Тестовое задание
Оценка качества бинарного классификатора

Вы работает специалистом по контролю качества программного обеспечения в крупной компании, которая занимается разработкой продуктов в области информационной безопасности.

Ваш отдел работает над системой анализа сетевого трафика и автоматического обнаружения сетевых атак. В качестве основного подхода к решению поставленной задачи предложено использовать методы бинарной классификации сетевых соединений на основе различных характеристик.

Для первичного анализа было предложено опробовать метод на данных, взятых из KDD cup 1999.

Команда разработки предложила два различных алгоритма классификации:

- логистическая регрессия;
- наивный байесовский классификатор.

Оба классификатора были обучены на некоторой подвыборке для случая, когда все возможные атаки отнесены к одному классу (обозначим его как 1), а нормальные сетевые соединения — к другому (обозначим как 0).

Далее была подготовлена тестовая выборка. По ней были получены вероятности принадлежности объекта к классу сетевых атак (файлы `LogisticRegression_pred.csv` и `NaiveBayes_pred.csv`). Также была подготовлена выборка с правильными ответами (`test_labels.csv`).

Ваша задача — *написать программу на языке Python, которая определит качество работы каждого из классификаторов.* На вход программе в качестве аргументов командной строки подаются пути к файлам с результатами классификации и ответами. Программа должна напечатать число — метрику качества работы классификатора. Выбрать метрику качества предлагается Вам.

С помощью написанной программы сравните два предложенных алгоритма и укажите наилучший.

В ходе решения Вам помогут следующие источники

Список литературы

[1] Классификация

[2] Логистическая регрессия

[3] Байесовский классификатор

[4] Python для анализа данных

[5] Метрики качества классификации