**Universitatea “DUNĂREA DE JOS” Galați**

Facultatea de Automatică, Calculatoare, Inginerie Electrica și Electronică,

Specializarea Calculatoare

**Tema 3**

**Obiceiurile alimentare ale Romanilor**

Învățarea nesupravegheata

S.l.dr.inf. Student:

Crăciun Viorel Marian Abeaboeru Alexandru Cristian

Cuprins

[Introducere 3](#_Toc153143652)

[Rezultate și analiză 5](#_Toc153143653)

[Concluzie 9](#_Toc153143654)

[Bibliografie: 10](#_Toc153143655)

# Introducere

Învățarea nesupravegheata

Sarcinile rețelelor neuronale sunt adesea clasificate în două categorii: discriminative (recunoaștere) și generative (imaginație). Deși în general sarcinile discriminative folosesc metode supervizate și cele generative folosesc metode nesupravegheate, această separare nu este strictă. **[2]**

De exemplu, recunoașterea obiectelor, deși favorizează învățarea supervizată, poate fi abordată și prin învățare nesupravegheată. În plus, unele sarcini combină ambele metode de învățare, precum în cazul recunoașterii imaginilor, care a evoluat de la o abordare predominant supervizată, la una hibridă, și apoi din nou la supervizare, prin integrarea tehnologiilor precum dropout, ReLU și ratelor de învățare adaptabile. **[2]**

Învățarea nesupravegheată reprezintă o metodă de învățare automată în care algoritmii sunt antrenați pe seturi de date neetichetate, fără instrucțiuni specifice sau rezultate predefinite. Principalul obiectiv al acestui tip de învățare este descoperirea de modele și structuri ascunse în cadrul datelor. În acest proces, sistemul analizează caracteristicile de intrare și determină în mod independent caracteristicile și modelele cele mai relevante.**[3]**

Spre deosebire de învățarea supravegheată, care implică clasificarea datelor în categorii sau clase predefinite, învățarea nesupravegheată se concentrează pe identificarea asemănărilor naturale dintre instanțe, grupând datele pe baza caracteristicilor comune. Prin urmare, în loc să se bazeze pe etichete externe, algoritmii de învățare nesupravegheată formează propriile grupuri sau clase bazate pe structura intrinsecă a datelor. Această abordare permite sistemului să dezvăluie și să înțeleagă complexitatea și variația internă a seturilor de date, fără a se baza pe cunoștințe prealabile sau pe orientări specifice.**[1]**

Diferiți algoritmi de învățarea nesupravegheata

K-means Clustering:

Scop: Gruparea datelor într-un număr predefinit de clustere, bazată pe similaritatea caracteristicilor.

Funcționare: Inițializează centroizi aleatorii, apoi iterează prin mutarea acestor centroizi până când gruparea datelor este optimizată.

Utilizare: Util în segmentarea pieței, organizarea documentelor, și în multe alte aplicații de grupare.

Analiza Componentelor Principale (PCA):

Scop: Reducerea dimensionalității datelor prin păstrarea componentelor cele mai semnificative.

Funcționare: Identifică direcțiile (componentele principale) în care datele variază cel mai mult.

Utilizare: Folosit pentru simplificarea seturilor de date, vizualizare și în preprocesarea pentru alte algoritmi de învățare automată.

Gaussian Mixture Models (GMM):

Scop: Un model probabilistic pentru reprezentarea prezenței subpopulațiilor în cadrul unui ansamblu mai mare.

Funcționare: Presupune că datele provin dintr-un amestec de mai multe distribuții gaussiene (normale).

Utilizare: Util în probleme de clasificare, în analiza grupării și în învățarea modelului generativ.

Apriori Algorithm:

Scop: Descoperirea regulilor de asociere frecvente în seturi mari de date tranzacționale.

Funcționare: Identifică seturi de elemente care apar împreună frecvent în tranzacții și formează reguli bazate pe aceste asocieri.

Utilizare: Larg folosit în analiza coșului de cumpărături pentru a găsi produse care sunt adesea cumpărate împreună.

