MINISTERUL EDUCAŢIEI ȘI CERCETĂRII AL REPUBLICII MOLDOVA

Universitatea Tehnică a Moldovei

Facultatea Calculatoare, Informatică și Microelectronică

Departamentul Ingineria Software și Automatică

**RAPORT**

Lucrarea de laborator nr. 1

Disciplina: Inteligenta Artificiala

**Tema: Regresia liniară**

|  |  |
| --- | --- |
| A efectuat: | st. gr. TI-214  Reguș Ruslan |
| A verificat: | asist.univ.  Mariana Rusu |
|  |  |

Chișinău, 2024

**Setul de date personal:**

Sarcină individuală: Consultați https://www.kaggle.com/datasets pentru a alege și a descărca un set de date în scopul creării unui model liniar. Dacă este necesar curățați mai întâi datele și argumentați metoda aleasă. Parcurgeți pașii 1-20 ca în exemplu de mai sus - argumentați corelația dintre coloanele setului de date ales și evaluați predicția

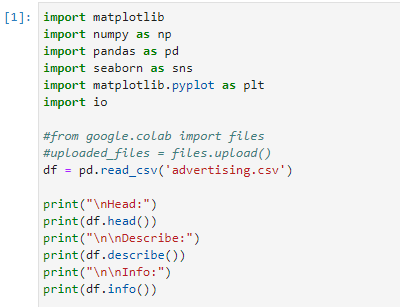
Pasul 1: Importați panda, numpy, matplotlib și seaborn. Apoi setați %matplotlib inline (veți importa apoi sklearn după necesitate).

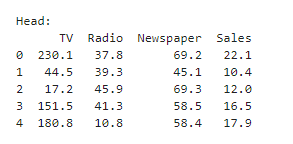
1. Obținerea datelor

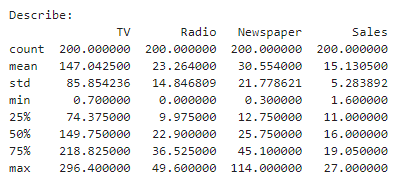
Vom lucra cu fișierul Ecommerce Customers CSV de la companie. Are informații despre clienți, cum ar fi e-mail, adresă și avatar de culoare. Apoi are și coloane cu valori numerice: - Avg. Session Length: durata medie a sesiunilor de consiliere stil în magazin. - Time on App: timpul mediu petrecut pe aplicație în minute. - Time on Website: timpul mediu petrecut pe site în minute - Length of Membership: de câți ani este membru clientul.

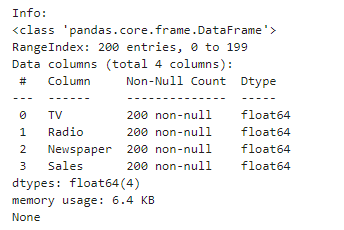
Pasul 2: Citiți fișierul Sleep\_Efficiency CSV ca DataFrame numit df.

Pasul 3: Citiți primele linii (head) despre clienți și aflați informații despre ei folosinf funcțiile info() și describe().



