



Министерство науки и высшего образования Российской Федерации
федеральное государственное бюджетное образовательное учреждение
высшего образования

«Российский экономический университет имени Г.В. Плеханова»
Кафедра информатики



ИЦЭиИТ
РЭУ им.Г.В.ПЛЕХАНОВА

Выпускная квалификационная работа на тему: **«Разработка подсистемы для оценки вероятности повторной покупки клиента строительного гипермаркета»**

Выполнил:
студент 4 курса 291-БИ/09/16
Асенбаев Маманбий Зайытбаевич

Научный руководитель:
к.т.н., доцент
Паклин Николай Борисович



Цель, объект и предмет и задачи исследования



ИЦЭИИТ
РЭУ им.Г.В.ПЛЕХАНОВА

Задачи

Цель:

Оценка вероятности повторной покупки клиента для совершенствования коммуникаций с клиентской базой

Объект исследования:

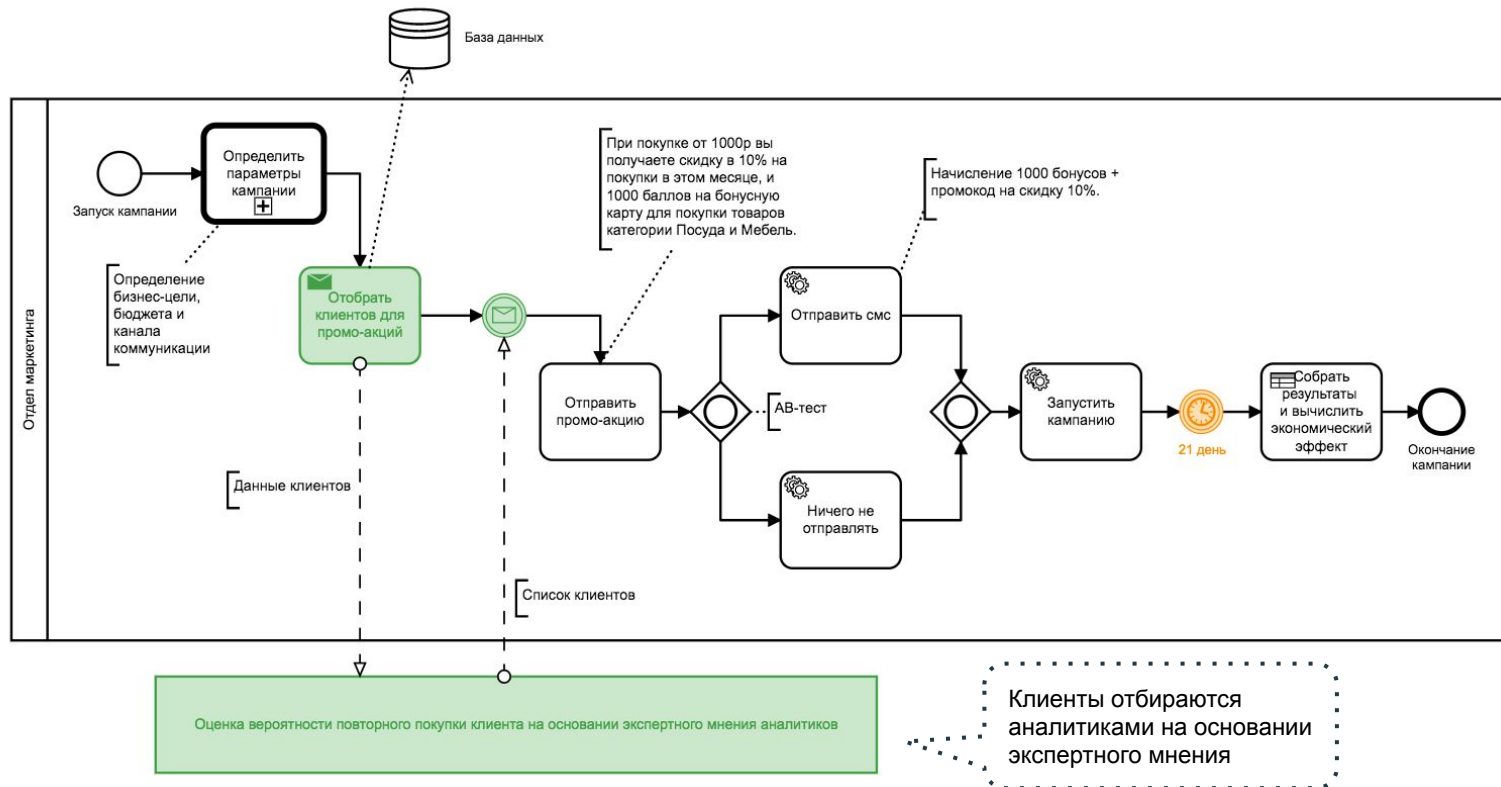
Строительная розничная сеть

Предмет исследования:

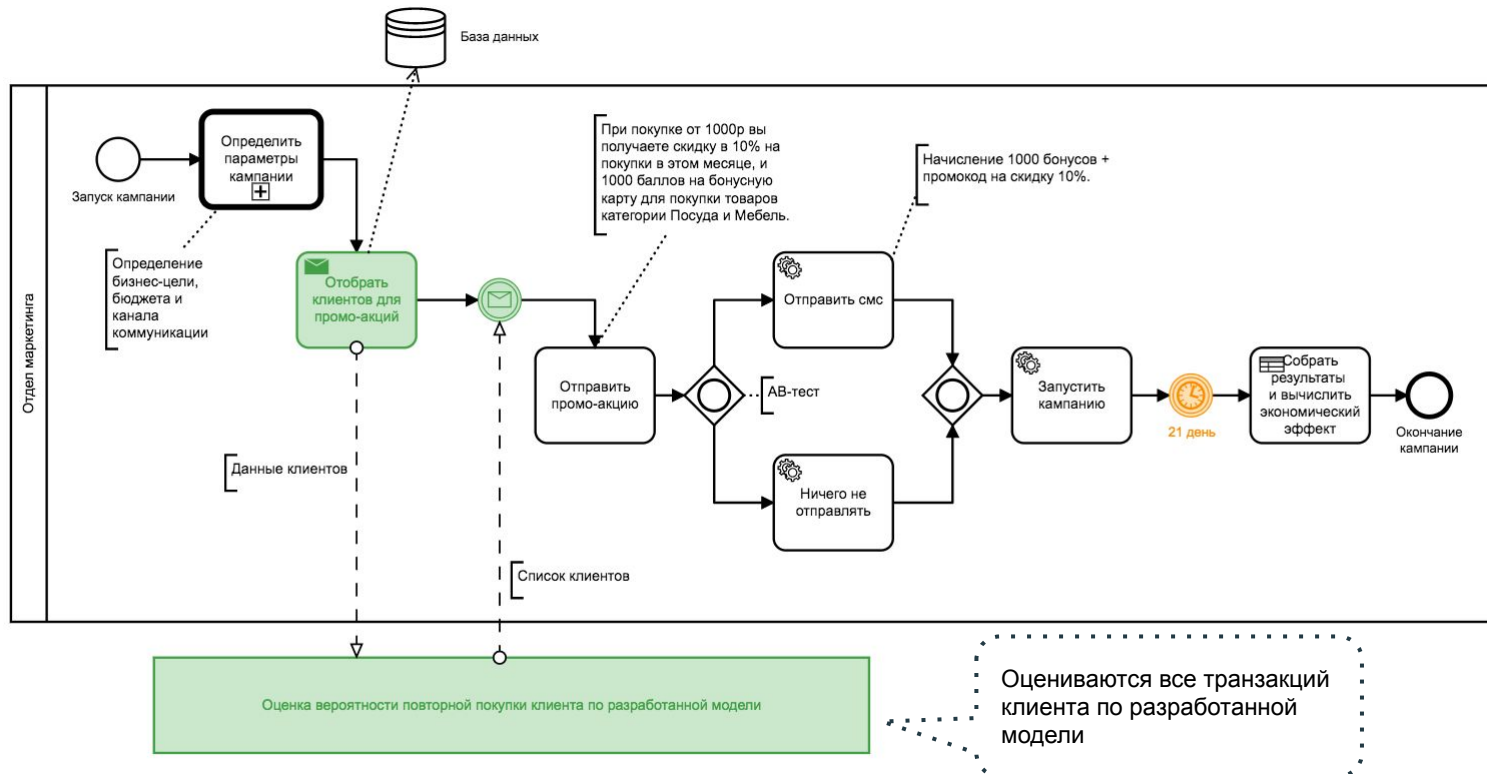
Клиентские базы данных, ETL-процессы, алгоритмы машинного обучения, аналитические веб-сервисы

