Исследуемая проблема заключается в том, что, учитывая данные об интересах клиентов, можно повысить прибыль, через улучшения эффективности маркетинга. Эти данные можно собирать и анализировать, тем самым извлекая из них полезную информацию.

Целью дипломного проекта является помощь маркетологам в понимании предпочтений владельцев автомобилей через автоматизированный сбор и анализ данных, что позволит разрабатывать более точные и персонализированные маркетинговые стратегии.

Задачи, которые необходимо решить:

* автоматизация процесса сбора данных;
* распознавание интересов посетителей парковки;
* прогнозирование интересов посетителей парковки.

Сбор и анализ данных о предпочтениях владельцев автомобилей - важный аспект для улучшения обслуживания и удовлетворения потребностей клиентов. Ключевым моментом является выбор методов сбора данных, их обработка. Для начала нужно определиться, каким образом собирать данные.

В зависимости от характера сбора данных их можно разделить на два основных типа, а именно:

* + - первичный метод;
    - вторичный метод.

Первичные данные исследователи получают самостоятельно и впервые в контексте своего исследования. Существуют несколько способов сбора первичных данных:

* интервью;
* метод Дельфи;
* проективные методы;
* интервью в фокус-группе;
* метод анкетирования.

Вторичные данные означают данные, которые исследователь собирает не самостоятельно. Фактически, вторичные данные уже доступны и их необходимо собирать из различных источников. ???? (обобщенные данные не нужны, значит не подходит)

Наиболее оптимальным будет метод анкетирования, потому что он не подразумевает прямого взаимодействия с клиентом, то есть источником информации. Значит нам нужен модуль, который будет обрабатывать заполненные клиентами анкеты и получать оттуда информацию. Сюда же можно включить компонент, который будет находить модель автомобиля, для того чтобы связать модель, возраст владельца с его интересами.   
 Для дальнейшего анализа собранных данных и предсказания интересов используется машинное обучение. Само обучение можно разделать на 2 основных класса: обучение с учителем и без.

В случае обучения с учителем модели предоставляются правильные ответы для обучающих данных.

В случае обучения без учителя модели не предоставляются правильные ответы. Модель обучается только на основе входных данных, без информации о правильных ответах.

В данном же случае используется обучение с учителем, так как необходимо именно предсказать интересы клиентов. То есть, также необходим второй модуль, который будет отвечать за обучение нейросети, а также за работу с обученной моделью

Разработанная система имеет следующую структуру. Она состоит из двух модулей: «Сбор информации» и «Предсказание интересов».

Первый модуль распознает номер автомобиля и его модель, а также анализирует бланк анкеты, который заполняет клиент. Для этого используется предобученные модели Tensorflow, OpenCV. Поиск модели происходит на сайте Autoteka.ru.

Второй модуль позволяет обучить нейросеть и работать с обученной моделью. Модель обучается с помощью Tensorflow и Keras, для автоматизированного подбора параметром используется GridSearchCV.

Вот модель нейросети, которая у меня получилась при обучении на пробном наборе данных.

Рассмотрим пример. С точки зрения первого модуля. Сбор информации об интересах. Есть функции загрузки видео, разбиения на кадры, выбор номера, то есть одного кадра, и распознавание номера. / Вот выбран один кадр. / Дальше распознается номер / и происходит поиск модели авто. / Выбор анкеты. Тут определяются ответы клиента и его возраст. / Собранные данные объединяются в одну строку и записываются в датасет, который является входными данными для второго модуля. / С точки зрения второго модуля. Обучение нейросети. Можно выбрать файл и запустить процесс обучения. / Предсказание интересов. / Тут необходимо ввести модель машины и возраст, а также выбрать файл обученной модели. А результаты выводятся в таблицу.

В результате работы была разработана программная система, которая может быть использована в реальных задачах, направленных на анализ интересов посетителей парковок. Разработанная система принесет пользу для маркетологов и других людей, которые работают с клиентами. Она помогает легко собирать и анализировать информацию о том, что нравится посетителям парковки. Это позволяет более точно нацеливать рекламные компании и предложения, что, в свою очередь, улучшает впечатления клиентов и делает маркетинг более эффективным.

