***ANALIZA ȘI PREDICȚIA COEFICIENȚILOR DE NUTRIȚIE***

Caracaci Dan1

1Universitatea Tehnică a Moldovei, Facultatea Calculatoare, Informatică și Microelectronică,

Departamentul Informatica și Ingineria Sistemelor

# ABSTRACT

În acest studiu, ne-am concentrat asupra analizei și predicției malnutriției la nivel global, folosind setul de date "Malnutrition across the globe", disponibil pe platform[ online pentru seturi de date. Scopul nostru a fost de a identifica factorii de risc relevanți și de a dezvolta modele pentru a estima probabilitatea de malnutriție în diferite țări. Am început analiza prin explorarea setului de date. Am investigat relațiile dintre variabile și am identificat modele semnificative care pot influența riscul de malnutriție. De asemenea, am realizat vizualizări grafice, precum distribuția severității de underweight în funcție de clasificarea veniturilor și evoluția a overweight în timp. Modelarea a constituit o componentă crucială a studiului nostru. Am dezvoltat mai multe modele, explorând diverse combinații de variabile, cum ar fi clasificarea veniturilor, severitatea de underweight, overweight, stunting și altele. Am utilizat analiza VIP (Variable Importance in Projection) pentru a identifica caracteristicile cheie asociate cu probabilitatea de malnutriție. Rezultatele noastre au demonstrat că anumite variabile, precum clasificarea veniturilor, severitatea a underweight și overweight, sunt semnificativ asociate cu probabilitatea de malnutriție în diferite țări. Modelele noastre au furnizat o înțelegere mai profundă a interacțiunilor dintre aceste variabile, sprijinind procesul de predicție și identificare a zonelor cu risc crescut. În concluzie, această analiză detaliată a setului de date "Malnutrition across the globe" a evidențiat factorii de risc implicați în malnutriție la nivel global. Studiul nostru oferă perspective valoroase asupra modului în care diferiți factori pot influența probabilitatea malnutriției în diferite țări și poate servi drept bază pentru dezvoltarea ulterioară a modelelor de predicție și a strategiilor de intervenție pentru îmbunătățirea stării nutriționale la nivel mondial.

# INTRODUCERE

În ultimele decenii, preocupările cu privire la malnutriție s-au accentuat odată cu creșterea accesului la date relevante și progresele tehnologice în domeniul sănătății. Cu toate acestea, există încă lacune semnificative în înțelegerea complexității interacțiunilor dintre factorii de risc asociati malnutriției, precum venitul, severitatea a underweight, overweight, stunting și alți parametri relevanți.

Malnutriția reprezintă o problemă globală, afectând grav starea de sănătate și calitatea vieții în numeroase țări. În acest context, setul de date "Malnutrition across the globe" aduce în prim-plan oportunități semnificative pentru explorarea și înțelegerea profunzimii acestei probleme la nivel mondial. În mod specific, malnutriția este asociată cu multiple consecințe asupra sănătății, inclusiv severitatea a underweight, prezența overweight și stunting-ul la copii. În anumite regiuni, aceste aspecte pot fi amplificate de condiții economice precare și acces limitat la resurse alimentare de calitate.

Studiul nostru se concentrează pe analiza și predicția riscului de malnutriție, utilizând setul de date "Malnutrition across the globe". Scopul principal este identificarea factorilor de risc cheie și dezvoltarea unui model eficient pentru a estima probabilitatea de malnutriție în diferite țări și regiuni.

Prin acest studiu, ne propunem să aducem contribuții semnificative în înțelegerea complexității malnutriției la nivel global. Vom explora relațiile dintre variabilele relevante și vom dezvolta modele predictive pentru a evalua și determina impactul acestora asupra probabilității de malnutriție. Identificarea factorilor de risc în cadrul acestui studiu poate oferi baze solide pentru elaborarea strategiilor personalizate de prevenție și intervenție în domeniul nutriției.

În perspectivă mai largă, această cercetare poate contribui la consolidarea cunoștințelor despre malnutriție la nivel global și la dezvoltarea unor abordări mai eficiente pentru prevenirea și gestionarea acestei probleme de sănătate publică. Astfel, această investigație se aliniază eforturilor globale de îmbunătățire a stării nutriționale a populației și promovare a sănătății la nivel mondial.[3]

**MATERIALE ȘI METODE**

***Setul de date***

Studiul nostru a utilizat setul de date " Malnutrition across the globe" disponibil pe platforma Kaggle[1]. Acesta conține informații variate despre țări, coeficienții precum overweight, underweight, vasting, stunting și sexul, aspecte precum anul colectării datelor și PIB pe cap de locuitor.

