

УДК 004

Кирик М.М.¹, Ільяшенко М.Б.²

¹ студ. гр. КНТ-214м НУ «Запорізька політехніка»

² доц. НУ «Запорізька політехніка»

ДОСЛІДЖЕННЯ ТА ПРОГРАМНА РЕАЛІЗАЦІЯ МЕТОДІВ ВІДБОРУ ОЗНАК

Об'єкт дослідження – процеси обчислень, пов'язані з відбором інформативних ознак.

Предмет дослідження – методи відбору інформативних ознак.

Мета роботи – дослідження та реалізація методів відбору ознак.

Матеріали, методи та технічні засоби: мова програмування Python, середовище розробки PyCharm.

В ході виконання роботи було проаналізовано та досліджено процеси обчислень, пов'язані з відбором інформативних ознак.

Визначено, що відбір ознак є підходом до скорочення розмірності, що передбачає вибір обмеженої підмножини релевантних ознак з первинного набору даних шляхом усунення шумових, надлишкових або малозначущих елементів. Виконання такого відбору здатне підвищити ефективність навчання моделей, забезпечити більшу точність прогнозування, зменшити обчислювальні витрати та полегшити інтерпретацію результатів. Відзначено, що надмірність даних у сучасних великих наборах змушує застосовувати розумні алгоритми для виявлення критично важливої інформації, що забезпечує ефективне формування моделей у задачах класифікації, регресії або кластеризації. Використання методів відбору ознак дозволяє скоротити обсяг даних для обробки, зменшити складність моделей, прискорити навчання та знизити ризик перенавчання. Проаналізовано методи відбору інформативних ознак.

За результатами проведеного аналізу зроблено висновок, що у наш час існує досить багато програмних засобів для аналізу даних та відбору ознак. Проте деякі програмні засоби, що реалізують методи відбору ознак, попри свою практичну цінність, мають низку обмежень, які впливають на якість та ефективність їх використання. Складність деяких методів призводить до значних обчислювальних витрат, що стає критичним при роботі з великими масивами даних. Крім того, реалізовані програмні рішення не завжди

забезпечують достатню гнучкість, оскільки часто орієнтовані на використання певних підходів і можуть мати обмежені можливості налаштування для специфічних завдань. Ще однією суттєвою проблемою є ризик втрати важливої інформації, коли алгоритми відбору зменшують простір ознак надмірно агресивно, і, як наслідок, модель втрачає здатність до адекватного відображення складних залежностей. В окремих випадках труднощі викликає також інтерпретація результатів, оскільки автоматизовані системи не завжди надають зрозуміле пояснення того, чому були відкинуті певні характеристики. Обмеженість інтеграційних можливостей із зовнішніми аналітичними програмними платформами та недостатня оптимізація під конкретні архітектури обчислювальних систем також ускладнюють практичне застосування таких рішень. Тому актуальною є розробка програмного забезпечення, що реалізує методи відбору ознак.

Сформульовано функціональні вимоги до програмного забезпечення, що реалізує методи відбору ознак.

Для реалізації програмного забезпечення, що реалізує методи відбору ознак обрано мову програмування Python, яка характеризується розвинутою екосистемою бібліотек і фреймворків, що забезпечують реалізацію алгоритмів машинного навчання, статистичного аналізу та роботи з великими масивами даних.

Для створення програмного забезпечення, що реалізує методи відбору ознак обрано середовище розробки PyCharm завдяки поєднанню інструментів, орієнтованих на ефективну роботу з мовою Python та бібліотеками для аналізу даних. PyCharm забезпечує інтелектуальне автодоповнення коду, зручну навігацію по проєкту та можливості швидкого рефакторингу, що істотно прискорює процес розробки й мінімізує кількість помилок. Важливою перевагою є інтеграція з віртуальними середовищами та системами керування залежностями, що дозволяє легко організувати роботу з різними бібліотеками машинного навчання та інструментами для обробки даних. PyCharm підтримує зручні засоби налагодження та профілювання, що є необхідними при оптимізації алгоритмів відбору ознак, особливо коли обробляються великі масиви інформації.

Новизна роботи полягає у тому, що створено метод відбору ознак, який полягає у поступовій підготовці даних, виділенні ключових характеристик, застосуванні адаптивного алгоритму, перевірці отриманих результатів та їх інтерпретації, що дозволяє підвищити точність аналізу, зменшити вплив випадкових похибок та забезпечити більшу практичну значущість кінцевих висновків.

Практична цінність роботи полягає у тому, що розроблено програмне забезпечення, що реалізує методи відбору ознак.

Запропоновано структуру програмного забезпечення, що реалізує методи відбору ознак. Визначено, що структура програмного забезпечення, що реалізує методи відбору ознак, організована у вигляді п'яти взаємопов'язаних модулів: модуль завантаження та підготовки даних, модуль управління ознаками, модуль методів відбору, модуль оцінки якості відібраних ознак, модуль візуалізації результатів.

Описано функціонування програмного забезпечення, що реалізує методи відбору ознак. Визначено, що у процесі функціонування програми спочатку здійснюється підготовка вхідних даних, після чого активуються алгоритмічні процедури, здатні оцінювати внесок кожної змінної у загальну результативність. Оцінювання може базуватися на статистичних показниках, навчанні моделей чи пошуку комбінацій, що мінімізують надлишковість та шум, поступово простір ознак звужується, і програмне забезпечення формує підмножину характеристик, яка найкраще описує сутність досліджуваного явища.

Описано особливості реалізації програмного забезпечення, що реалізує методи відбору ознак. Виконано проектування інтерфейсу взаємодії користувача з програмним забезпеченням, що реалізує методи відбору ознак.

Виконано тестування розробленого програмного забезпечення, що реалізує методи відбору ознак. Результати тестування програмного забезпечення показали, що розроблена програма дозволяє забезпечити автоматизацію процесів вибору інформативної комбінації ознак.

Галузь використання – розпізнавання образів, прогнозування, аналіз даних.