

УДК 621.326

Бонар В. - ст. гр. СНа-12

*Тернопільський національний технічний університет імені Івана Пулюя*

## **ЗАСТОСУВАННЯ АЛГОРИТМУ НЕКЕРОВНОГО МАШИННОГО НАВЧАННЯ DBSCAN ДЛЯ ЗНАХОДЖЕННЯ АНОМАЛІЙ В OBD-II ДАНИХ**

Науковий керівник: к.т.н., доцент Готович В. А.

Bonar V.

*Ternopil Ivan Puluj National Technical University*

## **USAGE OF THE UNSUPERVISED MACHINE LEARNING ALGORITHM DBSCAN FOR ANOMALY DETECTION IN OBD-II DATA**

Supervisor: Hotovych V. A., PhD

Ключові слова: DBSCAN, OBD-II, виявлення аномалій.

Keywords: DBSCAN, OBD-II, anomaly detection.

Дане дослідження фокусується на розробці та вдосконаленні методу виявлення аномалій в даних OBD-II за допомогою алгоритму некерованого машинного навчання DBSCAN (Density Based Spatial Clustering of Applications with Noise) [1].

Основна ідея DBSCAN полягає в тому, що кластери — це області високої щільності точок, розділені областями низької щільності. Алгоритм оцінює щільність сусідства кожної точки даних за допомогою параметра радіуса, який називається епсилон, і вважає точку основною, якщо вона має достатню кількість сусідів у цьому радіусі.

Після того, як основна точка знайдена, DBSCAN розширює кластер, рекурсивно включаючи всі сусідні точки, які також задовольняють вимогам щільності. Завдяки цьому, алгоритм здатний виявляти кластери довільної форми, що відрізняє його від більш жорстких алгоритмів, таких як метод k-середніх. Точки, які не відповідають критеріям входження до будь-якого кластера, оскільки вони надто ізольовані, позначаються як аномалії.

Перевагою алгоритму DBSCAN є відсутність вимоги визначення кількості кластерів. DBSCAN визначає кількість кластерів на основі фактичної структури даних. Це робить його особливо ефективним у таких програмах, як виявлення аномалій, де аномальні точки не є частиною жодної щільної групи і тому природним чином виключаються алгоритмом.

DBSCAN має 2 параметра для зміни:

1) Епсилон (epsilon) - визначає максимальну відстань між двома точками, для визначення їх сусідніми. Менший епсилон робить алгоритм суворішим, у результаті чого більше точок класифікується як аномальні, тоді як більший епсилон може об'єднати різні кластери та зменшити чутливість до локальних варіацій.

2) Мінімальна кількість точок для формування кластеру. Збільшення цього значення робить алгоритм більш консервативним у формуванні кластерів, що зменшує кількість виявлених аномальних точок. Зменшення параметру дозволяє DBSCAN виявляти більш дрібні кластери, ризикуючи додати більше аномалій.

Було проведено експеримент використовуючи зібрані OBD-II дані з автомобіля Honda CR-V 2013 з дизельним двигуном. Дані було нормалізовано та за допомогою мови програмування Python та бібліотеки sklearn навчено модель некерованого машинного навчання DBSCAN.

Для тестування алгоритму вихідні дані було спотворено шляхом додання в них шуму використовуючи алгоритм Гауссівського шуму [2].

Результати тестування зображено на рис. 1. Епсилон (threshold) було визначено експериментально на рівні 5. Мінімальна кількість сусідів у кластері визначено як 5. Червоні точки - аномалії. Сині - не аномальні дані.

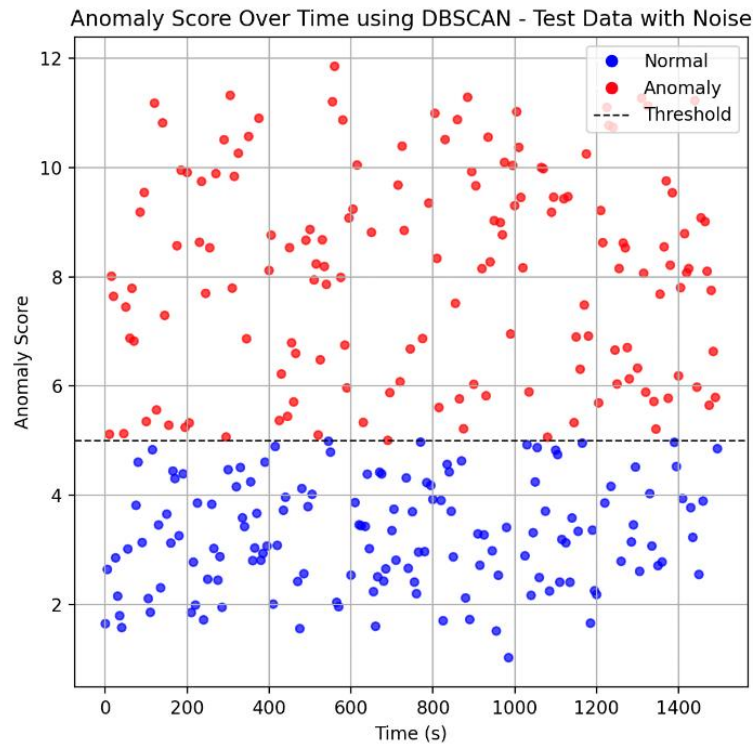


Рис. 1 - Визначення аномалій OBD-II алгоритмом DBSCAN

Проведений експеримент дозволяє більше зрозуміти перспективу і можливість аналізу великих даних OBD-II. Використаний алгоритм DBSCAN показав досить високу точність у знаходженні аномалій. В майбутніх дослідженнях планується розглянути більше методів з їхнім порівнянням.

### Література

1. Rehman, Saif ur & Asghar, Sohail & Fong, Simon. (2014). DBSCAN: Past, present and future. 10.1109/ICADIWT.2014.6814687.
2. Zhang, Xinhua. (2016). Gaussian Distribution. 10.1007/978-1-4899-7502-7\_107-1.