Sănătatea somnului și stilul de viață

NUME: TOMOIAGĂ IOANA-GABRIELA

SPECIALIZARE: CALCULATOARE

Anul III

Cuprins

[1.Introducere 3](#_Toc167995073)

[2. Motivația alegerii bazei de date 3](#_Toc167995074)

[3. Obiectivele proiectului 3](#_Toc167995075)

[4. Aspecte teoretice relevante 3](#_Toc167995076)

[4.1. Starea curentă a domeniului 4](#_Toc167995077)

[5. Descrierea bazei de date 6](#_Toc167995078)

[6. Încărcarea și analiza bazei de date 6](#_Toc167995079)

[7. Curățare și eliminare anomalii 6](#_Toc167995080)

[8. Corelații 7](#_Toc167995081)

[9. Modele 11](#_Toc167995082)

[10. Testare și validare 14](#_Toc167995083)

[11. Rezultate 14](#_Toc167995084)

[12. Concluzii 15](#_Toc167995085)

[13. Surse 16](#_Toc167995086)

## 1.Introducere

Sănătatea somnului și stilul de viață au devenit subiecte tot mai importante, pe măsură ce înțelegerea lor contribuie la îmbunătățirea calității vieții și la prevenirea unor afecțiuni grave. În acest context, analiza datelor are un rol în identificarea modelelor și a factorilor care influențează sănătatea somnului și calitatea vieții. Analiza acestui set de date ne va ajuta să găsim legături importante între diverse variabile și starea de sănătate a somnului sau stilul de viață al participanților. Aceste concluzii ar putea avea un impact important asupra sănătății și pot sa contribuie la îmbunătățirea acestea.

## 2. Motivația alegerii bazei de date

Am ales baza de date "Sănătatea somnului și stilul de viață" deoarece îmi doresc să înțeleg mai bine relația dintre sănătatea somnului și stilul de viață, având în vedere impactul major pe care aceste aspecte le au asupra sănătății. Din punctul meu de vedere, această temă este importantă în contextul actual, în care stresul și ritmul de viață agitat pot afecta calitatea și durata somnului, având consecințe pe termen lung asupra sănătății fizice și mentale.

## 3. Obiectivele proiectului

Proiectul are ca scop predicția sănătății si calității somnului având în vedere prezența sau absența tulburărilor de somn. Scopul este de a analiza cum variabilele asociate sănătății somnului și stilului de viață sunt afectate de tulburările de somn și cum acestea, la rândul lor, pot influența prezența sau agravarea acestor tulburări.

## 4. Aspecte teoretice relevante

Somnul este o stare fiziologică periodică și reversibilă, caracterizată prin suprimarea temporară a conștienței, prin abolirea parțială a sensibilității și încetinirea funcțiilor vieții organice (ritm respirator, ritm cardiac, relaxare musculară, scăderea temperaturii cu aproximativ 0,5 grade C, scăderea funcțiilor secretorii). În timp ce mulți dintre noi cred că somnul este o întindere temporală în care nu se întâmplă nimic, somnul este, de fapt, cel puțin din punct de vedere neurologic, o perioadă de timp foarte aglomerată. Deși importanța somnului nu poate fi discutată, oamenii de știință nu cunosc cu exactitate de ce este atât de important pentru supraviețuirea noastră. (wikipedia, 2023)

Există mai multe ipoteze în această privință: (wikipedia, 2023)

* Ipoteza adaptivă: Această teorie ne spune că somnul îmbunătățește șansele de supraviețuire ale unui animal. Cele care obișnuiesc să doarmă conform modului lor de viață au cele mai mari șanse de supraviețuire. Speciile nocturne au somnul diferit de vânătorii diurni, spre exemplu.
* Ipoteza Conservării Energiei: Animalele ce au un stil de viață alert și cu metabolismul rapid dorm mai mult decât cele care ard caloriile mai încet, astfel conservându-și energia pentru partidele de vânătoare.
* Ipoteza Restaurării: Conform acestei teorii, corpul se restaurează pe parcursul somnului. Cercetătorii cunosc faptul că neuro-toxinele sunt neutralizate în timpul somnului și faptul că celulele se divid, țesuturile se sintetizează și hormonii de creștere sunt eliberați în timpul etapelor de somn cu emisii de unde slabe (sau non-REM). Atleții, spre exemplu, petrec mai mult timp în etapele 3 și 4 (unde slabe) decât alții, iar copiii și tinerii de asemenea.
* Ipoteza Programării-Reprogramării: Această teorie susține că informația inutilă este "ștearsă" și că informația importantă este întipărită într-o memorie mai puternică. Copii, care acumulează informație la o rată mult mai mare decât în orice altă etapă evolutivă a vieții, dorm cel mai mult. Cu toate acestea, somnul poate să nu fie similar întipăririi informațiilor învățate la școală, spre exemplu.

