НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ «КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ імені Ігоря СІКОРСЬКОГО» НАВЧАЛЬНО-НАУКОВИЙ ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ
Кафедра математичного моделювання та аналізу даних

«До захисту допущено»

В.о. завідувача кафедри

_____Іван ТЕРЕЩЕНКО

« »_____2023 р.

ДИПЛОМНА РОБОТА

на здобуття ступеня бакалавра

зі спеціальності: 113 Прикладна математика
на тему: **«Математичні методи аналізу пошкоджених територій аграрного сектору України»**

Виконала: студентка 4 курсу, групи ФІ-91
Вакулюк Вероніка Василівна
Керівник: ст. викладачка кафедри ММАД ННФТІ, д-р філософії Яйлимова Г.О.
Рецензент: д-р філософії, с.н.с. ІКД Лавренюк М.С.

Засвідчую, що у цій дипломній роботі немає запозичень з праць інших авторів без відповідних посилань.
Студент _____

Київ — 2023

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ «КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ імені Ігоря СІКОРСЬКОГО» НАВЧАЛЬНО-НАУКОВИЙ ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ
Кафедра математичного моделювання та аналізу даних

Рівень вищої освіти — перший (бакалаврський)
Спеціальність (освітня програма) — 113 Прикладна математика,
ОПП «Математичні методи моделювання, розпізнавання образів та комп'ютерного зору»

ЗАТВЕРДЖУЮ
В.о. завідувача кафедри
_____Іван ТЕРЕЩЕНКО
« »_____2023 р.

ЗАВДАННЯ

на дипломну роботу

Студентка: Вакулюк Вероніка Василівна

1. Тема роботи: *«Математичні методи аналізу пошкоджених територій аграрного сектору України»,*

керівник: старший викладач, д. ф. Яйлимова Г.О.,

затверджені наказом по університету №__ від « »_____ 2023 р.

2. Термін подання студентом роботи: « »_____ 2023 р.

3. Вихідні дані до роботи: *опубліковані джерела за тематикою дослідження.*

4. Зміст роботи: *досліджуючи супутникові знімки, побудовано логістичну регресійну модель, яка класифікує пошкоджені території аграрного сектору України.*

5. Перелік ілюстративного матеріалу (із зазначенням плакатів, презентацій тощо): *презентація доповіді*.

6. Дата видачі завдання: 6 лютого 2023 р.

Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання	Примітка
1	Узгодити тему роботи з науковим керівником	6-17 лютого 2023 р.	Виконано
2	Оглянути опубліковані джерела за тематикою дослідження	Лютий - березень 2023 р.	Виконано
3	Дослідити аграрний сектор України	5-17 березня 2023 р.	Виконано
4	Визначити види пошкоджень територій	17 - 30 березня 2023 р.	Виконано
5	Зібрати дані для обчислень	5 - 13 квітня 2023 р.	Виконано
6	Провести аналіз та побудувати модель	Квітень 2023 р.	Виконано
7	Організувати дані, оцінити результати розрахунків	Травень 2023 р.	Виконано
8	Оформити дипломну роботу	Червень 2023 р.	Виконано

Студент _____ Вакулюк В.В.

Керівник _____ Яйлимова Г.О.

РЕФЕРАТ

Кваліфікаційна робота містить: 54 стор., 12 рисунків, 1 таблицю, 8 джерел.

Метою дослідження є застосування математичних методів аналізу для прогнозування результатів впливу замінованих та обстріляних територій України на агропромисловий комплекс держави. Об’єктом дослідження є пошкоджені території аграрного сектору України, внаслідок збройної агресії росії.

Розглянуто супутникові знімки з супутника Sentinel-2 Лисичанської об’єднаної територіальної громади, на основі якої побудовано логістичну регресійну модель, яка класифікує території на наявність пошкоджень.

МАТЕМАТИЧНІ МЕТОДИ АНАЛІЗУ, ЛОГІСТИЧНА РЕГРЕСІЙНА МОДЕЛЬ, ПОШКОДЖЕНІ ТЕРИТОРІЇ, АГРАРНИЙ СЕКТОР УКРАЇНИ.

ABSTRACT

Qualification work contains: 54 pages, 12 figures, 1 table, 8 sources.

The purpose of the study is to apply mathematical methods of analysis to predict the results of the impact of mined and shelled territories of Ukraine on the agro-industrial complex of the state. The object of the study is the damaged territories of the agricultural sector of Ukraine as a result of Russia's armed aggression.

Satellite images from the Sentinel-2 satellite of the Lysychansk united territorial community are considered, on the basis of which a logistic regression model is built, which classifies the territories for damage.

MATHEMATICAL METHODS OF ANALYSIS, LOGISTIC REGRESSION MODEL, DAMAGED AREAS, AGRICULTURAL SECTOR OF UKRAINE.

ЗМІСТ

ВСТУП.....7
РОЗДІЛ 1. ВПЛИВ ОБСТРІЛЯНИХ ТА ЗАМІНОВАНИХ ТЕРИТОРІЙ НА АГРАРНИЙ СЕКТОР УКРАЇНИ..... 10
1.1 Роль аграрного сектору в економіці України..... 10
1.2 Основні галузі аграрного сектору..... 11
1.3 Проблеми та виклики, що стоять перед аграрним сектором України, спричинені війною росії проти України..... 12
1.4 Заміновані території України.....14
1.5 Види пошкоджень сільськогосподарських земель..... 16
Висновки до Розділу 1.....18
РОЗДІЛ 2. РЕГРЕСІЙНИЙ АНАЛІЗ..... 19
2.1 Загальні поняття про регресійний аналіз..... 19
2.2 Лінійна регресійна модель.....20
2.3 Поліноміальна регресія.....25
2.4 Логістична регресія..... 27
2.5 Порівняння лінійної та логістичної регресій.....29
2.5 Регресія з використанням дерев рішень..... 30
2.6 Гребінна регресія..... 31
Висновки до Розділу 2.....33
РОЗДІЛ 3. ПОБУДОВА МАТЕМАТИЧНОЇ МОДЕЛІ НА ОСНОВІ СУПУТНИКОВИХ ЗНІМКІВ..... 34
3.1 Збір даних для аналізу.....34
3.2 Навчання моделі логістичної регресії.....37
3.3 Побудова графіків логістичної регресії.....40
3.4 Обчислення функції втрат та оптимізація параметрів..... 42
3.5 Оцінка моделі.....43
Висновки до Розділу 3.....43
ВИСНОВКИ.....45
ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ.....46
ДОДАТОК А ТЕКСТИ ПРОГРАМ..... 48
А.1 Програма 1.....48

ВСТУП

Актуальність дослідження полягає у тому, що повномасштабне вторгнення росії в Україну не лише знищує інфраструктуру сіл та міст, а також має згубні наслідки для екосистем та їх компонентів. Військова діяльність веде до поширеної і тривалої деградації природного середовища. Війна має серйозний вплив на грунти України. Україна, як аграрна країна, має значну частку своєї економіки, пов'язану з сільським господарством. Тому будь-які конфлікти та воєнні дії можуть суттєво позначитися на стані ґрунтів і, відповідно, на сільськогосподарському секторі.

Один з основних впливів війни на ґрунти - це використання важкої військової техніки та артилерії, які негативно впливають на структуру ґрунтів. Вибухи та пошкодження від важкої техніки можуть призвести до стиснення, забруднення та засмічення ґрунтів. Це може спричинити зниження родючості грунтів і негативно позначитися на врожайності. Тому аналіз пошкоджених територій є дуже важливим, щоб у майбутньому можна було оцінити збитки, завдані війною, та розробити план відновлення сільськогосподарських земель і не тільки.

Метою дослідження є застосування математичних методів аналізу для прогнозування результатів впливу замінованих та обстріляних територій України на агропромисловий комплекс держави. Для досягнення мети необхідно розв’язати **задачу дослідження**, яка полягає у побудові регресійної моделі для виявлення пошкоджених територій. Для розв’язання задачі необхідно вирішити такі завдання:

1) дослідити аграрний сектор України і як він впливає на економіку держави.

2) визначити види пошкоджень та чинники, які найбільше вразили території аграрного сектору.

3) розглянути математичні методи, які застосовуються для побудови математичних моделей.

4) збір даних для обчислень.

5) аналіз за обраними показниками та побудова моделей.

6) організація даних, результати розрахунків та їх оцінка.

Об’єктом дослідження є пошкоджені території аграрного сектору України, внаслідок збройної агресії росії.

Предметом дослідження є фактори впливу на території аграрного сектору.

Наукова новизна отриманих результатів полягає в тому, що навчання логістичних регресійних моделей дуже корисне у розрізі дослідження впливу воєнних дій на земельні ділянки аграрного сектору України.

Практичне значення результатів полягає в тому, що навчання логістичної регресійної моделі дозволяє дослідити та класифікувати різні види пошкоджень територій аграрного сектору та порахувати як ці пошкодження у майбутньому вплинуть на врожайність тощо.

РОЗДІЛ 1. ВПЛИВ ОБСТРІЛЯНИХ ТА ЗАМІНОВАНИХ ТЕРИТОРІЙ НА АГРАРНИЙ СЕКТОР УКРАЇНИ

1.1 Роль аграрного сектору в економіці України

Аграрний сектор грає важливу роль в економіці України. Він є одним з ключових секторів господарства країни та має великий потенціал для розвитку. Роль аграрного сектору полягає в кількох аспектах:

- виготовлення харчових продуктів: забезпечує значну частку продуктів харчування для внутрішнього споживання та експорту. Україна вирощує широкий спектр сільськогосподарських культур, таких як зернові, олійні культури, овочі, фрукти, м'ясо, молоко та інші. Харчова безпека країни частково залежить від продуктивності аграрного сектору;
- експорт: аграрний сектор є одним з найбільших експортерів України, країна активно експортує зерно, соняшникову олію, рибу, м'ясо, яйця та інші сільськогосподарські продукти. Експорт сільськогосподарської продукції сприяє покращенню торговельного балансу країни та залученню іноземних валютних коштів;
- зайнятість: аграрний сектор є важливим джерелом зайнятості для населення, особливо в сільських районах. Він забезпечує робочі місця для сільськогосподарських працівників, фермерів, агрономів та інших фахівців. Значна кількість населення отримує дохід від сільськогосподарської діяльності;
- регіональний розвиток: аграрний сектор сприяє розвитку сільських територій і підтримці соціально-економічної активності в сільській місцевості. Він включає в себе не тільки сільське господарство, але й

агропромисловий комплекс, який включає переробну промисловість, логістику та інфраструктуру.

Утримуючи значну роль у внутрішньому ринку та на зовнішніх ринках, аграрний сектор сприяє сталому економічному розвитку України. Продуктивний та конкурентоспроможний аграрний сектор є важливим фактором для підтримки економічної стабільності та процвітання країни.

1.2 Основні галузі аграрного сектору

Основні галузі аграрного сектору України включають рослинництво, тваринництво та рибальство. Кожна з цих галузей має свої особливості і внесок у загальний обсяг сільськогосподарського виробництва.

