

Київський національний університет імені Тараса Шевченка

Економічний факультет

Кафедра економічної кібернетики

КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

«Моделювання доходів добувної та виробничої діяльності підприємства
«ЮНІКОМ-ПРОМ» на ринку декоративного каменю України»

студента 4 курсу

спеціальності 051 «Економіка»

ОПП «Економічна кібернетика»

денної форми навчання

Чарнюк Максим Валерійович

Науковий керівник:

Кандидат економічних наук, доцент

Шпирко Віктор Васильович

Засвідчую, що в цій роботі немає
запозичень із праць інших авторів без
відповідних посилань

Студент

(підпис)

Роботу допущено до захисту перед ЕК
рішенням кафедри економічної кібернетики
від 05.06.2024р., протокол № 15

Завідувач кафедри: доктор

економічних наук, професор

Ляшенко Олена Ігорівна

(підпис)

КИЇВ-2024

РЕФЕРАТ

Кваліфікаційна робота бакалавра містить: 70 ст., 22 рис., 8 табл., 36 джерел, додатки.

Ключові слова: декоративний камінь, граніт, промисловість, виробництво, комбінації назв товарів, обсяги реалізації продукції, розподіл цін.

Об'єкт дослідження: ринок декоративного каменю та підприємство «ЮНІКОМ-ПРОМ»

Мета дослідження: прогнозування доходів виробництв на ринку декоративного каменю України, оцінки ключових трендів й небезпек ринку.

Методи дослідження: кореляційний аналіз, статистичний аналіз, регресійний аналіз, прогнозування Random Forest, автокореляція, гомоскедастичність.

Наукова новизна, теоретична значимість дослідження: оцінка ринку промисловості, детальний аналіз підприємств по добуванню та виготовленню продукції з декоративного каменю, застосування ML та NLP для аналізу збуту підприємства.

Практична цінність: визначено тенденції обсягів реалізації продукції для різних КВЕДів на ринку декоративного каменю України, визначено ключові точки для формування запасів підприємства.

RESUME

Taras Shevchenko National University of Kyiv, Faculty of Economics, Department of Economic Cybernetics

Key words: decorative stone, granite, industry, production, combinations of product names, product sales, price distribution.

The graduation research of student forecasts revenues of production in the decorative stone market of Ukraine LLC «UNIKOM-PROM», assessing key trends and market dangers.

This work deals with assessment of the industry market, detailed analysis of enterprises for the extraction and manufacture of decorative stone products, the use of ML and NLP for the analysis of enterprise sales.

The work is interesting for determining the trend of sales volumes for various NACEs in the decorative stone market of Ukraine, determining key points for the formation of enterprise reserves.

Pages 70, tables 8, bibliog. 36, append.

ЗМІСТ

ВСТУП	5
РОЗДІЛ 1. ВИКОРИСТАНІ ІНСТРУМЕНТИ ДЛЯ АНАЛІЗУ. ОЦІНКА РИНКУ ПО ВИДОБУВАННЮ ТА ВИГОТОВЛЕННЮ ВИРОБІВ З ДЕКОРАТИВНОГО КАМЕНЮ.	10
1.1.Бібліотеки Python	10
1.2.Основні етапи виробництва граніту.....	14
1.3.Стан промисловості в Україні	19
1.4.Ринок декоративного каменю	22
РОЗДІЛ 2. ДОСЛІДЖЕННЯ РИНКУ ДЕКОРАТИВНОГО КАМЕНЮ	26
2.1.Дослідження обсягів реалізації продукції та тенденцій ринку промисловості.....	26
2.2.Діяльність підприємств за агрегаціями.....	30
2.3.Спеціалізація підприємств	39
РОЗДІЛ 3. АНАЛІЗ ДІЯЛЬНОСТІ ПІДПРИЄМСТВА «ЮНІКОМ-ПРОМ». МОВНЕ МОДЕЛЮВАННЯ ПРОДАЖІВ ТА ІНТЕРПРЕТАЦІЯ ВИСНОВКІВ ЩОДО ПОКРАЩЕННЯ ДІЯЛЬНОСТІ ПІДПРИЄМСТВА.....	47
3.1.Оцінка загальних статистичних показників бухгалтерської звітності.....	47
3.2.Використання NLP задля оцінки ефективності продажів та визначення тенденцій продажів підприємства.	51
3.3.Аналіз розподілу цін за родовищами та видами продукції граніту. Оцінка ефективності прогнозування ціни Random Forest.....	56
ВИСНОВКИ.....	64
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	67
ДОДАТКИ.....	71

ВСТУП

Актуальність теми дослідження

Моделювання доходів виробництва на ринку декоративного каменю України є актуальною темою дослідження з кількох причин. По-перше, ринок декоративного каменю має значний економічний потенціал для України, враховуючи багаті природні ресурси та відновлення попиту на продукцію як на внутрішньому, так і на міжнародному ринках. Розуміння структури та тенденцій цього ринку може сприяти підвищенню ефективності виробництва, поліпшенню управління та планування, а також залученню інвестицій.

Аналіз доходів виробництва є ключовим елементом економічного прогнозування та стратегічного планування. Використання сучасних методів економетричного моделювання дозволяє більш точно оцінювати доходи, виявляти основні фактори, що впливають на прибутковість, і прогнозувати майбутні тенденції. Це особливо важливо для декоративного каменю, де ціни можуть бути високими, а ринок конкурентним.

Станом на сьогоднішній день проблема моделювання доходів на ринку декоративного каменю не є достатньо дослідженою. Більшість робіт зосереджена на загальних питаннях виробництва та маркетингу без детального аналізу доходів. Наприклад, дослідження інших вчених, таких як Смирнова та ін. (2019), зосереджуються на технологічних аспектах виробництва або загальних економічних показниках без конкретного моделювання доходів. Таким чином, наше дослідження заповнює цю прогалину, пропонуючи нові методи аналізу та прогнозування.[1]

Крім того, використання методів обробки природної мови (NLP) для аналізу структурних продажів товарів є інноваційним підходом, який дозволяє більш глибоко зрозуміти тенденції та взаємозв'язки на ринку. Інші дослідження в цій галузі, наприклад, роботи Смирнова та ін. (2019) та Коваленко та ін. (2018), демонструють потенціал NLP для аналізу текстових даних, але не застосовують ці методи до аналізу ринку декоративного каменю.

Відтак, наше дослідження пропонує унікальний підхід до аналізу ринку декоративного каменю, що базується на NLP [1][3].

Об'єкт та предмет дослідження

Об'єктом дослідження є ринок декоративного каменю України, зокрема процеси виробництва, продажу та ціноутворення на цьому ринку. Вибір цього об'єкта обумовлений його значенням для економіки України та необхідністю покращення ефективності управління виробництвом і продажами в цій галузі. Промислове виробництво у час війни знаходиться у критичному становищі, знаходячись в умовах загальносвітового зниження доходів та кризового становища національних ринків праці. Зокрема, добування та виробництво мінералів знаходиться на критичному роздоріжжі у боротьбі зі штучними матеріалами-субститутами.

Предметом дослідження є моделювання доходів від виробництва декоративного каменю та структурний аналіз продажів товарів за допомогою методів обробки природної мови (NLP). Це включає вивчення впливу різних факторів на доходи, аналіз тенденцій зміни цін та кластеризацію товарів за їх характеристиками.

Мета та завдання дослідження

Метою даного дослідження була розробка та застосування економетричних моделей для прогнозування доходів виробництва на ринку декоративного каменю України, оцінки ключових трендів й небезпек ринку та проведення структурного аналізу продажів товарів за допомогою NLP, що детальніше оцінити товарну складову попиту на ринку, а також структурувати виробничі потреби для підприємства

Для досягнення цієї мети були виконані наступні завдання:

1. Вивчити існуючі методи економетричного моделювання доходів та адаптувати їх до специфіки ринку декоративного каменю України.
2. Описати структуру та динаміку ринку декоративного каменю, визначити основні фактори, що впливають на доходи.

3. Побудувати економетричні моделі для прогнозування обсягів реалізації виробництва на ринку декоративного каменю. Зробити прогнози щодо зміни цін та доходів на основі побудованих моделей.

4. Використати методи обробки природної мови (NLP) для аналізу текстових даних про продажі товарів та визначити релевантність використання різних специфікацій каменю для підприємства «ЮНІКОМ-ПРОМ»

У дослідженні використовувалися такі методи наукового дослідження:

- Економетричне моделювання: використання регресійних моделей для аналізу факторів, що впливають на доходи та ціни на ринку декоративного каменю.

- Обробка природної мови (NLP) та машинне навчання Random Forest: виділення ключових слів та їх комбінацій у назвах товарів, аналіз текстових даних.

- Кластерний аналіз: використання методу K-середніх для групування товарів за їх характеристиками.

- Описова статистика: розрахунок середніх значень, мод, медіан та дисперсій для різних категорій товарів.

- Візуалізація даних: побудова графіків для аналізу трендів та виявлення закономірностей у зміні цін.

Наукова новизна роботи полягає в застосуванні сучасних методів економетричного моделювання та обробки природної мови для аналізу ринку декоративного каменю. Проведено комплексний аналіз доходів та цін на ринку декоративного каменю України з використанням методів NLP для структурного аналізу продажів. Отримані результати дозволяють більш точно оцінювати доходи та робити прогнози щодо майбутніх тенденцій на ринку.

Практична новизна роботи полягає в можливості застосування отриманих результатів для покращення управління виробництвом та продажами на ринку декоративного каменю. Використання розроблених моделей дозволяє

виробникам та продавцям більш ефективно планувати свою діяльність, враховуючи прогнозовані зміни на ринку.

Результати даного дослідження можуть бути застосовані для реструктуризації виробничих планів підприємства по виробництво декоративного каменю, зокрема, граніту. Планування ремонтних робіт за певними видами обладнання, а також організація постачання та відбору продукції за її характеристиками задля збільшення економічної ефективності підприємства

Інформаційна база дослідження включає офіційні статистичні дані про ринок декоративного каменю України, наукові статті та монографії, а також дані з відкритих джерел. Основними джерелами інформації були:

Державна служба статистики України: дані про виробництво та продажі декоративного каменю.

Наукові статті з баз даних Scopus, Web of Science: дослідження з економетрії та обробки природної мови. Літературними джерелами були статі, монографії та підручники з економетрії, маркетингу та NLP.

Відкриті дані з інтернет-ресурсів: додаткові дані про ринок та ціни на декоративний камінь. Дані, що надало підприємство, що діє у сфері виробничої діяльності продукції з декоративного каменю.

Використання широкого спектра джерел інформації дозволило забезпечити комплексний підхід до дослідження та отримати надійні результати.

Структура роботи

Дане дослідження буде поділено на смислові розділи, що допоможуть детальніше зрозуміти, як нішу промисловості та становище ринку, методика дослідження, а також діяльність підприємства на ринку.

Першим розділом буде теоретична інформація щодо видобування граніту, його класифікації та географічні особливості розміщення підприємств та родовищ в Україні. Також, будуть описані деякі дослідження, як ринку праці загалом, становища української економіки, а також ринку будівельних

матеріалів та будівництва, як сфери. Також, виділено певні акценти щодо поточного стану ринку декоративного каменю.

У другому ж розділі, я проведу аналіз ринку декоративного каменю на основі зібраних мною даних, та спробую прогнозувати доходи підприємств за агрегаціями та інтерпретувати дані результати. Далі на основі розуміння умов діяльності, ми оцінимо діяльність підприємства протягом 2013-2022 років У третьому розділі зробимо оцінку продажів компанії у 2021-2023 роках, за допомогою NLP.

Останній розділ матиме функціональну мету підведення підсумків, визначення ключових точок, як самого аналізу, так і об'єкту дослідження.

РОЗДІЛ 1. ВИКОРИСТАНІ ІНСТРУМЕНТИ ДЛЯ АНАЛІЗУ. ОЦІНКА РИНКУ ПО ВИДОБУВАННЮ ТА ВИГОТОВЛЕННЮ ВИРОБІВ З ДЕКОРАТИВНОГО КАМЕНЮ.

1.1. Бібліотеки Python

Проведемо огляд основних бібліотек Python, що будуть використані при аналізі даних у розділах 2 та 3. Розглянемо особливості бібліотеки, а також опишемо чому саме цю бібліотеку було обрано для аналізу:

- **Pandas**

Pandas — це бібліотека для роботи з даними, що забезпечує високоефективні, зручні у використанні структури даних і засоби аналізу даних для мови програмування Python.

DataFrame: основна структура даних, що нагадує таблиці в базах даних чи електронні таблиці. Вона дозволяє маніпулювати даними, як у рядках, так і у стовпцях.

Інтеграція з іншими бібліотеками: легко інтегрується з бібліотеками для обчислень (`numpy`), візуалізації (`matplotlib`) та машинного навчання (`scikit-learn`).

Імпорт/експорт даних: підтримує зчитування та запис даних з/до різних форматів файлів (CSV, Excel, SQL, JSON тощо).

Обробка відсутніх значень: надає функції для заповнення або видалення відсутніх даних.

- **NumPy**

NumPy — це бібліотека для наукових обчислень у Python. Вона надає підтримку для багатовимірних масивів і матриць, а також функції для виконання високорівневих математичних операцій над ними.

Масиви (arrays): забезпечує потужні N-вимірні масиви для обробки великих обсягів даних.

Математичні функції: містить функції для виконання базових арифметичних операцій, алгебри, тригонометрії, статистики тощо.

Швидкодія: завдяки оптимізованим реалізаціям операцій над масивами значно прискорює виконання коду порівняно зі стандартними структурами Python.

- Matplotlib

Matplotlib — це бібліотека для створення статичних, анімаційних та інтерактивних візуалізацій у Python.

Гнучкість Matplotlib дозволяє створювати різні типи графіків і візуалізацій, від простих лінійних графіків до складних багатовимірних діаграм. Легко інтегрується з іншими бібліотеками, такими як Pandas та NumPy, для візуалізації даних з цих джерел. Забезпечує широкі можливості для налаштування зовнішнього вигляду графіків (кольори, шрифти, лінії тощо).

- Seaborn

Seaborn — це бібліотека для візуалізації даних на основі Matplotlib, яка надає високоінформативні статистичні графіки.

Містить функції для створення графіків, що відображають розподіли, відносини між змінними та статистичні дані. Стильні графіки: забезпечує більш естетичний і зрозумілий дизайн графіків порівняно з Matplotlib. Інтеграція з Pandas дозволяє безпосередньо працювати з DataFrame об'єктами з Pandas.

- Scikit-learn

Scikit-learn — це бібліотека для машинного навчання в Python, яка включає інструменти для класифікації, регресії, кластеризації та зниження розмірності. Різноманітні алгоритми надають широкий спектр алгоритмів для машинного навчання, включаючи лінійну регресію, логістичну регресію, дерева рішень, випадкові ліси, кластеризацію K-середніх та інші. Попередня обробка даних містить інструменти для масштабування, нормалізації, заповнення відсутніх значень та інших процедур попередньої обробки. Оцінка

моделей надає методи для крос-валідації, метрики точності, середньоквадратичної помилки та інші інструменти для оцінки якості моделей.

- Statsmodels

Statsmodels — це бібліотека для статистичного моделювання в Python. Вона дозволяє проводити оцінку статистичних моделей та виконувати тестування гіпотез. Регресійний аналіз підтримує різні типи регресійного аналізу, включаючи лінійну, логістичну регресію та аналіз часових рядів.[2] Містить функції для проведення різних статистичних тестів, таких як тест Бройша-Пагана для гетероскедастичності та тест Дарбіна-Уотсона для автокореляції. Забезпечує докладні звіти про результати моделювання, включаючи оцінки параметрів, їх стандартні помилки та рівні значущості.

- XGBoost

XGBoost — це ефективна і гнучка бібліотека для градієнтного бустингу дерев рішень. Вона широко використовується для завдань класифікації та регресії. Висока продуктивність оптимізована для швидкості і використання пам'яті, забезпечуючи високу продуктивність. Обробка відсутніх значень проводиться автоматично для відсутніх значення в даних. Тюнінг параметрів надає широкі можливості для налаштування параметрів моделі, що дозволяє покращувати її якість.

- NLTK (Natural Language Toolkit)

Опис: NLTK — це бібліотека для обробки природної мови (NLP) в Python. Вона надає інструменти для роботи з текстом, аналізу та обробки природної мови. Токенізація забезпечує розбиття тексту на слова, речення або інші одиниці. Містить списки стоп-слів для різних мов, що дозволяє виключати часто вживані слова з аналізу. Лематизація та стемінг включає інструменти для приведення слів до їх базових форм.

- SciPy

SciPy — це бібліотека для наукових і технічних обчислень у Python. Вона базується на NumPy і надає додаткові функції для оптимізації, інтеграції,

інтерполяції, власних значень та інших завдань. Статистичні функції містять великий набір статистичних функцій, включаючи розподіли, тестування гіпотез та функції випадкових чисел. Забезпечує інструменти для чисельного інтегрування та розв'язання диференціальних рівнянь.

Оптимізація включає алгоритми для знаходження мінімумів і максимумів функцій. Дана робота використовує дані методи комплексно, що допомагає об'єктивізувати висновки із дослідження.

Кожна з цих бібліотек має свої унікальні особливості та можливості, що робить їх незамінними для аналізу даних, побудови моделей та візуалізації результатів. Використання цих бібліотек разом дозволяє проводити комплексний аналіз даних, оцінювати ефективність моделей та робити прогнози на основі великих обсягів даних.

1.2. Основні етапи виробництва граніту

Задля загального розуміння витрат, які потребує процес виробництва декоративного каменю, варто детально розглянути особливості його етапів, а також ключові технологічні особливості. Тоді варто почати з розуміння, що основними каменями, які називають декоративними є мрамур, граніт, анортозити, лабрадорит, а також габро (останні вважаються підвидами граніту сформованими за іншим магматичним принципом)

Методи видобування граніту:

- Вибуховий метод

Цей метод передбачає використання вибухівки для отримання великих блоків граніту. Його основною перевагою є швидкість добування великої кількості каменю. Однак, метод має значні недоліки: велика кількість відходів та низька якість отриманих блоків, оскільки вибухові роботи часто призводять до тріщин і дефектів каменів. Крім того, цей метод має високі витрати на вибухівку та пов'язані з нею заходи безпеки.

- Бурово-вибуховий метод

Цей метод включає буріння отворів у граніті, до яких встановлюють вибухівку. Він дозволяє більш контрольовано відокремлювати блоки каменю та зменшувати відходи порівняно з чисто вибуховим методом. Точність буріння дозволяє мінімізувати тріщини та інші пошкодження каменю, що призводить до вищої якості кінцевого продукту. Однак, цей метод все ще потребує значних витрат на вибухівку та обладнання для буріння, а також забезпечення високих стандартів безпеки.

- Використання алмазного дроту

Використання алмазного дроту для різання граніту є одним з найсучасніших методів, який забезпечує високу точність відокремлення блоків і мінімальну кількість відходів. Дротяні пили, обладнані алмазними сегментами, дозволяють різати камінь без зайвих пошкоджень і тріщин, що є великою перевагою для подальшої якісної обробки граніту. Незважаючи на

вищу початкову вартість обладнання та потребу в технічному обслуговуванні, довгострокові переваги цього методу з точки зору зменшення відходів та покращення якості блоків є значними.

Порівняння методів

З точки зору ефективності, використання алмазного дроту є найбільш прогресивним і ефективним методом, оскільки воно забезпечує високу якість блоків і мінімізує кількість відходів. Хоча початкові витрати на обладнання високі, вони виправдовуються довгостроковою економією на відходах та зниженням витрат на переробку.

Вибухові методи, хоч і менш витратні в короткостроковій перспективі, призводять до вищої відходності та потенційно нижчої якості продукції, що може негативно позначитися на загальній рентабельності виробництва через необхідність додаткової обробки пошкоджених блоків.

У сумі, аналіз методів добування граніту показує, що інвестиції в сучасні технології та обладнання можуть значно підвищити якість продукції та ефективність виробництва, хоча й вимагають значних початкових капіталовкладень.

Методи первинної обробки граніту

1. Розпилювання алмазними пилами

Це найбільш поширений метод, що використовується для первинної обробки граніту. Алмазні пили забезпечують точне та чисте різання, мінімізуючи кількість відходів. Цей метод вимагає значних інвестицій у високоякісне обладнання, але його ефективність та здатність зменшувати тріщини та інші пошкодження каменю роблять його оптимальним вибором для великих виробничих операцій.

2. Водяне різання

Водяне різання використовує струмінь води під високим тиском, часто з додаванням абразивних матеріалів, для різання граніту. Цей метод є більш екологічним і зменшує кількість пилу та інших викидів. Водяне різання також

забезпечує високу точність, але може бути повільнішим порівняно з алмазними пилами та вимагає додаткових витрат на воду та абразиви.

Порівняльний аналіз ефективності, витрат і відходності

Алмазне різання є більш ефективним з точки зору швидкості обробки та здатності обробляти більші обсяги каменю за одиницю часу. Однак, витрати на утримання і заміну алмазних дисків можуть бути високими, що потребує значних вкладень. Водяне різання має більші експлуатаційні витрати, але забезпечує більшу гнучкість у формуванні складних форм та зменшує кількість відходів.

Середні ціни на обробку поверхні:

- Термообробка: \$5-10 за квадратний метр.
- Полірування: \$10-20 за квадратний метр.
- Браширування: \$5-10 за квадратний метр.
- Галтування: \$15-25 за квадратний метр.

Хоча дані показник цін можуть відрізнитись залежно від структури породи каменю, періоду його утворення і тд. З такими характеристиками можна відділити кілька видів такого декоративного каменю, гранітного сімейства: граніт, анортозити, лабрадорити та габро. Анорзити є рідкісними для України та у меншій мірі використовуються для будівництва, тому давайте оцінимо інші підвиди.

Характеристика будівельного ринку під час війни

Відповідно до дослідження проведеного ГКМ та на думку директора компанії Rauta - будівельний ринок у 2023 році повинен був зрости на 25%. Так, і відбулось, адже обсяг виконаних будівельних робіт в Україні у 2022 році скоротився на 65,1% порівняно з 2021 роком – до 113,8 млрд грн. Проте у січні-червні 2023 року цей показник збільшився на 18,4% порівняно з аналогічним періодом 2022 року – до 53,2 млрд. грн. Драйвером зростання сфери будівництва у першому півріччі стали інженерні споруди, де збільшення

будівельних робіт становило 34% та нежитлове будівництво, яке зросло на 9,5%. [6]

За оцінками KSE Institute, загальні прямі збитки від військової агресії РФ станом на червень 2023 становлять понад \$150 млрд. Зокрема, найбільших збитків зазнав житловий фонд (\$56 млрд), інфраструктура (\$37 млрд) та промисловість (\$12 млрд). За два роки повномасштабного вторгнення близько 15% виробничих потужностей будівельних матеріалів зазнали руйнувань. При цьому найбільших втрат зазнали виробництво металопрокату і сухих гіпсових сумішей.[7]

Кардинально змінилася структура попиту на нові об'єкти житлової нерухомості України. Прифронтові регіони зазнали найбільшого падіння обсягів будівництва – майже на 90%, центральна частина – зменшення до 70%, а на заході будівництво зросло на 15%, що пов'язано з релокацією бізнесу та внутрішньо переміщених осіб, а також активним розвитком курортної нерухомості у Карпатах.[5]

У 2022-2023 роках собівартість будівництва збільшилася на 53%, що призвело до аналогічного зростання цін на ринку первинної нерухомості. [12]

Наприкінці 2023 року виникли ускладнення з ввезенням сировини та матеріалів через блокування польськими та словацькими перевізниками пунктів пропуску на кордоні з Україною. Це також спричинило збільшення цін і термінів поставки імпортованих будматеріалів. Через блокування кордону вартість логістики збільшилася вдвічі, що призвело до подорожчання імпортованих будматеріалів на 5-10%.