Cercetări recente indică faptul că somnul REM ar putea fi soluția. Copiii și tinerii parcurg în timpul somnului o perioada REM mai mare decât a adulților, iar adulții care sunt la școală sau trec prin perioade de solicitare intelectuală acumulativă intensă cunosc o creștere temporală a somnului REM. Când oamenii sunt privați de somnul REM sunt mai puțin pricepuți la a rezolva probleme creative. (wikipedia, 2023)

### 4.1. Starea curentă a domeniului

Somnul este esențial vieții. Animalele de laborator private de somn au murit. Nici oamenii nu par sa fie imuni la efectele lipsei de somn care le amenință traiul. Un sondaj condus de **American Cancer Society** concluzionează faptul că oamenii ce dorm mai puțin de 6 ore pe noapte sau care dorm mai mult de 9 ore, au avut o rată a morții cu 30% mai mare decât cei care dorm în mod obișnuit 7 - 8 ore. Chiar și cei care au dormit mai puțin de 6 ore și în general nu au avut probleme de sănătate dar au avut o rată a morții de 1,8 ori mai mare decât cei ce au dormit orele "normale". (wikipedia, 2023)

În unele cazuri, „tulburarea somnului” a fost măsurată în funcție de un simptom specific (de exemplu, dificultate la inițierea somnului cel puțin 15 zile/lună ) și, în unele cazuri, a fost agravat (de exemplu, dificultăți de a adormi, de a rămâne adormit sau de a dormi prea mult, timp de cel puțin 6 din 14 zile). În unele cazuri, a existat o componentă de cronicitate, iar uneori a implicat doar o măsură subiectivă a severității. (Grandner, 2014)

Timp de aproximativ 50 de ani și în aproximativ 50 de studii, atât durata scurtă, cât și cea lungă a somnului au fost asociate cu risc crescut de mortalitate. Aceste studii au folosit o varietate de metode și modele, au studiat o varietate de populații diferite și au avut niveluri diferite de controale statistice. Luate împreună, totuși, se pare că apare o constatare solidă care leagă atât durata scurtă, cât și cea lungă de somn cu mortalitatea. Mecanismele acestei relații sunt neclare, deși mai multe studii au arătat că rezultatele adverse sunt asociate atât cu somnul scurt, cât și cu cel lung. (Grandner, 2014)

Simptomele de insomnie au fost raportate de mai mult de o treime din populația de 65 de ani și peste. Modelele multivariate au arătat că vârsta nu a fost un factor predictiv al simptomelor de insomnie atunci când se controlează starea activității și satisfacția vieții sociale. Nivelul de activitate și interacțiunile sociale nu au avut nicio influență asupra somnului, dar s-a constatat că vârsta are un efect pozitiv semnificativ asupra somnului. Aceste rezultate indică faptul că procesul de îmbătrânire în sine nu este responsabil pentru creșterea insomniei adesea raportată la persoanele în vârstă. În schimb, inactivitatea, nemulțumirea față de viața socială și prezența bolilor organice și a tulburărilor psihice au fost cei mai buni predictori ai insomniei, vârsta fiind nesemnificativă. Persoanele în vârstă sănătoase (adică, fără patologii organice sau mentale) au o prevalență a simptomelor de insomnie similară cu cea observată la persoanele mai tinere. (Ohayon M.M., 2001)

În cadrul studiului multicentrat al Institutului Național pentru Îmbătrânire, intitulat „Populații stabilite pentru Studii Epidemiologice ale Vârstnicilor” (EPESE), frecvența a cinci plângeri comune de somn a fost evaluată la peste 9.000 de participanți cu vârsta de 65 de ani și peste. Aceste plângeri includ dificultăți de a adormi, trezire frecventă pe parcursul nopții, trezire prea devreme dimineața, nevoia de a dormi în timpul zilei și senzația de a nu se simți odihnit după somn. Mai puțin de 20% dintre participanții din fiecare comunitate au avut rar sau niciodată plângeri, în timp ce peste jumătate au raportat că cel puțin una dintre aceste plângeri a apărut de cele mai multe ori. Între 23% și 34% au avut simptome de insomnie, iar între 7% și 15% la sută rar sau niciodată nu s-au simțit odihniți după trezire dimineața. În analizele multivariate, plângerile de somn au fost asociate cu un număr crescut de simptome respiratorii, dizabilități fizice, utilizarea medicamentelor fără prescripție, simptome depresive și o percepție de sănătate slabă. Tulburările de somn la persoanele în vârstă sunt adesea secundare bolilor coexistente. Pentru a determina prevalența specifică a tulburărilor de somn, sunt necesare măsuri mai sofisticate și obiective. (Daniel J. Foley, 1995)

