

**COLLECTION OF RESEARCH PAPERS**

of the 8th International Research and Practical Conference

**CHEMICAL TECHNOLOGY:  
SCIENCE, ECONOMY AND PRODUCTION**

**ЗБІРНИК НАУКОВИХ ПРАЦЬ**

VIII Міжнародної науково-практичної конференції

**ХІМІЧНА ТЕХНОЛОГІЯ:  
НАУКА, ЕКОНОМІКА ТА ВИРОБНИЦТВО**



МІНІСТЕРСТВО  
ОСВІТИ І НАУКИ  
УКРАЇНИ



Фармак



ISSN 2786-4898

Міністерство освіти і науки України  
Сумський державний університет  
Шосткинський інститут Сумського державного університету  
Центральний науково-дослідний інститут  
озброєння та військової техніки збройних сил України  
Публічне акціонерне товариство «Фармак»  
Управління освіти Шосткинської міської ради  
Виконавчий комітет Шосткинської міської ради

## COLLECTION OF RESEARCH PAPERS

of the 8th International Research and Practical Conference

### CHEMICAL TECHNOLOGY: SCIENCE, ECONOMY AND PRODUCTION



## ЗБІРНИК НАУКОВИХ ПРАЦЬ

VIII Міжнародної науково-практичної конференції  
**ХІМІЧНА ТЕХНОЛОГІЯ:  
НАУКА, ЕКОНОМІКА ТА ВИРОБНИЦТВО**

(м. Шостка, 27-29 листопада 2024 року)



Суми

Сумський Державний Університет

2024

УДК 66.01

Редакційна колегія:

Головний редактор Закусило Р.В., доцент кафедри хімічної технології високомолекулярних сполук, к.т.н., доцент.

Заступник головного редактора Павленко О.В., завідувач кафедри хімічної технології високомолекулярних сполук, к.т.н.

Відповідальний секретар Скуба Ю.Г. фахівець кафедри економіки та управління Шосткинського інституту Сумського державного університету.

Члени редакційної колегії:

Кравець В.Г. – професор кафедри хімічної технології високомолекулярних сполук, д.т.н., професор;

Худолей Г.М. – завідувач кафедри системотехніки і інформаційних технологій, к.т.н;

Тур О.М. – доцент кафедри економіки та управління, к.е.н.;

Тимофіїв С.В. – ст. викладач кафедри хімічної технології високомолекулярних сполук, к.х.н.;

Пригара І.О. – ст. викладач кафедри економіки та управління, к.е.н.

Збірник наукових праць VIII Міжнародної науково-практичної конференції «Хімічна технологія: наука, економіка та виробництво», м. Шостка, 27 - 29 листопада 2024 року. – Суми : Сумський державний університет, 2024. – 242 с.

ISSN 2786-4898.

Збірник містить наукові праці учасників VIII Міжнародної науково-практичної конференції «Хімічна технологія: наука, економіка та виробництво», що складаються з узагальнених матеріалів науково-дослідних робіт науковців різних галузей виробництва та наукових закладів України.

У збірнику висвітлюються актуальні питання спеціальної хімічної технології і виробництва боєприпасів, утилізації відходів виробництва різних галузей, енергозбереження, моделювання технологічних процесів, соціально-економічні аспекти виробництва та природокористування в умовах війни.

Збірник корисний робітникам хімічної промисловості, науковим співробітникам, аспірантам і студентам спеціальностей хіміко-технологічного та соціально-економічного профілів, фахівцям інформаційних технологій виробництва.

Наукові праці учасників конференції подаються в авторській редакції.

© Шосткинський інститут  
Сумського державного університету, 2024  
© Сумський державний університет, 2024

## АКТУАЛЬНІСТЬ ВИКОРИСТАННЯ BIG DATA В ХІМІЧНІЙ ТЕХНОЛОГІЇ

А.Л. Баланюк

Шосткинський інститут Сумського державного університету  
[b.n.y@ukr.net](mailto:b.n.y@ukr.net)

**Вступ.** Актуальність використання Big Data в хімічній технології.

Використання Big Data в хімічній технології є надзвичайно актуальним, оскільки дозволяє обробляти великі обсяги даних, що генеруються під час досліджень, виробничих процесів та експериментів. Це сприяє точнішому прогнозуванню поведінки хімічних систем, оптимізації технологічних процесів та створенню нових матеріалів. Інтеграція Big Data у хімічну галузь відкриває нові можливості для підвищення ефективності, зменшення витрат та зниження екологічного навантаження.

