

Расширяя круг

Разработка интеллектуального алгоритма для привлечения родственников текущих клиентов Альфа-Банка



Построение look-alike модели позволит оптимизировать маркетинговую стратегию и сформировать индивидуальные коммуникации для целевой аудитории

Таргет

1

если клиент оформил карту на другого человека

0

если клиент оформил карту только для себя

в течение следующего месяца после даты сбора фичей

Источники внешних данных

- ✓ Социальные сети
- ✓ Сервисы с семейными подписками
- ✓ Авиакомпании
- ✓ Сотовые операторы
- ✓ Региональные данные

Алгоритм построения финальной модели

Очистка данных

CatBoostEncoder + OneHotEncoder

ADASYN + UPSAMPLING

MinMaxScaler

VotingClassifier
(LGBM+XGB+CatBoost Classifiers)

Качество на тестовой выборке

F2
0,521

F1
0,371

Gini
0,747

5 кластеров целевой аудитории

Формирование ценностных предложений для сегментов

Определение эффективности подобранных коммуникаций во время пилота для сегментов целевой аудитории

A/B (/C) тестирование
Многорукие бандиты
Uplift-моделирование

Модель отклика на коммуникацию

Модель оттока

Только 6,5% клиентов когда-либо оформляли карту Альфа-Банка на другого человека, остальные пользуются продуктами Банка индивидуально

Таргет

1

если клиент оформил карту на другого человека

0

если клиент оформил карту только для себя

в течение следующего месяца после даты сбора фичей



Какие клиенты чаще других оформляют карты другим людям?