## 5. Descrierea bazei de date

Pentru realizarea proiectul, am ales o bază de date care conține informații referitoare la sănătatea somnului și stilul de viață cum ar fi: valori cuprinzătoare ale somnului, factori de stil de viață, sănătatea cardiovasculară și analiza tulburărilor de somn. Această bază de date are înregistrate 400 de persoane notate cu câte un index unic și are următoarele câmpuri: un ID pentru fiecare persoană (**Person ID**), genul persoanei (**Gender**), vârsta persoanei în ani (**Age**), numărul de ore pe care persoana le doarme pe zi (**Sleep Durationn**), o evaluare subiectivă a calității somnului, variind de la 1 la 10 (**Quality of Sleep**), numărul de minute în care persoana se angajează zilnic în activitate fizică (**Physical Activity Level**), o evaluare subiectivă a nivelului de stres experimentat de persoană, variind de la 1 la 10 (**Stress Level**), Categoria IMC a persoanei (indicele de masă corporală a persoanei) (**BMI Category**), măsurarea tensiunii arteriale a persoanei, indicată ca presiune sistolica peste presiunea diastolică (**Blood Pressure**) ritmul cardiac de repaus al persoanei în bătăi pe minut (**Heart Rate**), numărul de pași pe care o face o persoană pe zi (**Daily Steps**) și prezența sau absența unei tulburări de somn la persoană clasificată în categorii precum „Niciuna”, „Insomnie”, „Apnee în somn” (**Sleep Disorder**). Această bază de date este utilă pentru îmbunătățirea sănătății și calitatea vieții.

## 6. Încărcarea și analiza bazei de date

Pentru început, am încărcat baza de date în RapidMiner, importând datele dintr-un document Excel. Am analizat întreaga baze de date și mi-am dat seama că există câteva probleme pe care va trebui să le rezolv. O primă problemă am observat-o la câmpul **Blood Pressure** unde presiune sistolica și presiunea diastolică sunt în aceeași coloană delimitate prin semnul ,,/”, ceea ce ne împiedică să mergem mai departe, să calculăm corelațiile și să aplicăm modele, iar la câmpurile **Gender**, **BMI Category**, **Sleep Disorder**, **Blood Pressure\_1** și **Blood Pressure\_2** fiind și ele de tip string.

## 7. Curățare și eliminare anomalii

Pentru rezolvarea uneia dintre problemele identificate în baza de date, am folosit operatorul Split din RapidMiner ce creează noi atribute din atributele nominale selectate prin împărțirea valorilor nominale în părți conform criteriului de împărțire care este specificat prin parametrul modelului de împărțire sub forma unei expresii regulate. Am adăugat un split cu ajutorul căruia am reușit să separ informațiile din coloana **Blood Pressure** în câmpuri individuale pentru a îmbunătății corectitudinea datelor având ca rezultat următoarele coloane: **Blood Pressure\_1** și **Blood Pressure\_2**.

Pe lângă operatorul Split, am utilizat și operatorului Nominal to Numerical care schimbă tipul de atribute nenumerice selectate într-un tip numeric. Am folosit acest operator pentru urmatoarele 5 câmpuri: **Gender**, **BMI Category**, **Sleep Disorder**, **Blood Pressure\_1** și **Blood Pressure\_2** dându-le câte un index unic.

Modificările pe care le-am făcut asupra bazei de date le-am cuprins într-un subprocess pentru o mai bună organizare în RapidMiner. Această operație am făcut-o prin selectarea bazei de date și ale operatorilor Split si Nominal to Numerical, și selectând Move into new subprocess.