Входной информацией является совокупность данных для сбора информации и данных для предсказания интересов. О ни дальше расскажу подробнее. На выходе системы получаем датасет для обучения нейросети, обученную модель нейросети и предсказания интересов. Система состоит из 2-х модулей: «Сбор данных» и «Предсказание интересов». Первый модуль предназначен для того, чтобы получать данные о владельце (возраст и его интересы) и модель автомобиля. Модуль получает на вход данные, куда входят изображения автомобилей с видимым номерным знаком и бланки опроса, и извлекает из них необходимую информацию (далее расскажу как). Таким образом, на выходе получаем информацию об интересах клиента, его возраст и модель автомобиля. Полученные данные соединяются в строку и записываются в датасет. Второй модуль предназначен для обучения нейросети и предсказания интересов. Он использует для обучения сформированный первым модулем дадасет. Далее полученную обученную модель можно использовать для предсказания интересов на основе модели автомобиля и возраста клиента. Таким образом, на выходе у модуля получаем обученную модель (их можно обучить несколько штук) и предсказания интересов.

Модуль сбора информации состоит из 3 компонентов: «Анализ бланков», «Распознавание номера», «Определение модели». Первый компонент проводит обработку бланков опроса. Бланк содержит в себе набор вопросов и кружки с ответами «ДА» и «НЕТ». Клиенту необходимо закрасить кружок с верным для него ответом и указать свой возраст в специальном поле. Пример бланка представлен в ПЗ. При загрузке изображения бланка компонент распознает ответы и сохраняет их в виде последовательности чисел 1 и 0, где 1 соответствует ответу «ДА», а 0 – ответу «НЕТ». Также происходит распознавание возраста. Компонент «Распознавание номера», как понятно по названию, необходим для определения номера автомобиля. На вход ему подается изображение авто. Компонент использует 2 предобученные модели из библиотеки TensorFlow, одна из которых находит номерной знак, а другая распознает текст на нем. Тем самым, на выходе компонента получаем текст с номерного знака авто. Полученный текст передается в компонент «Определение модели», который находит названия машины, на которой приехал клиент. На данном этапе программно открывается сайт Autoteka.ru, вбивается в поиск номер автомобиля. По номеру сайт выдает краткий отчет об автомобиле при условии, что он стоит на учете в ГИБДД. Из этого отчета вытаскивается название модели. Дальше все полученные данные объединяются в одну строку и записываются в датасет.

2-ой модуль – «Обучение нейросети и предсказание интересов». Как понятно по названию, он позволяет обучить нейросеть на сформированном датасете и спрогнозировать инетресы посетителей. Обучение нейросети происходт с помощью библиотеки GridSearchCV, которая автоматизирует процесс подбора параметров. То есть можно указать какие параметры и какие значения могут принимать, а дальше библиотека позволяет запустить перебор всех возможных вариантов параметров и их значений, что, конечно занимает длительны промежуток времени. Далее сохраняется лучшая модель. Структура модели, которая получилась при обучении на пробном датасете приведена дальше. Второй компонент предназначен для предсказания интересов. По своей сути он представляет интерфейс для работы с обученной моделью, имеет поля для ввода названия модели автомобиля и возраста, а также возможность выбора модели нейросети, так как можно обучить несколько штук. Интересы предсказываются в виде вероятности. То есть в таблицу по каждой из областей интересов для каждого предсказания выводятся вероятности вида 45.36% и тд. Это показывает, на сколько вероятно присутствие того или иного интереса у человека.

Полученная мной структура нейросети. Она состоит из входного слоя, трех полносвязных слоев с 20 нейронами и трех слоев отсеивания для предотвращения переобучения нейросети. На вход модели подаются возраст и модель авто, а на выходе 7 вероятностей.

Окно определения модели авто. Можно загрузить видео, разбить его на кадры, выбрать один кадр и распознать номер. После этого номер отобразиться в окне. Дальше откроется сайт autoteka.ru и произойдет автоматический поиск модели авто. Как только система найдет модель, будет предложено выбрать файл бланка, которые необходимо обработать. После обработки бланка строка с полученными данными запишется в датасет. На следующем слайде представлено окно для обучения нейросети. Тут можно выбрать датасет, на котором будет происходить обучение. На этом слайде мы видим окно для предсказания. Тут можно выбрать обученную модель, ввести данные и получить предсказание.

В результате работы была разработана программная система, которая может быть использована в реальных задачах, направленных на анализ интересов посетителей парковок. Разработанная система принесет пользу для маркетологов и других людей, которые работают с клиентами. Она помогает легко собирать и анализировать информацию о том, что нравится посетителям парковки. Это позволяет более точно нацеливать рекламные компании и предложения, что, в свою очередь, улучшает впечатления клиентов и делает маркетинг более эффективным.