***Preprocesarea datelor***

Înainte de analiză, am efectuat o preprocesare a datelor. Variabilele ce indicau coeficienții, precum vasting, overweight, stunting și anul au fost transformate în numerice pentru a permite o analiză eficientă. Am utilizat funcții precum `*summary*` și `*glimpse*` pentru a obține o înțelegere detaliată a distribuției și caracteristicilor setului de date.

***Analiza exploratorie***

Analiza exploratorie a fost realizată pentru a evidenția relațiile dintre variabile și pentru a identifica tendințe semnificative. Cu ajutorul pachetelor ggplot2, tidyverse, am generat vizualizări informative, cum ar fi bar-plot care ilustreză coeficienții de nutriție pe parcursul a mai multor ani penru o țară și distribuții a coeficientului de stunting în dependență de PIB.

***Modelarea liniară***

Setul de date a fost divizat într-un set de antrenare și unul de testare folosind funcția `*initial\_split*` din pachetul rsample. Am construit apoi mai multe modele liniare, considerând combinații diferite de variabile. Modelele au fost dezvoltate folosind funcția `*lm*` pentru regresie liniară.

***Evaluarea performanței modelelor***

Funcția summary() oferă informații detaliate despre fiecare model, inclusiv coeficienți, valori p și valori R pătrat. De asemenea poate fi folosită funcția anova() pentru a efectua analiza a variației dintre două sau mai multe modele. Un model perfect ar avea valoare p foarte mică, ceea ce va indica că variabilile alese sunt excelente pentru model. Un alt aspect important este valoare R pătrat care indică performanța modelului, cu cât valoarea acestui coeficient este mai aproape de 1, cu atât modelul este mai perfect.

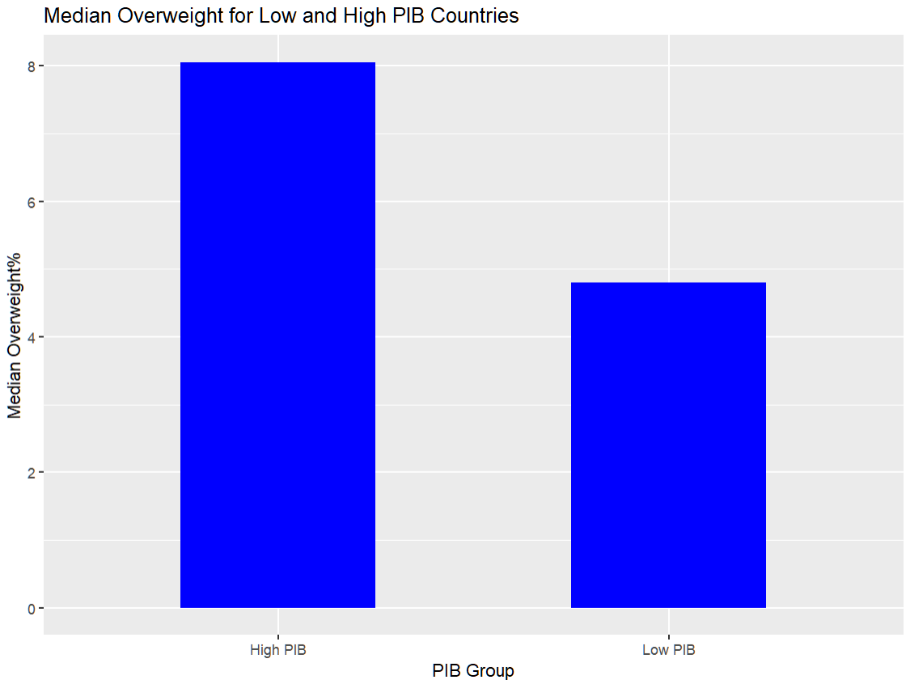
# REZULTATE

În setul de date analizat avem o variabilă dependentă *stunting,* și mai multe variabile independente precum *severe wasting, wasting, overweight, underweight și PIB per capita*. Variabila independentă *severe wasting* are un min de 0 procente și un max de 12 procente, iar o medie pentru setul de date analizat este 2.224, ce indică că coeficientul este destul de mic. De exemplu, variabila *wasting* are un min de 0.3 și un max de 25.3. Dacă ne referim la variabila ce conține observații privitor la malnutriție, atunci min este de 0.3 procente, iar max este de 66.8 procente. Penru a ilustra modificarea coeficienților și pentru a construi modelul mai eficient, datele au fost împărțite pe țări în dependență de observația ce necesită demonstrare.

