

Міністерство освіти і науки України
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Інститут прикладного системного аналізу
Кафедра математичних методів системного аналізу

Лабораторна робота № 2

з дисципліни «Байєсівський аналіз даних в наукових дослідженнях»

Виконав аспірант 2 курсу

групи КН-31ф

Кузнєцов О.А.

Перевішив

д.т.н., доц. Терентьев О. М.

Київ – 2024

Мета роботи: Ознайомлення з програмою Netica. Побудова ймовірнісного висновку в мережах Байєса – обчислення значень ймовірностей станів вершин.

Завдання

1. **Ознайомтеся з інформацією** наведеною в файлах:

02_Info_01.doc (приділіть увагу прикладу обчислення значень ймовірностей станів вершин при наявності в мережі інстанційованих вершин);

03_Manual_Netica.doc (інструкція з експлуатації програмою Netics).

2. **Побудуйте мережу Байєса** із використанням комп'ютерної програми Netica по даним наведеним в 03_Manual_Netica.doc призначеної оцінки кредитоспроможності клієнта.

3. **Завантажте готові приклади** мереж Байєса:

Chestclinic – мережа для діагностування пацієнта (в літературі ця мережа нерідко зустрічається під назвою Asia);

Car diagnostic – мережа для визначення ймовірності того, що автомобіль заводиться.

4. Для мережі Chestclinic задайте наступні інстанційованні значення: Smoking = smoker, Visit to Asia = visit, та обчисліть значення ймовірностей вершин **X Ray Result** та **Dyspnea** за допомогою програми **Netica** та вручну.

5. Для здачі лабораторної роботи необхідно **оформити протокол**, який повинен включати:

– Структуру мережі Байєса призначеної оцінки кредитоспроможності клієнта, Chestclinic та Car diagnostic

– Для кожної мережі змодельовати 2-3 ситуації наявності інстанційованих значень вершин в мережі. Результати представити у вигляді таблиці.

- Результати обчислення вручну та із використанням програми для пункту 4 представити у вигляді таблиці:
- Висновки по роботі.

Хід роботи

Завдання 2

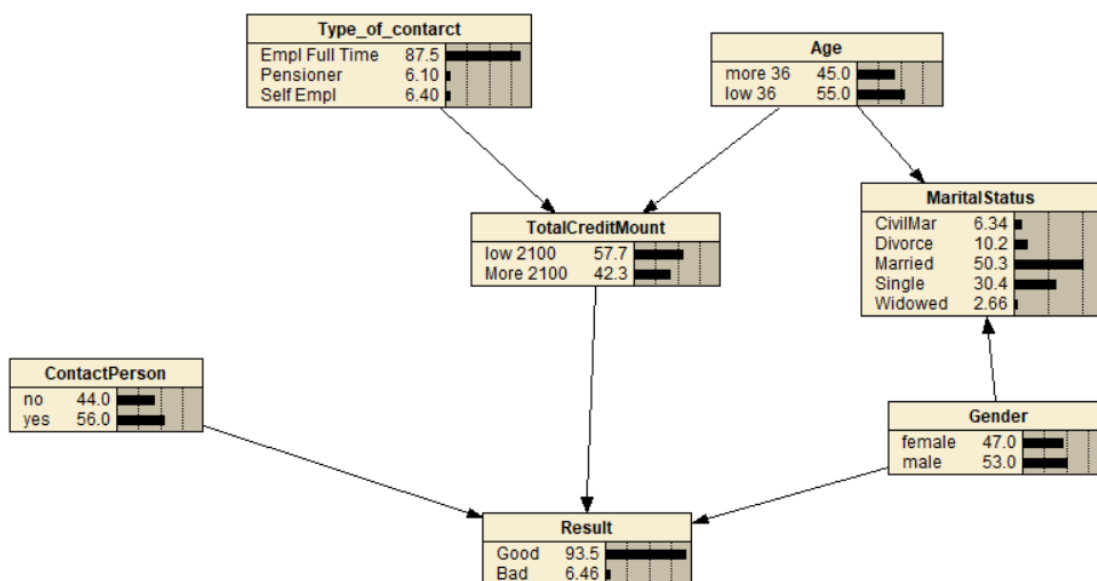


Рис. 1 – Структура мережі призначеної оцінки кредитоспроможності клієнта

Табл. 1 – Приклад роботи мережі призначеної оцінки кредитоспроможності клієнта

№	Інстанційовані вершини	Результат моделювання
1	TypeOfContract = Empl_Full_Time Age = low_36 Gender = male	$P(\text{Result} = \text{bad}) = 9,19\%$ $P(\text{Result} = \text{Good}) = 90,81\%$
2	ContactPerson = yes MaritalStatus = Widowed Gender = female	$P(\text{Result} = \text{bad}) = 7,79\%$ $P(\text{Result} = \text{Good}) = 92,21\%$

