

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ПРИРОДОКОРИСТУВАННЯ
ФАКУЛЬТЕТ МЕХАНІКИ, ЕНЕРГЕТИКИ ТА ІНФОРМАЦІЙНИХ
ТЕХНОЛОГІЙ
КАФЕДРА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ**

КВАЛІФІКАЦІЙНА РОБОТА

другого (магістерського) рівня вищої освіти

на тему: **«Інформаційна система прогнозування вартості
транспортування насіння зернових культур»**

Виконав: студент групи Іт-61

Спеціальності 126 «Інформаційні системи та
технології»

(шифр і назва)

Литвин Олег Русланович

(Прізвище та ініціали)

Керівник: к.е.н., доцент Желєзняк А.М.

(Прізвище та ініціали)

Рецензент: к.т.н., доцент Бабич М.І.

(Прізвище та ініціали)

ДУБЛЯНИ-2022

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ПРИРОДОКОРИСТУВАННЯ
ФАКУЛЬТЕТ МЕХАНІКИ, ЕНЕРГЕТИКИ ТА ІНФОРМАЦІЙНИХ
ТЕХНОЛОГІЙ
КАФЕДРА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

Другий (магістерський) рівень вищої освіти
Спеціальність 126 «Інформаційні системи та технології»

«ЗАТВЕРДЖУЮ»

Завідувач кафедри _____

д.т.н., проф. А.М. Тригуба

« ____ » _____ 2022 р.

ЗАВДАННЯ

на кваліфікаційну роботу студенту

Литвину Олегу Руслановичу

1. Тема роботи: «Інформаційна система прогнозування вартості транспортування насіння зернових культур»

Керівник роботи Железняк Алла Михайлівна, доцент
затверджені наказом по університету від 12.05.2022 року № 62/к-с.

2. Строк подання студентом роботи 15.12.2022 р.

3. Вихідні дані до роботи: база даних щодо процесів транспортування зернових культур та їх вартості; алгоритми машинного навчання; методика дослідження моделей машинного навчання; методика проектування інформаційних систем.

4. Зміст розрахунково-пояснювальної записки (перелік питань, які необхідно розробити)
Вступ.

1. Аналіз стану використання інформаційних систем під час транспортування вантажів та завдання кваліфікаційної роботи.

2. Обґрунтування інструментарію для інформаційної системи прогнозування вартості транспортування насіння зернових культур.

3. Результати проектування інформаційної системи прогнозування вартості транспортування насіння зернових культур.

4. Охорона праці та безпека у надзвичайних ситуаціях.

5. Визначення ефективності від використання інформаційної системи.
Висновки та пропозиції.

Список використаної літератури.

5. Перелік ілюстраційного матеріалу (з точним зазначенням обов'язкових слайдів): аналіз інформаційних систем для планування автомобільних вантажних перевезень; особливості використання машинного навчання для транспортних підприємств; вибір методу машинного навчання; результати підготовки та аналізу даних для прогнозування вартості транспортування насіння зернових культур; масштабування функцій та навчання моделі; компоненти інформаційної системи; структурно-функціональне моделювання діяльності інформаційної системи; результати проектування головного вікна користувача; економічна ефективність.

6. Консультанти з розділів:

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
1, 2, 3, 5	<i>Желсзняк А.М., доцент кафедри інформаційних технологій</i>		
4	<i>Городецький І.М., доцент кафедри управління проектами та безпеки виробництва</i>		

7. Дата видачі завдання

12 травня 2022 р.

Календарний план

№ з/п	Назва етапів кваліфікаційної роботи	Терміни виконання етапів роботи	При-мітка
1	<i>Написання першого розділу</i>	<i>12.05-20.06.22</i>	
2	<i>Виконання другого розділу та аркушів ілюстраційного матеріалу до нього</i>	<i>21.06-14.08.22</i>	
3.	<i>Виконання третього розділу та аркушів ілюстраційного матеріалу до нього</i>	<i>15.08-31.10.22</i>	
4.	<i>Написання розділу «Охорона праці та безпека у надзвичайних ситуаціях»</i>	<i>01.11-10.11.22</i>	
5.	<i>Оцінення ефективності запропонованої системи</i>	<i>20.11-30.11.22</i>	
6.	<i>Завершення оформлення розрахунково-пояснювальної записки та аркушів ілюстраційного матеріалу</i>	<i>01-04.12.22</i>	
7.	<i>Завершення роботи в цілому</i>	<i>05-15.12.22</i>	

Студент _____ Литвин О.Р.
(підпис)

Керівник роботи _____ Желсзняк А.М.
(підпис)

УДК 004.9 : 656.033

Інформаційна система прогнозування вартості транспортування насіння зернових культур.

Литвин О.Р. Кафедра інформаційних технологій – Дубляни, ЛНУП, 2022.

Кваліфікаційна робота: 77 с. текст. част., 15 рис., 5 табл., 12 арк. ілюстраційного матеріалу, 34 джерела.

Проведено аналіз стану перевезень вантажів в Україні та існуючих інформаційних систем для планування залізничних вантажних перевезень. Проаналізовано існуючі інформаційні системи для планування автомобільних вантажних перевезень. Подано особливості використання машинного навчання для транспортних підприємств. Сформульовано завдання кваліфікаційної роботи. Наведено особливості вирішення задачі прогнозування вартості транспортування насіння зернових культур на підставі машинного навчання. Здійснено опис підготовки та аналіз даних для прогнозування вартості транспортування насіння зернових культур. Вибрано метод машинного навчання для прогнозування вартості транспортування насіння зернових культур та інструментарій мови Python для побудови лінійних регресій.

Подано результати підготовки та аналізу даних для прогнозування вартості транспортування насіння зернових культур. Здійснено масштабування функцій та навчання моделі. Підібрано компоненти інформаційної системи прогнозування вартості доставки насіння зернових. Виконано структурно-функціональне моделювання діяльності інформаційної системи. Подано результати проектування головного вікна користувача.

Розроблено заходи стосовно охорони праці та безпека у надзвичайних ситуаціях. Визначено ефективність використання інформаційної системи прогнозування вартості транспортування насіння зернових культур.

ЗМІСТ

ВСТУП.....	7
1. АНАЛІЗ СТАНУ ВИКОРИСТАННЯ ІНФОРМАЦІЙНИХ СИСТЕМ ПІД ЧАС ТРАНСПОРТУВАННЯ ВАНТАЖІВ ТА ЗАВДАННЯ КВАЛІФІКАЦІЙНОЇ РОБОТИ.....	9
1.1. Стан перевезень вантажів в Україні	9
1.2. Аналіз існуючих інформаційних систем для планування залізничних вантажних перевезень.....	10
1.3. Аналіз існуючих інформаційних систем для планування автомобільних вантажних перевезень	14
1.4. Використання машинного навчання для транспортних підприємств.....	21
1.5. Завдання кваліфікаційної роботи.....	22
2. ОБҐРУНТУВАННЯ ІНСТРУМЕНТАРІЮ ДЛЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ ПРОГНОЗУВАННЯ ВАРТОСТІ ТРАНСПОРТУВАННЯ НАСІННЯ ЗЕРНОВИХ КУЛЬТУР.....	23
2.1. Особливості вирішення задачі прогнозування вартості транспортування насіння зернових культур на підставі машинного навчання.....	23
2.2. Підготовка та аналіз даних для прогнозування вартості транспортування насіння зернових культур	26
2.3. Вибір методу машинного навчання для прогнозування вартості транспортування насіння зернових культур.....	30
2.4. Продуктивність лінійної регресії та її різновиди.....	31
2.4.1. Проста лінійна регресія.....	32
2.4.2. Множинна лінійна регресія.....	33
2.4.3. Поліноміальна лінійна регресія.....	34
2.5. Інструментарій мови Python для побудови лінійних регресій.....	35
3. РЕЗУЛЬТАТИ ПІДГОТОВКИ ДАНИХ ТА РОЗРОБКИ ІНФОРМАЦІЙНОЇ СИСТЕМИ ПРОГНОЗУВАННЯ ВАРТОСТІ ТРАНСПОРТУВАННЯ НАСІННЯ ЗЕРНОВИХ КУЛЬТУР.....	40

3.1. Результати підготовки та аналізу даних для прогнозування вартості транспортування насіння зернових культур.....	40
3.2. Масштабування функцій та навчання моделі.....	46
3.3. Компоненти інформаційної системи прогнозування вартості доставки насіння зернових культур.....	51
3.4. Структурно-функціональне моделювання діяльності інформаційної системи.....	53
3.5. Результати проектування головного вікна користувача.....	56
4. ОХОРОНА ПРАЦІ ТА БЕЗПЕКА У НАДЗВИЧАЙНИХ СИТУАЦІЯХ	58
4.1. Аналіз небезпечних і шкідливих виробничих чинників та розробка заходів щодо покращення умов праці	58
4.2. Розробка логічно-імітаційної моделі травматизму під час монтажу інформаційної систем	58
4.3. Розробка заходів щодо безпеки у надзвичайних ситуаціях	63
6. ВИЗНАЧЕННЯ ЕФЕКТИВНОСТІ ВІД ВИКОРИСТАННЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ.....	65
ВИСНОВКИ І ПРОПОЗИЦІЇ.....	69
СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ.....	74

ВСТУП

У сучасному світі інформаційні системи призначені для вирішення конкретних завдань практичного характеру. При побудові організаційних процесів логістичний підхід може допомогти підвищити стабільність і знизити витрати в цілому, особливо на окремі ресурси.

Структура інформаційного потоку в логістиці складна і насичена даними. Це важливо для прийняття правильних управлінських рішень, постійного моніторингу та аналізу. Транспортна інфраструктура значною мірою впливає на процес виробництва – його тривалість, частоту і розмір поставок, місткість складів тощо. Тому інформаційні системи розробляються під конкретне підприємство, створюючи умови для оптимальної реалізації процесу та встановлюючи межі розширення діяльності на конкретний період роботи. Завдання інформаційних систем полягає в тому, щоб максимально швидко і точно використовувати існуючі можливості транспортної інфраструктури.

Інформаційні системи у транспортування зернових культур створюють можливості для аналізу та відбору даних, швидкого доступу та перевірки стану доставки зернових культур, планування процесів, зокрема прогнозування вартості транспортування насіння зернових культур. Інформаційні системи інтегровані у діяльність транспортних підприємств, отже, вибираються відповідно до конкретного транспортного підприємства, його масштабів, технічних і фінансових можливостей.

Логістична інфраструктура транспортних підприємства в цілому, потребує розробки інформаційних систем прогнозування вартості транспортування насіння зернових культур. Використання зазначених інформаційних систем під час транспортування зернових культур може вирішити наступні задачі:

- забезпечити своєчасну доставку зернових культур;
- знизити транспортні витрати під час доставки зернових культур;
- володіти своєчасною інформацією щодо процесу доставки зернових культур;

- пришвидшити виконання замовлень щодо доставки зернових культур.

Сьогодні багато транспортних підприємств намагаються впровадити передові рішення машинного навчання та індивідуальні рішення обчислювального інтелекту, щоб залишатися конкурентоспроможними на ринку. Аналітика даних при цьому необхідна для покращення результатів і підвищення ефективності проєктованих інформаційних систем. Завдяки машинному навчанню та обчислювального інтелекту легко розробити інформаційні системи прогнозування вартості транспортування насіння зернових культур, які допоможуть вам успішно вести транспортний бізнес.

Усе сказане вказує на те, що виконана кваліфікаційна робота «Інформаційна система прогнозування вартості транспортування насіння зернових культур» має актуальне теоретичне та практичне значення.

Об'єктом дослідження є процес транспортування насіння зернових культур та алгоритми машинного навчання, які забезпечують вирішення задач прогнозування.

Предмет дослідження є залежність якості прогнозування вартості транспортування насіння зернових культур від вибраних алгоритмів машинного навчання.

Метою роботи є підвищення ефективності процесу транспортування насіння зернових культур завдяки розробці інформаційної системи, яка базується на обчислювальному інтелекті.

РОЗДІЛ 1.

АНАЛІЗ СТАНУ ВИКОРИСТАННЯ ІНФОРМАЦІЙНИХ СИСТЕМ ПІД ЧАС ТРАНСПОРТУВАННЯ ВАНТАЖІВ ТА ЗАВДАННЯ КВАЛІФІКАЦІЙНОЇ РОБОТИ

1.1. Стан перевезень вантажів в Україні

На даний час добре функціонуючий транспортний комплекс України є важливим для стабільності країни та її безпеки, а також забезпечення потреб економіки. Україна має розвинуту транспортну систему, яка забезпечує перевезення понад 621 млн тонн вантажів [15]. При цьому близько 87% цих транспортних перевезень належить залізничному та автомобільному транспорту (рис. 1).



Рис. 1.1. Структура перевезень вантажів окремими видами транспорту

Одним із основних напрямків розвитку транспортної галузі, визначених Національною транспортною стратегією [19], є впровадження сучасних інформаційних технологій та електронного документообігу в транспортному процесі. Водночас, за даними 2021 року [4], лише 17,6% українських компаній, зайнятих у транспорті, складському господарстві, пошті та кур'єрській діяльності, користуються Інтернетом, з них 21,7% мають швидкісний доступ 30 Мбіт/с. Крім того, лише 21,4% зареєстрованих компаній мають власний веб-

сайт, з яких лише 6,3% дозволяють обробляти замовлення онлайн і лише 5,7% дозволяють відстежувати статус виконання. Ця статистика може опосередковано свідчити про відносно обмежене використання сучасних інформаційних технологій для допомоги та електронного обміну даними при транспортуванні вантажів в Україні.

Останні досягнення в інформатизації економічної діяльності та нові розробки в області штучного інтелекту, блокчейну, Інтернету речей, автоматизації та хмарних технологій стають все більш важливими для транспорту. Це забезпечує вдосконалення існуючих процесів, створення нових можливостей для бізнесу, трансформації ланцюжків поставок і бізнес-сфер.

Отже, використання інформаційних систем для планування та прогнозування діяльності транспортних підприємств є досить актуальним на даний час.

1.2. Аналіз існуючих інформаційних систем для планування залізничних вантажних перевезень

На даний час для регулювання транспортно-експедиційної діяльності на залізницях України (УЗ) використовується Єдина автоматизована система управління транспортно-експедиційною діяльністю Укрзалізниці (АСК ВП УЗЕ). Система об'єднує інформаційну базу шести регіональних залізничних вузлів:

- виконує контроль та аналіз стану поїзда;
- формує розклад руху поїздів;
- планує змінний добовий розклад;
- забезпечує технічний регламент управління виконанням перевезень;
- здійснює контроль за наявністю, станом і експлуатацією вагонного парку;

- описує та організовує роботу локомотивів і локомотивних бригад.

Зокрема, нею виконується аналіз економічних показників транспортного процесу. Ця система може бути використана для створення кількох звітів (довідок), необхідних для оперативного управління залізничними операціями та планування. При цьому більшість інформації, що відображається, доступна лише працівникам залізниці через АРМ, що ускладнює обмін оперативною інформацією з вантажовласниками та експедиторами.

В даний час документообіг і обмін даними між власником вантажу і залізницею здійснюється в електронному вигляді за допомогою залізничних інформаційних систем і кількох комерційних програмних продуктів. Для планування перевезень вантажів залізницею в обов'язковому порядку виконується попереднє планування за допомогою інформаційної системи «МЕСПЛАН» [1].

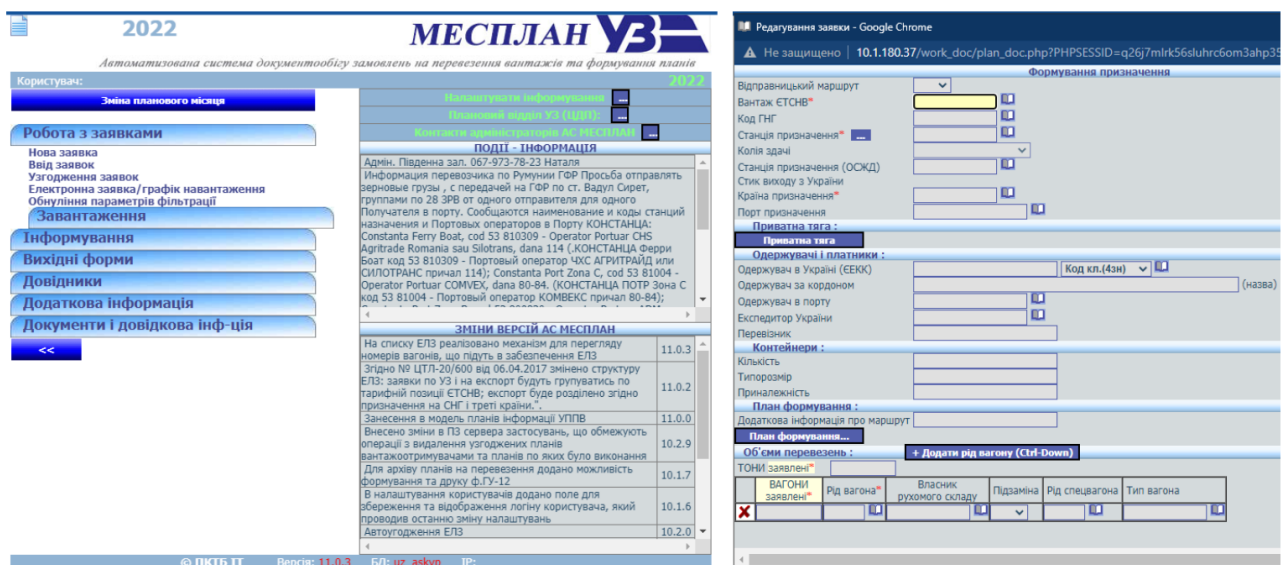


Рис. 1.2. Інтерфейс користувача інформаційної системи «МЕСПЛАН»

Інтерфейс інформаційної системи «МЕСПЛАН» забезпечує огляд переліку заяв (записів, погоджень) на плановий місяць; забезпечує виготовлення облікових карток; здійснює базовий контроль за правильністю введених даних, а також забезпечує формування координаті заявки на рівні УЗ.

Договір на організацію залізничних перевезень та надання послуг з оформлення транспортної та супутньої документації укладається за допомогою

інформаційної системи «Клієнт УЗ» (рис. 1.3). Він доступний як для власників вантажів, так і для експедиторів в онлайн-режимі через інтернет-сторінки.

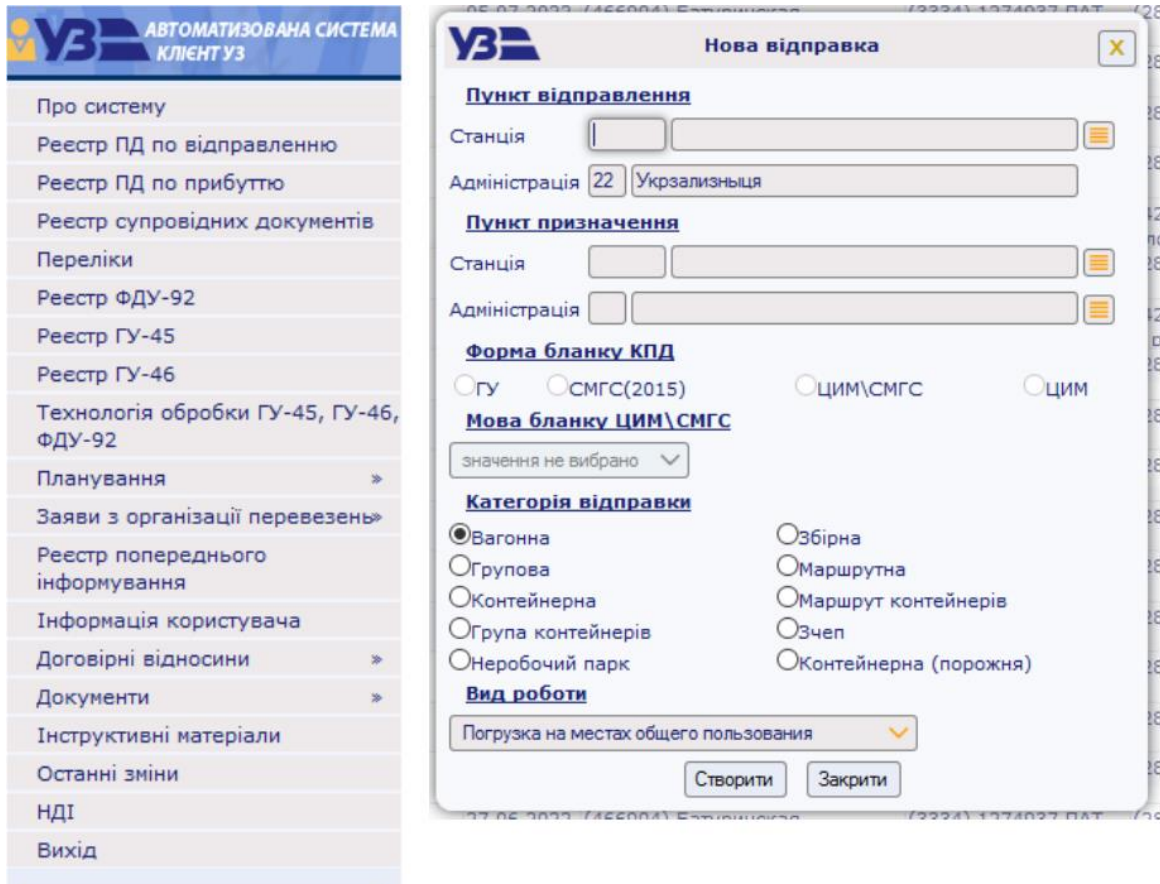


Рис. 1.3. Інтерфейс користувача інформаційної системи «Клієнт УЗ»

Інформаційна система «Клієнт УЗ» може виконувати такі операції [1]:

- загальне прийняття пропозиції (акцепту) договору перевезення;
- створення електронних даних транспортного документа;
- формування транспортного документа з використанням електронних даних;
- передача електронних транспортних документів до ВП УЗ-Е АСК;
- забезпечує отримати паперові транспортні документи або електронні дані електронних транспортних документів для всіх осіб на етапі оформлення;
- додаткова робота з нормативно-довідковою інформацією.