Проблеми, які вирішуються за допомогою аналізу великих даних.

Аналіз великих даних дозволяє вирішувати ряд ключових проблем у різних сферах, зокрема в хімічній технології:

- Оптимізація виробничих процесів: Завдяки обробці великих обсягів даних можна вдосконалити технологічні параметри, зменшити енергоспоживання, підвищити вихід продукції та знизити витрати на сировину.
- Прогнозування поведінки хімічних систем: Використання алгоритмів для моделювання та прогнозування дозволяє передбачати результати експериментів та реагувати на зміни умов у реальному часі.
- Інноваційні розробки та створення нових матеріалів: Аналіз великих даних допомагає ідентифікувати закономірності, які можуть бути використані для розробки нових матеріалів з покращеними властивостями.
- Зменшення екологічного впливу: Аналіз даних дозволяє виявляти можливості для зменшення відходів, викидів та підвищення ефективності використання ресурсів у процесах.
- Управління ризиками: Аналіз великих даних сприяє точнішому виявленню потенційних ризиків у виробництві, що дозволяє оперативно вживати заходів для їх мінімізації.

Мета статті – дослідити методи аналізу Big Data та їх використання для класифікації бетону.

### Теоретичний огляд

Що таке Big Data: основні характеристики та технології збору даних.

Big Data – це великі обсяги структурованих та неструктурованих даних, що надходять з різноманітних джерел. Основні характеристики: обсяг (Volume), швидкість (Velocity), різноманітність (Variety), достовірність (Veracity) і цінність (Value). Технології збору даних включають використання сенсорів, Інтернету речей (IoT), соціальних мереж, а також автоматизованих систем моніторингу та збору даних у реальному часі.

Сучасні інструменти та методи аналізу даних включають [1]:

1. Машинне навчання (ML): Використовується для побудови моделей, що прогнозують чи класифікують дані. Алгоритми, як-от регресія, дерева рішень, нейронні мережі та кластеризація, допомагають у виявленні патернів і тенденцій у великих даних.

2. Хмарні платформи: Платформи, як-от AWS, Google Cloud, Microsoft Azure, надають інфраструктуру для зберігання та обробки великих обсягів даних, дозволяючи гнучко масштабувати ресурси та використовувати потужні обчислювальні можливості для аналізу.

3. Python-бібліотеки:

- Pandas – для обробки і маніпулювання даними.
- NumPy – для числових обчислень.
- SciPy – для наукових і технічних обчислень.
- Scikit-learn – для машинного навчання та моделювання.
- TensorFlow та Keras – для глибокого навчання.
- Matplotlib та Seaborn – для візуалізації даних.
- Dask – для обробки великих даних, що не вміщуються в оперативну пам'ять.

Ці інструменти дають можливість ефективно аналізувати, обробляти і візуалізувати великі обсяги даних з високою точністю.

Роль Big Data в хімічній технології: приклади застосувань (прогнозування реакцій, оптимізація складів, зменшення витрат).

Роль Big Data в хімічній технології полягає в покращенні точності та ефективності процесів. Приклади застосувань:

1. Прогнозування реакцій: Аналіз великих обсягів даних дозволяє передбачити поведінку хімічних реакцій і оптимізувати умови для досягнення бажаного результату [2].

2. Оптимізація складів: Використання Big Data для аналізу складів матеріалів допомагає розробляти нові рецептури з покращеними властивостями та зменшувати витрати на сировину.

3. Зменшення витрат: Обробка даних з виробничих процесів дозволяє знизити енергоспоживання, скоротити час виробництва і зменшити кількість відходів.

**Методологія.** Опис процесу аналізу великих даних. Процес аналізу великих даних включає:

1. Збір даних: Використання сенсорів, лабораторних експериментів і відкритих джерел для збору даних.

2. Попередня обробка: Очищення та нормалізація даних для підготовки до аналізу.