Рослинництво є однією з найбільших галузей аграрного сектору України. Країна вирощує широкий спектр сільськогосподарських культур, зокрема зернові (пшениця, кукурудза, ячмінь), олійні культури (соняшник, соя), цукрові буряки, овочі, фрукти, ягоди та інші. Рослинництво забезпечує необхідність у харчових продуктах, включаючи зерно, олію, овочі та фрукти, як для внутрішнього споживання, так і для експорту.

Тваринництво є важливою галуззю аграрного сектору України. Вона включає в себе вирощування та утримання худоби (крупної рогатої, малих рогатих, свиней, птиці та інших тварин). Україна виробляє значні обсяги м'яса, молока, яєць та інших тваринних продуктів. Тваринництво забезпечує внутрішнє споживання та експорт м'яса та молочних продуктів.

Україна має значний потенціал у рибальстві завдяки своїм водним ресурсам. Рибальство здійснюється як в прісних водах (річки, озера), так і в морях (Чорне море, Азовське море). Україна виробляє рибні продукти, такі як риба, морепродукти, ракоподібні, які використовуються як для внутрішнього споживання, так і для експорту.

Ці галузі аграрного сектору взаємодіють між собою та сприяють створенню інтегрованої сільськогосподарської системи. Розбудова та розвиток цих галузей є важливим завданням для стабільного розвитку аграрного сектору та економіки України в цілому.

1.3 Проблеми та виклики, що стоять перед аграрним сектором України, спричинені війною росії проти України

Війна Росії проти України суттєво позначилася на аграрному секторі країни і викликала ряд проблем і викликів, таких як:

1) Порушення територіальної цілісності: російська агресія призвела до анексії Криму та втрати контролю над окремими районами на сході та півдні України, де знаходиться значна частина сільськогосподарських земель. Ці землі були важливими для вирощування різноманітних культур і виробництва сільськогосподарської продукції. Втрата земель створює серйозні перешкоди для розвитку аграрного сектору та зменшує загальний обсяг виробництва.

2) Руйнування і збитки: постійні обстріли спричинили знищення та пошкодження сільськогосподарської інфраструктури: багато сільськогосподарських будівель, таких як амбари, сараї, фермерські споруди та інші, піддалися зруйнуванню в результаті бойових дій. Це створює значні втрати для фермерів та підприємств, позбавляючи їх житлових та робочих приміщень, а також необхідної інфраструктури для зберігання та обробки сільськогосподарської продукції. Пошкодження систем поливу та іригації, які є життєво важливими для забезпечення зростання та розвитку рослинництва. Пошкодження цих систем може призвести до погіршення родючості грунту та зменшення врожайності. Зруйнування зерносховищ та складів призводить до великих втрат зерна та інших сільськогосподарських продуктів. Це створює проблеми зі зберіганням та збутом продукції, а також

впливає на її якість та ціну. Пошкодження сільськогосподарської техніки та обладнання, такого як трактори, комбайни, засоби поливу та інші. Це ускладнює проведення сільськогосподарських робіт та збільшує витрати на ремонт або заміну обладнання.

3) Втрати земельного фонду: війна супроводжувалася незаконним захопленням та зміною власності на земельні ділянки, що належать аграрному сектору. Це створило невизначеність щодо власності на землю, порушення прав власності та перешкоди для впровадження сільськогосподарської діяльності, що ускладнює ведення бізнесу і втягує аграрний сектор в правові конфлікти. Відомо, що в Луганській області зембанк та елеватори потужністю 235 тис. т зараз контролює російська компанія «Агрокомплекс» ім. М.І. Ткачова, яка належить ексміністру сільського господарства рф Олександру Ткачову. Ця ж агрофірма вивезла минулого року з елеваторів «Агротону» (найбільший диверсифікований вертикально-інтегрований сільськогосподарський виробник на сході України) 20 тис. т урожаю 2021 року, а також 200 тис. т урожаю 2022 року.

4) Зниження економічного потенціалу: окупація територій веде до економічного виключення цих регіонів з загального економічного простору України. Втрата ринків збуту та обмеження можливостей торгівлі та експорту суттєво впливають на доходи аграрного сектору. Крім того, втрата інвестицій та погіршення бізнес-клімату на окупованих територіях також негативно позначаються на аграрному секторі.

5) Загроза безпеці працівників: обстріли територій України створили загрозу безпеці для аграрних працівників, зокрема фермерів, сільськогосподарських робітників, інших фахівців та сільських жителів. Бойові дії та наявність незаконних збройних формувань ускладнюють нормальне функціонування аграрного сектору та створюють ризики для

життя і майна людей. Ризик поранень, втрати життя та психологічні травми погіршують умови праці та безпеку працівників.

6) Втрата земель та урожаїв: військові дії можуть призвели до втрати сільськогосподарських земель та урожаїв. Пошкодження земельних ділянок через обстріли, мінування або військову окупацію може призвести до неможливості їх використання для вирощування культур, зменшити врожайність та вже призвели до збитків для сільськогосподарських підприємств. Наприклад, земельний банк «НІБУЛОН» (один із лідерів українського аграрного ринку) зменшився з 76,5 тис. га до 51 тис. га. Більша частина земель, на яких компанія наразі не може вести господарську діяльність, знаходиться в Луганській області — приблизно 20 тис. га. Близько 5,5 тис. га потребують розмінування у Снігурівському районі Миколаївської області.

7) Переселення населення: війна призвела до переселення місцевого населення, зокрема сільськогосподарських працівників та фермерів. Це призводить до зупинки або обмеження сільськогосподарської діяльності, зменшення працездатного населення та втрати людських ресурсів, необхідних для ведення сільського господарства.

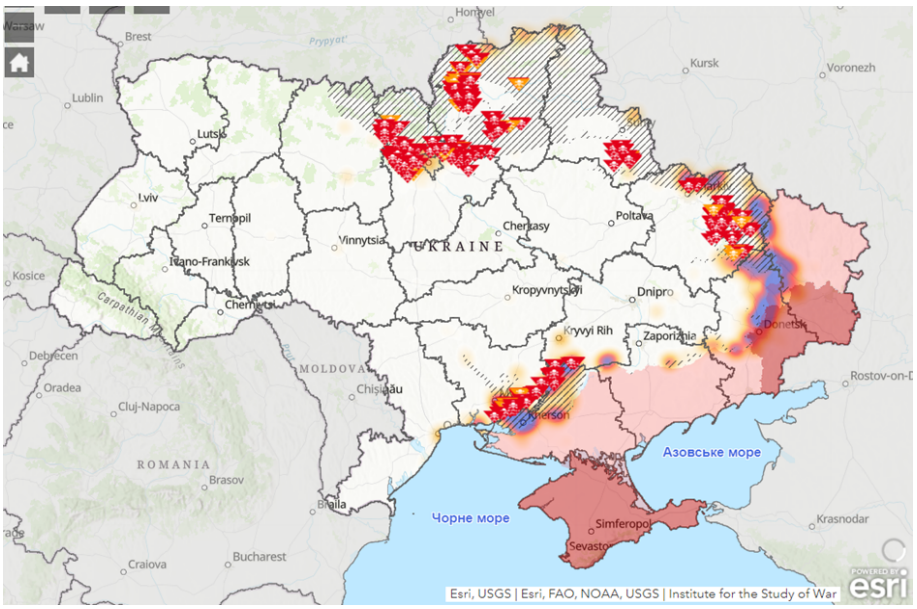
1.4 Заміновані території України

Згідно з даними ДСНС України частина замінованих та забруднених вибухонебезпечними предметами територій держави з 2014 року складає близько 30%. Більше ніж 200 тисяч гектарів територій засмічені снарядами, мінами та уламками боєприпасів. Це – сотні тисяч протипіхотних та протитанкових мін, авіаційних бомб, мінометних снарядів та боєприпасів, які не розірвалися і застрягли в полях.

За різними оцінками, не спрацьовують до 20% випущених боєприпасів. Крім того, якщо росіяни тривалий час перебувають на певній території, то ставлять міни в лісах і на полях.

Регіони, на яких ведуться (велися) бойові дії: Київщина, Чернігівщина Сумщина, Харківщина, Дніпропетровщина, Луганщина, Донеччина, Запоріжчина, Миколаївщина, Херсонщина. Території, такі як Херсонщина, Миколаївщина і Запоріжжя, є важливими регіонами для української сільськогосподарської продукції. Вони служать основними джерелами зерна, овочів і фруктів в Україні. У червні саме починається збір урожаю в цих регіонах, але вода Каховського водосховища просто знищила непереробленний та незібраний урожай. Загальна площа, яка може бути потенційно вкрита водою разом з руслом і гирлом Дніпра сягає 1126 км^2. У 2023 році на правому березі Дніпра працюють лише 13 зрошувальних систем. Підрив Каховської ГЕС фактично залишив без джерела води 94% зрошувальних систем в Херсонській, 74% — в Запорізькій та 30% — в Дніпропетровській областях.

Далі представлена мапа (Рис.1) територій, які постраждали внаслідок російської агресії.



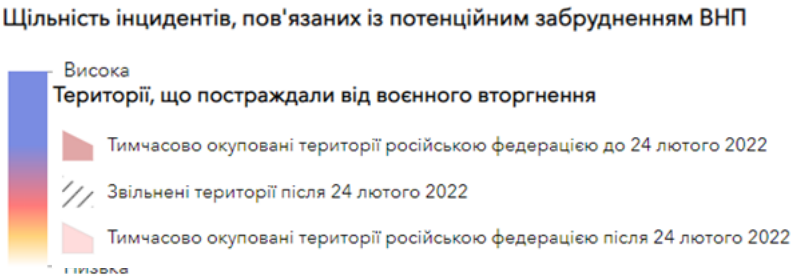


Рисунок 1– Мапа забруднених територій України.

1.5 Види пошкоджень сільськогосподарських земель

Сільськогосподарські землі уражені двома значними типами пошкоджень - мінним забрудненням і прямим фізичним пошкодженням. У зонах, де тривали або продовжують тривати бойові дії та на окупованих територіях, сільськогосподарські угіддя мають високий ризик мінного забруднення. Крім мінних полів, території, що постраждали від активних бойових дій, також заражені нерозірваними боєприпасами, створюючи смертельну загрозу для українських фермерів під час сільськогосподарських робіт. Всі сільськогосподарські угіддя у зоні активних бойових дій або окуповані території потребують ретельного обстеження, а деякі з них потребують розмінування, щоб зробити землю знову придатною для землеробства.