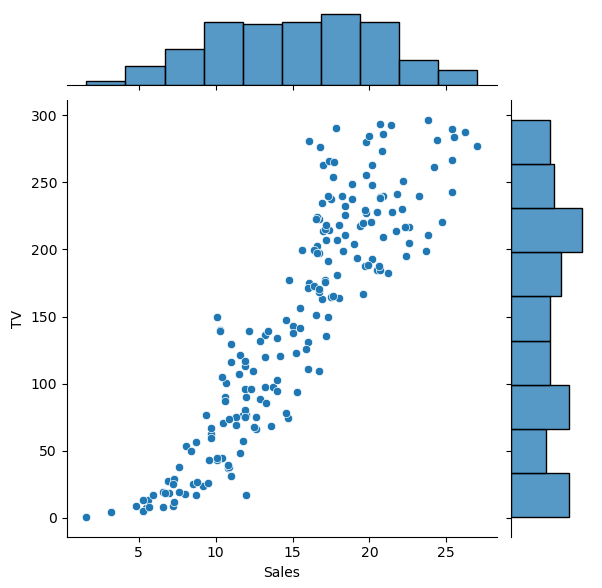
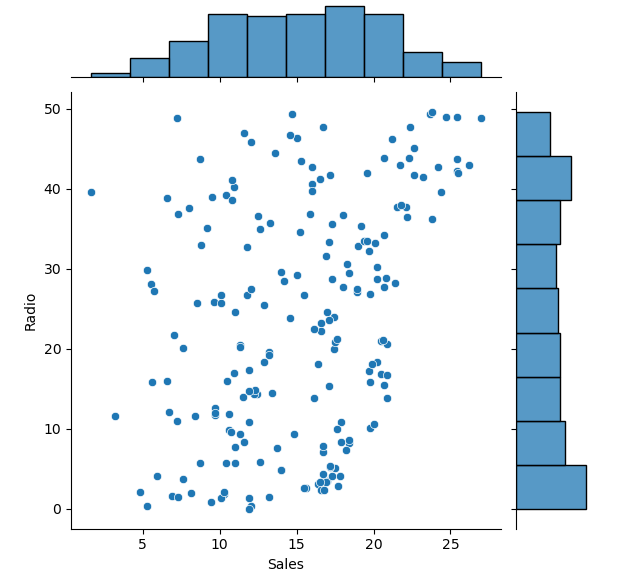






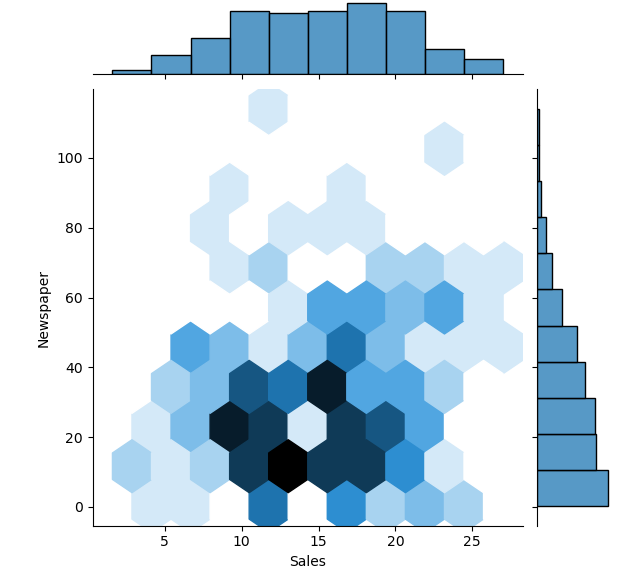
1. Analiza exploratorie a datelor Pentru restul exercițiului vom folosi doar datele numerice ale fișierului csv. Utilizați seaborn pentru a crea o diagramă comună (jointplot) pentru a compara coloanele Deep sleep percentage și Caffeine Consumption.
2. Pasul 4: Faceți același lucru, dar cu coloana Alcohol consumption



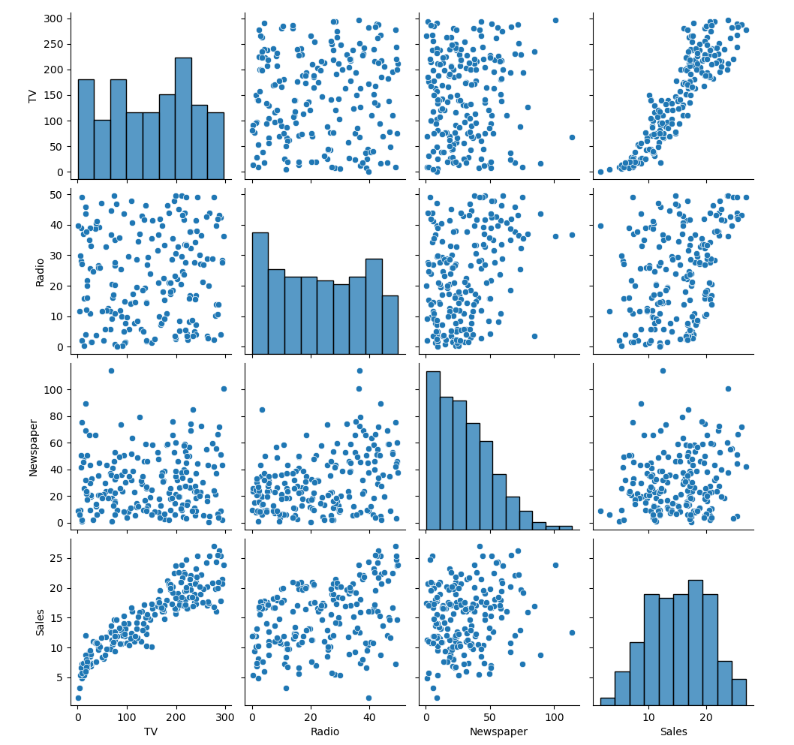
Pasul 5: Folosiți jointplot() pentru a crea o diagramă 2D ce compară Exercise frequency și Deep sleep percentage



