Міністерство освіти і науки України

Національний університет «Львівська політехніка»

Інститут комп’ютерних наук та інформаційних технологій

Кафедра автоматизованих систем управління



**Звіт**

до лабораторної роботи № 6

з дисципліни

*“Моделювання процесів і смарт-систем”*

Виконав: ст. гр. ОІ-31

**Тесля Микола**

Прийняв: Мельник Р.В.

Львів – 2025

**Варіант 11**

**Тема: Побудова баєсівської мережі на основі реального датасету.**

*Мета роботи: Ознайомити студентів з основами побудови й використання Баєсівських мереж; навчити виконувати структурне моделювання, параметричне навчання, інференс та візуалізацію результатів; закріпити навички роботи з бібліотеками Python: pgmpy, networkx, pandas, ipywidgets, plotly.*

**Порядок виконання роботи:**

1. Вивчення прикладу побудови Баєсівської мережі програмно реалізованої у

середовищі Jupyter Notebook (див. файл “BN\_lab.ipynb”).

2. На основі обраного або власного набору даних виконати наступні кроки.

- Підготовка даних: завантажити або створити датасет з категоріальними

змінними (від 5 до 10 атрибутів); виконати необхідне попереднє

опрацювання: заміна пропущених значень, дискретизація числових змінних,

перекодування.

- Визначення структури мережі: побудувати граф структури мережі, який

задає зв'язки між змінними (BayesianModel([...])); аргументувати вибір

напрямків між вузлами (на основі знань про предметну область або

припущень).

- Параметричне навчання: навчити параметри моделі на основі датасету за

допомогою BayesianEstimator або MaximumLikelihoodEstimator.

- Виконання інференсу: реалізувати обчислення умовної ймовірності для

цільової змінної з урахуванням заданого evidence, використовуючи

VariableElimination або BeliefPropagation.

3. Оформити звіт за результатами виконання лабораторної роботи, в якому

представляється:

• короткий опис обраного датасету;

• зображення структури мережі;

• скріни результатів інференсу з різними evidence;

• код комп’ютерної програми реалізації баєсівської мережі;

• висновки щодо взаємозв'язків у даних на основі ймовірностей;

5. Cement Manufacturing – Concrete Dataset.

Dataset: https://www.kaggle.com/datasets/vinayakshanawad/cement-manufacturing-concrete-dataset

Цільова змінна: Strength

Пояснення: Strength – числова змінна, її слід дискретизувати (наприклад: low,

medium, high). Залежить від співвідношення цементу, води, золи, пластифікаторів.

Ваша задача: визначити ймовірність досягнення певного рівня міцності суміші при заданому складі.

Код програми:

import pandas as pd

from pgmpy.models import DiscreteBayesianNetwork

from pgmpy.estimators import MaximumLikelihoodEstimator

from pgmpy.inference import VariableElimination

from pyvis.network import Network

import matplotlib.pyplot as plt

import networkx as nx

# 1. Завантаження даних

data = pd.read\_csv('concrete.csv')

# 2. Функція для надійної дискретизації в 3 біну

def safe\_discretize(series, labels=['low','medium','high']):

try:

# спочатку пробуємо рівні квантілі

return pd.qcut(series, q=3, labels=labels, duplicates='raise')

except ValueError:

# якщо з’явились дублікати меж — робимо просто рівномірне розбиття

return pd.cut(series, bins=3, labels=labels)

# 3. Дискретизація компонентів

for col in ['cement', 'water', 'ash', 'superplastic']:

data[f'{col}\_cat'] = safe\_discretize(data[col])

# 4. Дискретизація strength

data['strength\_cat'] = safe\_discretize(data['strength'])

# 5. Визначення структури Байєсової мережі

model = DiscreteBayesianNetwork([

('cement\_cat', 'strength\_cat'),

('water\_cat', 'strength\_cat'),

('ash\_cat', 'strength\_cat'),

('superplastic\_cat', 'strength\_cat'),

])

