

Министерство образования и науки Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ
Н.Г.ЧЕРНЫШЕВСКОГО»

Кафедра дискретной математики
и информационных технологий

Информационные технологии анализа социальных сетей

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

Студента 5 курса 521 группы
направления 09.03.01 «Информатика и вычислительная техника»
факультета компьютерных наук и информационных технологий
Внуковского Владислава Сергеевича

Научный руководитель

д. э. н., профессор

подпись, дата

Л.В. Кальянов

Зав. кафедрой

к. ф.-м.н., доцент

подпись, дата

Л.Б. Тяпаев

Саратов 2017

В последние годы Интернет стал источником всевозможной информации. Веб-контент все чаще и чаще используется для того чтобы узнавать о предпочтениях клиентов и реальных людей. Сегодня социальные сети в мире занимают нишу самых высокодоходных сайтов в Интернете и хранят огромный объем данных. Задача состоит в том, чтобы получить доступ к этим данным и преобразовать их в что-то действительно полезное. Чаще всего, организации хотят использовать данные социальных сетей чтобы понимать поведение и интересы своих клиентов или конкретных целевых групп лиц в отношении текущих или будущих продуктов, или услуг организаций.

Изначально социальные сети задумывались как инструмент для комфортного общения. Человек мог в любое время посмотреть, что нового в жизни у его друзей, оставить комментарий, написать. Но с течением времени социальные сети превратились в серьезную медийную силу, которую стало возможно использовать в своих интересах. Если посмотреть, что происходило в этой сфере в течение последних нескольких лет, то можно заметить постепенный рост практического использования социальных медиа. Реклама, маркетинг, продвижение брендов, онлайн-продажи и даже политические кампании – все это постепенно уходит в мир социальных сетей. Социальные медиа на данный момент являются самым актуальным трендом в интернете, этот сегмент развивается очень быстро. ВКонтакте, Facebook, Twitter, YouTube и другие социальные сети создают мощные информационные потоки, которые влияют на жизнь огромного числа людей.

В работе данные анализировались при помощи аналитической платформы Knime Analytics Platform.

KNIME предоставляется на бесплатной основе и представляет собой модульную платформу анализа данных. Позволяет реализовывать полный цикл анализа данных встроенными средствами включающий чтение данных из различных источников, преобразование и фильтрацию, статистику, анализ социальных и медиа-данных, визуализацию и экспорт.

Среда позволяет пользователю визуально создавать поток данных, выборочно выполнять анализ шагов, а затем исследовать результаты посредством интерактивного просмотра данных и моделей.

Потоки данных могут быть запущены через интерактивный интерфейс пользователя и обрабатываться в пакетном режиме, упрощая интеграцию процесса анализа данных для менеджмента и выполнения на периодической основе.

Основная часть среды включает в себя модули для обработки данных:

- узлы доступа к данным – для чтения и записи данных из БД и файлов;
- узлы управления данными – для управления таблицами данных, которые используются между узлами;
- инструменты для построения графиков, диаграмм и графических изображений;
- узлы для статистики и интеллектуального анализа данных.

Данная система предоставляет довольно гибкий подход к построению алгоритмов анализа, преобразования и визуализации данных, но все же при реализации сложных манипуляций с данными можно столкнуться со следующими проблемами:

- Недостаточно гибкое поведение узлов – они работают по определенным алгоритмам и нет возможности обработать какие-либо нюансы.

- Программирование сложных алгоритмов может сложиться в длинную и сложную диаграмму с циклами, условиями, что приведет к сложности чтения и отладки.

KNIME хорошо подойдет для людей, не имеющих большого опыта программирования. С его помощью можно быстро создавать простые и средней сложности проекты.

KNIME показывают хорошее быстродействие при загрузке из файлов и обработке данных, построении графиков. Узким местом является получение данных через различные сервисы и интерфейсы.

Веб-интерфейсы социальных сетей являются источниками данных реального времени и предназначены для просмотра и взаимодействия со страницами социальной сети в веб-браузере либо для использования данных пользователей специализированными приложениями. Поскольку сценарии использования интерфейсов социальных сетей не предполагают автоматического сбора данных множества пользователей, то возникает ряд проблем:

1. приватность данных - зачастую доступ к данным пользователей разрешён только для зарегистрированных и авторизованных участников сети, что требует поддержки эмуляции пользовательской сессии с помощью специальных учётных записей (аккаунтов). В большинстве социальных сетей для работы с методами API необходимо передавать в запросе `access_token` — специальный ключ доступа. Такой ключ позволяет работать с API от имени группы или пользователя;

2. слабая структурированность данных - во многих случаях программные интерфейсы (API) социальных сетей имеют ограниченный функционал, что требует поддержки получения с помощью пользовательского веб-интерфейса статических копий HTML-страниц, корректной обработки их динамической части, извлечения нужных данных с помощью алгоритма и/или шаблона и построения их структурированного представления, удобного для дальнейшей автоматической обработки;

3. ограничения доступа и блокировки - с целью предотвращения несанкционированного автоматического сбора данных и ограничения нагрузки на инфраструктуру сервиса социальной сети владельца сервисов зачастую вводят явные или скрытые ограничения на допустимое количество запросов от одного пользовательского аккаунта и/или IP-адреса в единицу времени, что требует учёта количества посылаемых запросов.