Сьогодні ринкова вартість зведення швидкокомтованих комерційних будівель складає 18-22 тис. грн за кв. м. Сам процес будівництва в середньому триває 6-9 місяців.

При зведенні котеджів за безкаркасною технологією з сендвіч-панелей ціна «під ключ» з інженерними комунікаціями та внутрішнім оздобленням складає 20-40 тис. грн за кв. м, а термін будівництва – всього два тижні.[5] [12]

За час повномасштабного вторгнення скорочення чисельності працездатного населення України склало приблизно 35%. Мобілізація та еміграція кваліфікованих фахівців спричинили дефіцит кадрів у будівельній галузі, який відчувається навіть на фоні значного падіння ринку. В майбутньому це може призвести до збільшення заробітних плат, що додатково вплине на зростання собівартості будівництва, але все одно не вирішить проблему нестачі кваліфікованих кадрів. Наразі, новий етап мобілізації, на жаль, може позбавити підприємства в Україні кваліфікованої спеціалізованої робочої сили.

Щодо оцінки конкурентних продукції видів товарів для декоративного каменю, то головною продукцією можливо виділити лише бетонні вироби, адже вони також часто використовують для благоустрою, як у плитці, так сходах, та інших цільних та збірних виробів. Хоча, використання сухих будівельних сумішей на цементній основі, як для клею, так і прокладення основи для благоустрою, є спільним для бетонних та кам'яних виробів.



Рис. 1.1. Структура ринку сухих будівельних сумішей

Джерело: [8]

На рисунку ми можемо побачити, що, внаслідок воєнних дій, активно відбувається відновлення дорожнього пішохідного покриття, адже воно дуже легко пошкоджується, внаслідок обстрілів. Також, можна помітити, що загалом більша частина СБС – це і є плиткові клеї, що нормальним показником для міжнародного будівництва, починаючи з 2010-х років, адже використання легких матеріалів, або частин конструкцій для облицювання стін, підлог, пішохідних зон та зон паркування є логічнішим, а у декоративних каменях у ЄС є нормованим стандартом, адже забезпечує рівномірність шару покриття, відповідно до поступовості висихання та відносно високої густини.

Варто оцінити загальне становище промисловості, щоб відносно оцінити критичність змін у окремих сферах.

1.3. Стан промисловості в Україні

Загалом, варто вказати, що почалось відновлення показників, що було втрачено у 2022 році. Так, звичайно, велика кількість підприємств закрилась, або перейшла у вигляд ФОП, яким досі було легше існувати в умовах війни. В той ж самий час, тренд є досить схожим на процес активізації виробництв, внаслідок децентралізації з 2015 року. Хоча й можливо очікувати загальний тренд до падіння промисловості, але в Україні, досить важливим стає фактор експорту продукції закордон, для чого ми повинні використовувати усі промислові потужності, а також модернізувати їх за можливості. Інноваційність виробництв є критичним фактором початку воєнних дій, але намагання навіть цифровізації та відповідності міжнародним стандартам стає зараз достатньою умовою для отримання грантів. Розвиток малих промислових підприємств, а також у поєднанні з появою інфраструктури для освіти та кооперації, що виникає з появою нових індустріальних парків (понад 13 у 2023 році). [8][10]

Галузевий аналіз промисловості варто провести для уточнення специфіки взаємодії підприємств у сфері видобутку каменю та порівняти загальні

тенденції, що притаманні економіці України, і гранітним видобувним та виробничим підприємствам окремо.

Легка промисловість: близько 62% підприємств працюють майже на повну потужність.

Енергетика отримала великі пошкодження енергетичної інфраструктури та процес її відновлення може становити довгий час, що сприяє появі енергонезалежності у окремих підприємств.

Релокація підприємств: за даними інституту економіки та прогнозування у 2022 році на західну Україну релоковано 800 підприємств, з них: у Львівську область – 24% (легка промисловість, металообробка, машинобудування, виробництво комплектуючих), Закарпатську – 14,5% (легка промисловість, машинобудування, меблева, оброблення деревини), Івано-Франківську – 8,3% (швейна промисловість, машинобудування, металообробка), Чернівецьку – 9,8% (легка промисловість, виробництво високовольтного обладнання, модульних будинків, деревообробна галузь), Хмельницьку – 7,3% (швейна промисловість, машинобудування, виробництво електричного устаткування, СІП панелей), Тернопільську – 6,3% (харчова промисловість, виробництво будівельних матеріалів, холодильного обладнання, металевих виробів).

Важливим для аналізу є виділення похідної продукції з каменю, адже навіть при великих добувних та виробничих потужностях, відходність породи каменю може бути до 75%. Увесь інший сировинний потенціал, або складається на промислових звалищах, або йде у переробку на мілку продукцію. Це може бути бутовий камінь, або мілко перебитий щебінь, що також використовується не лише, як дорожнє покриття, а й складова, що заливається сухими будівельними сумішами та іншими клейовими речовинами, як основа робіт з облицювання та благоустрою.[9]

Отже, на рис. 1.2. ми можемо побачити, що у 2023 році, як і у всіх інших сферах пішло відновлення переробних підприємств для щебеню, хоча й у меншій мірі, ніж загальні темпи ринку, хоча ріст є й значним. Великою проблемою є той втрачений потенціал на Сході України, що ми на даний

момент не можемо повернути. Загалом, кількість щебню, що могла вироблятися у прифронтових зонах є ваговою, але не критичною, адже використання для щебню певних порід каменю не потрібно.[10]

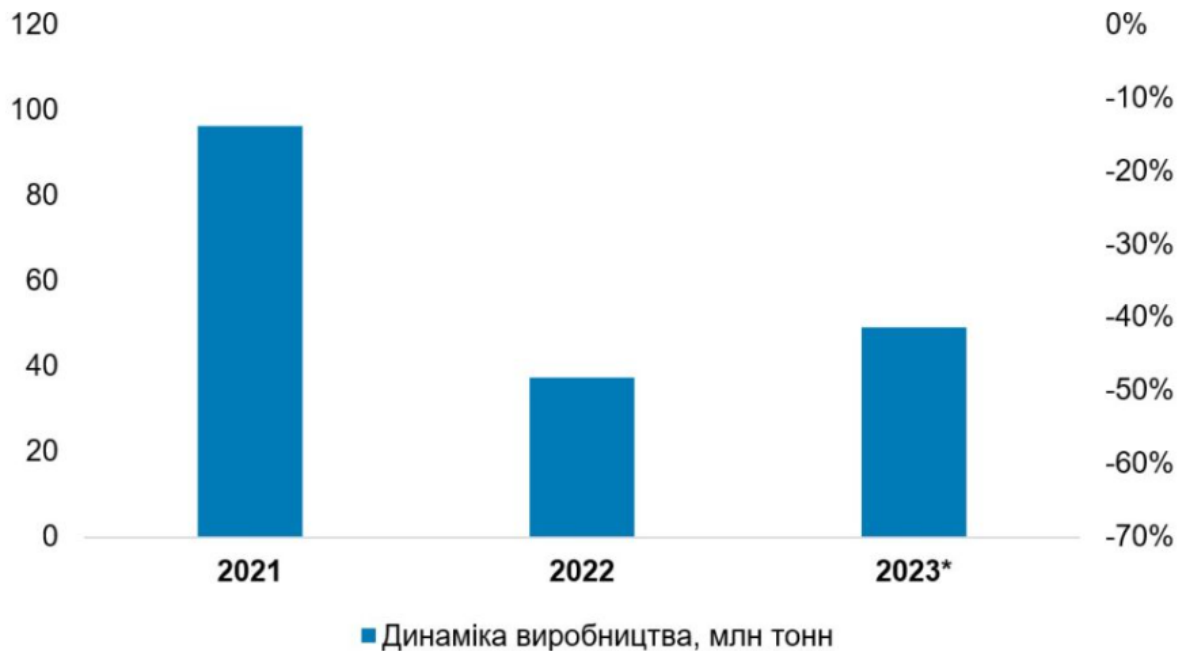


Рис. 1.2. Динаміка виробництва щебню

Джерело: [10]

Ринок лише потребує загальної реструктуризації та релокації, адже технічні вимоги таких виробництв є низькими, а втрата якості робочої сили не буде позначатись на якості продукції, тому варто враховувати лише регіональні особливості. Зокрема, варто відзначити, що ріст відбувся переважно у Закарпатській та Житомирській областях (де і так знаходилась найбільша кількість побідних виробництв), а от щодо проміжних регіонів варто чекати виготовлення аналогів у Центральній Україні, де переважно піщаний ґрунт вимагатиме більших витрат на добування.

1.4. Ринок декоративного каменю

Україна має значний мінерально-сировинний потенціал, особливо у виробництві високоміцних порід таких як граніт, габро та лабрадорит, які формують основу ринку декоративного каменю країни.

Граніт є найбільш поширеним декоративним каменем, використовується в архітектурі та будівництві завдяки своїй міцності та естетичній привабливості. Україна має великі запаси граніту, особливо у західних регіонах. Габро відомий своєю темною текстурою та використовується переважно в якості облицювального матеріалу та для скульптурних робіт. Лабрадорит славиться своїми унікальними оптичними властивостями, які включають іризуючий відблиск, і також активно застосовується в декоративних та облицювальних роботах.[4]

Ці породи становлять приблизно 75% частки ринку природного каменю України. Виробництво граніту складає 32 714,4 тис. тонн, лабрадориту — 665 тис. тонн, а габро — близько 287 тис. тонн на рік.

Порівняння з Іншими Видами Декоративного Каменю

У порівнянні з мармуром, травертином та іншими видами декоративного каменю, граніт, габро та лабрадорит вирізняються вищою міцністю та довговічністю, що робить їх більш привабливими для використання в зовнішніх роботах і великих будівельних проектах. Відсоток експорту цих каменів є відносно високим, особливо на ринках, де високо цінують їх унікальні характеристики.[20][21]

Потенціал Виробничих Потужностей

Україна володіє значними запасами високоякісного декоративного каменю, такого як граніт, габро, та лабрадорит. За оцінками, обсяг запасів декоративного каменю в Україні становить понад 500 мільйонів м³. Така мінерально-сировинна база надає Україні потенціал стати одним із світових лідерів у виробництві та експорті декоративного каменю.

На території країни розташовані численні кар'єри та заводи з обробки каменю, які володіють потужностями для виробництва великої кількості готової продукції. Виробничі лінії, зокрема, включають сучасне італійське обладнання з номінальною потужністю 200-250 тис. м² облицювальних плит на рік на кожній лінії.[18]

Поточний технічний та технологічний стан

Хоча в країні й існує солідна база для виробництва декоративного каменю, технічний та технологічний стан багатьох кар'єрів та заводів залишається застарілим. Більшість обладнання в галузі ще датується радянськими часами або раннім пострадянським періодом. Найпоширенішими є вибухові методи видобутку, що негативно впливає на якість кінцевого продукту та зумовлює великі втрати корисного матеріалу.

До числа технологічних проблем можна віднести недостатню автоматизацію процесів обробки каменю. В більшості випадків обладнання потребує ручного керування, що знижує продуктивність праці та підвищує ризик виробничих помилок. Пневматичні системи приводу, які домінують на українських підприємствах, характеризуються високим рівнем амортизації та обмеженою ефективністю порівняно з сучасними електричними та гідравлічними системами.[11]

Числові показники виробництва та експорту (обсяги реалізації):

- Граніт: 32714,4 тис. т
- Лабрадорит: 665 тис. т
- Габро: 287 тис. Т

Середня ціна експорту на готову продукцію на 2021 рік:

- Граніт: 364 \$/т
- Лабрадорит: 665 \$/т
- Габро: 287 \$/т

Ціни на внутрішньому ринку зазвичай нижчі порівняно з експортними через високі витрати на логістику та менший податковий тягар для місцевих

операцій. Експортні показники за 2019-2021 роки свідчать про різке падіння обсягів експорту природного каменю, що частково пояснюється пандемією COVID-19. Надалі показники далі падали, хоча й систематизованого аналізу у відкритих джерелах не було знайдено.

Продаж готової продукції є досить логічним для мілких виробництв у вигляді цільних та полірованих слябів продукції, що використовується, як у внутрішньому, так і зовнішньому оздобленні по всьому світу. Бруківка ж, або інша продукція може вимагати більшої кількості робочої сили та терміни виготовлення. Складні ж продукти, що виготовляються індивідуальними майстрами можуть бути дорожчим за середні за 2021 рік, але така продукція є індивідуальною, та часто виробляється локально місцевими виробниками безпосередньо в країні. Тоді на експорт буде проданий похідна сировина у вигляді добутих слябів без попередньої обробки.

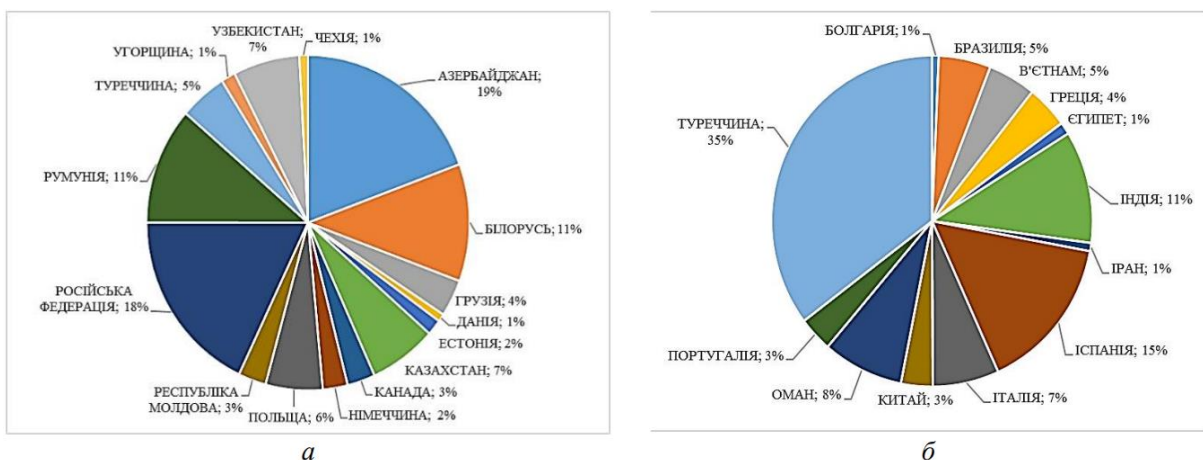


Рис. 1.3. Експорт (а) та імпорт (б) природного каменю у 2021 році

Джерело: [13]

Сама ж структура міжнародних економічних відносин, як бачимо на рис. 1.3., має основним напрямком країни Європи, а також деякі виключення із Азійського регіону. Наприклад, Азербайджан, який активно використовує граніт на вулицях міст. Хоча й велику частину експорту до війни займала РФ та Білорусь, наразі, цей попит можливо знайти у Західній Європі, не втрачаючи

переваг у відстані доставки. Основний ж імпорт каменю є піщані та мармурові породи, що не є поширеними в Україні. Також, можемо виділити взаємовідносини з Китаєм, куди наразі продає велика кількість локальних та великих виробництв із України, а вони нам ж імпортують його у вигляді готової продукції. У цьому напрямку відбувається рух в напрямку Північної Америки, що має схожі логістичні показники, але і має кілька країн, що активно використовують декоративні камені. Зокрема, Канада навіть у 2021 році, яка мала 3% експорту. Також, варто відзначити великі країни-конкуренти, які можуть бути нашими партнерами, адже я великими, як виробника, так і посередниками у ринку каменю – це Індія та Туреччина. Обидві держави експортують велику кількість товарів із граніту та мрамору, але важливим є те, що вони активно закупають сляби та іншу породу етапу первинної обробки.

Для зміцнення позицій на світовому ринку та підвищення якості продукції рекомендується вжити наступні кроки:

1. Модернізація обладнання – Заміна застарілого обладнання на сучасні автоматизовані системи, що дозволять підвищити точність обробки, зменшити відходи та збільшити швидкість виробництва. Застосування передових технологій у видобутку та обробці каменю, таких як водяне різання та лазерне сканування, що дозволяє оптимізувати виробничі процеси.

2. Розвиток науково-дослідної бази – Інвестиції в наукові дослідження та розробку для вдосконалення виробничих технологій та створення нових товарів та матеріалів з використанням декоративного каменю.

РОЗДІЛ 2. ДОСЛІДЖЕННЯ РИНКУ ДЕКОРАТИВНОГО КАМЕНЮ

2.1. Дослідження обсягів реалізації продукції та тенденцій ринку промисловості

Підприємство, що буде проаналізоване за державною класифікацією відноситься до малих підприємств, адже має від 10 до 50 робітників з річним доходом у межах 10 млн. доларів. Тому варто оцінювати діяльність підприємства саме на рівні малих підприємств. Хоча, в той самий час, основна частина ринку гранітних виробів та виробів з декоративного каменю перенасичений індивідуальними підприємцями з місцевими цехами-виробництвами, але як діяльність на ринку даних підприємств абсолютно відрізнятиметься від методів доставки, так і до об'ємів та цін закупленої сировини. Дані були взяті із державного сайту статистики [14], а далі проаналізовані автором задля отримання висновків. Таблиця та код дослідження надано у Додатку А, Таблиця А.1. та Додатку Б, Код Б.1.. Було проаналізовано 3 показники діяльності підприємств за масштабом: кількість суб'єктів господарювання, кількість суб'єктів господарювання на 10 тис. осіб, а також обсяг реалізованої продукції підприємствами у млн. гривень.

Таблиця 2.1. Статистичні показники Кількість суб'єктів господарювання на 10 тисяч осіб наявного населення, одиниць.

	Усього	Підприємства усього	Малі підприємства
Середнє значення	432.92	83.25	78.83
Медіана	443.50	82.50	77.50
Мода	462.00	80.00	76.00
Дисперсія	1878.27	29.48	26.15
Стандартне відхилення	43.34	5.43	5.11

Продовження таблиці 2.1.

Стандартне відхилення	43.34	5.43	5.11
Логарифмічне відхилення	0.10	0.07	0.00

Джерело: розрахунки автора на основі Додатку Б(табл. Б.1.)

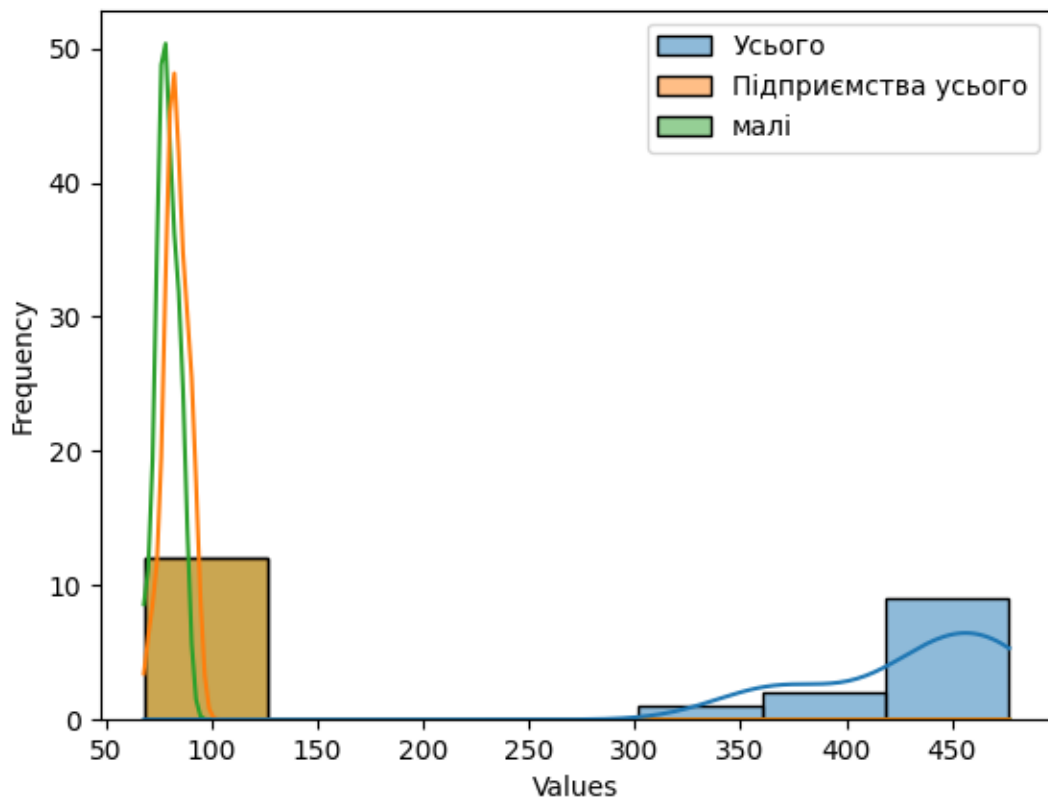


Рис. 2.1. Стандартний розподіл кількості підприємств на 10 тис.осіб.

Джерело: дослідження автора на основі Додаток Б (табл. Б.1.)

Середнє значення вказує, що в середньому на 10 тисяч осіб наявного населення припадає близько 432,92 суб'єктів господарювання. Підприємства складають близько 83,25, з яких більшість є малими – 78,83. Невелике значення стандартного відхилення свідчить про незначну варіативність показників. Позитивні значення коефіцієнтів кореляції Пірсона та Спірмена вказують на позитивну, майже абсолютну кореляцію між кількістю

підприємств і малих підприємств. Причиною для цього є, вже раніше описана, розшарованість діяльності ФОПів, які фактично складають більшу частину кількості учасників ринку, відповідно до рис. 2.1. близько 75% ринку.

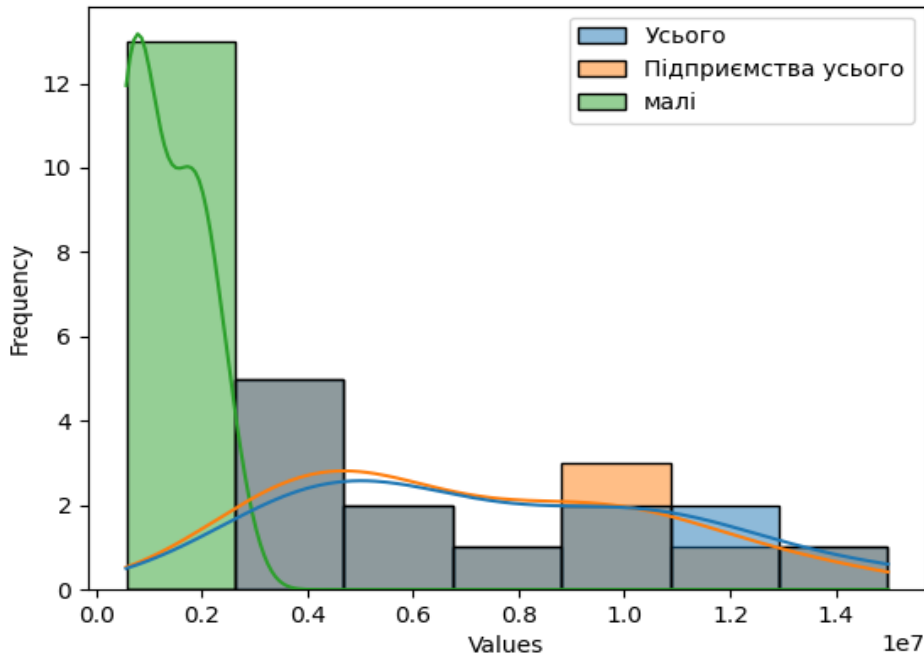


Рис. 2.2. Стандартний розподіл обсягів реалізованої продукції.