Завдання з планування доставки та інтеграція обробки транспортних документів виконуються у платних сторонніх системах Freight.

АРМ «Вантажовідправника» [20] (рис. 1.4) має спрощений інтерфейс. Теперішня реалізація АРМ «Вантажовідправника» дозволяє працювати з

даними в автономному режимі з вбудованим функціоналом та електронними архівами документів.



Сервіс Довідки Допомога

APM
ВАНТАЖОВІДПРАВНИКА

Виправка №32039596 (збоюму) Збірня ГУ шв. Грузовий

Статус документа: Розподілено 17.01.2014 09:49:12 | Сторінка: 1 | Тестовий клієнт | Перейти П. 1

Відправник: (1860) ТОВ ТЕХ "ТРАНС-МАГІСТР"	Район: 2600208159	Експедитор:	Одержувач: (1860) ТОВ ТЕХ "ТРАНС-МАГІСТР"	Район: 2600208159
Платіж: (6408541) ТОВ ТЕХ "ТРАНС-МАГІСТР"	Експедитор:	Платіж: (6408541) ТОВ ТЕХ "ТРАНС-МАГІСТР"	Експедитор:	
Станція відправлення: (322407) Світлоино	(22) УЗ	План навантаження: 04	Станція призначення: (320505) Киев-Вольноий	(22) УЗ
Станція навантаження:	Траноплан №:	Вантаж одержав: 7458	за довіреністю №: 1111	від: 03.01.2014

Вид роботи станції відправлення: Торгує на путях клієнту

Вид роботи станції призначення: Витрує на путях клієнту

Прийнято до перевезення: 17.01.2014 09:41

Термін доставки: 20.01.2014 00:00

Прйбути: 17.01.2014 09:45

Повідомлення: 17.01.2014 09:45

Розподілено: 17.01.2014 09:48

Грузовий склад та закріплено-бульовий реєстр

№ вагону	Рід	Об'	Тара	в/л, т	Адм.	Приналежність	Використовується	№	Тип ЗПП	Знак	Власник ЗПП	Станція накладення ЗПП
55549521	Дитячий	4	27500	60.0	(22) Учр...	собственный	підгруз	1	Варта-Укр...	П748596	Отправитель	(322407) Світлоино

Вантаж

№	Вантаж	Упаковка	Кількість	Маса, кг
1	(232395) Щобель гранцій не пошкодований в алаците	(1) Навалом (насилля твердые мелкие частицы шор...	120	21000
2	(415141) Масариби ручные бытовые	(2) Насыпью, твердые гранулированные частицы (ге...	160	18000
Всього 2			280	39000

Заяви і відписки відправника

Відомості про вантаж та особливості його перевезення

Країна відправлення: (804) Україна

Країна призначення: (804) Україна

Спосіб визначення маси: Негабаритність вантажу

Відомості про навантаження: Знаки / Марки / Номери

Оголошена ціна вантажу: Оцінка попереднього оцілювача

Осорожа:

Додаткові відомості:

Клас/наклад небезпечки: N/AK Код ООН

Від-ітти записки

Відомості вносення записки: Марка записки

Відомості вносення записки:

Код Назва штемпеля/відписки Текст штемпеля/відписки

611 Власний вагон. Власник: ... Власний вагон. Власник: ПАО "Вольно-Це..."

705 Упаковка вантажу НАСЫПЬ

208 Вантаж розміщено й закріплено згідно з в... Вантаж розміщено й закріплено згідно з 4...

Розрахунок за перевезення

Платіж за розрахунок: Вид розрахунку: Тарифна сітка: Тарифна вартість: Код виключеного тарифу

Платіж при відправленні: На станції отримання через 2.0 5

Код	Найменування платіж	Сума	Валюта	ГУ57
001	Провізна плата, сплачена при відправленні	753.00	ГРИВНЯ	
Всього сплачено при відправленні		753.00	ГРИВНЯ	

Платіж по прибутті

Код	Найменування платіж	Сума	Валюта	ГУ57
Всього сплачено при відправленні			ГРИВНЯ	

Рис. 1.4. Інтерфейс користувача АРМ «Вантажовідправника»

Однак слід зазначити, що система АРМ «Вантажовідправника» лише частково може вирішити проблему обміну інформацією між підприємствами та системами автоматизації залізниці. Дані, імпортовані з інформаційної системи «Клієнт УЗ», можна лише вручну передати у бази даних підприємства через файли XML. Здійснення повного обміну інформацією між системами досягається спеціальним продуктом Coordination Module Server (SMU). Цей

модуль складається з серверної частини (функція зв'язку з інформаційною системою «Клієнт УЗ») та клієнтської частини (документообіг).

1.3. Аналіз існуючих інформаційних систем для планування автомобільних вантажних перевезень

Однією з причин впровадження сучасних інформаційних технологій на ринку автомобільних вантажних перевезень є розвиток онлайн транспортних бірж. Загалом біржі доставки – це онлайн-агрегати інформації про доставку та поточні пропозиції вантажів і безкоштовної доставки. Користувачами цих бірж можуть бути прямі перевізники (транспортні компанії), товаровласники (виробники та торгові компанії), експедитори, юридичні чи фізичні особи.

Наразі в Україні є десятки транзитних бірж, які мають відкритий або обмежений доступ. Найпопулярнішими українськими онлайн-біржами є Lardi-Trans.com, Della.ua, Degruz.com, Lading.eu, особливо популяризують те, що вони пропонують, такі іноземні ресурси, як Trans.eu, Timocom.ru, Transinfo.by, Cargo.lt.

Більшість цих онлайн-сервісів розповсюджені на внутрішньому та міжнародному ринках вантажних перевезень. За даними опитування [2], платформа Lardi-trans.com є провідною українською транспортною онлайн-біржею та користується популярністю серед вантажовідправників та перевізників України та Прибалтики.

The screenshot shows the Lardi-Trans website interface. At the top, there is a navigation bar with the LARDI logo and menu items: Вантаж, Транспорт, Послуги та ціни, Зона надійності, Страхування, Форум, Корисне. There are also buttons for 'Додати Вантаж' and 'Додати Транспорт'. Below the navigation bar, there is a search bar with the text 'Пошук' and a 'Форум' dropdown menu. The main content area is titled 'Пошук транспорту для вантажоперевезень'. It contains two columns of input fields for search criteria: 'Крайна/Країни звідки' and 'Крайна/Країни куди', both with 'Будь-яка область' and 'Будь-яке місто' options and a '+радіус, км' button. Below these are 'Маса' and 'Об'єм' sections, each with 'Від' and 'До' fields and units (т and м³). There are also 'Дата з' and 'Дата по' fields. At the bottom, there are buttons for 'Розширений пошук' and 'Знайти'. On the left side, there is a promotional banner for 'ЗНИЖКИ ДО 50%' with images of discount cards and the text 'ПЕРЕГЛЯНУТИ ВСІ ПРОПОЗИЦІЇ'.

Рис. 1.5. Онлайн платформа «Lardi-trans»

Ця платформа (рис. 1.5) існує з 1999 року. Щодня на сайт компанії додається понад 50 000 актуальних пропозицій. При цьому не менше 50% замовлень спрямовується на перевезення в Україні, а іншу половину – на міжнародні перевезення, зокрема до Європи, Центральної Азії та країн СНД. Платформа «Lardi-Trans» публікує транспортні пропозиції, які зараз використовуються для вантажних перевезень, і рейтинг успішності компанії [14].

Рейтинги складаються на основі відгуків про партнерство (як позитивних, так і негативних), результатів оцінки юридичних документів, дати реєстрації та історії змін на сайті компанії. Однією з переваг цієї платформи є те, що на ній є форум на актуальні теми транспортної галузі та «чорний список» клієнтів і перевізників, який постійно оновлюється та заслуговує на особливу увагу. Крім того, вбудовані інструменти моніторингу GPS, такі як SmartGPS [22] і TrucksNearMe [34], надають безкоштовні послуги пошуку вантажівок на картах у режимі реального часу та доступні для зареєстрованих користувачів платформи. Інші переваги транзитної біржі «Ларді-транс» включають онлайн-послуги страхування вантажів і відповідальності перевізника, а також можливість проведення тендерів. Реєстрація необхідна для доступу користувачів до всіх функцій онлайн-біржі, включаючи вибір типу реєстрації

(фізична особа, підприємець або юридична особа) та виду діяльності (перевізник, вантажовідправник, експедиція).

Реєстрація та публікація додатків на сайті безкоштовна, мінімальна вартість доступу до контактної інформації для додатків та профілів користувачів по Україні становить 150 грн на місяць.

Другою за популярністю в Україні є транзитна біржа Della.ua, яка працює з 1995 року[30]. Ця транспортно-інформаційна платформа не тільки доступна на території України, але й доступна в Казахстані, Узбекистані та інших країнах Європи. Діяльність Della Exchange обмежується публікацією та переглядом поточних запитів на доставку вантажів. Інформацію про потенційних клієнтів на сайті можна шукати лише за назвою компанії, податковим номером, номером телефону чи електронною адресою. Однак вона не дає можливості відобразити відгуки про співпрацю із користувачами Della Exchange.

Четвер 17 листопада 2022 р.

Україна укр

Вхід або Реєстрація

Автоперевезення

додати вантаж | додати транспорт | пошук вантажів і транспорту

Вантажі Україна | Міжнародні Вантажі | Транспорт України | Міжнародний Транспорт

Вантажі для перевезення Україна - Україна [показати всі вантажі Україна - Україна](#)

* – безпосередньо Замовник автоперевезення (не надає експедиторські або диспетчерські послуги)

17.11 12:01	* Бровари (UA) — Маньківка (UA) ~ 218 км, тент, можл. дозавантаження	метал на палетах будь-яка	1 т	
17.11—18.11 12:01	Благодатне (UA) — Мостиська (UA) ~ 355 км, тент	тнв	3 т	30 м³
17.11 12:01	* Дніпро (UA) — Луцьк (UA) ~ 886 км, крита, кільк. палет: 2, тип палет: EUR 1,2 x 0...	запчастини на п... 4 000 грн готівка	2,4 т	2 м³

Вантажні перевезення України. Статистика цін **тент 20 тонн**

тент 20т, ціна грн/км

лис гру січ лют бер кві тра чер лип сер вер жов

— графік зміни ціни — тренд

18.11 12:01	Харків (UA) — Луцьк (UA) ~ 876 км, крита	комплектуючі 4 700 грн готівка, при розвантаж...	1,3 т	6 м³
17.11 12:01	Хмельницький (UA) — Полтава (UA) ~ 663 км, крита	тнв	0,6 т	3 м³
17.11 12:01	* Калинівка (UA) — Шостка (UA) ~ 341 км, крита	труби металеві 3 000 грн готівка	0,25 т	2 м³
17.11 12:01	Баранівка (UA) — Житомир (UA) ~ 77 км, крита	вагонка 2 000 грн	2 т	3 м³
17.11 12:01	* Шостка (UA) — Харків (UA) ~ 371 км, крита	пластикові вироби	6 т	45 м³

Потрібен перевізник?
Безкоштовно!
Додайте ваш вантаж і тисячі перевірених перевізників зможуть зв'язатися з вами **прямо зараз!**
[ДОДАТИ ВАНТАЖ](#)

Курси валют України
USD **37,02 грн**
EUR **38,45 грн**
[Курс валют](#)

Пошук попутних вантажів
Звідки: Україна (UA)
всі обл.
Куди: Польща (PL)
всі обл.
[Знайти](#)

Пошук попутних машин
Звідки: Німеччина (DE)
всі обл.
Куди: Україна (UA)
всі обл.
[Знайти](#)

Рис. 1.6. Онлайн платформа «Della Exchange»

При цьому платформа Della дозволяє відстежувати динаміку ціноутворення на внутрішні вантажні перевезення та популярні міжнародні напрямки, на відміну від Ларді-транс. Ще однією особливістю платформи є можливість розпізнавати заявки на доставку від прямих вантажовласників, недоступні на інших онлайн-біржах.

Доступ до замовлень обмежений реєстрацією, а доступ до контактної інформації для кожного замовлення враховує лише підписку на тарифний план, ціна від 130 грн на місяць.

Ще однією українською транспортною біржею є платформа Degruz.com [29], заснована в 2006 році і позиціонується як біржа, що працює переважно з прямими вантажовласниками. Вся інформація та послуги на сайті є безкоштовними та можуть використовуватися користувачем за винятком зв'язку з власником товару та перевізником. Контакти компанії починаються від 45 грн на місяць в залежності від обраних умов оплати та типу користувача.

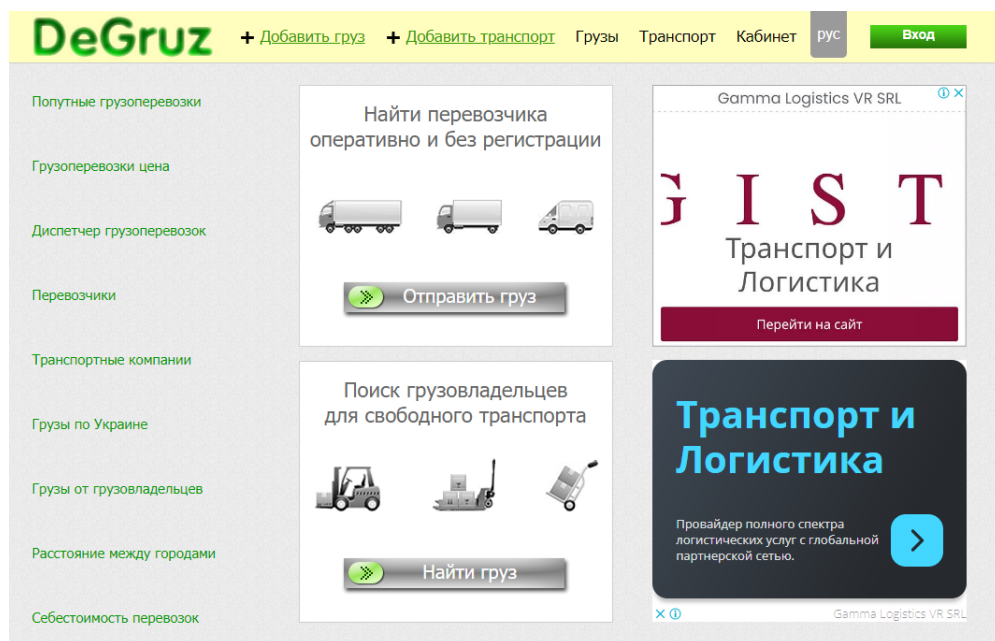


Рис. 1.7. Онлайн платформа «Degruz»

Мінусом даної біржі є повна відсутність сервісів, крім калькуляторів для розрахунку вартості проїзду вантажівок, статистики вартості проїзду для різних типів рухомого складу та відстані між населеними пунктами. Загалом інтерфейс

і можливості цього обміну застаріли і не користуються попитом серед вантажовласників і перевізників, але це доступний тарифний план.

У 2019 році запустили в Україні транспортну платформу нового покоління Lading.ua [32], яка зарекомендувала себе як сучасний і зручний український сервіс для пошуку товарів і транспорту з близько 6 тисячами користувачів.



Рис. 1.8. Онлайн платформа «Lading»

Платформа Lading – це сучасне SaaS-рішення для управління аутсорсинговими автомобільними перевезеннями, що значно заощаджує витрати внутрішніх і міжнародних транспортних організацій.

Різниця між транспортною платформою доставки полягає в тому, що приймаються та переглядаються лише замовлення на вантажні перевезення, а пропозиції безкоштовного транспортування не можна додавати. Слід зазначити, що всі операції в транспортних додатках повністю автоматизовані та оцифровані [7]. Система самостійно розраховує відстань, об'єм і масу і допомагає визначити вартість кожного відправлення. Усі учасники інформаційної системи доставки та платформи Lardi-trans проходять багаторівневу перевірку після первинної реєстрації та постійно контролюються службами сервісної інспекції. Для доступу до контактної інформації кожної програми достатньо зареєструватися на платформі. Платформа безкоштовна і

до кінця 2022 року не стягується абонплата за користування сервісом. Комісія стягується лише тоді, коли вантажовідправник подає заявку на сайті, приймає пропозицію про роботу перевізника з перевагою та анулює прийнятну пропозицію про роботу. Вартість цих послуг коливається у межах 10...20 грн. на запит.

Крім того, відома українська онлайн-платформа Sovtes.ua [33], якою користуються більшість компаній, таких як Метінвест, Інтерпайп, ArcelorMittal, Centraviv, GlobalSpirits, АТБ, Varus, Agrosem та інші. Ці платформи мають усі функції онлайн-обмінів, швидкий електронний документообіг, розвантаження в точці завантаження тощо.

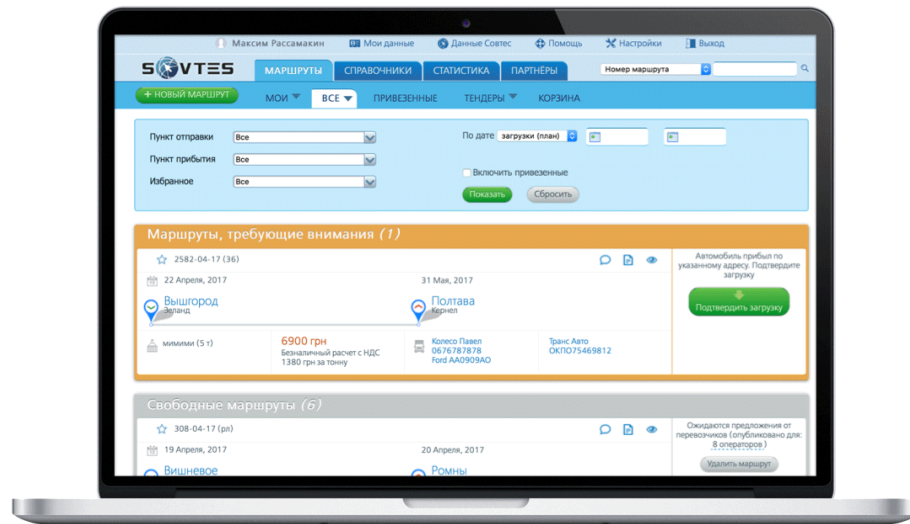


Рис. 1.9. Онлайн платформа «Sovtes»

Сильна сторона платформи «Sovtes» полягає в автоматизації всього логістичного процесу аж до відділення. На платформі «Sovtes» користувачі також мають багато можливостей для аналізу основних показників роботи, формування звітів за категоріями, відстеження місце знаходження транспортних засобів під час транспортування тощо.