Другим типом пошкоджень є фізичне пошкодження родючого шару грунту. Військові конфлікти впливають на ґрунтовий покрив шляхом спричинення механічних, фізичних та хімічних змін. Ці зміни призводять до руйнування структури та функцій ґрунтової екосистеми, що призводить до погіршення фізико-геохімічних характеристик. Наслідки військових дій на ґрунтовий покрив можуть варіюватись залежно від типу військових об'єктів і характеру бойових дій, таких як пошкодження рельєфу поверхні та ґрунтів (унаслідок вибухів, руху військової техніки), пошкодження захисних споруд (таких як окопи, бліндажі, вогневі точки, протитанкові споруди) та

забруднення верхніх шарів ґрунту продуктами військової діяльності, а також залишками військової техніки, захисних споруд, осколками та іншими матеріалами на поверхні. Внаслідок цього виникає знищення рослинності, порушення ґрунтового покриву, зменшення природного зволоження та опустелювання, що є поширеними наслідками воєнно-техногенного впливу. Такі пошкодження в районах активних військових дій вимагатимуть подальшого відновлення земель, включаючи рекультивацію та вирівнювання поверхні.

Південні регіони, які зараз страждають від окупації РФ та жорстоких бойових дій, також мають найбільш розвинену іригаційну інфраструктуру. На Херсонщині налічується 426,8 тис. га зрошуваних земель, або 21,6 % від загальної площі сільгоспугідь, у тому числі від державних зрошувальних систем — 384,5 тис. га, місцевого зрошення — 42,3 тис. га. ще 10% - на території частково окупованої Запорізької області. Також екологічні наслідки після підриву Каховської ГЕС російськими бойовиками у довгостроковій перспективі будуть страшні. Система водосховища і його каналів - це зрошування територій сільського господарства. Загалом зрошення буде неможливим в Херсонській області протягом 2-3 років [5]. Може впасти продуктивність грунтів та врожайність. Ці території стануть більш посушливими , бо не буде системи меліорації. Через зміни грунтів, ерозію берегів і зміну ландшафту ситуація з біорізноманіттям буде ще складнішою. Багато забруднювальних речовин, таких як: нафтопродукти, вибухівка, - потрапило у води Дніпра та Чорного моря, через руйнування споруд.

Внаслідок вибухів ракет, артилерійських снарядів різних типів, фугасних авіабомб, безпілотників, снарядів різних типів РСЗО, «вакуумних» бомб та інших подібних обставин, стверджується, що верхній родючий шар ґрунту, що нагромаджувався протягом багатьох століть, втрачається. Варто

враховувати, що за останні сто років втрати гумусу вітчизняних ґрунтів складають близько 30%. Цей процес ще більше прискорюється через війну, оскільки змінюються фізичні, хімічні та фізико-хімічні властивості, що впливають на родючість ґрунту.

Крім прямих пошкоджень земель - окупації, військових дій та мінного забруднення - фермерам ускладнюється доступ до полів і можливостей збору врожаю. Це призводить до того, що багато районів, постраждалих від війни, не можуть зібрати озимі культури. Орієнтовна площа 2,4 мільйона гектарів озимих культур загальною вартістю 1,435 мільярдів доларів США залишилася не зібраною у 2022 році через агресію РФ.

Висновки до Розділу 1

Війна росії проти України характеризується використанням повного спектру зброї, військової техніки та боєприпасів. Усі види воєнно-техногенного навантаження спричиняють значне забруднення та руйнування сільськогосподарських територій, а саме - грунтів. Окупація, заміновані та обстріляні території України негативно впливають на аграрний сектор держави, створюючи значні перешкоди для розвитку і ефективного функціонування агропромислової галузі. Військові дії та обстріли на аграрній території України суттєво ускладнюють ситуацію в аграрному секторі, призводять до значних втрат, зниження продуктивності землі та загрози безпеці.

РОЗДІЛ 2. РЕГРЕСІЙНИЙ АНАЛІЗ

У сучасному світі наукові підходи, які базуються на новітніх методиках та обчисленнях, є ключовими рушіями прогресу. Багато прикладних задач вимагають дослідження факторних причинно-наслідкових зв'язків, що передбачає встановлення функціонального вигляду та математичного опису цих зв'язків. Один з підходів до таких досліджень полягає у застосуванні регресійного аналізу.

2.1 Загальні поняття про регресійний аналіз

Регресійний аналіз — це статистичний метод для дослідження та моделювання зв'язку між змінними. Застосування регресії численні і відбуваються майже в усіх сферах, в т.ч. інженерія, фізико-хімічні науки, економіка, управління, життя та біологічні науки, а також соціальні науки. У програмуванні регресійний аналіз використовується для обробки даних, оцінки прогнозів на основі функцій однієї або декількох змінних, а також для опису системних процесів. Використання регресійного аналізу обумовлене потребою отримати характеристики змін від випадкових значень, процесів або подій.

Регресійний аналіз є одним з найскладніших та найвідповідальніших аспектів статистичного моделювання у сучасних дослідженнях суспільно-економічних явищ. Його основна мета полягає у побудові математичної лінії регресії, яка наближається з високою точністю до фактичних значень вивченої змінної на основі наявних даних. Більш того, регресійний аналіз дозволяє перенести результати дослідження на загальну

популяцію та прогнозувати дану змінну на основі очікуваних (планових) значень факторів.

Регресійний аналіз використовує відповідні критерії для визначення окремого та загального впливу факторів на залежну змінну, а також для кількісної оцінки цього впливу. Цей метод ґрунтується на побудованому рівнянні регресії та визначає, який внесок кожна незалежна змінна вносить у варіацію залежної змінної, яку ми досліджуємо або прогнозуємо.

Розробка математичної моделі та її моделювання є основою прикладного регресійного аналізу. Математичне моделювання включає процес створення математичної моделі для реального об'єкта, яка може бути представлена системою рівнянь або комп'ютерною програмою. Кожна математична модель наближено описує реальний об'єкт. Дослідження моделі дозволяє встановити характеристики реального об'єкта. Математичне моделювання може бути аналітичним, імітаційним або комбінованим.

Аналітичне моделювання передбачає запис системи математичних рівнянь, які описують її функціонування, у вигляді аналітичних співвідношень (алгебраїчних, диференціальних, інтегральних тощо) та додаткових умов (початкових, крайових, обмежень). Імітаційне моделювання, як правило, використовує комп'ютерну програму, яка відтворює поведінку системи в часі, імітуючи елементарні явища збереженням їх логічної структури та послідовності. Комбіноване (аналітико-імітаційне) моделювання поєднує переваги аналітичного та імітаційного моделювання, де аналітичні моделі використовуються там, де це можливо, і для інших підпроцесів будуються імітаційні моделі.

2.2 Лінійна регресійна модель

Лінійна регресія – це вид регресійної моделі, яка моделює лінійну залежність між кількома пояснювальними змінними ($X_1, X_2, ..., X$□) і однією

залежною змінною (*Y)*. Модель припускає, що ця залежність може бути описана лінійною комбінацією цих змінних. Формула для множинної лінійної регресії має вигляд:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n + \varepsilon, \quad (1)$$

Кожному регресорові в лінійній моделі надається числова вага – β біля кожного *x* у рівнянні 1, яка називається коефіцієнтом регресії, нахилом регресії, або просто вагою регресії, яка визначає, наскільки рівняння використовує значення цієї змінної для отримання оцінки *Y*. Ці ваги регресії виводяться за допомогою алгоритму, який створює математичне рівняння або модель для *Y*, що найкраще відповідає даним, використовуючи певний критерій для визначення "найкращого".

Метою лінійної регресії є знаходження оптимальних значень для коефіцієнтів $\beta_0, \beta_1, \beta_2, ..., \beta_n$, які найкраще апроксимують дані. Це може бути зроблено також за допомогою методу найменших квадратів, де ми мінімізуємо суму квадратів розбіжностей між спостережуваними значеннями Y і прогнозованими значеннями $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n$.

У методі МНК ми повинні вибрати значення β_p так, щоб загальна сума квадратів різниці між обчисленим і спостережуваним значеннями у мінімізується.

Формула для МНК:

$$S = \Sigma_{i=1}^{n} (y_i - \bar{y}_i)^2 = \Sigma_{i=1}^{n} (y_i - \beta_1 x_1 - \beta_0)^2 = \Sigma_{i=1}^{n} (\hat{\varepsilon_i})^2 = min, \quad (2)$$

де:

- $\bar{y}_i$ – прогнозоване значення для і-го спостереження
- y_i – фактичне значення для і-го спостереження

- ε_i – помилка/залишок для і-го спостереження

- n – загальна кількість спостережень

Щоб отримати значення β_p, які мінімізують *S*, ми можемо взяти часткову похідну для кожного коефіцієнта та прирівняти її до нуля.

У результаті, модель лінійної регресії дає нам гіперплощину або гіперплощини в просторі, які найкраще підходять до набору даних. Ми можемо використовувати цю модель для прогнозування значень *Y* на основі вхідних значень $X_1, X_2, ..., X$□.

Важливою особливістю лінійної регресії є можливість моделювання багатьох пояснювальних змінних одночасно. Це дозволяє враховувати взаємодію та вплив кожної змінної на залежну змінну. Крім того, за допомогою множинної лінійної регресії можна включати категоріальні змінні, якщо вони кодуються у вигляді факторних змінних.

Лінійна регресія є одним з базових і широко використовуваних методів в статистиці та машинному навчанні, оскільки вона проста, інтерпретована та ефективна для багатьох задач прогнозування.

Базовий метод лінійної моделі накладає шість вимог:

1) Як і в будь-якому статистичному аналізі, має бути набір "учасників", "випадків" або "одиниць".

2) Кожен з цих учасників повинен мати значення або виміри для двох або більше змінних, кожна з яких є числовою, дихотомічною або багатокатегорійною. Таким чином, вихідні дані для аналізу формують прямокутну матрицю даних з учасниками в рядках і змінними в стовпчиках.

3) Кожна змінна повинна бути представлена одним стовпчиком чисел. Наприклад, дихотомія статі може бути представлена таким чином

число 1 позначає чоловічу стать, а 0 – жіночу, таким чином, стать 100 осіб можна представити стовпчиком зі 100 чисел, кожне з яких дорівнює 0 або 1. Багато категорійна змінна з, скажімо, п'ятьма категоріями може бути представлена стовпчиком чисел, кожне з яких дорівнює 1, 2, 3, 4 або 5. Як для дихотомічних, так і для багато категоріальних змінних, числа що представляють категорії, є просто кодами і є довільними. Вони не несуть ніякого значення щодо кількості і можуть бути замінені будь-яким іншим набором чисел, не змінюючи результатів аналізу, якщо використовуються використовуються правильні методи кодування. І, звичайно, числову змінну, наприклад, вік, можна представити у вигляді стовпчика вікових категорій.

4) Кожен аналіз повинен мати лише одну залежну змінну, хоча він може мати кілька незалежних змінних і кілька коваріант.

5) Залежна змінна повинна бути числовою. Числова змінна – це щось на кшталт віку або доходу з інтегральними властивостями, так що значення можуть бути осмислено усереднені.

6) Статистичні висновки з лінійних моделей часто вимагають кількох додаткових припущень.