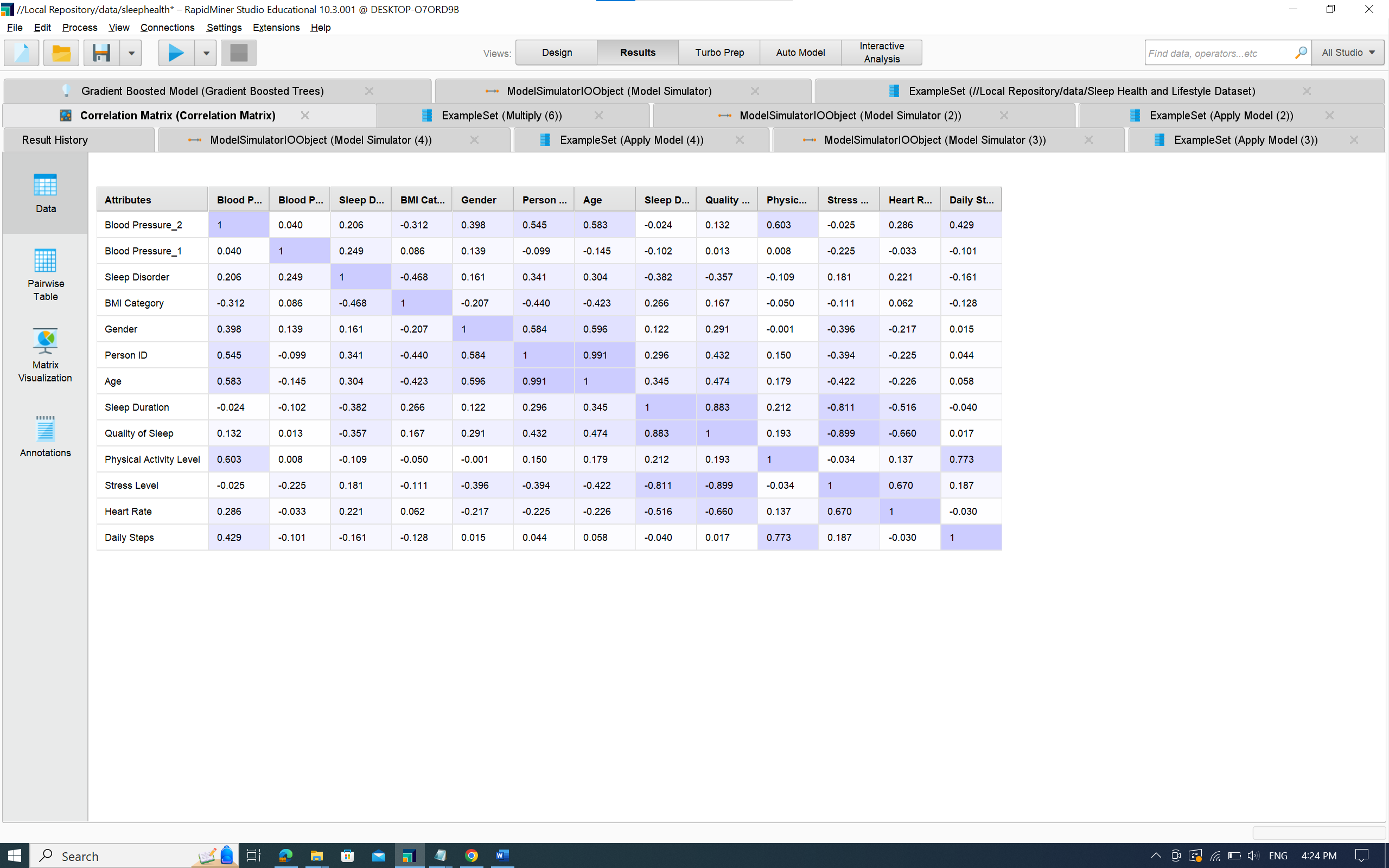
**O imagine care conține text, captură de ecran, software, Pictogramă computer

Descriere generată automat**

După modificările făcute, baza de date o să conțină următoarele câmpuri: **Person ID, Gender, Age, Sleep Durationn, Quality of Sleep, Physical Activity Level, Stress Level, BMI Category, Blood Pressure\_1, Blood Pressure\_2, Heart Rate, Daily Steps** și **Sleep Disorder.**

## 8. Corelații

Am determinat matricea de corelație folosind operatorul Correlation Matrix din Rapid Miner. Acest operator determină corelația dintre toate atributele și poate produce un vector de ponderi pe baza acestor corelații. Corelația este o tehnică statistică care poate arăta dacă și cât de puternic sunt legate perechile de atribute. Valorile prezente cuprinse în intervalul (-1, 1) sugerează corelația dintre atribute. O valoare mai mare de 0.3 indică o corelație puternică între cele două atribute, deci existența unei legi de variație între acestea, iar o valoare mai mică de 0.3 determină lipsa unei relații între respectivele variabile. După determinarea matricei de corelație, am analizat în detaliu toate datele afișate pentru a identifica care corelațiile puternice.



