

За результатами, Isolation Forest та Автоенкодер показали найкращі показники для виявлення аномальних записів у вибраному наборі даних.

- Isolation Forest забезпечив високу точність у класифікації аномалій, зберігаючи баланс між Precision та Recall. Цей метод виділився завдяки здатності ідентифікувати аномалії за допомогою ізольованих дерев, що автоматично враховують рідкість аномальних точок.
- Автоенкодер, реалізований на базі глибокої нейронної мережі, показав хороші результати завдяки можливості "вчитися" на нормальних даних та виявляти аномалії за значенням помилки реконструкції. Велика різниця між вихідними та відновленими даними для аномалій є ознакою того, що автоенкодер розпізнає нехарактерні для навчальної вибірки патерни.

Аналіз роботи М.Д.

Методи показали різну ефективність на цьому наборі даних, зокрема через особливості алгоритмів та структуру даних:

- Isolation Forest добре справляється з великими наборами даних і здатен ідентифікувати аномалії завдяки побудові випадкових дерев. Цей метод використовує специфіку розподілу вибірки, що дозволяє йому ефективно виділяти окремі точки з низькою щільністю.
- One-Class SVM не завжди дає хороші результати для великих обсягів даних, оскільки він обчислювально затратний, особливо при високій розмірності ознак. Через це він менш підходить для масових аномалій у великих наборах даних.
- Local Outlier Factor підходить для задач, де аномальні точки значно відрізняються за щільністю від своїх сусідів. Однак його ефективність чутлива до вибору гіперпараметрів, зокрема до кількості сусідів, що може призводити до менш стабільних результатів.
- Автоенкодер показав хорошу ефективність завдяки адаптації до складних і нерегулярних даних, адже нейронна мережа "навчається" відновлювати нормальні зразки, а для аномалій отримує високу помилку реконструкції.

На додачу: Автоенкодери мають значний потенціал для задач виявлення аномалій, особливо на великих і складних наборах даних. Глибокі нейронні мережі можуть адаптуватися до складних нелінійних зв'язків, що дозволяє їм ефективно навчатися відновлювати нормальні зразки та розпізнавати аномалії. Використання автоенкодерів стає особливо ефективним на високорозмірних або нерегулярних наборах даних, де інші методи можуть зазнавати труднощів.

Узагальнення: Isolation Forest та автоенкодер є перспективними методами для вирішення задачі виявлення аномалій на вибраному наборі даних. Обидва методи показали високу ефективність завдяки своїм характеристикам, і вибір конкретного методу залежить від особливостей завдання, структури даних і доступних обчислювальних ресурсів.