

## ЛАБОРАТОРНА РОБОТА № 3

### ДОСЛІДЖЕННЯ МЕТОДІВ РЕГРЕСІЇ ТА НЕКОНТРОЛЬОВАНОГО НАВЧАННЯ

**Мета:** використовуючи спеціалізовані бібліотеки і мову програмування Python дослідити методи регресії та неконтрольованої класифікації даних у машинному навчанні.

**Хід роботи:**

**Завдання 10**

**Лістинг програми:**

```
import datetime
import json
import numpy as np
import matplotlib.pyplot as plt
from sklearn.covariance import GraphicalLassoCV
from sklearn import cluster
import yfinance as yf

# Вхідний файл із символічними позначеннями компаній
input_file = 'company_symbol_mapping.json'

# Завантаження прив'язок символів компаній до їх повних назв
with open(input_file, 'r') as f:
    company_symbols_map = json.loads(f.read())
symbols, names = np.array(list(company_symbols_map.items())).T

# Вказані дати
start_date = datetime.datetime(2003, 7, 3)
end_date = datetime.datetime(2007, 5, 4)

# Вибір котирувань
quotes = [yf.download(symbol, start=start_date, end=end_date) for symbol in
symbols]
```

					ДУ «Житомирська політехніка».22.122.4.000 – Лр3			
Змн.	Арк.	№ докум.	Підпис	Дата				
Розроб.		Дяченко В.В.			Звіт з лабораторної роботи		Лім.	Арк.
Перевір.								1
Керівник							ФІКТ Гр. КН-20-1(1)	
Н. контр.								
Зав. каф.								

```

# Вилучення котирувань, що відповідають відкриттю та закриттю біржі
opening_quotes = np.array([quote['Open'] for quote in quotes]).astype(float)
closing_quotes = np.array([quote['Close'] for quote in quotes]).astype(float)

# Обчислення різниці між двома видами котирувань
quotes_diff = closing_quotes - opening_quotes

# Нормалізація даних
X = quotes_diff.copy().T
X /= X.std(axis=0)

# Створення моделі графа
edge_model = GraphicalLassoCV()

# Навчання моделі
with np.errstate(invalid="ignore"):
    edge_model.fit(X)

# Створення моделі кластеризації на основі поширення подібності,
використовуючи щойно навчену крайову модель
_, labels = cluster.affinity_propagation(edge_model.covariance_)
num_labels = labels.max()

# Виведення результатів
for i in range(num_labels + 1):
    print("Cluster", i + 1, ', ', ', ', '.join(names[labels == i]))

```

```

[*****] 1 of 1 completed
[*****] 1 of 1 completed
[*****] 1 of 1 completed
[*****] 1 of 1 completed
Cluster 1 : Exxon, Chevron, ConocoPhillips, Valero Energy
Cluster 2 : Toyota, Ford, Honda, Boeing, Mc Donalds, Apple, SAP, Caterpillar
Cluster 3 : Kraft Foods
Cluster 4 : Coca Cola, Pepsi, Kellogg, Procter Gamble, Colgate-Palmolive, Kimberly-Clark
Cluster 5 : Time Warner, Comcast, Marriott, Wells Fargo, JPMorgan Chase, AIG, American express, Bank of America, Goldman Sachs, Xerox, Wal-Mart, Home Depot, Hyder, DuPont de Nemours
Cluster 6 : Microsoft, IBM, HP, Amazon, 3M, General Electric, Cisco, Texas Instruments
Cluster 7 : Northrop Grumman, Lockheed Martin, General Dynamics
Cluster 8 : Walgreen, CVS
Cluster 9 : GlaxoSmithKline, Pfizer, Sanofi-Aventis, Novartis

```

Рис.1. Результат роботи програми.

		Дяченко В.В.			ДУ «Житомирська політехніка».22.122.4.000 – Лр3	Арк.
						2
Змн.	Арк.	№ докум.	Підпис	Дата		