Smarandescu Antonio-Stefan

Grupa 333AA

**Estimarea si predictia consumului de energie pentru incalzirea cladirilor**

* INTRODUCERE

**\*\*Context si Motivatie :\*\***

Intr-o lume tot mai preocupata de sustenabilitate si eficienta energetica, cladirile joaca un rol esential. O parte semnificativa din consumul global de energie provine din incalzirea spatiilor interioare, iar estimarea, predictia si nu in ulrimul rand, optimizarea acestui consum poate reduce costurile si amprenta de carbon.

**\*\*Descrierea generala si scopul proiectului :\*\***

Acest proiect isi propune sa prezica necesarul de energie termica pentru incalzirea in parametri normali ai unei cladiri, pe baza caracteristicilor sale constructive (dimensiuni, orientare, suprafata vitricata). Prin furnizarea unei estimari rapide, modelul poate ajuta :

* + - Arhitectii si proiectantii : pot identifica mai usor parametrii cheie ce influenteaza consumul pentru a proiecta cladiri mai eficiente termic
    - Inginerii : pot anticipa costurile si adopta strategii de optimizare a consumului

**\*\*Alegerea algorimului :\*\***

Pentru acest proiect, am folosit un algoritm de regresie liniara, un instrument simplu, dar eficient, care ne arata cum se schimba consumul de energie odata cu modificarea caracteristicilor cladirii in contextul specific al datelor din setul de date ENB2012\_data(Energy Efficiency in Buildins).

De asemenea, acest algoritm necesita putina putere de calcul si ofera o baza pentru extinderea ulterioara cu metode mai complexe, daca este cazul.

Totodata, regresia liniara este usor de evaluat folosind metrici comune, cum ar fi MSE(Mean Squared Error), R-squared, etc. care sunt calculate pentru seturile de antrenare, validare si test. Aceste metrici ajuta la evaluarea eficientei modelului si la cat de bine se pot face predictiile.

* **Metodologie**

\*\* Descriere implementare si particularitati algoritm :\*\*

* Setul de date

Datele provin dintr-un fisier EXCEL (ENB2012\_data.xslx) ce contine 8 caracteristici pentru fiecare cladire :

# X1 = Relative Compactness

# !X2 = Surface Area!

# !X3 = Wall Area!

# X4 = Roof Area

# X5 = Overall Height

# X6 = Orientation

# X7 = Glazing Area

# X8 = Glazing Area Distribution , si 2 tinte: Y1(Heating Load) si Y2(Cooling Load). Ne vom concentra pe Y1(necesarul de incalzire).

* Implementare algoritm

1. Importam bibliotecile : Pandas(Citirea datelor din fisiere, manipularea datelor), Numpy(calcul numeric), Matplotlib(vizualizarea datelor- grafice),Scikit-learn(algoritmi de invatare automata – folosit pentru crearea unui model de regresie liniara).
2. Citim datele dintr-un fisier Excel(‘df = pd.read\_excel(data\_path’) – ca si cum ai deschide o carte cu informatii despre mai multe casute.
3. Separam ce vrem sa aflam(tinta -Y1- cata energie trebuie pentru ca incaperea sa fie calda) de infomatiile de intrare(detalii despre incapere).
4. Variabilele categorice (X6, X8) le transformam in numere simple(0,1), folosind OneHotEncoder.
5. Scalarea caracteristicilor numerice, folosind ‘StandardScaler’ pentru a aduce toate valorile la aceeasi scara, imbunatatind stabilitatea numerica a modelului.
6. Impartirea Datelor:

\*\*Datele sunt impartite in 3 subseturi:

1.Antrenare(60%) – pentru invatarea modelului(ex. cand desenezi un cerc, repeti de mai multe ori pana stii)

2.Validare(20%) – pentru ajustarea modelului(ex. te verifici cu un prieten sa vezi daca ai desenat bine cercul)

3.Testare(20%) – pentru evaluarea finala, pe date complet noi(ex. un strain te pune sa faci un cerc, iar tu trebuie sa i-l arati si el iti spune daca e bine sau nu).

1. Antrenarea modelului

Se foloseste regresia liniara din scikit-learn. Acest model ajusteaza coeficientii pentru a minimiza eroarea patratica medie(MSE) intre valorile prezise si cele reale ale lui Y1(Heating Load). Aici, modelul invata: “ Daca o casuta are atata suprafata, atata inaltime, atata orientare, atunci probabil are nevoie de atata caldura”.

Facem predictii: Dupa ce modelul a invatat, ii dam casutele din setul de test si il intrebam : ‘Cata caldura crezi ca trebuie pentru acestea’, iar el ghiceste.

1. Evaluarea modelului prin calcularea erorilor(MSE,MAE,RMSE, R²). R² arata cat de bine explica modelul datele. ‘R² mare e ca o nota mare la scoala: ai inteles bine lectia’ . Daca R² mare si erorile mici, atunci modelul deseneaza un cerc aproape perfect ( ghiceste bine – in cazul nostru).
2. Realizam graficele

1.Vizualizare (Date Scalate): Surface Area vs Heating Load

#In graficul scalat, Surface Area este in unitati standardizate (aproape de 0)

A graph of a graph showing a surface area and heating load

Description automatically generated

2. Vizualizare (Date Nescalate): Surface Area vs Heating Load

#In graficul nescalat, Surface Area este in unitatile originale

A graph of a graph showing a comparison of a surface area and heating load

Description automatically generated

3. # Regresie univariată: Wall Area (X3) vs Y1

A graph with a red line

Description automatically generated

* **Rezultate obtinute si interpretare**

\*\*Performanta pe antrenare:\*\*

* MSE: 7.40
* MAE: 1.95
* RMSE: 2.72
* R²: 0.93
* **MSE (Mean Squared Error):**  
  Este media pătratelor diferențelor dintre valorile reale ale țintei (Heating Load) și valorile prezise de model. Un MSE de 7.40 înseamnă că, în medie, pătratul erorii (diferența dintre valoarea prezisă și cea reală) este 7.40.