Наприклад, клієнти та перевізники мають доступ до маршруту транспортування, можуть переглядати час навантаження та розвантаження, терміни митного оформлення, перетину кордону, непродуктивні простой тощо. Однак усі можливості, доступні в сервісі Sovtes, доступні лише за умови підключення користувача до тарифного плану від 425 грн/міс.

Таблиця 1.1. Порівняльні характеристики онлайн сервісів щодо вантажних автомобільних перевезень на українському ринку

Функціонал	Інформаційна система				
	Lardi-Trans	Della	Degruz	Lading	Sovtes
Пропозиції вантаж вільний транспорт	Δ	Δ	Δ	©	©
	Δ	Δ	Δ		©
Перегляд пропозиції вантаж вільний транспорт	Δ	Δ	Δ	Δ	©
	Δ	Δ	Δ		©
Пропозиції міжнародних перевезень	©	©	©	Δ	©
Контактна інформація (пропозиції, користувачі)	©	©	©	Δ	©
Тендери	©				©
Електронний документообіг	©				©
Рейтинги компаній	©	©		Δ	©
Відгуки та претензії	©				©
Інтеграція з корпоративними ІС через API	©				©
Розрахунок пройденого шляху	Δ	Δ	Δ		©
Мобільні додатки	Δ				
Додаткові функції	GPS-трекінги, страхування, месенджер, зворотній зв'язок, форум	статистичні дані, вартість доставки		–	статистика та звітність, месенджер, зворотній зв'язок
Вартість тарифного плану, грн/місяць	150...3500	130...375	45...100	10...20 грн/заяв ка	425...500

Примітка: Δ – зареєстрований користувач; © – зареєстрований користувач із підписаним тарифним планом

Слід зазначити, що в більшості випадків великі виробничі компанії, роздрібні продавці та дистриб'ютори зобов'язані використовувати визнані

платформи електронних торгів, такі як ProZorro, SmartTender, Salesbook тощо для тендерів на транспортні послуги.

Порівняльні характеристики онлайн сервісів щодо вантажних автомобільних перевезень на українському ринку представлено в таблиці 1.1.

1.4. Використання машинного навчання для транспортних підприємств

Ринок інформаційних систем та технологій із використанням машинного навчання (ML) стрімко зростає. Починаючи із 2016 року він перевищив \$1 млрд та, за прогнозами аналітиків, може зрости до \$39,98 млрд до 2025 року.

Наприкінці 2016 року компанії «MIT Technology Review» та «Google Cloud» виконали спільні дослідження щодо машинного навчання. Вони встановили, що машинне навчання дає новий спосіб отримати конкурентну перевагу. Було опитано декілька сотень респондентів із усього світу, які дочетні та мають досвід роботи у великих та малих компаніях із різного напрямку (транспорт, промисловість, сфера послуг, фінанси тощо). Встановлено, що 60 % цих компаній уже використовують ML, і одна третина цих компаній перейшли від інноваційної технології до зрілості із використанням ML. Крім того, 26% компаній вже отримують конкурентну перевагу завдяки ML. Четверть компаній інвестували понад 15% своїх доларів на розвиток інформаційних систем та технологій у ML і повернули більшу частину своїх інвестицій.

Машинне навчання, особливо нейронні мережі, слід використовувати для вирішення задач транспортних підприємств, коли:

- накопичується велика кількість різноманітних даних, але немає програми, яка б їх обробила та систематизувала;
- доступні дані викривлені, неповні або неструктуровані;
- дані настільки розрізнені, що між ними важко помітити зв'язки та закономірності;

- існують задачі транспортних підприємств, які можна вирішити за допомогою машинного навчання та нейронних мереж;
- є можливість виконати прогноз: попиту, обсяг перевезень, вартість надання транспортних послуг;
- є можливість виявлення тенденції, приховані зв'язки, аномалії, елементи, що повторюються.

1.5. Завдання кваліфікаційної роботи

Проведений аналіз стану розробки та використання інформаційних систем для транспортних підприємств дав можливість встановити, що існує актуальна задача прогнозування вартості транспортування насіння зернових культур. Для цього слід розробити інформаційну систему, яка базується на машинному навчанні та забезпечить зростання точності прогнозування і усуне недоліки існуючих інформаційних систем. Це зумовлює необхідність розроблення інформаційних систем прогнозування вартості транспортування насіння зернових культур. У виконаній роботі слід розв'язати такі завдання:

- провести аналіз стану використання інформаційних систем під час транспортування вантажів та сформулювати завдання кваліфікаційної роботи;
- обґрунтувати інструментарій для інформаційної системи прогнозування вартості транспортування насіння зернових культур;
- вибрати ефективну модель машинного навчання для прогнозування вартості транспортування насіння зернових культур;
- виконати проектування інформаційної системи прогнозування вартості транспортування насіння зернових культур;
- розробити заходи із охорони праці, а також безпеки у надзвичайних ситуаціях;
- визначити економічну ефективність від використання інформаційної системи.

РОЗДІЛ 2.

ОБҐРУНТУВАННЯ ІНСТРУМЕНТАРІЮ ДЛЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ ПРОГНОЗУВАННЯ ВАРТОСТІ ТРАНСПОРТУВАННЯ НАСІННЯ ЗЕРНОВИХ КУЛЬТУР

2.1. Особливості вирішення задачі прогнозування вартості транспортування насіння зернових культур на підставі машинного навчання

Прогностична аналітика – це аналіз поточних і історичних даних для прогнозування ймовірності майбутніх подій, результатів або значень у контексті прогнозування вартості транспортування насіння зернових культур. Для прогнозової аналітики потрібні численні статистичні методи, такі як інтелектуальний аналіз даних (виявлення закономірностей у даних) і машинне навчання.

Метою машинного навчання є створення систем, здатних знаходити шаблони в даних, навчатися на них без втручання людини та явного перепрограмування. Щоб вирішити задачу прогнозування вартості транспортування насіння зернових культур, дослідники з обробки даних спочатку повинні зрозуміти, які дані використовувати для навчання моделей машинного навчання, і саме тому потрібна описова аналітика.

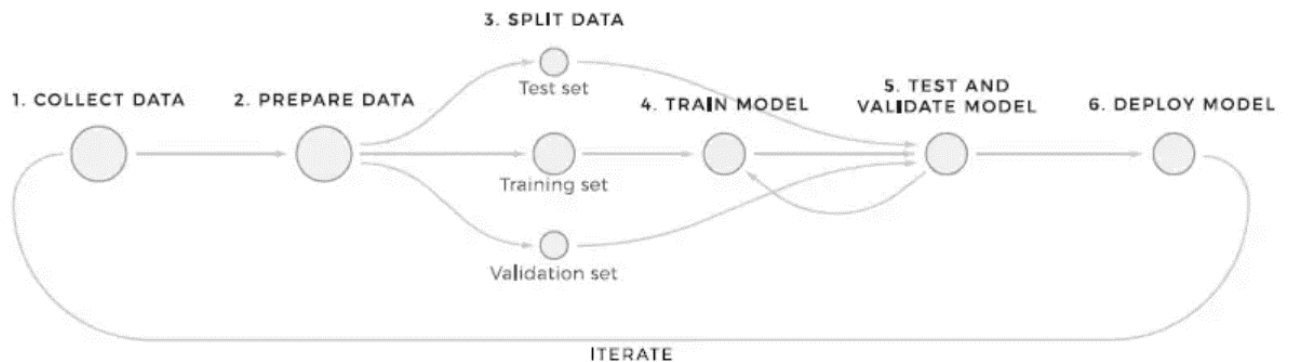


Рис. 2.1. Структура процесу машинного навчання для прогнозування вартості транспортування насіння зернових культур

На початковому етапі слід зібрати, відібрати, підготувати, попередньо обробити і трансформувати історичні дані про транспортування насіння зернових культур. Після завершення цього етапу приступають до побудови прогнозних моделей. Модель, яка прогнозує вартість транспортування насіння зернових культур із найвищим рівнем точності, буде обрана для відповідної інформаційної системи. Отже, структура задачі прогнозування вартості транспортування насіння зернових культур буде наступна:

- постановка задачі;
- розуміння особливостей процесу транспортування насіння зернових культур (відповідь на запитання: *Які фактори впливають на вартість транспортування насіння зернових культур?*);
- збір, підготовка та попередня обробка даних;
- моделювання та тестування;
- розгортання моделі в програмній системі або додатку.

Після того, коли відома типова дорожня карта проекту прогнозування вартості транспортування насіння зернових культур, розглянемо його особливості.

Перш ніж розробляти модель прогнозування вартості транспортування насіння зернових культур, необхідно встановити дві речі: яка інформація доступна на момент складання прогнозу (особливості), а також період часу, на який потрібно отримати прогнозовані значення (ціль).

Початок прогнозу – це час, коли розпочинається процес виконання прогнозу. На практиці вважають початком прогнозу останній час, коли у дослідника є навчальні дані для прогнозованого вартості транспортування насіння зернових культур. При цьому можна їх використовувати для створення функцій.

Горизонт прогнозу – це час, на який виконується прогноз. На практиці досить часто виконується прогноз за кількістю часових кроків у його горизонті: наприклад, «1-кроковий» прогноз вартості транспортування насіння зернових культур або «5-кроковий». Горизонт прогнозу описує ціль прогнозування.

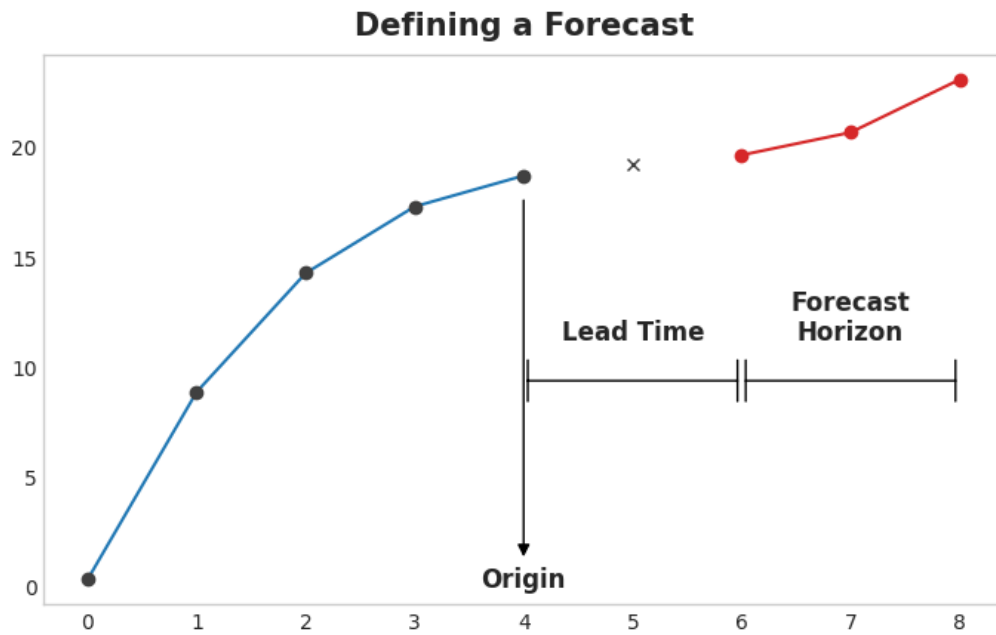


Рис. 2.2. Приклад трьохетапного горизонту прогнозу з двокроковим часом випередження із використанням чотирьох функцій затримки

Час між початком та горизонтом – це час випередження (або іноді затримки) прогнозу. Час очікування прогнозу описується кількістю кроків від початку до горизонту: скажімо, прогноз «на 1 крок попереду» або «на 3 кроки попереду». На практиці може знадобитися, щоб прогноз починався на декілька кроків перед датою початку прогнозу через затримки в отриманні чи обробці даних.

Після цього виконується підготовка даних для прогнозування[¶] вартості транспортування насіння зернових культур. Щоб виконати таке прогнозування за допомогою алгоритмів машинного навчання, насамперед слід перетворити ряд у масив даних, який можна використовувати для навчання за цими алгоритмами. Це справедливо за умови, що не використовуються детерміновані функції, такі як тренд та сезонність.

Кожен окремий рядок у масиві даних представляє один прогноз. Індекс окремого рядка є першим у горизонті прогнозу, але слід розташовувати значення для однієї події (процесу) в одному рядку. За умови виконання багатоетапних прогнозів потрібна модель для отримання кількох результатів, по одному для кожного окремого кроку.

2.2. Підготовка та аналіз даних для прогнозування вартості транспортування насіння зернових культур

Підготовка даних може бути одним із найскладніших кроків у будь-якому проєктів машинного навчання. Причина в тому, що кожен набір даних є різним і дуже специфічним для проєкту. Тим не менш, у проєкті прогнозування вартості транспортування насіння зернових культур є достатньо спільних ознак із іншими проєктами машинного навчання із використанням регресії. При цьому можна визначити окрему послідовність кроків і підзавдань, які виконуватимемо у подальшому.

Процес підготовки та аналізу даних для прогнозування вартості транспортування насіння зернових культур забезпечує один із основних етапів машинного навчання. У цьому етапі виконується підготовка даних, необхідних для прогнозування вартості транспортування насіння зернових культур, враховуючи особливості проєкту, а також виконання як підготовки даних, так і отримання результатів проведеної оцінки алгоритмів машинного навчання.

Підготовка даних для створення моделей машинного навчання – це набагато більше, ніж просто очищення та структурування даних. У багатьох випадках корисно почати з відступу від даних, щоб подумати про основну задачу, яку намагаються вирішити. Щоб побудувати успішну модель машинного навчання слід зрозуміти детально задачу, щоб знати яким чином її вирішувати.

У нашій роботі існує задача прогнозування вартості транспортування насіння зернових культур, яка зумовлюється низкою чинників – вид вантажу, обсяг вантажу, віддаль транспортування, вантажообіг, марка транспортного засобу тощо.

Після того, як було сформульовано задачу машинного навчання, яку потрібно розв'язати, для неї необхідно проаналізувати потенційні джерела даних на підприємстві та від зовнішніх джерел.

Дані для виконання наших досліджень було взято у компанії «Вест Агро Груп», яка має польські інвестиції та із 2015 року займається вирощуванням, закупівлею та зберіганням сільськогосподарських культур на території України та Польщі.

У процесі збору даних необхідно враховувати не лише те, що дані мають представляти особливості вирішуваної задачі, а й те, чому вони були зібрані та що вони можуть означати, особливо коли використовуються в іншому контексті. Під час вирішення нашої задачі також важливо врахувати чинники, які могли вплинути на вартість транспортування насіння зернових культур.

У подальшому проводиться дослідження даних. При цьому слід повністю розуміти наявні дані, з якими працюємо на ранніх стадіях процесу створення моделі прогнозування вартості транспортування насіння зернових культур, щоб мати розуміння їх значення та застосовності. Однією із найбільш поширених помилок під час виконання зазначеного етапу є те, що розпочинають створення моделі, не приділивши часу для розуміння суті даних та їх дослідження.

Дослідження даних означає перегляд таких їх характеристик, як тип і розподіл даних, що містяться в кожній змінній, зв'язки між змінними та те, як вони змінюються відносно результату, який ви прогнозуєте або бажаєте досягти – вартість транспортування насіння зернових культур.

Цей етап може висвітлити такі складові задачі, як колінеарність – змінні, які впливають на бажаний результат і є динамічними, а також випадки коли потрібна стандартизація наборів даних та інші перетворення даних. Саме це також може виявити можливості для покращення продуктивності моделі прогнозування вартості транспортування насіння зернових культур, зменшення розмірності набору даних тощо.

Візуалізація даних також може допомогти покращити цей процес. Це може здатися додатковим етапом, який вважають непотрібним, але для дослідника це є важливий процес виявляти шаблони разом із даними, які не відповідають шаблону. Дослідники даних можуть легко бачити тенденції та правильно досліджувати дані, створюючи відповідні візуалізації, перш ніж

робити висновки. До популярних засобів візуалізації даних належать Tableau, Microsoft Power BI, D3.js і бібліотеки Python, такі як Matplotlib, Bokeh і стек HoloViz.

Наступний етап стосується очищення та перевірки даних. При цьому етапі використовують різні методи очищення та перевірки даних. Вони можуть допомогти аналітикам виявити та виправити невідповідності, викиди, аномалії, відсутні дані та вирішити інші задачі аналізу даних. Наприклад, відсутні окремих значень у даних, які часто можна вирішити за допомогою наявних інструментів, які заповнюють порожні поля статистично релевантними замінниками.

Для очищення та перевірки даних для машинного навчання та забезпечення високої якості даних можна використовувати широкий спектр комерційних інструментів і інструментів з відкритим кодом. Технології з відкритим кодом, такі як Great Expectations і Pандера, наприклад, призначені для перевірки атрибутів даних, які зазвичай використовуються для організації аналітичної обробки даних у двовимірних масивах. Також доступні інструменти, які перевіряють код і робочі процеси обробки даних. Одним із них є pytest, який дослідники даних можуть використовувати для модульного тестування розробки програмного забезпечення та вручну писати тести своїх робочих процесів.

На наступному етапі проводиться структурування даних. Коли дослідники, які проводять обробку даних, будуть задоволені отриманими даними, їм потрібно розглянути алгоритми машинного навчання, які будуть у подальшому використовуватися. Більшість алгоритмів, наприклад, працюють краще, коли дані розбиваються на категорії, наприклад діапазони обсягів транспортування зернових культур, а не залишаються як необроблені числа.

Двома прийомами попередньої обробки даних, які часто дослідники пропускають, є групування даних і згладжування безперервних функцій. Ці методи регуляризації даних можуть зменшити дисперсію моделі машинного

навчання, запобігаючи зміні незначними статистичними флуктуаціями в наборі даних.

Об'єднання даних у різні групи можна здійснювати рівновіддаленим способом, з однаковою «шириною» для кожного діапазону, або еквістатистичним методом із приблизно однаковою кількістю вибірок у кожному контейнері. Це також може служити передумовою для локальної оптимізації даних у кожному контейнері для створення моделей машинного навчання з низьким упередженням.

Згладжування безперервних функцій може допомогти у «позбавленні» необроблених даних. Цей етап також можна використовувати для створення причинно-наслідкових припущень щодо процесу генерації даних, представляючи зв'язки в упорядкованих наборах даних як монотонні функції, які зберігають порядок між елементами даних.

Інші дії, які спеціалісти з обробки даних часто виконують для структурування даних для машинного навчання, включають наступне:

- скорочення даних за допомогою таких методів, як вибірка атрибутів або записів і агрегація даних;
- нормалізація даних, що включає зменшення розмірності та масштабування даних;
- створення окремих наборів даних для навчання та тестування моделей машинного навчання.

Після цього виконується розробка та вибір функцій. Це є останнім етапом підготовки даних перед розробкою моделі машинного навчання. Розробка функцій, яка передбачає додавання або створення нових змінних для покращення виходу моделі, є основним здобутком фахівців із обробки даних і має різні форми. Приклади включають вилучення окремих змінних із набору даних, розкладання змінних на окремі ознаки, агрегування змінних і перетворення ознак на основі розподілу ймовірностей.

Дослідники даних також повинні звернути увагу на вибір функцій – вибір відповідних функцій для аналізу та усунення нерелевантних. Багато функцій

можуть виглядати багатообіцяючими, але призводять до задач, які потребують розширене навчання моделі та їх зміну, що обмежує здатність моделі точно аналізувати нові дані. Такі методи, як ласо-регресія та автоматичне визначення релевантності, можуть допомогти у виборі ознак.

2.3. Вибір методу машинного навчання для прогнозування вартості транспортування насіння зернових культур

Лінійна регресія є одним із найважливіших та широко використовуваних методів машинного навчання. Водночас, це один з найпростіших методів машинного навчання. Однією з його головних переваг є простота інтерпретації результатів.

Під час машинного навчання та реалізації моделі лінійної регресії деякої залежної змінної y на множині незалежних змінних:

$$x = (x_1, x_2, \dots, x_r), \quad (2.1)$$

де r – кількість предикторів, ви припускаєте лінійний зв'язок між y та x .

При цьому отримане рівняння є рівнянням регресії:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_r x_r + \varepsilon, \quad (2.2)$$

де $\beta_0, \beta_1, \dots, \beta_r$ – відповідно коефіцієнти регресії, а ε – випадкова похибка.