Мужчины

62% > 50%

Класс 1

Класс 0

Зарплатные клиенты

98% > 75%

Класс 1

Класс 0

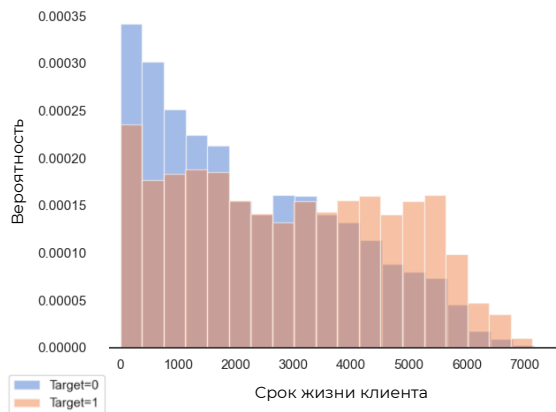
Активные клиенты

92% > 80%

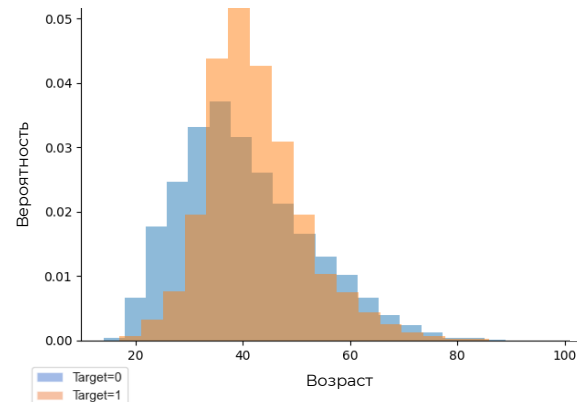
Класс 1

Класс 0

Клиенты, которые дольше пользуются услугами Альфа-Банка



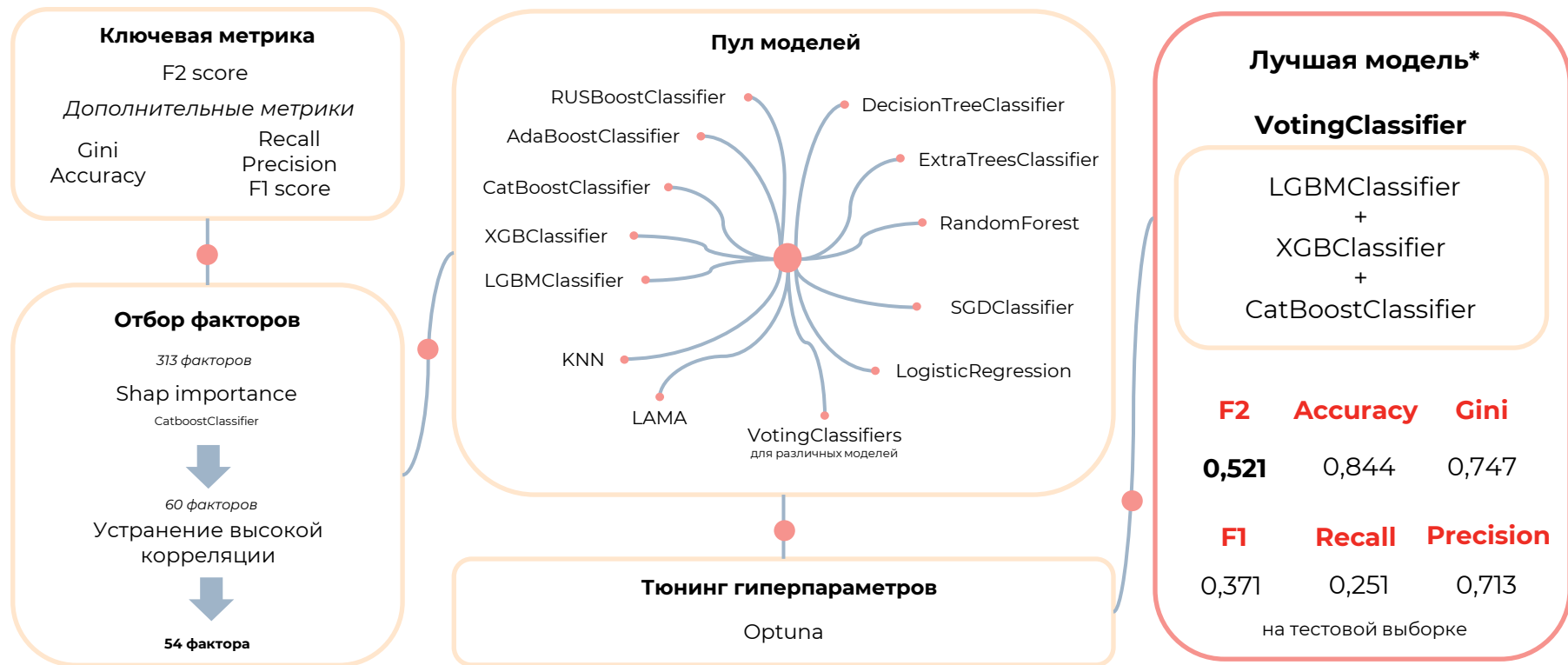
Клиенты старшего возраста



В результате предобработки данные были очищены, масштабированы и сбалансированы



Лучший результат по F2 метрике показал VotingClassifier на основе классических бустингов с 54 факторами



С помощью кластеризации Kmeans было выделено 5 сегментов целевой аудитории для формирования персональных коммуникаций



Активный мужчина, много зарабатывает и тратит, пользуется пакетом услуг

Акцент на удобство, скорость и функциональность использования семейного аккаунта



Представительницы активные и много тратят, склонны тратить в минус

Акцент на то, чтобы держать под контролем семейные доходы и расходы: статистика, советы



Среднестатистические мужчины, склонны тратить в минус

Акцент на систему отслеживания и накопление общего бюджета



Мужчины со средним доходом, чаще всех покупают фаст-фуд, меньше других покупают спорт. одежду и оплачивают в супермаркетах

Мотивация на то, чтобы копить сбережения на семейном счете: вместе копите, оплачивайте товары для дома, поездки и любые общие покупки
Делать акцент на выгоды



Женщины со средним доходом, покупают фастфуд, иногда покупают спорт. одежду и оплачивают в супермаркетах

приоритет

Каналы коммуникации

Баннеры в приложении
Push-уведомления

Предложение о продукте в чат-боте

Спонсорство мероприятий семейной тематики
(8 июля 2023 - День семьи)

Основные ценностные предложения*

Кэшбек и всем членам семьи
Повышенный кэшбек по 3 любимым категориям для всей семьи
Обслуживание от 0 руб.

Шагами по развитию проекта являются обогащение данных, проведение пилота, определение эффективных коммуникаций и построение новых моделей

Возможные источники внешних данных для построения модели

Социальные сети

VK, Одноклассники и др.:
Семейный статус
Фото для определения наличия семьи

Сервисы с семейными подписками

Яндекс
Онлайн-кинотеатры (ОККО и др.)
Сотовые операторы (MTC Premium и др.)
Стриминговые сервисы (Apple Music, YouTube Premium)

Авиакомпании

Информация о количестве пассажиров в одной регистрации,
приобретении дополнительных услуг и др.

