

№	Содержание	Вопросы по кейсу
Кейс № 1	<p>Банку необходимо эффективно управлять имеющейся краткосрочной ликвидностью на НОСТРО счете в ЦБ.</p> <p>Если совсем просто, то можно сформулировать следующие условия:</p> <p>1) Потоки на этом счете состоят из платежей и зачислений клиентов Банка – Физических и Юридических лиц (например, вы переводите деньги из зеленого банка в желтый – происходит списание, вы снимаете деньги с Банкомата – происходит списание, вы платите налоги или совершаете платеж государству – происходит списание. И наоборот – нашему клиенту юр. лицу совершают платеж – происходит зачисление и т.д.).</p> <p>2) Нам нужно построить модель, которая будет прогнозировать сумму всех поступлений и списаний на конец следующего дня, причем делать это полностью автоматически.</p> <p>3) По результатам прогноза, мы можем целый день крутить потенциальные поступления на фин. рынках и получим примерно ключевую ставку.</p> <p>4) Если мы перепрогнозим – в конце дня размер перепрогноза приходится занимать у других банков под ставку, выше ключевой.</p> <p>5) Если недопрогнозим – в конце дня мы можем сами одолжить эту сумму, но по ставке чуть ниже ключевой.</p>	<p>Вопросы:</p> <ul style="list-style-type: none"> ● Как бы вы сформировали признаковое пространство? ● Как бы вы выбирали метрику качества? ● Какую предобработку данных вы бы сделали? ● Как бы вы выбирали значимые факторы? ● Какие алгоритмы вы бы попробовали применить? ● Что необходимо учесть, учитывая, что модель будет применяться автоматически (без участия разработчика, пользователь будет видеть только число - прогноз)?
Кейс № 2	<p>Требуется создать алгоритм машинного обучения для прогнозирования количества ядер CPU кластера на сутки вперед с дискретизацией 10 мин (т.е. 144 шага вперед).</p> <p>Какие данные есть: Есть данные по количеству использованных ядер CPU кластера с дискретизацией 1 мин. За 1 год назад, в данных бывают пропуски но не более 5 мин.(5 точек в данных), как подряд так и в разнбой.</p> <p>Что нам известно: Нагрузка представляет из себя временной ряд без какого-либо экзогенного влияния, без «разладок».</p>	<p>Вопросы:</p> <ul style="list-style-type: none"> ● Как будете анализировать ряд, какие его свойства нам могут быть интересны? ● Какие предикторы(фичи) можно сделать из ряда и какие Таргеты? ● Каким образом будете разбивать выборку на Test и Train? ● Какие алгоритмы можно использовать? Их преимущества и недостатки по вашему мнению. ● Какую метрику выберете для оценки качества и почему? <p><i>Опционально (Оценку не понижает, но может повысить)</i></p> <ul style="list-style-type: none"> ● Какие нейросетевые алгоритмы по вашему мнению можно использовать и какие модификации ряда для этого надо сделать? ● Какие модификации вы бы сделали, если я скажу, что хочу прогнозировать 95%

		квантиль числа необходимых ядер, а не фактическое потребление?
Кейс № 3	<p>Банк принял решение снизить затраты на опросы удовлетворенности клиента. Для этого решено построить модель, которая предсказывает удовлетворенность.</p> <p>В наличии: e-mail опросы клиентов с марта 2022 года (примерно 150 ответов респондентов в месяц по 5ти бальной шкале, где 60% оценок 5, остальные оценки распределены равномерно).</p> <p>Данные по звездам (примерно 10000 кликов на звезды, появляющиеся после оформления продуктов в мобильном приложении, 5ти бальная шкала, где 95% оценок 5).</p>	<p>Вопросы:</p> <ul style="list-style-type: none"> • Что будет являться целевой переменной для данной задачи? Достаточно ли измерений в опросах и является ли их распределение проблемой? • Как будет решаться задача? • Какие данные будете использовать для предсказания? • Какие алгоритмы подойдут для задачи?
Кейс № 4	<p>Рекомендательная система для магазина приложений</p> <p>Представьте, что для магазина приложений на смартфоны вам необходимо обучить модель, которая бы предсказывала, какое приложение хочет установить пользователь, когда он открывает ваш магазин.</p>	<p>Вопросы:</p> <ul style="list-style-type: none"> • Как решать такую задачу? • Какие данные необходимо собрать? • Что делать, если о пользователе нет информации (к примеру, он только что купил смартфон)? • Как измерять качество такой модели? • Как можно улучшить качество ее предсказания?
Кейс № 5	<p>Юридические лица пользуются услугами банка часть из которых не имеет формальной даты окончания договора. Вам поставили задачу предсказать для оттока клиентов для одного из таких продуктов. В данном случае отток - прекращение юридическим лицом использования продукта. Стандартно у каждой услуги есть, как минимум, количественный (штуки) и объемный (рубли) показатели, распределенные во времени (по месяцам).</p>	<p>Вопросы:</p> <ul style="list-style-type: none"> • Формализуйте задачу. • С чего вы начнете решение задачи? • Какие данные Вам понадобятся? • На что будете обращать внимание при сборе и подготовке данных? • Какие недостатки данных вам могут встретиться и как бы Вы их исправили? • Какие методы машинного обучения будете использовать, а какие нет? • Какие метрики будете использовать для оценки качества? • Какую вспомогательную модель / модели Вы могли построить, и чем они Вам помогут? • Что изменится, если вы будете решать аналогичную задачу для пользователей Интернет-приложения онлайн-игры?
Кейс № 6	<p>Представьте, что для банковского приложения для смартфона вам необходимо обучить модель, которая бы предсказывала цель запуска приложения (т.е. что хочет сделать клиент в приложении, какую операцию совершить или каким продуктом воспользоваться). На основании результатов этой модели предполагается наполнять часть экрана "быстрыми" кнопками.</p>	<p>Вопросы:</p> <ul style="list-style-type: none"> • Формализуйте задачу. • Какие данные необходимо собрать? • На что будете обращать внимание при сборе и подготовке данных? • Что делать, если о пользователе нет информации (к примеру, это новый клиент)? • Какие методы машинного обучения будете использовать?