Лінійна регресія обчислює оцінки коефіцієнтів регресії або просто прогнозованих ваг, позначених b_0, b_1, \dots, b_r .

Ці оцінки визначають оцінену функцію регресії:

$$f(\mathbf{x}) = b_0 + b_1 x_1 + \dots + b_r x_r. \quad (2.3)$$

Функція (2.3) має достатньо добре фіксувати залежності між входами та виводом моделі.

Розрахункова або прогнозована відповідь функції $f(x_i)$, для кожного спостереження $i = 1, 2, \dots, n$, повинна бути якомога ближчою до відповідної фактичної реакції y_i .

Різниця $y_i - f(x_i)$ для всіх спостережень $i = 1, 2, \dots, n$ називаються залишками. Регресія стосується визначення найкращих прогнозованих ваг, тобто ваг, що відповідають найменшим залишкам.

Щоб отримати найкращі ваги, ви зазвичай мінімізуєте суму квадратів залишків (SSR) для всіх спостережень $i = 1, 2, \dots, n$:

$$SSR = \sum_i (y_i - f(x_i))^2. \quad (2.4)$$

Описаний вище підхід називається методом звичайних найменших квадратів.

2.4. Продуктивність лінійної регресії та її різновиди

Варіація фактичних відповідей y_i , $i = 1, 2, \dots, n$, відбувається частково через залежність від предикторів x_i . Однак існує також додаткова притаманна дисперсія результату.

Коефіцієнт детермінації, позначений як R^2 , вказує на те, яке кількісне значення варіації y можна відображається залежністю від x , використовуючи конкретну модель регресії. Більший R^2 вказує на кращу відповідність між показниками і означає, що модель може краще пояснити варіацію виходу з різними вхідними даними.

Значення $R^2 = 1$ відповідає $SSR = 0$. Це є повна відповідність, оскільки значення прогнозованих і фактичних відповідей повністю збігаються одне з одним.

2.4.1. Проста лінійна регресія

Проста або однофакторна лінійна регресія є найпростішим випадком лінійної регресії, оскільки вона має одну незалежну змінну $X = x$. Наступний рис. 2.3 ілюструє просту лінійну регресію.

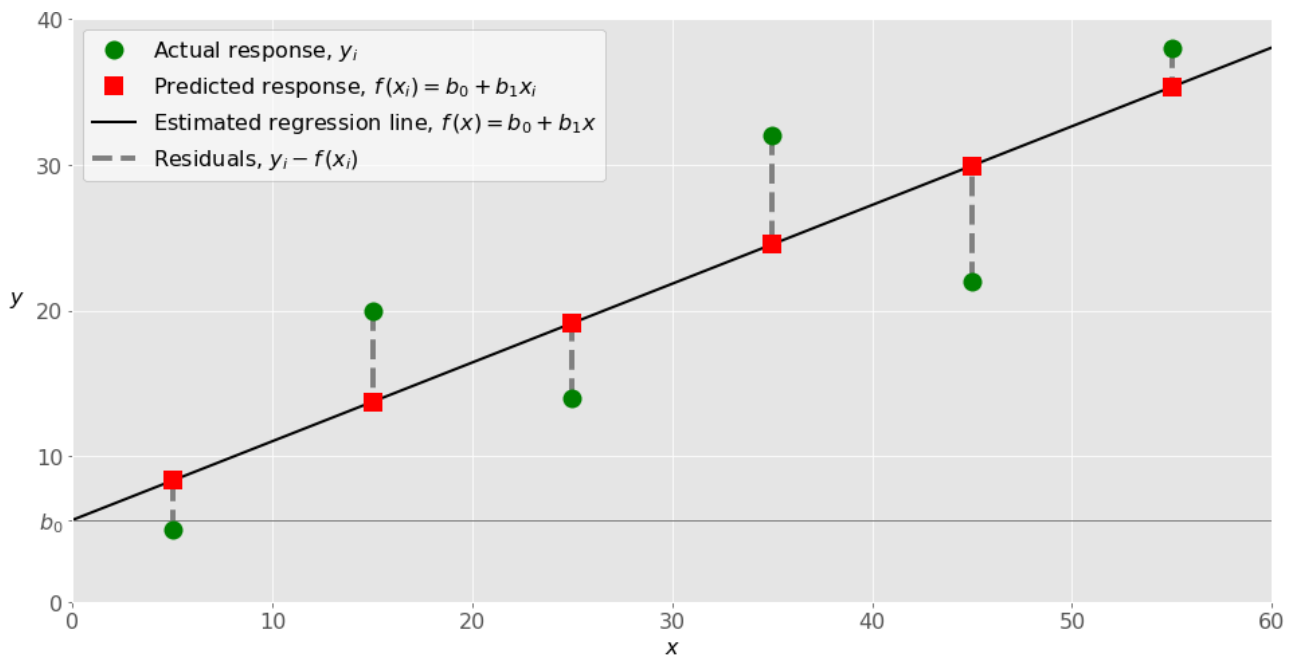


Рис. 2.3. Приклад простої лінійної регресії

Реалізуючи просту лінійну регресію, зазвичай починають із заданого набору пар даних для входу-виходу ($x - y$). Ці пари даних є вибірками спостереженнями, які зображено зеленими кружечками на рис. 2.3. Наприклад, крайнє ліве спостереження має вхід $x = 5$ та має фактичний вихід, або відповідь, $y = 5$. Наступне спостереження має $x = 15$ та $y = 20$ і так далі.

Оцінена функція регресії, представлена чорною лінією, має рівняння[^]

$$f(x) = b_0 + b_1x. \quad (2.5)$$

Метою дослідження є обчислити оптимальні значення прогнозованих ваг b_0 та b_1 , які мінімізують SSR , і визначити оцінену функцію регресії.

Значення b_0 , яке також називають перетином, показує точку, де оцінена лінія регресії перетинає вісь y . Це значення оціненої відповіді $f(x)$ для $x=0$. Значення b_1 визначає нахил розрахункової лінії регресії.

Прогнозовані відповіді, показані у вигляді червоних квадратів, є точками на лінії регресії, які відповідають вхідним значенням. Наприклад, для вхідних даних $x=5$ прогнозована відповідь $f(5)=8,33$, що представляє крайній лівий червоний квадрат.

Вертикальні пунктирні сірі лінії представляють залишки, які можна обчислити як:

$$y_i - f(x_i) = y_i - b_0 - b_1x_i. \quad (2.6)$$

Вираз (2.6) справедливий для $i=1,2,\dots,n$. Це відстані між зеленими колами та червоними квадратами. Коли застосовується лінійна регресія, то насправді намагаються мінімізувати ці відстані та зробити червоні квадрати якомога ближчими до попередньо визначених зелених кіл.

2.4.2. Множинна лінійна регресія

Багатофакторна або багатовимірна лінійна регресія – це випадок лінійної регресії з двома або більше незалежними змінними. Якщо є лише дві незалежні змінні, тоді оцінена функція регресії дорівнює:

$$f(x_1, x_2) = b_0 + b_1x_1 + b_2x_2. \quad (2.7)$$

Отримана залежність являє собою площину регресії в тривимірному просторі. Метою цієї регресії є визначення значень ваг b_0 , b_1 і b_2 так, щоб ця площина була якомога ближчою до фактичних відповідей, водночас даючи мінімальну SSR .

Вибірки, що мають більше ніж дві незалежні змінні подібні, але більш загальні. Розрахункова функція регресії дорівнює:

$$f(x_1, x_2, \dots, x_r) = b_0 + b_1 x_1 + \dots + b_r x_r. \quad (2.8)$$

При цьому $r+1$ це вагові коефіцієнти, які потрібно визначити, коли кількість вхідних даних дорівнює r .

2.4.3. Поліноміальна лінійна регресія

Можна розглядати поліноміальну регресію як узагальнений випадок лінійної регресії. При цьому припускають, що поліноміальна залежність між виходом і вхідними даними має поліноміальну оцінену функцію регресії. Іншими словами, окрім лінійних термінів, таких як $b_r x_r$, функція регресії $f(x_1, x_2, \dots, x_r)$ може містити нелінійні терміни, такі як $b_2 x_1^2$, $b_3 x_1^3$ або навіть $b_4 x_1 x_2$ та $b_5 x_1^2 x_2$.

Найпростіший приклад поліноміальної регресії має одну незалежну змінну, а оцінювана функція регресії є поліномом другого ступеня:

$$f(x) = b_0 + b_1 x + b_2 x^2. \quad (2.9)$$

Виходячи із виразу (2.9) слід пам'ятати, що потрібно обчислити b_0 , b_1 та b_2 , щоб мінімізувати SSR .

Порівнюючи попередню функцію регресії з функцією (2.7), яка використовується для лінійної регресії, можна сказати що вони виглядають дуже схожими і є лійними функціями невідомих b_0 , b_1 та b_2 . Ось чому можна розв'язати задачу поліноміальної регресії як лінійну задачу із членом x^2 , який розглядається як вхідна змінна.

У випадку двох змінних та полінома другого ступеня функція регресії має вигляд:

$$f(x_1, x_2) = b_0 + b_1x_1 + b_2x_2 + b_3x_1^2 + b_4x_1x_2 + b_5x_2^2. \quad (2.10)$$

Порядок вирішення задачі ідентичний попередньому випадку. При цьому застосовується лінійна регресія для п'яти вхідних даних: x_1 , x_2 , x_1^2 , x_1x_2 і x_2^2 .

У результаті регресії отримується значення шести ваг, які мінімізують SSR – b_0 , b_1 , b_2 , b_3 , b_4 і b_5 .

Звичайно, є більш конкретні задачі регресії, які мають свою специфіку, але описаних достатньо щоб означити їх особливості.

2.5. Інструментарій мови Python для побудови лінійних регресій

Для обґрунтування лінійних регресій вибираємо мову Python. Вона є дуже популярною сьогодні та потужною мовою програмування. Python є універсальною, і її можна використовувати для створення потужних і корисних програм. Python можна використовувати для створення різних типів інформаційних систем, веб-додатків, настільних програм. Python також дуже активно використовується в науці про дані та аналізі даних.

Також нами обрано Jupyter Notebook – це середовище, яке можна використовувати для інтерактивних експериментів із Python. Це дозволяє ділитися живим кодом Python з іншими учасниками проектування інформаційних систем.

Jupyter Notebook є важливий інструмент для даних, створення прототипів і навчання, який є в наборі інструментів кожного розробника Python. Із використанням Jupyter Notebook можна прослідкувати навігацію інтерфейсом, а також прослідкувати рамки коду та доповнити блокноти багатим вмістом, таким як розмітка, візуалізація та форматування даних. Окрім того можна зібрати необхідні блокноти Jupyter, які допоможуть створити регресійну модель машинного навчання для прогнозування вартості транспортування насіння зернових культур.

Окрім того, нам потрібно використати низку бібліотек та їхні функції та класи.

Бібліотека NumPy – це фундаментальний науковий пакет Python, який дозволяє застосовувати багато високопродуктивних операцій над одновимірними та багатовимірними масивами. Цей пакет також пропонує багато математичних процедур, які мають відкритий код.

Безкоштовна програмна бібліотека машинного навчання scikit-learn – це широко використовувана бібліотека Python для машинного навчання, побудована на основі NumPy та деяких інших бібліотеках. Вона надає засоби для попередньої обробки даних, зменшення розмірності, реалізації регресії, класифікації, кластеризації тощо. Як і NumPy, scikit-learn також має відкритий код.

За умови побудови лінійних регресій слід використовувати функціональності, що виходить за рамки scikit-learn. Для цього підходить бібліотека statsmodels. Це потужний пакет Python для оцінки статистичних моделей, виконання тестів тощо, який також має відкритий код.

Для побудови лінійних регресій першим кроком є імпорт пакета numpy та класу LinearRegression із sklearn.linear_model:

```
Python
```

```
>>>
```

```
>>> import numpy as np
>>> from sklearn.linear_model import LinearRegression
```

Після цього є усі функції, необхідні для впровадження лінійної регресії. Основним типом даних NumPy є тип масиву під назвою `numpy.ndarray`.

Клас `sklearn.linear_model.LinearRegression` використовують для обґрунтування лінійної та поліноміальної регресії та виконання відповідних прогнозів.

Наступним етапом для прогнозування вартості транспортування насіння зернових культур є визначення даних для роботи. Вхідні дані (регресори, x) і вихідні дані (відповідь, y) мають бути масивами або подібними об'єктами. Це найпростіший спосіб надання даних для регресії.

Після цього є два масиви – вхідний x , вихідний y . При цьому слід викликати `.reshape()`, оскільки цей масив має бути двовимірним, точніше, він повинен мати один стовпець та стільки рядків, скільки потрібно. Саме це вказує аргумент `(-1, 1)`.`.reshape()`.

Наступним кроком є створення моделі лінійної регресії та підгонка її за наявними даними. Для цього слід створити екземпляр класу `LinearRegression`, який представлятиме модель регресії:

```
Python >>>
>>> model = LinearRegression()
```

Цей оператор створює змінну `model` як екземпляр `LinearRegression`. При цьому можна надати кілька додаткових параметрів для `LinearRegression`:

- `fit_intercept` є логічним значенням, яке, якщо `True`, вирішує обчислити відрізок b_0 або, якщо `False`, вважає його рівним нулю. За замовчуванням `True`.

- `normalize` є логічним значенням, яке, якщо `True`, вирішує нормалізувати вхідні змінні. За умовчанням він має значення `False`, і в цьому випадку він не нормалізує вхідні змінні.

- `copy_X` є логічним значенням, яке вирішує, копіювати (`True`) чи перезаписувати вхідні змінні (`False`). Це `True` за замовчуванням.

➤ `n_jobs` є або цілим числом, або `None`. Він представляє кількість завдань, які використовуються в паралельних обчисленнях. За замовчуванням це `None`, що зазвичай означає одну роботу. `-1` означає використовувати всі доступні процесори.

Після цього починають використовувати модель. Для початку необхідно звернутися `.fit():model`

```
Python >>>
>>> model.fit(x, y)
LinearRegression()
```

За допомогою `.fit()` можна обчислювати оптимальні значення ваг b_0 і b_1 , використовуючи наявні вхідні та вихідні дані x та y , як аргументи. Іншими словами, `.fit()` відповідає моделі. Він повертає `self`, що є самою змінною `model`. Ось чому можна замінити останні два твердження цим:

```
Python >>>
>>> model = LinearRegression().fit(x, y)
```

Цей оператор робить те саме, що й два попередні. Просто він коротший.

Після встановлення моделі можна отримати результати, щоб перевірити, чи модель працює задовільно та інтерпретувати її. При цьому можна отримати коефіцієнт детермінації, R^2 , за `.score()` допомогою `model`:

```
Python >>>
>>> r_sq = model.score(x, y)
>>> print(f"coefficient of determination: {r_sq}")
```

Коли застосовується `.score()`, аргументи також є предиктором x_i відповіддю y , а поверненим значенням є R^2 .

У `scikit-learn`, за бажанням, кінцеве значення вказує на атрибут що оцінюється. У цьому прикладі `.intercept_` і `.coef_` є оціночними значеннями.

Отримавши задовільну модель, можна використовувати її для прогнозів із наявними чи новими даними. Щоб отримати прогнозовану відповідь, використовується `.predict()`:

Python

>>>

```
>>> y_pred = model.predict(x)
>>> print(f"predicted response:\n{y_pred}")
```

При застосуванні `.predict()` передається регресор як аргумент і отримується відповідна прогнозована відповідь.

РОЗДІЛ 3.

РЕЗУЛЬТАТИ ПІДГОТОВКИ ДАНИХ ТА РОЗРОБКИ ІНФОРМАЦІЙНОЇ СИСТЕМИ ПРОГНОЗУВАННЯ ВАРТОСТІ ТРАНСПОРТУВАННЯ НАСІННЯ ЗЕРНОВИХ КУЛЬТУР

3.1. Результати підготовки та аналізу даних для прогнозування вартості транспортування насіння зернових культур

На підставі отриманого аналітичного набору даних компанії «Вест Агро Груп» щодо процесів транспортування насіння зернових культур автомобілями DAF FT XF 105 виділено наступні їх атрибути, що характеризують вартості транспортування насіння зернових культур (рис. 3.1).

```
In [1]: import pandas as pd
df = pd.read_excel("F:\Dipl_2022\Mar_IT\Литвин\Data_DAF_model_1.xlsx", engine="openpyxl")
df.to_csv('F:\Dipl_2022\Mar_IT\Литвин\Data_DAF_model_000.csv', encoding='utf-8-sig')
# show the dataframe
print(df)
```

	Вантаж	Відстань, км (загальна)	Фактичні витрати палива, літрів	\
0	Пшениця	580	97.810	
1	Ріпак	55	10.412	
2	Ріпак	81	15.424	
3	Пшениця	92	17.756	
4	Пшениця	202	39.065	
...	
9697	Ріпак	10	9.694	
9698	Пшениця	16	15.526	
9699	Ріпак	3	2.931	
9700	Пшениця	5	4.888	
9701	Ріпак	16	15.726	

	Витрати ДП, літрів/100км	Вантажообіг, т.км	Обсяг вантажу, тон	\
0	16.863793	14882.80	25.66	
1	18.930909	5457.10	99.22	
2	19.041975	2030.67	25.07	
3	19.300000	149.04	1.62	
4	19.339109	5561.06	27.53	
...	
9697	96.940000	478.30	47.83	
9698	97.037500	426.56	26.66	
9699	97.700000	294.96	98.32	
9700	97.760000	137.10	27.42	
9701	98.287500	7200.00	450.00	

	Вартість доставки, грн
0	37207.000
1	13642.750
2	5076.675
3	372.600
4	13902.650
...	...
9697	1195.750
9698	1066.400
9699	737.400
9700	342.750
9701	18000.000

[9702 rows x 7 columns]

Рис. 3.1. Аналітичний набір даних компанії «Вест Агро Груп» щодо процесів транспортування насіння зернових культур

Із представленого аналітичного набору даних компанії «Вест Агро Груп» щодо процесів транспортування насіння зернових культур (рис. 3.1) можна сказати, що він містить 9702 спостереження за 8 атрибутами (рис. 3.2).

```
In [4]: data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9702 entries, 0 to 9701
Data columns (total 8 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                -
0   Unnamed: 0                            9702 non-null   int64
1   Вантаж                                9702 non-null   object
2   Відстань, км (загальна)                9702 non-null   int64
3   Фактичні витрати палива, літрів        9702 non-null   float64
4   Витрати ДП, літрів/100км               9702 non-null   float64
5   Вантажобіг, т.км                       9702 non-null   float64
6   Обсяг вантажу, тон                     9702 non-null   float64
7   Вартість доставки, грн                 9702 non-null   float64
dtypes: float64(5), int64(2), object(1)
memory usage: 606.5+ KB
```

Рис. 3.2. Характеристика даних

Для прогнозування вартості транспортування насіння зернових культур зазначений набір даних має 6 вхідних змінних ($X_1 \dots X_6$) та 1 вихідну змінну (Y_1 – Вартість доставки, грн). Зазначені дані лежать в основі вирішення задачі лінійної регресії. При цьому використано середовище розробки Jupyter Notebook.

Наявний набір даних для прогнозування вартості транспортування насіння зернових культур після завантаження, опрацьовуємо та виводимо у вигляді масиву даних (рис. 3.3).

```
In [2]: # Read in data and display first 5 rows
data = pd.read_csv('F:\Dipl_2022\Mar_IT\Литвин\Data_DAF_model_000.csv')
data.head(9703)
```

Out[2]:

	Unnamed: 0	Вантаж	Відстань, км (загальна)	Фактичні витрати палива, літрів	Витрати ДП, літрів/100км	Вантажообіг, т.км	Обсяг вантажу, тон	Вартість доставки, грн
0	0	Пшениця	580	97.810	16.863793	14882.80	25.66	37207.000
1	1	Ріпак	55	10.412	18.930909	5457.10	99.22	13642.750
2	2	Ріпак	81	15.424	19.041975	2030.67	25.07	5076.675
3	3	Пшениця	92	17.756	19.300000	149.04	1.62	372.600
4	4	Пшениця	202	39.065	19.339109	5561.06	27.53	13902.650
...
9697	9697	Ріпак	10	9.694	96.940000	478.30	47.83	1195.750
9698	9698	Пшениця	16	15.526	97.037500	426.56	26.66	1066.400
9699	9699	Ріпак	3	2.931	97.700000	294.96	98.32	737.400
9700	9700	Пшениця	5	4.888	97.760000	137.10	27.42	342.750
9701	9701	Ріпак	16	15.726	98.287500	7200.00	450.00	18000.000

Рис. 3.3. Результати формування масиву початкових даних

Насамперед виконаємо аналіз виду транспортованих вантажів, до яких належать окремі види насіння зернових культур (рис. 3.4).