3. Аналіз і моделювання: Використання методів, як-от регресія, кластеризація та нейронні мережі, для виявлення патернів та створення моделей.

Використані інструменти: програмне забезпечення, алгоритми, платформи.

Використані інструменти включають [3]:

1. Програмне забезпечення: Python, R для аналізу даних та моделювання.

2. Алгоритми: Регресія, кластеризація, нейронні мережі, дерева рішень.

3. Платформи: Хмарні сервіси (AWS, Google Cloud, Azure), бібліотеки (Pandas, Scikit-learn, TensorFlow).

### **Практична частина**

Аналіз набору даних для прогнозування результату хімічного процесу.

Ми будемо прогнозувати, до якої категорії відноситься бетон за допомогою гіперпараметра "Value". В даному випадку, "Value" буде використовуватись як одна з ознак для моделювання, з метою класифікації зразків бетону на різні категорії. Це може включати визначення міцності, складу чи інших властивостей бетону, що дозволить автоматично відносити кожен зразок до певної категорії на основі його

числових характеристик. Для цього буде застосовано машинне навчання, зокрема методи класифікації, такі як Random Forest, для навчання моделі прогнозування.

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score

# Завантаження даних
df = pd.read_csv(r"C:\Users\User\Desktop\CAX_Test_Quality\CAX_Train_Quality.csv")

# Перетворення Timestamp_Shifted на datetime
df['Timestamp_Shifted'] = pd.to_datetime(df['Timestamp_Shifted'])

# Створення ознак з datetime (за потреби)
df['hour'] = df['Timestamp_Shifted'].dt.hour
df['dayofweek'] = df['Timestamp_Shifted'].dt.dayofweek

# Кодування параметра
encoder = LabelEncoder()
df['Parameter_encoded'] = encoder.fit_transform(df['Parameter'])

# Вибір ознак та цільової змінної
X = df[['Value', 'hour', 'dayofweek']] # Можна додавати більше ознак
y = df['Parameter_encoded']

# Розбиття на тренувальний та тестовий набори
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Навчання моделі
model = RandomForestClassifier()
```

Рисунок 1 – Приклад коду

```
model.fit(X_train, y_train)