1. Изучить процессы таргетирования в персонализированном маркетинге и объекте исследования.
2. Сформировать новый бизнес-процесс для маркетолога с использованием модели оценки вероятности.
3. Разработать архитектуру информационно-аналитической подсистемы и выбрать программные инструменты.
4. Подготовить и собрать данные объекта исследования – транзакции клиентов розничной DIY-сети.
5. Построить модель оценки вероятности на основе прошлых транзакций клиентов и оценить её качество.
6. Опубликовать модель как аналитический веб-сервис.

Модель AS-IS



Модель ТО-ВЕ



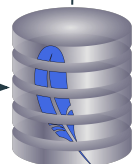


ИЦЭиИТ
РЭУ им.Г.В.ПЛЕХАНОВА

Предполагаемая архитектура аналитической подсистемы

ОФФЛАЙН

Транзакций



БД SQLite

SQL запрос транзакций
клиента

Веб-сервис



Отправляет в режиме
реального времени id
клиентов и получает
профиль клиента

WWW



Канал
коммуникаций



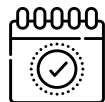
профили клиентов
экспортируются в



Исходные данные для анализа

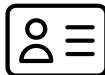


ИЦЭиИТ
РЭУ им.Г.В.ПЛЕХАНОВА



Дата

31



ID клиента

ab



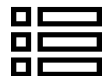
ID чека

ab



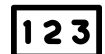
Товар

ab



Товарная группа

ab



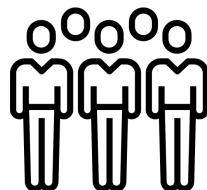
Количество

12

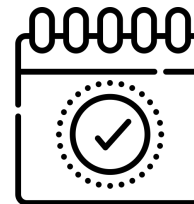


Сумма

9.0



Количество
клиентов:
42 738



Период
анализа:
01.09.2017 -
31.10.2019

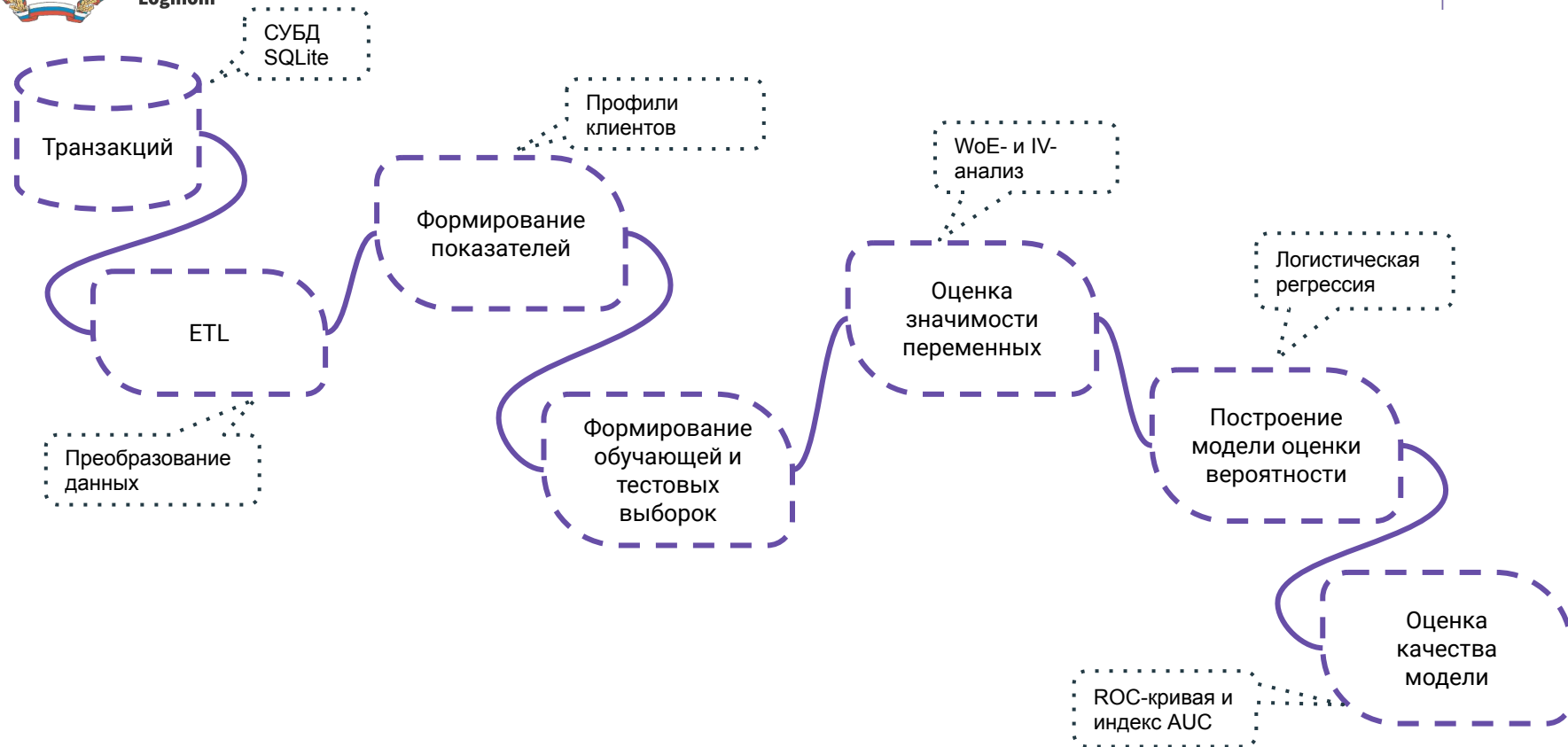


Количество
транзакций:
1 010 000

Какие данные
хранятся в базе
данных



Процесс построения модели





Этап формирования показателей

Исходные показатели

	Дата	31
	ID клиента	ab
	ID чека	ab
	Товар	ab
	Товарная группа	ab
	Количество	12
	Сумма	9.0



Показатели,
связанные с датой,
временем, кол-во,
суммой транзакций

Показатели,
связанные с
товарной
группой

Показатели,
связанные с
товаром

Сформированные показатели

- 12 Кол-во транзакций
- 9.0 Сумма транзакций
- 9.0 Средняя сумма транзакции
- 9.0 Доля визитов в выходные
- 12 Время сна, дни
- 12 Дней с первой транзакции
- 12 Дней между последними двумя транзакциями
- 9.0 Средний промежуток между транзакциями, дни
- 9.0 Минимальная транзакция
- 9.0 Максимальная транзакция
- ab Любимая категория
- 9.0 Среднее кол-во категорий в транзакции
- 12 Разнообразие категорий
- 9.0 Среднее число позиций в транзакции
- 12 Разнообразие позиций
- 9.0 Средняя стоимость позиции
- 9.0 Минимальная стоимость позиции
- 9.0 Максимальная стоимость позиции

18

показателей



Формирование обучающей и тестовой выборки

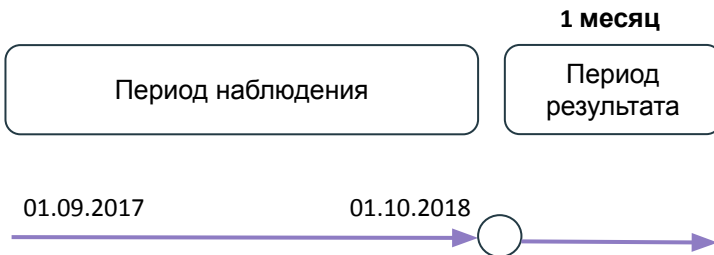


ИЦЭиИТ
РЭУ им.Г.В.ПЛЕХАНОВА

Метод аналогий: модель обучается на исторических профилях клиентов, и применяется для оценки вероятности события текущих клиентов.