Cu cât MSE este mai mic, cu atât modelul este mai precis în medie (erorile sunt mai mici).

* **MAE (Mean Absolute Error):**  
  Este media valorilor absolute ale erorilor. Un MAE de 1.95 înseamnă că, în medie, modelul greșește cu aproximativ 1.95 unități la predicția Heating Load.  
  Față de MSE, MAE este mai ușor de interpretat deoarece se află pe aceeași scală ca și variabila țintă.

Un MAE mic înseamnă că diferența medie dintre predicții și valori reale este mică.

* **RMSE (Root Mean Squared Error)**:  
  Este radicalul din MSE. Aici, 2.72 reprezintă o măsură similară cu MSE, dar readusă la scara originală a datelor. RMSE este adesea mai intuitiv, deoarece are aceeași unitate ca ținta (Heating Load).

Un RMSE mic indică erori mici.

* **R² (R-squared):**  
  Indică proporția din variația valorilor reale care este explicată de model. Un R² de 0.93 înseamnă că 93% din variația valorilor Heating Load din datele de antrenare este explicată de model. Un R² apropiat de 1 sugerează un model foarte bun pe datele de antrenare.

Pe scurt, aceste valori arată că modelul a învățat bine pe datele de antrenare (R² mare și erori relativ mici), ceea ce înseamnă că liniile și coeficienții obținuți se potrivesc destul de bine cu exemplele pe care le-a văzut.

\*\*Performanta pe validare :\*\*

* MSE: 8.87
* MAE: 2.15
* RMSE: 2.98
* R²: 0.92

R² de 0.92 față de 0.93 pe antrenare arată că modelul încă explică o mare parte din variația datelor și pe setul de validare. Deci, modelul generalizează destul de bine, pierzând doar puțin din puterea explicativă.

Performanța pe validare este foarte apropiată de cea de antrenare. Aceasta este o veste bună: modelul nu pare să fie supra-antrenat și se descurcă bine chiar și pe date noi (de validare).

\*\*Performanta pe testare :\*\*

* MSE: 7.89
* MAE: 2.04
* RMSE: 2.81
* R²: 0.92

**R²: 0.92** - Similar cu valorile de pe setul de validare (0.92) și foarte apropiat de cel pe antrenare (0.93), semn că modelul explică aproape la fel de bine variația datelor noi ca și pe cele cu care a fost obișnuit.

Performanța pe testare este consistentă cu cea pe antrenare și validare, ceea ce arată că modelul generalizează bine și nu s-a supra-antrenat. Pe scurt, modelul este stabil și robust, reușind să prezică necesarul de încălzire cu precizie aproape la fel de bună și pe date complet noi.

Matricea de confuzie nu este aplicabila direct, deoarece nu avem clasificare.

Ecuația dreptei pentru Wall Area vs Heating Load: y = 0.11 \* x + -11.26

* **0.11** este panta (coeficientul) dreptei. Asta înseamnă că, dacă Wall Area crește cu 1 unitate, modelul prezice că Heating Load va crește, în medie, cu 0.11 unități.
* **-11.26** este interceptul, punctul în care linia ar tăia axa Y dacă Wall Area ar fi 0. Practic, dacă nu am avea deloc perete (Wall Area = 0), modelul ar prezice un Heating Load de -11.26. Acest rezultat, deși nerealist din punct de vedere fizic (nu poți avea necesar de încălzire negativ), apare deoarece linia încearcă să se potrivească cel mai bine datelor existente. În realitate, interceptul nu are întotdeauna o interpretare practică directă, mai ales dacă domeniul de valori analizat nu include acel punct (Wall Area = 0).

Pe scurt, această ecuație arată cum variază aproximativ Heating Load atunci când modificăm numai Wall Area, ignorând restul factorilor. Este o simplificare care ajută la înțelegerea relației dintre aceste două variabile.

* **Concluzii**
* Modelul de regresie liniară folosit în acest proiect a demonstrat o capacitate solidă de a prezice necesarul de încălzire (Heating Load) pentru clădiri pe baza parametrilor analizați. Valorile R² ridicate (peste 0,9) și erorile moderate (MAE în jur de 2 unități și RMSE sub 3 unități) sugerează că, în practică, acest model poate oferi o estimare utilă a consumului energetic pentru încălzire, încă de la etapa de proiectare a unei clădiri.
* Acest lucru înseamnă că, fără a efectua măsurători reale și costisitoare după construirea efectivă a clădirii, se pot lua decizii timpurii privind dimensiunile, suprafața vitrificată sau orientarea pentru a optimiza consumul. Astfel, proiectanții pot ajusta designul și materialele, managerii pot planifica bugete și resurse, iar autoritățile pot încuraja standarde mai înalte de eficiență energetică.
* Totuși, merită subliniat că modelul liniar, deși simplu și interpretabil, nu poate surprinde întotdeauna relații nelineare sau factori externi (ex. variații climatice extreme, materiale de izolație atipice, comportamentul real al ocupanților). Pentru o precizie și mai bună, s-ar putea investiga în modele mai complexe (de ex. tehnici de machine learning avansate) sau în extinderea setului de date cu caracteristici suplimentare. În ansamblu, lucrarea oferă un punct de plecare concret și util pentru estimarea cerințelor de încălzire și luarea deciziilor informate în domeniul eficienței energetice a clădirilor.