

# Exploring ADR of diabetes medicine using social media analytics and interactive visualizations

Pritish Naik (192422IT010)

Rohit Meshram (192090IT017)

# Motivation

- Why social media makes sense?
  - 60% of physicians believe that the transparency and authenticity of social media actually helps improve the quality of care delivered to patients.
  - 41% of online users agree that their choice of a specific doctor, hospital, or medical facility will be influenced by word of mouth on social media
- How online health community (OHC) helps?
  - traditional data analysis
    - manually identify new perspectives from patient data
    - are not feasible or efficient
  - OHC is social networking platform where patients can discuss their health concerns and share their experience with other patients or health professionals

# Literature review

- Issues related to Adverse drug reactions
  - ADRs have been monitored by the FDA in US and ADR surveillance relies on FAERS
    - **Pre-marketing stage:** Each drug is investigated by clinical trials.
    - **Post-marketing stage:** The surveillance continues by self-reporting from pharmaceutical companies, hospitals, and consumers.
  - FEARS is restricted by its passive nature and only covers a fraction of knowledge available.
  - The insufficiency of ADR knowledge, is addressed by intelligently extracting new insights to understand drugs' side effects from patient centric online communities on social media
  - As per Hazell & Shakir, 2006; Ji et al., 2011, estimated reporting rate of ADRs in the existing system < 10%
  - As per Harpaz et al., 2012, results in > 2 million injuries, hospitalizations, and deaths per year in the US alone and an increase in costs estimated at \$75 billion annually
- The three most commonly used information extraction techniques are
  - text classification, lexicon-based entity recognition and ADR relation extraction.

# Outcome of Literature Survey

- traditional statistical methods are insufficient
  - regression models or data mining approaches are used to discover the pattern of ADRs
  - Data extraction and classification approaches are used to automatically obtain meaningful ADR information
  - However, ADR information stored in OHC is rarely discovered, analyzed and visualized
  - We need
    - automatic information extraction,
    - real-time text analytics
    - and automatic topic modeling on forum and social media data.
- Data visualisation is not emphasized
  - The developed data mining frameworks lack the capability to present the results in a visual, interactive, and real-time way.
  - Therefore, it is difficult for healthcare practitioners to interpret the results from social media to obtain new insights into ADR surveillance

# What does the paper propose?

- This paper proposes an automatic and real-time social media analytics framework with interactive visualizations to support effective exploration of knowledge about ADR surveillance.
- A case study on a longitudinal diabetes patient social media platform
  - AskaPatient.com
  - FDA Adverse Event Reporting Systems (FAERS) data
- The dataset from these sources are used as a benchmark to evaluate author's approach performance.
- The contents provided from patient drug experiences of a particular diabetes drug with respect to time, gender of patients, age group of patients and associations are analysed

# Research methodology

- Data extraction

- To understand the ADRs, The author collected patient-provided comments on a specific diabetes drug (i.e., Glucophage) from the FAERS and online health community
- FEARS
  - The FAERS datasets are available for the public form.
  - Each data point is an individual report of an adverse drug event.
- AskaPatient.com
  - participants share their opinions and experiences about specific diseases with respect to diagnosis, symptom management, drug usage and treatment as well as a five-point scale rating.
- Preliminary data cleaning was performed using python programming language with cleaned historical data as reference.

# Research methodology

- Data analysis

- Data analysis tools were used to process all kinds of data
  - Ratings
  - Comments
- The Askapatient.com dataset was cleaned up and modeled to extract ADRs at this stage.
  - NLP used for data processing and manipulations.
  - the python **pattern** module developed by the CLiPS was used in this phase.
- Step 1: stop words removal and text normalization
  - comments can be colloquial and the words have no info. ( stop words )
  - data provided by users on Askapatient.com is text, so text normalization is performed
- Step 2: text tokenization and information gain model
  - The text is tokenized into 3g.
  - After this tokenization, the information gain model is applied.

# Research methodology

- Data Analysis

- Step 3: information extraction for the top ADRs
  - Based on the information gain that each n gram provides in accurately predicting the ratings given by Askapatient.com
    - rating 1 [very dissatisfied with the medicine] and 5 [very satisfied]),
    - the presence of these top ADRs in particular reviews are captured by forming binary flags for each of these side effects.
    - If the particular side effect is present in the review, then the corresponding binary flag in the data is updated as 1, else 0.