Завдання 3

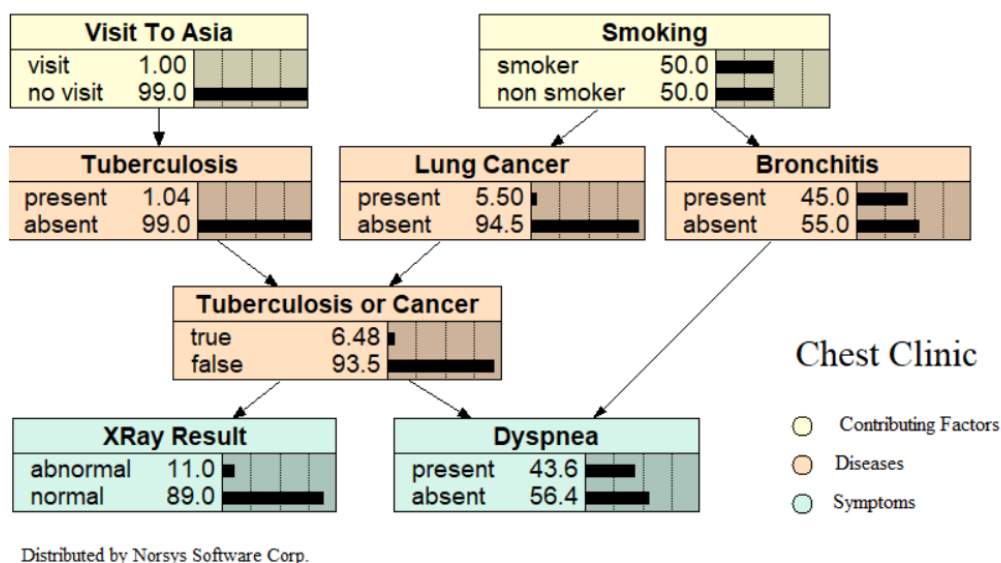


Рис. 2 – Структура мережі Chestclinic

Табл. 2 – Приклад роботи мережі Chestclinic

№	Інстанційовані вершини	Результат моделювання
1	Lung Cancer = present Bronchitis = present	$P(\text{XRay Result} = \text{abnormal}) = 98\%$ $P(\text{XRay Result} = \text{normal}) = 2\%$ $P(\text{Dyspnea} = \text{present}) = 90\%$ $P(\text{Dyspnea} = \text{absent}) = 10\%$
2	Lung Cancer = absent Bronchitis = absent	$P(\text{XRay Result} = \text{abnormal}) = 5.97\%$ $P(\text{XRay Result} = \text{normal}) = 94.03\%$ $P(\text{Dyspnea} = \text{present}) = 10.6\%$ $P(\text{Dyspnea} = \text{absent}) = 89.4\%$

