



Отчёт о выполненных этапах разработки  
клиент-маркетинговой системы поддержки  
принятия решений (вероятность покупки клиента)

4 августа 2021 г.

<b>СОДЕРЖАНИЕ</b>	<b>1</b>
<b>1. Цели и задачи.</b>	<b>2</b>
<b>2. Процессная схема этапов разработки</b>	<b>2</b>
2.1. Общая схема.	2
2.2. Выполненные работы за предыдущие отчетные периоды.	3
<b>3. Выполненные работы за текущий отчетный период.</b>	<b>4</b>
3.1. Общее описание.	4
3.2. Система предикторов.	4
3.2. Опытный прототип модели.	6
3.3. План A/B-теста.	8
<b>4. Заключение.</b>	<b>11</b>

## 1. Цели и задачи.

**Цель** - повышение эффективности процесса взаимодействия с клиентами.

Под эффективностью будем понимать отношение среднего дохода с продаж к средним затратам на взаимодействие с клиентом (в относительных единицах).

Для достижения выбранной цели поставлены следующие **задачи**:

1. Увеличение конверсии потенциальных покупателей в сделки.
2. Разработка маркетинговой стратегии взаимодействия с клиентами на основе предсказанной вероятности покупки.

## 2. Процессная схема этапов разработки СППР.

### 2.1. Общая схема.

Общая процессная схема разработки СППР представлена на рис. 2.1.

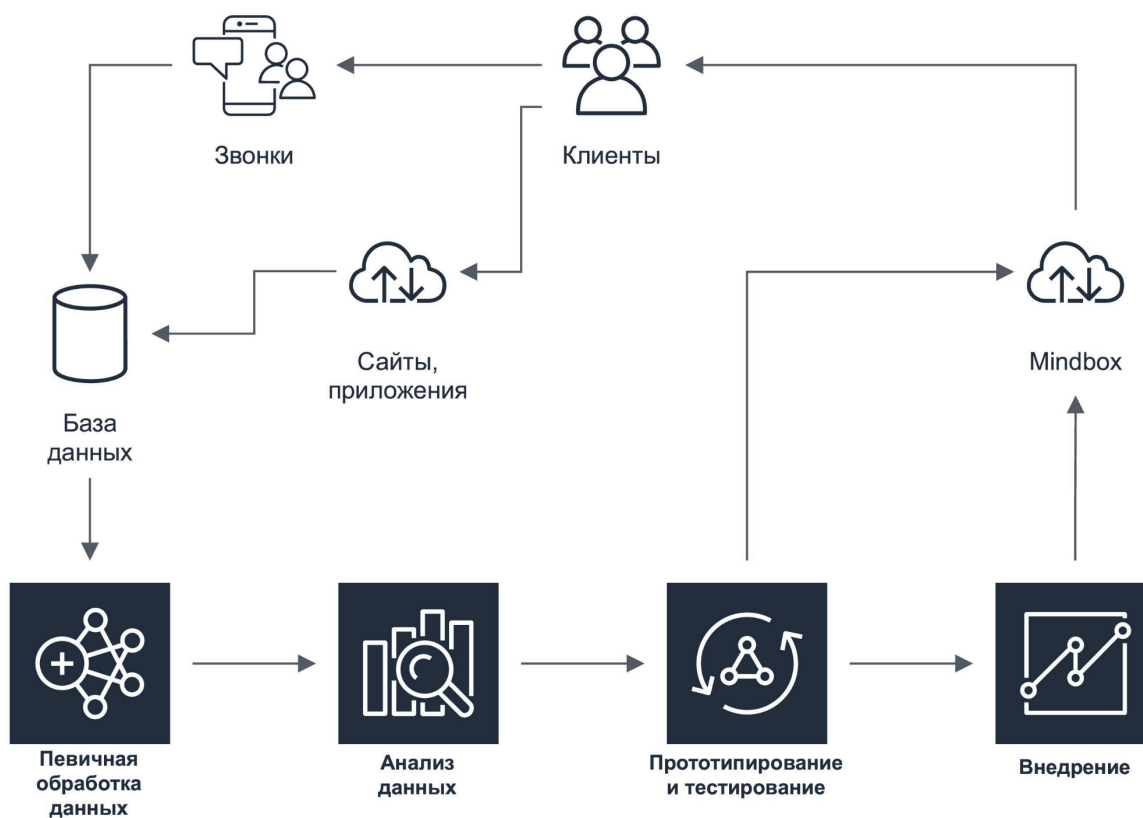


Рис. 2.1. Процессная схема разработки СППР.