Событие: повторный визит в течение 1 месяца.

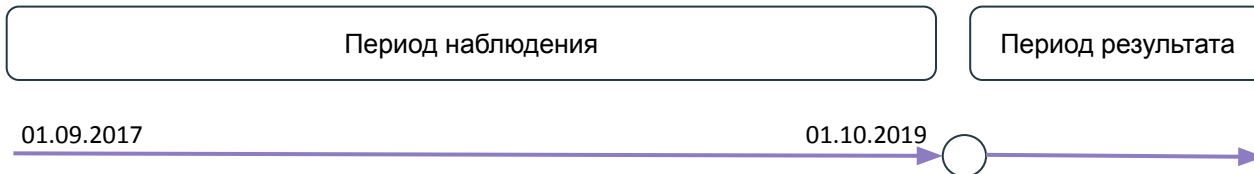
Исторические профили



Обучающая
выборка
(23 728
клиентов)

Тестовая
выборка
(41 197
клиентов)

Оцениваемые профили





Оценка значимости переменных



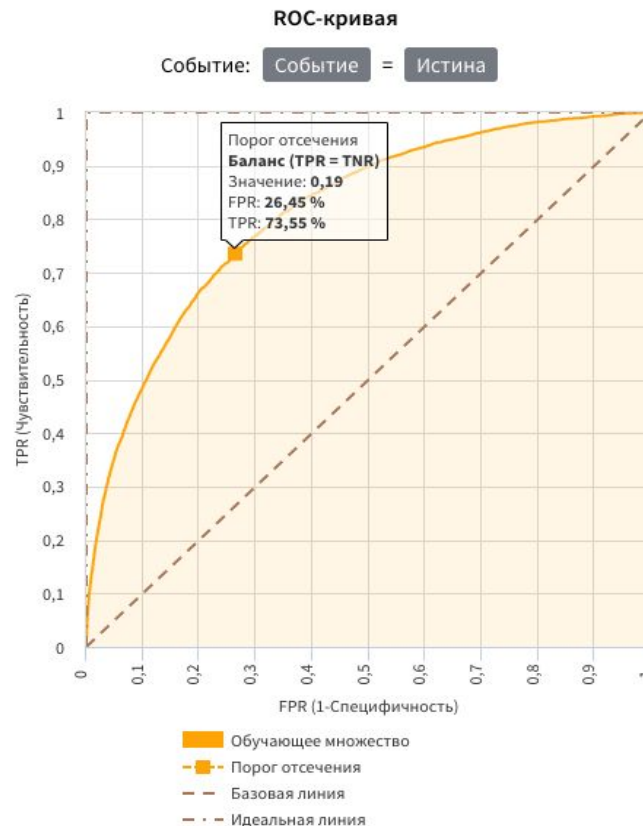
ИЦЭИИТ
РЭУ им.Г.В.ПЛЕХАНОВА



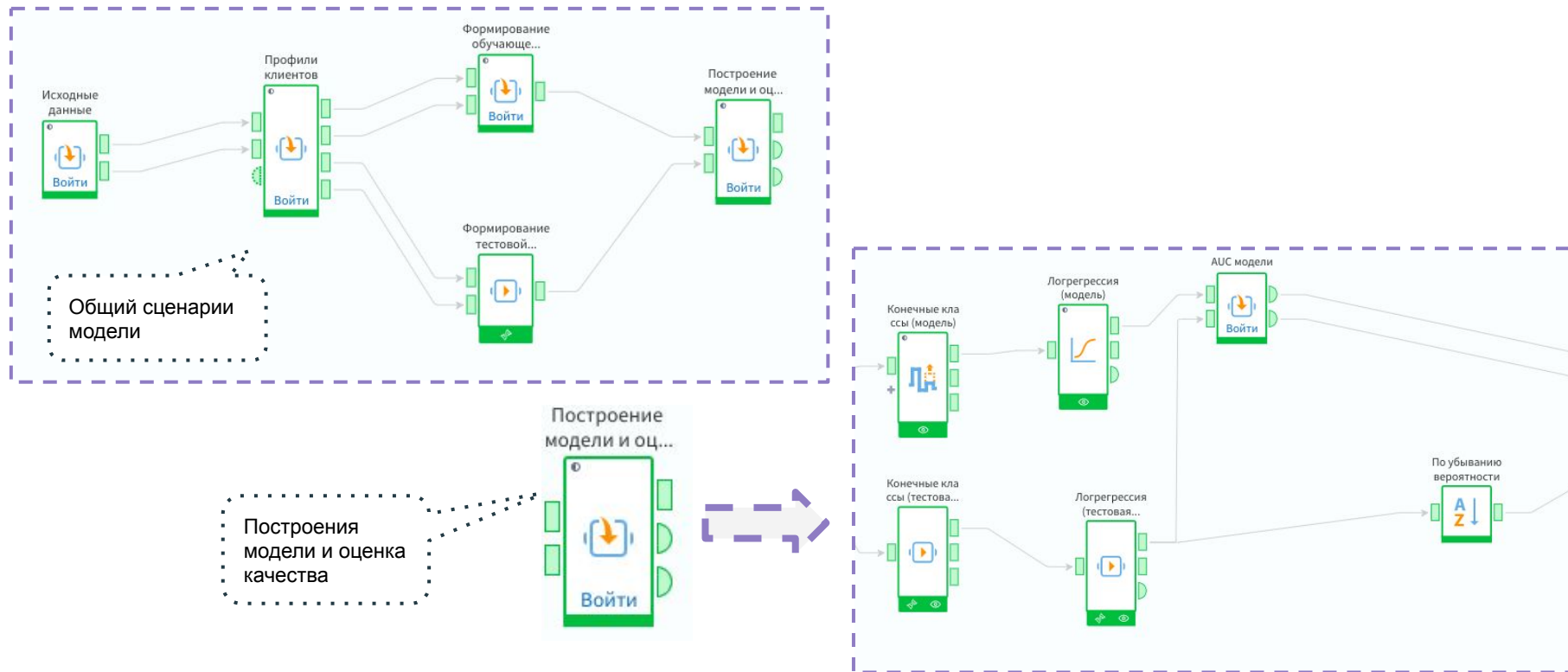
Информационный индекс (от англ. Informational Value – IV) отвечает за предсказательную силу переменной.

IV < 0,02 – отсутствует;
0,02 <= IV < 0,1 – низкая;
0,1 <= IV < 0,3 – средняя;
IV >= 0,3 – высокая.

- Метод – логистическая регрессия с L2-регуляризацией (Ридж)
- Модель статистически значима
- Площадь под кривой AUC на тестовой выборке 0.81 -> хорошая предсказательная способность



Сценарий построения модели в Loginom





Конкретный пример



ИЦЭиИТ
РЭУ им.Г.В.ПЛЕХАНОВА



Сформированный показатель	Значение
Кол-во транзакций	20
Сумма транзакций	60354 руб.
Доля визитов в выходные	0,25%
Средняя сумма транзакции	3017,7 руб.
Минимальная транзакция	170 руб.
Максимальная транзакция	21768 руб.
Время сна, дни	5 дней
Дней с первой транзакции	309 дней
Дней между последними двумя транзакциями	9 дней
Средний промежуток между транзакциями, дни	15,45 дней
Любимая категория	Инструменты
Среднее кол-во категорий в транзакции	2,5
Разнообразие категорий	8
Средняя стоимость позиции	434,1746 руб.
Минимальная стоимость позиции	19
Разнообразие позиций	85
Максимальная стоимость позиции	15291 руб.
Среднее число позиций в транзакции	4,25