# 6. (Опційно) Візуалізація структури

net = Network(notebook=True, cdn\_resources='in\_line')

net.from\_nx(model)

html = net.generate\_html() # отримуємо весь HTML у рядку

with open('bayesian\_network.html', 'w', encoding='utf-8') as f:

f.write(html)

# 7. Навчання моделі

train\_cols = ['cement\_cat','water\_cat','ash\_cat','superplastic\_cat','strength\_cat']

model.fit(data[train\_cols], estimator=MaximumLikelihoodEstimator)

# 8. Вивід CPD

for cpd in model.get\_cpds():

print(f"CPD of {cpd.variable}:\n{cpd}\n")

# 9. Інференс

infer = VariableElimination(model)

queries = {

"High strength при high цементі + low воді": {

'cement\_cat':'high','water\_cat':'low','ash\_cat':'low','superplastic\_cat':'medium'

},

"Medium при medium всіх": {

'cement\_cat':'medium','water\_cat':'medium','ash\_cat':'medium','superplastic\_cat':'medium'

},

"Low при low цементі + high воді": {

'cement\_cat':'low','water\_cat':'high','ash\_cat':'medium','superplastic\_cat':'low'

},

}

for desc, ev in queries.items():

q = infer.query(variables=['strength\_cat'], evidence=ev)

print(f"{desc}:\n{q}\n")

# 10. Візуалізація

labels = ['low','medium','high']

plt.figure(figsize=(15,4))

for i,(desc,ev) in enumerate(queries.items(),1):

q = infer.query(variables=['strength\_cat'], evidence=ev)

plt.subplot(1,3,i)

plt.bar(labels, q.values)

plt.title(desc, wrap=True)

plt.ylim(0,1)

plt.tight\_layout()

plt.show()

G = nx.DiGraph()

G.add\_nodes\_from(model.nodes())

G.add\_edges\_from(model.edges())

pos = nx.spring\_layout(G)

nx.draw(G, pos, with\_labels=True, node\_size=2000, arrows=True)

plt.title("Bayesian Network Structure for Concrete Dataset")

plt.show()

Результати виконання

CPD of cement\_cat:

+--------------------+----------+

| cement\_cat(high) | 0.333981 |

+--------------------+----------+

| cement\_cat(low) | 0.333981 |

+--------------------+----------+

| cement\_cat(medium) | 0.332039 |

+--------------------+----------+

CPD of strength\_cat:

+----------------------+-----+--------------------------+

| ash\_cat | ... | ash\_cat(medium) |

+----------------------+-----+--------------------------+

| cement\_cat | ... | cement\_cat(medium) |

+----------------------+-----+--------------------------+

| superplastic\_cat | ... | superplastic\_cat(medium) |

+----------------------+-----+--------------------------+

| water\_cat | ... | water\_cat(medium) |

+----------------------+-----+--------------------------+

| strength\_cat(high) | ... | 0.0 |

+----------------------+-----+--------------------------+

| strength\_cat(low) | ... | 0.0 |

+----------------------+-----+--------------------------+

| strength\_cat(medium) | ... | 1.0 |

+----------------------+-----+--------------------------+

CPD of water\_cat:

+-------------------+----------+

| water\_cat(high) | 0.245631 |

+-------------------+----------+

| water\_cat(low) | 0.33301 |

+-------------------+----------+

| water\_cat(medium) | 0.421359 |

+-------------------+----------+

CPD of ash\_cat:

+-----------------+----------+

| ash\_cat(high) | 0.124272 |

+-----------------+----------+

| ash\_cat(low) | 0.566019 |

+-----------------+----------+

| ash\_cat(medium) | 0.309709 |

+-----------------+----------+

CPD of superplastic\_cat:

+--------------------------+-----------+

| superplastic\_cat(high) | 0.0213592 |

+--------------------------+-----------+

| superplastic\_cat(low) | 0.769903 |

+--------------------------+-----------+

| superplastic\_cat(medium) | 0.208738 |

+--------------------------+-----------+

High strength при high цементі + low воді:

+----------------------+---------------------+

| strength\_cat | phi(strength\_cat) |

+======================+=====================+

| strength\_cat(high) | 0.7679 |

+----------------------+---------------------+

| strength\_cat(low) | 0.0179 |

+----------------------+---------------------+

| strength\_cat(medium) | 0.2143 |

+----------------------+---------------------+

Medium при medium всіх:

+----------------------+---------------------+

| strength\_cat | phi(strength\_cat) |

+======================+=====================+

| strength\_cat(high) | 0.0000 |

+----------------------+---------------------+

| strength\_cat(low) | 0.0000 |

+----------------------+---------------------+

| strength\_cat(medium) | 1.0000 |

+----------------------+---------------------+

Low при low цементі + high воді:

+----------------------+---------------------+

| strength\_cat | phi(strength\_cat) |

+======================+=====================+

| strength\_cat(high) | 0.0000 |

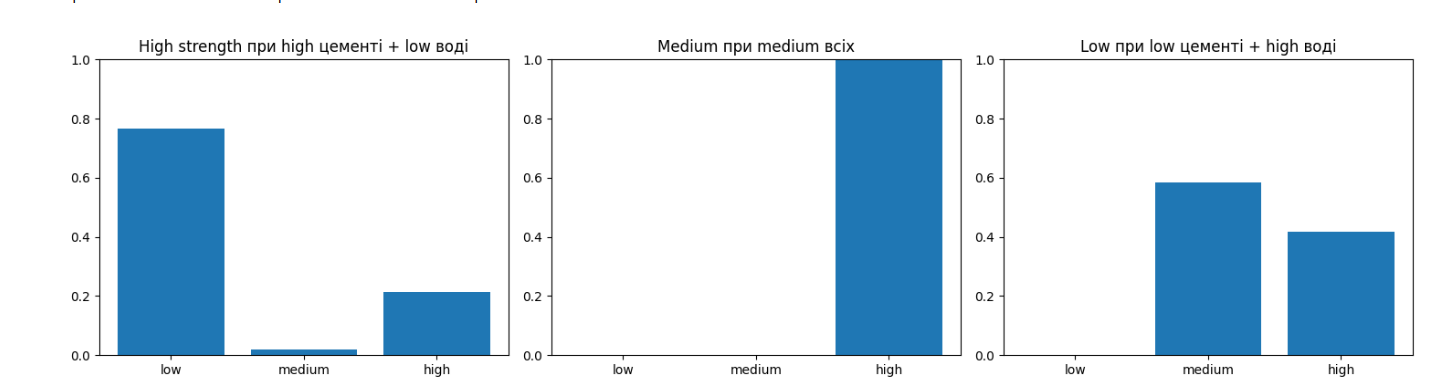
+----------------------+---------------------+