4. права доступа - все методы делятся на секции — friends, photos, video и так далее. Каждая секция соответствует определенному разделу на сайте. Часто приложения создаются для работы с конкретным разделом. Например, для просмотра фотографий или для редактирования видео. При этом логично, что если есть доступ только с аудиозаписями пользователя, остальные разделы будут не доступны.

В данной работе данные получаются из социальной сети vk.com при помощи интерфейса API ВКонтакте. В первой части проекта получают данные из профиля пользователя участников сообщества. Основные данные которые получить из профиля пользователя это: имя, фамилия, дата рождения, город, пол и количество подписчиков. Метод «users.get» позволяет получать все поля из профиля пользователя, но это делать не целесообразно т.к. такие поля как «О себе», «О карьере», «Жизненная позиция» заполняются очень редко. В результате запроса сервер вернет объект в формате JSON с запрошенными данными. JSON Schema — это распространенный стандарт описания структуры данных. Схема создана для описания JSON-данных, но и сама она при этом является JSON-объектом. С помощью ключевых слов в схеме создаются правила валидации структуры объекта и типов его полей.

Получив данные о пользователях можно составить среднестатистические данные. Большинство подписчиков указали в профиле город Саратов и возраст от 20 до 30 лет.

В системах интернет-маркетинга и рекомендаций особую важность представляет определение демографических атрибутов пользователя для

таргетированного продвижения товаров и услуг в группах пользователей с одинаковыми значениями атрибутов. Помимо интернет-сервисов, такие демографические характеристики находят применение в различных дисциплинах: социология, психология, экономика, управление персоналом и др.

Для работы в социальных сетях важно отслеживать аналитику по демографическим и географическим данным аудитории, активность в сообществе по часам и дням недели, наиболее популярные материалы и т. д.

Именно на эти значения нужно обращать внимание при составлении контент-плана. Например, с каждой возрастной категорией придется искать разные точки взаимодействия. Это главное преимущество, которое поможет находить общие интересы и создавать тот контент, который будет резонировать с каждой группой.

Для разных сообществ и социальных сетей нужны разные материалы, поэтому простое копирование текста из блога в ВКонтакте и Facebook может стать крайне проигрышной стратегией. Важно понимать, что не получится заметно повысить показатели страницы или сообщества, не разобравшись в том, какой контент действительно интересен аудитории.

Также была получена статистика подписок участников сообщества.

Для анализа сообщества были получены последние 10000 записей. определим уровень вовлеченности подписчиков. Для этого нужно подсчитать средние показатели по сообществу. В среднем пользователи ставят запись отметки мне нравится – 120, комментариев – 20 и репостов – 6.

Таким образом можно анализировать эффективность постов, отфильтровав записи больше или меньше средних показателей.

В работе записи сообщества анализировались при помощи хештегов. Хештег— обозначает ключевое слово сообщения. Их использование значительно упрощает поиск записей по теме или содержанию. Кроме того, можно легко определить основной контент и уровень увлеченности участников по основным темам. Хештеги так же используют в рекламной

продукции, в политических акциях или городских мероприятиях в качестве отсылки к появившемуся тренду в интернете или в попытке создать такой тренд.

