**UNIVERSITATEA TEHNICĂ „GHEORGHE  ASACHI”  IAȘI**

**FACULTATEA DE AUTOMATICĂ ȘI CALCULATOARE**

**DOMENIUL: Calculatoare și Tehnologia Informației**

**SPECIALIZAREA: Tehnologia Informației**

Inteligență Artificială - Proiect

**Inferența prin enumerare în rețele bayesiene**

**Coordonator,**

Asist. drd. ing. Codruț-Georgian Artene

**Autori,**

Dima Raul Andrei, 1411B

 Melinte Alexandru-Gicu, 1411B

Nistor Paula-Alina, 1411B

**IAȘI**

**2021-2022**

**Cuprins**

[**1. Descrierea problemei considerate 1**](#_Toc92629977)

[**2. Aspecte teoretice privind algoritmul 1**](#_Toc92629978)

[2.1 Probabilitatea 1](#_Toc92629979)

[2.2 Probabilitatea condiționată 1](#_Toc92629980)

[2.3 Teorema lui Bayes 1](#_Toc92629981)

[2.4 Independență și independență condiționată 2](#_Toc92629982)

[2.5 Distribuție comună de probabilitate 2](#_Toc92629983)

[2.6 Rețele Bayesiene 2](#_Toc92629984)

[2.7 Sortarea topologică 2](#_Toc92629985)

[2.8 Inferență prin enumerare 2](#_Toc92629986)

[**3. Modalitatea de rezolvare 3**](#_Toc92629987)

[**4. Cod sursă, explicații, comentarii - blocuri semnificative 5**](#_Toc92629988)

[4.1 Metoda EnumerationAsk 5](#_Toc92629989)

[4.2 Metoda EnumerateAll 6](#_Toc92629990)

[4.3 Metoda KahnSorting – Sortare topologică 7](#_Toc92629991)

[4.4 Constructorul rețelei bayesiene 8](#_Toc92629992)

[**5. Rezultate obținute 9**](#_Toc92629993)

[**6. Concluzii 11**](#_Toc92629994)

[**7. Rolul membrilor din echipă 11**](#_Toc92629997)

[**8. Bibliografie 11**](#_Toc92629998)

# 1. Descrierea problemei considerate

Proiectul vine ca și suport pentru descrierea unei aplicații ce primește date stocate într-un fișier .txt despre nodurile, legăturile dintre noduri și evidențele acestora. Aceste date reprezintă structura și parametrii unei rețele bayesiene, iar pe baza acestor date se dorește a se determina, cu ajutorul algoritmului de inferență prin enumerare, rezultatele probabilistice dorite pentru orice tip de rețea.

Mai exact se dorește ca aplicația să aibă posibilitatea interogării unui anumit nod din rețeaua bayesiană pentru a extrage date din divereste situații (prin setarea evidențelor) în care acesta este implicat.

# 2. Aspecte teoretice privind algoritmul

**2.1 Probabilitatea**

Dacă un eveniment **A**, ce face parte dintr-o mulțime de evenimente **Ω ,** se poate realiza în *s* probe dintr-un total de *n* încercări echiprobabile pe care experimentul le poate produce, atunci probabilitatea evenimentului A se poate defini prin formula:

**P(A) = =** și 0 **P(A)** 1, A Ω               (2.1)

**2.2 Probabilitatea condiționată**

Dacă A şi B sunt două evenimente arbitrare, prin **probabilitatea condiţionată** a lui A de către B, notată **P(A|B)**, se înţelege probabilitatea de a se realiza evenimentul A dacă în prealabil s-a realizat evenimentul B. Prin definiție:  
 **P(A|B) =**                                         (2.2)

Putem exprima probabilitatea intersecției în două moduri și de aici deducem expresia lui P(B|A) în funcție de P(A|B):

**P(AB)** = P(A|B) ∙ P(B),     (2.3)

**P(AB)** = P(B|A) ∙ P(A),     (2.4)

**2.3 Teorema lui Bayes**

**Teorema lui Bayes** descrie probabilitatea unui eveniment date fiind condițiile ce ar putea duce la apariția evenimentului.  
 **P(A|B)** = , **P(E|I)** = , (2.5)

**I  -** ipoteza;

**E -** evidența (provenită din datele observate);

**P(I) -** probabilitatea a-priori a ipotezei (gradul inițial de încredere în ipoteză);

**P(E)** - probabilitatea evidenței;

**P(E|I) -** verosimilitatea datelor observate (măsura în care s-a observat evidența în condițiile îndeplinirii ipotezei);

**P(I|E)** - probabilitatea a-posteriori a ipotezei, dată fiind evidența.

Cu ajutorul teoremei, putem calcula astfel probabilitățile cauzelor, date fiind efectele. Este mai simplu de cunoscut când o cauză determină un efect, dar invers, când cunoaștem un efect, probabilitățile cauzelor nu pot fi cunoscute imediat.