У цих умовах лінійні моделі є гнучкими у багатьох відношеннях:

1) Змінна може бути природною властивістю учасника, наприклад, вік або стать, або може бути властивістю, якою маніпулюють в експерименті, наприклад, яка з двох або більше експериментальних умов, в які потрапляє учасник потрапляє через процедуру випадкового розподілу. Змінні, якими маніпулюють, зазвичай є категоріальними, але можуть бути і числовими, наприклад, кількість годин практики для виконання завдання, яку отримують учасники або кількість актів насильства на телебаченні, які людина бачить під час експерименту.

2) Можна провести серію аналізів на основі однієї і тієї ж прямокутної матриці даних, і та сама змінна може бути залежною змінною в

одному аналізі і незалежною або коваріантою в іншому. Наприклад, якщо матриця містить змінні вік, стать, роки освіти та заробітна плата, то в одному аналізі можна дослідити роки освіти як функцію віку та статі, а в іншому – заробітну плату як функцію віку, статі та освіти.

3) Різниця між незалежними змінними та коваріатами може бути нечіткою, оскільки програми лінійного моделювання не бачать різниці між ними. Програма обчислює міру зв'язку між залежною змінною та кожною іншою змінною в аналізі, одночасно здійснюючи статистичний контроль за всіма іншими змінними, включаючи коваріанти та інші незалежні змінні. Незалежні змінні - це ті, зв'язок яких із залежною змінною ви хочете обговорити або які є предметом вашого дослідження, тоді як коваріати - це інші змінні, які ви хочете контролювати або включити в модель з якоюсь іншою метою. Таким чином, різниця між цими двома поняттями визначає, як ви будете описувати результати аналізу, але не використовується при написанні комп'ютерних команд, які визначають аналіз або математику, що лежить в його основі.

4) Кожна незалежна змінна або коваріація може бути дихотомічною, багатокатегоріальною або числовою. Всі три типи змінних можуть зустрічатися в одній і тій самій проблемі. Наприклад, якщо ми вивчаємо заробітну плату в професійній фірмі як функцію статі, етнічної приналежності та віку, контролюючи при цьому стаж роботи, громадянство та тип вищої освіти (бізнес, мистецтво, інженерія тощо), ми б мали одну незалежну змінну та по одній коваріантній змінній для кожного з трьох типів шкал.

5) Незалежні змінні та коваріанти можуть бути взаємопов'язаними, як це, швидше за все, буде в усіх наведених прикладах. Насправді, необхідність контролювати контролювати коваріату зазвичай виникає через те, що вона корелює з однією або кількома незалежними змінними, або залежною змінною, або з обома.

6) На додаток до кореляції між собою, незалежні змінні і коваріанти можуть взаємодіяти, впливаючи на залежну змінну.

7) Незважаючи на назви "лінійна регресія" та "лінійна модель", ці методи можуть бути легко поширені на велику кількість проблем, пов'язаних з криволінійними зв'язками між змінними. Наприклад, фізична сила криволінійно залежить від віку, досягаючи піку у 20 років. Але лінійна модель може бути використана для вивчення зв'язку між віком і силою або навіть для оцінки віку, в якому сила досягає піку.

8) Припущення, необхідні для статистичного висновку, не є надзвичайно обмежуючими. Існує кілька способів обійти обмеження, що накладаються цими припущеннями.

Логістична регресія, пробіт-регресія та багаторівневе моделювання є близькими родичами лінійного регресійного аналізу. У логістичній та пробіт-регресії залежна змінна може бути дихотомічною або порядковою, наприклад, чи досягає людина успіху або провалу у виконанні завдання, діє або не діє певним чином у певній ситуації, не любить, відчуває себе нейтрально або любить стимул. Багаторівневе моделювання використовується, коли дані мають "вкладену" структуру, наприклад, коли різні підгрупи учасників дослідження мають щось спільне, наприклад, район або житловий комплекс, в якому вони живуть, або будівлю в місті, в якому вони працюють.

2.3 Поліноміальна регресія

Поліноміальна регресія - це вид регресійної моделі, в якій використовуються поліноми в якості функцій відгуку для моделювання залежності між пояснювальними змінними (X) і залежною змінною (Y). Вона дозволяє моделі бути більш гнучкою при апроксимації складних форм залежностей.

Формула для поліноміальної регресії має наступний вигляд:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \ldots + \beta_n X^n + \varepsilon, \quad (3)$$

де:

- Y - залежна змінна, яку ми хочемо прогнозувати,

- X - пояснювальна змінна,

- $\beta_0, \beta_1, \beta_2, \ldots, \beta_n$ - коефіцієнти регресії, які визначають форму полінома,

- ε - помилка, яка представляє випадкову нев'язку або шум, що не може бути пояснений моделлю,

- n - ступінь полінома, який обирається відповідно до складності залежності.

У поліноміальній регресії коефіцієнти $\beta_0, \beta_1, \beta_2, \ldots, \beta_n$ визначають форму полінома і впливають на кривизну і форму апроксимації. Зазвичай для побудови поліноміальної регресійної моделі використовуються методи, такі як метод найменших квадратів або метод максимальної правдоподібності, для знаходження оптимальних значень коефіцієнтів.

Важливою особливістю поліноміальної регресії є те, що вона дозволяє моделювати нелінійні залежності між змінними. За допомогою поліноміальної регресії можна виразити криві, які апроксимують даний набір даних краще, ніж проста лінійна регресія.

Для вибору оптимальної ступеня полінома (n) важливо враховувати баланс між складністю моделі і перенасиченістю. Модель з надмірно високим ступенем полінома може призвести до перенасичення, що призводить до поганої узагальнюючої здатності на нових даних.

Поліноміальна регресія є хорошим інструментом для моделювання нелінійних залежностей та дослідження взаємодій між змінними. Вона знаходить своє застосування в різних галузях, включаючи фізику, економіку, біологію та інші науки.

2.4 Логістична регресія

Класифікаційні методи є важливою частиною прикладних програм машинного навчання та обробки даних. Приблизно 70% проблем у галузі науки даних – це проблеми класифікації. Існує багато проблем класифікації, але логістична регресія є загальною та може бути корисним методом регресії для вирішення проблеми бінарної класифікації.

Логістична регресія є одним із найпростіших і найчастіше використовуваних алгоритмів глибокого машинного навчання. Вона особливо корисна, коли залежна змінна є бінарною (два можливих значення) або категоріальною з відомими ймовірностями. Логістична регресія надає нам інтерпретовані коефіцієнти та дозволяє моделювати нелінійні залежності між змінними. Логістична регресія надає нам інтерпретовані коефіцієнти та дозволяє моделювати нелінійні залежності між змінними.

Логістична регресія може стати статистичною процедурою для прогнозування бінарних класів. Результат чи цільова змінна має дихотомічну природу. Дихотомічний означає, що існує лише два можливі класи. Наприклад, він часто використовується для проблем із виявленням раку. Він обчислює можливість виникнення випадку.

Це особливий випадок прямолінійної регресії, коли цільова змінна має категоріальний характер. Вона використовує журнал ймовірностей, тому що змінна .Logistic Regression у Python передбачає можливість виникнення бінарної події, використовуючи функцію logit.

Формула для логістичної регресії використовує логістичну функцію (сигмоїду) для перетворення лінійної комбінації пояснювальних змінних (X) у ймовірність (P) виникнення події:

$$P = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n)}}, \quad (4)$$

де:

- P - ймовірність виникнення події,

- $X_1, X_2, \ldots, X_n$ - пояснювальні змінні,

- $\beta_0, \beta_1, \beta_2, \ldots, \beta_n$ - коефіцієнти регресії, які визначають вплив кожної змінної на ймовірність,

- e - експонента (приблизне значення 2.71828).

Логістична функція приймає значення між 0 і 1, що відповідає ймовірності. Значення, близькі до 0, вказують на низьку ймовірність, тоді як значення, близькі до 1, вказують на високу ймовірність.

Оптимальні значення коефіцієнтів $\beta_0, \beta_1, \beta_2, \ldots, \beta_n$ визначаються за допомогою методів оптимізації, таких як метод максимальної правдоподібності або метод градієнтного спуску. Метою є максимізувати ймовірність спостережень, що спостерігаються у навчальній вибірці.

Переваги та недоліки логістичної регресії:

- переваги: завдяки своїй ефективності та простоті, не вимагає високої обчислювальної потужності, легко реалізується, легко інтерпретується, широко використовується аналітиками даних та вченими. Крім того, не потребує масштабування функцій. Логістична регресія дає оцінку ймовірності спостережень;

- недоліки: логістична регресія не готова обробляти надмірну кількість категоричних ознак/змінних. вона схильна до перепідбору. Крім того, не може вирішити нелінійну проблему з логістичною регресією, тому вона потребує зміни нелінійних ознак. Логістична регресія не буде добре працювати з незалежними змінними, які не корелюються з цільовою змінною і дуже схожі чи корелюються один з одним.

Логістична регресія потужний інструмент для моделювання і прогнозування ймовірності виникнення події на основі даних з використанням пояснювальних змінних. Вона широко використовується у багатьох галузях, включаючи медицину, епідеміологію, маркетинг, фінанси та інші, де цікаво прогнозувати ймовірність виникнення подій або класифікувати спостереження на основі ймовірності.

2.5 Порівняння лінійної та логістичної регресій

Лінійна регрессія Vs. Логістична регресія

Лінійна регресія дає нескінченний вивід, а логістична регресія – безперервний вивід. Прикладом безперервного видобутку є ціна будинку та вартість акцій. Приклад дискретного виходу – передбачення того, чи є у пацієнта рак чи ні. Прямолінійна регресія оцінюється методом найменших квадратів (OLS), тоді як логістична регресія оцінюється з допомогою підходу Maximum Likelihood Estimation (MLE).

Максимальна оцінка ймовірності Vs. Метод найменших квадратів

MLE може бути методом максимізації “ймовірності”, тоді як OLS може бути методом апроксимації з мінімізацією відстані. Функція максимізації ймовірності визначає параметри, які, ймовірно, забезпечують дані, що спостерігаються. З статистичної точки зору, MLE задає середнє значення та дисперсію як параметри при визначенні точних параметричних

значень для даної моделі. Цей набір часто використовується для прогнозування інформації, необхідної при гаусійвському розподілі .

Звичайний метод оцінки найменших квадратів обчислюється шляхом підгонки регресійної кривої за заданими точками даних, яка має мінімальну суму квадратних відхилень (похибка найменшого квадрата). Обидва методи не дозволяють оцінити параметри моделі прямолінійної регресії. Модель MLE передбачає функцію маси ймовірності, тоді як OLS не вимагає стохастичних припущень мінімізації відстані.

2.5 Регресія з використанням дерев рішень

Регресія з використанням дерев рішень - це метод прогнозування, який використовує деревоподібну структуру для моделювання відношення між пояснювальними змінними (X) і залежною змінною (Y). Вона відрізняється від класичної лінійної регресії тим, що використовує деревоподібну структуру замість лінійної функції для прогнозування.