***Analiza distibuțiilor variabilelor***

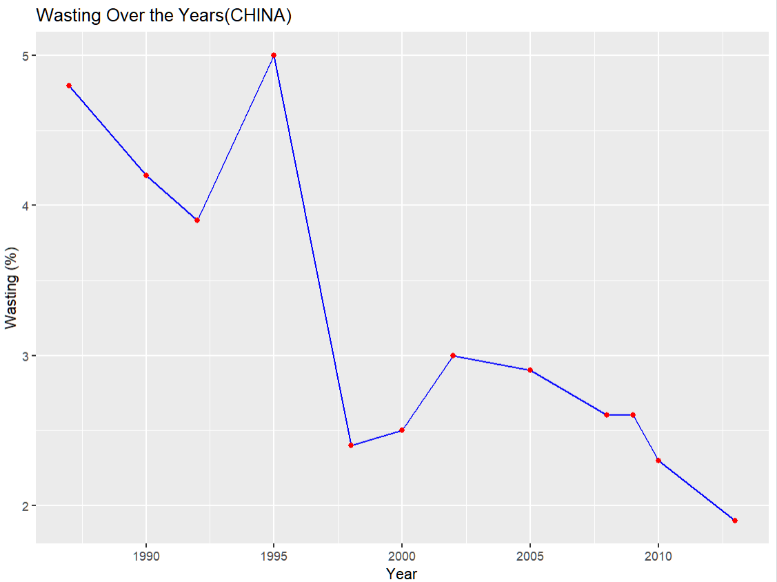
În încercarea de a înțelege relația dintre obezitate și PIB-ul pe cap de locuitor, am analizat setul de date. Una din presupuneri este dacă venitul înalt este factor care indică nivelul înalt de obezitate.

Prin urmare, ne-am concentrat pe variabilele '*overweight*' și '*PIB.per.capita*' pentru a explora trendurile posibile. Pentru a ilustra această relație și pentru a facilita înțelegerea rezultatelor noastre, am pregătit o figură informativă care sintetizează corelația între obezitate și venit, care a fost categorizat în venit înalt și scăzut în baza valorii medie (*Figura 1*).

---

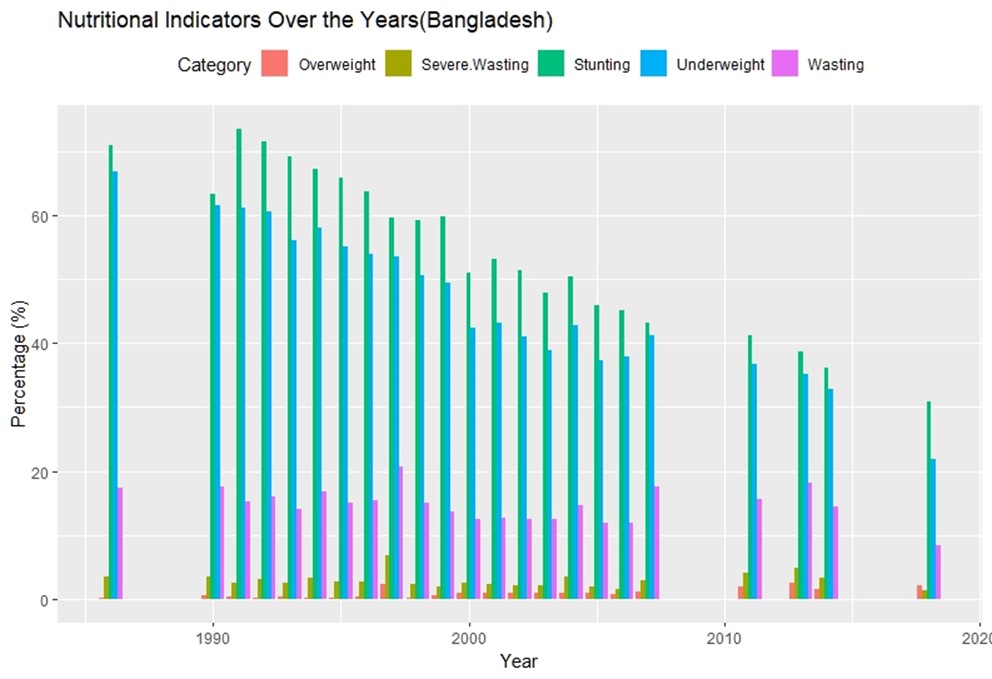
**Figura 1.** *Nivelul de obezitatel raportat la venit*