Сотовые операторы

Платежи по счету
Пополнение счетов других абонентов
Подписки, услуги

Региональные данные

Флаг региона проживания клиента
Уровень фин. грамотности в регионе
Средний уровень доходов в регионе и т.д.

Проведение пилота

~ 6 мес.

Сбор таргета по откликам на различные
типы коммуникаций

Подходы к определению эффективных коммуникаций для сегментов

A/B (/C) тестирование

По окончании пилота получим
интерпретируемые статистически значимые
данные для выбора коммуникации

Многорукие бандиты

Максимизация прибыли за счет выбора
оптимальной коммуникации уже в ходе пилота

Uplift-моделирование

Оценка чистого эффекта от коммуникации,
если продукт будет достаточно популярным

Мониторинг качества модели и характеристик целевой аудитории

Модели отклика на различные типы коммуникаций

Построение модели с корректной разметкой
для повышения эффективности
маркетинговых коммуникаций

Модель оттока пользователей для продукта Семейный счет

Поиск клиентов, склонных к оттоку, для
проведения дополнительной коммуникации

Параллель 55



**Ивершинь
Анастасия**

Team Lead

НГУ. Экономический факультет

ivershin_anastasia@mail.ru

+7 983 123 7442



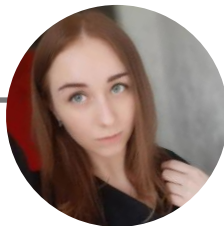
**Адель
Чернятов**

Data Scientist

КФУ. Прикладная математика

adelworkspace@mail.ru

+7 937 488 7147



**Сибгатуллина
Инна**

Data Scientist

Росбиотех. Инженерия и
технология пищевых производств

ophelials12@gmail.com

+7 968 719 4080



**Бредихин
Арсентий**

Data Scientist

ЮГУ. Институт цифровой
экономики

bredihin.igorr@yandex.ru

+7 951 974 1803



**Полковникова
Елизавета**

Data Scientist

Финансовый Университет при
Правительстве РФ. Маркетинг
GeekBrains. Data Science курс

elizavetapolk@gmail.com

+7 960 141 7722



Приложение 1. Сравнение результатов предобработки датасета на CatBoostClassifier

Предобработка	F2	F1	Precision	Recall	Accuracy	Gini
Upsampling + MinMaxScaler	0.314	0.383	0.280	0.605	0.874	0.721
Upsampling + StandardScaler	0.320	0.390	0.286	0.611	0.876	0.722
SMOTE + MinMaxScaler	0.243	0.311	0.586	0.212	0.939	0.728
SMOTE + StandardScaler	0.245	0.312	0.580	0.214	0.939	0.726
SMOTE + Upsampling + MinMaxScaler	0.314	0.383	0.280	0.605	0.874	0.718
SMOTE + Upsampling + StandardScaler	0.310	0.378	0.277	0.597	0.873	0.718
ADASYN + Upsampling + MinMaxScaler	0.348	0.402	0.316	0.551	0.894	0.726
ADASYN + Upsampling + StandardScaler	0.347	0.402	0.318	0.547	0.895	0.725
ADASYN + MinMaxScaler	0.245	0.314	0.600	0.213	0.940	0.728
ADASYN + StandardScaler	0.245	0.314	0.599	0.213	0.940	0.729