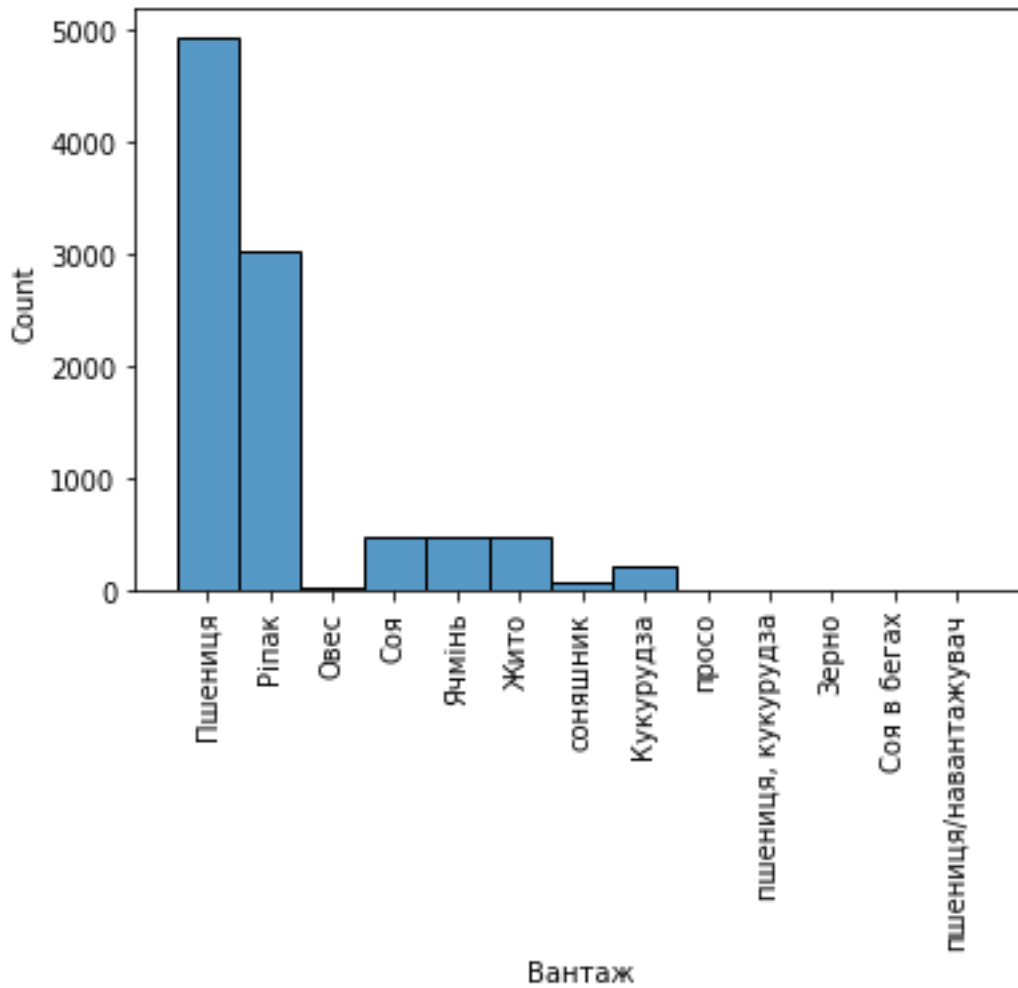


Рис. 3.4. Кількість замовлень на транспортування окремих видів насіння зернових культур

Встановлено, що найбільше замовлень на транспортування припадає на пшеницю – 4920 од та ріпак – 3026 од.

На наступному етапі виконано побудову гістограм зміни основних даних за їх встановленими атрибутами (рис. 3.5).

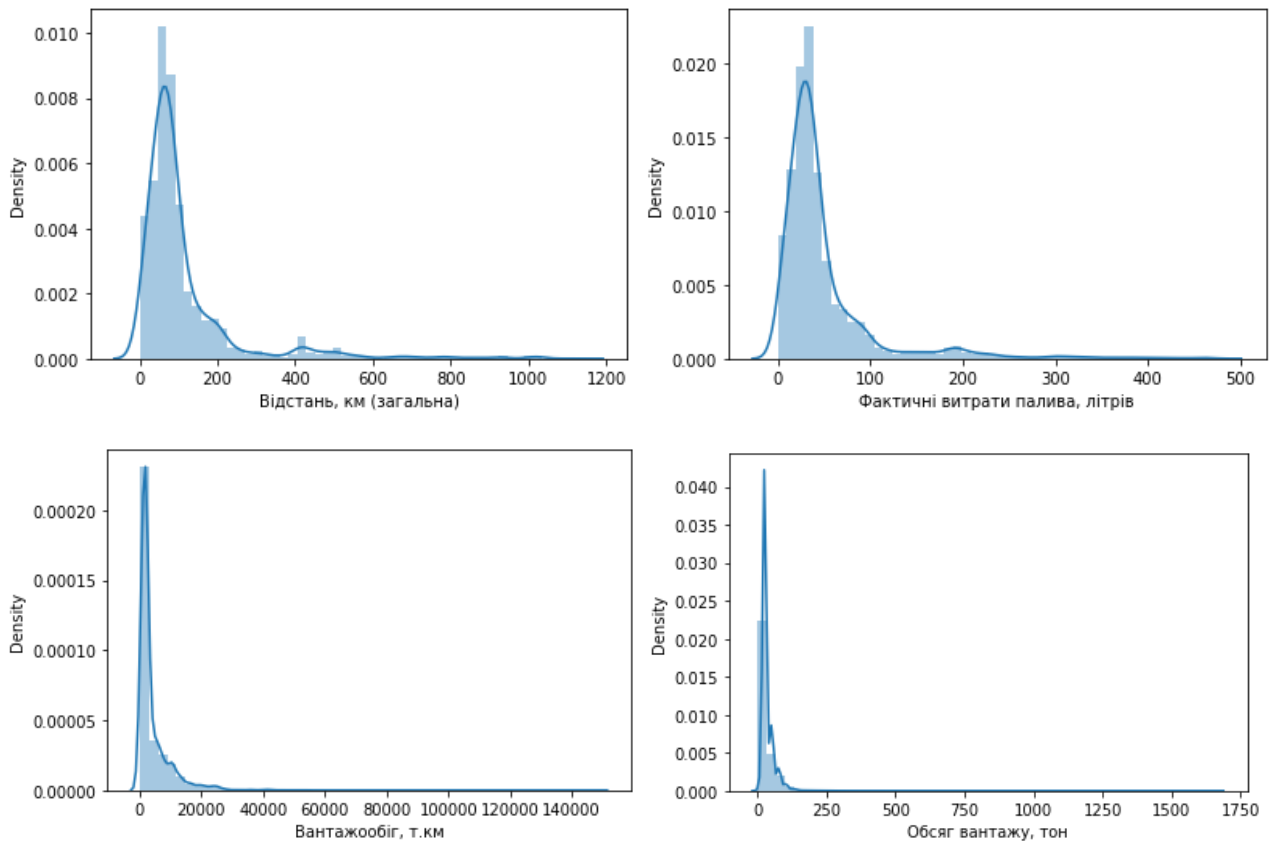


Рис. 3.5. Результати оцінення мінливості основних даних за їх встановленими атрибутами

Окрім того, нами виконано оцінення вартості доставки насіння зернових культур (рис. 3.6).

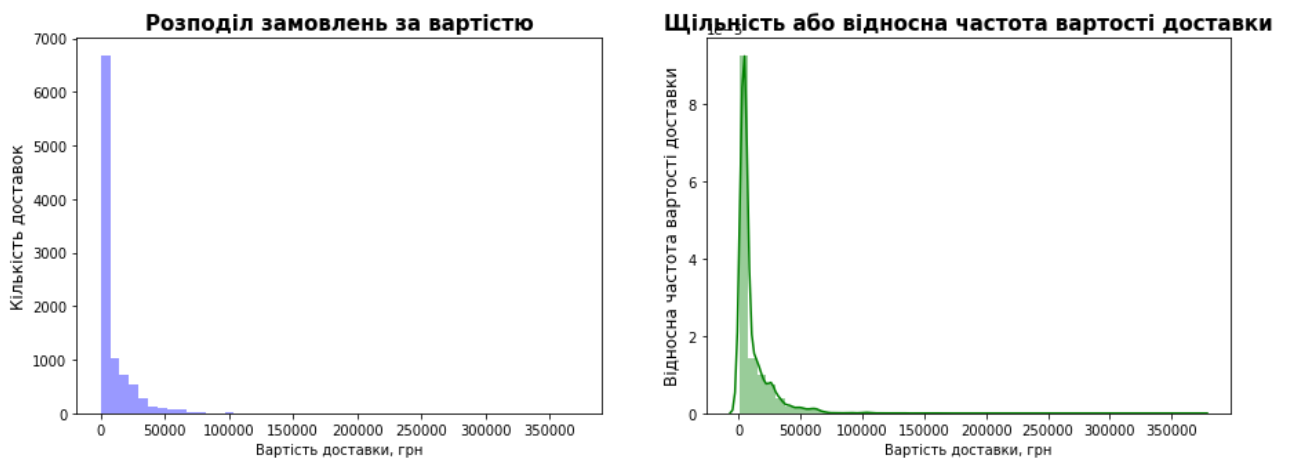


Рис. 3.6. Результати оцінення вартості доставки насіння зернових культур

Для встановлення взаємозв'язків між окремими даними за встановленими атрибутами побудовано усі можливі залежності між ними (рис. 3.7).

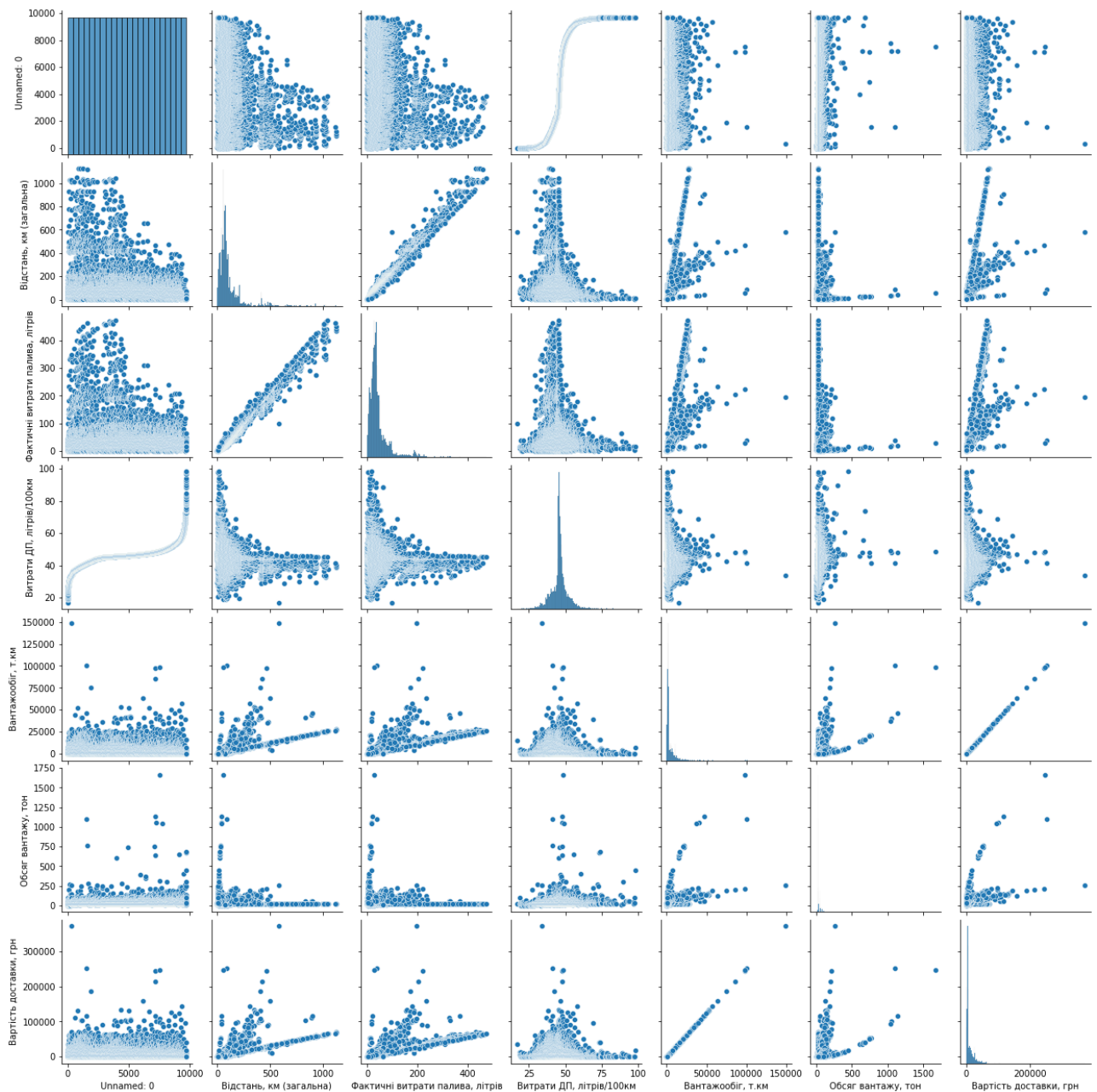


Рис. 3.7. Взаємозв'язки між даними доставки насіння зернових культур за встановленими атрибутами

На підставі попереднього аналізу даних щодо доставки насіння зернових культур встановлено, що є між ними сильні взаємозв'язки (наприклад, між фактичними витратами палива та відстанню, вантажообігом та вартістю доставки), а також такі, де зв'язки відсутні (наприклад, питомими та фактичними витратами палива).

Для кількісного оцінення тісноти зв'язків між даними доставки насіння зернових культур нами побудовано кореляційну матрицю (рис. 3.8).



Рис. 3.8. Кореляційна матриця між даними доставки насіння зернових культур

Отримана кореляційна матриця відображає коефіцієнти кореляції для різних змінних. Встановлено, що найвищий коефіцієнт кореляції між вантажообігом, відстанню, фактичними витратами палива та вартістю доставки насіння зернових культур відповідно становить – 1,0, 0,7 та 0,75. Водночас, найнижчий коефіцієнт кореляції між питомими витратами палива та вартістю доставки насіння зернових культур відповідно становить – 0,099.

Нами виконано заміну текстових змінних, які стосуються виду транспортованих зернових культур, на числові (рис. 3.9).

```
In [30]: X=data.drop(columns=['Вартість доставки, грн', 'Unnamed: 0'])
y=data['Вартість доставки, грн']
```

```
In [31]: # створення фіктивних змінних для категоріальних змінних
cars_categorical = X.select_dtypes(include=['object'])
cars_categorical.head()
```

```
Out[31]:
```

	Вантаж
0	Пшениця
1	Ріпак
2	Ріпак
3	Пшениця
4	Пшениця

Рис. 3.9. Заміну текстових змінних на числові

3.2. Масштабування функцій та навчання моделі

Виконаний аналіз набору даних для прогнозування вартості доставки насіння зернових культур за встановленими атрибутами лежить в основі проведення масштабування функцій, підбору моделі та виконання її навчання.

```
In [36]: X.describe()
#усі засоби = 0 або нуль і std=1
```

	Відстань, км (загальна)	Фактичні витрати палива, літрів	Витрати ДП, літрів/100км	Вантажообіг, т.км	Обсяг вантажу, тон	Вантаж_Зерно	Вантаж_Кукурудза	Вантаж_Овес	Вантаж_Пшениця	Вантаж_Рі
count	9.702000e+03	9.702000e+03	9.702000e+03	9.702000e+03	9.702000e+03	9.702000e+03	9.702000e+03	9.702000e+03	9.702000e+03	9.702000e
mean	1.709940e-15	-2.757821e-16	-2.038544e-15	2.219531e-16	-6.842141e-16	-8.582443e-16	1.306634e-14	-8.297964e-16	1.592762e-15	1.991558e
std	1.000052e+00	1.000052e+00	1.000052e+00	1.000052e+00	1.000052e+00	1.000052e+00	1.000052e+00	1.000052e+00	1.000052e+00	1.000052e
min	-8.107733e-01	-8.631856e-01	-4.472543e+00	-7.001458e-01	-8.683819e-01	-1.015294e-02	-1.512555e-01	-4.657464e-02	-1.017049e+00	-6.729266e
25%	-4.628020e-01	-4.817422e-01	-3.277512e-01	-4.677736e-01	-2.935877e-01	-1.015294e-02	-1.512555e-01	-4.657464e-02	-1.017049e+00	-6.729266e
50%	-2.888164e-01	-2.842051e-01	-3.427275e-02	-3.635010e-01	-2.730593e-01	-1.015294e-02	-1.512555e-01	-4.657464e-02	9.832368e-01	-6.729266e
75%	-2.058854e-02	5.086339e-03	3.020938e-01	1.095422e-01	1.542929e-01	-1.015294e-02	-1.512555e-01	-4.657464e-02	9.832368e-01	1.486046e
max	7.337554e+00	7.314617e+00	7.941196e+00	2.443575e+01	3.938412e+01	9.849365e+01	6.611329e+00	2.147091e+01	9.832368e-01	1.486046e

```
In [37]: # розділити на навчальні та тестові
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y,
                                                    train_size=0.7,
                                                    test_size = 0.3, random_state=100)
```

Рис. 3.10. Результати масштабування функцій та підготовки вибірок даних для навчання моделі

З метою проведення навчання моделі для прогнозування вартості доставки насіння зернових культур підготовлені дані розбито на навчальну вибірку – 70% та тестову вибірку – 30%.

У подальшому виконуємо побудову моделі та здійснюємо вибір функцій за допомогою RFE. Оскільки вартість доставки насіння зернових культур, яка є залежною змінною, і описується із більшістю незалежних змінних лінійними функціями, використовуємо лише лінійну регресію, а не інші типи регресії, такі як поліноміальна, випадковий ліс тощо. На підставі цього представляємо модель з усіма функціями, використовуючи бібліотеку Sklearn із моделлю машинного навчання LinearRegression (рис. 3.11).

```
In [38]: # Модель з усіма функціями
from sklearn import linear_model
from sklearn.linear_model import LinearRegression

lm=LinearRegression()
lm.fit(X_train,y_train)

y_pred_test=lm.predict(X_test)
y_pred_train=lm.predict(X_train)
```

Рис. 3.11. Створення моделі лінійної регресії

Враховуючи проведену зовнішню оцінку моделі, яка призначає ваги окремим її характеристикам (коефіцієнтам лінійної моделі) виконується рекурсивне виключення ознак (RFE) (рис. 3.12).

```
In [40]: from sklearn.feature_selection import RFE
import statsmodels.api as sm

n_features_list = list(range(4, 31)) #перевірка оптимальної кількості функцій від 4 до 30
train_adjusted_r2 = []
train_r2 = []
test_r2 = []
train_RMSE=[]
test_RMSE=[]

for n_features in range(4, 31):

    # RFE з n функціями
    lm = LinearRegression()

    # визначення кількості функцій
    rfe_n = RFE(estimator=lm, n_features_to_select=n_features)

    # відповідати n функціям
    rfe_n.fit(X_train, y_train)

    # вибір функцій, вибраних rfe_n
    col_n = X_train.columns[rfe_n.support_] #rfe_n.support_: повертає масив із логічними значеннями, щоб вказати, чи є
    #атрибут було вибрано за допомогою RFE

    # навчальні та тестові дані для n вибраних стовпців
    X_train_rfe_n = X_train[col_n]
    X_test_rfe_n = X_test[col_n]

    # додати константу до моделі
    X_train_rfe_n = sm.add_constant(X_train_rfe_n)

    X_test_rfe_n = sm.add_constant(X_test_rfe_n, has_constant='add')

    # підгонка моделі з n ознаками
    lm_n = sm.OLS(y_train, X_train_rfe_n).fit()

    ## Складання прогнозів
    y_pred_test = lm_n.predict(X_test_rfe_n)
    y_pred_train = lm_n.predict(X_train_rfe_n)

    #Розрахунок оціночних метрик

    #R-square
    train_adjusted_r2.append(lm_n.rsquared_adj)
    train_r2.append(lm_n.rsquared)
    test_r2.append(r2_score(y_test, y_pred_test))

    #RMSE/stan. error
    error_test=y_pred_test-y_test
    error_train=y_pred_train-y_train

    test_RMSE.append((((error_test**2).mean())**0.5)
    train_RMSE.append((((error_train**2).mean())**0.5)
```

Рис. 3.12. Рекурсивне виключення ознак (RFE)

Рекурсивне виключення ознак (RFE) являє собою вибір функцій на підставі рекурсивного розгляду менших наборів атрибутів. При цьому модель навчається на початковому наборі даних, а водночас вага кожної окремої функції визначається конкретними атрибутами. Потім найменш важливі функції видаляються із набору функцій, що повторюється доки не буде досягнуто бажаної кількості функцій (рис. 3.13).