Identificarea grupurilor

K-means Clustering este un algoritm folosit în învățarea automată nesupravegheată, cu scopul de a grupa automat punctele de date în funcție de caracteristicile lor similare. Această metodă analizează și împarte datele în grupuri bazate pe modelele identificate în setul de date. Prin urmare, K-means are capacitatea de a forma propriile sale clase din date neetichetate.**[1]**

Etape principale ale algoritmului K-means Clustering

1. Atribuirea Centroizilor: Algoritmul începe prin alegerea unui număr predefinit de centroizi, reprezentând centrele grupurilor, de obicei selectați aleatoriu din setul de date. Acest număr este indicat de "K" în numele algoritmului.

2. Alocarea Punctelor de Date: Fiecare punct din setul de date este alocat grupului cu centroidul cel mai apropiat, determinând apropierea prin distanța euclidiană.

3. Actualizarea Pozitiei Centroizilor: Centroizii sunt recalculați pentru a reprezenta centrul de masă al punctelor din grupul lor. Algoritmul repetă acest proces de alocare și actualizare până când pozițiile centroizilor se stabilizează.

4. Convergența: Se atinge atunci când centroizii nu se mai modifică semnificativ între iterații, sugerând o poziție optimă unde distanța dintre centroizi și punctele de date din grupurile lor este minimizată.

K-means Clustering este eficient pentru organizarea datelor în grupuri ce reflectă structura naturală și relațiile dintre punctele de date. Este important în diverse domenii, dar necesită atenție la alegerea valorii K și la sensibilitatea față de pozițiile inițiale ale centroizilor.

# Rezultate și analiză

Graf 1.

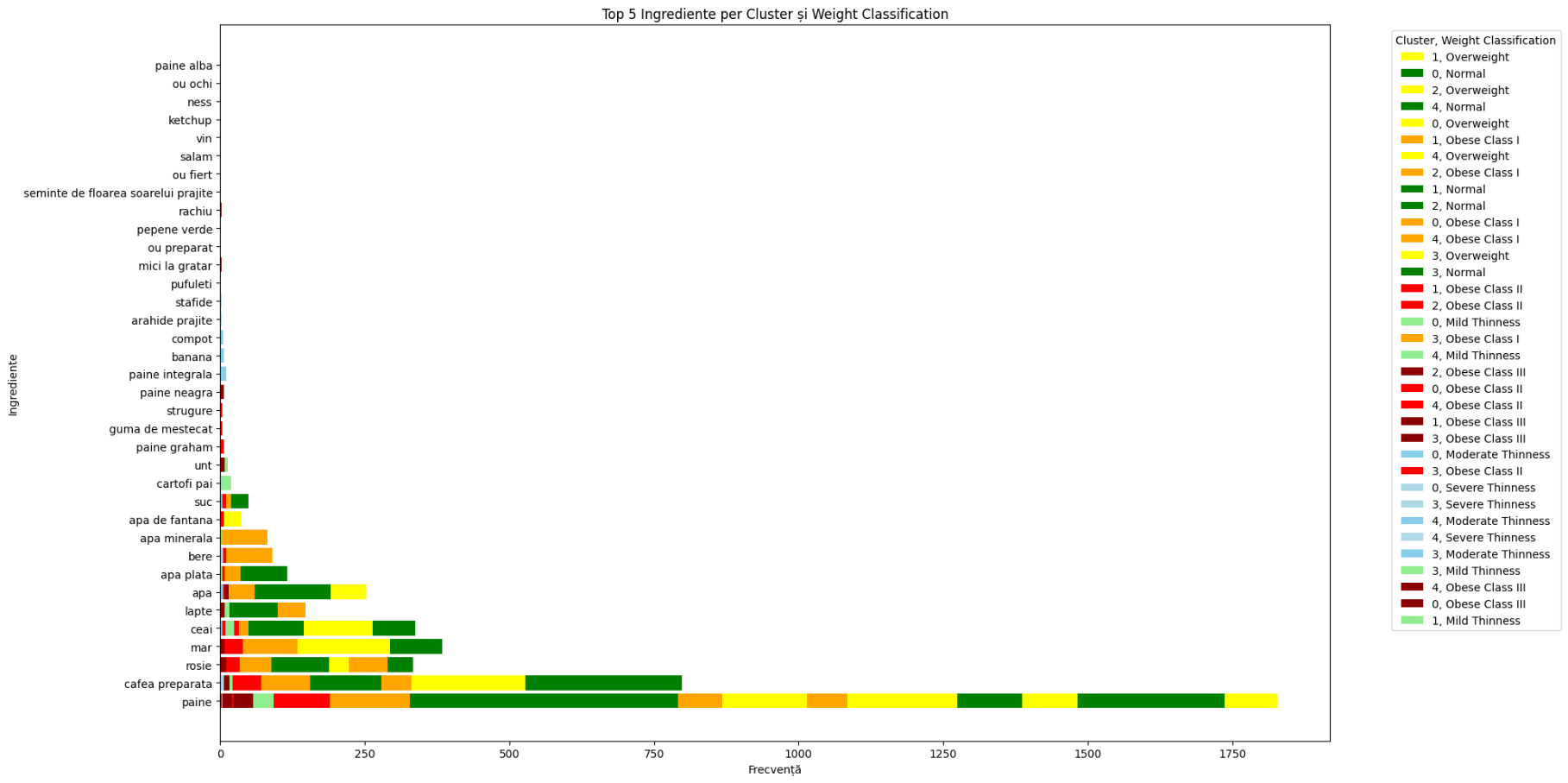
|  |  |
| --- | --- |
| Procentajul pentru fiecare rețetă: | |
| paine | 14.26% |
| cafea preparata | 5.45% |
| rosie | 2.83% |
| mar | 2.43% |
| ceai | 2.36% |
| lapte | 1.73% |
| apa | 1.31% |
| brânza | 1.11% |
| strugure | 1.09% |
| unt | 1.08% |