Джерело: дослідження автора, [14], Додаток Б (табл. Б.1.)

У той ж час, оцінюючи вже економічну діяльність підприємств, варто зазначити, що наявний рівномірний спадний розподіл реалізованої малими підприємствами продукції та послуг має досить чіткі межі, але в той ж час має великі відхилення значень за рис. 2.2. До того ж при великій кількості таких підприємств маємо не порівнюємо малі показники діяльності, що означає, що більшість таких малих підприємств знаходяться в стані монотонного підтримання рівня діяльності, а не розвитку, як технологічного, так і фінансового. В той ж час, і ФОПи, що становлять основну масу ринку, не мають такого економічного впливу на загальні показники, у той ж час підприємства в цілому, відповідаючи загальній тенденції на високе відхилення абсолютних показників, забезпечують практично усі фінансові кошти, що може свідчити про велику монополізованість, або олігополізованість ринку.

По-перше, це можливо аргументувати, справді, великою кількістю мережевих послуг в Україні, і не дивлячись, на підтримку малого бізнесу та підприємців, це змінити не вдається. По-друге, централізація державного бюджету, а більше того зосередження ВВП в руках державних органів, що почалось з 2022 року, і це вже видно на показниках, адже % підприємств у реалізованій продукції досягає практично 100%, що означає, або про підлеглисть послуг ФОПів та малих підприємств великим, а також про вузьконаправленість державних витрат на великі підприємства.

Таблиця 2.2. Обсяг реалізованої продукції за масштабом підприємства (товарів, послуг) млн. грн.

	Усього	Підприємства усього	Малі підприємства
Середнє значення	7,639,348.00	7,110,116.00	1,320,684.00
Медіана	6,726,739.80	6,237,535.20	1,177,385.20
Мода	3,596,646.40	3,366,228.20	568,267.10
Стандартне відхилення	3,611,558.00	3,320,181.00	689,073.90
Логарифміч не відхилення	0.48	0.47	0.55

Джерело: розрахунки автора на основі Додатку Б(табл. Б.1.)

Із таблиці 2.2. маємо, що середнє значення показує, що загальний обсяг реалізованої продукції складає близько 7,639,348 млн грн, з яких більшість припадає на підприємства 7,110,116 млн грн, тоді як малі підприємства реалізують продукції на 1,320,684 млн грн. Значення t-статистики та р-значень підтверджують статистичну значущість отриманих результатів. Загальний

аналіз показників підприємств за масштабом свідчить про значну варіативність у кількості суб'єктів господарювання та обсязі реалізованої продукції. Малі підприємства складають значну частку як у загальній кількості суб'єктів, але вже не у реалізованій продукції. Позитивні кореляції між різними категоріями підприємств вказують на взаємозв'язок між різними масштабами діяльності підприємств. Результати підтверджують необхідність подальшого аналізу та моніторингу для забезпечення ефективного управління та підтримки малих підприємств, більш детальна оцінка в порівнянні з ФОПами.

2.2. Діяльність підприємств за агрегаціями

Діяльність у видобутку каменю та виготовленню виробів із декоративного каменю, хоча і є технологічно важкою, а застаріле обладнання може мати критичний вплив на ефективність підприємства, але саме такі технології називають технологіями низького рівня, адже каменевидобувна діяльність на землі почалась майже 3 млн. років тому. Саме ж підприємство здійснює промислову діяльність, а також має відділ наукових досліджень породи каменю, щоб забезпечити безпеку власних працівників, а також збільшити ефективність виробництва через зменшення відходності. [12]

Тепер, знову ж на основі даних із державної служби статистики, було обрано ключові діяльності підприємства агрегації, та проаналізовано аналогічні показники: кількість суб'єктів господарювання, кількість суб'єктів господарювання на 10 тис. осіб, а також обсяг реалізованої продукції підприємствами у млн. гривень, а також: кількість найманих та зайнятих показників.

Зважаючи, на рис. 2.3, легко помітити лінійний рівномірний ріст обсягів виробництва та доданої вартості за витратами виробництва, що вагомо підсилює аналогічний поступовий ріст обсягу реалізованої продукції. Падіння ж у 2022 році в межах 20% показує стійкість й малу вибагливість сфери

діяльності до економічних потрясінь, адже не потребує такої кількості, як інвестицій у підтримання стану підприємств, так і у навчання робітників, а самі вимоги до робітників є нижчими.

Таблиця 2.3. Описові статистика для показників виробництв з використанням технологій низького рівня

	Середнє значення	Медіана	Мода
Кількість діючих підприємств, одиниць	16,405.31	16,280.00	1,686.74
Кількість зайнятих працівників, тис. осіб	581.66	550.20	72.10
Обсяг реалізованої продукції (товарів, послуг), млн. грн	644,914.74	638,848.90	308,349.82
Обсяг виробленої продукції (товарів, послуг), млн. грн	972,158.63	777,058.10	856,675.88
Додана вартість за витратами виробництва, млн. грн	273,589.44	177,424.10	294,945.21

Джерело: дослідження автора на основі Додаток Б (табл. Б.1.)

Виробництва під час війни не втрачають, але недореалізують, що в майбутньому вплине на задоволеність робітників, що недоотримують коштів за відповідну роботу, а також стан діяльності підприємств загалом.

Виробництва технологій низького рівня мають відносно стабільні показники з невеликою варіацією, що підтверджується високим R^2 Score - 0.8781. Квадратичне відхилення не є значущим 0.0210, навіть менше за 0.05, що є чудовим показником. Високі значення доданої вартості свідчать про значний економічний внесок цього сектору. Низьке р-значення 0.0973 тесту Бройша-Пагана вказує на можливі проблеми з гетероскедастичністю, що означає прийняття гіпотези про гомоскедастичність, а тому дисперсія залишків немає критичного тенденційного відхилення для прогнозу, але загалом модель має високу точність.



Рис. 2.3. Нормований тренд показників для виробництв з використанням технологій низького рівня.

Джерело: дослідження автора на основі Додатку А(Код А.1)

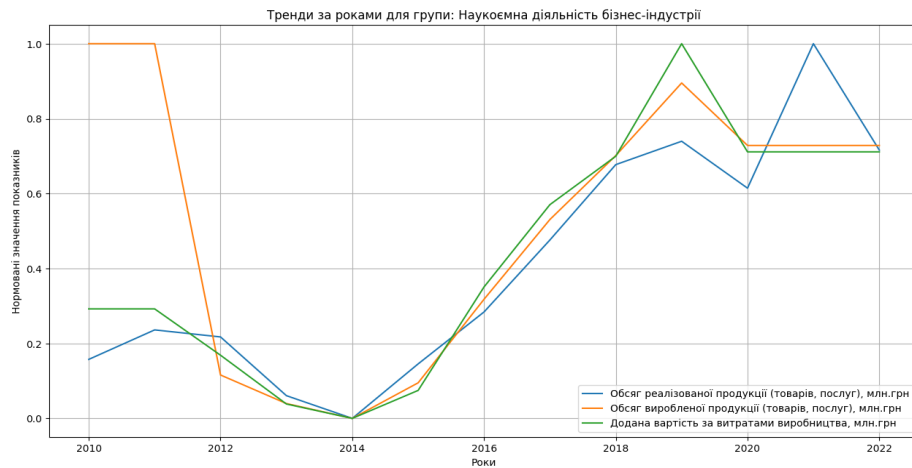


Рис. 2.4. Нормований тренд для показників наукоємної діяльності бізнес-індустрії.

Джерело: дослідження автора на основі Додатку А(Код А.1)

За рис.2.4. відновлення та активний позитивний тренд розвитку сфери з 2014 року був досить сильно – практично на 25% загалом порушений у всіх напрямках. Лаги показників, що переходять з виробництва у збут із затримкою у 2 роки, варто помітити, а також можна вказати, що наявна рівномірна дисперсія значень (наявна гомоскедастичність 0.097 за тестом Бройша-Паган). Для наукоємної діяльності бізнес-індустрії високе MSE 0.0210 показників свідчить про значну варіацію між підприємствами в цій групі. Негативний R² Score вказує на неточність моделі прогнозування для цієї групи. Проте високі показники середньої доданої вартості підтверджують значний вклад у економіку. Як помітно у таблиці 2.4, модельне може бути надійною з такими відхиленнями у значенням між модою та медіаною.

Таблиця 2.4. Описові статистика для показників виробництв з використанням технологій низького рівня

	Середнє значення	Медіана	Мода
Кількість діючих підприємств, одиниць	54,379.62	56,414.00	6,138.30
Кількість зайнятих працівників, тис. осіб	631.42	578.10	104.88
Обсяг реалізованої продукції (товарів, послуг), млн. грн	659,267.08	576,580.50	206,064.90
Обсяг виробленої продукції (товарів, послуг), млн. грн	598,353.87	700,957.00	221,044.48
Додана вартість за витратами виробництва, млн. грн	254,020.72	229,373.40	96,747.44

Джерело: дослідження автора на основі Додаток А(Код А.1)

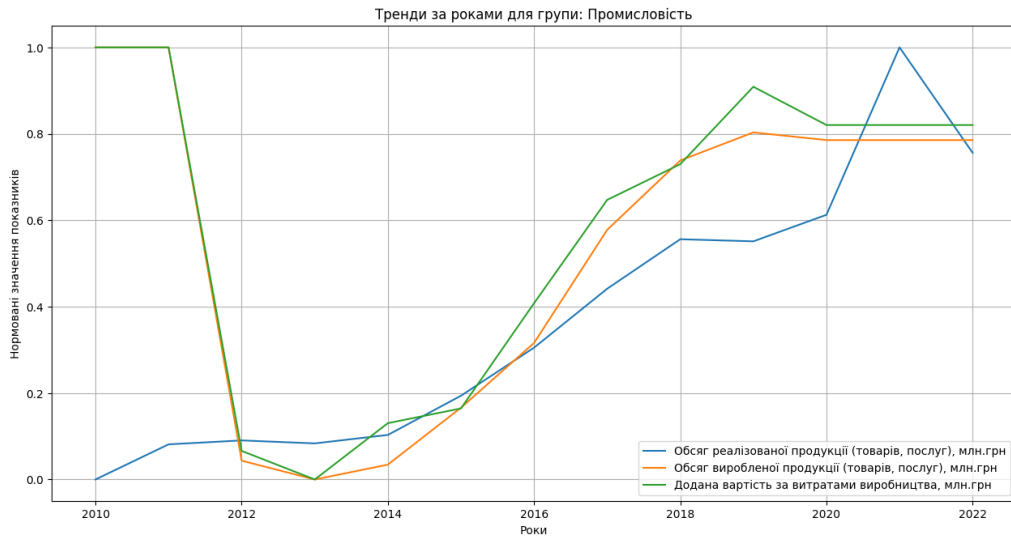


Рис. 2.5. Нормований тренд показників для промисловості.

Джерело: дослідження автора на основі Додаток А(Код А.1)

Далі, можемо побачити статистичні показники для промисловості у табл. 2.5., що вказують на надмірну варіативність показників у даній групі, хоча й рівномірні значення доданої вартості показують загальне стратегічне зниження фактичних прибутків у виробників.

Група промисловості демонструє значні показники у всіх аспектах діяльності, що вказує на важливість цього сектору для економіки. Бачимо, що падіння не є критично високим у 2022 році, але й так, як відображає узагальнену діяльність усієї промисловості, то вказує на критичні втрати не лише в стані ринку, а потенціалі наявних підприємств. Відповідність моди у обсягу виробленої продукції середньому значенню показує плановість роботи подібних підприємств, що може мати позитивний вплив за умов якісного розподілу поточних витрат і відсутності значного падіння попиту на ринку. Гетероскедастичність наявна, але не висока, фактично можливо не розраховувати на ріст дисперсії надалі, адже р-значення тесту Бройша-Пагана – 0.13. Невисоке квадратичне відхилення свідчать про малу варіативність

показників між підприємствами промисловості, хоча й він $0.0536 > 0.05$. R^2 Score - 0.6796 вказує на середню точність моделі прогнозування для цієї групи.

Таблиця 2.5. Описові статистика для показників промисловості

	Середнє значення	Медіана	Мода
Кількість діючих підприємств, одиниць	44,501.15	44,425.00	4,085.86
Кількість зайнятих працівників, тис. осіб	2,380.66	2,176.90	477.31
Обсяг реалізованої продукції (товарів, послуг), млн. грн	2,539,180.00	2,305,695.90	1,170,290.00
Обсяг виробленої продукції (товарів, послуг), млн. грн	2,843,090.00	2,746,608.10	1,748,908.00
Додана вартість за витратами виробництва, млн. грн	1,048,287.00	868,069.90	842,646.60

Джерело: дослідження автора на основі Додаток А(Код А.1)

Загальні показники діяльності підприємств відображають високу активність та продуктивність у всіх групах. Висока дисперсія та стандартне відхилення свідчать про значну варіативність між підприємствами. Високий R^2 Score - 0.7503 вказує на добру точність моделі прогнозування для загальної групи. У той ж час, бачимо, що високе значення р-значення тесту Бройша-Пагана – 0.8736 відображає критичну нестабільність загальних економічних тенденцій, адже наявна гетероскедастичність, тому прогноз може мати великі

відхилення у результатах. Квадратичне відхилення знаходиться в межах норми – 0.0423.

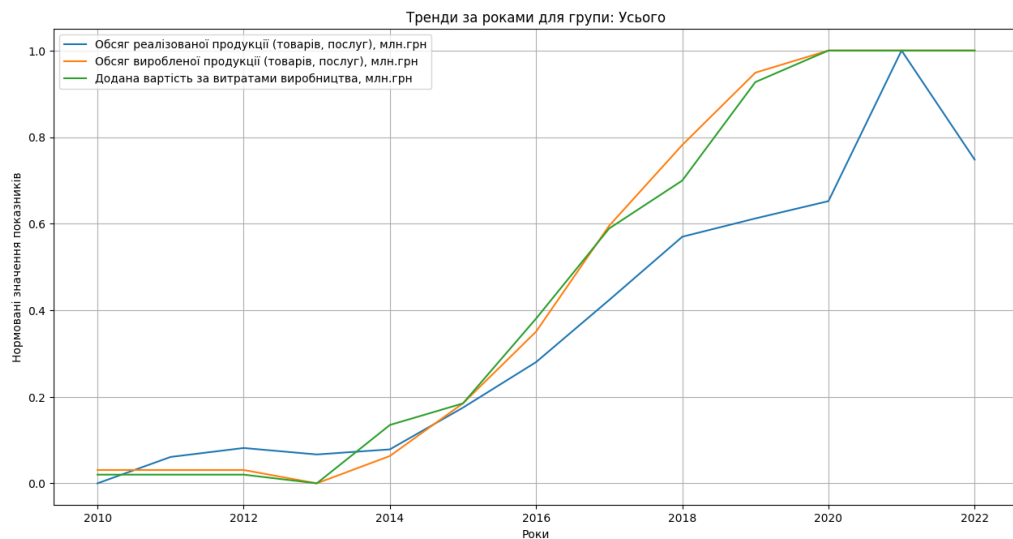


Рис.2.6. Нормований тренд показників для усіх підприємств України
Джерело: дослідження автора, [14], Додаток А (Код А.1)

Розраховувати на прогнозування у даній моделі можливо, з огляду на підвищення плановості виробленої продукції з 2020 року, в умовах пандемії та війни варто відзначити прямий позитивний вплив, який це може мати у подальшому на економіку України, адже попередні такі періоди стабільності означали стрімке довгострокове зростання, але ж у той ж час, недостатність коштів на ринку та великий вплив держави у розподіл іноземних коштів може призводити до подальшого зменшення обсягів реалізації, за відсутності ринку для такої реалізації.

Прогнози на 2023 рік:

- Виробництво з використанням технологій низького рівня: -0.11% приросту
- Наукоємна діяльність бізнес-індустрії: -0.10% приросту
- Промисловість: -0.11% приросту
- Загалом: -0.11% приросту

Таблиця 2.6. Описові статистика для показників підприємств України

	Середнє значення	Медіана	Мода
Кількість діючих підприємств, одиниць	352683.6	364935	35814.5
Кількість зайнятих працівників, тис. осіб	6549.77	6366.1	862.8
Обсяг реалізованої продукції (товарів, послуг), млн. грн	7110116	6237535.2	3320181
Обсяг виробленої продукції (товарів, послуг), млн. грн	4338295	3884617.6	1747057
Додана вартість за витратами виробництва, млн. грн	1853368	1702670.5	806226

Джерело: дослідження автора на основі Додаток А (Код А.1)

Прогнози на 2024 рік:

- Виробництво з використанням технологій низького рівня: 0.00% приросту
- Наукоємна діяльність бізнес-індустрії: 0.00% приросту
- Промисловість: 0.00% приросту
- Усього: 0.00% приросту

Звичайно, кожна з моделей маючи свою надійність має відповідну надійність прогнозу, тому кожен з них мав би бути різним, але за процес відновлення попереднього тренду діяльності бізнесу в державі на 2 роки для умов непередбачуваного впливу (як зараз війна), так і для відновлення тренду регресії, можна вказати, що прогнозування для моделі XGBoost є практично ідентичним для всіх агрегація. Їхнє загальне падіння під час воєнних дій було схожим, тому й вихід у поступовий розвиток також буде схожим. До кінця 2023 року можна було б очікувати ще невеликого падіння ринку, але у реальності, окрилені частково успішним наступом, а також максимальною задіяністю в економічній діяльності всього населення, показники 2023 показали ріст у практично 7%. Прогноз ж на 2024 залишається відповідною точкою сталось, і очікуваний нульовий приріст в реальності замінюється від місяця до місяця, то падінням, то ростом на 2-3%.

Прогнози на 2023 рік вказують на незначне зниження приросту у всіх групах, тоді як на 2024 рік передбачається стабілізація без змін. Це свідчить про потенційну стабілізацію економічної ситуації та можливу стабільність у майбутньому, за умов визначення довгострокового плану розвитку.

Найкращі прогнози отримані для групи "Виробництво з використанням технологій низького рівня", що підтверджується найнижчим MSE та високим R² Score. Група "Наукоємна діяльність бізнес-індустрії" має найгірші прогнози, що свідчить про необхідність удосконалення методології прогнозування для цього сектору.

2.3. Спеціалізація підприємств

Цей розділ аналізу поглинається глибше у сферу діяльності підприємства та оцінює діяльність за КВЕДами 8.01 (добування каменю, піску та глини), 8.11 (добування декоративного та будівельного каменю, вапняку, гіпсу, крейди та глинистого сланцю), 23.70 (різання, оброблення та оздоблення декоративного та будівельного каменю). полягала у визначенні впливу різних факторів на цільову змінну та виявлення ключових закономірностей для подальшого прогнозування та інтерпретації результатів прогнозування. Дані містять інформацію про спеціалізацію підприємств, включаючи обсяг реалізованої продукції серед суб'єктів мікро підприємництва за роки з 2010 до 2022, і саме цей показник визначений ключовим у моделі, адже під час війни діяльність багатьох підприємств обмежилась, а також фактично зникли малі підприємства у сфері добування та виготовлення каменю, а середні та великі поступово зникали ще до цього. Основні показники для аналізу включають кількість діючих підприємств, витрати на виробництво, капітальні інвестиції та інші фінансові та виробничі метрики.

Попередня обробка даних потребувала зміни базового датасету для підвищення точності дослідження, адже багато значень, або відсутні загалом за певний період, або ж не заповнені у даних держстату [14], додаток 1.3. Тому я вирішив, що варто заповнити відсутні показники – прогнозованими. Для заповнення пропущених значень використовувався KNNImputer з параметром ``n_neighbors=5``. Це дозволило мінімізувати вплив пропущених даних на

результати аналізу, заповнивши їх на основі схожих рядків у наборі даних.

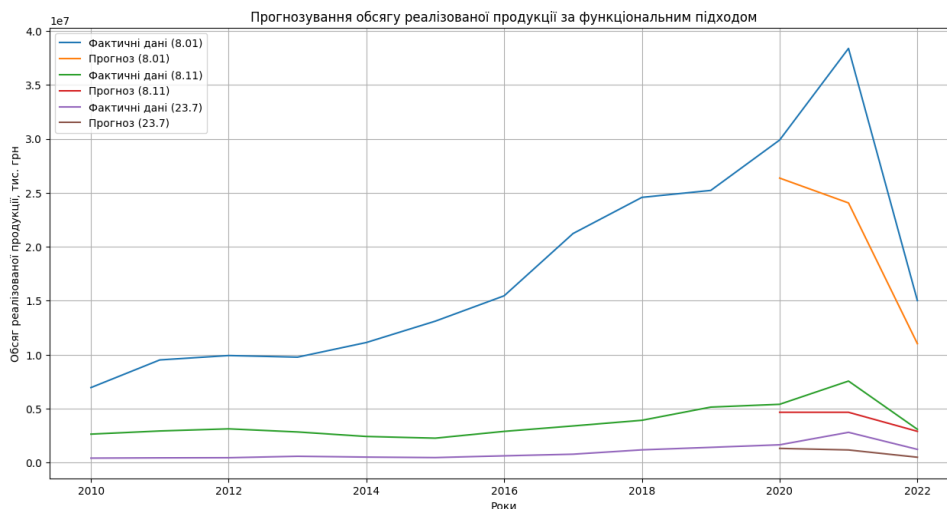


Рис. 2.7. Тренд обсягу реалізованої продукції за КВЕДами

Джерело: дослідження автора на основі Додаток А (Код А.1)

За рис. 2.7. можемо побачити, що інтенсивний тренд на ріст мала лише видобувна промисловість загалом, особливо за КВЕДом 8.01, адже спектр застосування продукції є ширшим, ніж у декоративного та будівельного каменю. Пісок та глина має більшу аудиторію попиту, так і сферу застосування, отже інвестиції були вагоміші. Під час війни, велика частина відповідних кар'єрів були знищені, але значення не впали нижче вихідних з 2010. Хоча, ріст промисловості загалом у 2021 році був непомірним, але КВЕД 23.70 має більшу стабільність та прогнозованість показників, відповідно до меншої волатильності значень. В той ж самий час, виготовлення продукції напряму залежить від видобувної сфери й рідко може переганяти її, за умов рівномірного росту доданої вартості. Для нормалізації даних використовувався MinMaxScaler, що привело всі показники до діапазону $[0, 1]$. Це було необхідно для забезпечення коректної роботи алгоритмів машинного навчання та порівнянності різних показників.

Для кожної групи була побудована модель лінійної регресії. Ось основні результати для кожної групи:

- КВЕД 8.01

Ця група має помірний рівень пояснювальної здатності моделі (R^2 Score), що свідчить про те, що модель може пояснити близько 65% варіації цільової змінної – високе. Значення MSE вказує на помірну середню похибку прогнозу, що майже знаходиться в межах надійного прогнозу – 0,05. Загалом, можна вважати дані цього прогнозу релевантними, але не надійними.

