**1 слайд**

Добрый день, уважаемые председатели и члены государственной аттестационной комиссии. Вашему вниманию представляется выпускная квалификационная работа на тему «Сбор данных пользователей социальной сети Twitter с помощью технологии OSINT и прогнозирование событий на их основе».

**2 слайд**

Актуальность темы данной выпускной квалификационной работы обусловлена следующими причинами:

* Twitter – универсальная коммуникационная платформа, которая является источником актуальной информации. Эту информацию можно использовать для прогнозирования результатов различных событий в важных жизненных сферах.
* прогнозирование играет роль инструмента планирования, который позволяет компаниям подготовиться к неопределенности, которая может возникнуть в будущем;
* сама компания, Twitter, переживает новый этап своего существования, в связи с чем важно просмотреть изменения в их политике и документации.

**3 слайд**

Целью данной работы является представление инструмента сбора данных микроблогов внутри социальной сети Twitter.

Для достижения поставленной цели необходимо выполнить следующие задачи:

1. Рассмотреть механику работы социальной сети: ее архитектуру, пользователей, структуру сообщений, политику конфиденциальности.
2. Провести сравнительный анализ существующих инструментов сбора данных сообщений Twitter.
3. Провести метаанализ статей по прогнозированию результатов событий с помощью микроблогов внутри интернет-ресурса Twitter.
4. Рассмотреть способы анализа данных, полученных при сборе, для определения результата прогнозирования.

**4 слайд**

В рамках данной работы использовались следующие инструменты:

Языком программирования был выбран Python. Данное решение было обусловлено тем, что python является высокоуровневым, легко читаемым, объекто-ориентированным, кроссплатформенным и интроспективным языком программирования.

Средой разработки приложения был выбран Visual Studio Code.

Запись данных идет в MongoDB. Выбор данной СУБД был обусловлен высокой доступностью, масштабируемостью и безопасностью. Главные особенности MongoDB то, что это кроссплатформенная документоориентированная база данных NoSQL. Она не требует описания схемы таблиц, как в реляционных БД.

Фреймворком для работы веб-приложения был выбран Streamlit. По сравнению с другими фреймами streamlit:

1. Практически не требуется обучения.
2. Удобен как для пользователя, так и для разработчика
3. Требует меньше времени на сборку
4. Поддерживает стороннюю интеграцию графиков

Статьи для метаанализа методов прогнозирования были взяты с сайта arxiv.org

**5 слайд**

В первом разделе было приведено описание технологии OSINT, а также были рассмотрены история и архитектура социальной сети Twitter, структура твитов, проведен краткий сравнительный анализ ее политики конфиденциальности, а также рассмотрены виды данных, которые возможно получить из Twitter.

**6 слайд**

Во втором разделе был проведен анализ существующих инструментов сбора данных в Twitter, некоторые из которых представлены на слайде, рассмотрены их положительные и негативные качества. Также в работе уделено внимание сбору данных через API: потоковый, поисковый и REST API.

**7 слайд**

В третьем разделе был проведен метаанализ статей о прогнозировании результатов политических событий, проходивших в разное время на основании мнений, представленных пользователями социальной сети Twitter в своих микроблогах. Было проведено сравнение результатов исследований с базовой линией, которая была получена уже по итогам самых событий.

**8 слайд**

В рамках работы было выбрано 8 различных исследований, результаты которых были представлены в приложении А данной работы. Отдельно были рассмотрены два наиболее популярных способа прогнозирования: путем простого подсчета, а также путем анализа настроений, выведены параметры их эффективности. В конце главы была рассмотрена и обратная сторона – недостатки прогнозирования событий результатов с помощью Twitter

**9 слайд**

В четвертом разделе был предложен собственных инструмент сбора данных микроблогов пользователей социальной сети Twitter. Интерфейс самого приложения представлен на слайде. Данное приложение собирает данные на основе трех параметров: начала и окончания интервала выбора твитов, а также ключевого слова.

Результатом работы данного приложения является список, в котором содержатся следующие данные: дата/время написания твита, его ID, текст и домен пользователя. Данные записываются в виде документов в коллекцию MongoDB. Выборка данных из базы представлена на слайде.

**10 слайд**

В соответствии с поставленной целью были решены следующие задачи:

1. Рассмотрена механика работы социальной сети: ее архитектура, пользователи, структура сообщений, политика конфиденциальности.
2. Проведен сравнительный анализ существующих инструментов сбора данных сообщений Twitter.
3. Проведен метаанализ статей по прогнозированию результатов событий с помощью микроблогов внутри интернет-ресурса Twitter.
4. Рассмотрен способы анализа данных, полученных при сборе, для определения результата прогнозирования.

Цель представления инструмента сбора данных микроблогов внутри социальной сети Twitter была достигнута.