Формули в регресії з використанням дерев рішень виражаються через розбиття даних на групи та середні значення залежної змінної для кожної групи. Дерево рішень будується шляхом рекурсивного поділу даних на підгрупи залежно від значень пояснювальних змінних і обчислення середнього значення залежної змінної для кожної підгрупи.

В загальному вигляді, формула для регресії з використанням дерев рішень може бути подана наступним чином:

$$Y = f(X), \qquad (5)$$

де:

- Y - залежна змінна, яку ми хочемо прогнозувати,
- X - пояснювальні змінні,

- f - функція, яка відображає відношення між X і Y.

Для кожної гілки дерева рішень, формула може мати такий вигляд:

$$Y = c, \quad (6)$$

де:

- Y - залежна змінна,

- c - константа або середнє значення залежної змінної для даної підгрупи даних.

Дерево рішень розбиває вхідні дані на підгрупи, використовуючи правила розбиття, які базуються на значеннях пояснювальних змінних. Кожна підгрупа представляє вузол дерева, а подальше розбиття продовжується досягнення певного критерію зупинки, такого як максимальна глибина дерева або мінімальна кількість спостережень у вузлі.

У регресії з використанням дерев рішень, модель навчається шляхом знаходження оптимальних правил розбиття та обчислення середніх значень залежної змінної для кожної підгрупи. Це може виконуватися за допомогою різних алгоритмів, таких як CART (Classification and Regression Trees) або Random Forests.

Регресія з використанням дерев рішень є потужним інструментом для прогнозування залежної змінної, особливо в тих випадках, коли існує нелінійна залежність між пояснювальними змінними та залежною змінною. Вона дозволяє отримувати інтерпретовані результати та розуміння важливості пояснювальних змінних у моделі.

2.6 Гребінна регресія

Гребінна регресія, також відома як регуляризована лінійна регресія або *L2*-регуляризація, є методом регресійного аналізу, який використовується для зменшення перенасиченості моделі та врахування мультиколінеарності (коли пояснювальні змінні корелюють між собою). Гребінна регресія додає штрафний член до функції втрат, який залежить від квадратів коефіцієнтів регресії, змушуючи їх до менших значень.

Формула для гребінної регресії включає штрафний член, який контролює розмір коефіцієнтів регресії:

$$minimize: RSS + \lambda\sum\beta^2, \quad (7)$$

де:

- *RSS* - сума квадратів відхилень між спостережуваними значеннями залежної змінної і прогнозованими значеннями,
- λ - параметр регуляризації (також відомий як гіперпараметр), який контролює міру регуляризації. Більші значення λ призводять до більшої регуляризації і зменшують значення коефіцієнтів регресії,
- $\sum\beta^2$ - сума квадратів коефіцієнтів регресії, яка включається в штрафний член. Це забезпечує зменшення значень коефіцієнтів регресії, що сприяє уникненню перенасиченості та контролює мультиколінеарність.

Головна ідея гребінної регресії полягає в тому, що штрафний член змушує модель шукати компроміс між точністю прогнозу та складністю моделі. Чим більше значення параметра регуляризації λ, тим сильніше зменшуються значення коефіцієнтів регресії, що призводить до більш спрощених моделей з меншою чутливістю до малих змін у пояснювальних змінних.

Для розв'язання задачі гребінної регресії існують різні методи, такі як метод найменших квадратів з гребінною регуляризацією (Ridge Regression) і метод стохастичного градієнтного спуску з гребінною регуляризацією (Stochastic Gradient Descent with Ridge Regression). Обидва методи використовуються для знаходження оптимальних значень коефіцієнтів регресії, що мінімізують суму квадратів відхилень та штрафного члена.

Гребінна регресія є ефективним методом для зменшення перенасиченості та контролю мультиколінеарності в моделях регресії. Вона дозволяє покращити прогнозні здібності моделі та забезпечує більш стійкі та зрозумілі результати.

Висновки до Розділу 2

Вибір правильної регресійної моделі важливий для досягнення точних та надійних прогнозів на основі ваших даних. Найкраще підходити до задачі, ретельно аналізувати характеристики даних та використовувати відповідні методи для досягнення оптимальних результатів.

Лінійна регресія може бути корисною для простих задач, де є лінійна залежність між змінними. Поліноміальна регресія може бути використана для моделювання складних нелінійних залежностей в даних. Регресія з регуляризацією може бути ефективною, коли потрібно уникнути перенавчання та покращити узагальнювальні здібності моделі. Логістична регресія є корисною для бінарної класифікації та прогнозування ймовірності належності до певного класу.

РОЗДІЛ 3. ПОБУДОВА МАТЕМАТИЧНОЇ МОДЕЛІ НА ОСНОВІ СУПУТНИКОВИХ ЗНІМКІВ

Сьогодні наявність численних геоінформаційних систем (ГІС) та додатків дозволяє легко отримувати супутникові знімки в режимі онлайн. Ці знімки можна використовувати для багатьох цілей, таких як моніторинг стану посівів, оцінка вигорілих площ, відстеження ураганів, забезпечення безпеки та вирішення інших актуальних завдань.

Супутникові знімки знаходять широке практичне застосування в різноманітних індустріях. Метод, який лежить в основі супутникових карт, дозволяє розпізнавати типи земного покриву, виявляти зміни та оцінювати стан посівів це — технологія дистанційного зондування.

Супутники використовують дистанційні датчики для отримання інформації про різні характеристики поверхні Землі, включаючи рослинний покрив, будівлі, водні поверхні, температуру повітря, висоти й ухили, і багато інших характеристик.

3.1 Збір даних для аналізу

Супутники Sentinel-2 — космічна місія дистанційного зондування Землі, запущена Європейським космічним агентством (ESA) у рамках програми «Copernicus» для здійснення дистанційного спостереження і підтримки таких сервісів, як моніторинг лісів, фіксування змін покриву Землі, відстеженням наслідків стихійних лих. Sentinel-2 має 12 спектральних каналів з просторовою вирізняльною здатністю 10 м, 20 м та 60 м залежно від каналу.

Для дослідження було завантажено серію мозаїк супутникових знімків для тестової території дослідження Лисичанської об’єднаної територіальної громади. Проведено обробку завантажених мозаїк та підготовлено три зображення супутникових знімків.

Геометрична та спектральна корекція. Сформовані знімки модельної території дослідження охоплюють два періоди: ведення активних бойових дій (21 – 28 червня, 1 – 3 липня 2022 року), коли природне середовище зазнало пошкоджень внаслідок воєнних дій, та довоєнний період ведення господарської діяльності (9 вересня 2021 року).

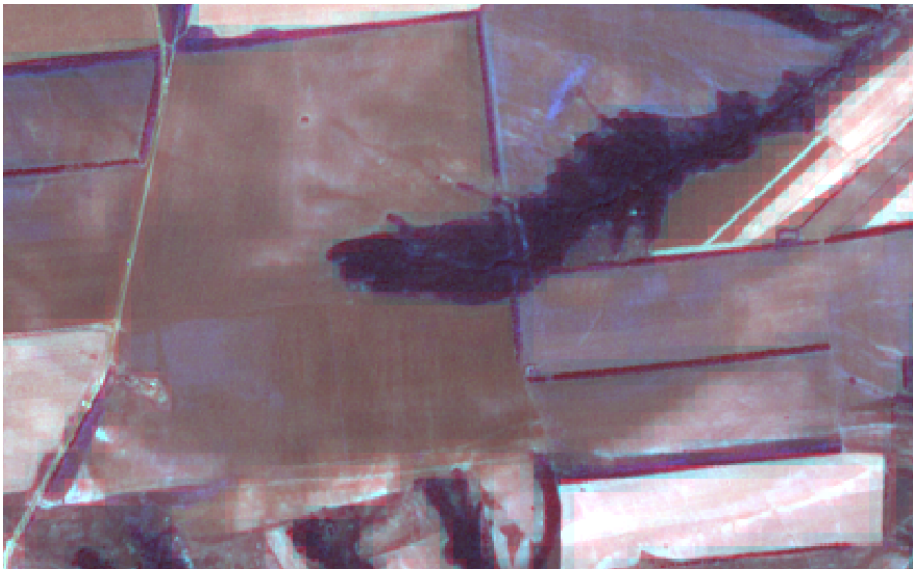


Рисунок 3.1 – Досліджувана територія на 10 вересня 2021 р.

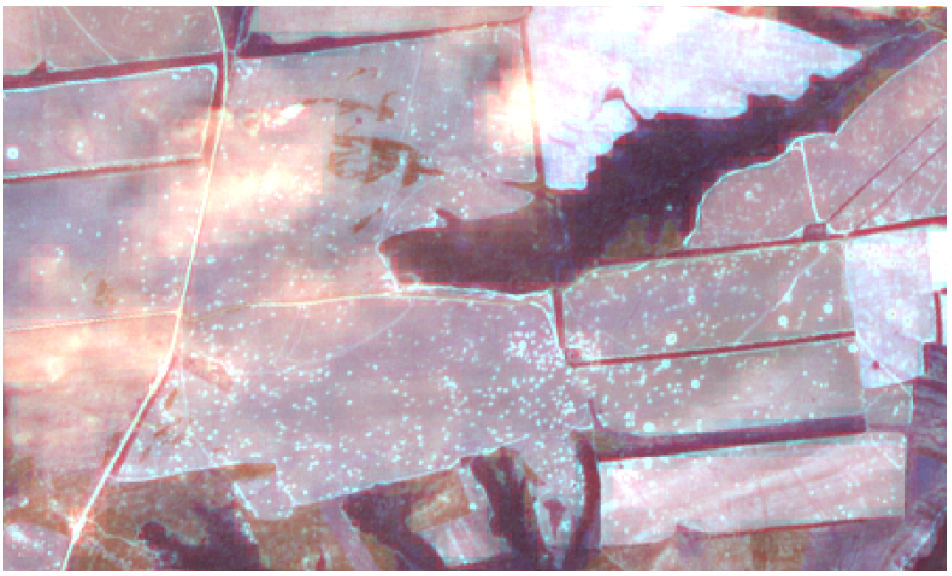


Рисунок 3.2 – Досліджувана територія на 21 – 28 червня 2022 р.