- КВЕД 8.11

Mean Squared Error (MSE): 0.1235, R^2 Score: 0.1417. Для цієї групи модель має низьку пояснювальну здатність, що вказує на необхідність додаткового аналізу або використання інших моделей для поліпшення результатів.

- КВЕД 23.7

Mean Squared Error (MSE): 0.1492, R^2 Score: 0.1126. Подібно до групи 8.11, ця група також має низьку пояснювальну здатність моделі, що потребує подальшого вдосконалення моделей або додаткового аналізу даних, адже загальне відхилення є вищим за здатність прогнозу.

Високий рівень кореляції між обсягом реалізованої продукції та собівартістю реалізованої продукції (0.992), накопиченою амортизацією (0.951), первісною вартістю (0.916) вказує на визначеність цінової політики підприємств у сферах, відповідно до умов ринку, рідко утворюється додаткова вартість за допомогою реклами.

Кореляційна матриця:

Код за КВЕД	Код за КВЕД ...	Обсяг реалізованої продукції з них суб'єкти мікропідприємництва, тис.грн
Код за КВЕД	1.000000e+00 ...	0.678643
Рік	-9.866771e-15 ...	0.454046
Кількість діючих підприємств за однорідними вид...	-5.824171e-01 ...	-0.537847
Витрати на виробництво продукції	-5.341656e-01 ...	-0.321720
матеріальні витрати та витрати на оплату послуг...	-5.068331e-01 ...	-0.299318
амортизація	-6.019872e-01 ...	-0.368995
витрати на оплату праці	-5.458337e-01 ...	-0.323766
відрахування на соціальні заходи	-6.274891e-01 ...	-0.437203
інші витрати	-6.765361e-01 ...	-0.474253
Капітальні інвестиції - усього	-5.155158e-01 ...	-0.315236
капітальні інвестиції у матеріальні активи	-5.073426e-01 ...	-0.309978
капітальні інвестиції у землю	-4.144909e-01 ...	-0.235974
капітальні інвестиції в існуючі будівлі	-2.100546e-01 ...	-0.120565
капітальні інвестиції у будівництво та перебудо...	-5.804938e-01 ...	-0.419055
капітальні інвестиції у машини та обладнання	-4.652043e-01 ...	-0.290535
капітальні інвестиції у нематеріальні активи	-5.572962e-01 ...	-0.345023
капітальні інвестиції у концесії, патенти, ліц...	-4.748549e-01 ...	-0.391680
капітальні інвестиції у придбання програмного з...	-4.471909e-01 ...	-0.313939
Обсяг реалізованої продукції за інституціональн...	-5.138771e-01 ...	-0.305870
Обсяг реалізованої продукції за функціональним ...	-5.085942e-01 ...	-0.300957
суб'єкти малого підприємництва, тис. грн	-3.605599e-01 ...	-0.128201
суб'єкти мікропідприємництва, тис. грн	4.907740e-01 ...	0.825297
к-ть зайнятих прац. суб'єкти малого підприємств...	4.989732e-01 ...	0.111168
к-ть зайнятих прац. суб'єкти мікропідприємств...	9.787975e-01 ...	0.681468
к-ть найманих прац. суб'єкти малого підприємств...	-7.952981e-02 ...	-0.217749
к-ть найманих прац. суб'єкти мікропідприємств...	9.482164e-01 ...	0.686865
Обсяг реалізованої продукції суб'єкти малого пі...	-2.308494e-01 ...	0.167916
Обсяг реалізованої продукції з них суб'єкти мік...	6.786429e-01 ...	1.000000

Рис. 2.8. Кореляційна матриця для моделі з цільовою функцією обсягу реалізованої продукції за функціональним підходом.

Джерело: дослідження автора, [14], Додаток А (Код А.1.)

Середній рівень кореляції за рис. 2.8. між обсягом реалізованої продукції та іншими фінансовими показниками, такими як капітальні інвестиції та витрати на виробництво, що означає, що все-таки контроль виробничого процесу важливіший за інтенсивність розширення, або покращення умов роботи на підприємствах.

Високі значення на рис. 2.9. VIF (>10) вказують на наявність мультиколінеарності, що може впливати на стійкість і інтерпретацію регресійних коефіцієнтів. Наприклад, VIF для капітальних інвестицій у матеріальні активи склав 732, що вказує на високу мультиколінеарність. У той ж час, достатньо низький VIF для капітальних інвестицій у землю, або ж патенти показує на стабільність, утворених раніше локальних торгових союзів та інших подібних відносин у підприємств даних КВЕДів.

Мультиколінеарність (VIF):

	feature	VIF
0	Код за КВЕД	3.972443e+02
1	Рік	3.365419e+02
2	Кількість діючих підприємств за однорідними ви...	2.513362e+02
3	Витрати на виробництво продукції	inf
4	матеріальні витрати та витрати на оплату послу...	inf
5	амортизація	inf
6	витрати на оплату праці	inf
7	відрахування на соціальні заходи	inf
8	інші витрати	inf
9	Капітальні інвестиції - усього	inf
10	капітальні інвестиції у матеріальні активи	inf
11	капітальні інвестиції у землю	2.327516e+01
12	капітальні інвестиції в існуючі будівлі	1.238497e+01
13	капітальні інвестиції у будівництво та перебуд...	1.186546e+02
14	капітальні інвестиції у машини та обладнання	7.502771e+02
15	капітальні інвестиції у нематеріальні активи	inf
16	капітальні інвестиції у концесії, патенти, лі...	2.127621e+01
17	капітальні інвестиції у придбання програмного ...	1.955761e+01
18	Обсяг реалізованої продукції за інституціональ...	3.452799e+04
19	Обсяг реалізованої продукції за функціональним...	4.165528e+04
20	суб'єкти малого підприємництва, тис. грн	2.620757e+02
21	суб'єкти мікропідприємництва, тис. грн	8.334844e+02
22	к-ть зайнятих прац. суб'єкти малого підприємни...	9.792149e+02
23	к-ть зайнятих прац. суб'єкти мікропідприємницт...	3.356372e+02
24	к-ть найманих прац. суб'єкти малого підприємни...	1.574907e+03
25	к-ть найманих прац. суб'єкти мікропідприємницт...	2.660195e+02
26	Обсяг реалізованої продукції суб'єкти малого п...	1.100294e+02
27	Обсяг реалізованої продукції з них суб'єкти мі...	6.285766e+02

Рис. 2.9. Мультиколінеарність для моделі з цільовою функцією обсягу реалізованої продукції за функціональним підходом.

Джерело: дослідження автора, [14], Додаток А (Код А.1.)

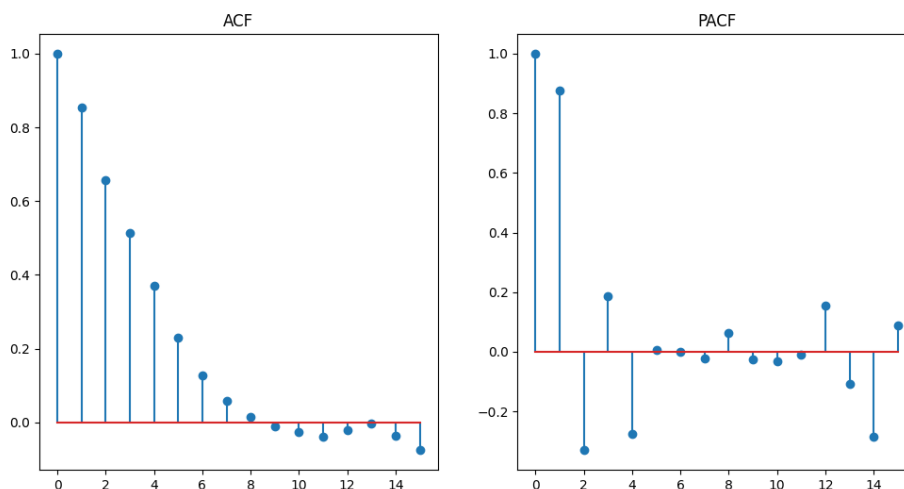


Рис. 2.10. Модель ARIMA для показників за КВЕДами.

Джерело: дослідження автора, [14], Додаток А (код А.1.)

Інтерпретація рис. 2.10. - графіків ACF та PACF часто використовуються для ідентифікації параметрів моделей ARIMA (Autoregressive Integrated Moving Average).

- AR(p) (авторегресійна частина): Параметр (p) визначається за кількістю значущих піків на графіку PACF, які виходять за межі довірчого інтервалу.
- MA(q) (частина ковзного середнього): Параметр (q) визначається за кількістю значущих піків на графіку ACF, які виходять за межі довірчого інтервалу.

Значущі піки на початкових лагах: Вказують на наявність короткострокових залежностей у даних. Швидке зменшення кореляції може свідчити про відсутність довгострокових трендів або циклів у ряді, отже прогнозування може бути циклічним. Підвищена кореляція на довгих лагах може вказувати на сезонні або трендові компоненти у ряді, які потребують додаткового моделювання.

Значущі кореляції на перших кількох лагах PACF можуть вказувати на те, що модель AR(p) з невеликим значенням (p) буде добре підходити для даних. Аналогічно, значущі кореляції на графіку ACF можуть допомогти визначити відповідне значення (q) для частини ковзного середнього.

Загалом, ці графіки є критичними інструментами для розуміння структури даних і вибору оптимальних моделей для їх аналізу та прогнозування.

Прогноз на 2023-2025 роки були виконані за допомогою моделі XGBoost, що мала визначену мету оцінити найвідповідніший період із від'ємним розрив у прирості за попередні роки значень. Таким виявився період з 2015 по 2018, який визначений, як трендовий для даного прогнозу, який введено, як ключовий фактор для покращення моделі.

КВЕД 8.01

- 2023: Приріст 15.09%
- 2024: Приріст 5.59%
- 2025: Приріст 0.23%

Бачимо, що тут спадний приріст доходів, що повинен зупинитися у 2025 році. Дійсно, у даній сфері є певне технічне обмеження з військовими діями, що

вимушує відкривати нові кар'єри для піску, адже багато було захоплено. Підвищення показників також реально за умов використання цих матеріалів активно у відновлювальному будівництві.

КВЕД 8.11

- 2023: Приріст 11.78%
- 2024: Приріст 1.43%
- 2025: Приріст -4.76%

КВЕД 23.7

- 2023: Приріст -6.03%
- 2024: Приріст -28.41%
- 2025: Приріст -37.48%

Надійність обох даних моделей була низькою, навіть за умови введення циклічності трендів, як ключового показника – ми не можемо зважати на цей прогноз, як на достовірні дані. У той ж час, зображені тренди є досить логічними. Так, в Україні є певний список експортерів, які залюбки купуватимуть українське декоративне каміння, зокрема, граніт. Більш того, після критичного зниження цін на продукцію, ми стаємо ще й надзвичайно вигідними партнерами, єдиною проблемою яких є надійність (як транспортна, так і безпекова, так і енергетична, адже підприємства відстають від графіків більше, ніж на 7% через відключення [8]). Також, його використання в Україні зростатиме із забудовою, що зустрічатиме все більше європейські стандарти у дорожніх роботах та організації благоустрою. Тому видобуток каменю можливо й зростатиме, а от попит на таку готову продукцію в Україні, скоріше за все, ні. Дешевизна не є сильною стороною таких виробництв, автоматизація ж практично ж неможлива, а кількість некваліфікованої робочої сили знижується з кожним етапом мобілізації, а кваліфікованої прямує до нуля. У той час, коли товари-субститути, такі як бетонна бруківка, мають вже достатню кількість виробництв, що мають найновішу технічну базу, гранітні виробництва потребують постійного підтримання та покращення, але попиту такого не отримують. Тому єдиними

шансами для підприємств КВЕДу 23.70 є виготовлення продукції могильного характеру, або ж використання жорстких маркетингових важелів задля залучення інвестиції, зміни ставлення на ринку до уже існуючої продукції та створення унікальної ніші.

Отже, підведемо короткі підсумки та перейдемо до власного аналізу автора:

1. Високий рівень пояснювальної здатності моделі був досягнутий для групи 8.01 (R^2 Score = 0.6545), що свідчить про адекватну модель для цієї групи.

2. Низька пояснювальна здатність моделей для груп 8.11 та 23.7 вказує на необхідність додаткового аналізу даних та використання інших алгоритмів для покращення результатів.

3. Висока мультиколінеарність деяких показників може впливати на стабільність моделей, що вимагає подальшого коригування або використання методів зниження мультиколінеарності.

4. Прогнози на 2023-2025 роки вказують на різні тренди для різних груп, зокрема надійний позитивний приріст для групи 8.01 та невизначені негативні тренди для групи 23.7, що потребує особливої уваги при плануванні діяльності підприємств цих груп.

РОЗДІЛ 3. АНАЛІЗ ДІЯЛЬНОСТІ ПІДПРИЄМСТВА «ЮНІКОМ-ПРОМ».

МОВНЕ МОДЕЛЮВАННЯ ПРОДАЖІВ ТА ІНТЕРПРЕТАЦІЯ

ВИСНОВКІВ ЩОДО ПОКРАЩЕННЯ ДІЯЛЬНОСТІ ПІДПРИЄМСТВА

3.1. Оцінка загальних статистичних показників бухгалтерської звітності

У даному розділі розглядаються дані надані ТОВ «ЮНІКОМ-ПРОМ», а зокрема, у цьому розділі проведено економетричний аналізу бухгалтерського статистичного звіту. Для аналізу взято усі показники бухгалтерської звітності за Кодами від 1000 до 2350. Результативною ознакою виступає чистий дохід від реалізації продукції (код рядка 2000), а факторними ознаками є загальні показники активів підприємства, зокрема, усі засоби виробництва, заборгованості, запаси, баланс, витрати, податки та інше.

Далі перейдемо до результатів регресійного аналізу:

Таблиця 3.1. Коефіцієнти моделі

Параметр	Коефіцієнт	Стандартна помилка	t-статистика	P-значення	Довірчий інтервал (95%)
const	-1.467e+07	7e+06	-2.097	0.069	-3.08e+07
x1	7308.6988	3467.930	2.108	0.068	-688.363

Джерело: дослідження автора на основі Додаток В (Код В.1.)

Показник R-squared (0.357) показує, що модель пояснює 35.7% варіації залежної змінної. Це відносно низьке значення, що вказує на те, що існують інші фактори, які впливають на залежну змінну, але не включені в модель.

Adjusted R-squared (0.277): скоригований R-squared враховує кількість предикторів у моделі та зменшує значення R-squared, якщо додавання нових предикторів не покращує модель значно. Значення 0.277 підтверджує, що модель має обмежену пояснювальну силу.

F-statistic (4.442) та Prob (F-statistic) (0.0681): F-статистика та її P-значення показують, що модель загалом має значний вплив на залежну змінну, але з рівнем значущості 0.0681, що трохи більше стандартного рівня значущості 0.05.

Log-Likelihood рівний -116.65, тобто логарифмічна ймовірність використовується для порівняння моделей. Негативне значення лог-імовірності є звичайним для моделей цього типу.

AIC - 237.3: це критерій Акаїке (AIC), використовується для оцінки моделей, де менші значення вказують на кращу модель. Значення 237.3 вказує на те, що ця модель має певну придатність, але може бути місце для покращень.

BIC - 237.9: це критерій Байєса (BIC), також використовується для оцінки моделей, де менші значення вказують на кращу модель. Значення 237.9, близьке до AIC, підтверджує загальну адекватність моделі.

Durbin-Watson (1.206): цей показник тестує наявність автокореляції залишків. Значення близьке до 2 свідчить про відсутність автокореляції. Значення 1.206 вказує на деяку позитивну автокореляцію. Коефіцієнти та їх P-значення: P-значення для коефіцієнтів вказують на те, що обидва коефіцієнти значущі на рівні 0.068, що є трохи вище стандартного рівня значущості 0.05.

Розглянемо графіки автокореляції та часткової автокореляції на рис. 3.1., де маємо, що графік ACF демонструє кореляції між значеннями часового ряду на різних лагах (відставаннях).

Автокореляція на нульовому лагу завжди дорівнює 1, оскільки це кореляція значення із самим собою. Автокореляція на першому лагу (близько 0.7) показує, що існує сильний позитивний зв'язок між поточними значеннями та значеннями попереднього періоду. Це свідчить про те, що значення у часовому ряді мають тенденцію залишатися схожими з періоду в період.

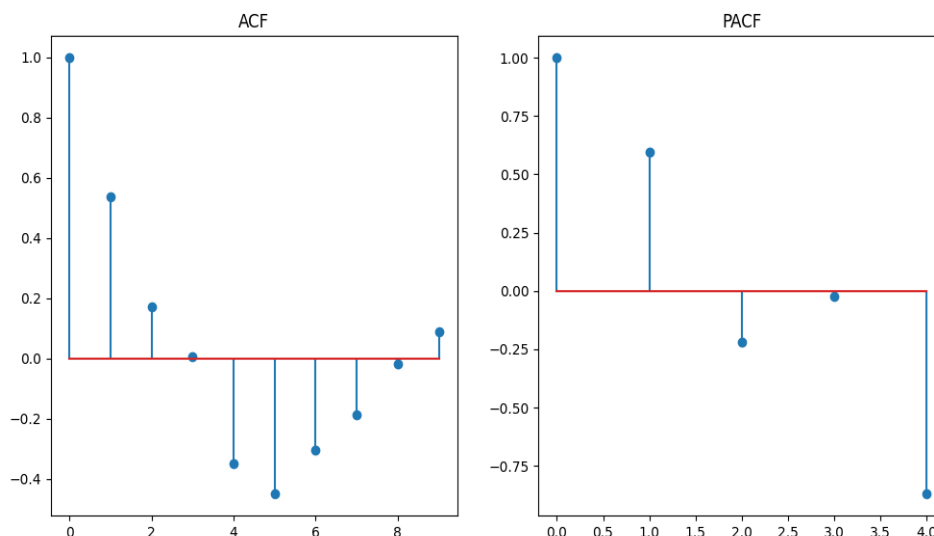


Рис. 3.1. Модель ARIMA показників «ЮНІКОМ-ПРОМ».

Джерело: дослідження автора, Додаток В (код В.1.)

Автокореляція на другому лагу (близько 0.4) все ще є позитивною, але зменшується, що вказує на зменшення залежності між поточними значеннями та значеннями двох попередніх періодів.

Lag 3 і більше помітно, що збільшенням лагу автокореляція стає близькою до нуля або навіть негативною. Це означає, що вплив попередніх періодів на поточні значення поступово зменшується і зрештою зникає.

Графік ACF демонструє, що значення у часовому ряді мають значні позитивні залежності на коротких лагах, але ці залежності швидко зникають зі збільшенням лагу.

Графік PACF показує кореляції між значеннями часового ряду на різних лагах, виключаючи вплив проміжних лагів.

Lag 1: Часткова автокореляція на першому лагу (близько 0.7) є високою і позитивною, що підтверджує сильну пряму залежність між поточними значеннями та значеннями попереднього періоду. Lag 2: Часткова автокореляція на другому лагу (близько 0.5) залишається позитивною, але зменшується, що вказує на помітну залежність між значеннями через один період. Lag 3 і більше:

На третьому та четвертому лагах часткова автокореляція різко зменшується і стає негативною на четвертому лагу. Це свідчить про те, що прямий вплив попередніх періодів зникає, коли враховуються проміжні періоди.

Графік PACF підтверджує, що короткострокові залежності є значними, але довгострокові залежності швидко зменшуються і зникають.

На основі аналізу графіків ACF та PACF для підприємства можна зробити наступні висновки:

1. . Сильна короткострокова залежність: Існує значна автокореляція на коротких лагах (до 2-3 періодів), що свідчить про те, що значення у часовому ряді мають тенденцію залишатися схожими у короткостроковій перспективі.

2. . Зменшення довгострокової залежності: Автокореляція швидко зменшується зі збільшенням лагу, що вказує на зменшення впливу попередніх періодів на поточні значення у довгостроковій перспективі.

3. . Прямі залежності: Графік PACF підтверджує існування прямої залежності на коротких лагах, але показує відсутність такої залежності на довгих лагах.

Топ 3 фактори за значенням кореляції

- Собівартість реалізованої продукції : 0.991977
- Накопичена амортизація: 0.951042
- Первісна вартість: 0.916469

Відсотковий вплив факторів

- Собівартість реалізованої продукції : 34.690721%
- Накопичена амортизація: 33.259171%
- Первісна вартість: 32.050107%

Ці показники відображають відсотковий вплив кожного фактору на цільову змінну. Вони свідчать про те, що всі три фактори мають значний і майже рівномірний вплив на цільову змінну. Як ми вже помітили у попередньому розділі, що саме дані показники є ключовими у визначенні трендів за КВЕДами, а отже й зараз, на реальних даних, вони мають відповідно ключовий вплив.

Ще одне підтвердження прогнозованості, плановості та циклічності підприємств у галузі.

Як результат аналізу даних підприємства «ЮНІКОМ-ПРОМ», можна виділити кілька опорних. По-перше, обмежена пояснювальна здатність моделі: R-squared та Adjusted R-squared вказують на те, що модель пояснює лише частину варіації залежної змінної. Це свідчить про необхідність додаткового дослідження та включення інших потенційно важливих факторів. По-друге, собівартість реалізованої продукції, накопичена амортизація та первісна вартість є ключовими факторами, які впливають на цільову змінну, і вони мають високу кореляцію та відсотковий вплив.

Також, наявна потенційна автокореляція, адже Durbin-Watson показує деяку позитивну автокореляцію, що може вимагати додаткового аналізу та, можливо, корекції моделі задля отримання надійних прогнозів. Наявна мультиколінеарність, тому що високі значення кореляції та відсоткового впливу можуть вказувати на наявність мультиколінеарності, що підтверджується великим значенням умови числа моделі ($1.42e+06$). Це може впливати на стабільність та інтерпретацію моделі.

Ці результати вказують на необхідність подальшого дослідження та вдосконалення моделі для досягнення більш точних та надійних прогнозів.

3.2. Використання NLP задля оцінки ефективності продажів та визначення тенденцій продажів підприємства.

У аналізі було використано датасет із усіма продажами (сумами продажу, кількість проданих одиниць та ціною продажу), що були здійснені на підприємстві протягом 2021, 2022 та 2023 років. Перші 1040 продажів були здійснені до початку військових дій в Україні, інші 700 протягом військових дій. Товари описані за видом продукції, назвою родовища, специфікацією товару та одиницею виміру ціни.

Для початку було проведено описову статистику для товарів, що закінчуються на "шт", "т", "компл", "пог.м.", "м2". Статистика у таблиці 3.2. включає середнє значення, моду, медіану та дисперсію цін для кожної категорії.

Таблиця 3.2. Описова статистика для показників за номенклатурною одиницею

	Середня ціна	Мода ціни	Медіана ціни	Дисперсія ціни
шт	2595,508002	1346,585	986,15	42972089,31
т	2660,35183	1000	1784,58	22535521,57
компл	13208,87	117,08	1380,29	270482910,5
м2	1338,692628	1600	1232,498596	484144,0846

Джерело: дослідження автора на основі Додаток В (код В.1.)