|  |  |
| --- | --- |
| Weight Classification | |
| Overweight | 41.22% |
| Normal | 34.97% |
| Obese Class I | 18.67% |
| Obese Class II | 3.18% |
| Mild Thinness | 1.15% |
| Obese Class III | 0.58% |
| Moderate Thinness | 0.15% |
| Severe Thinness | 0.05% |

Tabela 2.

Tabela 1.

Analizând setul de date despre obiceiurile alimentare ale românilor și clasificarea lor de greutate, observăm din Graf 1. si Tabelele 1. si 2. că majoritatea persoanelor se încadrează în categoria "Overweight" (41,22%) și "Normal" (34,97%). Cele mai populare alimente sunt de bază și includ "pâinea", "cafea preparată" și "roșiile".

Graf 2.

|  |  |
| --- | --- |
| Procentajul pentru fiecare rețetă: | |
| paine | 19.09% |
| cafea preparata | 7.30% |
| rosie | 3.78% |
| mar | 3.25% |
| ceai | 3.16% |
| lapte | 2.32% |
| apa | 1.76% |
| brânza | 1.49% |
| strugure | 1.46% |
| unt | 1.45% |

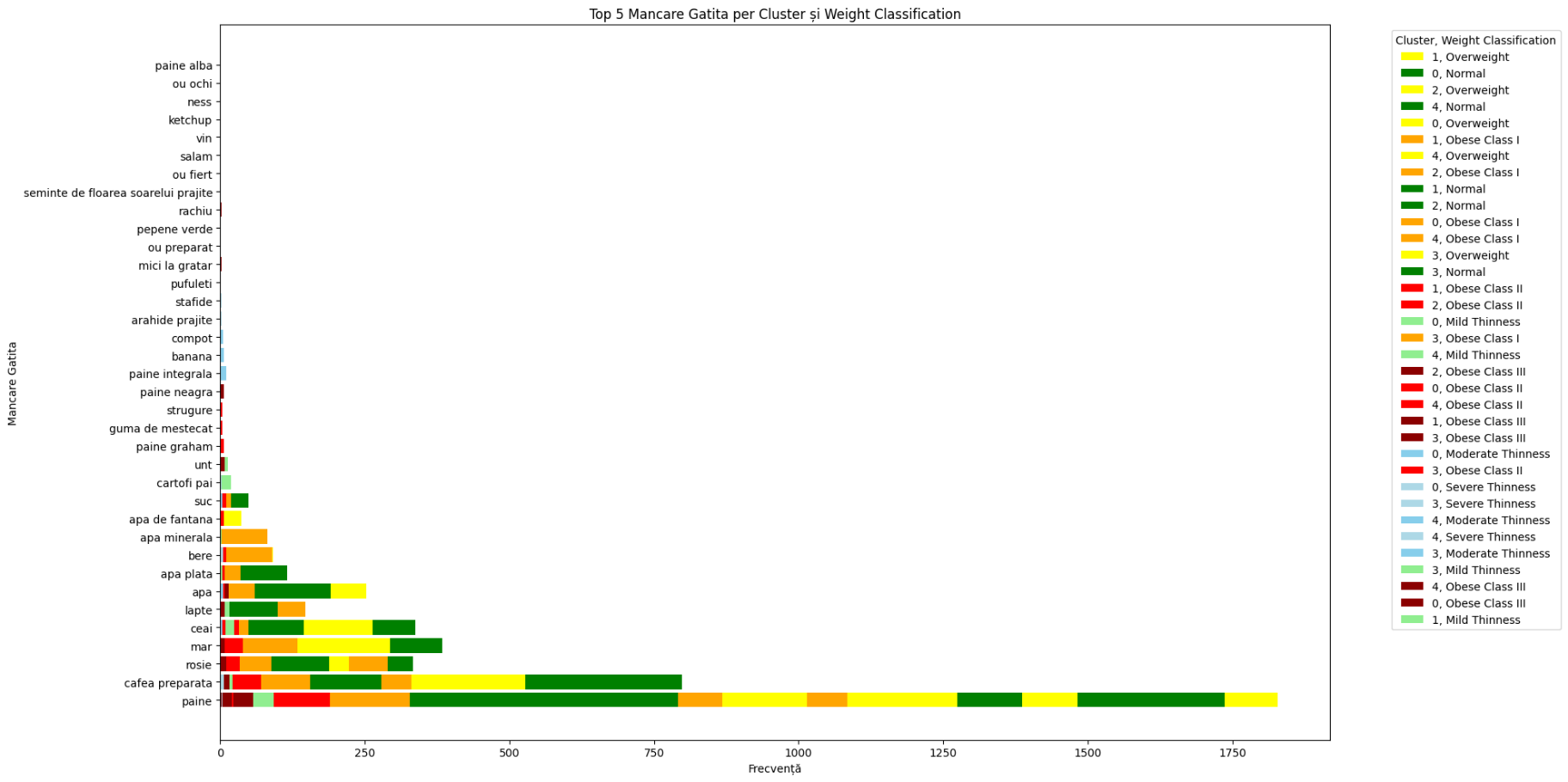
|  |  |
| --- | --- |
| Weight Classification | |
| Overweight | 41.16% |
| Normal | 34.93% |
| Obese Class I | 18.73% |
| Obese Class II | 3.17% |
| Mild Thinness | 1.18% |
| Obese Class III | 0.59% |
| Moderate Thinness | 0.15% |
| Severe Thinness | 0.06% |

.

Tabela 4

Tabela 3.

Analizând setul de date care se concentrează pe ingredientele alimentare consumate de români, observăm din Graf 2. si Tabelele 3. si 4. că majoritatea persoanelor sunt clasificate ca "Overweight" (41,16%) și "Normal" (34,93%). Printre ingredientele cel mai des consumate, "pâinea" conduce cu un procent semnificativ de 19,09%, urmată de "cafea preparată" și "roșii".

Graf 3.