## 2.4 Independență și independență condiționată

Două evenimente A si B sunt **independente** dacă probabilitatea realizării (sau nerealizării) oricăruia din cele două evenimente nu se modifică în funcție de realizarea, nerealizarea sau ignorarea celuilalt. Matematic, aceste noțiuni s-ar traduce în formula următoare:

**P(AB)** = P(A) ⋅ P(B) (2.6)

**Independența condiționată** reprezintă independența bazată pe îndeplinirea unor condiții.  
 **P(B|A) = P(B) = P(B|)** (2.7)

## 2.5 Distribuție comună de probabilitate

În [probabilitate](https://koaha.org/wiki/Probabilit%C3%A0), având în vedere două [variabile aleatorii](https://koaha.org/wiki/Variabile_aleatoria) *X* și *Y* , definite pe același [spațiu de probabilitate](https://koaha.org/wiki/Spazio_di_probabilit%C3%A0), **distribuția** lor **comună** este definită ca [distribuția de probabilitate](https://koaha.org/wiki/Distribuzione_continua) asociată vectorului (*X*, *Y*).

## 2.6 Rețele Bayesiene

Literatura de specialitate oferă multe definiţii generale ale unei rețele Bayesiene, redate în forme destul de diferite:

* O **rețea bayesiană** reprezintă o modalitate de vizualizare grafică şi de analiză a modelelor care implică incertitudinea, incertitudine gestionată într-un mod matematic riguros, eficient şi simplu.
* O **rețea bayesiană** este o reprezentare a unei distribuţii comune de probabilitate a unui set de variabile cu posibile legături mutuale cauzale între ele.
* O **rețea bayesiană** este o reprezentare grafică a unor mărimi (şi decizii) incerte care, în mod explicit, relevă dependenţa cauzală între variabile ca şi circulaţia informaţiei în modelul procesului analizat.
* O **rețea bayesiană** se referă la o metodă sistematică şi concretă pentru structurarea informaţiei cu caracter probabilistic într-un mod coerent şi cu ajutorul unor algoritmi de inferenţă.
* O **rețea bayesiană** este un graf orientat aciclic (engl. “directed acyclic graph”), în care evenimentele sau variabilele se reprezintă ca noduri, iar relațiile de corelație sau cauzalitate se reprezintă sub forma arcelor dintre noduri.
* O **rețea bayesiană** este un model grafic probabilistic, adică un graf cu o mulțime de noduri, care reprezintă evenimente aleatorii, conectate de arce, care reprezintă dependențe condiționate între evenimente.

## 2.7 Sortarea topologică

**Sortarea topologică** a unui graf este o ordonare liniară a nodurilor sale astfel încât, pentru fiecare arc A → B, A apare înaintea lui B. Pentru o **rețea bayesiană**, **sortarea topologică** asigură faptul că nodurile părinte vor apărea înaintea nodurilor fiu. Dacă există cicluri în graf, sortarea topologică este imposibilă.

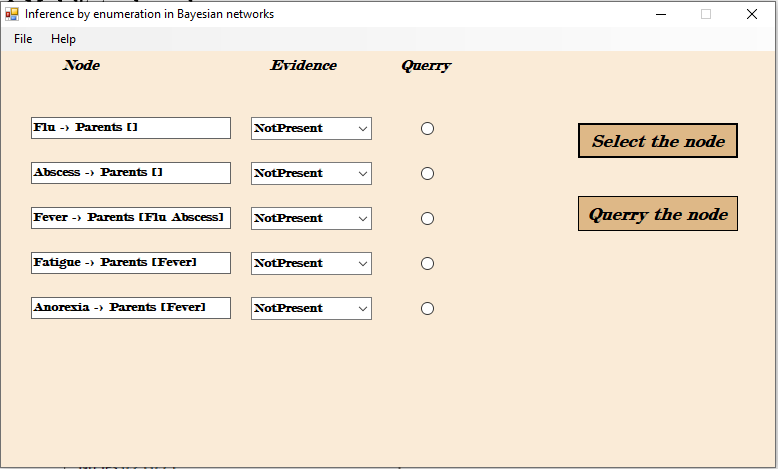
## 2.8 Inferență prin enumerare

**Inferența** este un proces [logic](https://ro.wikipedia.org/wiki/Logic%C4%83) care derivă o [concluzie](https://ro.wikipedia.org/wiki/Concluzie) dintr-o [premisă](https://ro.wikipedia.org/wiki/Premis%C4%83), adică extrage o consecință necesară, o informație specifică, dintr-o descriere de stare dată.