Бачимо, що найдорожчою є комплектна продукція, яка переважно є монументами, стелами, або могильними конструкціями, також у такі набори входять цільні сходи. У той ж час продукція поштучна, яка може бути складовою комплектної продукції має велике падіння у вартості, при тому ж маючи низьке медіанне значення, можемо визначити даний клас, як фактично дешеві, залишкові товари, адже вони продаються за заниженими цінами, й рідко за надвисокими.

Малим є відхилення цін за квадратний метр, що може означати надійність та визначеність ринку саме на такій номенклатурі, а також моду ціни 1600 грн/кв.м., як ключову ціну для проданої продукції. Також важливо відзначити, що номенклатурної одиниці тонн мода у 1000 гривень може означати лише або збут залишкових товарів, або ж великі продажі у довоєнний період, адже ціна виробництва будь-якої продукції є вищою за дану.

Графік трендів цін на рис. 3.2. для кожної категорії товарів був побудований та обмежений за віссю ціни до 20000, але без нормування значень. Цей графік показує загальний тренд зміни цін для кожної категорії товарів протягом періоду дослідження. Великий показник дисперсії означає не лише велику варіативність

цін, але у більшій мірі показує, що товари у т, шт, або комплектах можуть бути різного рівня складності. Так як переважно у тоннах продається бруківка, чи вихідна порода каменю у слябах, то складна, пиленна та колото-тесана продукція може коштувати значно вище. Рівномірність цін варто відзначити ще раз для кв.м., адже саме ця міра є ключовою, а також, як і тонни, має тренд на ріст із часом.

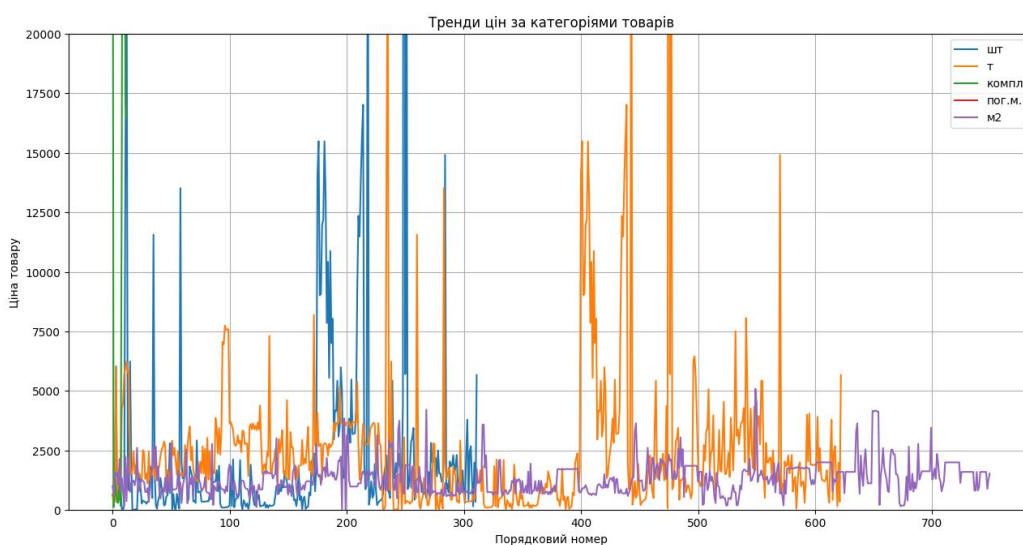


Рис. 3.2. Тренди цін для номенклатурних одиниць товарів.

Джерело: дослідження автора на основі Додаток В (код В.1.)

Далі з використанням бібліотеки nltk було виділено ключові слова з назв товарів. Важливою частиною була смислова класифікація слів, а саме визначення ключових послідовних зв'язків від товару до виду товару.

По-перше, повинні бути виділені окремо: вид продукції, родовище, габарити, номенклатурна міра. Для кожного ключового слова була розрахована статистика, включаючи середню суму покупки, середню кількість проданих товарів, середню ціну товару, моду ціни, медіану ціни та дисперсію ціни.

Наприклад, для ключового слова "губенківського": Середня сума покупки 247416.47 грн, Середня кількість проданих товарів 24.74, Середня ціна товару 980.0 грн, Мода ціни 985.0 грн, Медіана ціни 900.0 грн, Дисперсія ціни 4712929.0

Перехресне комбінування слів

Були створені всі можливі комбінації слів (пари) та розраховані середні значення, мода, медіана та дисперсія цін для кожної комбінації. Топ 50 найпопулярніших комбінацій слів були визначені на основі середньої суми покупки.

Наприклад, комбінація "корнинського бордюру": Середня сума покупки 1927271.0 грн, Середня кількість проданих товарів 19.27, Середня ціна товару 980.0 грн, Мода ціни 985.0 грн, Медіана ціни 900.0 грн, Дисперсія ціни 79930.0

Внаслідок, відповідного аналізу, можливо було виділити такі тенденції:

- Нижча ціна ціна продукцію провокує більші об'єми продажів.
- Масово продається пиленна, а не колота продукція, хоча вона й дорожча.
- Ціна на бордюру - визначає загальну середню ціну по виду каменю.
- Лабрадорит купують найбільше, на найбільші суми, але не завжди дешево.
- Слово «не» та «риф» є дорогим у багатьох комбінаціях.

Кластеризація слів і комбінацій слів

Були проведені кластеризації для ключових слів та комбінацій слів з використанням методу KMeans. Оптимальна кількість кластерів була визначена на основі значення Silhouette Score.

Кластеризація для ключових слів: Кількість кластерів = 2, Silhouette Score = 0.84

Кластеризація для комбінацій слів: Кількість кластерів = 2, Silhouette Score = 0.87

Внаслідок, кластеризації можна помітити, що надійні кластери утворені практично одними смисловими групами, а саме – до кластеру «1» віднесено товари із вищою ціноюю варіативність і туди потрапили переважно види товарів та родовищ, інколи види обробки, або методи застосування (прикметники).

Дисперсія при цьому вища для кластеру «0», до того ж, загалом, різниця у кластеризації комбінацій та слів окремо у тому, що ключові слова мають меншу дисперсія самі по собі, а самі ж кластери практично ідентичні у обох випадках.

Аналіз даних показав, що категорії товарів мають різні середні ціни, моди, медіани та дисперсії, що дозволяє зробити висновки про різноманітність цінових політик у різних товарних групах. NLP аналіз дозволив виділити ключові слова, що найчастіше зустрічаються в назвах товарів, та визначити їх вплив на показники продажів. Кластеризація слів і комбінацій слів дозволила виявити групи товарів, які мають схожі характеристики, що може бути корисним для оптимізації маркетингових стратегій.

Також, було визначено середні ціни у різних часових проміжках для різних слів, що дає яскравіше побачити трендовість установавання ціни моди 1600 гривень за квадратний метр. Як, бачимо, навіть на рисунку 4.3. - велика кількість слів мають широкий діапазон значень перший 10 періодів, а останні 5 ж мають загальний тренд на формування навколо значень 1400-1700, особливо, якщо частота появи цього слова висока.

	василівсько	кап	абрадорит	мм
0	688,4597			667,0202
1	1051,559			1605,755
2	1666,023			2516,671
3	2832,737			2151,023
4	2753,815	833,3325	815,8333	1545,458
5				1716,015
6				1160,077
7	2510,09			758,7377
8	681,4331			1000,792
9	6593,703			918,019
10	11731,75		29364	2268,42
11	1429,779			2496,573
12	772,104			1644,353
13			529,1569	1649,468
14				1475,946
15				1621,318

Рис. 3.3. Тренд у ціні слів для назви товару

Джерело: дослідження автора на основі коду Додаток В (код В.1.)

Після цього, отримавши загальні результати на рис. 3.3., варто оцінити, які комбінації товарів є найвигіднішими, та найчастіше продавались підприємством

у аналізований період часу. За допомогою коду з Додатку Г, було отримана описова статистика для текстових комбінацій усіх видів унікальної продукції за типом, а не специфікацією. Звичайно, отримані результати не враховують методи обробки та габаритні особливості продукції, але типологічна відповідність одного родовища іншому в кінцевому наборі товарів, що виготовляє підприємство дає можливість оцінити відносні значення популярних одиниць продукції.

Загалом було визначено більше 10 видів товарів, але саме бруківка та бордюр були найпоширенішими, як рівномірно по всьому датасету, так і кожен окремий продаж мав більшу кількість одиниць продукції.

3.3. Аналіз розподілу цін за родовищами та видами продукції граніту. Оцінка ефективності прогнозування ціни Random Forest.

Тут ми перейдемо вже до порівняння родовищ, адже, як ми бачимо на рисунку 4.3., різниця показників вагома, особливо виділяються високими цінами рідкісні та зараз недоступні родовища: Берестовецьке, Губенківське, Анівське. Ключовими родовищами є виділені зеленим та червоним серед бруківки родовища.

Особливо, можна відзначити продажі Покостівського граніту, родовищем якого володіє підприємство, і ціни на продукцію з якого є найстабільнішими серед усіх інших.

До того ж, виготовлення бордюрів саме з покостівського варто визначити, як найраціональніше, адже інші родовища мають, або значно нижчі середні ціни, або ж їх об'єми у середньому є значно нижчими, що може бути, взагалі, не вигідно підприємству, зважаючи на вартість підвозу каменю та дисконти на продукцію в залишках.

бордюр	лабрадорит	88361,10462	155,84	445,9520185	45	350	78684,48137
	габро	59391,52711	107,4657778	941,7154198	1346,585	751,5817143	533903,5804
	покоствівського	243287,9343	479,5032432	660,1843462	197	562,5	196549,6981
	корецького	166054,5775	501,0625	359,0199624	306,117259	355,4612953	2522,398011
	невирівського	373979,4144	602,4222222	536,0520287	128,0357143	600	70992,22092
	букінського	65546,05	188,3027273	545,2838377	280,2825091	506,007	78056,31789
	губенківського	13998,865	18,2	734,7889423	654,5701923	734,7889423	12870,0957
	василівського	58830,595	68,25	700,7879077	423,9871667	583,88625	137343,0346
	лезниківського	3864,875	5,86	667,2549834	384,5099668	667,2549834	159889,4888
	корнинського	1927270,97	3160,035	555,4808225	355,5678783	555,4808225	79930,37056
	маславського	31248,61	68,4	599,1833544	431,412	629,1426087	24018,47961
	осниківського	77585,396	181,058	467,0757278	247,0607285	549,8224981	30914,64241
	лабрадориту	869930,5667	1444,666667	529,1568856	380	600	16699,78507
правого	2755,8	1,5	2239,0875	1033,425	2239,0875	2907244,128	
бруківка	лабрадорит	99124,25545	41,02909091	2622,216128	550,52	2770,883961	2104265,844
	габро	179703,8884	103,1156178	2064,303547	1000	1756,928594	1555803,197
	покоствівського	212541,4814	122,4362329	2073,504114	735	1750	2648598,007
	корецького	169492,7314	112,8466327	2150,298793	1500	1644,25	1560572,928
	анівського	127790,455	47,0575	3028,336501	850	2946,558034	1732527,348
	невирівського	113253,1614	85,92019048	1518,738582	737,0803429	1208,844694	818840,6948
	букінського	175833,0691	98,91826087	2391,680663	705,8007494	2285,435523	1307584,973
	губенківського	364940,9525	61,44	4782,188755	2333,663905	4643,769308	8044791,261
	бистрівського	62228,61846	34,40307692	1944,015446	771,4768182	1533,919355	1301291,149
	василівського	129861,4693	85,81242857	1796,156913	500	1496,982411	1387414,768
	лезниківського	32342,79188	32,0525	1283,845837	450	949,4585	720339,6734
	корнинського	260815,6193	143,0314286	1473,134824	190	1001,982514	1859297,561
	капустинського	20077,93333	15,33333333	1055,273241	866,705	866,705	106673,9442
	осниківського	21875,54	33,5225	1162,024543	550,52	680,338	1106906,253
	юліївського	51921,985	41,25	1418,237452	1172,244559	1418,237452	121025,0068
	берестовецького	129012,0758	32,54166667	4019,764916	1057,033125	4383,998876	4346964,486
	писарівського	183079,3671	93,32117647	1806,724562	915,15	1542,115778	575862,6309
	дерибасівського	144358,6333	91,30333333	1628,67003	838,2975	1592,467804	273942,2781
	сліпчицького	96897,97571	50,81285714	2374,994029	544,6713507	2572,086957	1356503,05
	дідковицького	2108,736667	1,781666667	2126,029704	500	2649,825	1738496,374
човнівського	2365,245	5,75	2031,125	400	2031,125	5321137,531	
іванівського	218301,2867	148,0033333	1987,152205	1291,366452	1306,844043	1420285,733	

Рис. 3.4. Середні показники для найбільш поширених видів продукції

Джерело: розрахунки автора на основі коду Додаток Г (код Г.1.)

Тепер, варто подивитись на рис. 3.4., а саме на червоні комірки серед бруківки, ми помітимо, що практично усі ці родовища є родовищами червоного каменю, але вони не є дешевими де-факто, тому можливо припустити, що продажі були здійснені давно, або за дисконтними цінами (адже колота продукція для цих родовищ не є пріоритетною, а лише це могло б знизити вартість).Тоді, як інший червоний камінь – Корецьке має чудові показники саме серед бруківки, тим паче, має високі показники моди та медіани.

Серед темного каменю можна відзначити габро та Букинське (одне з родовищ габро), що продається за стабільними цінами вище середнього у різноманітних видах продукції, хоча й не великих об'ємах. Лабрадорит ж у той час (Осниківське, Невирівське родовища включно) не має великих об'ємів саме у бруківці, але великі продажі бордюрів, хоча й по низьких цінах (хоча відносно сировини вони є адекватними), може вказувати, що бордюри є ключовою

продукцією для граніту з цього родовища, на що варто розраховувати, при складанні виробничого плану.

		Середня кількість проданих товарів
бордюр	корнинського	3160.035000
	лабрадориту	1444.666667
вироби	натурального	386.134500
бордюр	невирівського	602.422222
бруківка	губенківського	61.440000
	корнинського	143.031429
бордюр	покостівського	479.503243
бруківка	іванівського	148.003333
	покостівського	122.436233
	писарівського	93.321176
	габро	103.115618
	букінського	98.918261
	корецького	112.846633
бордюр	корецького	501.062500
бруківка	дерibasівського	91.303333
	василівського	85.812429
	берестовецького	32.541667
	анівського	47.057500
	невирівського	85.920190
	лабрадорит	41.029091
	сліпчицького	50.812857
бордюр	лабрадорит	155.840000
	осниківського	181.058000
	букінського	188.302727
бруківка	бистріївського	34.403077
бордюр	габро	107.465778
	василівського	68.250000
бруківка	юлівського	41.250000
блок	невирівського	3.706667
	покостівського	29.354000
сходи	габро	22.166667
блок	лабрадорит	3.030000

Рис. 3.5. Середня кількість проданих товарів для комбінації виріб-родовище
Джерело: дослідження автора на основі коду Додаток Г (код Г.1.)

На рис. 3.5. можна помітити, що найбільша кількість за профільними товарами була продана саме бордюрів. Шкала одиниць була переведена у уніфіковані одиниці для бордюру – це пог.м., бруківки - кв.м., або ж тонни, а тоді обчислена відносна кількість одиниць для продажів. Однозначно найбільша кількість продано бордюрів Корнинського граніту, хоча продажів загалом було багато, але для лабрадориту, корецького бордюрів загальна кількість продажів більша. Можливо очікувати вищу точність прогнозів для даних товарів. Об'єми партій бруківки переважно коливаються до об'єму одного вантажного

перевезення (близько 150 кв.м.). Покоствівське має загальні найбільші продажі на рівні з габро, а об'єми є середніми для усієї категорії.

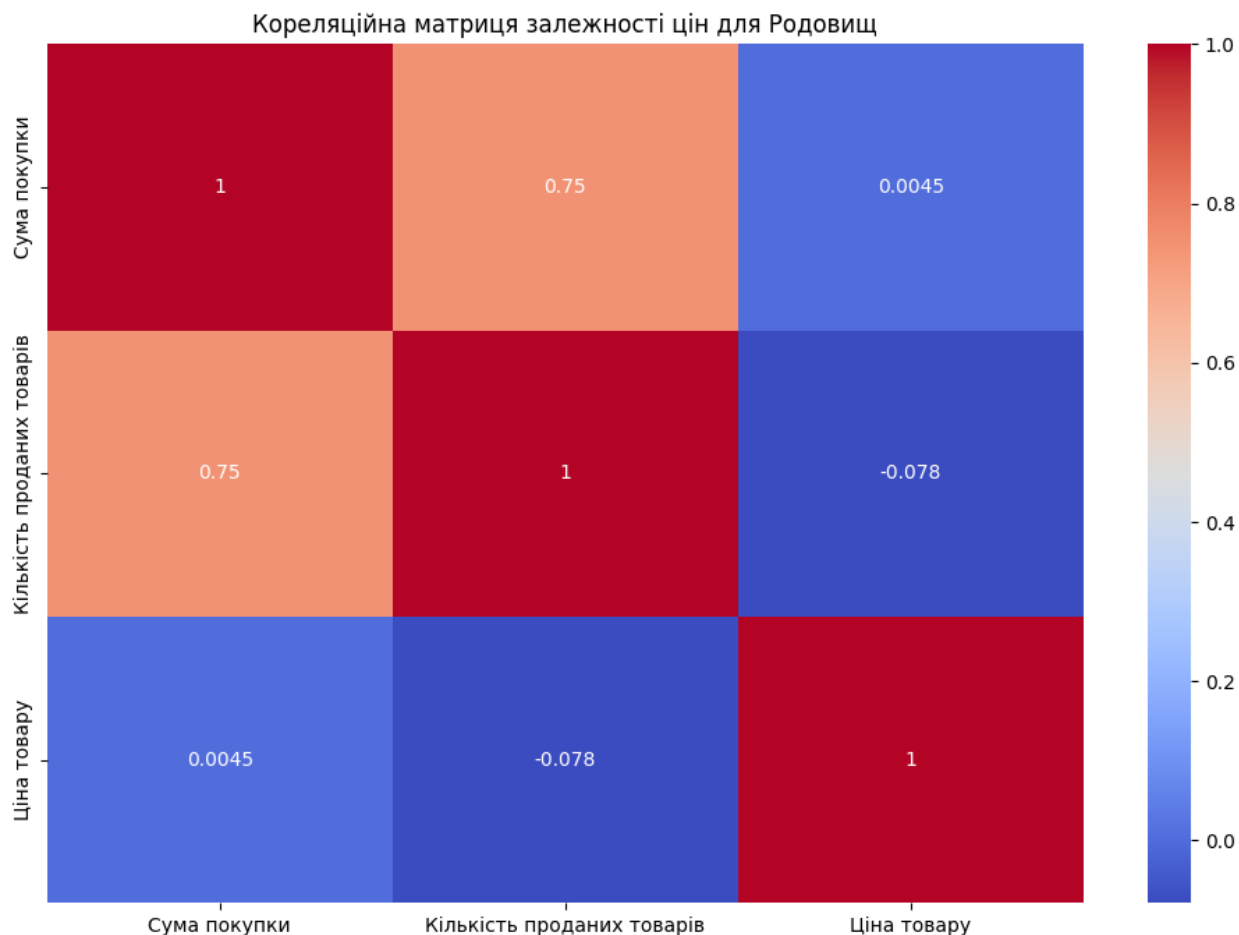


Рис. 3.6. Кореляційна матриця залежності цін товарів за родовищами

Джерело: дослідження автора на основі Додаток Г (код Г.1.)

За рис. 3.6. сума покупок і кількість проданих товарів має кореляційний коефіцієнт між цими двома показниками становить 0.75. Це свідчить про сильний позитивний зв'язок. Тобто, збільшення кількості проданих товарів веде до зростання суми покупок. Цей висновок є логічним і очікуваним, оскільки з більшим обсягом продажів зростає й загальна сума виручки. Сума покупок і ціна товару має кореляцію дуже слабку, близько 0.0045 - це може свідчити про те, що загальна сума продажів не сильно залежить від ціни окремого товару. Іншими словами, незалежно від зміни ціни, загальна сума покупок залишається приблизно на одному рівні, що може свідчити про стабільний попит на

продукцію. Для кількості проданих товарів і ціни товару присутня від'ємна кореляція (-0.078) між кількістю проданих товарів і ціною вказує на слабку зворотну залежність. Це може свідчити про те, що при зростанні ціни кількість проданих товарів незначно зменшується. Хоча цей показник є слабким, він все ж вказує на те, що ціна може впливати на обсяги продажів, але не критично.

Результати кореляційного аналізу дозволяють оцінити вплив цінової політики на обсяги продажів і виручку. Слабка залежність суми покупок від ціни свідчить про те, що компанія має певну цінову еластичність попиту, що може використовуватися для збільшення маржі. Водночас, сильна залежність між кількістю проданих товарів і сумою покупок підкреслює важливість обсягу продажів у формуванні доходів.

Далі стає логічним оцінити наскільки ж буде ефективно виконано прогноз моделлю Random Forest, яка показала найкращі показники серед інших та виділилась за критерієм приросту інформації, що дало можливість позбутися варіації між прогнозами та значеннями в межах відхилення 1000 грн у ціні для товарів, адже для таких товарів була більша кількість спостережень і для них повторне навчання було ефективним, хоча й не виникало значного покращення моделі.

Модель Random Forest показала значний рівень точності. Значення R^2 , що дорівнює 0.7182, вказує на те, що модель здатна пояснити приблизно 71.82% варіацій у цінових даних. Цей показник дає сподівання, що в умовах плановості обсягу виробленої продукції ринку, яку ми розглянули у розділах 2.2 та 2.3, збут даної продукції можливо прогнозувати в межах ризиків бюджету, адже високий показник для економетричної моделі, особливо враховуючи складність та різноманіття факторів, що впливають на ціноутворення в галузі декоративного каменю.

		Середня ціна товару	Прогнозована ціна
бордюр	корнинського	555.488823	341.745302
	лабрадориту	529.156886	475.508052
вироби	натурального	1110.999924	2151.271699
бордюр	невирівського	536.052029	438.315893
бруківка	губенківського	4782.188755	776.637085
	корнинського	1473.134824	363.273013
бордюр	покоствівського	660.184346	712.416354
бруківка	іванівського	1987.152205	583.194138
	покоствівського	2073.504114	745.083060
	писарівського	1806.724562	745.083060
	габро	2064.303547	734.351212
	букінського	2391.680663	552.181277
	корецького	2150.298793	445.948025
бордюр	корецького	359.019962	340.821356
бруківка	дерибасівського	1628.670030	773.835981
	василівського	1796.156913	594.510316
	берестовецького	4019.764916	475.223108
	анівського	3028.336501	583.194138
	невирівського	1518.738582	513.745367
	лабрадорит	2622.216128	506.294048
	сліпчицького	2374.994029	745.083060
бордюр	лабрадорит	445.952019	475.508052
	осниківського	467.075728	712.416354
	букінського	545.283838	576.109995
бруківка	бистрівського	1944.015446	601.455861
бордюр	габро	941.715420	701.862699
	василівського	700.787908	638.303602
бруківка	юлівського	1418.237452	689.952715

Рис. 3.7. Прогнозування моделі Random Forest для товару та родовища
Джерело: дослідження автора на основі Додаток Г (код Г.1.)