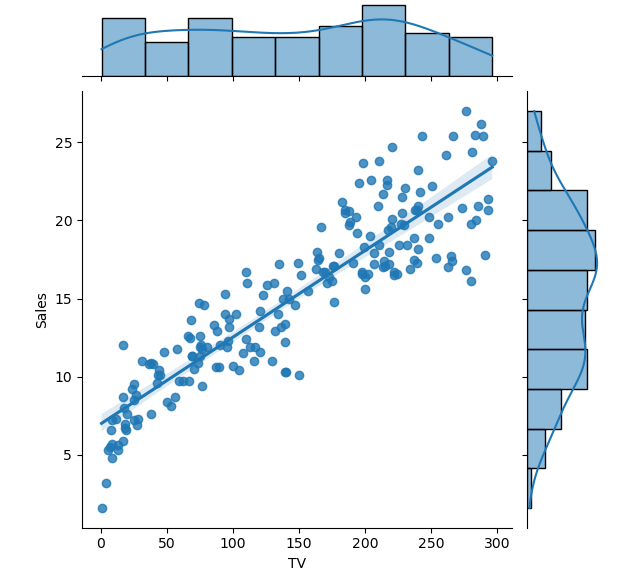


Pasul 6: Să explorăm aceste tipuri de relații în întregul set de date. Utilizați pairplot pentru a recrea graficul de mai jos. (Nu vă faceți griji pentru culori)



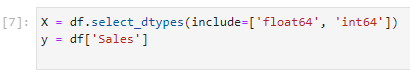


Pasul 7: Creați un model liniar grafic (folosind Implot din seaborn) al datelor: Sleep efficiency vs. Age



**Date de instruire și testare. Să împărțim datele în seturi de antrenament și de testare.**

Pasul 8: Setați o variabilă X egală cu caracteristicile numerice ale clienților și o variabilă y egală cu coloana Sleep efficiency



**Antrenarea modelului Acum este timpul să ne instruim modelul pe datele noastre de antrenament!**

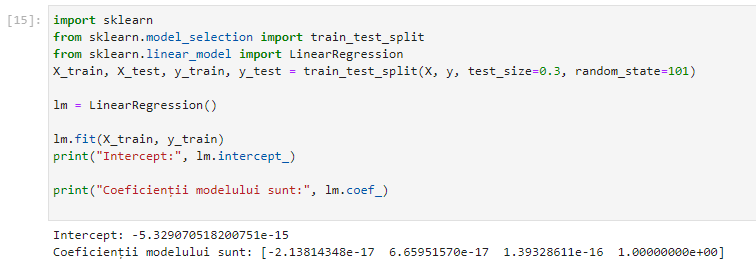
Pasul 9: Utilizați model\_selection.train\_test\_split din sklearn pentru a împărți datele în seturi de antrenament și de testare. Setați test\_size=0,3 și random\_state=101.