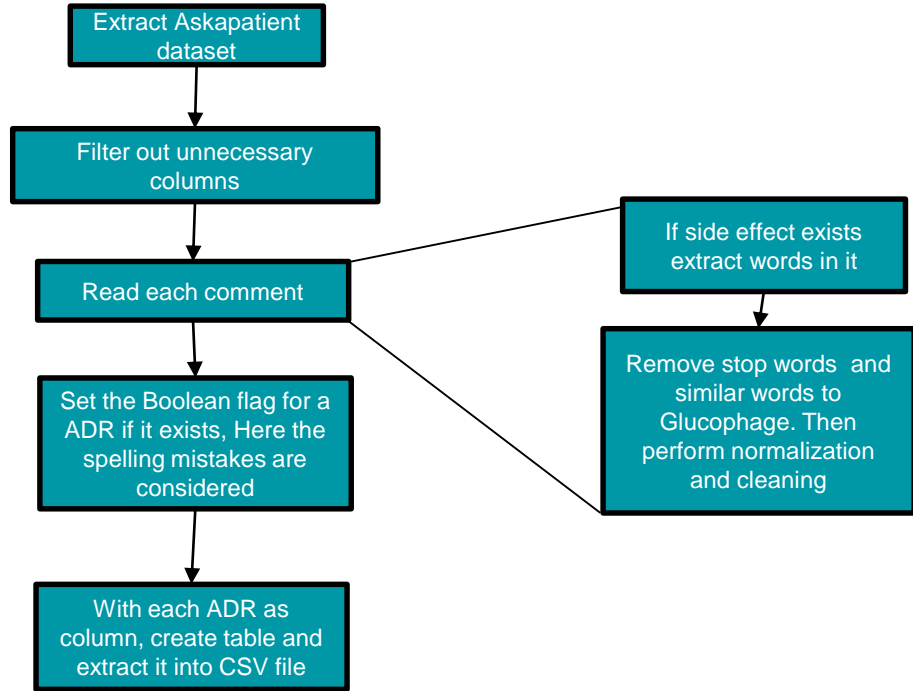
- Data visualization

- The iterative visualizations page is made possible by Tableau,
- Interactive visualizations are made from the different data sets combined.

