



3. Huang, G. Densely Connected Convolutional Networks / G. Huang, Z. Liu, L. van der Maaten, K.Q. Weinberger // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) — 2017. — P. 4700-4708.
4. Palanisamy K. Rethinking CNN Models for Audio Classification / K. Palanisamy, D. Singhanian, A. Yao // ArXiv — 2020. — Vol. abs/2007.11154.
5. Guzhov A. ESResNet: Environmental Sound Classification Based on Visual Domain Models / A. Guzhov, F. Raue, J. Hees, A. Dengel // 25th International Conference on Pattern Recognition (ICPR) — 2020. — P. 4933-4940.
6. L. Wyse. Audio Spectrogram Representations for Processing with Convolutional Neural Networks // Proceedings of the First International Workshop on Deep Learning and Music joint with IJCNN. — 2017. — Vol.1. — No.1. — P. 37-41.

А.А. Медведева, В.А. Соболев

ПЕРСОНАЛИЗАЦИЯ ПРЕДЛОЖЕНИЙ НА ОСНОВЕ АБОНЕНТОВ «МЕГАФОН»

(Самарский государственный технический университет)

Цель проекта: разработать модель, которая сможет формировать персональные предложения абонентам «МегаФона».

Задача проекта: Разработка модели, которая принимает файл test.csv из корневой папки и записывает в эту же папку файл answers_test.csv. В этом файле должны находиться четыре столбца: id, vas_id, buy_time и target.

Актуальность проекта

Искусственный интеллект проникает во все сферы бизнеса и индустрий. Теоретические знания в области машинного обучения и нейронных сетей удастся эффективно применять на практике для решения прикладных бизнес-задач. Огромное количество данных о своих клиентах дает телеком-компания «Мегафон» возможность использовать методы искусственного интеллекта для создания дополнительных ценностей и получения конкурентного преимущества

Основные этапы работы

Чтобы решить данную задачу надо сделать правильную оценку и выбор модели. На данный момент есть множество доступных моделей машинного обучения, мы решили попробовать в работе несколько и посмотреть, какой на практике будет лучше.

Для сравнения мы взяли 4 метода:

- Логистическая регрессия
- Случайный лес
- Градиентный бустинг
- Экстремальное усиление градиента

После исследования выяснилось что метод экстремального усиления градиента работает эффективнее, так как точность его больше.

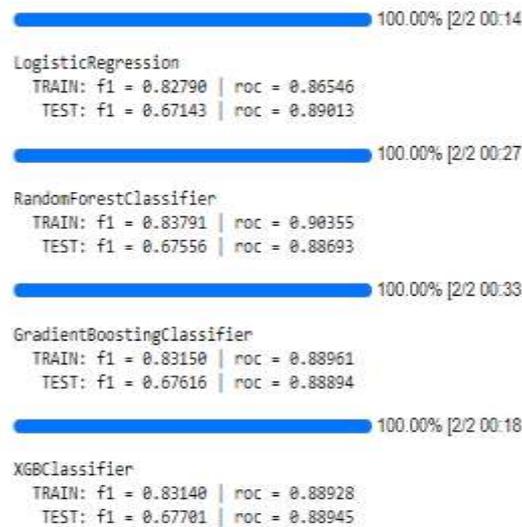


Рис. 1

Конструирование и выбор признаков

- **Проектирование признаков.** Процесс создания новых функций из необработанных данных для повышения эффективности возможностей прогнозирования алгоритма обучения. Для конструирования признаков требуется дополнительная информация, которая не так очевидна в исходном наборе признаков.
- **Выбор признаков.** В этом процессе выбирается ключевое подмножество признаков с целью сокращения размерности задачи обучения.

Выбираем базовый уровень

- Мы очистили данные, провели разведочный анализ и сконструировали признаки. И прежде чем перейти к созданию модели, нужно выбрать исходный базовый уровень (naive baseline) — некое предположение, с которым мы будем сравнивать результаты работы моделей. Если они окажутся ниже базового уровня, мы будем считать, что машинное обучение неприменимо для решения этой задачи, или что нужно попробовать иной подход.

Для регрессионных задач в качестве базового уровня разумно угадывать медианное значение цели на обучающем наборе для всех примеров в тестовом наборе. Эти наборы задают барьер, относительно низкий для любой модели.

Локальные интерпретируемые моделезависимые объяснения.