O corelație de **-0.889** între calitatea somnului (**Quality of Sleep**) și nivelul de stres (**Stress Level**) sugerează că există o relație puternică și inversă între aceste două variabile. În acest caz, o valoare negativă indică că atunci când una dintre variabile crește, cealaltă tinde să scadă și invers. Stresul poate avea un impact semnificativ asupra calității somnului. Persoanele care experimentează niveluri mai ridicate de stres pot avea dificultăți în adormire, pot avea un somn fragmentat sau pot suferi de insomnie, ceea ce poate reduce calitatea generală a somnului. Pe de altă parte, calitatea somnului influențează capacitatea organismului de a gestiona și de a face față stresului. Un somn odihnitor și de calitate poate contribui la reglarea nivelului de stres și la reducerea reacțiilor excesive la factorii stresanți din viața de zi cu zi. În concluzie, o corelație puternică și inversă între calitatea somnului și nivelul de stres sugerează că îmbunătățirea calității somnului poate fi benefică pentru gestionarea stresului și invers.

O corelație de **0.883** între calitatea somnului (**Quality of Sleep**) și durata somnului (**Sleep Duration**) indică o asociere semnificativă și pozitivă între aceste două variabile. Aceasta înseamnă că, în general, o creștere a duratei somnului este asociată cu o îmbunătățire a calității somnului, și invers. O perioadă adecvată de somn poate influența pozitiv calitatea somnului. Persoanele care dorm mai mult pot avea ocazia să treacă prin mai multe cicluri de somn și să ajungă la stadiile mai profunde și mai odihnitoare ale somnului, ceea ce poate duce la o calitate mai bună a somnului. În concluzie, corelația pozitivă și semnificativă între durata somnului și calitatea acestuia sugerează că îmbunătățirea duratei somnului ar putea fi asociată cu o îmbunătățire a calității somnului și invers. Totuși, este important să se ia în considerare și alți factori care ar putea influența această relație pentru a înțelege mai bine dinamica acestor variabile.

O corelație de **-0.811** între nivelul de stres (**Stress Level**) și durata somnului (**Sleep Duration**) sugerează o relație semnificativă și inversă între aceste două variabile. Asta înseamnă că, în general, pe măsură ce nivelul de stres crește, durata somnului tinde să scadă și invers. Stresul intens poate perturba calitatea și durata somnului. Persoanele care au un nivel ridicat de stres pot avea dificultăți în adormire, pot avea un somn fragmentat sau pot suferi de insomnie, ceea ce poate duce la o scădere a duratei somnului. Pe de altă parte, lipsa somnului sau calitatea scăzută a acestuia poate crește nivelul de stres perceput de o persoană. Un somn insuficient sau de proastă calitate poate afecta capacitatea organismului de a face față stresului și poate contribui la creșterea reacțiilor excesive la factorii stresanți din viața de zi cu zi. În concluzie, corelația inversă semnificativă între nivelul de stres și durata somnului sugerează că gestionarea stresului poate fi esențială pentru îmbunătățirea calității și duratei somnului, iar invers.

O corelație de **0.773** între numărul de pași zilnici (**Daily Steps**) și nivelul de activitate fizică (**Physical Activity Level**) sugerează o relație semnificativă și pozitivă între aceste două variabile. Aceasta înseamnă că, în general, o creștere a numărului de pași zilnici este asociată cu un nivel mai ridicat al activității fizice, și invers. Este de așteptat ca persoanele care sunt mai active să facă mai mulți pași în decursul unei zile. Activitățile fizice, cum ar fi mersul, alergarea, implică de obicei un număr mai mare de pași, contribuind astfel la o creștere a numărului de pași zilnici. Creșterea numărului de pași zilnici poate contribui la creșterea nivelului general de activitate fizică, ceea ce poate duce la îmbunătățirea stării de sănătate și a stării de bine. Acest lucru pozitiv poate încuraja persoanele să își mențină sau să își crească nivelul de activitate fizică pe termen lung. În concluzie, corelația pozitivă semnificativă între numărul de pași zilnici și nivelul de activitate fizică indică că monitorizarea și creșterea numărului de pași zilnici ar putea fi un mod eficient de a încuraja și de a menține un nivel ridicat al activității fizice, cu beneficii semnificative pentru sănătate.

O corelație de **0.670** între ritmul cardiac (**Heart Rate**) și nivelul de stres (**Stress Level**) sugerează o relație semnificativă și pozitivă între aceste două variabile. Aceasta înseamnă că, în general, o creștere a nivelului de stres este asociată cu o creștere a ritmului cardiac și invers. Atunci când o persoană se confruntă cu stresul, sistemul său nervos poate reacționa prin eliberarea de hormoni de stres, cum ar fi cortizonul și adrenalina, care pot crește ritmul cardiac. Acest lucru este o reacție naturală a organismului la situații de stres, menită să pregătească corpul pentru a face față acestuia. Stresul intens poate pune o presiune suplimentară asupra sistemului cardiovascular și poate contribui la creșterea frecvenței cardiace pe termen lung. Acest lucru poate avea efecte negative asupra sănătății inimii și poate crește riscul de afecțiuni cardiovasculare, cum ar fi hipertensiunea arterială sau boli de inimă. În concluzie, corelația semnificativă și pozitivă între ritmul cardiac și nivelul de stres indică o legătură strânsă între aceste două variabile, fiind importantă pentru sănătatea cardiovasculară și gestionarea stresului.