		<ul style="list-style-type: none"> • Какие метрики будете использовать для оценки качества? • Как измерять качество такой модели? • Как можно улучшить качество ее предсказания? • Если Вам для решения этой задачи предложат создать ансамбль моделей, то какие модели Вы бы построили дополнительно?
<p>Кейс № 7</p>	<p>Вы работаете с маркетинговым отделом банка. Рекламная и пиар-службы хотят понимать, как банк обсуждается в соц. сетях - в каких контекстах, насколько активно, и т.д. Эта информация поможет им проводить более эффективные рекламные кампании и оперативно реагировать на проблемы.</p> <p>В вашем распоряжении поток данных с сообщениями, в которых упоминается название банка (~150к сообщений в день); предполагается, что он репрезентативен относительно всего обсуждения в соц. сетях и достаточен по содержанию (других источников данных не требуется).</p> <p>Перед вами стоит задача создать пайплайн сюжетирования, позволяющий из всего потока ежедневно выделять 10-30 тем, наиболее активно осядающихся пользователями.</p>	<p>Вопросы:</p> <ul style="list-style-type: none"> • Какие дополнительные вопросы следует задать маркетологам? • Из каких этапов состоит пайплайн? • Какие алгоритмы/инструменты могли бы входить в пайплайн? • Как оценить качество пайплайна?
<p>Кейс № 8</p>	<p>Перед Вами стоит задача формирования пула клиентов ММБ с которыми необходимо провести какие-либо взаимодействия (тип взаимодействия за периметром кейса), чтобы клиент склонный к оттоку продолжил обслуживаться в банке.</p> <p>У вас уже есть: ежемесячная разметка всего портфеля клиентов по статусу на конец месяца (статусы: лояльный, прекратил транзакции, закрыл р/с, заблокирован по требованиям безопасности, бизнес ликвидирован).</p>	<p>Вопросы:</p> <ul style="list-style-type: none"> • Каким образом можно формализовать задачу? • Какие данные могут быть полезны для решения этой задачи? • Как Вы будете собирать разметку для обучения модели? • Какие промежуточные исследования будете проводить? • Какие метрики качества модели применимы для данной задачи? • Как оценить финансовый результат использования данной модели? • Какие проблемы могут возникнуть при моделировании?
<p>Кейс № 9</p>	<p>Представьте, что у вас есть лидарное облако с трассы – набор точек в 3-х мерном пространстве, полученный лидаром с машины. Также у вас есть информация о движении машины – координаты машины каждые 100 миллисекунд (то есть 10 измерений в секунду). Перед вами две задачи – найти в этом облаке отбойники и баннеры/дорожные знаки. Отбойники должны быть аппроксимированы непрерывной линией (набор отрезков), а баннеры/знаки – краями знаков.</p>	<p>Вопросы:</p> <ul style="list-style-type: none"> • Понадобятся ли дополнительные данные для решения задачи и если да, то какие? • Понадобятся ли алгоритмы машинного обучения и если да, то какие? • Как будете бороться с выбросами? С «дырками» в полосе отбойников? • Какие краевые случаи (конкретные сцены на дороге) стоит рассмотреть? Что с ними можно сделать? • Допустим, у вас есть возможность использовать уже готовые модели

		<p>детекции различных объектов в виде 3D боксов (прямоугольнички). Какие модели здесь могут пригодиться и почему?</p> <ul style="list-style-type: none"> ● Облаков точек с одного места у нас много (было совершенно несколько проездов с лидаром). Как будете объединять найденные на разных проездах отбойники и баннеры/знаки? <p><i>Опционально (Оценку не понижает, но может повысить)</i></p> <ul style="list-style-type: none"> ● У нас есть также фото с разных камер с машины, 10 фото в секунду. Теперь мы хотим также получать максимально качественное изображение с баннера/знака. Как будете решать задачу? ● Как распараллелить ваш алгоритм максимально эффективно?
<p>Кейс № 10</p>	<p>При подаче заявки на кредит в банк в обязательном порядке необходимо указывать свой доход. Ранее обязательно было предоставлять справку 2-НДФЛ для подтверждения уровня дохода. Банк движется в сторону цифровизации и ускорения процесса выдачи кредитов, поэтому хочет отказаться от строгой необходимости предоставлять справку 2-НДФЛ в момент подачи заявки. Знание о доходе клиента крайне важно для оценки кредитоспособности клиента, в связи с чем возникает необходимость в модельной оценке дохода заемщиков.</p>	<p>Вопросы:</p> <ul style="list-style-type: none"> ● Как бы ты решал данную задачу? ● Какой таргет выбрать для моделирования дохода? ● Какие данные можно использовать для решения данной задачи? ● Как оценить качество полученной модели?