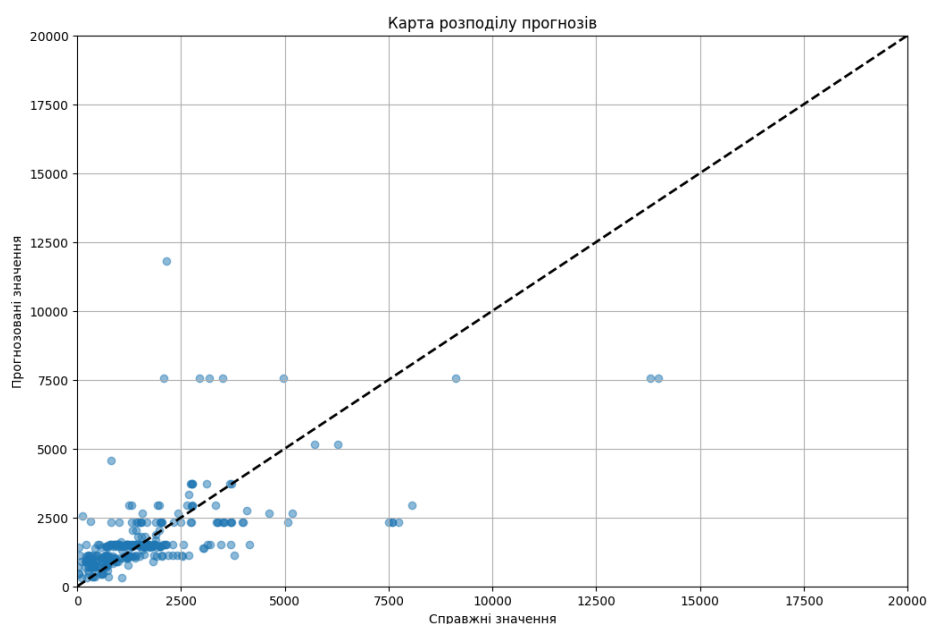


Рис. 3.8. Розподіл прогнозів Random Forest

Джерело: дослідження автора на основі Додаток Г (код Г.1.)

Аналіз рис. 3.8. графіку розподілу прогнозів демонструє, що більшість прогнозів сконцентрована у діапазоні нижче 2500 грн. Це вказує на ефективність моделі в прогнозуванні цін для товарів з нижчими цінами, що ми й можемо побачити у результатах на рис. 3.7. Проте існують окремі випадки, де модель суттєво недооцінює або переоцінює ціни. Це може бути обумовлено обмеженою кількістю тренувальних даних для цих товарів або специфічними

характеристиками самих товарів. Наприклад, товари з унікальними характеристиками або ті, що виготовляються на замовлення, можуть мати значні відхилення у ціні.

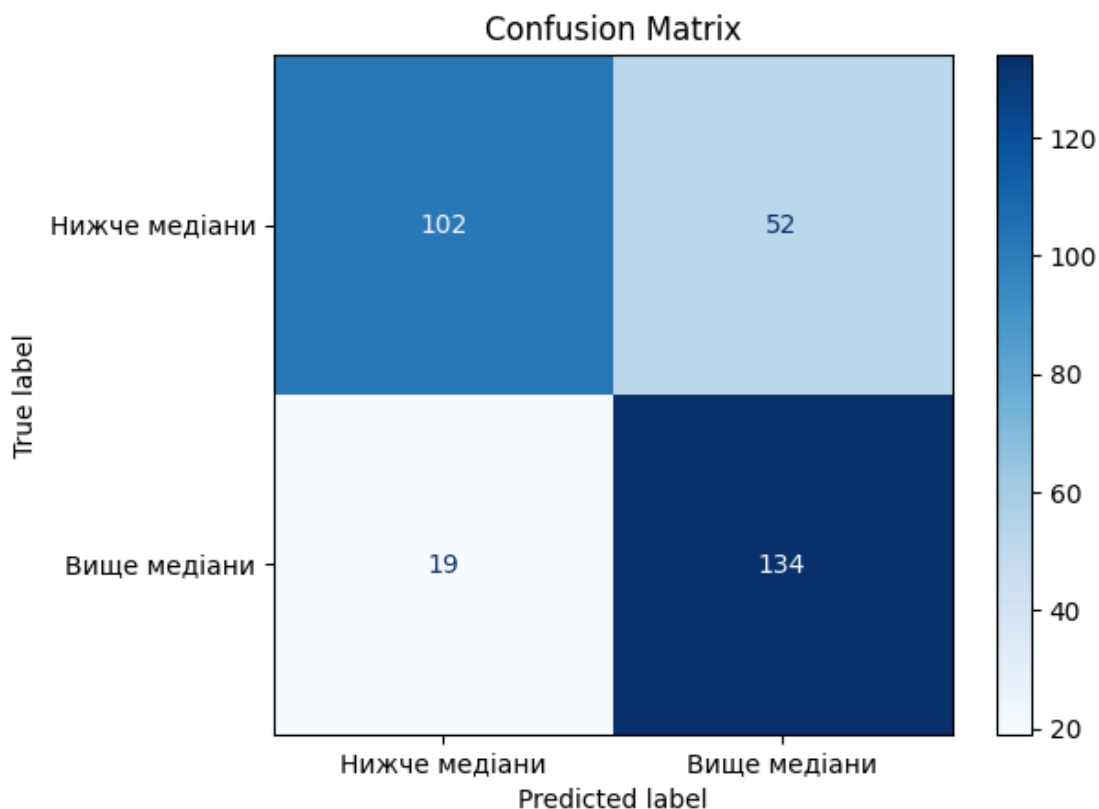


Рис. 3.9. Матриця сподівань (Confusion Matrix)

Джерело: дослідження автора на основі Додаток Г (код Г.1.)

Матриця на рис. 3.9. сплутування дає нам чітке уявлення про точність моделі у класифікації цін товарів, адже модель демонструє вищу точність у прогнозуванні цін для товарів, що є вище медіанного значення. Це може бути пов'язано з тим, що товари з вищими цінами зазвичай мають більш стабільні характеристики, адже відбувається завчасне планування постачань та виготовлення продукції, а також каталог таких продуктів у вибірці є обмеженим і це може бути одним з наслідків перенавчання. Модель прогнозів загалом є надійною, але присутнє завищення значень, що можливо виправити із більшим категоризованим об'ємом даних.

До того ж товари, що продавались регулярно протягом усього терміну, як бордюр з лабрадориту та корецького граніту, або ж бруківка Покостівська та габро, ми можемо побачити на рис. 3.7. Наприклад, камені з родовищ "лабрадорит" або "габро" мають стабільніші ціни, що сприяє більш точним прогнозам.

Точні здогадки моделі в основному сконцентровані на товарах з стабільними цінами та чітко вираженими характеристиками. Наприклад, вироби з власного родовища не є найбільш стабільними у цінах, що вимагає покращення політики управління ресурсами виробництва.

Найбільші неточності моделі спостерігаються для товарів, де ціни значно варіюються в залежності від додаткових характеристик, таких як специфікація виробу або обсяг замовлення. Наприклад, різні види обробки поверхні (термообробка, пилена, колота) можуть значно впливати на ціну товару, що ускладнює точне прогнозування.

Загалом, модель Random Forest демонструє високий потенціал для прогнозування цін товарів на ринку декоративного каменю, проте подальше вдосконалення може значно покращити її ефективність та точність.

ВИСНОВКИ

Хоча й велика частина виробничого потенціалу не може бути використана, а попит на продукцію низький, що не дозволяють сподіватися на формування безризикових умов для інвестицій у ринок промисловості загалом, ринок декоративного каменю через унікальність використовуваних ресурсів має обґрунтовані надії та шляхи розвитку. Наприклад, шляхом збільшення видобутку каменю та збільшення експорту дешевої продукції, адже питання нестачі робочих місць переходить у нестачу робітників, а тому варто інвестувати саме у сферу з більшою доданою вартістю.

Плановість української економіки знаходиться на критичному рівні, і хоча у довоєнний період такий метод виробництва давав результати, то дворічний лаг у промисловості вже зник, і зниження ціни продажу, а тобто і прибутків підприємств одразу означатиме зниження виробничого потенціалу.

Підприємства низького рівня технологій страждають у меншій мірі саме через непотребу формування освіченої групи робітників, хоча й теж не мають достатню базу розвитку. Потрібні інновації, а бажано автоматизація, адже процеси такого характеру, якщо й не можуть бути виконані без людей, то повинні мінімізувати їх вплив на вироблену продукцію. Можливо очікувати, що до кінця 2024 року ситуація не стане критичною у даній сфері, але спадний тренд нового характеру почався, адже уже почалась боротьба не з конкурентами, а за сферу.

Для досліджуваного підприємства ж критично важливим є перепрофілізація на певні товари, обмежені за своїм відображенням, адже несистемність цін призводила до меншої інформативності у результатах продажів, а також до продажу товарів за зниженими цінами, що останнім часом зменшилось, адже мода ціни періоду 2023 року є модою періоду всього дослідження. Продаж товарів комплектного порядку також є важливим для підприємства, тому потрібно формувати більш системні ціни на таку продукцію, хоча й не виготовляти подібних виробів завчасно, адже попит низький.

У той ж час, важливо правильно розподіляти ресурси, тому камінь лабрадорит (зокрема Невирівського та Осниківського родовищ) варто використовувати для виготовлення бордюрів для складського розміщення (що може стати у нагоді при відновленні транспортного покриття мостів і т.д.), адже об'єми замовлень на ці товари були великими, а тому пришвидшення видачі подібних товарів звільнить нагальні потужності підприємства. При цьому, Корецький граніт не варто використовувати для бордюрів, адже його відносна ціна у інших видах виробів є вищою, особливо, якщо брати бруківку та плитку, а ще він має найменше коливання у цінах. А от сировина Покостівського родовища є найціннішою для підприємства, адже є найстабільнішою, як за цінами, об'ємами, так і за видами продукції. Наявність габро на виробництві також дуже важлива, адже не маючи особливої цільової продукції, габро було продано найбільшу кількість разів (але при цьому дисперсія цін була менша за Покостівське родовище, хоча це може означати меншу кількість оптових знижок). Загальна середня ціна продажу продукції повинна становити 1600-1750 грн/кв.м., особливо для габро та Покостівського граніту.

Окрім цього, ефективним буде орієнтація на видобувну та експортну діяльність підприємства, а також залучення активів у сферу наукоємної діяльності задля отримання контролю над доданою вартістю товарів, а також сферою послуг, що супроводжує добувну та виробничу діяльність.

Внаслідок дослідження було ефективно описано структуру усього ринку промисловості, окреслено основні тенденції та особливості діяльності саме на ринку декоративного каменю. Побудовано економетричні моделі, виконано кореляційний та статистичний аналіз показників капітальних інвестицій, витрат підприємств та обсягів реалізованої та виробленої продукції задля прогнозування подальшого розвитку ринку. Також, оцінено діяльність підприємства «ЮНІКОМ-ПРОМ» та виконано NLP-аналіз бази даних продажів та зроблено прогноз Random Forest для ціни категоріального товару.

Дане дослідження у подальшому можливо розвивати у таких напрямках:

- Збір додаткових даних: збільшення обсягу тренувальних даних дозволить моделі краще навчитися і зменшити похибки прогнозів.
- Врахування додаткових змінних: включення додаткових характеристик товарів (наприклад, особливості замовлення, стану ринку, специфікації товару, сезонні фактори) може значно підвищити точність прогнозів.
- Регулярне оновлення моделі: модель повинна періодично оновлюватися на основі нових даних для підтримання її актуальності та точності, формування змінних бінарних характеристик товару задля підвищення деталізованості аналізу, а також надійності прогнозів.
- Оптимізація параметрів моделі: використання методів гіперпараметричної оптимізації, таких як Grid Search або Random Search, для налаштування параметрів моделі може допомогти зменшити MSE.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Крижановська Н. Я. Світло-кольоровий дизайн : конспект лекцій для студентів 5 курсу за спеціальністю 191 – Архітектура та містобудування, освітня програма підготовки магістрів «Дизайн архітектурного середовища» / Н. Я. Крижановська, О. В. Смірнова ; Харків. нац. ун-т міськ. госп-ва ім. О. М. Бекетова. – Харків : ХНУМГ ім. О. М. Бекетова, 2020
2. Мармоза А.Т. Теорія статистики А.Т. Мармоза, 2-ге вид. перероб. та доп. К.: "Центр учбової літератури", 2013
3. Аналіз сучасного стану каменедобувної та каменеобробної галузі України. URL: <http://ten.ztu.edu.ua/article/view/269366/264846>
4. Wastes production in dimension stones industry: resources, factors and solutions to reduce them. URL: https://www.researchgate.net/publication/353943463_Wastes_production_in_dimension_stones_industry_resources_factors_and_solutions_to_reduce_them
5. Аналітична стаття «АНАЛІЗ СПОСОБІВ ПОЛІПШЕННЯ ТЕПЛОТЕХНІЧНИХ ХАРАКТЕРИСТИК ПРИ БУДІВНИЦТВІ БУДІВЕЛЬ». URL: <https://elartu.tntu.edu.ua/bitstream/lib/32538/1/Text%20statti%20Ihnatieva-Tekin-Logos%205-10-20191105.pdf>
6. Аналітична стаття «Будівельний ринок у 2023 році може зрости на 25%». URL: <https://gmk.center/ua/opinion/budivelnij-rinok-u-2023-roci-mozhe-zrosti-na-25/>
7. Аналітична стаття «June Damages Ukraine Report». URL: https://kse.ua/wp-content/uploads/2023/09/June_Damages_ENG_-Report.pdf
8. Ринок будівельних сумішей. URL: <https://interfax.com.ua/news/press-conference/980842.html>
9. Сучасні тенденції розвитку ринку будівельних матеріалів в Україні. URL: <https://stroyobzor.ua/ru/news-ukraine/ekspert-rozpoviv-pro-tendencii-budivelnogo-rinku-ukraini>

10. Аналіз ринку керамічної плитки. URL: <https://www.sworldjournal.com/index.php/swj/article/download/swj11-04-031/2390>
11. Аналіз ринку щебеню в Україні у 2023 році. URL: <https://www.impulse-consulting.com.ua/kontakti/>
12. Вплив інвестицій на розвиток будівельної галузі України. URL: https://ev.nmu.org.ua/docs/2022/2/EV20222_078-085.pdf
13. Сучасні проблеми та перспективи розвитку будівельної галузі України. URL: <https://www.ukrinform.ua/rubric-presshall/3758174-problemi-ta-vikliki-budivelnoi-galuzi-ukraini.html>
14. Державна служба статистики України. Економічна статистика / Економічна діяльність / Промисловість URL: <https://www.ukrstat.gov.ua/>
15. Геєць В.М., Клебанова Т.С., Черняк О.І., Іванов В.В., Дубровіна Н.А., Ставицький А.В. «Моделі і методи соціально-економічного прогнозування, 2-ге вид., виправ.»
16. Статистика – Опря А. Т. – 6.2.4. Розподіл Хі – квадрат - Статистика - Опря А. Т | Підручники. Підручники для школярів та студентів. URL: <https://subj.ukr-lit.com/statistika-oprya-a-t-6-2-4-rozpodil-xi-kvadrat/>
17. Єріна А.М., Пальян З.О. Теорія статистики [Текст]: [практикум] / А.М. Єріна, З.О. Пальян – 6-те вид., стер. – К.: Знання, 2008. – 255 с.].
18. Ковтун Н.В. Теорія статистики [Текст]: [підручник] / Н.В. Ковтун, Е.В. Галицька. – К.: Видавничо-поліграфічний центр "Київський університет", 2008. – 336 с.
19. Ставицький А.В., Харламова Г.О. «Моделювання мікроекономічних процесів» Навч. посібник. – Київ, Аграр Медіа Груп, 2018 – 182 с.

20. ГЕОГРАФІЯ ТА ТУРИЗМ. Матеріали VII Всеукраїнської науково-практичної Інтернет-конференції Харківського національного педагогічного університету імені Г.С. Сковороди (28 лютого 2024 р., м. Харків). URL: <http://repositsc.nuczu.edu.ua/bitstream/123456789/19965/1/%D0%93%D0%95%D0%9E%D0%93%D0%A0%D0%90%D0%A4%D0%86%D0%AF%20%D0%A2%D0%90%20%D0%A2%D0%A3%D0%A0%D0%98%D0%97%D0%9C-2024.pdf#page=390>

21. Гориславець Р.О. Аналіз сучасних підходів до конструювання деформаційних швів у транспортному будівництві / Р.О. Гориславець, І.А. Єрмакова // Тези 76-ї наукової конференції професорів, викладачів, наукових працівників, аспірантів та студентів університету (Полтава, 14 травня – 23 травня 2024 р.). – Полтава : Нац. ун-т ім. Юрія Кондратюка, 2024. – Т. 1. – С. 53–54. URL: <https://reposit.nupp.edu.ua/handle/PoltNTU/15645>

22. Золотарьова Оксана, Мережко Ніна, Шульга Ольга «Ринок будівельного каменю в Україні», 2018 URL: [https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwiSsveZwbWGAxW1VPEDHcfbDioQFnoECCAQAQ&url=http%3A%2F%2Ftr.knute.edu.ua%2Ffiles%2F2018%2F01\(25\)%2F07.pdf&usg=AOvVaw290LANZ_o6PjTehtrWiNuD&opi=89978449](https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwiSsveZwbWGAxW1VPEDHcfbDioQFnoECCAQAQ&url=http%3A%2F%2Ftr.knute.edu.ua%2Ffiles%2F2018%2F01(25)%2F07.pdf&usg=AOvVaw290LANZ_o6PjTehtrWiNuD&opi=89978449)

23. Дослідження впливу машинного навчання на прогнозування економічних показників в Україні. URL: http://visnyk-onu.od.ua/journal/2020_25_3/23.pdf

24. Аналіз ефективності використання методів машинного навчання для оцінки ризиків інвестицій в українську економіку. URL: <https://moodle.znu.edu.ua/mod/resource/view.php?id=78011>

25. Розробка моделі прогнозування курсу гривні за допомогою машинного навчання. URL: https://theses.oa.edu.ua/DATA/3543/%D0%AF%D1%86%D1%8E%D0%BA_%D0%B4%D1%80%D1%83%D0%BA.pdf

26. Використання машинного навчання для виявлення та запобігання шахрайству в державних фінансах України. URL: <https://cyberleninka.ru/article/n/vikoristannya-metodiv-mashinnogo-navchannya-dlya-viyavlennya-shahrayskih-vakansiy>
27. Застосування машинного навчання для оптимізації податкових зборів в Україні. URL: https://prometheus.org.ua/course/course-v1:IRF+ML101+2016_T3
28. Огляд методів машинного навчання для аналізу економічних даних. URL: <http://journals.ksauniv.ks.ua/index.php/tech/article/view/417>
29. Використання машинного навчання для прогнозування попиту та ціноутворення на товари та послуги. URL: <https://machinelearningmastery.com/>
30. Застосування машинного навчання для виявлення закономірностей у фінансових ринках. URL: https://www.wunu.edu.ua/opp/fkit/kompyterni_nayku/kompyterni_nayku_bakalavr/vubirkovi/Mashynne_navchannia/Sylabus.pdf
31. Розробка систем машинного навчання для автоматизації економічних завдань. URL: https://www.researchgate.net/publication/376878645_Modeli_ta_metodi_masinnogo_navcanna_dla_rozpiznavanna_fejkovogo_kontentu
32. Етичні аспекти використання машинного навчання в економіці. URL: <https://dergipark.org.tr/tr/download/article-file/254134>
33. Аналіз ринку декоративного каменю в Україні: тенденції, проблеми та перспективи розвитку. URL: <http://ten.ztu.edu.ua/article/view/283476/277680>
34. Аналіз світового ринку граніту: основні тенденції, виробники та споживачі. URL: <https://b-granum.com/ua/blog/doslidzennya-svitovogo-rinku-granitu-v-2023-roci>
35. Використання граніту в будівництві та дизайні. URL: <https://granit-pk.com/ua/articles/Granite-in-construction-architecture>
36. Видобуток та обробка граніту. URL: https://eiti.gov.ua/resursi-rozvidka-ta-vidobuvannya/kaminnya-budivelne_2022/

ДОДАТКИ

Додаток А

Код А.1. Аналіз та прогнозування показників обсягу реалізованої продукції за агрегаціями та КВЕДами. Аналіз бухгалтерської статистичної звітності.

```
from sklearn.model_selection import train_test_split, GridSearchCV
from scipy import stats
import xgboost as xgb
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.api as sm
from statsmodels.stats.diagnostic import het_breuschpagan
from statsmodels.stats.outliers_influence import variance_inflation_factor
from statsmodels.tsa.stattools import acf, pacf
from sklearn.impute import KNNImputer
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler
from sklearn.ensemble import RandomForestRegressor

# Шлях до файлу Excel
file_path = 'D:\KNU\ДИПЛОМ\Диплом укр.xlsx'

# Завантаження всіх листів в словник DataFrame
sheets_dict = pd.read_excel(file_path, sheet_name=None)

# Виведення назв листів
print("Назви листів у файлі:", sheets_dict.keys())

# Ініціалізація словника для зберігання DataFrame з індексами за номером листка
ues = {}
```

```
# Завантаження першого листа в DataFrame з індексом 1
first_sheet_name = list(sheets_dict.keys())[0]
ues[1] = sheets_dict[first_sheet_name]

# Виведення перших 5 рядків DataFrame для перевірки
print(f"\nЛист: {first_sheet_name}")
print(ues[1].head())

# Групування за унікальними значеннями з першої колонки
grouped = ues[1].groupby(ues[1].columns[0])

# Аналіз для кожної групи
for group_name, group_df in grouped:
    print(f"\nГрупа: {group_name}")

    # Орієнтація на значення років у 2-й колонці
    years = group_df.iloc[:, 1]

    # Показники з 3 до 12 стовпчиків
    data = group_df.iloc[:, 2:12]

    # Обчислення описової статистики
    mean = data.mean()
    median = data.median()
    mode = data.mode().iloc[0] # Враховуємо перше значення моди
    variance = data.var()
    std_dev = data.std()
    log_deviation = np.log(data + 1).std() # Логарифмічне відхилення

# Виведення результатів
print(f"Середнє значення:\n{mean}")
print(f"Медіана:\n{median}")
print(f"Мода:\n{mode}")
print(f"Дисперсія:\n{variance}")
print(f"Стандартне відхилення:\n{std_dev}")
```

```
print(f"Логарифмічне відхилення:\n{log_deviation}")

# Побудова стандартного розподілу
sns.histplot(data, kde=True)
plt.title(f'Standard Distribution for Group {group_name}')
plt.xlabel('Values')
plt.ylabel('Frequency')
plt.show()

# Розподіл Ст'юдента
t_stat, t_pvalue = stats.ttest_1samp(data, 0)
print(f"t-статистика:\n{t_stat}")
print(f"p-значення:\n{t_pvalue}")

# Коефіцієнт кореляції Пірсона
pearson_corr = data.corr(method='pearson')
print(f"Коефіцієнт кореляції Пірсона:\n{pearson_corr}")

# Коефіцієнт кореляції Спірмена
spearman_corr = data.corr(method='spearman')
print(f"Коефіцієнт кореляції Спірмена:\n{spearman_corr}")

# Завантаження другого листа в DataFrame
second_sheet_name = list(sheets_dict.keys())[1]
df = sheets_dict[second_sheet_name]

# Виведення перших 5 рядків DataFrame для перевірки
print(f"\nЛист: {second_sheet_name}")
print(df.head())

# Замінити "n/a" на NaN і заповнити пропущені значення
df.replace("n/a", np.nan, inplace=True)
df.ffill(inplace=True)
df.bfill(inplace=True)
```

```
# Вибір стовпчиків для виведення графіків, що закінчуються на "млн.грн"
graph_columns = [col for col in df.columns if col.endswith('млн.грн')]

# Групування за унікальними значеннями з першої колонки та орієнтація на роки у другій
колонці
grouped = df.groupby(df.columns[0])

# Ініціалізація списків для збереження результатів
results = []
forecasts = {}

# Нормування даних
scaler = MinMaxScaler()

# Аналіз для кожної групи
for group_name, group_df in grouped:
    print(f"\nГрупа: {group_name}")

    # Орієнтація на значення років у 2-й колонці
    years = group_df.iloc[:, 1].copy()

    # Показники для обчислення статистичних та економетричних показників
    data_stats = group_df.iloc[:, 2:].copy()
    # Показники для виведення графіків
    data_graph = group_df[graph_columns].copy()

    # Нормування даних для графіків
    data_graph_normalized = pd.DataFrame(scaler.fit_transform(data_graph),
columns=data_graph.columns,
index=data_graph.index)

    # Описова статистика
    mean = data_stats.mean()
    median = data_stats.median()
    variance = data_stats.var()
```

```

std_dev = data_stats.std()

# Економетричні показники: оцінка гомоскедастичності
X = sm.add_constant(data_stats)
y = years.loc[X.index] # Забезпечення узгодженості індексів
model = sm.OLS(y, X).fit()
_, pval, _, f_pval = het_breuschpagan(model.resid, model.model.exog)

print(f"Середнє значення:\n{mean}")
print(f"Медіана:\n{median}")
print(f"Дисперсія:\n{variance}")
print(f"Стандартне відхилення:\n{std_dev}")
print(f"Р-значення тесту Бройша-Пагана для оцінки гомоскедастичності: {pval}")

results.append({
    'group': group_name,
    'mean': mean,
    'median': median,
    'variance': variance,
    'std_dev': std_dev,
    'homoscedasticity_pval': pval
})

# Побудова графіка загальних трендів
plt.figure(figsize=(12, 8))
for column in data_graph_normalized.columns:
    plt.plot(years, data_graph_normalized[column], label=column)
plt.title(f"Тренди за роками для групи: {group_name}")
plt.xlabel('Роки')
plt.ylabel('Нормовані значення показників')
plt.legend()
plt.grid(True)
plt.show()

# Прогнозування