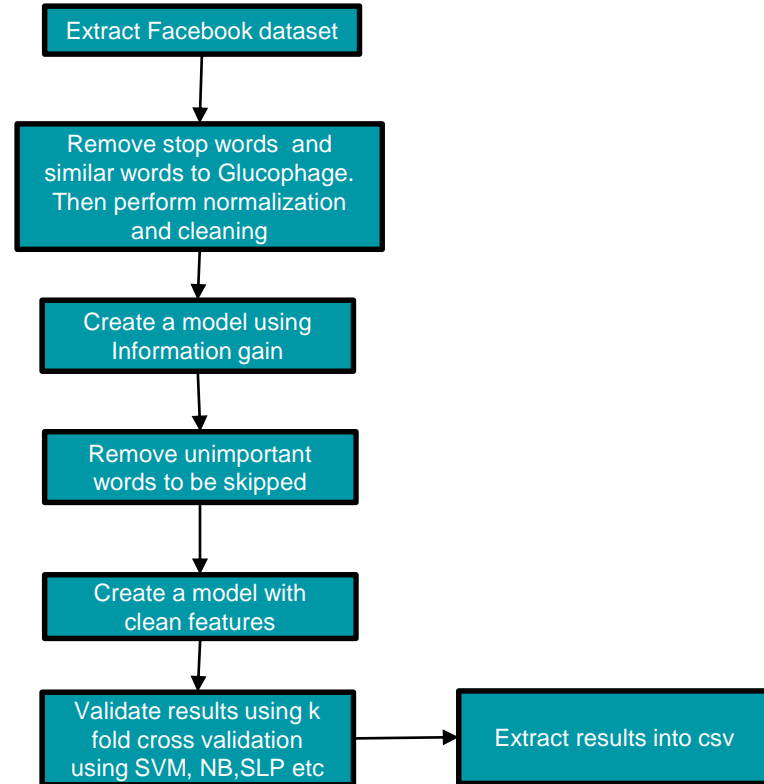


# Flowchart

## Askapatient.com OHC

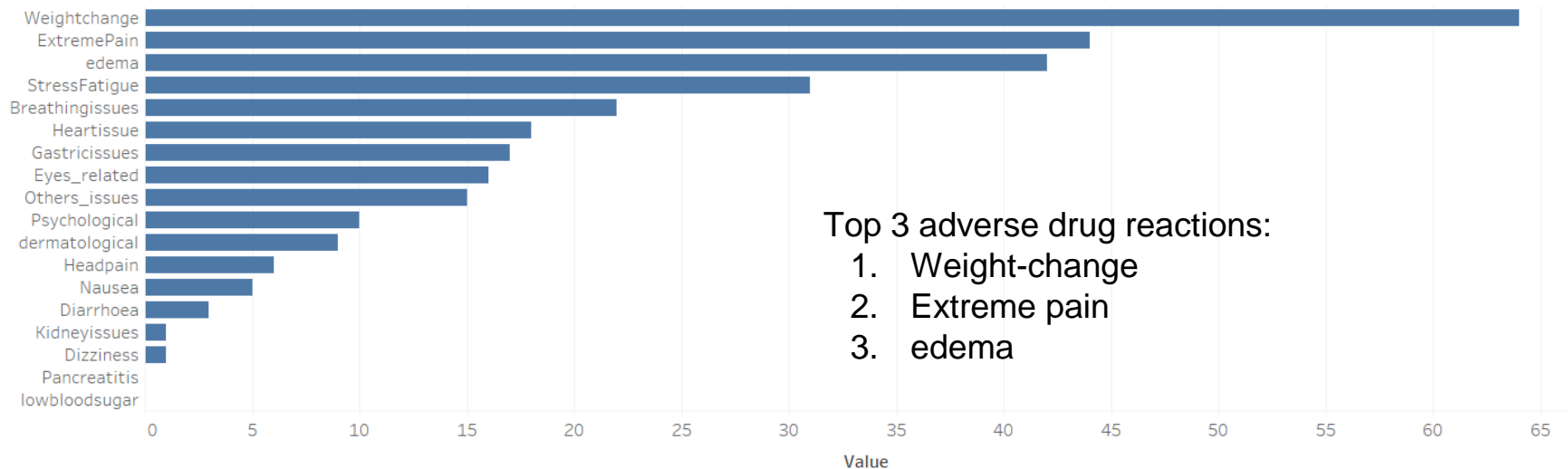


## Facebook



Implementation specifics [code]

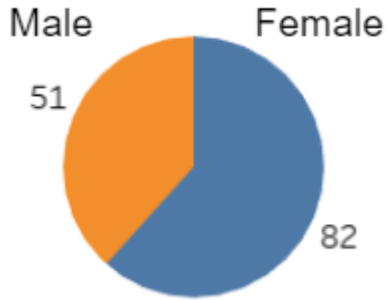
# Results - Askapatient ADR for Glucophage drug



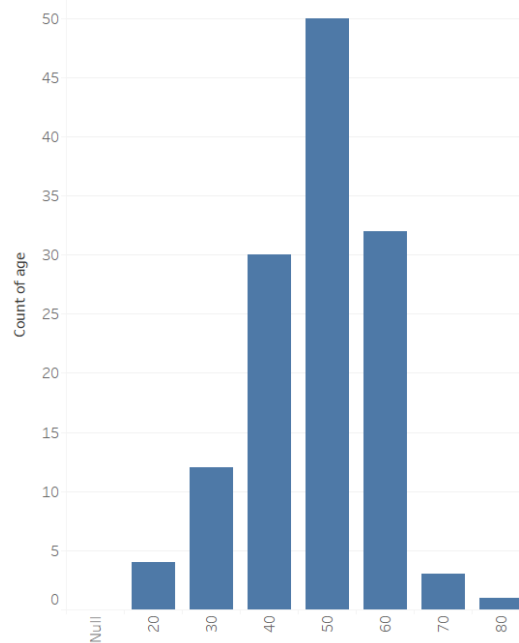
Low other issues means whatever adverse drug reactions that were considered