Приложение 2. Сравнение результатов моделей

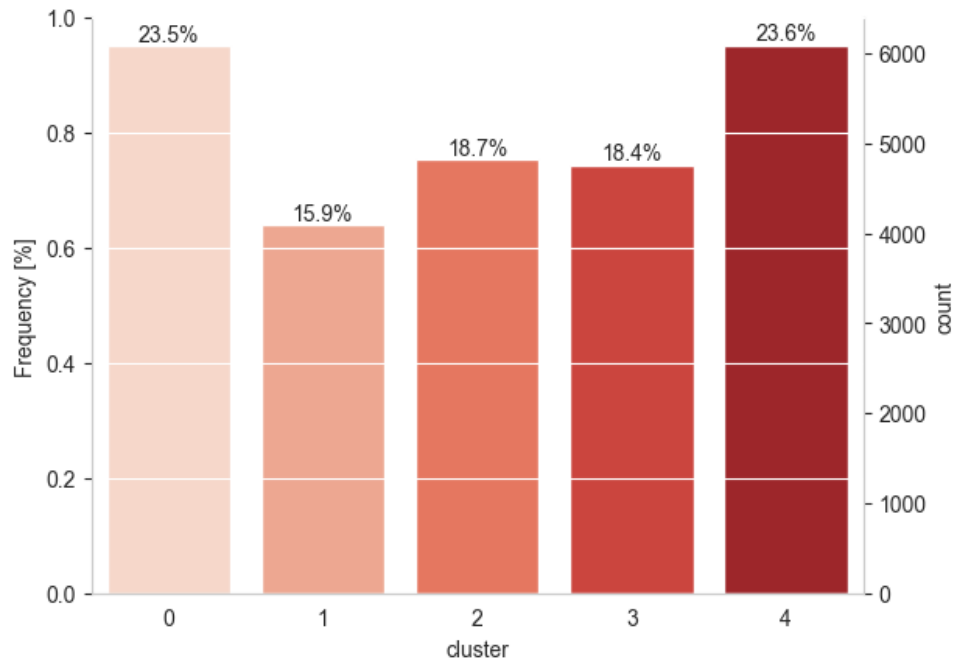
Модель	F2	F1	Precision	Recall	Accuracy	Gini
LogisticRegression	0,446	0,312	0,208	0,624	0,822	0,637
SGDClassifier	0,257	0,121	0,065	0,998	0,065	0,349
KNN	0,434	0,300	0,198	0,618	0,813	0,625
DecisionTreeClassifier	0,466	0,325	0,216	0,654	0,825	0,614
ExtraTreesClassifier	0,474	0,348	0,241	0,626	0,848	0,697
RandomForest	0,299	0,149	0,081	0,910	0,328	0,356
RUSBoostClassifier	0,500	0,360	0,245	0,675	0,845	0,723
AdaBoostClassifier	0,498	0,356	0,241	0,680	0,841	0,716
CatboostClassifier	0,430	0,336	0,246	0,528	0,602	0,602
XGBClassifier	0,512	0,366	0,248	0,698	0,844	0,741
LGBMClassifier	0,516	0,372	0,253	0,696	0,848	0,745
Lama	0,501	0,317	0,196	0,816	0,772	0,746
VC: LR+RF+LGBM	0,453	0,404	0,342	0,494	0,906	0,728
VC: CAT+XGB+LGBM	0,521	0,371	0,251	0,713	0,844	0,747

Приложение 3. Описание и ценностные предложения сегментам

Кластер 0. 1 приоритет - Самый прибыльный, активный сегмент. Мужчины Пользуются пакетом услуг - Сумма переводов больше, чем у остальных, чаще переводят по номеру телефона - Осуществляют внешние переводы в большей сумме - Чаще всех покупают в супермаркетах, фастфуд, спорт. одежду - Средний баланс текущих счетов и сумма ежедневных транзакций выше, чем у остальных сегментов - Меньше всех должны за кредит	Кластер 1. 5 приоритет - Менее всех прибыльный и активный. Женщины - Относятся к сегменту MAFFLUENT - Фастфуд покупают активно - Иногда покупают спорт. одежду и оплачивают в супермаркетах - После 1го сегмента меньше всех должны	Кластер 2. 4 приоритет Мужчины - Относятся к сегменту MAFFLUENT Ключевое: - Чаще всех покупают фаст-фуд - Меньше всех оплачивают в супермаркетах - Меньше всех покупают спорт. одежду Остальные признаки - на среднем уровне	Кластер 3. 3 приоритет Мужчины Средние показатели по всем признакам Ключевое: Фастфуд покупают активно Больше всех должны после 4-го кластера	Кластер 4. 2 приоритет - Второй по прибыльности и активности сегмент. Женщины - Часть из которых пользуются пакетом услуг - Вторые по всем признакам после 0-го сегмента. Ключевые различия: - Больше всех должны - Не так часто покупают фастфуд
В коммуникации делать акцент на удобство и функциональность использования семейного аккаунта, эффективность по времени. - Все необходимые контакты для переводов будут под рукой. - Возможность настроить автоматические переводы - Предложения о благотворительности - Контроль над финансовыми операциями членов семьи - Удобство оплаты детских потребностей	Делать акцент на выгоды использования сем. аккаунта. Мотивация на то, чтобы копить сбережения на семейном счете. Вместе копите, оплачивайте товары для дома, поездки и любые общие покупки. - Кэшбек с оплаты за ЖКХ - Контроль бюджета ребенка - Общий тариф: доступ к бонусам всех членов семьи, без переплат - Кэшбек по аналогии X5 - Возможность копить баллы и тратить их на товары для семьи, дома, детей	Опция для всей семьи по выгодной цене. Доступ к акциям, закрытым распродажам и мероприятиям Большое количество сервисов, партнеров - Возможность выбрать категории повышенного кэшбека Детям студентам бонусы	Склонны тратить в минус. - Акцент на систему отслеживания и накопление общего бюджета - Уведомления о тратах и пополнениях всех членов семьи. - Статистика, какие категории трат занимают главные места. - Общий тариф на всю семью выгоднее - Доход с накоплений больше (т.к. накопления общие) - Кэшбек за покупки со счета - Бесплатный перевод с карты на карту и на моб. телефон внутри семейного аккаунта	Представительница активные и много тратят, склонны тратить в минус. - Обучение пользователя и ребенка фин. грамотности для того, чтобы уметь беречь деньги, ставить цели и достигать - Акцент на то, чтобы держать под контролем доходы и расходы: статистика, советы - Пункты, аналогичные предложениям 0-го кластера

