

Модель

Вероятность подключения услуги

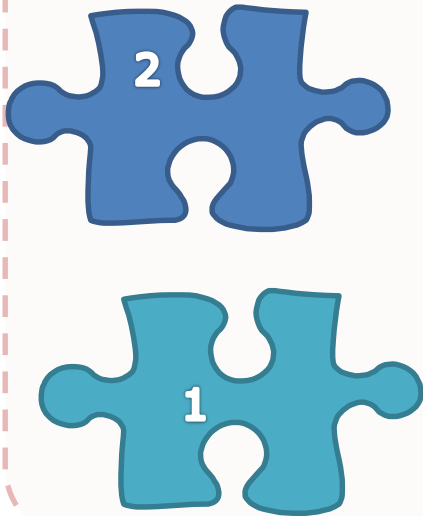
Задача

	Тренировочный сет	Тестовый сет	Профили пользователей
Время	4 месяца	1 месяц	-
Размер	831 653	71 231	4 512 528
Пропуски	Нет пропусков	Нет пропусков	Нет пропусков
Дубли	Строки с одинаковым id, но разными vas_id	Строки с одинаковым id, но разными vas_id	Строки с одинаковым id, но разной датой
Излишек			Строки с id, которых нет в тренировочном и тестовом сетах.

Проблема

- На одного пользователя может быть несколько предложений услуг (как одной и той же услуги, так и разных), в один и тот же день и в разные.
- Даты профиля пользователя не равны датам предложения пользователя

Предложения



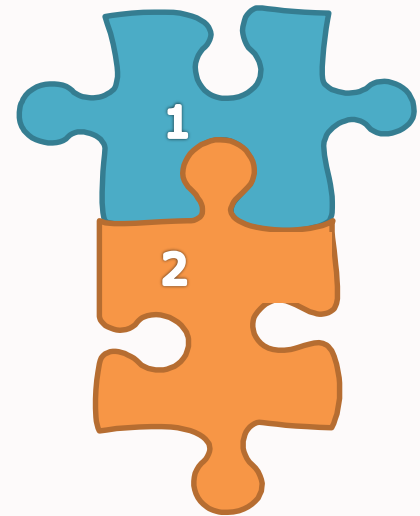
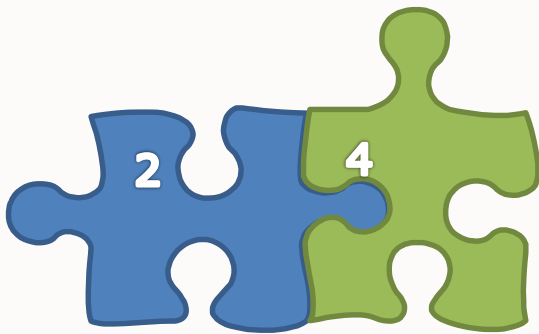
Профиль пользователя



Предположение

К предложению цепляется тот профиль, с которым минимальная разница по времени

Предложение услуги



Действия



Новые признаки

■ Даты

- ☐ Месяц
- ☐ День
- ☐ День года
- ☐ Неделя года
- ☐ День перед праздником
- ☐ День перед новым годом
- ☐ День после нового года
- ☐ Праздник

■ История

- ☐ Количественная история предложений для клиента
- ☐ Номер повторного предложения для клиента
- ☐ Разница по датам между повторными предложениями для одного клиента
- ☐ Разница по датам между всеми предложениями одного клиента

■ Группы (Kmeans)

- ☐ 15 (по числовым признакам)
- ☐ 7 (по числовым признакам)
- ☐ 3 (по числовым признакам)