Car Diagnosis 2

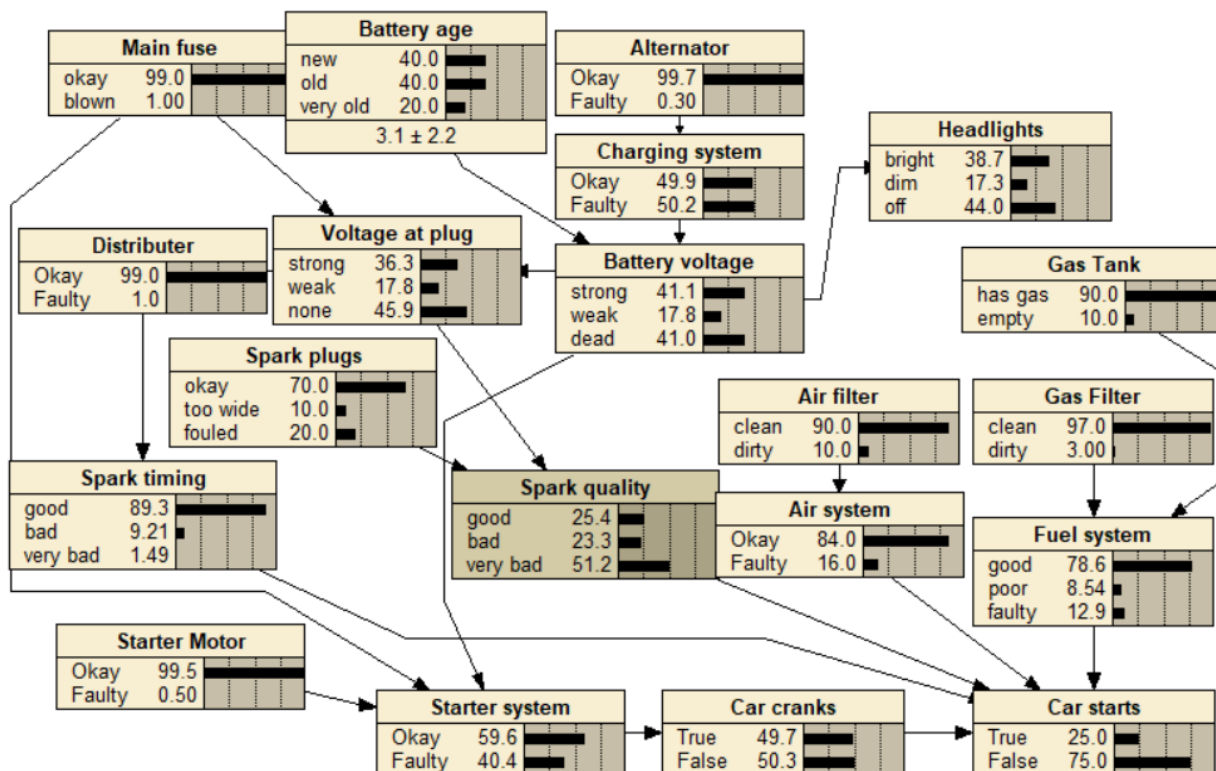


Рис. 3 – Структура мережі Car diagnostic

Табл. 3 – Приклад роботи мережі Car diagnostic

№	Інстанційовані вершини	Результат моделювання
1	Battery age = new Battery Voltage = strong Starter System = okay Car cranks = true	P(Car starts = true) = 64.9% P(Car starts = false) = 35.1%
2	Battery age = new Battery Voltage = strong Starter System = okay Car cranks = true Fuel system = poor	P(Car starts = true) = 22.5% P(Car starts = false) = 77.5%

Завдання 4

Наведемо розрахунки, які були зроблені вручну:

Якщо **Visit to Asia = visit**, то ймовірність **Tuberculosis = present** дорівнює 5%, таким чином:

$$P(\text{Tuberculosis}=\text{present} \mid \text{Visit to Asia}=\text{visit}) = 0.05$$

$$P(\text{Tuberculosis}=\text{absent} \mid \text{Visit to Asia}=\text{visit}) = 0.95$$

Розрахунок ймовірності для Lung Cancer

Якщо **Smoking = smoker**, то ймовірність **Lung Cancer = present** становить 10%, отже:

$$P(\text{Lung Cancer}=\text{present} \mid \text{Smoking}=\text{smoker}) = 0.10$$

$$P(\text{Lung Cancer}=\text{absent} \mid \text{Smoking}=\text{smoker}) = 0.90$$

Розрахунок ймовірності для Bronchitis

Якщо **Smoking = smoker**, то ймовірність **Bronchitis = present** становить 60%, тому:

$$P(\text{Bronchitis}=\text{present} \mid \text{Smoking}=\text{smoker}) = 0.60$$

$$P(\text{Bronchitis}=\text{absent} \mid \text{Smoking}=\text{smoker}) = 0.40$$

Розрахунок ймовірності для Tuberculosis or Cancer

Tuberculosis or Cancer буде істинним, якщо присутній хоча б один з діагнозів: **Tuberculosis** або **Lung Cancer**. Використовуємо формулу для об'єднання незалежних подій:

$$P(\text{Tuberculosis or Cancer}=\text{true}) = P(\text{Tuberculosis}=\text{present}) +$$

$$P(\text{Lung Cancer}=\text{present}) - P(\text{Tuberculosis}=\text{present}) * P(\text{Lung Cancer}=\text{present})$$

Підставимо значення:

$$P(\text{Tuberculosis or Cancer}=\text{true}) = 0.05 + 0.10 - (0.05 * 0.10) = 0.145$$

$$P(\text{Tuberculosis or Cancer}=\text{false}) = 1 - 0.145 = 0.855$$

Розрахунок ймовірності для XRay Result

Ймовірність аномального результату рентгену залежить від **Tuberculosis or Cancer**:

$$P(\text{XRay Result}=\text{abnormal} \mid \text{Tuberculosis or Cancer}=\text{true}) = 0.885$$

$$P(\text{XRay Result}=\text{abnormal} \mid \text{Tuberculosis or Cancer}=\text{false}) = 0.04$$