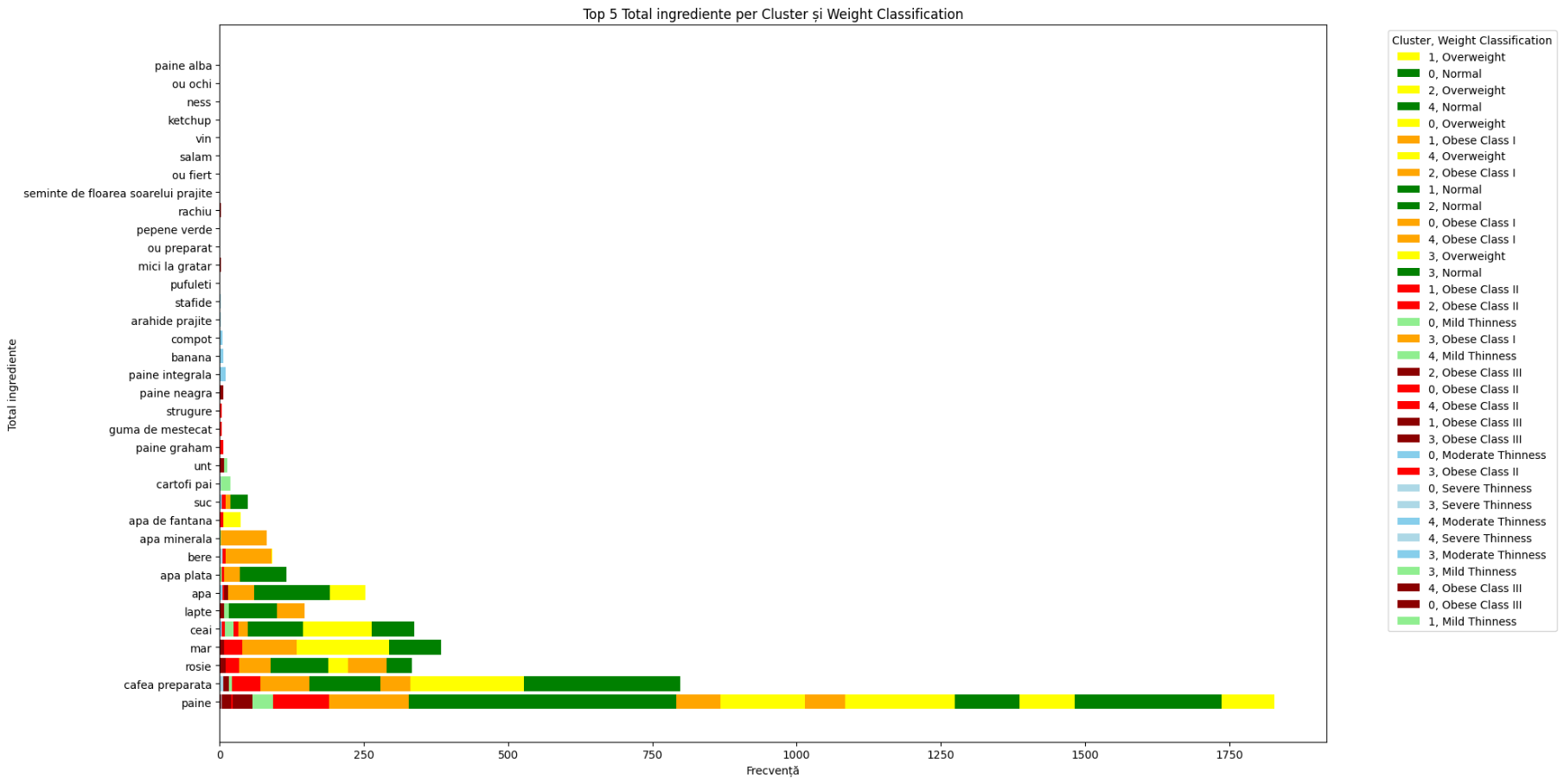
|  |  |
| --- | --- |
| Procentajul pentru fiecare rețetă: | |
| piure de cartofi | 3.30% |
| ciorba taraneasca | 2.71% |
| bors de gaina | 2.36% |
| salata de vinete | 2.23% |
| Prajitura de casa | 1.85% |
| salata de rosii | 1.71% |
| ciorba de legume | 1.68% |
| ciorba de perisoare | 1.67% |
| bors de fasole | 1.67% |
| supa de pasare | 1.66% |

|  |  |
| --- | --- |
| Weight Classification | |
| Overweight | 41.39% |
| Normal | 35.10% |
| Obese Class I | 18.45% |
| Obese Class II | 3.20% |
| Mild Thinness | 1.07% |
| Obese Class III | 0.55% |
| Moderate Thinness | 0.16% |
| Severe Thinness | 0.05% |

Tabela 6.