### **Первичная обработка данных.**

На данном этапе производится очистка данных (выделяются данные, которые могут быть использованы в процессе машинного обучения) и создание датасетов, содержащих всю необходимую информацию.

### **Анализ данных.**

На данном этапе осуществляется подготовка предикторов (предсказывающих переменных) для последующего обучения алгоритмов, включающая выявление статистических взаимосвязей, анализ временных рядов, видов распределений, кластеризацию и выявление наиболее важных признаков.

### **Прототипирование и тестирование.**

На данном этапе выполняется циклический синтез и совершенствование алгоритмов с последующим тестированием как на тестовых данных, так и с проведением экспериментов в реальных условиях (A/B тесты).

### **Внедрение.**

После разработки алгоритмов и успешно проведённых экспериментальных исследований производится внедрение разработанных информационно-аналитических решений в бизнес-процесс (обычно в виде приложения или функционального модуля).

## **2.2. Выполненные работы за предыдущие отчетные периоды.**

За предшествующие отчетные периоды были выполнены следующие работы:

1. Произведена первичная обработка данных (исследованы структура и содержание БД Заказчика).
2. Собран уникальный профиль клиента.
3. Выполнена кластеризация действий клиентов.
4. Разработана первая версия системы предикторов будущей модели.

### 3. Выполненные работы за текущий отчетный период.

#### 3.1. Общее описание.

За текущий отчетный период была окончательно сформирована система предикторов для опытной модели, включающая наиболее важные факторы, влияющие на вероятность совершения сделки в течение 40 дней после заведения первого интереса.

Частично выполнен этап прототипирования и тестирования, - разработана опытная модель, определяющая вероятность совершения сделки клиентом на основе системы предикторов, разработан A/B-тест для пробного тестирования модели.

#### 3.2. Система предикторов.

Разработанная система предикторов для опытной модели представлена в таблице 1.

Таблица 1 - Модель системы предикторов (версия 2)

№	Наименование полей	Описание
1	mainId	уникальный идентификатор клиента
2	lead_split	первый интерес появился раньше/позже даты сокрытия встреч
3	dynamic_messages_client	открытие и клики клиента по сообщениям об изменениях цен и ставок - от lead до term_point
4	dynamic_messages_pik	отправка клиенту сообщений об изменениях цен и ставок от - от lead до term_point
5	contact_act	контактные действия клиента при выборе - от lead до term_point
6	adv_unsubscribe	действия клиента по отказу от рекламы, опросов, рассылки - от lead до term_point
7	flat_choice	действия клиента по выбору (просмотр, ...) - от lead до term_point

<b>№</b>	<b>Наименование полей</b>	<b>Описание</b>
8	mortgage_view	просмотр клиентом информации по..- от lead до term_point
9	adv_subscription	подписка клиентом на информирование и рекламу - от lead до term_point
10	not_delivered	недоставленные сообщения (в целом) - от lead до term_point
11	office_choice	действия клиента по выбору ...- от lead до term_point
12	adv_messages_client	открытие и клики клиента по рекламным сообщениям - от lead до term_point
13	offer_messages_client	открытие и клики клиента по сообщениям с выгодными предложениями и условиями - от lead до term_point
14	surv_messages_pik	отправка клиенту опросов - от lead до term_point
15	adv_messages_pik	отправка клиенту рекламных сообщений - от lead до term_point
16	surv_messages_client	открытие и клики клиента по сообщениям с опросами - от lead до term_point
17	quantity_calls	количество звонков - от lead до term_point
18	duration_calls	продолжительность всех звонков - от lead до term_point
19	waiting_connect	время ожидания по всем звонкам - от lead до term_point
20	appointment	факт записи на встречу, видеовстречу, экскурсию - от lead до term_point
21	appointment_done	число состоявшихся встреч - от lead до term_point
22	quantity_reserves	количество резервирований - от lead до term_point
23	background_life	разница в днях между датой самого раннего интереса и мин. датой первого интереса, превышающей дату первого действия
24	previous_life	разница в днях между датой первого действия и датой заведения первого интереса