```

```

target_column = "Обсяг реалізованої продукції (товарів, послуг), млн.грн"
data_target = group_df[target_column].copy()

# Нормування даних для прогнозування
data_scaler = MinMaxScaler()
data_stats_normalized = data_scaler.fit_transform(data_stats)
target_scaler = MinMaxScaler()
data_target_normalized = target_scaler.fit_transform(data_target.values.reshape(-1, 1))

X = data_stats_normalized # Нормовані дані характеристик
y = data_target_normalized.flatten() # Нормовані значення цільової змінної

# Замінити "n/a" на NaN і заповнити пропущені значення у X
X = np.nan_to_num(X)

# Розділення даних на тренувальну та тестову вибірки
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Навчання моделі XGBoost
model = xgb.XGBRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# Прогнозування
y_pred = model.predict(X_test)

# Оцінка моделі
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
print(f"Mean Squared Error для групи {group_name}: {mse}")
print(f"R2 Score для групи {group_name}: {r2}")

forecasts[group_name] = {
    'mse': mse,
    'r2': r2,
    'model': model,

```

```

    'last_year': years.max(),
    'last_data': data_stats.iloc[-1],
    'base_value': data_target.iloc[-1], # Останнє фактичне значення для розрахунку приросту
    'scaler': target_scaler,
    'data_scaler': data_scaler
}

# Прогноз на 2023 та 2024 роки для всіх груп
forecast_years = [2023, 2024]
overall_forecast = []

for year in forecast_years:
    year_forecast = []
    for group_name, forecast in forecasts.items():
        last_data_normalized = forecast['data_scaler'].transform([forecast['last_data']])
        next_year_pred = forecast['model'].predict(last_data_normalized)
        next_year_pred_original = forecast['scaler'].inverse_transform(next_year_pred.reshape(-1,
1))[0][0]
        base_value = forecast['base_value']
        growth_rate = ((next_year_pred_original - base_value) / base_value) * 100
        year_forecast.append(growth_rate)
        print(f"Прогноз на {year} рік для групи '{group_name}': {growth_rate:.2f}% приросту")
        # Оновлення базового значення для наступного прогнозу
        forecast['base_value'] = next_year_pred_original
    overall_forecast.append((year, np.mean(year_forecast)))

# Виведення середнього приросту для всіх груп на 2023 та 2024 роки
for year, avg_growth in overall_forecast:
    print(f"Середній приріст на {year} рік для всіх груп: {avg_growth:.2f}%")

# Порівняння ефективності прогнозу для різних груп
print("\nПорівняння ефективності прогнозу для різних груп:")
for group, metrics in forecasts.items():
    print(f"Група: {group}")
    print(f"Mean Squared Error: {metrics['mse']}")

```

```
print(f"R2 Score: {metrics['r2']}\n")
# Завантаження третього листа в DataFrame
df = pd.read_excel(file_path, sheet_name=2)

# Виведення перших 5 рядків DataFrame для перевірки
print(f"\nЛист: {df.head()}")

# Замінити "n/a" на NaN
df.replace("n/a", np.nan, inplace=True)

# Прогнозування відсутніх значень з використанням KNNImputer
imputer = KNNImputer(n_neighbors=5)
df.iloc[:, 3:] = imputer.fit_transform(df.iloc[:, 3:])

# Вибір цільової змінної та характеристик
target_column = "Обсяг реалізованої продукції за функціональним підходом, тис. грн"
df_grouped = df.groupby(df.columns[1]) # Групування за другим стовпчиком

# Ініціалізація списків для збереження результатів
results = []
forecasts = {}

# Нормування даних
scaler = MinMaxScaler()

# Аналіз для кожної групи
for group_name, group_df in df_grouped:
    print(f"\nГрупа: {group_name}")

    # Орієнтація на значення років у 3-й колонці
    years = group_df.iloc[:, 2].copy()

    # Показники для обчислення статистичних та економетричних показників
    data_stats = group_df.iloc[:, 3:].copy()

    # Показники для прогнозування
```

```
data_target = group_df[target_column].copy()

# Нормування даних для прогнозування
data_scaler = MinMaxScaler()
data_stats_normalized = data_scaler.fit_transform(data_stats)
target_scaler = MinMaxScaler()
data_target_normalized = target_scaler.fit_transform(data_target.values.reshape(-1, 1))

X = data_stats_normalized # Нормовані дані характеристик
y = data_target_normalized.flatten() # Нормовані значення цільової змінної

# Замінити "n/a" на NaN і заповнити пропущені значення у X
X = np.nan_to_num(X)

# Розділення даних на тренувальну та тестову вибірки
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Пошук найкращих гіперпараметрів для RandomForestRegressor
param_grid = {
    'n_estimators': [100, 200, 300],
    'max_depth': [4, 6, 8, 10],
    'criterion': ['squared_error', 'absolute_error', 'friedman_mse', 'poisson']
}

grid_search = GridSearchCV(RandomForestRegressor(random_state=42), param_grid, cv=5,
scoring='r2')
grid_search.fit(X_train, y_train)
best_model = grid_search.best_estimator_

# Прогнозування
y_pred = best_model.predict(X_test)

# Оцінка моделі
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
```

```

print(f"Mean Squared Error для групи {group_name}: {mse}")
print(f"R2 Score для групи {group_name}: {r2}")

# Перевірка відхилення та повторне навчання, якщо відхилення більше 10%
if np.abs((y_pred - y_test) / y_test).mean() > 0.1:
    # Повторне навчання з додатковими гіперпараметрами
    param_grid = {
        'n_estimators': [100, 200, 300, 500],
        'max_depth': [4, 6, 8, 10, 12],
        'criterion': ['squared_error', 'absolute_error', 'friedman_mse', 'poisson']
    }
    grid_search = GridSearchCV(RandomForestRegressor(random_state=42), param_grid, cv=5,
scoring='r2')
    grid_search.fit(X_train, y_train)
    best_model = grid_search.best_estimator_
    y_pred = best_model.predict(X_test)
    mse = mean_squared_error(y_test, y_pred)
    r2 = r2_score(y_test, y_pred)
    print(f"Перенавчання для групи {group_name}")
    print(f"Mean Squared Error після перенавчання для групи {group_name}: {mse}")
    print(f"R2 Score після перенавчання для групи {group_name}: {r2}")

# Додавання результатів у список
results.append({
    'group': group_name,
    'mse': mse,
    'r2': r2,
    'model': best_model,
    'years': years,
    'data_stats': data_stats,
    'data_target': data_target,
    'data_scaler': data_scaler,
    'target_scaler': target_scaler,
    'X_test': X_test,
    'y_test': y_test,

```

```

        'y_pred': y_pred
    })

# Кореляційний аналіз (виключити текстові стовпчики)
numeric_df = df.select_dtypes(include=[np.number])
correlation_matrix = numeric_df.corr()
print("\nКореляційна матриця:\n", correlation_matrix)

# Визначення мультиколінеарності за допомогою VIF
vif_data = pd.DataFrame()
vif_data["feature"] = numeric_df.columns
vif_data["VIF"] = [variance_inflation_factor(numeric_df.values, i) for i in
range(len(numeric_df.columns))]
print("\nМультиколінеарність (VIF):\n", vif_data)

# Визначення автокореляції
acf_values = acf(df[target_column])
pacf_values = pacf(df[target_column])
print("\nАвтокореляція (ACF):\n", acf_values)
print("\nЧасткова автокореляція (PACF):\n", pacf_values)

# Використання тренду 2022-2025 з тенденцією 2015-2018 років як найвагоміший показник
trend_2015_2018 = df[(df.iloc[:, 2] >= 2015) & (df.iloc[:, 2] <=
2018)][target_column].pct_change().dropna().values
trend_2022_2025 = trend_2015_2018.mean() # Середній приріст за період 2015-2018 років

# Прогноз на 2023-2025 роки для кожної групи
forecast_years = [2023, 2024, 2025]
overall_forecast = []

for result in results:
    group_name = result['group']
    model = result['model']
    last_data = result['data_stats'].iloc[-1].values.reshape(1, -1)
    data_scaler = result['data_scaler']

```

```

target_scaler = result['target_scaler']
data_target = result['data_target']

# Прогнозування на 2023-2025 роки
for year in forecast_years:
    last_data_normalized = data_scaler.transform(last_data)
    next_year_pred = model.predict(last_data_normalized)
    next_year_pred_original = target_scaler.inverse_transform(next_year_pred.reshape(-1,
1))[0][0]
    next_year_pred_growth = next_year_pred_original * (1 + trend_2022_2025) # Зростаючий
тренд
    relative_growth = (next_year_pred_growth - data_target.iloc[-1]) / data_target.iloc[
-1] * 100 # Відносний приріст
    print(f"Прогноз на {year} рік для групи '{group_name}': {relative_growth:.2f}%")
    # Оновлення даних для наступного прогнозу
    last_data = np.append(last_data[:, 1:], next_year_pred_original).reshape(1, -1)

# Збереження прогнозів для кожного року
forecasts.setdefault(group_name, []).append({
    'year': year,
    'forecast': next_year_pred_growth,
    'relative_growth': relative_growth
})

# Побудова графіків ACF і PACF
plt.figure(figsize=(12, 6))
plt.subplot(121)
plt.stem(acf_values, use_line_collection=True)
plt.title("ACF")
plt.subplot(122)
plt.stem(pacf_values, use_line_collection=True)
plt.title("PACF")
plt.show()

# Побудова графіків прогнозів

```

```

plt.figure(figsize=(12, 8))
for result in results:
    group_name = result['group']
    years = result['years']
    data_target = result['data_target']
    y_pred = result['y_pred']
    y_test = result['y_test']

    plt.plot(years, data_target, label=f'Фактичні дані ({group_name})')
    plt.plot(years.iloc[-len(y_test):], result['target_scaler'].inverse_transform(y_pred.reshape(-1,
1)),
            label=f'Прогноз ({group_name})')

plt.title('Прогнозування обсягу реалізованої продукції за функціональним підходом')
plt.xlabel('Роки')
plt.ylabel('Обсяг реалізованої продукції, тис. грн')
plt.legend()
plt.grid(True)
plt.show()

# Виведення прогнозів на 2023-2025 роки
for group_name, group_forecasts in forecasts.items():
    print(f"\nПрогнози для групи '{group_name}':")
    for forecast in group_forecasts:
        print(
            f'Рік: {forecast['year']}, Прогноз: {forecast['forecast']:.2f}, Відносний приріст:
{forecast['relative_growth']:.2f}%")

# Виведення результатів кореляційного аналізу
print("\nКореляційна матриця:")
print(correlation_matrix)

# Виведення результатів мультиколінеарності
print("\nМультиколінеарність (VIF):")
print(vif_data)

```

```
# Виведення автокореляції
print("\nАвтокореляція (ACF):")
print(acf_values)
print("\nЧасткова автокореляція (PACF):")
print(pacf_values)

# Шлях до файлу Excel
file_path = 'D:/KNU/ДИПЛОМ/Диплом.xlsx'

# Завантаження даних з листа "1" в DataFrame
df = pd.read_excel(file_path, sheet_name="1")

# Перетворення назви стовпців на роки (з 2-го по 11-й стовпчик)
df.columns = ['Актив'] + [str(year) for year in range(2022, 2012, -1)]
df.set_index('Актив', inplace=True)

# Транспонування DataFrame для аналізу
df = df.transpose()

# Вибір цільової змінної
target_asset = 'Чистий дохід від реалізації продукції'
df_target = df[target_asset].dropna().astype(float)

# Перевірка на пропущені значення та заповнення їх
df.replace("n/a", np.nan, inplace=True)
df = df.ffill().bfill()

# Побудова економетричної моделі
X = df.index.astype(int).values.reshape(-1, 1)
y = df_target.values
X = sm.add_constant(X)
model = sm.OLS(y, X).fit()
print(model.summary())
```

```
# Кореляційний аналіз
df_factors = df.drop(columns=[target_asset]).dropna().astype(float)
correlation_matrix = df_factors.corrwith(df_target)
top_3_factors = correlation_matrix.abs().sort_values(ascending=False).head(3)
top_3_factors_names = top_3_factors.index
print("\nТоп 3 фактори за значенням кореляції:\n", top_3_factors)

# Вплив факторів на цільову функцію
influence_percent = top_3_factors / top_3_factors.sum() * 100
print("\nВідсотковий вплив факторів:\n", influence_percent)
```

Додаток Б

Таблиця Б.1. Описові характеристики підприємств за масштабом

Показник	Рік	Усього	Підприємства усього	малі
Кількість суб'єктів господарювання, одиниць	2010	2 183 928,00	378 810,00	357 241,00
Кількість суб'єктів господарювання, одиниць	2011	1 701 620,00	375 695,00	354 283,00
Кількість суб'єктів господарювання, одиниць	2012	1 600 127,00	364 935,00	344 048,00
Кількість суб'єктів господарювання, одиниць	2013	1 722 070,00	393 327,00	373 809,00
Кількість суб'єктів господарювання, одиниць	2014	1 932 161,00	341 001,00	324 598,00
Кількість суб'єктів господарювання, одиниць	2015	1 974 318,00	343 440,00	327 814,00

Продовження таблиця Б.1.

Кількість суб'єктів господарювання, одиниць	2016	1 865 530,00	306 369,00	291 154,00
Кількість суб'єктів господарювання, одиниць	2017	1 805 059,00	338 256,00	322 920,00
Кількість суб'єктів господарювання, одиниць	2018	1 839 593,00	355 877,00	339 374,00
Кількість суб'єктів господарювання, одиниць	2019	1 941 625,00	380 597,00	362 328,00
Кількість суб'єктів господарювання, одиниць	2020	1 973 577,00	373 822,00	355 708,00
Кількість суб'єктів господарювання, одиниць	2021	1 956 248,00	370 834,00	352 722,00
Кількість суб'єктів господарювання, одиниць	2022	1 732 508,00	261 924,00	246 647,00
Кількість суб'єктів господарювання на 10 тис. осіб наявного населення, одиниць	2010	477,00	83,00	78,00
Кількість суб'єктів господарювання на 10 тис. осіб наявного населення, одиниць	2011	372,00	82,00	77,00
Кількість суб'єктів господарювання на 10 тис. осіб наявного населення, одиниць	2012	351,00	80,00	76,00

Продовження таблиця Б.1.

Кількість суб'єктів господарювання на 10 тис. осіб наявного населення, одиниць	2013	378,00	86,00	82,00
Кількість суб'єктів господарювання на 10 тис. осіб наявного населення, одиниць	2014	450,00	80,00	76,00
Кількість суб'єктів господарювання на 10 тис. осіб наявного населення, одиниць	2015	462,00	81,00	77,00
Кількість суб'єктів господарювання на 10 тис. осіб наявного населення, одиниць	2016	437,00	72,00	68,00
Кількість суб'єктів господарювання на 10 тис. осіб наявного населення, одиниць	2017	425,00	80,00	76,00
Кількість суб'єктів господарювання на 10 тис. осіб наявного населення, одиниць	2018	435,00	84,00	80,00
Кількість суб'єктів господарювання на 10 тис. осіб наявного населення, одиниць	2019	462,00	91,00	86,00
Кількість суб'єктів господарювання на 10 тис. осіб наявного населення, одиниць	2020	473,00	90,00	85,00

Продовження таблиця Б.1.

Кількість суб'єктів господарювання на 10 тис. осіб наявного населення, одиниць	2021	473,00	90,00	85,00
Кількість суб'єктів господарювання на 10 тис. осіб наявного населення, одиниць	2022	n/a	n/a	n/a
Обсяг реалізованої продукції (товарів, послуг), млн.грн	2010	3 596 646,40	3 366 228,20	568 267,10
Обсяг реалізованої продукції (товарів, послуг), млн.грн	2011	4 202 455,20	3 991 239,40	607 782,40
Обсяг реалізованої продукції (товарів, послуг), млн.грн	2012	4 459 818,80	4 203 169,60	672 653,40
Обсяг реалізованої продукції (товарів, послуг), млн.грн	2013	4 334 453,10	4 050 215,00	670 258,50
Обсяг реалізованої продукції (товарів, послуг), млн.грн	2014	4 459 702,20	4 170 659,90	705 000,50
Обсяг реалізованої продукції (товарів, послуг), млн.грн	2015	5 556 540,40	5 159 067,10	937 112,80
Обсяг реалізованої продукції (товарів, послуг), млн.грн	2016	6 726 739,80	6 237 535,20	1 177 385,20
Обсяг реалізованої продукції (товарів, послуг), млн.грн	2017	8 312 271,90	7 707 935,20	1 482 000,70

Продовження таблиця Б.1.

Обсяг реалізованої продукції (товарів, послуг), млн.грн	2018	9 966 804,50	9 206 049,50	1 766 150,40
Обсяг реалізованої продукції (товарів, послуг), млн.грн	2019	10 524 112,80	9 639 730,60	1 839 875,90
Обсяг реалізованої продукції (товарів, послуг), млн.грн	2020	11 062 297,10	10 049 870,80	2 064 120,70
Обсяг реалізованої продукції (товарів, послуг), млн.грн	2021	14 989 811,90	13 616 793,20	2 576 371,40
Обсяг реалізованої продукції (товарів, послуг), млн.грн	2022	11 119 875,60	11 033 018,10	2 101 912,00

Джерело: [14]

Додаток В

Код В.1. Застосування NLP для описової статистики товарів.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score
from sklearn.preprocessing import MinMaxScaler
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from collections import Counter
import itertools

# Завантаження файлу
file_path = 'D:\\KNU\\ДИПЛОМ\\Диплом.xlsx'
df = pd.read_excel(file_path, sheet_name='Продажі Юніком')

# Перевірка структури файлу
print(df.head())

# Виправлення помилок у колонках
df.columns = ['Назва товару', 'Сума покупки', 'Кількість проданих товарів', 'Ціна товару']

# Фільтрація даних за категоріями
categories = ["шт", "т", "компл", "пог.м.", "м2"]
category_stats = {}

for category in categories:
    category_data = df[df['Назва товару'].str.endswith(category)]
    if not category_data.empty:
        category_stats[category] = {
            "Середня ціна": category_data['Ціна товару'].mean(),
            "Мода ціни": category_data['Ціна товару'].mode()[0] if not category_data['Ціна
товару'].mode().empty else np.nan,
            "Медіана ціни": category_data['Ціна товару'].median(),
            "Дисперсія ціни": category_data['Ціна товару'].var()
        }

category_stats_df = pd.DataFrame(category_stats).transpose()
print("\nОписова статистика для категорій товарів:")
print(category_stats_df)

# Побудова трендів цін
plt.figure(figsize=(12, 8))
for category in categories:
    category_data = df[df['Назва товару'].str.endswith(category)]
    plt.plot(category_data['Ціна товару'].values, label=category)

plt.title('Тренди цін за категоріями товарів')
```

```

plt.xlabel('Порядковий номер')
plt.ylabel('Ціна товару')
plt.ylim(0, 20000)
plt.legend()
plt.grid(True)
plt.show()

# NLP аналіз для виділення сенсових ключових слів
nltk.download('punkt')
nltk.download('stopwords')

try:
    stop_words = set(stopwords.words('ukrainian'))
except OSError:
    stop_words = set(stopwords.words('english'))

# Виділення ключових слів
df['Назва товару'] = df['Назва товару'].astype(str)
df['Токени'] = df['Назва товару'].apply(lambda x: [word for word in word_tokenize(x.lower()) if
word.isalpha() and word not in stop_words])
all_tokens = list(itertools.chain(*df['Токени']))

# Статистика по токенах
token_stats = {}

for token in set(all_tokens):
    token_data = df[df['Токени'].apply(lambda x: token in x)]
    token_stats[token] = {
        "Середня сума покупки": token_data['Сума покупки'].mean(),
        "Середня кількість проданих товарів": token_data['Кількість проданих товарів'].mean(),
        "Середня ціна товару": token_data['Ціна товару'].mean(),
        "Мода ціни": token_data['Ціна товару'].mode()[0] if not token_data['Ціна
товару'].mode().empty else np.nan,
        "Медіана ціни": token_data['Ціна товару'].median(),
        "Дисперсія ціни": token_data['Ціна товару'].var()
    }

token_stats_df = pd.DataFrame(token_stats).transpose()
print("\nОписова статистика для ключових слів:")
print(token_stats_df)

# Перехресне комбінування слів
combinations = list(itertools.combinations(set(all_tokens), 2))
combo_stats = {}

for combo in combinations:
    combo_data = df[df['Токени'].apply(lambda x: all(word in x for word in combo))]
    combo_stats[combo] = {
        "Середня сума покупки": combo_data['Сума покупки'].mean(),
        "Середня кількість проданих товарів": combo_data['Кількість проданих товарів'].mean(),
        "Середня ціна товару": combo_data['Ціна товару'].mean(),
        "Мода ціни": combo_data['Ціна товару'].mode()[0] if not combo_data['Ціна

```