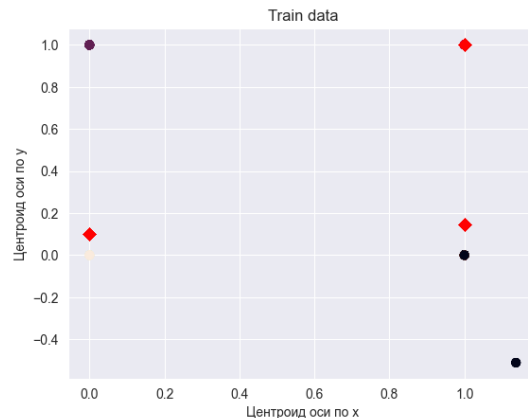
Приложение 4. Характеристики кластеров (сегментов) целевой аудитории

Распределение наблюдений между кластерами



Факторы для кластеризации

Пол
Флаг, Причина попадания в сегмент - Пакет услуг
Средняя сумма электронных операций в категории - Перевод по номеру телефона за месяц на протяжении года
Средняя сумма электронных операций в категории - Внешний перевод RUR за месяц на протяжении года
Сумма транзакций за 90 дней в категории - ФАСТФУД
Сумма транзакций за 90 дней в категории - СПОРТИВНАЯ ОДЕЖДА
Средняя дневная сумма транзакций за 90 дней
Средний баланс текущих счетов за последние 3 месяца
Сумма транзакций в категории Супермаркеты за 2 месяца
Разница между средним кредитным и дебетовым оборотом по текущим счетам

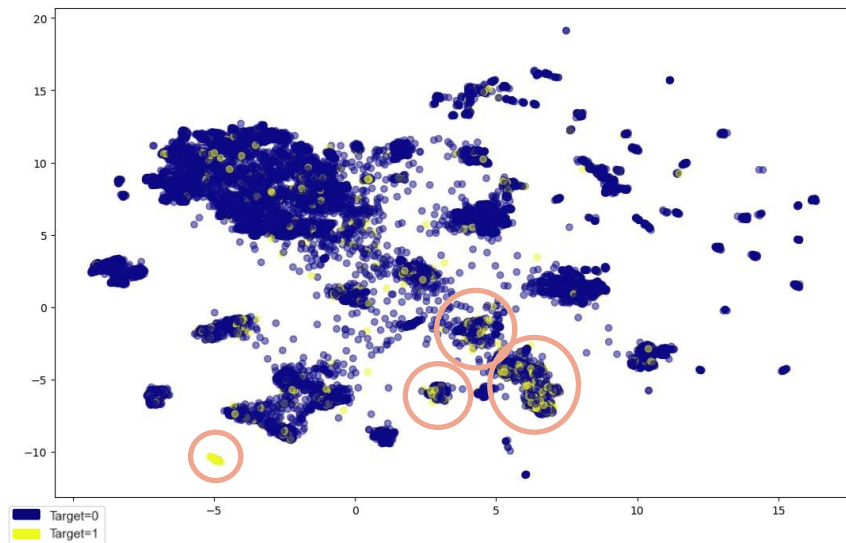


Приложение 5. Пример визуализации распределения скоров по признакам между кластерами