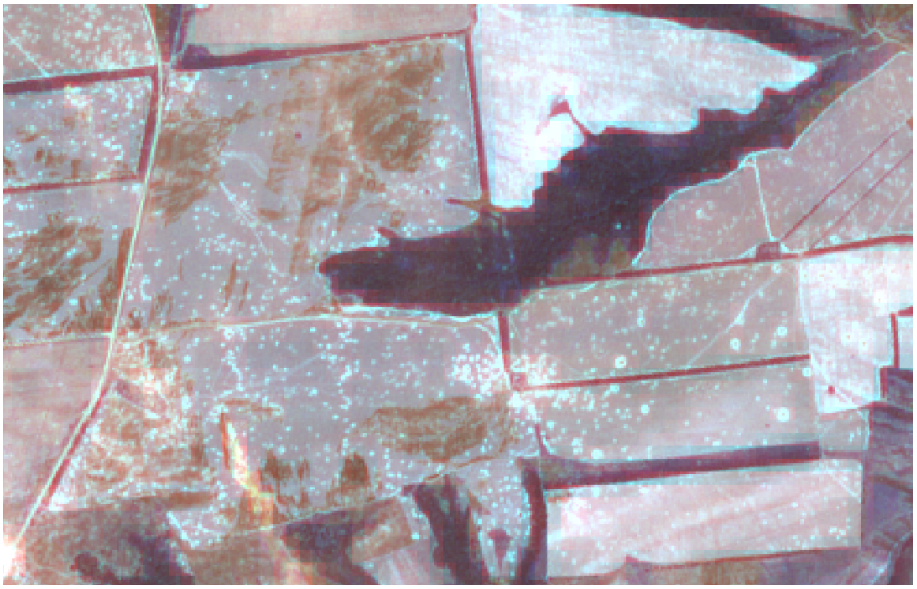


Рисунок 3.3 – Досліджувана територія на 1 – 3 липня 2022 р.

3.2 Навчання моделі логістичної регресії

Модель логістичної регресії навчається шляхом оптимізації параметрів за допомогою методу максимальної правдоподібності. Основна ідея полягає в тому, щоб підігнати модель таким чином, щоб вона максимально відповідала спостережуваним даним.

Основні кроки навчання моделі логістичної регресії:

- ініціалізація параметрів: початкові значення параметрів моделі (ваги) ініціалізуються випадковим чином або нулями.

- обчислення вагованої суми: для кожного прикладу даних обчислюється вагова сума, що представляє лінійну комбінацію вхідних ознак і вагових коефіцієнтів моделі.

- примінення функції активації: вагова сума передається через функцію активації, яка в даному випадку є сигмоїдною функцією. Функція активації перетворює ваговану суму в ймовірність належності до позитивного класу.

- обчислення функції втрат: використовуючи отримані ймовірності, обчислюється функція втрат, яка відображає розбіжність між передбаченими і спостережуваними значеннями.

- оптимізація параметрів: застосовується алгоритм оптимізації (наприклад, градієнтний спуск) для зміни параметрів моделі таким чином, щоб мінімізувати функцію втрат.

- оцінка моделі: Після навчання модель оцінюється за допомогою метрик оцінки, таких як точність(accuracy), чутливість (precision), повнота (recall) тощо.

Процес навчання моделі логістичної регресії базується на оптимізації параметрів моделі таким чином, щоб досягти максимальної відповідності між передбаченими значеннями і спостережуваними мітками класів.

Для навчання було використано знімки, на яких позначено пошкоджені (рожевий колір) та не пошкоджені (синій колір) точки (рис. 3.4, рис. 3.5), для тестування - інші знімки (рис. 3.6, рис. 3.7). Для проведення розрахунків пошкоджені території було визначено одиницями, а непошкоджені - нулями, тобто отримали два класи 0 та 1. На основі отриманих даних було створено два датасети, де один був навчальним, а інший - тестувальним.

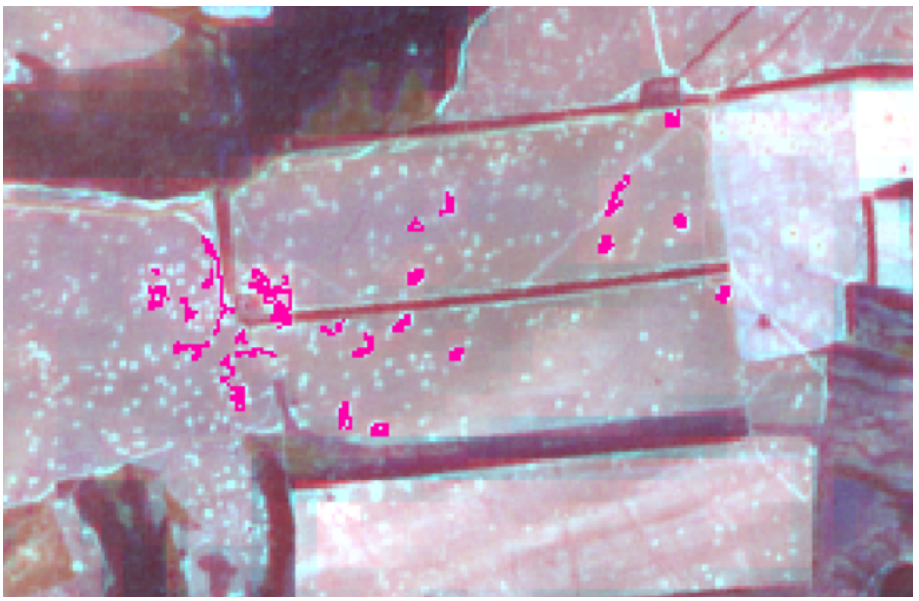


Рисунок 3.4 – Навчальний знімок з позначеними пошкодженими територіями

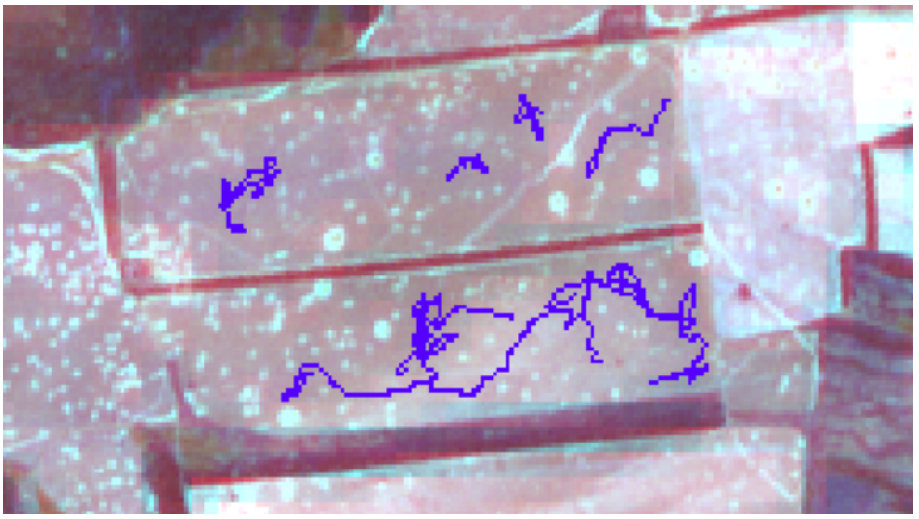


Рисунок 3.5 – Навчальний знімок з позначеними непошкодженими територіями

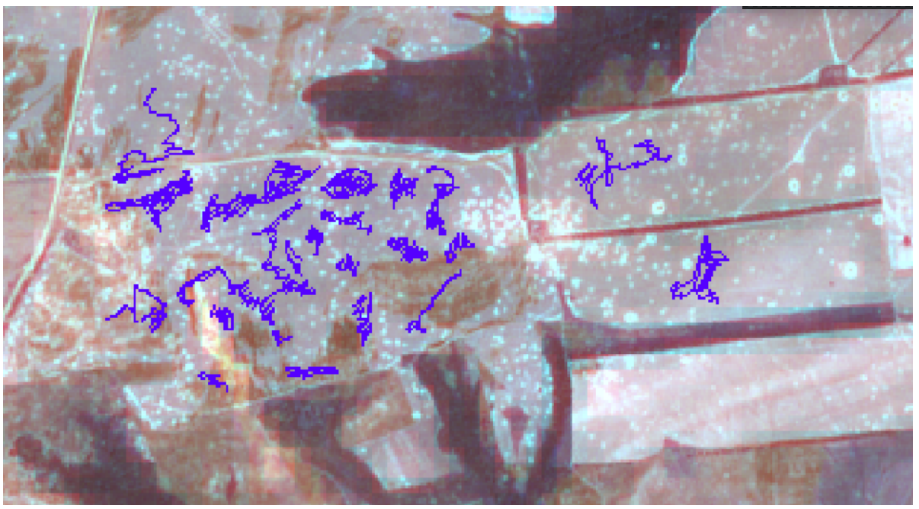


Рисунок 3.6 – Тренувальний знімок з позначеними непошкодженими територіями

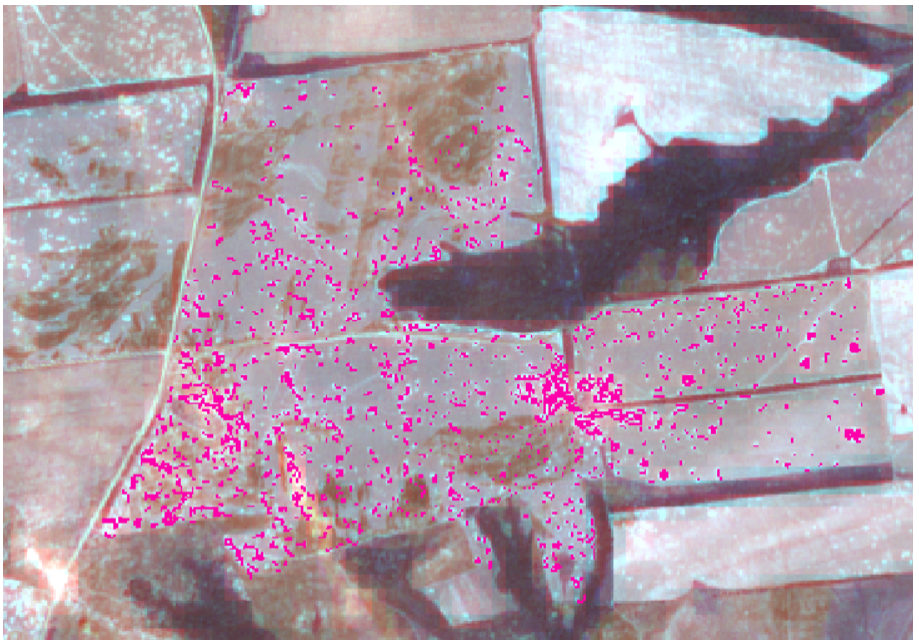


Рисунок 3.7 – Тренувальний знімок з позначеними пошкодженими територіями

3.3 Побудова графіків логістичної регресії

Далі побудовано графіки логістичної регресійної моделі для 12 каналів, але, щоб найкраще визначити результати роботи моделі, було

розглянуто найбільш інформативні канали для точок, такі як: 2 (рис. 3.8), 3 (рис. 3.9), 4 (рис. 3.10), 8 (рис. 3.11).