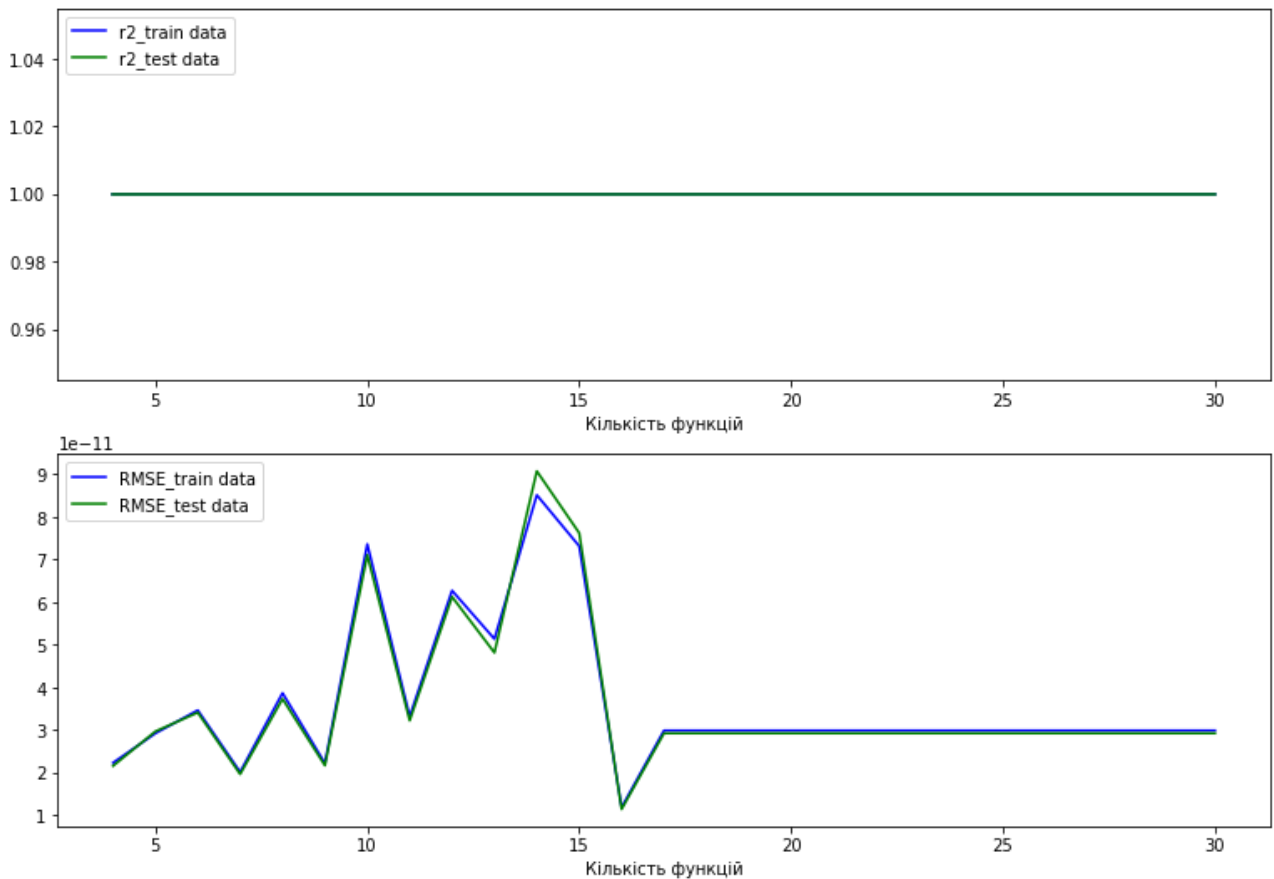


Рис. 3.13. Результати визначення раціональної кількості функцій

З наведених вище графіків (рис. 3.13) знаходимо, що R_{square} для тестових даних досягає максимуму на 13 функціях, і на цьому етапі модель добре прогнозує, оскільки значення R^2 однакове і не залежить від кількості функцій.

Середньоквадратична похибка RMSE для тестових даних є найнижчою за використання 13 функцій і зі зростання кількості функцій, далі зростає. Навчання RMSE на рівні 13 функцій також виглядає на належному рівні, а додавання додаткових функцій до навчання зменшує RMSE. При цьому завжди

є компроміс між видаленням функцій (зменшенням складності моделі) та продуктивністю моделі. У нашій моделі будемо використовувати 13 функцій.

Обґрунтована модель для прогнозування вартості доставки насіння зернових культур з оптимальною кількістю функцій представлена на рис. 3.14.

```

-----R-sqaure-----
R-sq for test data is 1.0
R-sq for train data is 1.0
-----STANDARD ERROR/RMSE-----
RMSE for test data is 4.8154992909791285e-11
RMSE for train data is 5.1371675571286393e-11
                OLS Regression Results
=====
Dep. Variable:   Вартість доставки, грн      R-squared:                1.000
Model:          OLS                        Adj. R-squared:           1.000
Method:         Least Squares              F-statistic:              4.418e+31
Date:          Thu, 24 Nov 2022             Prob (F-statistic):       0.00
Time:          17:00:05                     Log-Likelihood:          1.5126e+05
No. Observations: 6791                    AIC:                     -3.025e+05
Df Residuals:   6777                      BIC:                     -3.024e+05
Df Model:       13
Covariance Type: nonrobust
=====
                coef      std err      t      P>|t|      [0.025      0.975]
-----
const          1.038e+04  6.24e-13  1.66e+16  0.000      1.04e+04      1.04e+04
Відстань, км (загальна) -1.91e-11  5.89e-12  -3.243    0.001     -3.06e-11     -7.56e-12
Фактичні витрати палива, літрів 9.095e-11  5.85e-12  15.546    0.000      7.95e-11      1.02e-10
Витрати ДП, літрів/100км -3.411e-12  7.94e-13  -4.294    0.000     -4.97e-12     -1.85e-12
Вантажообіг, т.км 1.479e+04  1.13e-12  1.31e+16  0.000      1.48e+04      1.48e+04
Обсяг вантажу, тон 7.731e-12  8.01e-13  9.647     0.000      6.16e-12      9.3e-12
Вантаж_Кукурудза -3.602e-12  6.87e-13  -5.245    0.000     -4.95e-12     -2.26e-12
Вантаж_Овес -5.4e-13  7.57e-13  -0.714    0.476     -2.02e-12      9.44e-13
Вантаж_Ріпак 4.704e-12  6.47e-13  7.265     0.000      3.43e-12      5.97e-12
Вантаж_Соя 9.095e-13  6.42e-13  1.417     0.157     -3.49e-13      2.17e-12
Вантаж_Соя в бегах 4.002e-12  5.22e-13  7.664     0.000      2.98e-12      5.03e-12
Вантаж_Ячмінь -3.979e-13  6.37e-13  -0.625    0.532     -1.65e-12      8.5e-13
Вантаж_пшениця, кукурудза 1.215e-12  5.22e-13  2.327     0.020      1.91e-13      2.24e-12
Вантаж_соняшник 1.748e-12  6.69e-13  2.613     0.009      4.37e-13      3.06e-12
=====
Omnibus:          4638.118      Durbin-Watson:           1.918
Prob(Omnibus):    0.000         Jarque-Bera (JB):        159142.004
Skew:             -2.807         Prob(JB):                 0.00
Kurtosis:         26.041         Cond. No.                  22.3
=====

```

Рис. 3.13. Параметри обґрунтованої моделі для прогнозування вартості доставки насіння зернових культур з оптимальною кількістю функцій

Встановлено, що запропонована модель з оптимальною кількістю функцій із високою точністю виконує для прогнозування вартості доставки насіння зернових культур. Нижче на графіку (рис. 3.14-3.15) також показано, що прогнози здебільшого близькі до фактичних значень.

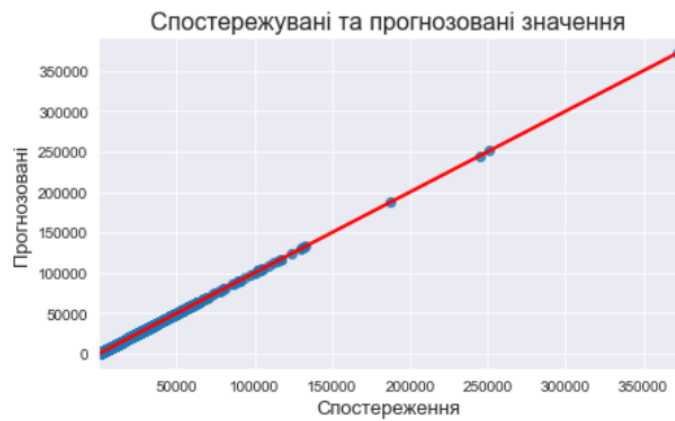


Рис. 3.14. Взаємозв'язки між прогнозованими та спостережуваними значеннями вартості доставки насіння зернових культур

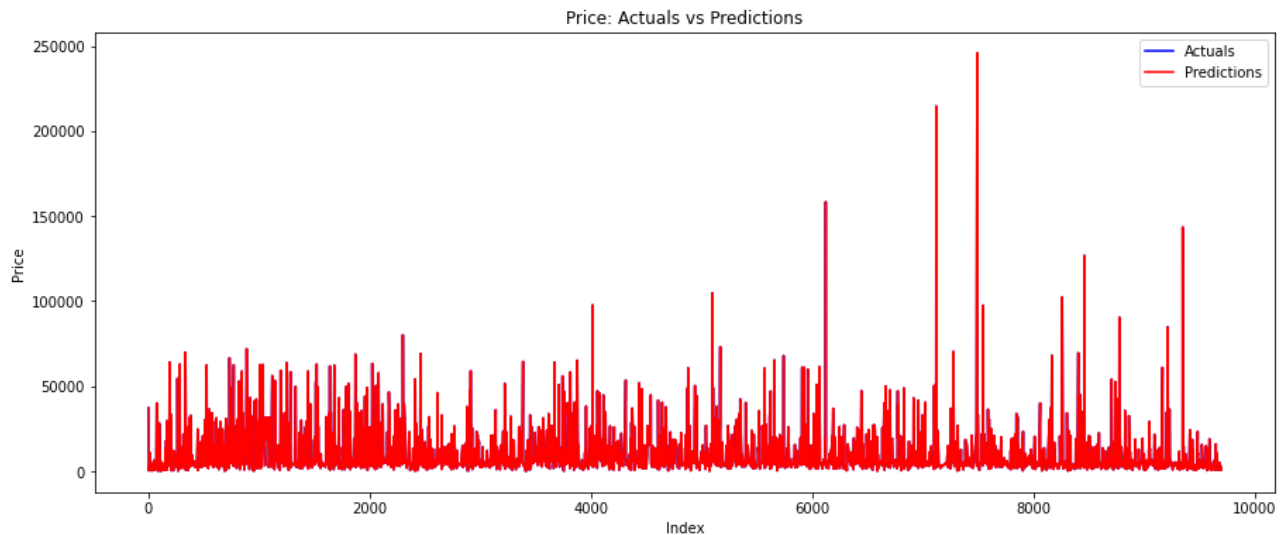


Рис. 3.15. Результати перевірки якості прогнозування вартості доставки насіння зернових культур із обґрунтованою моделлю

На підставі перевірки якості прогнозування вартості доставки насіння зернових культур із обґрунтованою моделлю встановлено, що вона забезпечує точність 96,5%. Отже, приймаємо, що у інформаційній системі прогнозування вартості доставки насіння зернових культур використовуватимемо модель лінійної регресії із обґрунтованими параметрами.

3.3. Компоненти інформаційної системи прогнозування вартості доставки насіння зернових культур

Основною метою проектування інформаційної системи прогнозування вартості доставки насіння зернових культур є вдосконалення процесу прийняття відповідних рішень та підвищення точності завдяки використанню обґрунтованої моделі машинного навчання. Цей процес включає розробку прикладного програмного забезпечення та навчання співробітників його використанню.

Прикладне програмне забезпечення, яке входить у зазначену систему, призначене для підтримки процесу прогнозування вартості доставки насіння зернових культур. Метою прикладного програмного забезпечення є перетворення даних на потрібну для менеджерів автотранспортних підприємств інформацію.

Окрім прикладного програмного забезпечення до складу інформаційної системи прогнозування вартості доставки насіння зернових культур входять:

- апаратне та системне програмне забезпечення, на якому працює прикладне програмне забезпечення. Зверніть увагу, що системне програмне забезпечення допомагає комп'ютеру функціонувати, тоді як прикладне програмне забезпечення допомагає користувачеві виконувати завдання, як підготовка електронної таблиці та підключення до Інтернету;
- документація та навчальні матеріали, які є матеріалами, створеними системним аналітиком, щоб допомогти користувачам використовувати програмне забезпечення, у створенні якого вони допомагали;
- елементи керування, які є частинами програмного забезпечення, створеного для запобігання шахрайству та крадіжці;
- виконавці, які використовують програмне забезпечення для виконання своєї роботи.

Компоненти інформаційної системи прогнозування вартості доставки насіння зернових культур представлено на рис. 3.16.



Рис. 3.16. Компоненти інформаційної системи прогнозування вартості доставки насіння зернових культур

Під час проектування інформаційної системи прогнозування вартості доставки насіння зернових культур слід розглядати усі аспекти системи, з особливим наголосом на розробці прикладного програмного забезпечення.

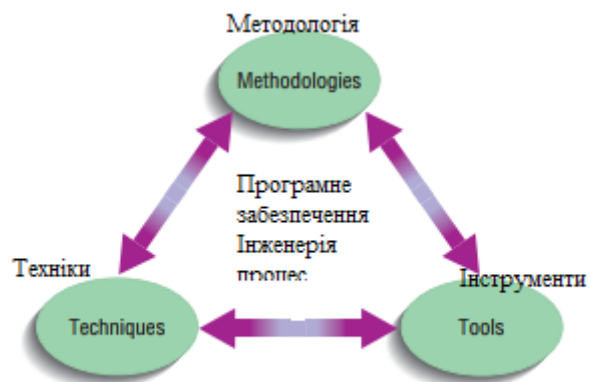


Рис. 3.17. Процес розробки програмного забезпечення для інформаційної системи прогнозування вартості доставки насіння зернових культур

Як показано на рис. 3.17, перевірені методології, техніки та інструменти займають центральне місце в процесах розробки програмного забезпечення.

3.4. Структурно-функціональне моделювання діяльності інформаційної системи

Техніка структурованого аналізу та проектування (SADT) – це методологія системної та програмної інженерії для опису систем як ієрархії функцій, які мають схематичне позначення, що розроблені для опису та розуміння системи. Модель SADT містить набір діаграм, які описують систему із визначеної точки зору та для певної мети. SADT містить блоки для представлення сутностей і дій, а також передбачає використання різноманітних стрілок, які відображають взаємозв’язані блоки із окремими пов’язаними діями. Їх можна використовувати як інструмент функціонального аналізу процесу прогнозування вартості доставки насіння зернових культур, використовуючи окремі послідовні рівні системи, яка деталізується.

Розглянемо роботу інформаційної системи прогнозування вартості доставки насіння зернових культур «FORECASTING THE COST OF GRAIN DELIVERY». Контекстна діаграма, яка відображає виконання зазначених процесів подана на рис. 3.17.



Рис. 3.17. Контекстна діаграма процесів у «FORECASTING THE COST OF GRAIN DELIVERY»

Окремі вхідні дані для функції «Прогнозування вартості доставки насіння зернових культур» є:

- вид насіння зернових;
- віддаль транспортування;
- обсяг вантажу.

На виході маємо прогнозовану вартості доставки насіння зернових культур.

Прогнозування вартості доставки насіння зернових культур керується окремими вимогами користувачів до процесу прогнозування, функціональними обмеженнями баз даних та знань, документацією Python та обґрунтованою моделлю машинного навчання для прогнозування вартості доставки насіння зернових культур.

До механізмів, а також використовуваних ресурсів належить Інтернет, бази даних та знань, а також розроблений програмний продукт.

Враховуючи те, що контекстна діаграма забезпечує загальний опис проєктованої системи, то у подальшому її потрібно виконати декомпозицію таким чином, щоб визначити її складові та зберегти цілісність структури. Цей процес забезпечує детальне знайомство із послідовно виконуваними роботами для досягнення мети – прогнозування вартості доставки насіння зернових культур.

Розроблена діаграма деталізується на три рівні, які пов'язані між собою. На першому із цих рівнів виконується декомпозиція батьківської діаграми, яка розчленовується на три окремі блоки, яку представлено на рис. 3.18. До цих блоків належать:

- заповнення бази даних;
- перевірка обмежень щодо виконання процесу доставки насіння зернових культур;
- оцінення виконання робіт із доставки насіння зернових культур.

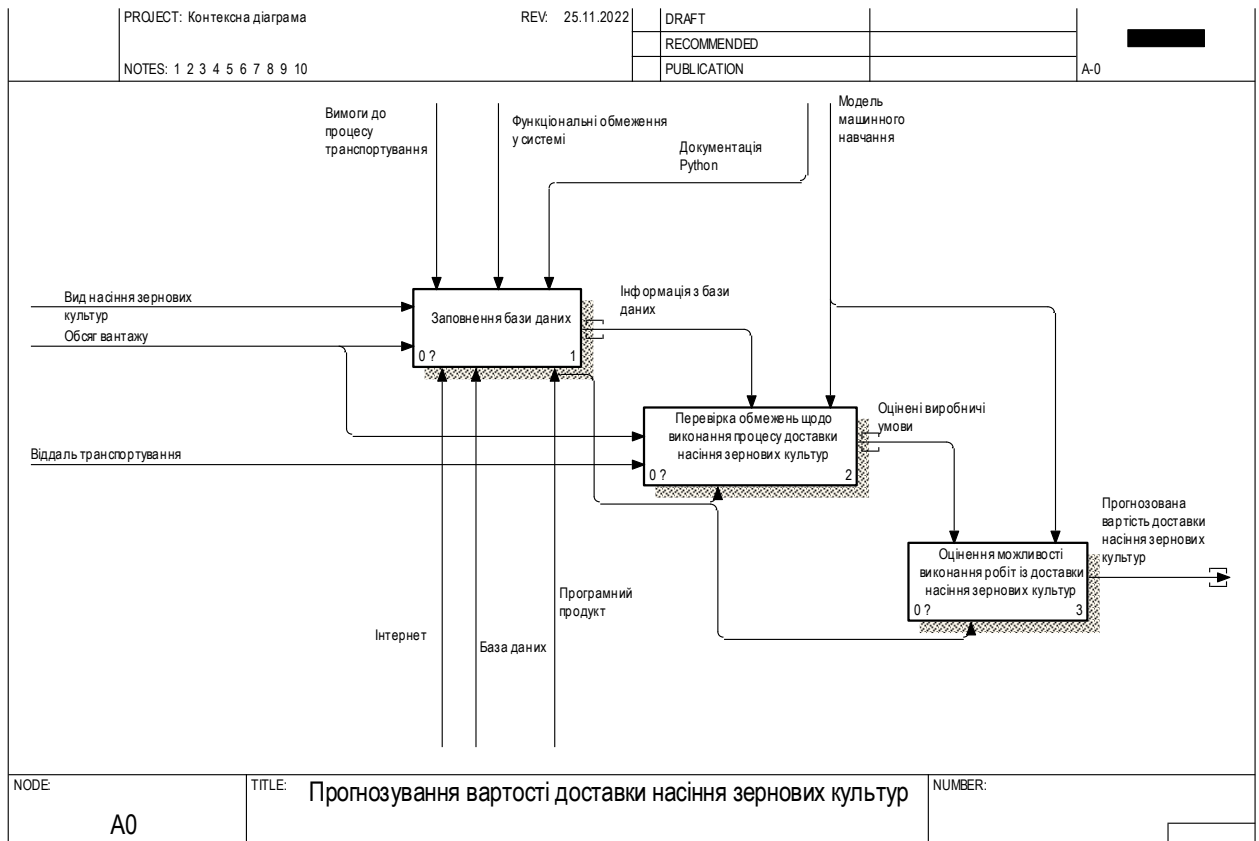


Рис. 3.18. Діаграма декомпозиції IDEF0 інформаційної системи прогнозування вартості доставки насіння зернових культур «FORECASTING THE COST OF GRAIN DELIVERY»

Входом для діяльності «Заповнення бази даних» є «Вид насіння зернових культур» та «Обсяг вантажу», а виходом – «інформація з бази даних». Процес впливу від управління включає «Вимоги до виконання процесу доставки насіння зернових культур», «Функціональні обмеження у системі» і «Документацію Python». Ресурси та механізми включають «Інтернет», «Бази даних» і «Необхідне програмне забезпечення».