O corelație de **-0.660** între ritmul cardiac (**Heart Rate**) și calitatea somnului (**Quality of Sleep**) sugerează o relație semnificativă și inversă între aceste două variabile. Aceasta înseamnă că, în general, o creștere a ritmului cardiac este asociată cu o scădere a calității somnului și invers. Un somn de proastă calitate sau perturbat poate duce la creșteri ale ritmului cardiac pe parcursul nopții. Tulburările de somn, cum ar fi insomniile sau apneea de somn, pot afecta negativ ritmul cardiac și pot duce la creșteri ale acestuia în timpul nopții. Persoanele care se confruntă cu probleme de somn sau insomnii pot experimenta un nivel crescut de stres și anxietate legate de somnul lor. Acest stres asociat cu calitatea scăzută a somnului poate duce la creșteri ale ritmului cardiac în timpul nopții. În concluzie, corelația inversă semnificativă între ritmul cardiac și calitatea somnului indică o legătură strânsă între aceste două variabile, având importanță pentru sănătatea cardiovasculară și calitatea somnului.

O corelație de **0.603** între nivelul de activitate fizică (**Physical Activity Level**) și tensiunea arterială (**Blood Pressure\_2**) sugerează o relație semnificativă și pozitivă între aceste două variabile. Aceasta înseamnă că, în general, o creștere a nivelului de activitate fizică este asociată cu o creștere a tensiunii arteriale și invers. Activitățile fizice intense pot duce la creșteri temporare ale tensiunii arteriale, pe măsură ce inima pompează mai mult sânge în timpul efortului fizic. Aceste creșteri ale tensiunii arteriale pot fi observate imediat după exercițiu. Cu toate acestea, pe termen lung, exercițiile fizice regulate pot duce la scăderi ale tensiunii arteriale în repaus, prin îmbunătățirea sănătății cardiovasculare și a funcției inimii și vaselor de sânge. Astfel, chiar dacă exercițiile fizice pot crește inițial tensiunea arterială, ele pot contribui la reducerea riscului de hipertensiune arterială pe termen lung. În concluzie, corelația semnificativă și pozitivă între nivelul de activitate fizică și tensiunea arterială indică o legătură complexă între aceste două variabile, cu implicarea atât a efectelor imediate, cât și a adaptărilor pe termen lung ale activității fizice asupra tensiunii arteriale.

O corelație de **0.583** între vârstă (**Age**) și tensiunea arterială (**Blood Pressure\_2**) sugerează o relație semnificativă și pozitivă între aceste două variabile. Aceasta înseamnă că, în general, o creștere a vârstei este asociată cu o creștere a tensiunii arteriale și invers. Pe măsură ce oamenii îmbătrânesc, este normal să se observe o creștere a tensiunii arteriale. Acest lucru se datorează adesea schimbărilor pe parcursul procesului de îmbătrânire. Persoanele mai în vârstă pot fi mai predispuse să prezinte alți factori de risc pentru tensiune arterială ridicată, cum ar fi obezitatea, nivelul ridicat de stres. Acești factori pot contribui la creșterea tensiunii arteriale și pot fi prezenți într-o măsură mai mare la persoanele mai în vârstă. În concluzie, corelația semnificativă și pozitivă între vârstă și tensiunea arterială indică o legătură strânsă între aceste două variabile, având importanță pentru sănătatea cardiovasculară și îmbătrânirea sănătoasă.

O corelație de **-0.516** între ritmul cardiac (**Heart Rate**) și durata somnului (**Sleep Duration**) indică o relație semnificativă și inversă între aceste două variabile. Aceasta sugerează că, în general, o creștere a duratei somnului este asociată cu o scădere a ritmului cardiac, iar o scădere a duratei somnului este asociată cu o creștere a ritmului cardiac. În timpul somnului, ritmul cardiac este influențat de sistemul nervos, care reglează funcțiile involuntare ale corpului, cum ar fi ritmul cardiac și respirația. O durată adecvată a somnului poate contribui la o reglare mai eficientă a acestui sistem, ducând la o scădere a ritmului cardiac în timpul somnului. Un somn de calitate și suficient poate contribui la reducerea nivelului de stres și anxietate a unei persoane. Scăderea stresului și a anxietății poate avea un impact pozitiv asupra ritmului cardiac în repaus. În concluzie, corelația semnificativă și inversă între ritmul cardiac și durata somnului sugerează că un somn de calitate și suficient poate fi asociat cu o reglare mai eficientă a ritmului cardiac în timpul repausului.