**Inferența prin enumerare** are scopul de a calcula probabilitatea unei variabile interogate (engl. “query”), date fiind variabilele observate (evidenţă).

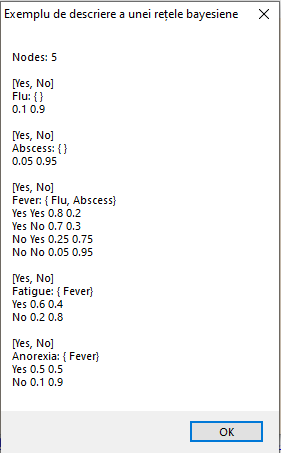
# 3. Modalitatea de rezolvare

Pentru interogarea unui nod dintr-o rețea bayesiană, utilizatorul dispune de o interfață unde poate vizualiza nodurile, evidențele și poate selecta nodul dorit. Se va încărca un fișier .txt ce conține parametrii rețelei, adică numărul de noduri, evidența fiecărui nod în parte, relațiile dintre noduri și probabilitățile.



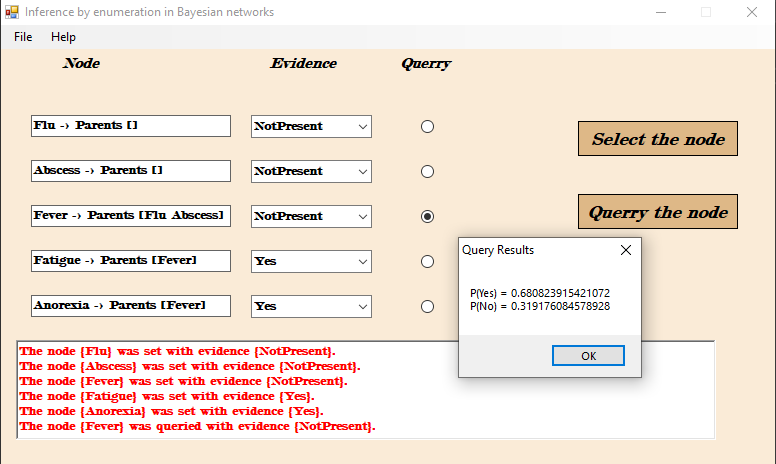
**Figura 1**

În secțiunea help a aplicației se regăsește un model .txt ce conține descrierea rețelei bayesiene (*Figura 2*).



**Figura 2**

Prin selectarea unui nod și/sau a unor evidențe, la apăsarea butonului *Querry the node* se va apela algoritmul de inferență prin enumerare și va furniza probabilitatea nodului selectat pentru rețeaua bayesiană.



**Figura 3**

În acest exemplu (*Figura 2*), se consideră că atât gripa (*Flu*) cât și abcesul (*Abscess*) pot determina febra (*Fever*). De asemenea, febra poate cauza o stare de oboseală (*Fatigue*) sau anorexie (*Anorexia*).

Fiecare variabilă are o mulțime de valori. În cazul cel mai simplu, variabilele au evidențe  
binare, de exemplu Da (Yes) și Nu (No). În general însă, o variabilă poate avea oricâte evidențe.

Asociate cu variabilele, o rețea bayesiană conține o serie de tabele de probabilități, precum  
cele din *Figura 2*. Pentru nodurile fără părinți se indică probabilitățile marginale ale fiecărei valori  
(adică fără a lua în considerare valorile celorlalte variabile). Pentru celelalte noduri, se indică  
probabilitățile condiționate pentru fiecare valoare, ținând cont de fiecare combinație de valori ale  
variabilelor părinte. În general, o variabilă binară fără părinți va avea un singur parametru independent, o variabilă cu 1 părinte va avea 2 parametri independenți iar o variabilă cu n părinți va avea parametri independenți în tabela de probabilități corespunzătoare.

Folosindu-ne de algoritmul de inferență prin enumerare, avem posibilitatea de a răspunde la orice întrebare privind nodurile din rețea setate cu diferite evidențe.