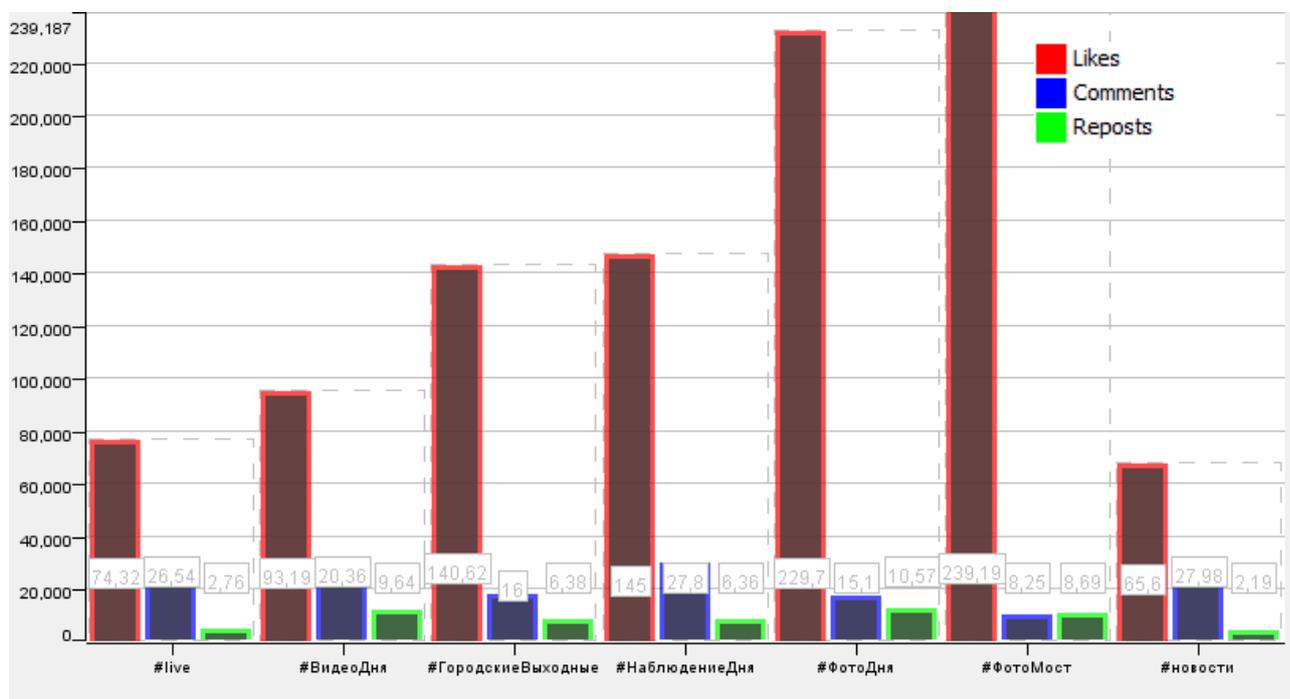


Рисунок 1 – средние показатели вовлеченности участников по хештегам.

Анализируя рисунок 1 видно, что повседневные новости с хештегами #Новости и #live показывают уровень вовлеченности ниже среднего, хотя и составляют основную часть контента. Исходя из статистики традиционные новости утомляют большинство читателей по большей части из-за большого количества текста. Самый высокий уровень вовлеченности участников группы показывают фотографии города и достопримечательностей, медиа контент участников группы не новостного характера и интересные факты и мероприятия из жизни города. Это и представляет интерес для целевой аудитории данного сообщества.

Так же можно сделать выводы о форме подачи контента. Оптимальным вариантом будет картинка или фотография с текстовым содержанием и небольшим комментарием. Даже если изображение не требуется, с ним пользователям будет привычнее просматривать ленту. Такие же требования будут относиться и к рекламе в сообществе. Если необходимо выделить

рекламу из общей массы постов, то лучше это делать явно не формой подачи. Например, рекламу в формате видео или длинного текста вряд ли кто-то будет смотреть.

API ВКонтакте позволяет получать информацию о том, содержит ли запись отметку «реклама». Отфильтровав записи по этой отметке получим, что в данном сообществе за последний год размещено около 600 записей с пометкой «Владелец сообщества разместил эту запись платно».