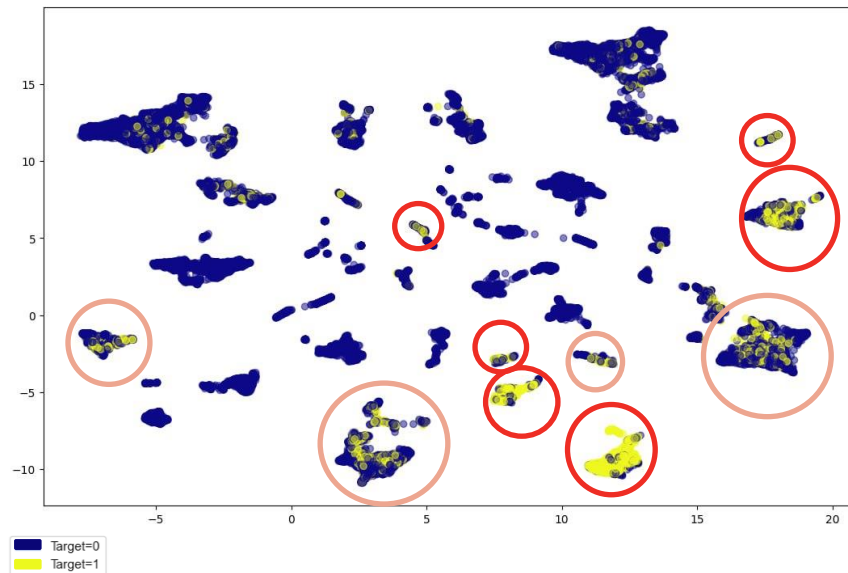
Приложение 6. Сегменты для таргета и предиктов на тестовой выборке (UMAP алгоритм)

Кластеры для таргета



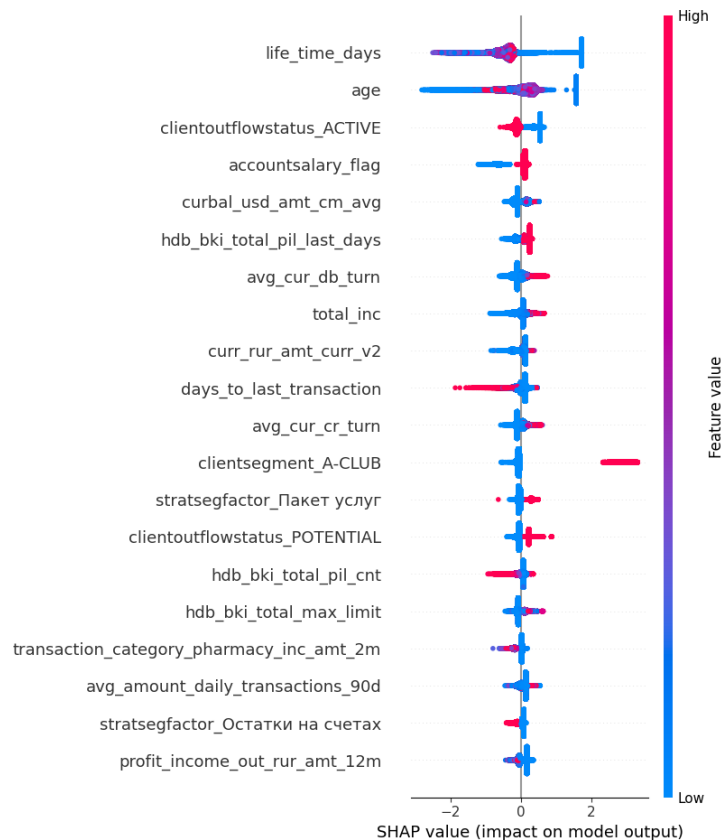
Определяются потенциальные группы похожих наблюдений, в которых части наблюдений присвоен таргет 1, остальные наблюдения в этих группах – потенциально тоже единички

Кластеры для предиктов



Модель смогла выявить разреженные группы клиентов, потенциально склонных к оформлению карты другому человеку

Приложение 7. Влияние наиболее важных фичей на score модели



Наиболее важные фичи

1. Время жизни клиента
2. Возраст
3. Активный статус клиента
4. Статус зарплатного клиента
5. Средний баланс текущих счетов за последние 3 месяца
6. Количество дней с последнего оформленного кредитного продукта типа Кредит
7. Средний кредитный оборот по текущим счетам
8. Общий среднемесячный доход
9. Сумма средств клиента на группе счетов Основные текущие счета на дату расчета
10. Количество дней до последней транзакции