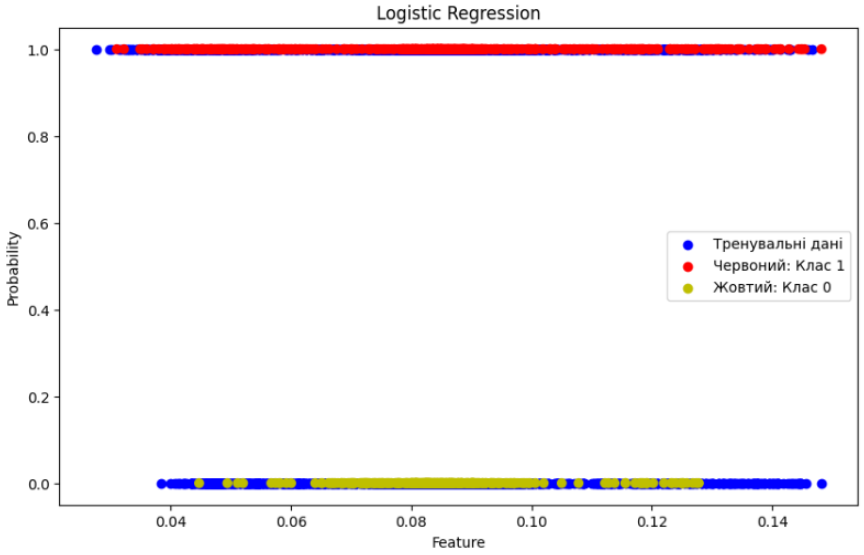


Рисунок 3.8 – Логістична регресія для каналу 2.

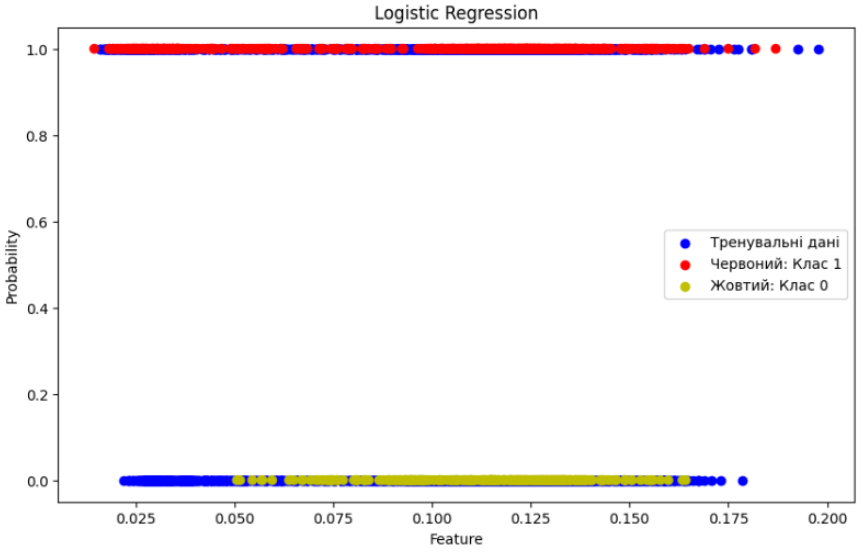


Рисунок 3.9 – Логістична регресія для каналу 3.

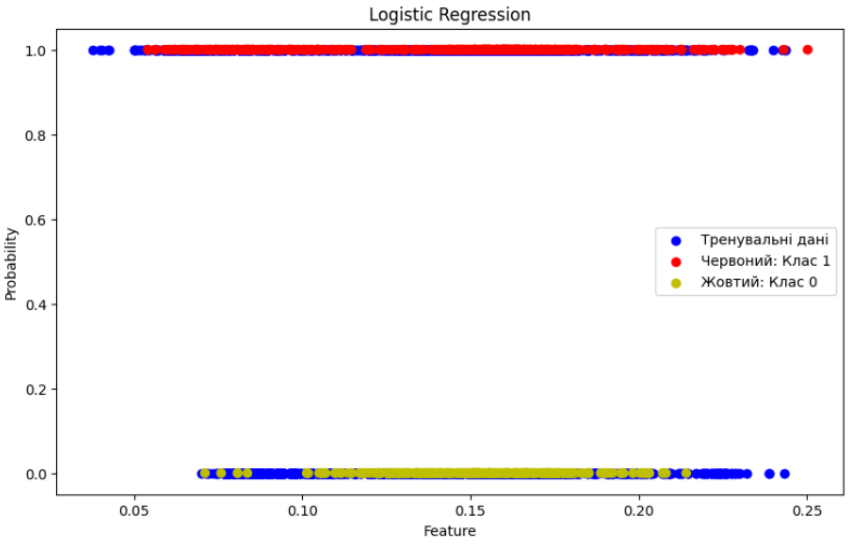


Рисунок 3.10 – Логістична регресія для каналу 4.

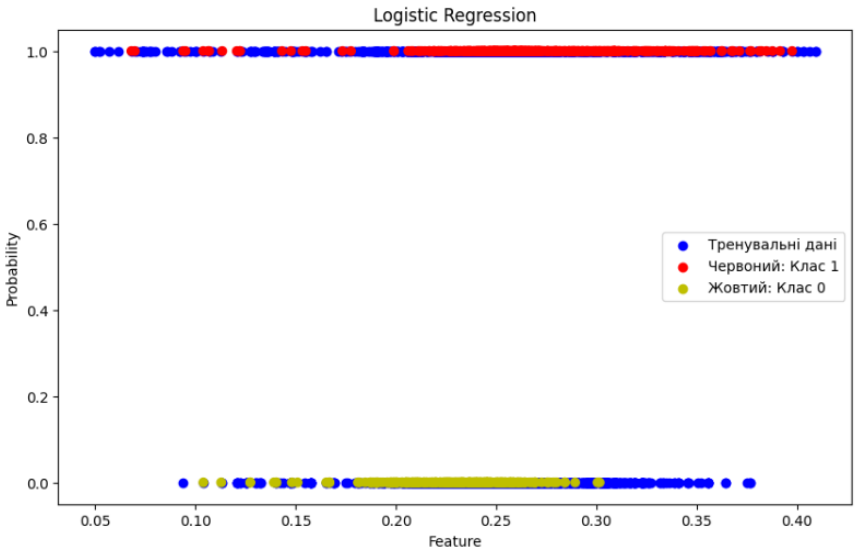


Рисунок 3.11 – Логістична регресія для каналу 8.

Щоб перевірити достовірність побудованих моделей, для кожного з класів було пораховано середнє (сер.зн.), мінімальне (мін.зн.) та максимальне значення (макс.зн.) (табл.1) та порівняно зі значеннями, які представлені на графіках.

a	c “0”	Н	ЗН	Н	М “1”	Н

		“0	“0	“1		“1
	0 48	4	0.14	62	0	0
	0 52	8	6	36	0	0
	0 45	0	1	33	0	0
	0 33	9	7	06	0	0

Таблиця 1 – Середнє, мінімальне та максимальне значення класів даних.

3.4 Обчислення функції втрат та оптимізація параметрів

Обчислення функції втрат та оптимізація параметрів є ключовими етапами в навчанні моделі машинного навчання. Обчислення значення функції втрат дозволяє оцінити, наскільки добре побудована модель прогнозує вихідні дані порівняно зі справжніми значеннями. Функція втрат вимірює різницю між прогнозованими і справжніми значеннями інформаційного критерію. Отримане значення функції втрат: 0.719896965483567, що є не зовсім хорошим результатом, тому далі була мета мінімізації значення цієї функції, щоб модель навчалась краще прогнозувати дані.

Оптимізація параметрів моделі, таких як ваги, полягає у знаходженні оптимальних значень цих параметрів, які мінімізують функцію втрат. Оптимізація використовує градієнтний спуск, щоб крок за кроком оновлювати значення параметрів, зменшуючи функцію втрат і покращуючи якість моделі. Виконано 1000 ітерацій оптимізації, щоб отримати достатньо хороше значення функції втрат. На останній ітерації отримано значення функції втрат 0.2545087874845312, а ваги виглядають так: [0.17379353, 0.20057366, 0.2934221, 0.35612184, 0.49651525, 0.65883357, 0.72975248,

0.75852759, 0.80971715, 0.81716313, 0.69827458, 0.50420139], що є досить хорошим результатом.

Таким чином, обчислення функції втрат і оптимізація параметрів дозволили моделі навчатись на вихідних даних та покращувати свої прогнозні здібності шляхом знаходження оптимальних параметрів.

3.5 Оцінка моделі

Оцінка моделі є важливим етапом після навчання, оскільки вона дозволяє оцінити якість роботи моделі на нових, раніше невикористаних даних. Це допомагає зрозуміти, наскільки добре модель здатна узагальнювати і робити прогнози на невідомих даних.

Побудована модель має такі характеристики:

- точність: 0.61761 (вимірює відсоток правильно класифікованих зразків);

- чутливість: 0.59819 (вимірює відсоток правильно класифікованих позитивних зразків серед усіх позитивних прогнозів);

- повнота: 0.73451 (вимірює відсоток правильно класифікованих позитивних зразків серед усіх справжніх позитивних зразків; використовується, коли важлива повнота виявлення певного класу);

- F-міра: 0.65938 (комбінована метрика, яка об'єднує точність та повноту; використовується для оцінки узгодженості між точністю та повнотою).

Ці метрики допомагають оцінити ефективність моделі з різних поглядів і з'ясувати, наскільки вона коректно класифікує дані. Це дозволяє прийняти рішення щодо використання моделі, вибрати оптимальні параметри або порівняти різні моделі між собою.

Висновки до Розділу 3

На графіках бачимо, що побудована логістична регресійна модель достатньо добре класифікує пошкоджені території з даних супутникових знімків, що підтверджують чисельні розрахунки для оцінки ефективності використання моделі. Код реалізації та всі обчислення для логістичної регресійної моделі наведено в Додатку А.

ВИСНОВКИ

Досліджено, що усі види воєнно-техногенного навантаження спричиняють значне забруднення та руйнування сільськогосподарських територій, а саме - ґрунтів. Щоб класифікувати різні види пошкоджень на територіях аграрного сектору потрібно створити модель, яка буде аналізувати дані та визначати, яких саме руйнувань зазнали ті чи інші земельні ділянки. Саме це було реалізовано у роботі. На основі даних супутникових знімків з супутника Sentinel-2 було завантажено серію мозаїк для модельної території дослідження Лисичанської об’єднаної територіальної громади, навчено та побудовано логістичну регресійну модель, яка класифікує пошкоджена ділянка чи ні. Обчислення функції втрат та оптимізація її параметрів дозволила моделі навчатись на вихідних даних та покращувати свої прогнозні здібності шляхом знаходження оптимальних параметрів, з чисельних розрахунків оцінок для моделі видно, що модель добре здатна узагальнювати і робити прогнози на невідомих даних.

У подальшому слід ще краще навчати модель визначати і класифікувати види пошкоджень за певними визначеними параметрами, рахувати кількість воронок від вибухів, площу забруднених ґрунтів, об’єм переміщеного ґрунту або ж навіть визначати види артилерійських снарядів, якими було уражено територію.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Шелестов А. Ю., Лавренюк М. С., Яйлимов Б. Я., Ткаченко О. М. Методи глибинного навчання для геопросторового аналізу та задач спостереження Землі, К.: “Наукова думка” – 2019. – 228 с.