Продолжение таблицы 1

№	Наименование полей	Описание
25	lead_split	метка 1/0 первый интерес появился раньше/позже даты сокрытия встреч (2020-04)
26	lead_days	число дней с момента заведения интереса до контрольной точки
27	lead_quantity	число интересов с момента заведения интереса до контрольной точки
28	LivingComplex_min_score	нижняя граница оценки
29	LivingComplex_mean_score	средняя оценка
30	LivingComplex_max_score	верхняя граница оценки

Условные сокращения:

lead - дата заведения первого интереса (в поле «Описание»),

term\_point - дата окончания условного периода при расчете значений предикторов (дата моделирования при практическом применении).

Целевая переменная (contract\_40) - факт совершения сделки от момента заведения первого интереса в течение 40 дней.

### 3.2. Опытный прототип модели.

В качестве опытного прототипа модели по предсказанию факта совершения сделки от момента заведения первого интереса в течение 40 дней была разработана модель типа RandomForest (Случайный лес).

Для оценки модели используются следующие метрики:

1. AUC (Area Under the Curve) - площадь, ограниченная ROC-кривой (Receiver Operating Characteristic) и осью FPR (False Positive Rate), см. рис. 3.1.

2. Recall (True Positive Rate) - доля истинно-положительных результатов классификации, характеризует способность алгоритма правильно определять целевой класс (чувствительность).

3. Precision (Positive Predictive Value) - доля истинно-положительных результатов классификации в отношении общего количества предсказанных положительных результатов, характеризует точность предсказания.

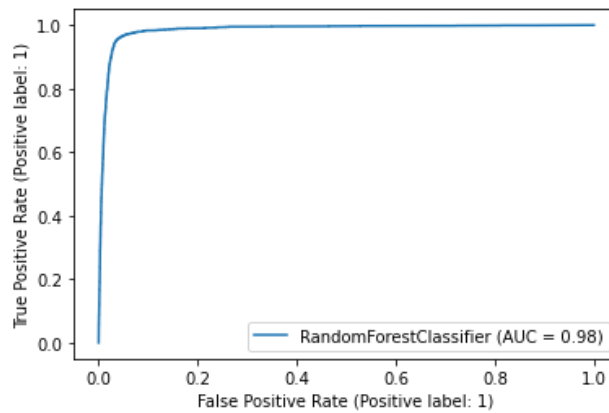


Рис. 3.1. ROC-кривая разработанной модели.

Показатели качества (метрики) модели:

AUC = 0,98,

Recall = 0,96,

Precision = 0,76.

Оптимальный порог по вероятности, обеспечивающий наилучшее соотношение Recall и FPR, составил 0.4726.

По результатам моделирования была оценена значимость предикторов, представленная на рис. 3.2.





Рис. 3.2. Значимость предикторов модели.

Здесь также следует отметить, что модель разработана с использованием ручной выгрузки данных из системы Mindbox по датам первых действий пользователей. Модель может быть существенно улучшена с привлечением дополнительной информации, включающей информацию о просмотрах, их площади, стоимости, а также аналогичную информацию..., привязанных к интересам клиентов, которая в настоящий момент не выгружается из системы Mindbox или выгружается некорректно и частично удалена из сервиса.

### 3.3. План А/В-теста.

А/В-тест - это эксперимент с двумя группами объектов (клиентов), отличающихся значениями одного или нескольких показателей, для определения изменений, улучшающих целевой показатель.

Целевой показатель - усредненный доход от продажи.

Проведение А/В-теста включает выделение тестовой и контрольной групп объектов, в нашем случае, клиентов компании. Разделяющим

показателем данных групп является распределение выгодных предложений на покупку.

Для тестирования разработанной модели в отношении контрольной группы применяется механизм стимулирования спроса, основанный на экспертном опыте сотрудников компании заказчика и существующих механизмах ВІ, внедренных в бизнес-процесс; в отношении тестовой - механизм стимулирования спроса на основе предсказанной вероятности совершения сделки и разработанных рекомендаций по взаимодействию с сегментами клиентов.