# Прогнозування на тестових даних
y_pred = model.predict(X_test)

# Оцінка моделі
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy: {accuracy:.2f}')
```

✓ 40s Python

Рисунок 2. – Приклад коду

	Actual	Predicted
0	Quality 8	Quality 8
1	Quality 10	Quality 10
2	Quality 6	Quality 6
3	Output Parameter	Quality 10
4	Quality 9	Quality 9

Рисунок 3 – Реальне і прогнозоване значення

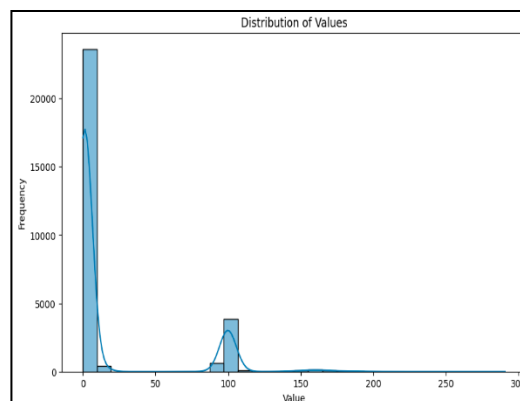


Рисунок 4 – Діаграма частоти значень

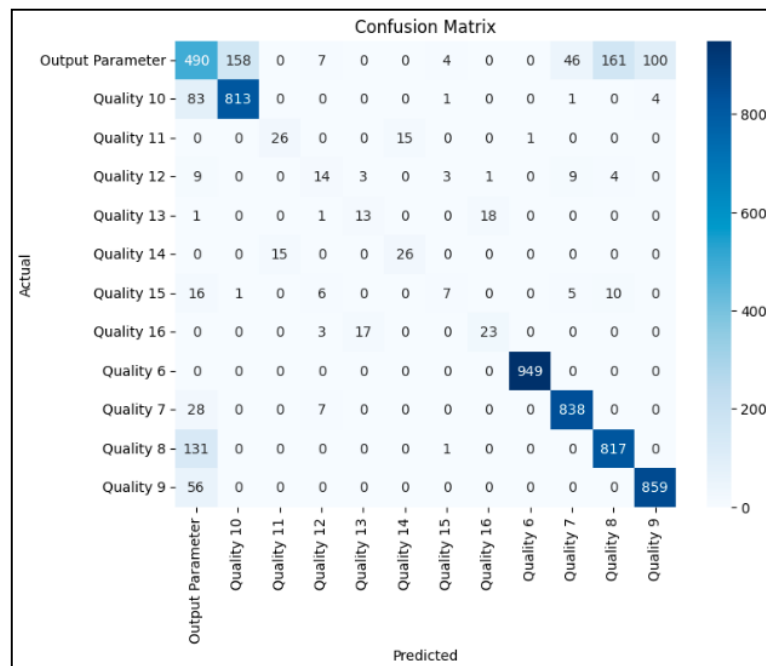


Рисунок 5 – Матриця частоти точності прогнозування

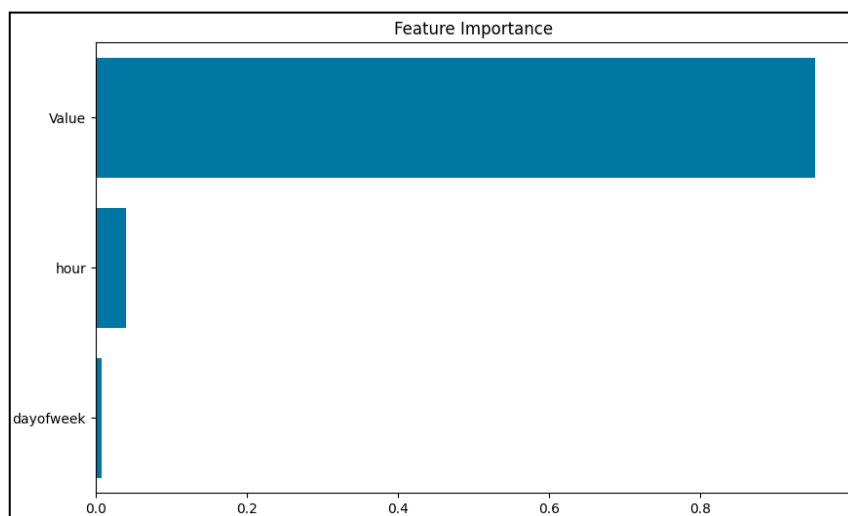


Рисунок 6 – Важливість гіпер параметрів

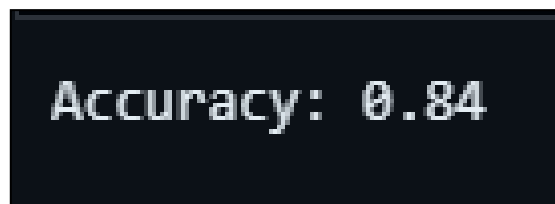


Рисунок 7 – Точність прогнозування 84%

### Обговорення результатів

1.Прогнозування за допомогою Random Forest: У вашому коді для прогнозування параметра використовується Random Forest Classifier, який є методом ансамблевого навчання, що базується на побудові кількох дерев рішень. Ця модель

комбінує прогнози з усіх дерев, що дозволяє зменшити варіативність та підвищити точність прогнозів. Вона є популярною завдяки своїй здатності обробляти великі та складні набори даних, а також не схильна до перенавчання.

## 2. Попередня обробка даних:

- перетворення `Timestamp_Shifted` на `datetime`: Спочатку перетворюється колонка з часом на формат `datetime`, щоб можна було виділити додаткові ознаки, такі як година та день тижня. Це дозволяє моделі враховувати часові фактори при прийнятті рішень;

- кодування параметра: Оскільки колонка `Parameter` містить категоріальні дані, вони кодуються за допомогою `LabelEncoder`, що дозволяє перетворити їх на числові мітки, з якими працює модель [4];

- навчання та оцінка моделі: Модель тренується на 80% даних (використовуючи `train_test_split`), а на 20% даних тестується. Ваша модель досягла 84% точності на тестовому наборі, що є високим результатом для задач класифікації.

## Пояснення таблиць та діаграм

### 1. Таблиця результатів (Actual vs Predicted):

- таблиця показує порівняння між реальними значеннями (Actual) і прогнозованими значеннями (Predicted) для кожного з тестових прикладів.

- це допомагає зрозуміти, які саме приклади були правильно класифіковані, а де модель помилилась. Такий аналіз дозволяє виявити можливі слабкі місця в прогнозах.

### 2. Гістограма для розподілу значень Value [5]:

- гістограма допомагає побачити, як розподіляються значення Value. Це дає зрозуміти, чи є які-небудь домінуючі значення або тренди, що можуть впливати на прогнозування.

- використання кривої ядерної оцінки щільності (KDE) на графіку дозволяє побачити розподіл більш плавно.

### 3. Матриця плутанини (Confusion Matrix):

- ця матриця допомагає зрозуміти, скільки правильних і неправильних прогнозів було зроблено для кожного класу.

- анотації на тепловій карті дозволяють оцінити, в яких категоріях модель мала найбільше помилок (наприклад, помилкові позитиви чи негативи).