Ideea de bază la algorimul de inferență prin enumerare este calcularea unu produs de probabilități condiționate, însă în cazul variabilelor despre care nu se cunoaște nimic (*NotPresent*), se sumează variantele corespunzătoare tuturor valorilor acestora.

Exemplu dat răspunde la întrebarea „Care este probabilitatea de a avea febra (*Fever*),  
dacă prezintă simptome de oboseală (*Fatigue*) și anorexie (*Anorexia*)? ”, obținându-se:  
 **P(Yes)** 0.69

**P(No)**  0.31

# 4. Cod sursă, explicații, comentarii - blocuri semnificative

## 4.1 Metoda EnumerationAsk

Text

Description automatically generated

Text

Description automatically generated

**Figura 4**

**Metoda EnumerationAsk** primește ca parametru de intrare nodul ce se dorește a fi interogat și furnizează probabilitățile în funcție de evidențe. Se construiesc variabilele funcției: lista de noduri sortată topologic (*nodes*), nodul din graf ce se dorește a fi interogat (*querryNode*) și un vector ce conține probabilitățile (*q*). În caz de nodul interogat nu există în graful rețelei, se va genera o excepție cu mesajul *Querry node – not found!.* Se vor calcula probabilitățile pentru nodul interogat în funcție de evidențele selectate. Dacă evidențele sunt setate pe valoarea *NOT\_PRESENT,* se vor genera probabilitățile marginale. În caz contrar, se vor genera probabilitățile calculate în situațiile alese. Se va returna vectorul de valori normalizat.

## 4.2 Metoda EnumerateAll

Text

Description automatically generatedText

Description automatically generated

**Figura 5**

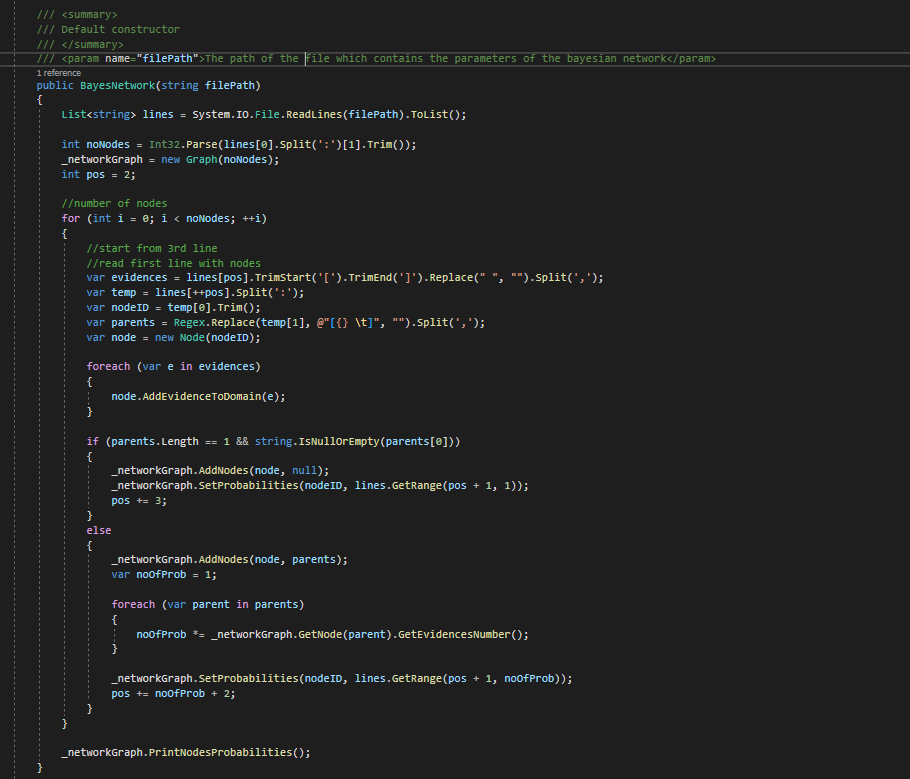
**Metoda EnumerationAll** este o metodă recursivă ceprimește ca parametrii de intrare lista de noduri și domeniul de evidențe și va returna probabilitatea nodului în situația dată (evidențele selectate). În caz de numărul de noduri este 0, se va returna valoarea reală *1.0*. Altfel, se ia fiecare nod în parte și se verifică daca are părinți, care este numărul lor și care sunte evidențele. În caz de evidențele selectate corespund valorii *NOT\_PRESENT*, se va calcula în funcție de probabilitățile marginale. In caz contrat, se calculeaza recursiv probabilitea fiecărui nod.