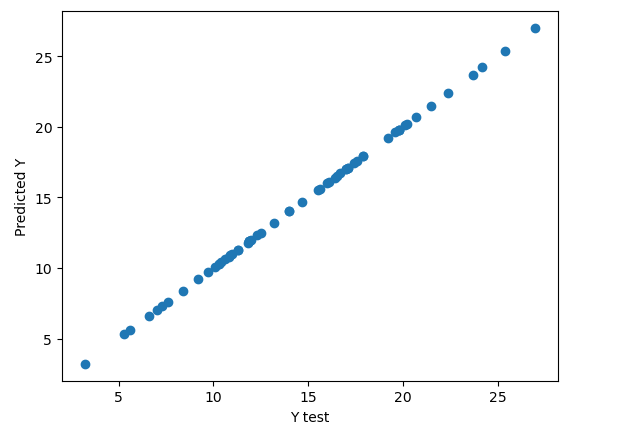
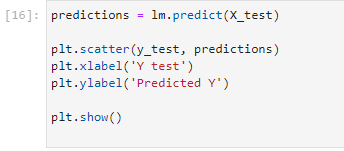
Pasul 10: Importați LinearRegression din sklearn.linear\_model.

Pasul 11: Creați o instanță a unui model LinearRegression() numit lm.

Pasul 12: Antrenați lm pe datele de antrenament. LinearRegression(copy\_X=True, fit\_i ntercept=True, n\_jobs=1, normalize=False)

Pasul 13: Imprimați coeficienții modelului.





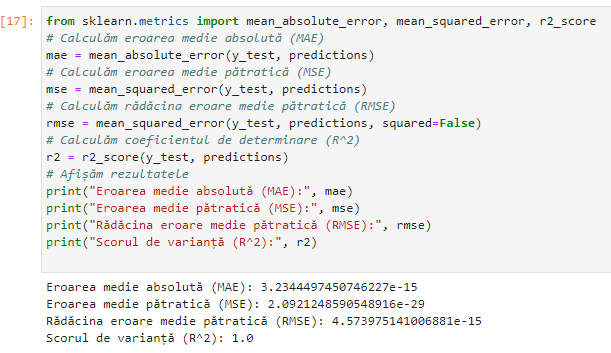
Prezicerea datelor de testare Să evaluăm performanța modelului prin prezicerea valorilor testului!

Pasul 14: Utilizați lm.predict() pentru a prezice setul X\_test de date.

Pasul 15: Creați o diagramă de dispersie a valorilor reale de test față de valorile prezise.

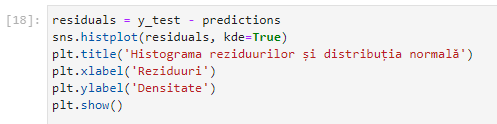
Evaluarea modelului Să evaluăm performanța modelului nostru calculând suma reziduală a pătratelor și scorul de varianță (R^2).

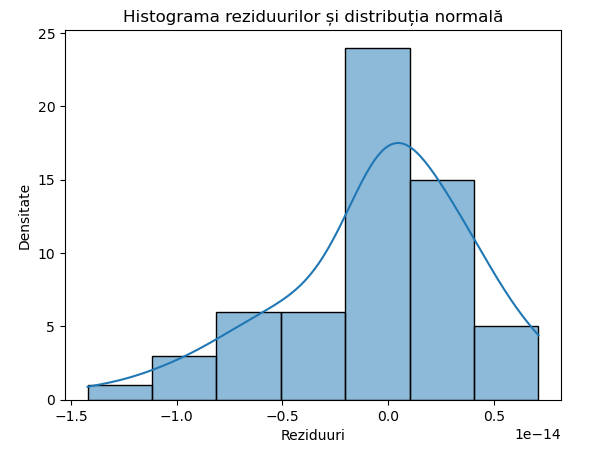
Pasul 16: Calculați eroarea medie absolută, eroarea medie pătratică și eroarea medie pătratică.



Reziduuri Să explorăm rapid reziduurile pentru a ne asigura că totul a fost în regulă cu datele noastre.

Pasul 17: Trasați o histogramă a reziduurilor și asigurați-te că are o distribuție normală. Utilizați fie seaborn distplot, fie doar plt.hist().





Concluzie: În urma analizei efectuate, am observat schimbarea folosirii reclamei intre gazete, TV si radio-u, cu cat mai accesibil a devenit folosirea TV-ul, cu atat mai multa reclama a fost introdusa cu scopul de a ridica popularitatea produselor.