Тоді повна ймовірність для **XRay Result** (використовуючи теорему повної ймовірності) буде:

$$\begin{aligned} P(\text{XRay Result}=\text{abnormal}) &= P(\text{XRay Result}=\text{abnormal} \mid \\ &\text{Tuberculosis or Cancer}=\text{true}) * P(\text{Tuberculosis or Cancer}=\text{true}) + \\ &P(\text{XRay Result}=\text{abnormal} \mid \text{Tuberculosis or Cancer}=\text{false}) * \\ &P(\text{Tuberculosis or Cancer}=\text{false}) \end{aligned}$$

Підставимо значення:

$$P(\text{XRay Result}=\text{abnormal}) = (0.885 * 0.145) + (0.04 * 0.855) = 0.128325 + 0.0342 = 0.162525$$

Отже:

$$P(\text{XRay Result}=\text{abnormal}) \approx 0.163$$

$$P(\text{XRay Result}=\text{normal}) = 1 - 0.163 = 0.837$$

Розрахунок ймовірності для Dyspnea

Ймовірність **Dyspnea** залежить від **Tuberculosis or Cancer** та **Bronchitis**. Використовуємо закон повної ймовірності, розглядаючи всі комбінації:

$$\begin{aligned}
P(\text{Dyspnea}=\text{present}) &= P(\text{Dyspnea}=\text{present} \mid \text{Tuberculosis or Cancer}=\text{true}, \\
&\text{Bronchitis}=\text{present}) * P(\text{Tuberculosis or Cancer}=\text{true}) * P(\text{Bronchitis}=\text{present}) + \\
&P(\text{Dyspnea}=\text{present} \mid \text{Tuberculosis or Cancer}=\text{true}, \text{Bronchitis}=\text{absent}) * \\
&P(\text{Tuberculosis or Cancer}=\text{true}) * P(\text{Bronchitis}=\text{absent}) + P(\text{Dyspnea}=\text{present} \mid \\
&\text{Tuberculosis or Cancer}=\text{false}, \text{Bronchitis}=\text{present}) * P(\text{Tuberculosis or Cancer}=\text{false}) \\
&* P(\text{Bronchitis}=\text{present}) + P(\text{Dyspnea}=\text{present} \mid \text{Tuberculosis or Cancer}=\text{false}, \\
&\text{Bronchitis}=\text{absent}) * P(\text{Tuberculosis or Cancer}=\text{false}) * P(\text{Bronchitis}=\text{absent})
\end{aligned}$$

Підставимо значення для кожного випадку:

1. Tuberculosis or Cancer = true, Bronchitis = present:

$$P(\text{Dyspnea}=\text{present})=0.9$$

2. Tuberculosis or Cancer = true, Bronchitis = absent:

$$P(\text{Dyspnea}=\text{present})=0.7$$

3. Tuberculosis or Cancer = false, Bronchitis = present:

$$P(\text{Dyspnea}=\text{present})=0.8$$

4. Tuberculosis or Cancer = false, Bronchitis = absent:

$$P(\text{Dyspnea}=\text{present})=0.1$$

Обчислюємо:

$$\begin{aligned}
P(\text{Dyspnea}=\text{present}) &= (0.9 * 0.145 * 0.6) + (0.7 * 0.145 * 0.4) + (0.8 * 0.855 * 0.6) \\
&+ (0.1 * 0.855 * 0.4) = 0.0783 + 0.0406 + 0.4104 + 0.0342 = 0.5635
\end{aligned}$$

Отже:

$$P(\text{Dyspnea}=\text{present}) \approx 0.564$$

$$P(\text{Dyspnea}=\text{absent}) = 1 - 0.564 = 0.436$$

Наведемо розрахунки, які були зроблені в програмі Netica:

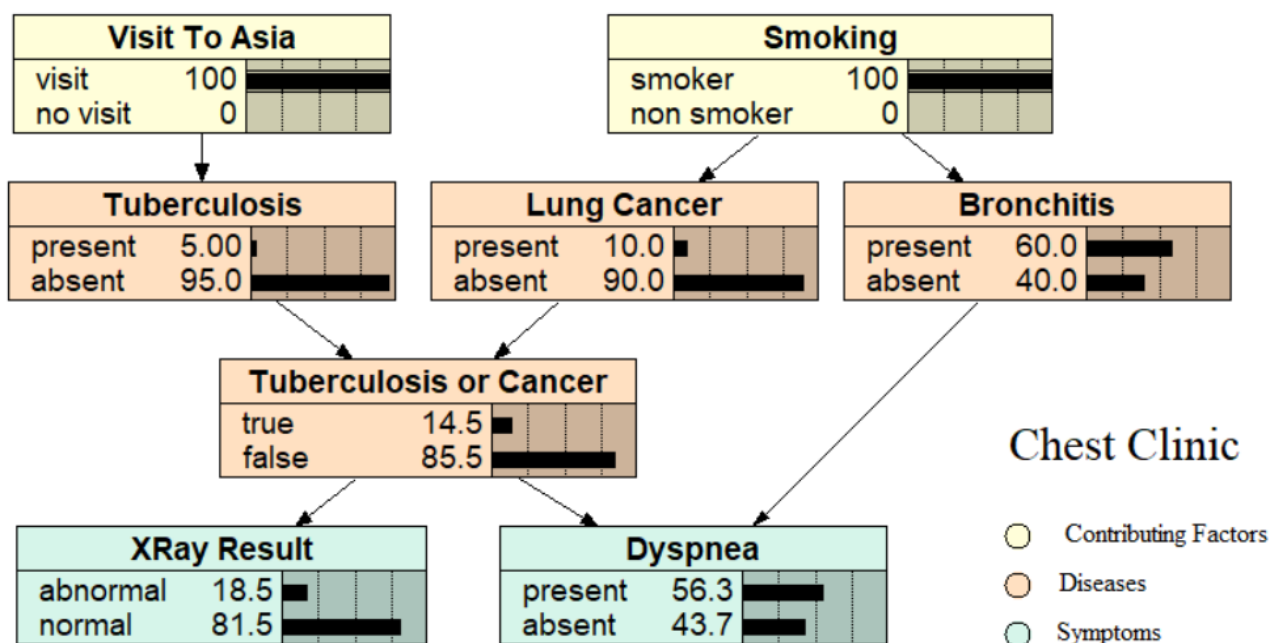


Рис. 4 – Результати обчислень у програмі Netica

Зведемо все у загальну таблицю:

Табл. 4 – Порівняльна таблиця отриманих результатів

Вершина	Результати Netica	Результати вручну
X Ray Result = abnormal	0.185	0.163
X Ray Result = normal	0.815	0.837
Dyspnea = present	0.563	0.564
Dyspnea = absent	0.437	0.436

Висновки

У ході виконання лабораторної роботи було досліджено принципи роботи мереж Байєса та проведено розрахунки умовних ймовірностей для різних вузлів мережі в умовах заданих факторів вручну, а також за допомогою програми Netica. На основі виконаної роботи можна сказати, що інстанційована вершина

мережі Байєса – це вузол, якому було присвоєно конкретне значення або стан. Інстанціювання дозволяє використовувати наявну інформацію для подальшого обчислення умовних ймовірностей інших змінних у мережі. Зокрема, у цій роботі було інстанційовано вершини, що відповідають за куріння та відвідування Азії, щоб дослідити їхній вплив на ймовірності інших подій, таких як захворювання і симптоми.

Порівняння результатів ручного розрахунку з результатами, отриманими у програмі Netica, показало, що результати можуть відрізнятися. Це пояснюється впливом округлення, ймовірними похибками в ручних обчисленнях та відмінностями в методології розрахунків. Програма Netica використовує високоточні алгоритми для обчислення ймовірностей, що забезпечує більшу точність і зменшує похибки порівняно з ручними розрахунками.

Ймовірнісний висновок у мережі Байєса – це процес обчислення умовних ймовірностей для деяких змінних, виходячи з відомих значень інших змінних. Наприклад, у рамках цієї роботи, знаючи, що пацієнт є курцем і відвідував Азію, ми виконали ймовірнісний висновок, щоб обчислити ймовірність виникнення задишки та аномального результату рентгену. Таким чином, ймовірнісний висновок дозволяє робити прогнози та діагностувати ймовірність подій на основі наявної інформації, що є корисним інструментом у медичних діагностичних системах та інших сферах.

Робота з мережею Байєса дозволила краще зрозуміти механізми ймовірнісного моделювання, а також важливість точності розрахунків і програмних інструментів для складних обчислень в аналізі умовних ймовірностей.