## 4.3 Metoda KahnSorting – Sortare topologică

**Figura 6**

**Metoda KahnSorting** este metoda echivalentă cu sortarea topologică, adică se vor sorta vârfurile unui graf orientat aciclic astfel încât, dacă există o legătura directă între nodul A și nodul B (adică un arc (A, B)), atunci nodul A va apărea înaintea nodului B în lista de noduri sortată.

## 4.4 Constructorul rețelei bayesiene

**Figura 7**

**Constructorul BayesNetwork** este constructorul clasei BayesNetwork (echivalentul rețelei bayesiene) și prin intermediul acestuia se contruiește rețeaua bayesiană. Acesta are ca parametru de intrare path-ul către fișierul *.txt* ce conține parametrii și evidențele unei rețele. Se va face parsarea din fișier pentru a extrage datele necesare creării unei rețele bayesiene și se vor afișa pe consolă.

# **5. Rezultate obținute**

**Întrebare**:Care este probabilitatea de a avea gripă (*Flu*) daca ai oboseală (*Fatigue*) și anorexie (*Anorexia*)?

**Rezultat: P(Yes)** 0.39

**P(No)**  0.61

**Vizualizare interfață:**

**Graphical user interface, application

Description automatically generated**

**Figura 8**

**Întrebare**:Care este probabilitatea de a avea gripă (*Flu*) daca ai oboseală (*Fatigue*) și anorexie (*Anorexia*)?

**Rezultat: P(Yes)** 0.2

**P(No)**  0.8

**Vizualizare interfață:**

Graphical user interface, application

Description automatically generated

**Figura 9**

**Întrebare**:Care este probabilitatea de a se întâmpla evenimentul alarmă (*Alarm*) dacă nu s-a întâmplat un eveniment de tip hoț (*Burglary*), cutremur (*Earthquake*), dar s-au întâmplat evenimentele de tip apel Ionuț (*JohnCalls*) și apel Maria (*MarryCalls*)?

**Rezultat: P(Yes)** 0.56

**P(No)**  0.44

**Vizualizare interfață:**

Graphical user interface, application

Description automatically generated

**Figura 10**

**Întrebare**:Dacă o mașină este pe segmentul S1 (*Sectiune1*), banda B1 (*Banda1*), care este probabilitatea ca după intersecție, pe drum (*Drum*) să meargă înainte (**Inainte**)?

**Rezultat: P(Banda1)** 0.36 **P(Banda3)** 0.10

**P(Banda2)**  0.54 **P(Banda4)** 0

**Vizualizare interfață:**

Graphical user interface

Description automatically generated

**Figura 11**

# **6. Concluzii**

Aplicația respectă cerințele impuse și anume, cu ajutorul algorimului de inferență prin enumerare se calculează interoghează un nod și se furnizează mai multe probabilități în funcție de situația aleasă și oferă posibilitatea utilizatorului de a crea propria rețea bayesiană prin intermediul unui fișier .txt.

Aplicația ar putea fi îmbunătățită prin crearea unei interfețe mai complexe, unde se pot desena automat și grafurile în funcție de parametrii rețelei din fișier.

# 7. Rolul membrilor din echipă

* **Dima Raul Andrei** – structurile rețelei (clasa Node, Graph, BayesNetwork), algoritmul de inferență prin enumerare.
* **Melinte Alexandru-Gicu** – interfața cu utilizatorul, documentație, comentarii cod.
* **Nistor Paula-Alina** – algoritmul de inferență prin enumerare (clasa InferenceByEnumeration), crearea domeniului de evidențe, creare fișiere de configurare a rețelei bayesiene.

Codul sursă este încărcat pe github (*Branch*: master), unde se pot vizualiza și commit-urile în funcție de sarcinile atribuite.

Link: [*https://github.com/dimaandrei/Project-AI/tree/master*](https://github.com/dimaandrei/Project-AI/tree/master)

# 8. Bibliografie

* <http://www.doctorat.tuiasi.ro/doc/SUSTINERI_TEZE/ETH/Ciobanu%20(Aionoae)/Ciobanu%20Aionoae%20Alexandra%20rezumat%20teza.pdf>
* <https://en.wikipedia.org/wiki/Bayesian_network>
* <http://florinleon.byethost24.com/Curs_IA/IA10_ReteleBayesiene.pdf?i=1>
* <https://koaha.org/wiki/Distribuzione_multivariata>