2. Kussul, N., Lavreniuk M., Shelestov, A., & Skakun, S. (2018). Crop inventory at regional scale in Ukraine: developing in season and end of season crop maps with multi-temporal optical and SAR satellite imagery. European Journal of Remote Sensing, 51(1), 627-636.

3. N. Kussul, A. Shelestov, M. Lavreniuk, B. Yailymov, A. Kolotii, H. Yailymova, S. Skakun, L. Shumilo, Y. Bilokonska (2021). SDG indicator 11.3.1 within HORIZON-2020 SMURBS. Space research in Ukraine. pp. 91-95.

4. Kussul, N., Lavreniuk, M., Skakun, S., & Shelestov, A. (2017) “Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data”, IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 5, pp. 778–782.

5. N. Kussul, A. Shelestov, M. Lavreniuk, B. Yailymov, A. Kolotii, H. Yailymova, S. Skakun, L. Shumilo, Y. Bilokonska (2021). SDG indicator 11.3.1 within HORIZON-2020 SMURBS. Space research in Ukraine. pp. 91-95.

6. [Електронний ресурс] / Негрей М., Тараненко А., Костенко І. "Аграрний сектор України в умовах війни: проблеми та перспективи". URL: https://economyandsociety.in.ua/index.php/journal/article/view/1474

7. [Електронний ресурс] / KSE Агроцентр. (2022). Agricultural war damages review Ukraine. Rapid Damage Assessment Режим доступу: https://kse.ua/wp-content/uploads/2022/06/Damages_report_issue1-1.pdf

8. [Електронний ресурс] / Сплодитель А., Голубцов О., Чумаченко С., Сорокіна Л. "Забруднення земель внаслідок агресії росії проти України". Режим доступу:: https://ecoaction.org.ua/wp-content/uploads/2023/03/zabrudnennia-zemel-vid-rosii2.pdf

ДОДАТОК А ТЕКСТИ ПРОГРАМ

А.1 Програма 1

Створення DataFrame на основі даних

```
from typing_extensions import Concatenate
!pip install rasterio
import cv2 as cv
import rasterio
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

img = cv.imread('/content/drive/MyDrive/diploma/22.11-1.png')
img1 = rasterio.open('/content/drive/MyDrive/diploma/damage_-2022-06-21-2022-06-28-S2_37n_cut.tif').read()
img2 = cv.imread('/content/drive/MyDrive/diploma/22.22-11.png')
img = np.float32(img)
img1 = np.float32(img1)

print(img1.shape)

image = img[:,:,0]
list_indexes = np.where(image > 0)
arr_data = image[ list_indexes[0], list_indexes[1]]

print(arr_data.shape)

arr_data = img1[ :, list_indexes[0], list_indexes[1]].swapaxes(0, 1)
print(arr_data.shape)

df1 = pd.DataFrame(arr_data)
df1['x'] = list_indexes[0]
df1['y'] = list_indexes[1]
df1['mask'] = 1
df1 = df1.replace([np.inf], 1000)
```

```
df1 = df1.replace([-np.inf], -1000)
df1 = df1.dropna()
print(df1.shape)

image = img2[:,:,0]
list_indexes = np.where(image > 0)
arr_data = image[ list_indexes[0], list_indexes[1]]

print(arr_data.shape)

arr_data = img1[ :, list_indexes[0], list_indexes[1]].swapaxes(0, 1)
print(arr_data.shape)

df2 = pd.DataFrame(arr_data)
df2['x'] = list_indexes[0]
df2['y'] = list_indexes[1]
df2['mask'] = 0
df2 = df2.replace([np.inf], 1000)
df2 = df2.replace([-np.inf], -1000)
df2 = df2.dropna()
print(df2.shape)

df3 = pd.concat([df1, df2])

pd.concat([df3])

img = cv.imread('/content/drive/MyDrive/diploma/stud2.png')
img1 =
rasterio.open('/content/drive/MyDrive/diploma/damage_-2022-06-21-2022-06-28-S2_37n_cut.tif').read()
img2 = cv.imread('/content/drive/MyDrive/diploma/stud1.png')
img = np.float32(img)
img1 = np.float32(img1)

print(img1.shape)

image = img[:,:,0]
list_indexes = np.where(image > 0)
arr_data = image[ list_indexes[0], list_indexes[1]]

print(arr_data.shape)
```

```
arr_data = img1[ :, list_indexes[0], list_indexes[1]].swapaxes(0, 1)
print(arr_data.shape)

df1 = pd.DataFrame(arr_data)
df1['x'] = list_indexes[0]
df1['y'] = list_indexes[1]
df1['mask'] = 1
df1 = df1.replace([np.inf], 1000)
df1 = df1.replace([-np.inf], -1000)
df1 = df1.dropna()
print(df1.shape)

image = img2[:,:,0]
list_indexes = np.where(image > 0)
arr_data = image[ list_indexes[0], list_indexes[1]]
print(arr_data.shape)

arr_data = img1[ :, list_indexes[0], list_indexes[1]].swapaxes(0, 1)
print(arr_data.shape)

df2 = pd.DataFrame(arr_data)
df2['x'] = list_indexes[0]
df2['y'] = list_indexes[1]
df2['mask'] = 0
df2 = df2.replace([np.inf], 1000)
df2 = df2.replace([-np.inf], -1000)
df2 = df2.dropna()
print(df2.shape)

df3_s = pd.concat([df1, df2])

pd.concat([df3])
```

Навчання логістичної регресії

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
```

```
# Розділяємо дані на навчальний та тестовий набори
X = df3[[0, 1 , 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 ]]
y = df3[['mask']]
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Ініціалізуємо та навчаємо модель логістичної регресії
model = LogisticRegression()
model.fit(X_train, y_train)

# Зробимо прогнози для тестового набору
y_pred = model.predict(X_test)

# Обчислюємо метрики оцінки моделі
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred)

# Виводимо метрики
print("Accuracy:", accuracy)
print("Precision:", precision)
print("Recall:", recall)
print("F1 Score:", f1)

# Перетворення імен стовпців на рядки
df3_s.columns = df3_s.columns.astype(str)

# Виконання навчання логістичної регресії
logreg = LogisticRegression()
logreg.fit(X_train, y_train)

# Використання навченої моделі для передбачення на новому наборі даних df3
predictions = logreg.predict(df3.drop(['x', 'y', 'mask'], axis=1))
```

Побудова графіків логістичної регресії

```
import numpy as np
import matplotlib.pyplot as plt
```

```
# Передбачення значень для нових даних
y_pred = logreg.predict_proba(X_test)[:, 1] > 0.5  # Передбачення ймовірності класу 1

# Побудова графіку
plt.figure(figsize=(10, 6))
plt.scatter(X_train[[8]], y_train, color='b', label='Тренувальні дані')  # Точки тренувального набору
plt.scatter(X_test[[8]], y_pred, c=np.where(y_pred, 'r', 'y'))  # Лінія логістичної регресії
plt.xlabel('Feature')
plt.ylabel('Probability')
plt.title('Logistic Regression')
plt.scatter([], [], color='r', label='Червоний: Клас 1')  # Позначка для червоного кольору
plt.scatter([], [], color='y', label='Жовтий: Клас 0')  # Позначка для жовтого кольору
plt.legend()
plt.show()
```

Обчислення середнього, мінімального та середнього значень

```
# Після створення df3

# Обчислення для mask = 1
df_mask_1 = df3[df3['mask'] == 1]
mean_mask_1 = df_mask_1.mean()
max_mask_1 = df_mask_1.max()
min_mask_1 = df_mask_1.min()

# Обчислення для mask = 0
df_mask_0 = df3[df3['mask'] == 0]
mean_mask_0 = df_mask_0.mean()
max_mask_0 = df_mask_0.max()
min_mask_0 = df_mask_0.min()

# Виведення результатів
print("Mask = 1:")
print("Середнє значення:", mean_mask_1)
print("Найбільше значення:", max_mask_1)
print("Найменше значення:", min_mask_1)
print()
print("Mask = 0:")
print("Середнє значення:", mean_mask_0)
```

```
print("Найбільше значення:", max_mask_0)
print("Найменше значення:", min_mask_0)
```

Обчислення метрик оцінки моделі

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
# Розділяємо дані на навчальний та тестовий набори
X = df3[[0, 1 , 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 ]]
y = df3[['mask']]
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Ініціалізуємо та навчаємо модель логістичної регресії
model = LogisticRegression()
model.fit(X_train, y_train)

# Зробимо прогнози для тестового набору
y_pred = model.predict(X_test)

# Обчислюємо метрики оцінки моделі
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred)

# Виводимо метрики
print("Accuracy:", accuracy)
print("Precision:", precision)
print("Recall:", recall)
print("F1 Score:", f1)
```

Обчислення значення та градієнта функції втрат

```
import numpy as np

# Функція втрат логістичної регресії (бінарна кросс-ентропія)
def logistic_regression_loss(y_true, y_pred):
```

```
    epsilon = 1e-8  # Додатковий термін для запобігання діленню на нуль
    loss = -np.mean(y_true * np.log(y_pred + epsilon) + (1 - y_true) * np.log(1 - y_pred + epsilon))
    return loss
# Передбачення ймовірностей класів
y_pred = model.predict_proba(X_test)
# Отримання ймовірностей для класу 1
y_pred_class1 = y_pred[:, 1]

# Зведення до однакового розміру
y_true = df3['mask'].values[:len(y_pred_class1)]
y_pred_class1 = y_pred_class1[:len(y_true)]

# Обчислення значення функції втрат
loss_value = logistic_regression_loss(y_true, y_pred_class1)
print("Значення функції втрат:", loss_value)



import numpy as np

# Функція втрат логістичної регресії (бінарна кросс-ентропія)
def logistic_regression_loss(y_true, y_pred):
    epsilon = 1e-8  # Додатковий термін для запобігання діленню на нуль
    loss = -np.mean(y_true * np.log(y_pred + epsilon) + (1 - y_true) * np.log(1 - y_pred + epsilon))
    return loss

# Ініціалізація параметрів моделі
weights = np.zeros(X.shape[1])

# Гіперпараметри градієнтного спуску
learning_rate = 0.01
num_iterations = 1000

# Градієнтний спуск
for i in range(num_iterations):
    # Обчислення передбачення
    y_pred = sigmoid(np.dot(X, weights))

    # Обчислення градієнта функції витрат
    gradient = np.dot(X.T, (y_pred - y_true)) / len(y_true)
```

```
    # Оновлення ваг моделі
    weights -= learning_rate * gradient

    # Оновлення передбачення
    y_pred = sigmoid(np.dot(X, weights))

    # Обчислення функції витрат для моніторингу збіжності
    loss = logistic_regression_loss(y_true, y_pred)
    print("Iteration:", i, "Loss:", loss)

# Остаточні ваги моделі
print("Final Weights:", weights)
```