```

товару'].mode().empty else np.nan,
    "Медіана ціни": combo_data['Ціна товару'].median(),
    "Дисперсія ціни": combo_data['Ціна товару'].var()
}

combo_stats_df = pd.DataFrame(combo_stats).transpose()
combo_stats_df.columns = ['Середня сума покупки', 'Середня кількість проданих товарів',
'Sередня ціна товару', 'Мода ціни', 'Медіана ціни', 'Дисперсія ціни']
combo_stats_df.dropna(inplace=True)

# Визначення топ 50 найпопулярніших комбінацій
top_50_combos = combo_stats_df.sort_values(by='Середня сума покупки',
ascending=False).head(50)
print("\nТоп 50 найпопулярніших комбінацій слів:")
print(top_50_combos)

# Тенденції зміни ціни для окремих слів
price_trend_tokens = {}

for token in set(all_tokens):
    token_data = df[df['Токени'].apply(lambda x: token in x)]
    if not token_data.empty:
        grouped = token_data.groupby(token_data.index // 100)['Ціна товару'].mean()
        price_trend_tokens[token] = grouped

price_trend_tokens_df = pd.DataFrame(price_trend_tokens)
print("\nТенденції зміни ціни для окремих слів (групи по 100):")
print(price_trend_tokens_df)

# Кластеризація слів і комбінацій слів
def optimal_kmeans(data):
    silhouette_scores = []
    K = range(2, 11)
    for k in K:
        kmeans = KMeans(n_clusters=k, random_state=42)
        kmeans.fit(data)
        score = silhouette_score(data, kmeans.labels_)
        silhouette_scores.append(score)
    optimal_k = K[np.argmax(silhouette_scores)]
    return optimal_k, silhouette_scores[np.argmax(silhouette_scores)]

# Кластеризація для токенів
token_stats_df[['Середня сума покупки', 'Середня кількість проданих товарів', 'Середня ціна
товару']] = token_stats_df[['Середня сума покупки', 'Середня кількість проданих товарів',
'Sередня ціна товару']].fillna(0)
token_k, token_score = optimal_kmeans(token_stats_df[['Середня сума покупки', 'Середня
кількість проданих товарів', 'Середня ціна товару']])
kmeans_tokens = KMeans(n_clusters=token_k, random_state=42)
token_labels = kmeans_tokens.fit_predict(token_stats_df[['Середня сума покупки', 'Середня
кількість проданих товарів', 'Середня ціна товару']])
token_stats_df['Кластер'] = token_labels

```

```

# Кластеризація для комбінацій слів
combo_stats_df[['Середня сума покупок', 'Середня кількість проданих товарів', 'Середня ціна товару']] = combo_stats_df[['Середня сума покупок', 'Середня кількість проданих товарів', 'Середня ціна товару']].fillna(0)
combo_k, combo_score = optimal_kmeans(combo_stats_df[['Середня сума покупок', 'Середня кількість проданих товарів', 'Середня ціна товару']])
kmeans_combos = KMeans(n_clusters=combo_k, random_state=42)
combo_labels = kmeans_combos.fit_predict(combo_stats_df[['Середня сума покупок', 'Середня кількість проданих товарів', 'Середня ціна товару']])
combo_stats_df['Кластер'] = combo_labels

# Збереження результатів у файли Excel
category_stats_df.to_excel("category_stats.xlsx", index=True)
token_stats_df.to_excel("token_stats.xlsx", index=True)
combo_stats_df.to_excel("combo_stats.xlsx", index=True)
top_50_combos.to_excel("top_50_combos.xlsx", index=True)
price_trend_tokens_df.to_excel("price_trend_tokens.xlsx", index=True)

print("\nОписова статистика для категорій товарів збережена у 'category_stats.xlsx'")
print("\nОписова статистика для ключових слів збережена у 'token_stats.xlsx'")
print("\nОписова статистика для комбінацій слів збережена у 'combo_stats.xlsx'")
print("\nТоп 50 найпопулярніших комбінацій слів збережено у 'top_50_combos.xlsx'")
print("\nТенденції зміни ціни для окремих слів збережено у 'price_trend_tokens.xlsx'")

# Виведення результатів кластеризації
print(f"\nКластеризація для ключових слів: Кількість кластерів = {token_k}, Silhouette Score = {token_score:.2f}")
print(f"\nКластеризація для комбінацій слів: Кількість кластерів = {combo_k}, Silhouette Score = {combo_score:.2f}")

print("\nСлова та комбінації слів, що належать до кластерів:")

# Виведення слів і комбінацій слів, що належать до кластерів
for i in range(token_k):
    cluster_words = token_stats_df[token_stats_df['Кластер'] == i].index.tolist()
    print(f"\nКластер {i} (ключові слова): {' '.join(cluster_words)}")

for i in range(combo_k):
    cluster_combos = combo_stats_df[combo_stats_df['Кластер'] == i].index.tolist()
    print(f"\nКластер {i} (комбінації слів): {' '.join([' '.join(combo) for combo in cluster_combos])}")

# Збереження результатів кластеризації у файли Excel
token_stats_df.to_excel("token_stats_clusters.xlsx", index=True)
combo_stats_df.to_excel("combo_stats_clusters.xlsx", index=True)

print("\nКластеризація для ключових слів збережена у 'token_stats_clusters.xlsx'")
print("\nКластеризація для комбінацій слів збережена у 'combo_stats_clusters.xlsx'")

# Виведення ключових слів та комбінацій слів, що належать до кластерів
for i in range(token_k):
    cluster_words = token_stats_df[token_stats_df['Кластер'] == i].index.tolist()

```

```
print(f"\nКластер {i} (ключові слова): {' '.join(cluster_words)}")

for i in range(combo_k):
    cluster_combos = combo_stats_df[combo_stats_df['Кластер'] == i].index.tolist()
    print(f"\nКластер {i} (комбінації слів): {' '.join([' '.join(combo) for combo in
cluster_combos])}")

# Збереження результатів кожного етапу аналізу в таблиці Excel
with pd.ExcelWriter("analysis_results.xlsx") as writer:
    category_stats_df.to_excel(writer, sheet_name='Category Stats', index=True)
    token_stats_df.to_excel(writer, sheet_name='Token Stats', index=True)
    combo_stats_df.to_excel(writer, sheet_name='Combo Stats', index=True)
    top_50_combos.to_excel(writer, sheet_name='Top 50 Combos', index=True)
    price_trend_tokens_df.to_excel(writer, sheet_name='Price Trend Tokens', index=True)
    token_stats_df.to_excel(writer, sheet_name='Token Clusters', index=True)
    combo_stats_df.to_excel(writer, sheet_name='Combo Clusters', index=True)

print("\nВсі результати аналізу збережені у 'analysis_results.xlsx'")
```

Додаток Г

Код Г.1. Використання NLP для визначення кореляції цін для Родовищ та прогнозування ціни одиниці товару

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, r2_score
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from collections import Counter
import itertools
import joblib

# Завантаження файлу
file_path = 'Диплом.xlsx'
df = pd.read_excel(file_path, sheet_name='Продажі Юніком')

# Перевірка структури файлу
print(df.head())

# Виправлення помилок у колонках
df.columns = ['Назва товару', 'Сума покупки', 'Кількість проданих товарів', 'Ціна товару']

# Фільтрація даних за категоріями
categories = ["шт", "т", "компл", "пог.м.", "м2"]
category_stats = {}

for category in categories:
    category_data = df[df['Назва товару'].str.endswith(category)]
    if not category_data.empty:
```

```
category_stats[category] = {
    "Середня ціна": category_data['Ціна товару'].mean(),
    "Мода ціни": category_data['Ціна товару'].mode()[0] if not category_data['Ціна
товару'].mode().empty else np.nan,
    "Медіана ціни": category_data['Ціна товару'].median(),
    "Дисперсія ціни": category_data['Ціна товару'].var()
}
```

```
category_stats_df = pd.DataFrame(category_stats).transpose()
print("\nОписова статистика для категорій товарів:")
print(category_stats_df)
```

```
# Побудова трендів цін
plt.figure(figsize=(12, 8))
for category in categories:
    category_data = df[df['Назва товару'].str.endswith(category)]
    plt.plot(category_data['Ціна товару'].values, label=category)
```

```
plt.title('Тренди цін за категоріями товарів')
plt.xlabel('Порядковий номер')
plt.ylabel('Ціна товару')
plt.ylim(0, 20000)
plt.legend()
plt.grid(True)
plt.show()
```

```
# NLP аналіз для виділення сенсових ключових слів
nltk.download('punkt')
nltk.download('stopwords')
nltk.download('averaged_perceptron_tagger')
```

```
try:
    stop_words = set(stopwords.words('ukrainian'))
except OSError:
    stop_words = set(stopwords.words('english'))
```

```

# Виділення ключових слів
df['Назва товару'] = df['Назва товару'].astype(str)
df['Токени'] = df['Назва товару'].apply(lambda x: [word for word in word_tokenize(x.lower()) if
word.isalpha() and word not in stop_words])
all_tokens = list(itertools.chain(*df['Токени']))

# Поділ на групи: Види виробів, Родовища, Інші
product_types = ['слеб', 'сляб', 'блок', 'валун', 'підсхода', 'плитка', 'бордюр', 'бруківка', 'плити',
'сходи', 'бордюра', 'стовпчики', 'перила', 'продукція', 'бутовий', 'камінь', 'кам', 'вироби',
'бруківки', 'бортир', 'постамент', 'зразок', 'брусок', 'комплект', 'підвіконня', 'камень', 'стовпчик']
deposits = ['лабрадорит', 'габро', 'лаброд', 'лабро', 'жежелів'] + [word for word in all_tokens if
word.endswith('y') or word.endswith('ого')]
other_words = [word for word in all_tokens if word not in product_types and word not in deposits]

# Створення DataFrame для статистики
group_stats = {
    "Види виробів": df[df['Токени'].apply(lambda x: any(item in x for item in product_types))],
    "Родовища": df[df['Токени'].apply(lambda x: any(item in x for item in deposits))],
    "Інші": df[df['Токени'].apply(lambda x: any(item in x for item in other_words))]
}

group_stats_data = {
    group: {
        "Середня сума покупки": data['Сума покупки'].mean(),
        "Середня кількість проданих товарів": data['Кількість проданих товарів'].mean(),
        "Середня ціна товару": data['Ціна товару'].mean(),
        "Мода ціни": data['Ціна товару'].mode()[0] if not data['Ціна товару'].mode().empty else np.nan,
        "Медіана ціни": data['Ціна товару'].median(),
        "Дисперсія ціни": data['Ціна товару'].var()
    } for group, data in group_stats.items()
}

group_stats_df = pd.DataFrame(group_stats_data).transpose()
print("\nОписова статистика для груп:")

```

```

print(group_stats_df)

# Описова статистика для окремих слів
word_stats = {}

for word in set(all_tokens):
    word_data = df[df['Токени'].apply(lambda x: word in x)]
    word_stats[word] = {
        "Середня сума покупки": word_data['Сума покупки'].mean(),
        "Середня кількість проданих товарів": word_data['Кількість проданих товарів'].mean(),
        "Середня ціна товару": word_data['Ціна товару'].mean(),
        "Мода ціни": word_data['Ціна товару'].mode()[0] if not word_data['Ціна
товару'].mode().empty else np.nan,
        "Медіана ціни": word_data['Ціна товару'].median(),
        "Дисперсія ціни": word_data['Ціна товару'].var()
    }

word_stats_df = pd.DataFrame(word_stats).transpose()
print("\nОписова статистика для окремих слів:")
print(word_stats_df)

# Перехресне комбінування слів лише для 'Види виробів' та 'Родовища'
combinations = list(itertools.product(product_types, deposits))
combo_stats = {}

for combo in combinations:
    combo_data = df[df['Токени'].apply(lambda x: all(word in x for word in combo))]
    combo_stats[combo] = {
        "Середня сума покупки": combo_data['Сума покупки'].mean(),
        "Середня кількість проданих товарів": combo_data['Кількість проданих товарів'].mean(),
        "Середня ціна товару": combo_data['Ціна товару'].mean(),
        "Мода ціни": combo_data['Ціна товару'].mode()[0] if not combo_data['Ціна
товару'].mode().empty else np.nan,
        "Медіана ціни": combo_data['Ціна товару'].median(),
        "Дисперсія ціни": combo_data['Ціна товару'].var()
    }

```

```

    }

    combo_stats_df = pd.DataFrame(combo_stats).transpose()
    combo_stats_df.columns = ['Середня сума покупки', 'Середня кількість проданих товарів',
    'Середня ціна товару', 'Мода ціни', 'Медіана ціни', 'Дисперсія ціни']
    combo_stats_df.dropna(inplace=True)

    # Визначення топ 50 найпопулярніших комбінацій
    top_50_combos = combo_stats_df.sort_values(by='Середня сума покупки',
    ascending=False).head(50)
    print("\nТоп 50 найпопулярніших комбінацій слів:")
    print(top_50_combos)

    # Тенденції зміни ціни для окремих слів
    price_trend_tokens = { }

    for word in set(all_tokens):
        word_data = df[df['Токени'].apply(lambda x: word in x)]
        if not word_data.empty:
            grouped = word_data.groupby(word_data.index // 15)['Ціна товару'].mean()
            price_trend_tokens[word] = grouped

    price_trend_tokens_df = pd.DataFrame(price_trend_tokens)
    print("\nТенденції зміни ціни для окремих слів (групи по 15):")
    print(price_trend_tokens_df)

    # Збереження результатів у файли Excel
    category_stats_df.to_excel("category_stats.xlsx", index=True)
    group_stats_df.to_excel("group_stats.xlsx", index=True)
    word_stats_df.to_excel("word_stats.xlsx", index=True)
    combo_stats_df.to_excel("combo_stats.xlsx", index=True)
    top_50_combos.to_excel("top_50_combos.xlsx", index=True)
    price_trend_tokens_df.to_excel("price_trend_tokens.xlsx", index=True)

    print("\nОписова статистика для категорій товарів збережена у 'category_stats.xlsx'")

```

```

print("\nОписова статистика для груп збережена у 'group_stats.xlsx")
print("\nОписова статистика для окремих слів збережена у 'word_stats.xlsx")
print("\nОписова статистика для комбінацій слів збережена у 'combo_stats.xlsx")

print("\nТоп 50 найпопулярніших комбінацій слів збережено у 'top_50_combos.xlsx")
print("\nТенденції зміни ціни для окремих слів збережено у 'price_trend_tokens.xlsx")

# Збереження результатів кожного етапу аналізу в таблиці Excel
with pd.ExcelWriter("analysis_results.xlsx") as writer:
    category_stats_df.to_excel(writer, sheet_name='Category Stats', index=True)
    group_stats_df.to_excel(writer, sheet_name='Group Stats', index=True)
    word_stats_df.to_excel(writer, sheet_name='Word Stats', index=True)
    combo_stats_df.to_excel(writer, sheet_name='Combo Stats', index=True)
    top_50_combos.to_excel(writer, sheet_name='Top 50 Combos', index=True)
    price_trend_tokens_df.to_excel(writer, sheet_name='Price Trend Tokens', index=True)

print("\nВсі результати аналізу збережені у 'analysis_results.xlsx")

# Побудова кореляційної матриці для Родовищ
deposits_data = group_stats["Родовища"]
numeric_deposits_data = deposits_data.select_dtypes(include=[np.number])
correlation_matrix = numeric_deposits_data.corr()
plt.figure(figsize=(12, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Кореляційна матриця залежності цін для Родовищ')
plt.show()

# Підготовка даних для моделі машинного навчання
df['Category'] = df['Назва товару'].apply(lambda x: next((cat for cat in categories if x.endswith(cat)),
'other'))
df['Product_Type'] = df['Токени'].apply(lambda x: next((ptype for ptype in product_types if ptype in
x), 'other'))
df['Deposit'] = df['Токени'].apply(lambda x: next((dep for dep in deposits if dep in x), 'other'))

# Переконайтесь, що всі можливі категорії враховані

```

```
all_categories =
list(set(df['Category']).union(set(top_50_combos.index.get_level_values(0).map(lambda x: next((cat
for cat in categories if x.endswith(cat)), 'other')))))
all_product_types =
list(set(df['Product_Type']).union(set(top_50_combos.index.get_level_values(0).map(lambda x:
next((ptype for ptype in product_types if ptype in x), 'other')))))
all_deposits =
list(set(df['Deposit']).union(set(top_50_combos.index.get_level_values(1).map(lambda x: next((dep
for dep in deposits if dep in x), 'other')))))
```

```
# Кодування категоріальних змінних
```

```
label_encoders = {
    'Category': LabelEncoder().fit(all_categories),
    'Product_Type': LabelEncoder().fit(all_product_types),
    'Deposit': LabelEncoder().fit(all_deposits)
}
```

```
df['Category'] = label_encoders['Category'].transform(df['Category'])
df['Product_Type'] = label_encoders['Product_Type'].transform(df['Product_Type'])
df['Deposit'] = label_encoders['Deposit'].transform(df['Deposit'])
```

```
# Вибір ознак та цільової змінної
```

```
features = ['Category', 'Product_Type', 'Deposit']
target = 'Ціна товару'
X = df[features]
y = df[target]
```

```
# Поділ даних на тренувальну та тестову вибірки
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
# Побудова моделі Random Forest
```

```
rf_model = RandomForestRegressor(n_estimators=100, random_state=42)
rf_model.fit(X_train, y_train)
```

```
# Прогнозування та оцінка моделі
```

```

y_pred = rf_model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
print(f"Mean Squared Error: {mse}")
print(f"R-squared: {r2}")

# Побудова карти розподілу прогнозів
plt.figure(figsize=(12, 8))
plt.scatter(y_test, y_pred, alpha=0.5)
plt.plot([y.min(), y.max()], [y.min(), y.max()], 'k--', lw=2)
plt.xlabel('Справжні значення')
plt.ylabel('Прогнозовані значення')
plt.title('Карта розподілу прогнозів')
plt.ylim(0, 20000)
plt.xlim(0, 20000)
plt.grid(True)
plt.show()

# Збереження моделі
joblib.dump(rf_model, 'rf_model.pkl')
print("Модель Random Forest збережена як 'rf_model.pkl'")

# Прогнози для топ 50 найпопулярніших комбінацій слів
def encode_new_data(row, column, label_encoders, categories, product_types, deposits):
    if column == 'Category':
        return next((cat for cat in categories if row.endswith(cat)), 'other')
    elif column == 'Product_Type':
        return next((ptype for ptype in product_types if ptype in row), 'other')
    elif column == 'Deposit':
        return next((dep for dep in deposits if dep in row), 'other')
    return 'other'

top_50_combos['Category'] = top_50_combos.index.map(lambda x: encode_new_data(x[0],
'Category', label_encoders, categories, product_types, deposits))

```

```

top_50_combos['Product_Type'] = top_50_combos.index.map(lambda x: encode_new_data(x[0],
'Product_Type', label_encoders, categories, product_types, deposits))
top_50_combos['Deposit'] = top_50_combos.index.map(lambda x: encode_new_data(x[1], 'Deposit',
label_encoders, categories, product_types, deposits))

top_50_combos['Category'] = label_encoders['Category'].transform(top_50_combos['Category'])
top_50_combos['Product_Type'] = label_encoders['Product_Type'].transform(top_50_combos['Product_Type'])
top_50_combos['Deposit'] = label_encoders['Deposit'].transform(top_50_combos['Deposit'])

top_50_combos_features = top_50_combos[['Category', 'Product_Type', 'Deposit']]
top_50_combos['Прогнозована ціна'] = rf_model.predict(top_50_combos_features)

print("\nПрогнози для топ 50 найпопулярніших комбінацій слів:")
print(top_50_combos[['Середня сума покупки', 'Середня кількість проданих товарів', 'Середня
ціна товару', 'Прогнозована ціна']])
from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay

# Прогнозування класів для тестової вибірки
y_test_classes = np.where(y_test > y_test.median(), 1, 0) # Класифікація на основі медіани
y_pred_classes = np.where(y_pred > y_test.median(), 1, 0) # Класифікація на основі медіани

# Побудова матриці плутанини
cm = confusion_matrix(y_test_classes, y_pred_classes)
disp = ConfusionMatrixDisplay(confusion_matrix=cm, display_labels=['Нижче медіани', 'Вище
медіани'])

# Відображення матриці плутанини
disp.plot(cmap=plt.cm.Blues)
plt.title('Confusion Matrix')
plt.show()

# Збереження результатів прогнозування у файл Excel
prediction_results = X_test.copy()
prediction_results['Actual Price'] = y_test
prediction_results['Predicted Price'] = y_pred

```

```
# Додавання колонки з класами
prediction_results['Actual Class'] = y_test_classes
prediction_results['Predicted Class'] = y_pred_classes

# Збереження у Excel
prediction_results.to_excel("prediction_results.xlsx", index=False)
print("Результати прогнозування збережені у 'prediction_results.xlsx'")
```

Київський національний університет імені Тараса Шевченка

Економічний факультет

Кафедра економічної кібернетики

Завдання

на кваліфікаційну роботу бакалавра

студента 4 курсу спеціальності 051 «Економіка»,

ОПП «Економічна кібернетика»

Чарнюка Максима Валерійовича

1. Тема роботи: Моделювання доходів добувальної та виробничої діяльності підприємства «ЮНІКОМ-ПРОМ» на ринку декоративного каменю України
2. Термін завершення роботи: 03.06.2024
3. Попередній захист роботи: 03.06.2024
4. Об'єкт дослідження: ринок декоративного каменю України.
5. Предмет дослідження: моделювання доходів від виробництва декоративного каменю та структурний аналіз продажів товарів за допомогою методів обробки природної мови (NLP).
6. Мета дослідження: розробка та застосування економетричних моделей для прогнозування доходів виробництва на ринку декоративного каменю України, оцінки ключових трендів й небезпек ринку та проведення структурного аналізу продажів товарів за допомогою NLP.
7. Завдання дослідження:
 - 7.1. Вивчити існуючі методи економетричного моделювання доходів та адаптувати їх до специфіки ринку декоративного каменю України.
 - 7.2. Описати структуру та динаміку ринку декоративного каменю, визначити основні фактори, що впливають на доходи.
 - 7.3. Побудувати економетричні моделі для прогнозування обсягів реалізації виробництва на ринку декоративного каменю. Зробити прогнози щодо зміни цін та доходів на основі побудованих моделей.

7.4. Використати методи обробки природної мови (NLP) для аналізу текстових даних про продажі товарів та визначити релевантність використання різних специфікацій каменю для підприємства «ЮНІКОМ-ПРОМ»

Науковий керівник:

Кандидат економічних наук, доцент

Шпирко Віктор Васильович

Студент:  (підпис)

Затверджено на засіданні кафедри економічної кібернетики
протокол № ... від..... 20 р.

Календарний план виконання кваліфікаційної роботи бакалавра

№	Етапи роботи	Терміни виконання	Відмітка керівника про виконання
1	Вибір теми кваліфікаційної роботи бакалавра	28.01.2024	
2	Розробка та затвердження завдання кваліфікаційної роботи бакалавра	01.02.2024	
3	Аналіз наявних наукових досліджень за темою	31.03.2024	
4	Вибір факторів, пошук відповідних даних	10.04.2024	
5	Проведення аналізу відкритих даних	22.04.2024	
6	Проведення аналізу даних підприємства	15.05.2024	
7	Оцінка отриманих результатів, аналіз прогнозів, формування висновків	25.05.2024	
8	Подання роботи до попереднього захисту	02.06.2024	