O corelație de **-0.468** între categoria IMC (**BMI Category**) și tulburările de somn (**Sleep Disorder**) indică o relație semnificativă, dar moderată, inversă între aceste două variabile. Aceasta sugerează că există o asociere între categoria IMC și prezența tulburărilor de somn, astfel încât cu cât IMC-ul este mai scăzut, cu atât este mai probabil să existe tulburări de somn, și invers. Persoanele cu IMC mai mic pot fi mai predispuse să aibă anumite tulburări de somn, cum ar fi insomniile sau apneea în somn. Greutatea corporală poate influența modul în care o persoană doarme, iar un IMC scăzut poate fi asociat cu o rată crescută a acestor tulburări. Tulburările de somn pot fi asociate cu o serie de afecțiuni de sănătate care sunt mai întâlnite la persoanele cu IMC mai scăzut sau mai mare. De exemplu, insomniile pot fi legate de stresul și anxietatea asociate cu îngrijorările legate de greutatea corporală, în timp ce apneea în somn poate fi mai frecventă la persoanele supraponderale sau obeze. În concluzie, corelația inversă între categoria IMC și prezența tulburărilor de somn sugerează că greutatea corporală poate fi un factor important de luat în considerare în evaluarea riscului de tulburări de somn.

O corelație de **0.474** între vârstă (**Age**) și calitatea somnului (**Quality of Sleep**) indică o relație semnificativă, dar moderată, între aceste două variabile. Aceasta sugerează că există o asociere pozitivă între vârstă și calitatea somnului, în sensul că, în general, o creștere a vârstei este asociată cu o calitate mai bună a somnului. Oamenii tind să aibă nevoi de somn diferite pe măsură ce îmbătrânesc. În general, adulții mai în vârstă pot avea tendința să aibă un somn mai fragmentat sau să se trezească mai des în timpul nopții, dar să se simtă totuși odihniți și refăcuți în timpul zilei.

## 9. Modele

Acuratețea reprezintă proporția de predicții corecte făcute de un model de clasificare sau de regresie. În cazul acesta, avem următoarele rezultate pentru trei modele de învățare automată: un arbore de decizie, un random forest și un gradient boosted trees.

Acuratețea modelului Decision Tree este de 85,35%. Arborele de decizie este un model simplu care construiește un set de reguli de decizie pe baza caracteristicilor setului de date. Acuratețea ridicată poate fi rezultatul faptului că modelul a reușit să identifice datele de antrenare într-un mod eficient.

O imagine care conține text, captură de ecran, software, Pictogramă computer

Descriere generată automat

Predicția modelului este 0. Valorile categoriei BMI nu susțin totuși această decizie. Root Mean Squared Error (RMSE) a tuturor predicțiilor realizate de acest model este 0,442. Iar eroarea relativă este de aproximativ 14,65%.

Acuratețea modelului Random forest este de 27,54%. Random forest este o metodă de ansamblu care utilizează mai mulți arbori de decizie pentru a face predicții. Fiecare arbore din random forest este antrenat pe un subgrup diferit din setul de date și utilizează un subset diferit de caracteristici pentru a face split-uri în cadrul nodurilor. Acuratețea scăzută a acestui model poate indica o lipsă de adaptabilitate la datele de testare.

O imagine care conține text, captură de ecran, software, număr

Descriere generată automat

Predicția modelului este 0,393. Valorile categoriei IMC nu susțin totuși această decizie. Root Mean Squared Error (RMSE) a tuturor predicțiilor realizate de acest model este 0,610. Iar eroarea relativă este de aproximativ 72,78%.

Acuratețea modelului Gradient boosted trees este de 27,22%. Gradient boosted trees este o altă metodă de ansamblu care construiește un ansamblu de arbori de decizie într-o manieră secvențială, concentrându-se pe exemplele care sunt greșit clasificate de modelele anterioare. Acest model este similar cu random forest, având o acuratețe scăzută.

O imagine care conține text, captură de ecran, software, număr

Descriere generată automat

Predicția modelului este 0,468. Cel mai mare sprijin pentru această decizie vine de la Age. Root Mean Squared Error (RMSE) a tuturor predicțiilor realizate de acest model este 0,453. Iar eroarea relativă este de aproximativ 72,46%.