Модели

XGBoost

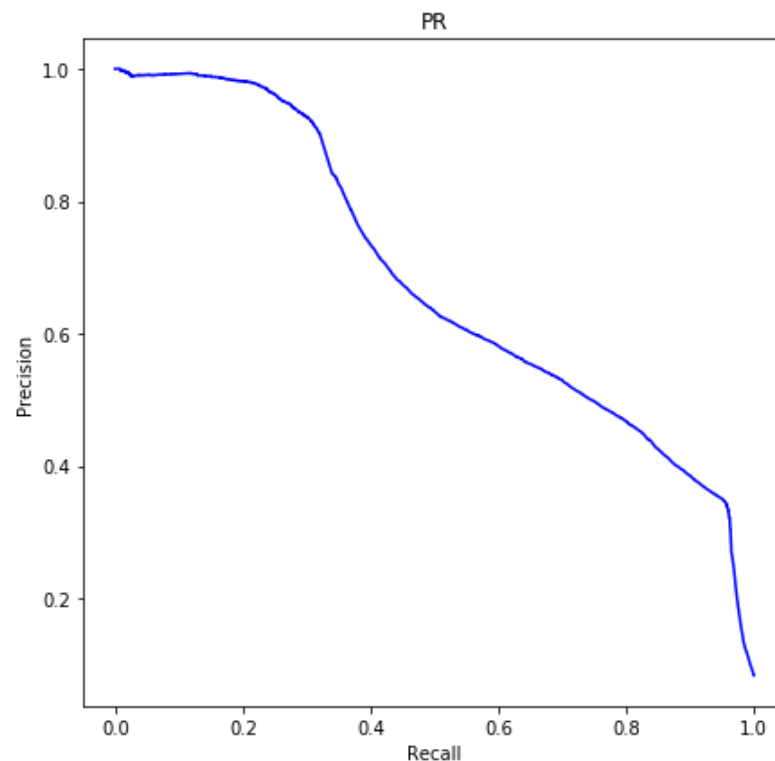
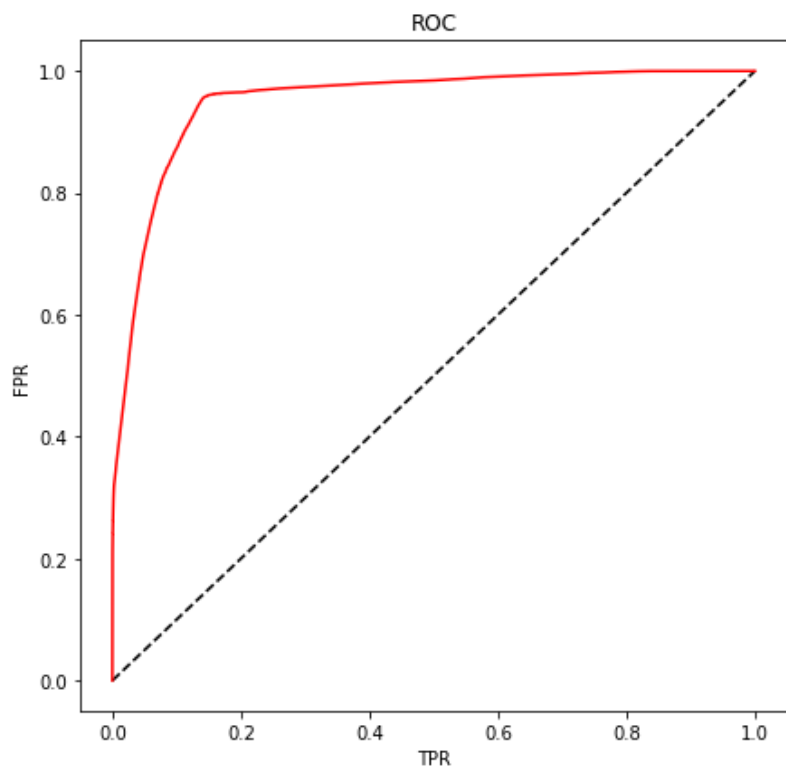
- max_depth=5
- n_estimators=50
- learning_rate=0.1
- subsample=0.5
- min_child_weight = 1
- reg_alpha=0.5
- reg_lambda=0.5

