

Задание по теме 3

Склифасовский Денис

Задание:

На основе представленных в файле credit.csv данных решить задачу скоринга. Возможные варианты инструментов: нейросеть, дерево решений, лес деревьев решений, логистическая регрессия. Построить не менее 2 сетей с разной архитектурой. Сделать обоснованный выбор в пользу одного из инструментов на основе метрик ошибки.

Возможные программные средства: Deductor, Loginom, R (RStudio), Python

Решение:

Нейронные сети, дерево решений, лес деревьев решений, логистическая регрессия были написаны на Python с использованием tensorflow и sklearn.

Решение можно посмотреть в прикрепленных файлах или на github:

https://github.com/MrSk1f/FA_IIS/blob/main/lab2/lab.ipynb

Вывод:

1. Model_nn1 (первая нейросеть) показывает низкую точность (0.648148), полноту (0.265152) и F1-меру (0.376344). Это означает, что модель имеет трудности в правильном определении проблемных клиентов, и доля ложных срабатываний высока.
2. Model_nn2 (вторая нейросеть) также демонстрирует низкую точность (0.717949), полноту (0.212121) и F1-меру (0.327485). В сравнении с Model_nn1, она показывает незначительное улучшение в точности, но все еще имеет слабую способность обнаруживать проблемных клиентов.
3. Дерево решений имеет низкую точность (0.281250), но относительно высокую полноту (0.340909) и F1-меру (0.308219). Оно также страдает от низкой точности в предсказании проблемных клиентов.
4. Лес деревьев решений показывает лучшие результаты с точностью (0.611111), полнотой (0.250000) и F1-мерой (0.354839) в сравнении с предыдущими моделями, но все еще далеко от идеальных показателей.
5. Логистическая регрессия демонстрирует наилучшие результаты с точностью (0.708333), полнотой (0.257576) и F1-мерой (0.377778). Она показывает более сбалансированные результаты и лучше справляется с предсказанием проблемных клиентов.

Логистическая регрессия показывает наилучшую производительность среди представленных моделей. Ее высокая точность и F1-мера указывают на хорошую способность модели предсказывать проблемных клиентов, а также учитывать баланс между точностью и полнотой.

Таким образом, на основе метрик ошибки, можно сделать выбор в пользу логистической регрессии как наиболее эффективного инструмента для решения задачи скоринга в данном случае.