Вхідними даними для дії «Перевірки обмежень щодо виконання процесу доставки насіння зернових культур» є «Інформація з бази даних», «Обсяг вантажу» та «Віддаль транспортування». При цьому першим елементом цієї діяльності є «Розрахункові значення обсягу виконаних транспортних робіт». Процес управлінського впливу включає «Вимоги до виконання процесу доставки насіння зернових культур», який відображено за допомогою стрілок

управління, «Розроблена модель машинного навчання», а ресурси та механізми – «необхідні програмні продукти».

Входом для діяльності «Оцінення виконання робіт із доставки насіння зернових культур» є «Очікувані виробничі умови». При цьому першою складовою цієї діяльності є «Фонд часу на виконання транспортних робіт». Процес управлінського впливу має «Розроблену модель машинного навчання», а ресурси та механізми є «необхідні програмні продукти».

3.5. Результати проектування головного вікна користувача

Під час створення головного модуля інформаційної системи прогнозування вартості доставки насіння зернових культур «FORECASTING THE COST OF GRAIN DELIVERY» насамперед слід виконати проектування головного вікна користувача. Для цього використовуємо бібліотеку Tkinter, що забезпечує створення інтерфейсу користувачів на мові програмування Python. У цій бібліотеці нами створено діалогове вікно користувача, яке представлено на рис. 3.19 із вікном отриманих результатів.

Forecasting the cost of grain delivery

FORECASTING THE COST OF GRAIN DELIVERY

Fri Nov 25 23:36:21 2022

Cars are available

Car: DAF_FT_XF_105 Cars, units: 2

Specify the characteristics of the transport order

	Distance, km	Cargo volume, tons	Type of cargo
Agricultural enterprise No. 1	10	20	Ріпак
Agricultural enterprise No. 2	52	15	Ріпак
Agricultural enterprise No. 3	47	47	Жито
Agricultural enterprise No. 4	32	12	Ячмінь
Agricultural enterprise No. 5	15	22	Пшениця

CALCULATE RESET

-----Results-----
forecasting the cost of grain delivery

Рис. 3.19. Вікно інформаційної системи прогнозування вартості доставки насіння зернових культур «FORECASTING THE COST OF GRAIN DELIVERY»

На основі розробленого шаблону створено діалогове вікно з відповідним функціоналом. Це діалогове вікно містить як поля введення, функціональні клавіші, так і вікно, яке використовується для відображення результатів прогнозування вартості доставки насіння зернових культур.

Спочатку нами створено блокнот у Tkinter (рис. 3.20). Для цього нами використано віджет Frame. Це дуже важливо для процесу групування та впорядкування інших віджетів. Він діє як контейнер, відповідальний за встановлення положення інших віджетів. Використано прямокутні області на екрані, щоб організувати свій макет і заповнити ці віджети. Фрейми також можна використовувати як базові класи для реалізації складних віджетів.

```

root = Tk()
root.geometry("1600x700+0+0")
root.title("Forecasting the cost of grain delivery")

Tops = Frame(root,bg="yellow",width = 1600,height=50,relief=SUNKEN)
Tops.pack(side=TOP)

f1 = Frame(root,width = 900,height=700,relief=SUNKEN)
f1.pack(side=LEFT)

f2 = Frame(root ,width = 600,height=400,relief=SUNKEN)
f2.pack(side=RIGHT)

#-----TIME-----
localtime=time.asctime(time.localtime(time.time()))

#-----INFO TOP-----
lblinfo = Label(Tops, font=( 'aria' ,30, 'bold' ),text="FORECASTING THE COST OF GRAIN DELIVERY",fg="gray22",bd=10,anchor='w')
lblinfo.grid(row=0,column=0)
lblinfo = Label(Tops, font=( 'aria' ,20, ),text=localtime,fg="gray22",anchor=W)
lblinfo.grid(row=1,column=0)

```

Рис. 3.20. Програмний код віджету Frame

Отриманий макет діалогового вікна користувача передбачає для них виведення результатів прогнозування вартості доставки насіння зернових культур, які подано у вигляді коду у форматі .py (рис. 3.21).

```

#-----Prediction results-----

status = Label(f2,font=('aria' ,16, 'bold'),width = 37, height=2, fg="steel blue", text="-----Results----- \n forecasting the cost of grain delivery", bd=2, relief=SUNKEN)
status.grid(row=0,column=0)

Rez=StringVar()

txtRez = Entry(f2,font=('ariel' ,16,'bold' ), textvariable=Rez , bd=6, width=35, insertwidth=4,bg="powder blue" , justify=LEFT)
txtRez.grid(row=1,column=0,
            padx=10,
            pady=10,
            ipadx=20,
            ipady=200
            )

```

Рис. 3.21. Програмний код результатів прогнозування вартості доставки насіння зернових культур

РОЗДІЛ 4.

ОХОРОНА ПРАЦІ ТА БЕЗПЕКА У НАДЗВИЧАЙНИХ СИТУАЦІЯХ

5.1. Аналіз небезпечних і шкідливих виробничих чинників та розробка заходів щодо покращення умов праці

Потенційно шкідливим чинником кабінету особи, яка приймає управлінські рішення, вважається небезпека враження людини електричним струмом. Важливим, але менш ймовірним чинником являється пожежна небезпека під час аварійної ситуації. Хімічні та біологічні джерела практично не мають впливу.

Перелік небезпечних та шкідливих виробничих чинників наведено у таблиці 5.1.

Таблиця 5.1. Небезпечні та шкідливі виробничі чинники

Фізичні	Електробезпека, пожежа, шум, мікроклімат
Хімічні	Відсутні
Біологічні	Відсутні
Психофізіологічні	Відсутні

В приміщенні кабінету особи, яка приймає управлінські рішення, присутні небезпечні чинники, та за умов дотримання заходів безпеки, вони не є критичним.

4.2. Розробка логічно-імітаційної моделі травматизму під час монтажу інформаційної системи

Проаналізувавши кожен із логічних моделей процесів формування та можливого виникнення травмонебезпечних та аварійних ситуацій, завжди

можна знайти подію з якої починається небезпечний процес ще до виникнення небезпечних наслідків.

Методикою оцінки рівня безпеки робочих місць, машин, виробничих процесів та окремих виробництв передбачено пошук об'єктивного критерію рівня безпеки для конкретного об'єкта. Таким показником вибрана ймовірність виникнення аварії, травми залежно від досліджуваного явища.

Для оцінки рівня безпеки певного об'єкта чи явища можна застосувати метод обчислення ймовірності виникнення будь-якого випадкового явища, який широко застосовують в зарубіжній інженерній практиці. Основні його принципи полягають в тому, що на основі обстеження робочого місця чи окремої машини виявляють виробничі небезпеки, можливі аварійні або травматичні ситуації. При оцінці ситуацій визначають події, які можуть стати головною подією при побудові логічно-імітаційної моделі травми. Після цього будують модель «дерева відмов і помилок оператора». При цьому важливе значення має правильний вибір головної події.

Головну подію (травма), модель якої нам необхідно побудувати, вибирають виходячи з оцінки відповідного об'єкта, виробництва чи окремої одиниці обладнання і змісту його найбільш небезпечного явища, яке за певних умов виробництва може виникнути.

Після вибору головного випадкового явища (події) розпочинаємо побудову моделі («дерева»). Використовуючи оператора «і» та «або», використовуємо набір ситуацій (відомих до цього), які можуть призвести до подій, вибраної як головна.

Після визначення відповідних травмонебезпечних ситуацій та їх кількості, визначаємо інші події, що входять до кожної такої ситуації, логічним аналізом із застосуванням операторів «і», «або» та інших. Процес побудови моделі триває, поки не будуть знайдені усі базові події, що визначають межу моделі.

Слід мати на увазі, що кожна випадкова подія, до якої входять базові події, може формуватися й виникати при входженні у неї двох, трьох і більше базових подій за допомогою відповідних операторів.

Повністю побудована і перевірена модель підлягає математичній обробці для визначення ймовірності кожної випадкової події, що увійшла до моделі, починаючи з базових і закінчуючи головною.

Ймовірність базових подій визначаємо за даними виробництва. Наприклад, базова подія «стан контролю з охорони праці». Для визначення ймовірності ми повинні встановити, наскільки (у відсотках) від ідеального рівня здійснюється відповідний контроль на об'єкті. Якщо буде встановлено, що такий рівень контролю становить 50% або 30%, то ймовірність відповідно дорівнює 0,5 і 0,3. При відсутності контролю ймовірність «не здійснення контролю» становитиме 1, якщо контроль ідеальний, то відповідно ймовірність дорівнює 0.

Після обчислення ймовірності всіх подій, розміщених у ромбах, і базових подій, починаючи з лівої нижньої гілки «дерева», позначаємо номерами всі випадкові події, що увійшли до моделі.

На цьому можна вважати, що певна модель підготовлена до математичних обчислень ймовірностей випадкових подій логічно-імітаційної моделі

Отже, для побудови логіко-імітаційної моделі процесу, формування і виникнення аварії та травми під час монтажу інформаційної системи прогнозування вартості доставки насіння зернових культур складемо список базових подій. Вони лежатимуть у основі даної моделі. Кожному пункту списку присвоюємо певне значення ймовірності виникнення. Нижче подано сам список:

- | | |
|--|---------------------------------|
| 1. Стан контролю з охорони праці | $P_1 = 0,2$; |
| 2. Несерйозне відношення до проходження ТО інструменту | $P_2 = 0,1$; |
| 3. Відсутність комплектуючих установки..... | $P_3 = 0,2$; |
| 4. Невисока міцність | $P_4 = 0,03$; |

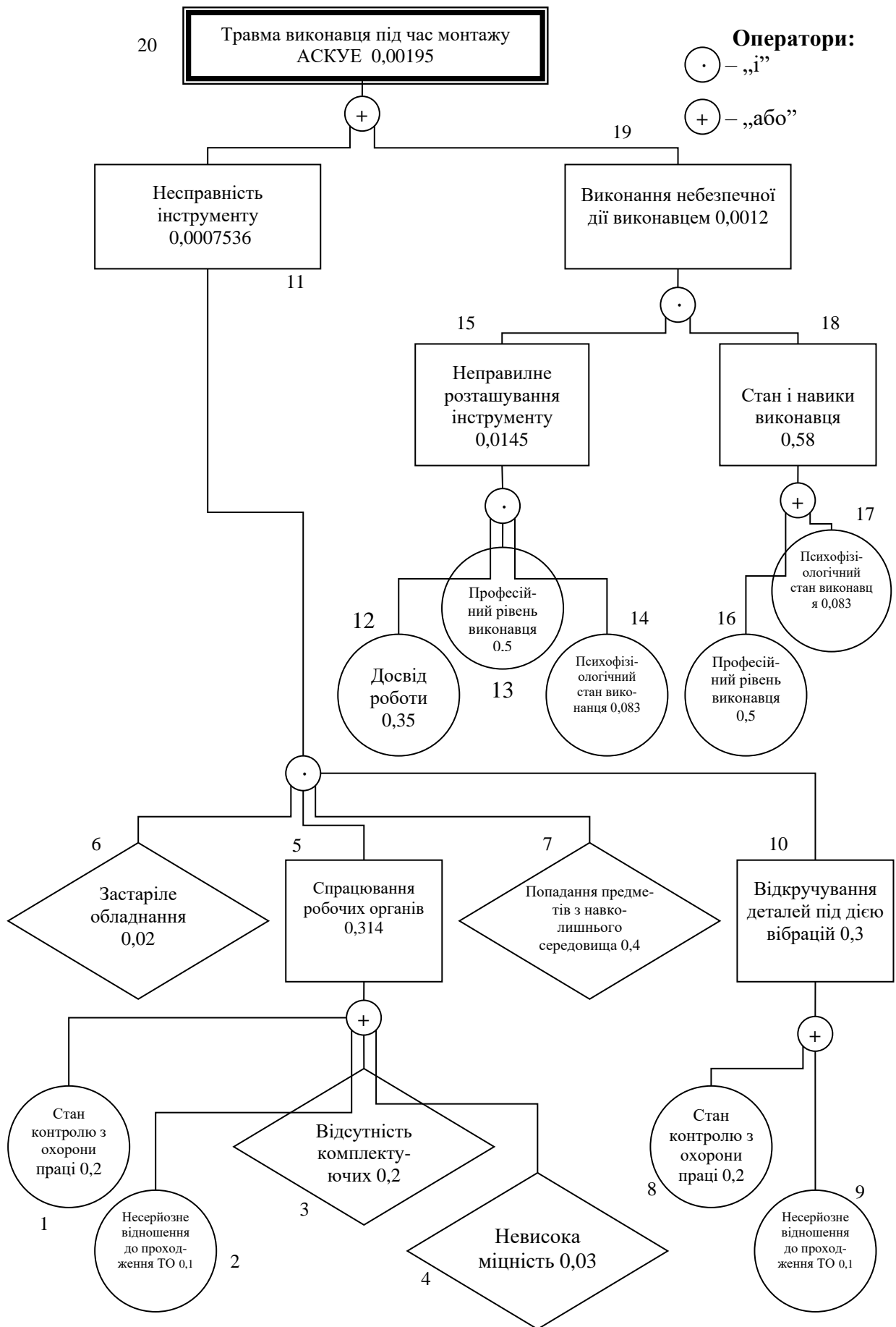


Рис. 4.1. Логіко-імітаційна модель процесу формування та виникнення аварії та травми під час монтажу інформаційної системи прогнозування вартості доставки насіння зернових культур

1. Використання застарілого обладнання..... $P_6 = 0,02$;
2. Попадання сторонніх предметів $P_7 = 0,4$;
3. Досвід роботи виконавця $P_{12} = 0,35$.
4. Професійний рівень виконавця $P_{13} = 0,5$;
5. Психофізіологічний стан виконавця..... $P_{14} = 0,083$;

На основі даного списку будуємо матрицю логічних взаємозв'язків між окремими пунктами, графічне представлення якої зображено на рис. 4.1.

Розрахуємо ймовірності виникнення подій, що входять у дану логіко-імітаційну модель процесу монтажу інформаційної системи прогнозування вартості доставки насіння зернових культур (на прикладі ймовірності отримання травми виконавця).

Ймовірність виникнення події P_5 визначаємо наступним чином:

$$P_5 = 0,2 + 0,1 + 0,2 + 0,003 - 0,2 \cdot 0,1 - 0,2 \cdot 0,03 - 0,2 \cdot 0,03 - 0,1 \cdot 0,2 - 0,1 \cdot 0,03 - 0,2 \cdot 0,03 + 0,2 \cdot 0,1 \cdot 0,2 + 0,1 \cdot 0,2 \cdot 0,03 + 0,2 \cdot 0,1 \cdot 0,2 + 0,2 \cdot 0,1 \cdot 0,03 - 0,2 \cdot 0,1 \cdot 0,2 \cdot 0,03 = 0,314$$

Ймовірність виникнення події P_{10} визначаємо так:

$$P_{10} = 0,2 + 0,1 = 0,3.$$

Ймовірність виникнення події P_{11} визначаємо:

$$P_{11} = 0,02 \cdot 0,314 \cdot 0,4 \cdot 0,3 = 0,00075.$$

Ймовірність виникнення події P_{15} визначаємо наступним чином:

$$P_{15} = 0,35 \cdot 0,5 \cdot 0,083 = 0,0145.$$

Ймовірність події P_{18} :

$$P_{18} = 0,5 + 0,083 = 0,58.$$

Ймовірність події P_{19} :

$$P_{19} = 0,0145 \cdot 0,083 = 0,0012.$$

Ймовірність події P_{20} :

$$P_{20} = 0,00075 + 0,0012 = 0,00195.$$

Ймовірність травми рівна ймовірності виникнення аварії, бо остання можлива лише за умови монтажу автоматизованої системи управління енергоспоживанням людиною.

Логіко-імітаційні моделі аварій і травм допомагають зменшити ймовірність виникнення аварійних та травмонезбезпечних ситуацій. Якщо необхідно оцінити рівень небезпеки будь-якого робочого місця, слід уважно вивчити і побудувати логічні моделі можливих небезпечних ситуацій, які охоплюють як стан обладнання і самого робочого місця, так і поведінку працюючого і обчислити ймовірність виникнення травми.

Після аналізу результатів моделювання ймовірність виникнення травми можна звести до дуже малої величини – достатньо зменшити вплив ймовірностей вихідних факторів, які до неї призводять.

4.3. Розробка заходів щодо безпеки у надзвичайних ситуаціях

Науково-технічний прогрес радикально змінив світ, породивши нові загрози для цивілізації. У житті сучасної людини все більше місце займають турботи, пов'язані з подоланням різних кризових явищ, що виникають в процесі розвитку земної цивілізації. В Україні, як і в усьому світі, в останні роки спостерігається зростання числа військових дій, катастроф природного та техногенного характеру. Це обумовлено, перш за все, прогресуючої урбанізацією територій, збільшенням щільності населення Землі, і, як наслідок, збільшенням антропогенного навантаження на навколишнє середовище. Захист природних систем і населення від надзвичайних ситуацій різного характеру сформувалася в останні роки як нагальна і об'єктивна потреба суспільства і держави.

Заходи щодо захисту цивільного населення плануються проводяться по населених пунктах де розміщені підприємства і охоплюють населення навколишніх сіл. Водночас характер та зміст захисних засобів встановлюються

вид ступеня загрози, місцевих умов з урахуванням важливості виробництва для безпеки населення і інших економічних і соціальних чинників.

Основні заходи щодо захисту населення плануються та здійснюються завчасно і мають випереджувальний характер, це стосується насамперед підготовки, підтримання у постійній готовності індивідуальних та колективних засобів захисту, їх накопичення, а також підготовки до проведення евакуації населення із зон підвищеного ризику.

Також раз в три роки проводяться навчання по підготовці близьких до військових дій, що в разі небезпеки могло би не дістати людину зненацька. Керівництво докладає максимум зусиль, щоб працівники підприємств були хоча би мінімально захищені в разі будь-якої небезпеки пов'язаної з тими чи іншими обставинами.

РОЗДІЛ 5.

ВИЗНАЧЕННЯ ЕФЕКТИВНОСТІ ВІД ВИКОРИСТАННЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ

Економічну ефективність від використання розробленої інформаційної системи отримують користувачі (диспетчери автотранспортних підприємств) завдяки точному прогнозуванню вартості доставки насіння зернових культур. Використання цієї інформаційної системи дає можливість знизити втрати автотранспортних підприємств із-за точного оцінення вартості доставки насіння зернових культур та відкидання замовлень, які призводять до збитків або недостатнього отримання прибутку.

Особливістю цієї інформаційної системи є потреба створення відповідної бази історичних даних дані для прогнозування вартості доставки насіння зернових культур. Вартість технічних засобів, а також програмного забезпечення для розробленої інформаційної системи прогнозування вартості доставки насіння зернових культур включають витрати на розробку програмних продуктів, придбання основних технічних компонентів та виконання монтажних робіт. На підставі розрахунків визначено вартість програмно-технічних засобів інформаційної системи прогнозування вартості доставки насіння зернових культур у сумі 38200 грн.