Plotul final este un bar-plot (*Figura 1*) pentru a evidenția corelația dintre obezitate și venit.. Putem observa că țările cu PIB pe cap de locuitor înalt au un coeficient de obezitate mai înalt comparativ cu țările cu PIB scăzut.



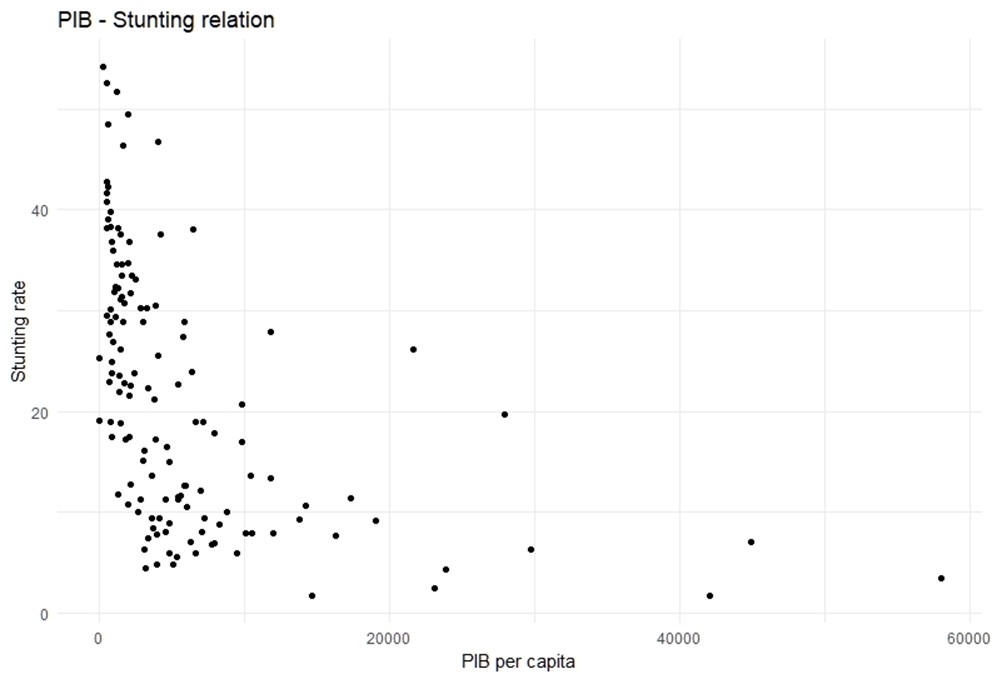
**Figura 2.** *Modificarea coeficientului Wasting pe parcursul anilor*

Pentru a evalua această observație, vom utiliza un grafic de tip line (*Figura 2*) pentru a ilustra cum se schimbă coeficientul wasting de la sfârșitul secolului trecut și până la ultimul an prezent în setul de date. Pentru o ilustrare mai eficientă au fost alese datele pentru China, căci conțin înregistrări mai vechi și se poate observa mai bine schimbarea drastică a atitudinii oamenilor față de irosirea produselor alimentare.