Tabela 5.

În acest set de date, care se concentrează pe mâncărurile gătite consumate de români, observăm din Graf 3. si Tabelele 5. si 6. că majoritatea persoanelor se încadrează în categoriile "Overweight" (41,40%) și "Normal" (35,11%). Printre mâncărurile gătite preferate, "piureul de cartofi" se află în frunte cu un procent de 3,30%, urmat de "ciorbă țărănească" și "borș de găină".

Graf  4.

|  |  |
| --- | --- |
| Procentajul pentru fiecare rețetă: | |
| sare | 7.96% |
| ulei | 6.66% |
| paine | 5.30% |
| ceapa | 4.98% |
| apa | 4.32% |
| morcovi | 3.89% |
| rosie | 3.02% |
| verdeata | 2.65% |
| faina | 2.28% |
| ou (crud) | 2.26% |

|  |  |
| --- | --- |
| Weight Classification | |
| Overweight | 41.30% |
| Normal | 34.86% |
| Obese Class I | 18.73% |
| Obese Class II | 3.21% |
| Mild Thinness | 1.09% |
| Obese Class III | 0.56% |
| Moderate Thinness | 0.16% |
| Severe Thinness | 0.05% |

Tabela 8.

Tabela 7.

Analizând acest set de date care se concentrează pe ingredientele prime utilizate de români în alimentația lor, observăm din Graf 4. si Tabelele 7. si 8. că majoritatea persoanelor se încadrează în categoriile "Overweight" (41,31%) și "Normal" (34,87%). Ingredientele prime cel mai frecvent consumate sunt "sarea" (7,96%), "uleiul" (6,66%) și "pâinea" (5,30%), urmate de "ceapa" și "apa".

Prezența ridicată a sării și a uleiului în topul listei poate indica o tendință spre alimente mai procesate sau gătite în moduri care pot contribui la o greutate corporală mai mare. De asemenea, clasificările de greutate reflectă o distribuție în care o proporție semnificativă a populației se confruntă cu probleme legate de greutate. Aceste rezultate sugerează că tipurile de ingrediente prime utilizate și modul lor de utilizare în alimentație ar putea avea un impact considerabil asupra sănătății și greutății populației.

# Concluzie

Analiza setului de date, centrată pe obiceiurile alimentare și clasificarea greutății a populației din România, a dezvăluit o legătură semnificativă între tipurile de alimente consumate și categoriile de greutate corporală. Prin filtrarea și concentrarea pe datele BMI și pe clasificarea greutății, am observat că o proporție mare a populației se încadrează în categoriile "Overweight" și "Obese", ceea ce indică o tendință generală spre supraponderalitate și obezitate. Această situație este reflectată și prin ingredientele și mâncărurile preferate, care variază de la alimente de bază, precum pâinea și ceapa, la cele cu conținut mare de grăsimi și sare, cum ar fi uleiul și sarea.

Utilizarea algoritmului K-Means pentru clustering a oferit o perspectivă valoroasă asupra modului în care obiceiurile alimentare se pot grupa în funcție de diferite categorii de greutate. Această abordare nesupravegheată a permis identificarea modelelor ascunse în date, sugerând relații specifice între tipurile de alimente consumate și problemele de greutate.

În ceea ce privește procentajele obținute prin testări multiple, acestea au subliniat faptul că anumite alimente, precum pâinea și uleiul, sunt consumate în mod regulat în toate categoriile de greutate, în timp ce alte alimente apar predominant în anumite categorii. De exemplu, prezența sării și a uleiului în dietele celor clasificați ca "Overweight" sau "Obese" sugerează o preferință pentru alimente mai procesate sau gătite în moduri potențial mai puțin sănătoase.

# Bibliografie:

1. <https://www.unite.ai/ro/%C3%AEnv%C4%83%C8%9Bare-supravegheat%C4%83-vs-nesupravegheat%C4%83/>
2. <https://en.wikipedia.org/wiki/Unsupervised_learning>
3. <https://d-data.ro/10-algoritmi-de-invatare-automata-nesupravegheati-ce-sunt-acestia-si-cum-sa-ii-cream/>