Для забезпечення рентабельності розробленої інформаційної системи прогнозування вартості доставки насіння зернових культур у обсязі $P_m = 10\%$ вартість запропонованої інформаційної системи становить:

$$C_m = C_n + C_n \cdot (P_m / 100). \quad (5.1)$$

Підставивши потрібні значення у формулу (5.1) зможемо визначити:

$$C_m = 38200 + 38200 \cdot (10 / 100) = 42020 \text{ грн.}$$

Балансову вартість інформаційної системи, розробленої для прогнозування вартості доставки насіння зернових культур, визначено з урахуванням її встановлення та налаштування:

$$C_{\text{бал}} = C \cdot K_{\text{mn}} \quad (5.2)$$

де C – вартість усіх елементів інформаційної системи; K_{mn} – коефіцієнт витрат на встановлення та налаштування ($K_{\text{mn}} = 1.1$).

Підставивши потрібні значення у формулу (5.2) зможемо визначити вартість інформаційної системи прогнозування вартості доставки насіння зернових культур:

$$C_{\text{бал}} = 42020 \cdot 1,1 = 46222 \text{ грн.}$$

Поточні витрати (річні витрати на обслуговування) необхідні для використання розробленої інформаційної системи прогнозування вартості доставки насіння зернових культур. Ці компоненти враховують заробітну плату, амортизаційні відрахування, рахунки за електроенергію та їх обслуговування.

Операційні витрати при обслуговуванні інформаційної системи становитимуть 26234 грн. Вартість розробки та тестування програмного забезпечення для інформаційних систем – 41 600 грн.

Загальна вартість інформаційної системи прогнозування вартості доставки насіння зернових культур для автотранспортного підприємства становить 79800 грн.

Економічна ефективність від інформаційної системи прогнозування вартості доставки насіння зернових культур для автотранспортного підприємства визначається за формулою:

Таблиця 5.1. Результати визначення економічної ефективності інформаційних систем прогнозування вартості доставки насіння зернових культур для автотранспортного підприємства

№ п/п	Показники	Одиниця виміру	Значення показників
1	Вартість програмно-технічних засобів інформаційної системи	грн	38200
2	Експлуатаційні витрати інформаційної системи	грн	26234
3	Вартість розробки та тестування програмного забезпечення для інформаційних систем	грн	41 600
4	Собівартість інформаційної системи	грн	79800
5	Приведені витрати на функціонування інформаційної системи	Грн / рік	34214
6	Економічна ефективність	Грн / рік	171381
7	Термін окупності капіталовкладень	років	0,46

$$E_{ICPI} = (P_1 - P_2) - Z_{ICPI} , \quad (5.3)$$

де P_1 – обсяг втрат автотранспортного підприємства із інформаційною системою прогнозування вартості доставки насіння зернових культур, грн.; P_2 – обсяг втрат автотранспортного підприємства без використання інформаційної системи прогнозування вартості доставки насіння зернових культур, грн.; Z_{ICPI} – річні приведені витрати на інформаційну систему прогнозування вартості доставки насіння зернових культур, Грн / рік.

Річні приведені витрати на використання інформаційної системи прогнозування вартості доставки насіння зернових культур становлять:

$$Z_{ICП} = E_n \cdot C_{бал} + B_p, \quad (5.4)$$

Підставивши потрібні значення у формулу (5.2) зможемо визначити річні приведені витрати на використання інформаційної системи прогнозування вартості доставки насіння зернових культур:

$$Z_{ICП} = 0,1 \cdot 79900 + 26234 = 34214 \text{ грн.}$$

Підставивши потрібні значення у формулу (5.2) зможемо визначити економічну ефективність від використання інформаційної системи прогнозування вартості доставки насіння зернових культур:

$$E_{ICП} = (238140 - 32545) - 34214 = 171381 \text{ грн.}$$

Термін окупності капіталовкладень у інформаційну систему прогнозування вартості доставки насіння зернових культур визначається за формулою:

$$T_{ок} = \frac{Z_{ICП}}{E_{ICП}}. \quad (5.5)$$

Підставивши потрібні значення у формулу (5.2) зможемо визначити термін окупності капіталовкладень:

$$T_{ок} = \frac{79800}{171381} = 0,46 \text{ року.}$$

ВИСНОВКИ І ПРОПОЗИЦІЇ

У сучасному світі інформаційні системи призначені для вирішення конкретних завдань практичного характеру. Логістична інфраструктура транспортних підприємства в цілому, потребує розробки інформаційних систем прогнозування вартості транспортування насіння зернових культур. Використання зазначених інформаційних систем під час транспортування зернових культур може вирішити наступні задачі зниження транспортних витрат під час доставки зернових культур та пришвидшити виконання замовлень.

Нами виконано аналіз стану перевезень вантажів в Україні. Встановлено, що існує обмежене використання сучасних інформаційних технологій для планування, водночас ведеться обмін даними при транспортуванні вантажів в Україні та наявні власні їх накопичувачі.

Виконаний аналіз існуючих інформаційних систем для планування вантажних перевезень свідчить про їх вагомість для практики. Існує доцільність використання інформаційних систем для планування та прогнозування діяльності транспортних підприємств із використанням наявних даних, що є досить актуальним на даний час. Однак вони мають низку недоліків. При цьому ринок інформаційних систем та технологій із використанням машинного навчання стрімко зростає. Починаючи із 2016 року він перевищив \$1 млрд та, за прогнозами аналітиків, може зрости до \$39,98 млрд до 2025 року.

Проведений аналіз стану розробки та використання інформаційних систем для транспортних підприємств дав можливість встановити, що існує актуальна задача прогнозування вартості транспортування насіння зернових культур. Для цього слід розробити інформаційну систему, яка базується на машинному навчанні та забезпечить зростання точності прогнозування і усуне недоліки існуючих інформаційних систем. Це зумовлює необхідність розроблення інформаційних систем прогнозування вартості транспортування насіння зернових культур.

Нами обґрунтовано особливості вирішення задачі прогнозування вартості транспортування насіння зернових культур на підставі машинного навчання. Метою машинного навчання є створення систем, здатних знаходити шаблони в даних, навчатися на них без втручання людини та явного перепрограмування. Щоб вирішити задачу прогнозування вартості транспортування насіння зернових культур, дослідники з обробки даних спочатку повинні зрозуміти, які дані використовувати для навчання моделей машинного навчання, і саме тому потрібна описова аналітика.

Підготовка даних може бути одним із найскладніших кроків у будь-якому проєкті машинного навчання. Причина в тому, що кожен набір даних є різним і дуже специфічним для проєкту. Тим не менш, у проєкті прогнозування вартості транспортування насіння зернових культур є достатньо спільних ознак із іншими проєктами машинного навчання із використанням регресії. При цьому можна визначити окрему послідовність кроків і підзавдань, які виконуватимемо у подальшому.

Нами здійснено вибір методу машинного навчання для прогнозування вартості транспортування насіння зернових культур. Встановлено, що лінійна регресія є одним із найважливіших та широко використовуваних методів машинного навчання. Водночас, це один з найпростіших методів машинного навчання. Однією з його головних переваг є простота інтерпретації результатів.

Нами проаналізовано існуючі види лінійних регресій. Проста або однофакторна лінійна регресія є найпростішим випадком лінійної регресії, оскільки вона має одну незалежну змінну. Багатофакторна або багатовимірна лінійна регресія – це випадок лінійної регресії з двома або більше незалежними змінними. Можна розглядати поліноміальну регресію як узагальнений випадок лінійної регресії. При цьому припускають, що поліноміальна залежність між виходом і вхідними даними має поліноміальну оцінену функцію регресії.

Для обґрунтування лінійних регресій вибираємо мову Python. Вона є дуже популярною сьогодні та потужною мовою програмування. Також нами обрано Jupyter Notebook як середовище, яке можна використовувати для інтерактивних

експериментів із Python. Також використана безкоштовна програмна бібліотека машинного навчання scikit-learn, побудована на основі NumPy та деяких інших бібліотеках. За умови побудови лінійних регресій слід використовувати функціональності, що виходить за рамки scikit-learn. Для цього підходить бібліотека statsmodels.

На підставі отриманого аналітичного набору даних компанії «Вест Агро Груп» щодо процесів транспортування насіння зернових культур автомобілями DAF FT XF 105 виділено наступні їх атрибути, що характеризують вартості транспортування насіння зернових культур (рис. 3.1). Із представленого аналітичного набору даних компанії «Вест Агро Груп» щодо процесів транспортування насіння зернових культур (рис. 3.1) можна сказати, що він містить 9702 спостереження за 8 атрибутами (рис. 3.2). Для прогнозування вартості транспортування насіння зернових культур зазначений набір даних має 6 вхідних змінних ($X_1 \dots X_6$) та 1 вихідну змінну (Y_1 – Вартість доставки, грн). Зазначені дані лежать в основі вирішення задачі лінійної регресії.

Встановлено, що найбільше замовлень на транспортування припадає на пшеницю – 4920 од та ріпак – 3026 од. На наступному етапі виконано побудову гістограм зміни основних даних за їх встановленими атрибутами (рис. 3.5) проведено оцінення вартості доставки насіння зернових культур (рис. 3.6).

На підставі попереднього аналізу даних щодо доставки насіння зернових культур встановлено, що є між ними сильні взаємозв'язки (наприклад, між фактичними витратами палива та відстанню, вантажообігом та вартістю доставки), а також такі, де зв'язки відсутні (наприклад, питомими та фактичними витратами палива).

Отримана кореляційна матриця відображає коефіцієнти кореляції для різних змінних. Встановлено, що найвищий коефіцієнт кореляції між вантажообігом, відстанню, фактичними витратами палива та вартістю доставки насіння зернових культур відповідно становить – 1,0, 0,7 та 0,75. Водночас, найнижчий коефіцієнт кореляції між питомими витратами палива та вартістю доставки насіння зернових культур відповідно становить – 0,099.

Виконаний аналіз набору даних для прогнозування вартості доставки насіння зернових культур за встановленими атрибутами лежить в основі проведення масштабування функцій, підбору моделі та виконання її навчання. З метою проведення навчання моделі для прогнозування вартості доставки насіння зернових культур підготовлені дані розбито на навчальну вибірку – 70% та тестову вибірку – 30%.

У подальшому виконуємо побудову моделі та здійснюємо вибір функцій за допомогою RFE. Оскільки вартість доставки насіння зернових культур, яка є залежною змінною, і описується із більшістю незалежних змінних лінійними функціями, використовуємо лише лінійну регресію, а не інші типи регресії, такі як поліноміальна, випадковий ліс тощо. На підставі цього представляємо модель з усіма функціями, використовуючи бібліотеку Sklearn із моделлю машинного навчання LinearRegression (рис. 3.11).

З наведених вище графіків (рис. 3.13) знаходимо, що R_square для тестових даних досягає максимуму на 13 функціях, і на цьому етапі модель добре прогнозує, оскільки значення R^2 однакове і не залежить від кількості функцій. Середньоквадратична похибка RMSE для тестових даних є найнижчою за використання 13 функцій і зі зростання кількості функцій, далі зростає. Навчання RMSE на рівні 13 функцій також виглядає на належному рівні, а додавання додаткових функцій до навчання зменшує RMSE. У нашій моделі будемо використовувати 13 функцій.

Обґрунтована модель для прогнозування вартості доставки насіння зернових культур з оптимальною кількістю функцій представлена на рис. 3.14. Встановлено, що запропонована модель з оптимальною кількістю функцій із високою точністю виконує для прогнозування вартості доставки насіння зернових культур. Нижче на графіку (рис. 3.14-3.15) також показано, що прогнози здебільшого близькі до фактичних значень.

На підставі перевірки якості прогнозування вартості доставки насіння зернових культур із обґрунтованою моделлю встановлено, що вона забезпечує точність 96,5%. Отже, приймаємо, що у інформаційній системі прогнозування

вартості доставки насіння зернових культур використовуватимемо модель лінійної регресії із обґрунтованими параметрами.

Основною метою проектування інформаційної системи прогнозування вартості доставки насіння зернових культур є вдосконалення процесу прийняття відповідних рішень та підвищення точності завдяки використанню обґрунтованої моделі машинного навчання. На обґрунтовано компоненти інформаційної системи прогнозування вартості доставки насіння зернових культур.

Нами виконано структурно-функціональне моделювання діяльності інформаційної системи. Розглянуто роботу інформаційної системи прогнозування вартості доставки насіння зернових культур «FORECASTING THE COST OF GRAIN DELIVERY». Це дало можливість побудувати контекстну діаграму, яка відображає виконання зазначених процесів подана на рис. 3.17. Розроблена діаграма деталізується на три рівні, які пов'язані між собою.

На основі розробленого шаблону створено діалогове вікно з відповідним функціоналом. Це діалогове вікно містить як поля введення, функціональні клавіші, так і вікно, яке використовується для відображення результатів прогнозування вартості доставки насіння зернових культур.

Розроблені заходи із охорони праці та безпеки у надзвичайних ситуаціях, які передбачають створення безпечних умов праці під час роботи з інформаційною системою прогнозування вартості доставки насіння зернових культур.

На підставі розрахунків показників економічної ефективності від використання інформаційної системи прогнозування вартості доставки насіння зернових культур дасть можливість автотранспортному підприємству економити кошти у розмірі 171381 грн. Термін окупності капіталовкладень у інформаційну систему становить 0,46.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. АС Клієнт УЗ. Автоматизована система з оформлення та обробки перевізних документів на перевезення вантажів залізничним транспортом України вантажовідправниками через мережу Інтернет [Текст]: керівництво користувача. К.:2016, 69с.
2. Вантажні автомобільні транспортні біржі в Україні [Електронний ресурс]: [Веб-сайт]. Електронні дані. Режим доступу: <https://logistics-ukraine.com/2017/09/18/вантажні-автомобільні-тран-спортні-б/> (дата звернення 12.08.2022)
3. Введення в машинне навчання за допомогою Python и Scikit-Learn. URL: <https://habr.com/ua/company/mlclass/blog/247751/> (дата звернення: 20.05.2022).
4. Використання інформаційно-комунікаційних технологій на підприємствах [Електронний ресурс]. Державна служба статистики України. Режим доступу: https://ukrstat.gov.ua/operativ/operativ2018/zv/ikt/arh_ikt_u.html (дата звернення 01.09.2022).
5. Галузеві рішення для автоматизації зернового і транспортного бізнесу перевезень [Електронний ресурс]: [Веб-сайт]. Електронні дані. Режим доступу: <https://artport.pro/> (дата звернення 14.09.2022)
6. Григорович О.В. Застосування багат шарових перцептронів для класифікації позичальників юридичних осіб. Нейронечіткі технології моделювання в економіці. Науково-аналітичний журнал. Київ, 2019. №8. С.48-64.
7. Ефективно, безпечно та автоматизовано! [Електронний ресурс]: [Веб-сайт]. Електронні дані. Режим доступу: https://trademaster.ua/dir_logistik/313476 (дата звернення 01.06.2022)
8. Жидецький В.Ц., Джигирей В.С., Мельников О.В. Основи охорони праці. Підручник. Вид. 5-е, доповнене. Львів: Афіша, 2012. 350с.

9. Класифікація в Python з Scikit-Learn та Pandas. URL: <https://stackabuse.com/classification-in-python-with-scikit-learn-and-pandas/> (дата звернення: 17.05.2022).
10. Лехман С.Д., Рублев В.І., Рябцев Б.І. Запобігання аварійності і травматизму у сільському господарстві. К.: Урожай, 1993. 267 с.
11. Лутц М. Программирование на Python. I том. СПб.: Символ-плюс, 2015. 992 с.
12. Марк Саммерфилд. Программирование на Python 3. Подробное руководство. Пер. с англ. СПб.: Символ-Плюс, 2013. 608 с.
13. Матвійчук А. В. Штучний інтелект в економіці: нейронні мережі, нечітка логіка: монографія. Київ, КНЕУ, 2011. 439 с.
14. Міжнародні вантажоперевезення онлайн Lardi-Trans [Електронний ресурс]: [Веб-сайт]. Електронні дані. Режим доступу: <https://lardi-trans.com/> (дата звернення 01.11.2022)
15. Обсяги перевезених вантажів за видами транспорту за 2021 рік [Електронний ресурс]. Державна служба статистики України. Режим доступу: https://ukrstat.gov.ua/operativ/operativ2022/tr/tr_rik/opvvt_22_ue.xlsx (дата звернення 22.08.2022).
16. Овчарук І. В., Боклаг Є. В. Інформаційні системи на залізничному транспорті: розвиток та перспективи [Текст]/ І. В. Овчарук, Є. В. Боклаг. Цифрова платформа: інформаційні технології в соціокультурній сфері. 2020. Том 3. №2. С. 170-182.
17. Плєскач В.Л., Рогущина Ю.В., Кустова Н.П. Інформаційні технології та системи. К.: Книга, 2004. 519 с.
18. Про затвердження Змін до Правил планування перевезень вантажів :Наказ; Мінтрансзв'язку України від 21.06.2007 №552 [Електронний ресурс] // База даних «Законодавство України» / Верховна Рада України. Режим доступу: <https://zakon.rada.gov.ua/go/z0787-07> (дата звернення 10.10.2022)
19. Про схвалення Національної транспортної стратегії України на період до 2030 року : Розпорядження Кабінету Міністрів України; Стратегія

від 30.05.2018 No 430-р[Електронний ресурс]. База даних «Законодавство України» / Верховна Рада України. Режим доступу: <https://zakon.rada.gov.ua/go/430-2018-%D1%80> (дата звернення: 01.09.2022).

20. Програмне забезпечення та послуги в галузі вантажних перевезень [Електронний ресурс]: [Веб-сайт]. Електронні дані. Режим доступу: <http://kpd-uz.com/ua/products/arm.php> (дата звернення 21.10.2022)

21. Прохоренок Н.А. Python 3 и PyQt. Разработка приложений. СПб.: БХВ-Петербург, 2012. 704 с.

22. Система GPS-моніторингу транспорту та вантажів [Електронний ресурс]: [Веб-сайт]. Електронні дані. Режим доступу: <https://smartgps.com> (дата звернення 11.10.2022)

23. Функціональні можливості системи «ТМкарта» [Електронний ресурс]: [Веб-сайт]. Електронні дані. Режим доступу: <http://tmkarta.com/uk/about/long.php> (дата звернення 01.10.2022)

24. Хахаев И.А. Практикум по алгоритмизации и программированию на Python. М.: АЛЬТ Линукс, 2010. 126 с.

25. Ahmed A.M., Rizaner A., and Ulusoy A.H., A Decision Tree Algorithm Combined with Linear Regression for Data Classification, 2018 Int. Conf. Comput. Control. Electr. Electron. Eng. ICCSEEE 2018, no. August, pp. 1–5, doi: 10.1109/ICCSEEE.2018.8515759.

26. Bevans R. Simple Linear Regression: An Easy Introduction & Examples. 2020. URL: <https://www.scribbr.com/statistics/multiple-linear-regression/>

27. Bylander T. Estimating generalization error on twoclass datasets using out-of-bag estimates. Machine Learning, 2002. 48, 18, pp. 287–297.

28. Crook J., Banasik J. Forecasting and explaining aggregate consumer credit delinquency behaviour, Int. J. Forecasting 28 (1) (2012) 145–160.

29. Degruz. Сайт вантажоперевезення Україна [Електронний ресурс]: [Веб-сайт]. Електронні дані. Режим доступу: <https://degruz.com> (дата звернення 07.10.2022)

30. DELLA. Вантажні перевезення [Електронний ресурс]: [Веб-сайт]. Електронні дані. Режим доступу: <https://della.ua> (дата звернення 10.08.2022)
31. Hertzmann A., Fleet D. J., and Brubaker M. Linear Regression. 2015. URL: <http://www.cs.toronto.edu/~mbrubake/teaching/C11/Handouts/LinearRegression.pdf>
32. Lading. Транспортна платформанового покоління [Електронний ресурс]: [Веб-сайт]. Електронні дані. Режим доступу: <https://lading.ua> (дата звернення 01.06.2022)
33. Sovtes-онлайн логістика, управління вантажоперевезеннями [Електронний ресурс]: [Веб-сайт]. Електронні дані. Режим доступу: <https://sovt.es.ua/> (дата звернення 03.09.2022).
34. TrucksNearMe [Електронний ресурс]: [Веб-сайт]. Електронні дані. Режим доступу: <https://trucksnearme.com/app/> (дата звернення 08.08.2022)