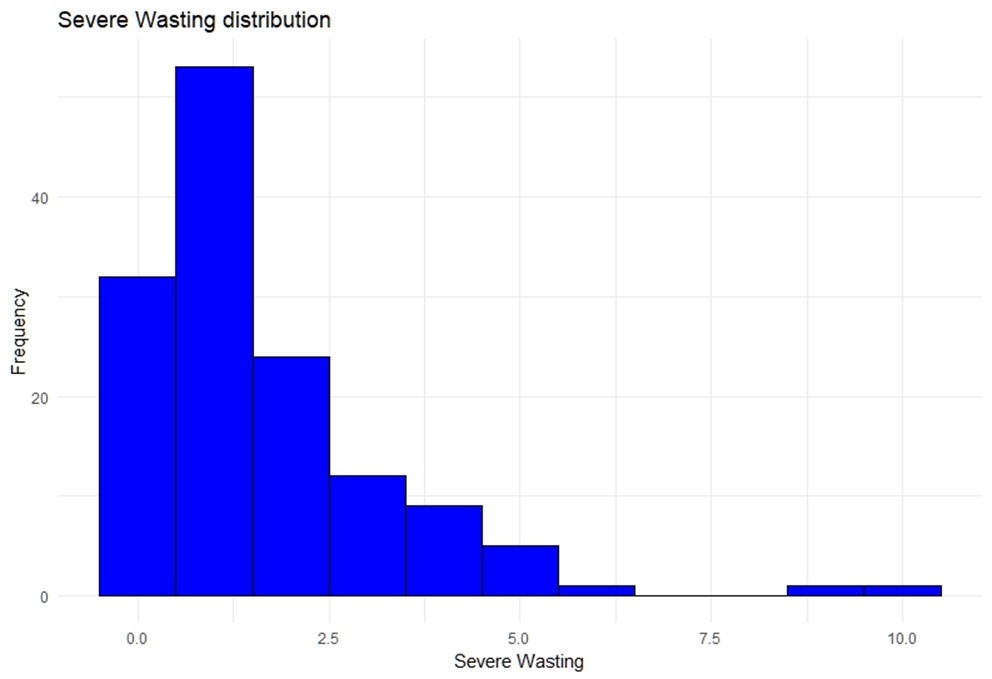
**Figura 3.** *Schimbarea coeficienților într-o țară în curs de dezvoltare*

Folosind un bar-plot, intenționăm să evidențiem evoluția coeficienților de nutriție în Bangladeș pe parcursul dezvoltării țării. Astfel putem observa (*Figura 3*) că cu trecerea timpului și dezvoltarea țării acești coeficienți negativ se micșorează semnificativ. Dar se observă că coeficienții stunting și underweight sunt mult mai mare comparativ cu ceilalți coeficienți.



**Figura 4.** *Relația dintre venit și stunting*

Cu ajutorul unui scatter-plot, dorim să observăm relația dintre venit înalt și coeficientul stunting. Scopul este identificarea unei relații direct proporționale dintre aceste două variabile. Plotul din (*Figura 4)* ne oferă claritate. În cazul venitului scăzut în țară, coeficientul stanting este foarte mare și inver, în cazul venitului pe cap de locuitor înalt coeficientul stanting este aproape de zero.



**Figura 5.** *Nivelul irosirii severe a produselor alimentare*

Pentru a demonstra această observație, vom folosi un bar-plot pentru a ilustra nivelul coeficientului severe wasting, care indică irosirea gravă a produselor alimentare, pentru toate țările conform celor mai recente date. Astfel, putem observa (*Figura 5*) că în majoritatea țărilor acest coeficient este mai mic de 2.5 procente. Putem conluziona că întradevăr majoritatea țărilor tind spre micșorarea irosirii grave.

O dată ce am terminat analiza exploratorie a datelor, pentru setul de date ales am creat modele de regresie liniară. Am selectat regresia liniară deoarece este o metodă de analiză statistică utilizată pentru a modela relația dintre variabilile numerice. Acest model este numit liniar deoarece analizează relația dintre o variabilă dependentă și una sau mai multe variabile independente și ilustrază această relație printr-o linie. Modelul liniar a fost creat pentru a prezice cum se va schimba coeficientul stunting în Bangladeș în dependență de schimbarea celorlalte coeficiente.

**Tabelul 1.** *Importanța variabilelor pentru model*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Coefficients | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 21.479172 | 7.087822 | 3.030 | 0.00755 |
| Underweight | 1.182909 | 0.128831 | 9.182 | 5.33e-08 |
| PIB.per.capita | -0.002127 | 0.002921 | -0.728 | 0.47643 |
| Severe.Wasting | 1.705407 | 1.611251 | 1.058 | 0.30467 |
| Wasting | -1.743416 | 0.536140 | -3.252 | 0.00469 |
| Overweight | -0.358723 | 2.559889 | -0.140 | 0.89020 |

Pentru a putea creia un model este necesar și de a ști care variabile sunt importante, care sunt mai puțin importante și care sunt redundante și antrenează modelul greșit. Pentru a putea spune cu precizie ce este util, am utilizat funcția *summary* care prezintă toți coeficienții pentru model. Cu cât mai mică este valoarea p cu atât mai importantă este variabila. (*Tabelul 1*).