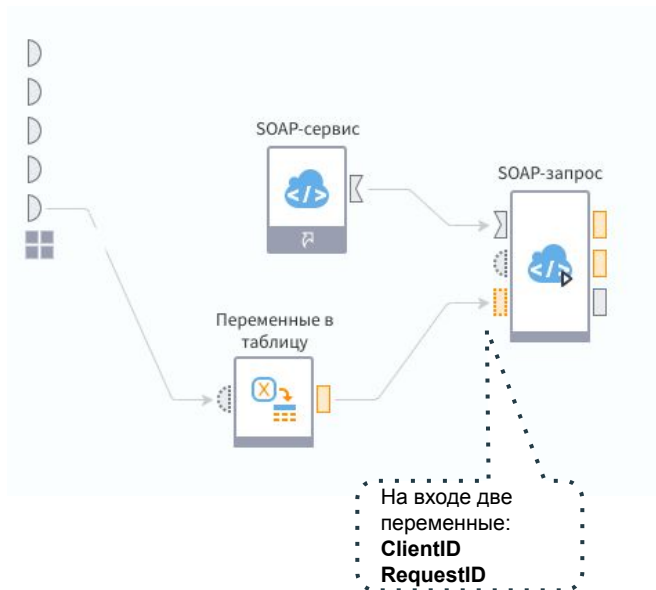
Фильтр строк • Соответствуют условию • Б		
#	ab ID клиента	9.9 Вероятность визита
1	client15771	0,79



Модель как веб-сервис



ИЦЭиИТ
РЭУ им.Г.В.ПЛЕХАНОВА



```
<?xml version='1.0' encoding='utf-8'>
<wsdl:definitions name="Service" targetNamespace="http://schemas.loginom.ru/integration">
  <wsdl:import namespace="http://schemas.loginom.ru/integration/Final4websrv" location="http://demo.loginom.ru/market/Igi/Service.svc?wsdl=wsdl2"/>
  <wsdl:types/>
</wsdl:definitions>

<wsdl:binding name="BasicHttpBinding_Final4websrv" type="t2:Final4websrv">
  <soap:binding transport="http://schemas.xmlsoap.org/soap/http"/>
</wsdl:binding>

<wsdl:operation name="VisitProbability">
  <soap:operation soapAction="http://schemas.loginom.ru/integration/Final4websrv/Final4websrv/VisitProbability" style="document"/>
</wsdl:operation>

<wsdl:input>
  <soap:body use="literal"/>
</wsdl:input>

<wsdl:output>
  <soap:body use="literal"/>
</wsdl:output>

</wsdl:operation>
</wsdl:binding>

<wsdl:service name="Service">
  <wsdl:port name="BasicHttpBinding_Final4websrv" binding="tns:BasicHttpBinding_Final4websrv">
    <soap:address location="http://demo.loginom.ru/market/Igi/Service.svc/Soap/Final4websrv"/>
  </wsdl:port>
</wsdl:service>
```

WSDL-
описание
сервиса



Заключение



ИЦЭиИТ
РЭУ им.Г.В.ПЛЕХАНОВА

Цель достигнута - разработана подсистема оценки вероятности повторной покупки клиента строительного гипермаркета и решены следующие задачи:

1. Изучены процессы таргетирования в персонализированном маркетинге и объекте исследования.
2. Сформирован новый бизнес-процесс для маркетолога с использованием модели оценки вероятности.
3. Разработана архитектура информационно-аналитической подсистемы и выбраны программные инструменты.
4. Подготовлены и собраны данные объекта исследования – транзакции клиентов розничной DIY-сети.
5. Построена модель оценки вероятности на основе прошлых транзакций клиентов и оценено её качество.
6. Опубликована модель как аналитический веб-сервис.



ИЦЭиИТ
РЭУ им.Г.В.ПЛЕХАНОВА

СПАСИБО