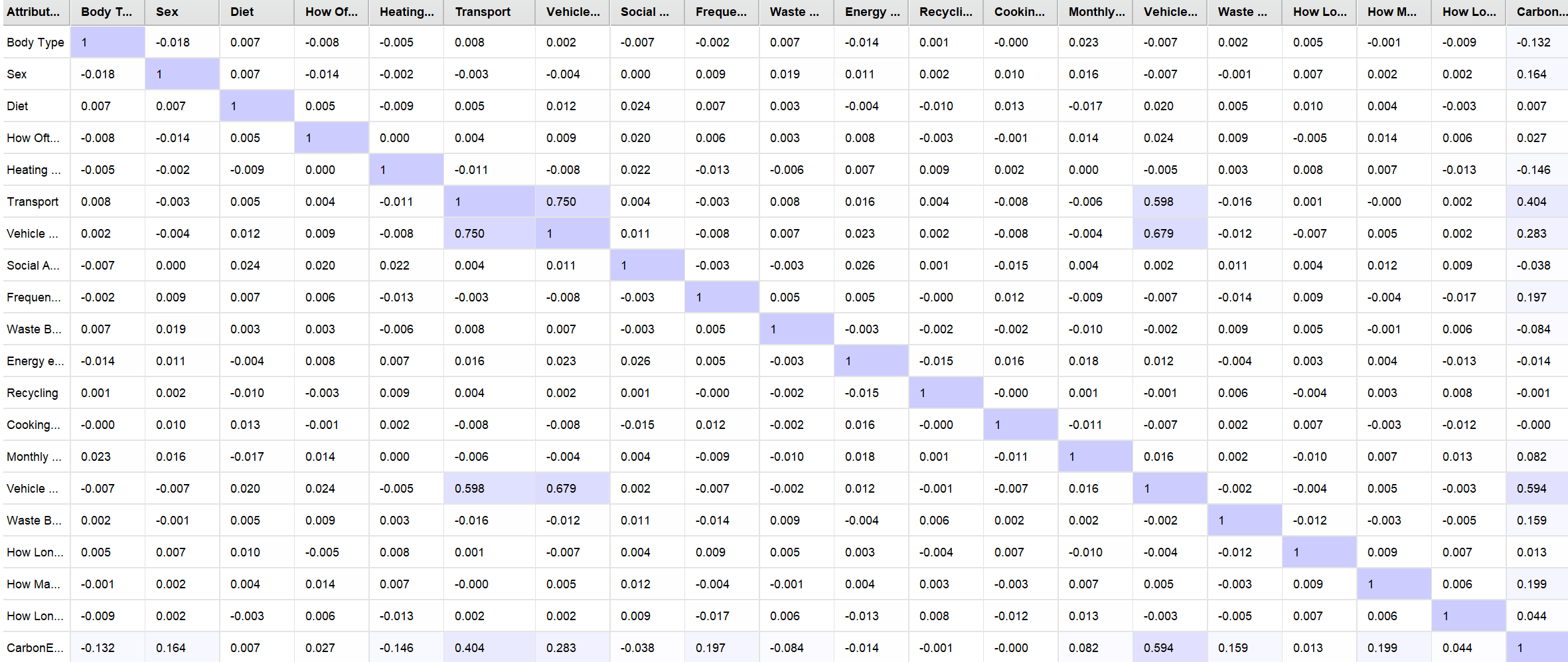
Factorii importanți care nu suportă predicția sunt: tipul vehicului, sex, cu ce se gătește mâncarea.



* 1. Matricea de corelații

Corelațiile importante extrase:

* Carbon Emission – Transport : 0.404 (corelație slabă)
* CarbonEmission – Vehicle Monthly Distance Km : 0.594 (corelație moderată)
* Carbon Emission – Vehicle Type : 0.283 (corelație slabă)
* Vehicle Type – Transport : 0.750 (corelație puternică)
* Vehicle Monthly Distance Km – Transport : 0.598 (corelație moderată)
* Vehicle Monthly Distance Km – Vehicle Type : 0.679 (corelație moderată)



# Rezultate

## Analiza performanței modelelor

Din rezultatele testării, observăm că modelul Random Forest a avut cea mai bună performanță, cu un RMSE mai mic și o eroare relativă mai mică în comparație cu Regression Tree și Gradient Boosted Model. Random Forest Model a oferit cea mai precisă predicție a amprentei de carbon.

## Interpretarea rezultatelor

Factorii cu cel mai mare impact asupra amprentei de carbon au fost frecvența călătoriilor aeriene, tipul corpului și distanța parcursă pe lună în km. Aceste variabile au fost constante în toate modelele testate, indicându-le ca fiind predictori cheie ai amprentei de carbon.

## Recomandări

Propuneri bazate pe rezultate pentru reducerea amprentei de carbon:

* Reducerea frecvenței călătoriilor aeriene: Având în vedere impactul major al acestui factor, reducerea zborurilor frecvente poate scădea semnificativ amprenta de carbon.
* Optarea pentru vehicule mai eficiente din punct de vedere al consumului de combustibil: Utilizarea vehiculelor cu consum redus sau a celor electrice.
* Reducerea consumului de energie și deșeuri: Implementarea unor măsuri pentru a minimiza deșeurile și a îmbunătăți eficiența energetică.

# Concluzii

În acest proiect, am explorat și analizat amprenta de carbon a indivizilor folosind o bază de date extinsă și diverse metode predictive. Scopul principal a fost de a dezvolta un sistem pentru monitorizarea și analiza amprentei de carbon, utilizând o bază de date adecvată pentru a colecta, stoca și analiza date referitoare la emisiile de carbon. Înțelegerea acestor date este esențială pentru a preveni și gestiona impactul negativ al schimbărilor climatice asupra mediului și sănătății umane.

Analiza rezultatelor a evidențiat că factorii cheie care influențează amprenta de carbon sunt frecvența călătoriilor aeriene, tipul corpului și distanța parcursă pe lună în kilometri. Aceste variabile au fost constante în toate modelele testate, indicându-le ca predictori esențiali ai amprentei de carbon. Identificarea acestor factori critici oferă o bază solidă pentru intervenții țintite care pot contribui la reducerea emisiilor de carbon.

Rezultatele acestui proiect au implicații semnificative pentru politicile de mediu și pentru strategiile individuale și organizaționale de reducere a amprentei de carbon. Studiile viitoare ar putea să se concentreze pe extinderea setului de date pentru a include mai multe variabile și pe utilizarea unor tehnici avansate de inteligență artificială pentru a îmbunătăți și mai mult acuratețea predicțiilor.

În concluzie, acest proiect demonstrează importanța și potențialul utilizării analizei de date și a modelării predictive pentru a înțelege și aborda provocările asociate cu emisiile de carbon. Prin aplicarea acestor tehnici, putem contribui la eforturile globale de combatere a schimbărilor climatice și la protejarea mediului pentru generațiile viitoare.

# Bibliografie

|  |  |
| --- | --- |
|  | -Baza de date : https://www.kaggle.com/datasets/dumanmesut/individual-carbon-footprint-calculation |
|  | -https://carbonexpert.ro/amprenta-co2/ce-este-amprenta-de-carbon/ |
|  | -https://infocons.ro/ce-este-amprenta-de-carbon-si-cum-se-calculeaza/ |
|  |  |
|  | -https://docs.rapidminer.com/latest/studio/index.html |
|  | -https://en.wikipedia.org/wiki/Greenhouse\_gas\_emissions  -https://en.wikipedia.org/wiki/Carbon\_footprint |