**Tabelul 2.** *Compararea modelelor*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Res.Df | RSS | Df | Sum of Sq |
| 1 | 21 | 220.7 |  |  |
| 2 | 21 | 1132.8 | 0 | -912.09 |
| 3 | 21 | 2806.3 | 0 | -1673.48 |
| 4 | 21 | 3336.4 | 0 | -530.12 |
| 5 | 21 | 1445.0 | 0 | 1891.35 |

În tabelul 2 putem observa compararea tuturor modelelor după scorul RSS, efectuată cu ajutorul funcției *anova.* Cel mai bun rezultat este obținut de către primul model care revizuie relația dintre stuning și underweight*.*

Folosind funcția *summary*, putem identifaca performața celui mai bun model în baza valori R pătrat, cu cât aceasta este mai aproape de 1 cu atât modelul este mai performant. În cazul nostru valoarea R pătrat este de 0.9345 ce este aproape de ideal și înseamnă că modelul prezice cu o probabilitate de 93%.

.

# DISCUȚII

Studiul nostru aduce în prim-plan complexitatea interacțiunilor dintre diferite variabile asociate malnutriției la nivel global, oferind o perspectivă detaliată asupra impactului acestora. Analiza detaliată a datelor relevante din setul "Malnutrition across the globe" dezvăluie variabilități semnificative, furnizând informații cruciale pentru dezvoltarea unor strategii personalizate de intervenție.

Explorarea vârstei și a clasificării veniturilor evidențiază diferențe semnificative în prevalența malnutriției, indicând necesitatea unor abordări specifice în funcție de categoriile de vârstă și nivelurile de venit. Aceste constatări pot orienta eforturile de prevenție și intervenție către grupurile cu risc crescut, contribuind astfel la îmbunătățirea stării nutriționale la nivel global.

O altă dimensiune a cercetării se referă la analiza severității a underweight și a overweight în funcție de regiune. Rezultatele indică variații semnificative în prevalența acestor condiții, subliniind importanța adoptării unor strategii personalizate pentru a aborda specificitățile locale și regionale ale malnutriției.

De asemenea, identificarea nepotrivirilor pe țări în ceea ce privește malnutriția reprezintă un aspect crucial. Aceste constatări subliniază nevoia de abordări diferențiate în prevenție și gestionare, luând în considerare caracteristicile specifice regiunilor. Implementarea unor strategii de sănătate publică eficiente trebuie să țină cont de aceste diferențe pentru a maximiza impactul intervențiilor.

Totuși, trebuie menționat că studiul are limite, inclusiv restricțiile impuse de sursa de date și variabilele disponibile în setul de date analizat. Aceste limitări subliniază importanța cercetărilor ulterioare pentru a valida și consolida constatările, precum și pentru a explora alte aspecte relevante ale malnutriției la nivel global.

În concluzie, cercetarea noastră aduce contribuții semnificative la înțelegerea complexității malnutriției la nivel mondial și oferă un fundament solid pentru dezvoltarea unor strategii eficiente de prevenție și intervenție. Prin abordarea detaliată a acestor aspecte, sperăm să stimulăm eforturile globale de combatere a malnutriției și îmbunătățire a stării nutriționale la nivel global.

# BIBLIOGRAFIE

1. Kaggle, platformă online pentru seturi de date, ‘Malnutrition across the globe’ [Accesat:04.12.23], disponibil: [Malnutrition across the globe (kaggle.com)](https://www.kaggle.com/datasets/ruchi798/malnutrition-across-the-globe/?select=malnutrition-estimates.csv)
2. RDocumentation, resursă web pentru documentație privitor la RStudio , ‘RDocumentation’ [Accesat:04.12.23], disponibil: <https://www.rdocumentation.org/>
3. *Malnutrition - World Health Organization (WHO),* resursă web pentru informare despre malnutriție, [Accesat:04.12.23], disponibil: [Malnutrition (who.int)](https://www.who.int/health-topics/malnutrition#tab=tab_1)