Нулевая гипотеза  $H_0$  заключается в том, что полученные результаты тестирования по двум группам клиентов принадлежат одной генеральной совокупности, т.е. их законы распределения вероятности равны:

$$H_0: F(Y_1) = G(Y_2),$$

где  $F(Y_1)$ ,  $G(Y_2)$  - законы распределения вероятностей целевого показателя контрольной  $Y_1$  и тестовой  $Y_2$  групп.

Гипотеза  $H_1$  соответственно:

$$H_1: F(Y_1) \neq G(Y_2).$$

Распределения  $F(Y_1)$  и  $G(Y_2)$  различны, если отличны их математические ожидания  $M[.]$  (средние значения выборок). Для определения статистически значимой разницы в значениях  $M[Y_1]$  и  $M[Y_2]$  используется  $t$ -статистика, выбранный уровень статистической значимости  $p = 0,95$  (означает, что с вероятностью 95% различия между целевыми показателями тестовой и контрольной групп не являются случайными). В соответствии с выбранным уровнем значимости размер каждой из выборок составляет 600 человек.

Условия проведения А/В-теста:

1. Отбор кандидатов в группы производится случайным образом из существующих клиентов, удовлетворяющих следующим условиям:

1.1. Клиент должен иметь хотя бы один интерес.

1.2. Число дней жизни клиента от момента заведения первого интереса до даты старта А/В-теста должно быть больше 0 и меньше 40.

1.3. У клиента не должно быть завершенных сделок к дате старта А/В-теста.

2. Для принятия решений по воздействию на испытуемые группы клиентов формируются условно две команды, каждая из команд может раздать ограниченное число выгодных предложений различного достоинства:

- с минимальной выгодой - 75% (450) клиентов в группе,
- со средней выгодой - 20% (120) клиентов в группе,
- с максимальной выгодой - 5% (30) клиентов в группе.

3. Срок действия предложений ограничен индивидуально для каждого клиента периодом от даты отправки (дата старта А/В-теста) до даты истечения 40 дней с момента заведения первого интереса.

4. Клиенты, участвующие в тестировании, не должны иметь более ранних предложений до даты старта А/В-теста.

План проведения А/В-теста:

1. Проинформировать команду Coffee\_Analytics о готовности к проведению А/В-теста не позднее, чем за пять рабочих дней до намеченной даты (необходимое время для возобновления работ и актуализации модели).

2. Выбор случайным образом двух групп клиентов (контрольной и тестовой) из удовлетворяющих условиям проведения теста.

3. Передача идентификаторов в Mindbox путем загрузки сформированного командой Coffee Analytics csv файла и создание на основе показателей из файла сегментов пользователей на основе значений тестовой и контрольной групп.

4. Для тестовой группы (сегмента) создание сообщений с выгодными предложениями на основе переданной с индикатором клиента

информации (тип выгодного предложения и количество дней). Для контрольной группы (сегмента) создание сообщений произвольным образом, сообщения должны рассылаться только клиентам, относящимся к сегменту контрольная группа (не рассылаться тестовой).

5. Оказание целевого воздействия на каждую из групп в соответствии с выбранными стратегиями каждой из команд, воздействие производится на дату старта А/В-теста.

6. По истечении 40 дней проверка заключенных сделок по тестовой и контрольной группам.

7. Расчет целевого показателя (при этом учитывается также доход с клиентов, не совершивших сделку, как нулевой).

8. Проверка равенства дисперсий целевого показателя в контрольной и тестовой группах по критерию Фишера.

9. Расчет критерия Беренса-Фишера  $t_{эм}$  (модифицированный критерий Стьюдента), если дисперсии статистически различны.

10. Расчет критерия Стьюдента  $t_{эм}$ , если дисперсии статистически равны.

11. Сравнение рассчитанного критерия (под п. 8 или 9) с критическим значением  $t_{кр} = 1.9639$ , если  $t_{эм} > t_{кр}$ , то различия в исследуемых выборках являются статистически значимыми и нулевая гипотеза отвергается.

#### **4. Заключение.**

В результате проведенной работы разработана опытная модель, предсказывающая совершение клиентом сделки от момента заведения первого интереса и в течение 40 дней, и разработан план А/В-теста. Следующим этапом является проведение А/В-теста и анализ результатов тестирования.