Pe baza rezultatelor obținute, reiese faptul că cel mai bun model de învățare pentru problema noastră este Decision Tree având acuratețea de 85,35%.

## 10. Testare și validare

Am împărțit baza de date în două subseturi folosind operatorul Split Data, alocând 70% pentru antrenament și 30% pentru testare, pentru a evalua corect performanța modelelor de învățare automată. Datele de antrenament sunt utilizate pentru a construi și ajusta modelele. După antrenament, modelele sunt testate pe setul de date de testare. Această împărțire este esențială pentru a evita supraantrenarea și pentru a asigura că modelele pot face predicții precise.

## 11. Rezultate

Analizând corelațiile care au loc între diferitele variabile din setul de date, am remarcat legături importante între factorii de sănătate și calitatea somnului. O corelație puternică inversă de -0.889 între calitatea somnului și nivelul de stres evidențiază impactul direct al stresului asupra somnului, evidențiind importanța gestionării acestuia pentru un somn odihnitor, iar corelația pozitivă de 0.883 între calitatea somnului și durata acestuia indică o legătură directă între o perioadă adecvată de somn și îmbunătățirea calității acestuia.

În urma evaluării diferitelor modele de învățare automată, am observat că modelul Decision Tree se evidențiază ca fiind cel mai eficient, obținând o acuratețe de 85.35%. Acest rezultat sugerează că modelul poate fi utilizat pentru a prezice sănătatea și calitatea somnului pe baza datelor disponibile în setul nostru de date. Însă, este important să subliniem faptul că, în ciuda preciziei ridicate, modelul prezintă o eroare de aproximativ 14.65%. Această eroare indică faptul că există încă aspecte care pot fi îmbunătățite sau variabile care ar putea fi luate în considerare pentru a perfecționa predicțiile.

## 12. Concluzii

Prin analiza bazei de date "Sănătatea somnului și stilul de viață" am observat corelații semnificative între diferite aspecte ale sănătății și calității somnului. Printre acestea, s-au evidențiat corelații între calitatea somnului și nivelul de stres, durata somnului și nivelul de activitate fizică, între durata somnului și calitatea acestuia. Persoanele cu niveluri mai ridicate de stres pot avea dificultăți în adormire și pot suferi de un somn fragmentat sau de insomnie. De asemenea, durata somnului este asociată pozitiv cu calitatea somnului, ceea ce sugerează că o perioadă adecvată de somn poate contribui la îmbunătățirea calității acestuia. Nivelul de activitate fizică este, de asemenea, un factor important, cu o corelație pozitivă între numărul de pași zilnici și nivelul de activitate fizică. Activitatea fizică regulată poate îmbunătăți calitatea somnului și poate contribui la un somn mai odihnitor. Stresul, durata somnului și nivelul de activitate fizică sunt doar câteva dintre variabilele din baza de date care pot influența sănătatea și calitatea somnului.

Pentru a prezice sănătatea și calitatea somnului pe baza acestor informații, am folosit mai multe modele de învățare automată, inclusiv Decision Tree, Random Forest și Gradient Boosted Trees. Cu toate acestea, modelul Decision Tree s-a dovedit a fi cel mai eficient, atingând o acuratețe de aproximativ 85,35%. Performanța acestui model sugerează că poate fi utilizat pentru a face predicții precise pe baza datelor disponibile.

Prin împărțirea bazei de date în subseturi de antrenament și testare, am evaluat performanța fiecărui model și am identificat modelul cu cea mai bună capacitate de predicție.

Proiectul realizat oferă informațiile necesare pentru înțelegerea și predicția sănătății somnului și stilului de viață. Rezultatele obținute au evidențiat importanța gestionării stresului, menținerii unei activități fizice regulate și asigurării unui somn adecvat pentru promovarea unei sănătăți optime.

13. Surse:

<https://ro.wikipedia.org/wiki/Somn>

<https://docs.rapidminer.com/latest/studio/operators/index.html>

<https://www.reginamaria.ro/articole-medicale/somnul-la-fel-de-necesar-ca-apa-si-mancarea>

<https://www.tandfonline.com/doi/full/10.3109/09540261.2014.911148>

<https://agsjournals.onlinelibrary.wiley.com/doi/full/10.1046/j.1532-5415.2001.49077.x?casa_token=X84-sG6JYdwAAAAA%3AaGVhm83fX5rMt-zWIxioKoebYy0CU-X8y_UuSzEUzjPQQu5Jc0j4rfBwY6XjDhYBMhgVrwmUHODEjLtR>

<https://academic.oup.com/>