У меня есть друг и я хочу вас с ним познакомить. Зовут его " [club43480339]Афиша. Саратов". Он знает обо всех тусовках города, мероприятиях, поездках, концертах и мастер-классах. Он не пропуск...
Самые атмосферные снимки Саратова в нашем аккаунте @SaratovInsta Подписывайся ☐ https://www.instagram.com/saratovinsta/
Сделай репост с 27 июля по 3 августа 2016 г. и получи подарки от лидера группы «Чайф»! До старта MegaFonLive 2016 в Саратове #megafonlive64 осталось 10 дней. У вас есть уникальная возможнос...
Доброго времени суток, Саратовцы! Сейчас на сайте аллен-славы.рф проходит акция, по результатам которой в 15 городах (по 5 городов в каждой категории) высадят 9000 деревьев...
Друзья, товарищи, знакомые, френды из социальных сетей! В Саратове есть отличный юрист Дмитрий Миненков. Он оспорил две сделки в между большими людьми и выиграл все суды в Саратове, К...
Самые атмосферные снимки Саратова в нашем аккаунте @SaratovInsta Подписывайся ☐ https://www.instagram.com/saratovinsta/
[club144142855]ТОП-7 событий «Городских выходных» 13-14 мая! До фестиваля осталось всего 5 дней! Мы составили для вас гид по событиям, которые нельзя пропустить. Наполним, вход на все мер...
Увлекательные прогулки по лесам и полям за рулём мощных квадроциклов! 1час-1500Р 2часа-2500Р Выезд от Гусельского моста. Тел. 53-41-59
Бортпроводники «Саратовских авиалиний» прошли по подиуму в московском аэропорту «Домодедово»
ПРОКАТ КВАДРОЦИКЛОВ! Увлекательный маршрут. Лучший отдых на свежем воздухе. Попробуй свои силы за рулём мощного квадроцикла. Не знаете что подарить друзьям на Новый год? Подароч...
Увлекательные прогулки по лесам и полям за рулём мощных квадроциклов! 1час-1500Р 2часа-2500Р Выезд от Гусельского моста. Тел. 53-41-59
Уже второй месяц как в Саратове вещает [club134636674]Наше радио, для кого-то эта новость стала предновогодним подарком. А вы любите «Наше»?
СВИДАНИЕ НА САМОЙ КРАСИВОЙ КРЫШЕ САРАТОВА! Устрой своей половинке незабываемый вечер! A SVIDANIE64 тебе в этом поможет. Под вами шумный город, над вами только звезды. Легкий ужин ...

Рисунок 2 – реклама в сообществе с высокими показателями вовлеченности.

Как видно из рисунка 2 популярной оказалась реклама различных мероприятий, а также заведений города, например, кафе. Таким образом вышеописанные выводы о интересах целевой аудитории совпадают с тематикой рекламы.

Так же стоит рассмотреть рекламу с уровнем увлеченности около нуля. На рисунке 3 видно, что рекламодатели выбрали сообщество с не подходящей для них целевой аудиторией. А трое из них ошиблись дважды, написав в записи около 200 слов сплошным текстом.

Комиссионный магазин "Звезда"☐ Срочно нужна большая сумма? Мы принимаем в залог мотоциклы, авто, недвижимость, зе...
Привет! Меня зовут Павел Фролов и я занимаюсь рекламой в сообществе Типичный Саратов. Сейчас я набираю новых клиентов на продвижен...
Почему плёнка от недобросовестных оклейщиков трескается и облезает через месяц? Чем специализированная защитная автоплёнка отлича...
cashbery.com - Лидер на рынке инвестирования среди крупнейших компаний России. Кэшбери имеет все необходимые лицензии и ведёт прозра...

Рисунок 3 – реклама в сообществе с нулевым показателем вовлеченности.

Вовлечение аудитории в активности бренда напрямую влияет на продажи и позволяет компаниям корректировать свою рекламную активность, исходя из реакции целевой аудитории.

Такой анализ сообществ позволяет определить показатели увлеченности участников. От этого показателя напрямую зависит стоимость рекламы в сообществе. Так же важно уметь определять основные интересы

целевой аудитории. Такой функционал будет полезен:

- Администраторам сообществ и контент менеджерам — для поиска трендового, самого интересного контента в том числе в других сообществах. Им не придется искать вручную вирусный контент и “предполагать”, является ли контент вирусным, т.к. уровень вовлеченности точно оценивает степень интересности контента.
- Бренд-менеджерам, маркетологам — для отслеживания интереса аудитории к конкретной торговой марке, рекламной кампании, вовлеченной в социальную сеть. Например, отслеживание хештегов, введенных компанией.
- Рекламные агенты, PR-менеджеры — при создании рекламы отправной точкой являются результаты маркетингового аудита. Прежде всего они должны иметь по возможности полное представление о том, что и кому рекламируется. Данные необходимы для определения эффективной рекламной политики и создание убедительной продающей информации для определенной целевой аудитории.

Таким образом пользователи сами добровольно выкладывают всю информацию, которая понадобится для эффективного анализа в социальных сетях: возраст, пол, образование и, что важно, интересы и увлечения. Главное – грамотно и четко понимать, кто же является потенциальными клиентами вашей компании, и двигаться вперед на поиск своей целевой аудитории. А продвижение брендов в социальных сетях – набирающий обороты эффективный инструмент, но только при грамотном использовании.

ЗАКЛЮЧЕНИЕ

Рассмотрены основные компоненты Knime Analytics Platform: чтение данных из различных источников, преобразование, фильтрация и визуализация данных. Кроме того, произведен сбор исходных данных о пользователях сообщества путём обращения к интерфейсу API ВКонтакте. Были получены данные демографических атрибутов пользователей, проведен анализ целевой аудитории сообщества. Так же получены и проанализированы записи в новостной ленте Саратова и получили представление о интересах целевой аудитории сообщества.