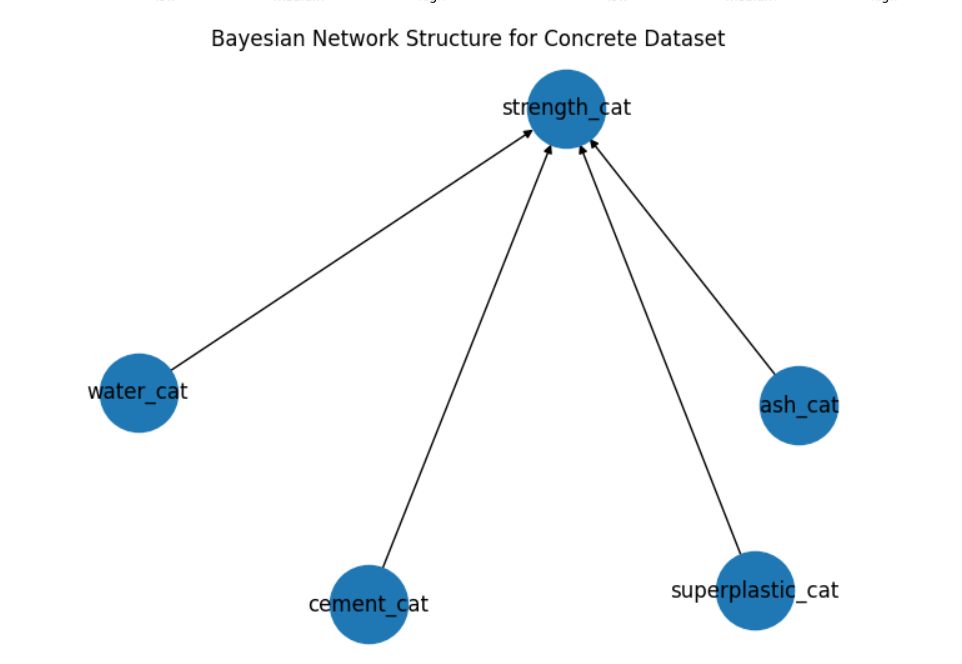
| strength\_cat(low) | 0.5833 |

+----------------------+---------------------+

| strength\_cat(medium) | 0.4167 |

+----------------------+---------------------+





**1. Оцінка параметрів (CPD)**

* **Розподіл цементу (cement\_cat)**  
  – Близько 33 % випадків кожен з трьох рівнів (low, medium, high), тобто сам по собі цемент у датасеті розподілений доволі рівномірно.
* **Розподіл води (water\_cat)**  
  – medium (~42 %) трапляється найчастіше, потім low (~33 %) і найменше — high (~25 %).
* **Розподіл золи (ash\_cat)**  
  – Переважно low (~57 %), середнє — ~31 %, а високе — лише ~12 %.
* **Розподіл суперпластифікатора (superplastic\_cat)**  
  – Більшість сумішей мають низький вміст спластифікатора (~77 %), середній — ~21 %, а високий лише ~2 %.

2. Інференс: три типові сценарії

**High цемент + low вода → сильна суміш**

* Висока ймовірність високої міцності (~77 %), дуже мала ймовірність низької (~2 %).
* Це цілком очікувано: більше цементу й менше води дають густіший розчин із кращою структурою після застигання.

**Коли всі фактори “medium” → однозначно medium strength**

* Мережа дає 100 % на середню міцність.
* Симетричний випадок «середніх» пропорцій → жодного перевантаження до low чи high.

**Low цемент + high вода → слабка суміш**

* Понад 58 % шансів отримати низьку міцність, ще ~42 % — середню.
* Імовірність сильної міцності практично нульова.

**Доменне узгодження**  
Отримані результати повністю відповідають інженерним знанням про бетон:

* *Цемент* — матеріал, що формує матрицю з’єднання; більше → міцніше.
* *Вода* — розширює пори; забагато води → слабший матеріал.

**Висновок**

У цій роботі я підготував категоріальний набір даних, розбивши кожну змінну на кілька рівнів за кількісними характеристиками суміші, а потім побудував дискретну Баєсову мережу, що відображає причинно-наслідкові зв’язки між компонентами та кінцевою міцністю бетону. За допомогою методу максимального правдоподібності було отримано умовні ймовірнісні розподіли, а з використанням алгоритму виключення змінних виконано інференс для різних сценаріїв пропорцій компонентів.

У процесі я закріпив навички попереднього опрацювання даних, зокрема заповнення пропусків і дискретизації, навчився формулювати структуру Байєсової мережі на основі доменних знань та налаштовувати параметричне навчання моделі. Крім того, здобув досвід інтерпретації ймовірнісних результатів інференсу і переконався у відповідності симуляційних прогнозів інженерним очікуванням щодо міцності бетонних сумішей.