- Инструмент, с помощью которого можно постараться разобраться в том, как «думает» наша модель. LIME позволяет объяснить, как сформирован одиночный прогноз любой модели машинного обучения. Для этого локально, рядом с каким-нибудь измерением на основе простой модели наподобие линейной регрессии создаётся упрощённая модель.



Результаты

Итоговая модель

```
Ввод [19]: report(MODEL(**PARAMS), TRAIN, FIELDS)
```

```
-----  
F1 (test) = 0.69650   roc (test) = 0.92272   roc (train) = 0.89830  
-----
```

	precision	recall	f1-score	support
0.0	1.00	0.85	0.91	77305
1.0	0.32	0.96	0.48	5795
accuracy			0.85	83100
macro avg	0.66	0.90	0.70	83100
weighted avg	0.95	0.85	0.88	83100

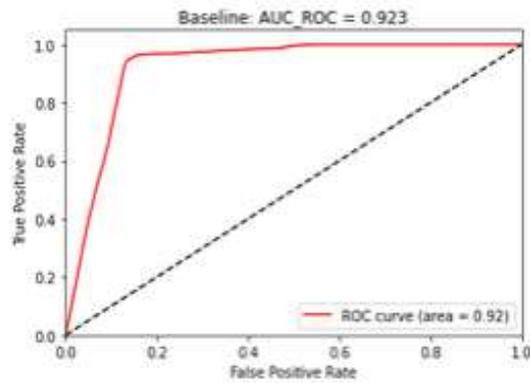
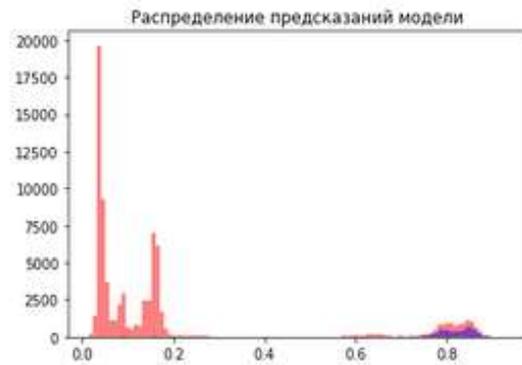
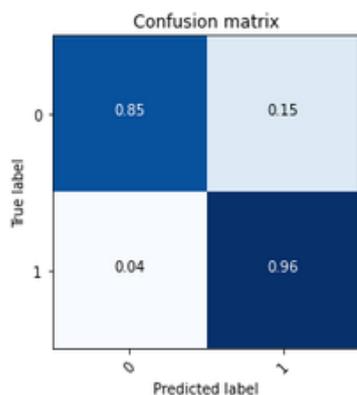


Рис. 2



Сохраняем результат

```
Ввод [0]: md = trainModel(MODEL(**PARAMS), TRAIN, FIELDS)
TEST['target'] = md.predict_proba(TEST[FIELDS])[:,1]
result = TEST[['index', 'target']].set_index('index')
```

```
Ввод [0]: dms = pd.read_csv(PATH+'data_test.csv', sep=',', index_col=0)
dms['target'] = np.NaN
dms.update(result)
dms.to_csv(PATH+'answers_test.csv', sep=',')
dms['target'] = np.round(dms['target'])
dms.to_csv(PATH+'answers_test_class.csv', sep=',')
```

Рис. 3

Заключение

В подведении итогов нашей работы можно выделить то, что проект нашей команды будет крайне полезным для телеком-компании «Мегафон». Метод машинного обучения способен выявить существующие взаимосвязи или продемонстрировать их отсутствие. При этом ситуация получения ошибочного результата (выявления «фантомных» связей и тенденций) практически невозможна. Такая гарантия от технических ошибок особенно важна в случае поиска неявных закономерностей в массивах исторических данных, поскольку исследователи-гуманитарии, как правило, испытывающие доверие к результатам «точных математических расчетов», способны выстроить убедительную интерпретацию фактически любых «фантомных зависимостей». Таким образом, залогом успешного анализа больших исторических данных является внимание к уровню точности выявляемых связей и зависимостей

Литература

1. Изучаем Python. Программирование игр, визуализации данных, веб-приложения. — СПб.: Питер, 2017. — 496с.: ил. — (Серия «Библиотека программиста»)
2. Николенко С., Кадурын А., Архангельская Е.Н. Глубокое обучение. — СПб.: Питер, 2018. — 480 с.: ил. — (Серия «Библиотека программиста»).