LightGBM

- num_leaves=8
- max_depth=4
- learning_rate = 0.1
- n_estimators = 100

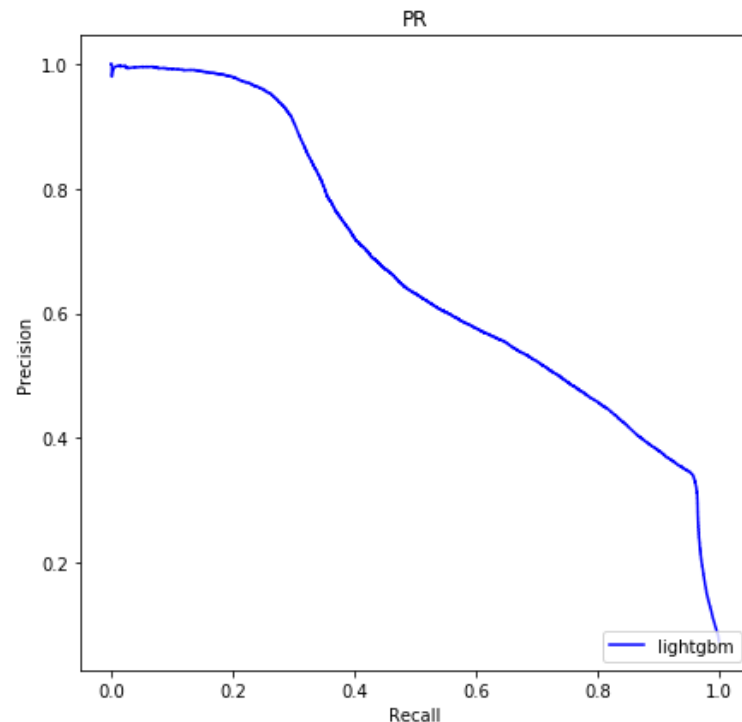
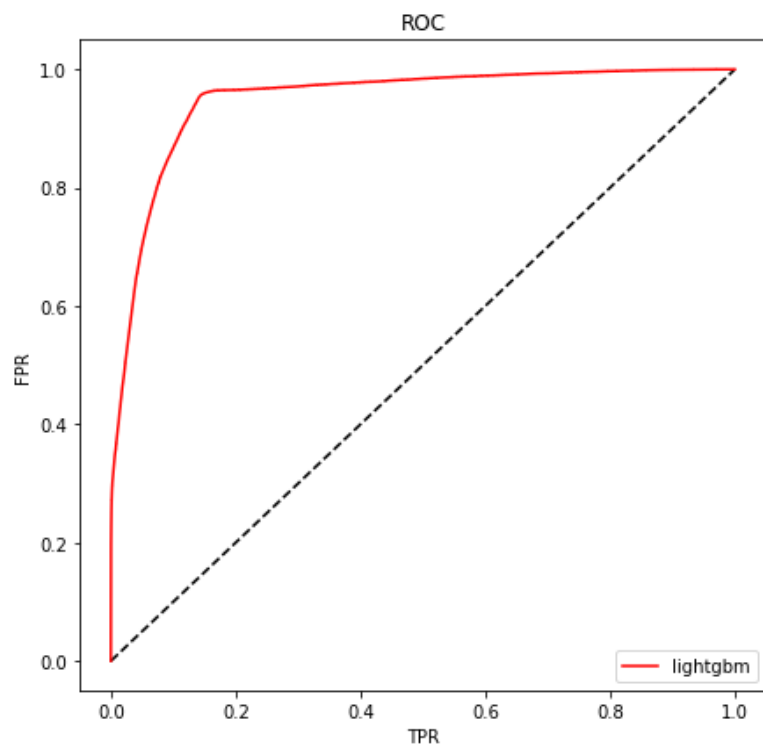
Результаты работы модели XGBoost