### 4. Важливість ознак (Feature Importance):

- графік важливості ознак показує, які ознаки (наприклад, `Value`, `hour`, `dayofweek`) найбільше впливають на прийняття рішень моделлю.

- це важливо для подальшої оптимізації моделі, бо можна побачити, які з ознак є найбільш важливими для точності прогнозу.

Висока точність (84%). 84% точності є високим результатом для даної моделі класифікації, що показує її здатність правильно передбачати значення параметра. Це свідчить про те, що модель добре навчається на даних, що мають складну структуру, і може ефективно прогнозувати категорії, навіть коли дані містять шум або неповні інформації.

Такий рівень точності є досить хорошим для багатьох застосувань, і в цьому випадку ми можемо зробити висновок, що модель адекватно справляється з прогнозуванням параметрів хімічного процесу.



## **Висновки**

Ключові висновки та рекомендації

1. Висока точність моделі: Модель Random Forest показала точність 84%, що є хорошим результатом для задачі прогнозування параметра хімічного процесу. Це свідчить про ефективність використаної технології машинного навчання для даних з хімії.

2. Покращення моделей: Для подальшого вдосконалення можна додати більше ознак, що описують хімічні процеси, або використовувати більш складні моделі, такі як нейронні мережі, щоб підвищити точність.

3. Оптимізація обробки даних: Рекомендується проводити додаткову обробку даних для виявлення можливих аномалій або шуму, що може впливати на точність прогнозів.

Перспективи розвитку та вдосконалення технологій Big Data у хімії

1. Прогнозування результатів хімічних процесів: З використанням Big Data можна ще точніше передбачати поведінку складних хімічних систем, що дозволить автоматизувати процеси та зменшити витрати на лабораторні експерименти.

2. Вдосконалення моделювання хімічних реакцій: Розвиток глибинного навчання та більш складних моделей допоможе передбачати складні хімічні реакції з великими наборами даних, що включають інформацію про температури, тиски, концентрації реагентів та інші параметри.

3. Інтеграція з іншими технологіями: Злиття Big Data з Інтернетом речей (IoT) дозволить в реальному часі збирати та обробляти дані з лабораторій та виробничих процесів, що сприятиме кращому моніторингу та управлінню процесами.

## **Список літературних джерел**

1. ACD/ChemSketch. Version 2012 for Microsoft Windows. Drawing Chemical Structures and Graphical Images Tutorial. / Advanced Chemistry

Development, Inc. 2013. – 156 p.

2. Bienz S. Short Manual to the Chemical Drawing Program ChemDraw / S. Bienz. University of Zurich. 2013. – 22 p.

3. Marvin Sketch User's Guide. Available at: <https://docs.chemaxon.com/display/docs/MarvinSketch+User%27s+Guide>.

4. Український науковий портал "Наука і техніка" – статті на тему Big Data в хімії, [[www.naukatech.com.ua](http://www.naukatech.com.ua)](<https://www.naukatech.com.ua>) (дата звернення 8.11.2024).

5. Біомедика, "Big Data у хімічних дослідженнях", [[www.biomedic.com.ua](http://www.biomedic.com.ua)](<https://www.biomedic.com.ua>) (дата звернення 10.11.2024).