	ROC AUC	PR AUC	Precision	Recall	F1
XGBoost	0.951	0.682	0.52	0.71	0.601



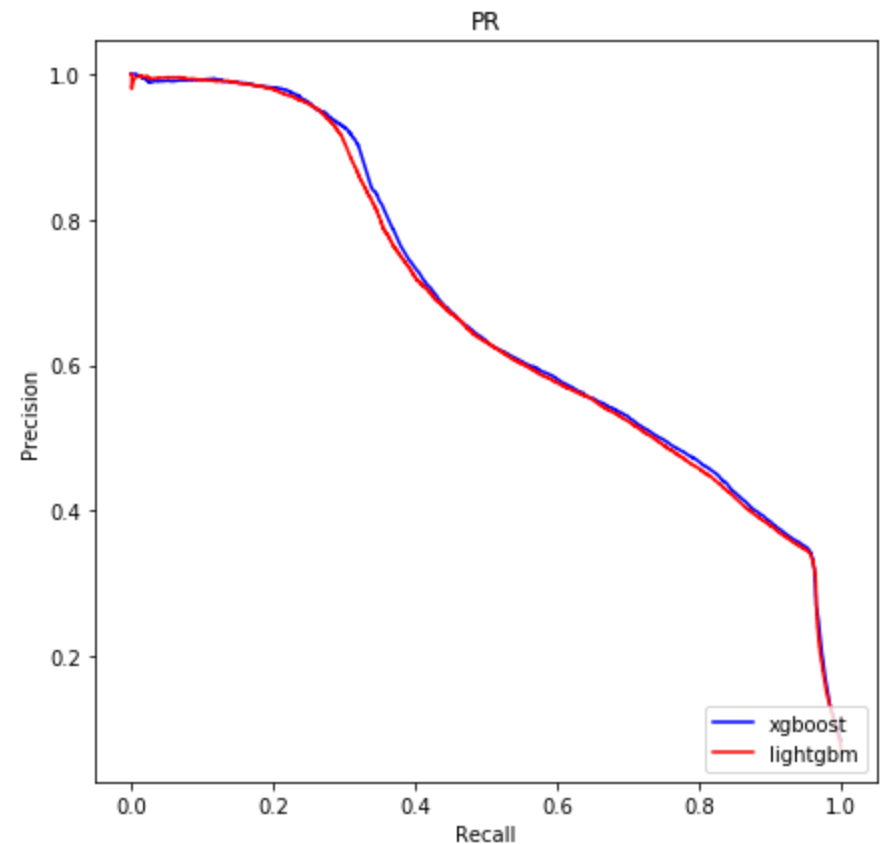
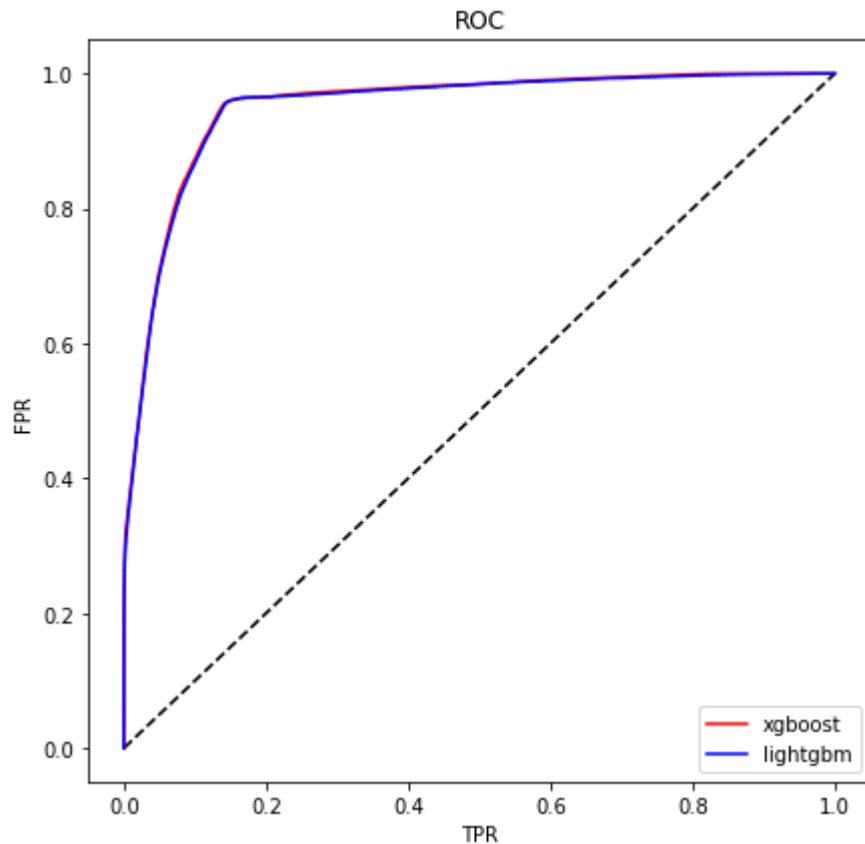
Результаты работы модели LightGBM

	ROC AUC	PR AUC	Precision	Recall	F1
LightGBM	0.949	0.676	0.54	0.67	0.598



Сравнение моделей

	ROC AUC	PR AUC	Precision	Recall	F1
XGBoost	0.951	0.682	0.52	0.71	0.601
LightGBM	0.949	0.676	0.54	0.67	0.598



Итоговый выбор: XGBoost

XGBoost

- Ниже Precision
- Выше Recall
- Чуть выше F1
- Чуть выше ROC AUC

LightGBM

- Выше Precision
- Ниже Recall
- Чуть ниже F1
- Чуть ниже ROC AUC