# Results - Askapatient ADR for Glucophage drug

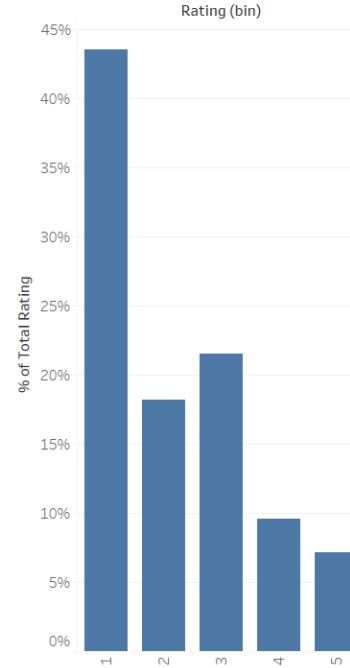
Gender analysis



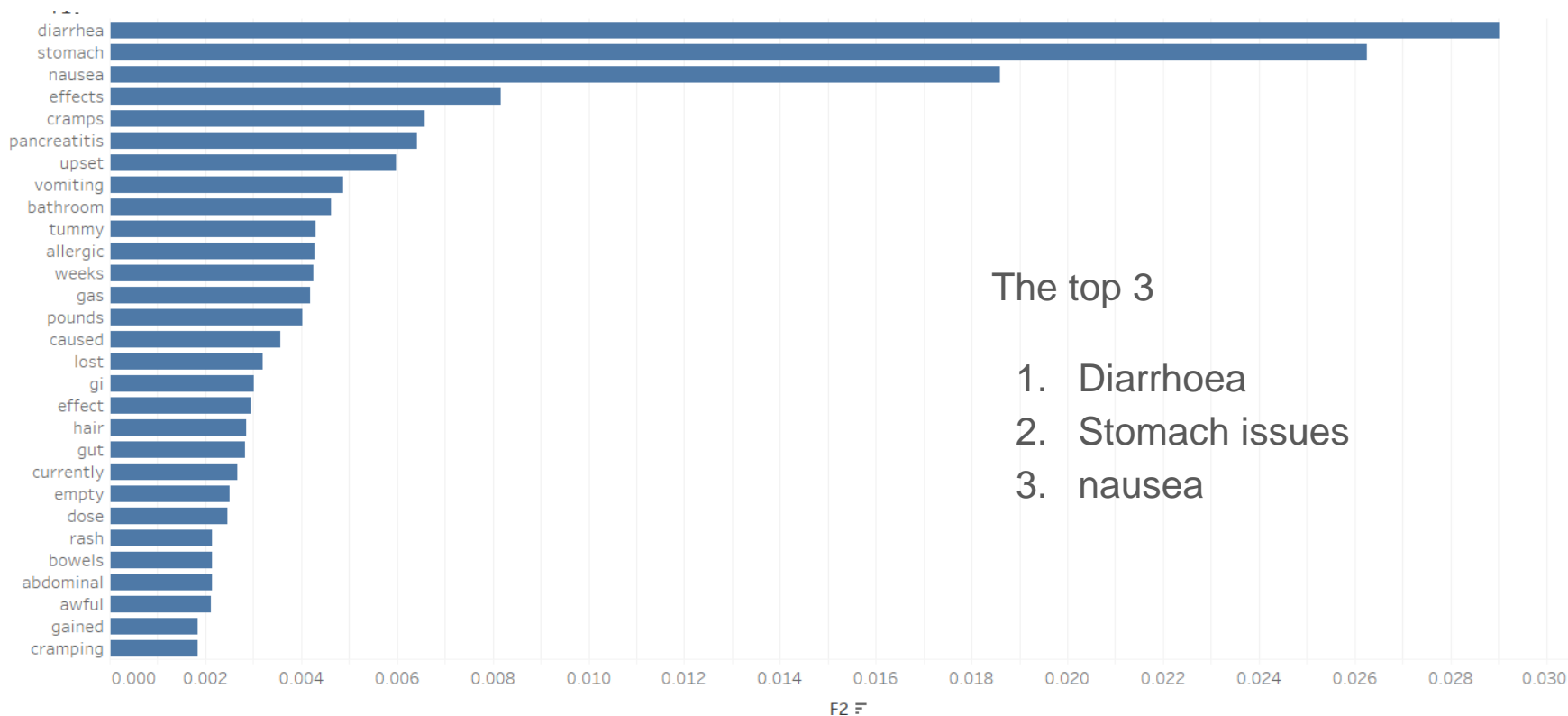
Age analysis



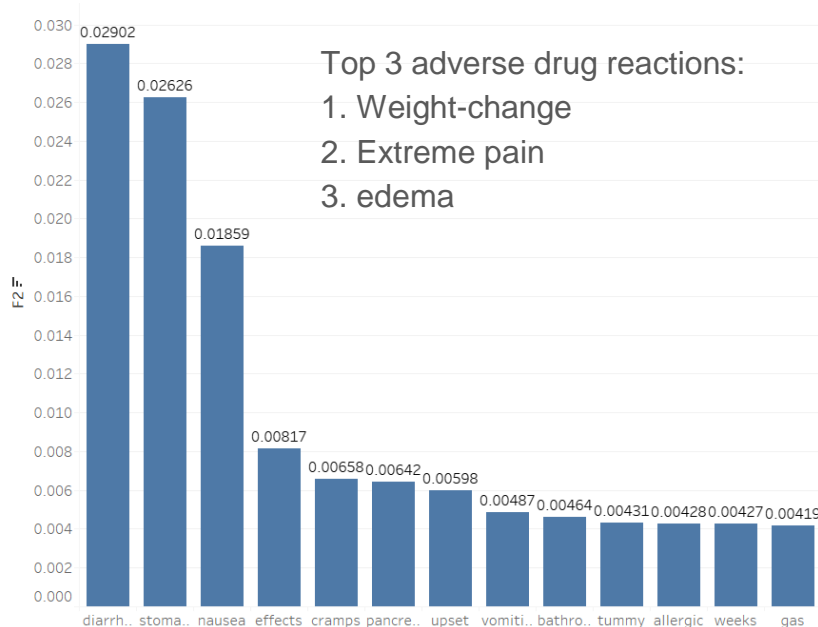
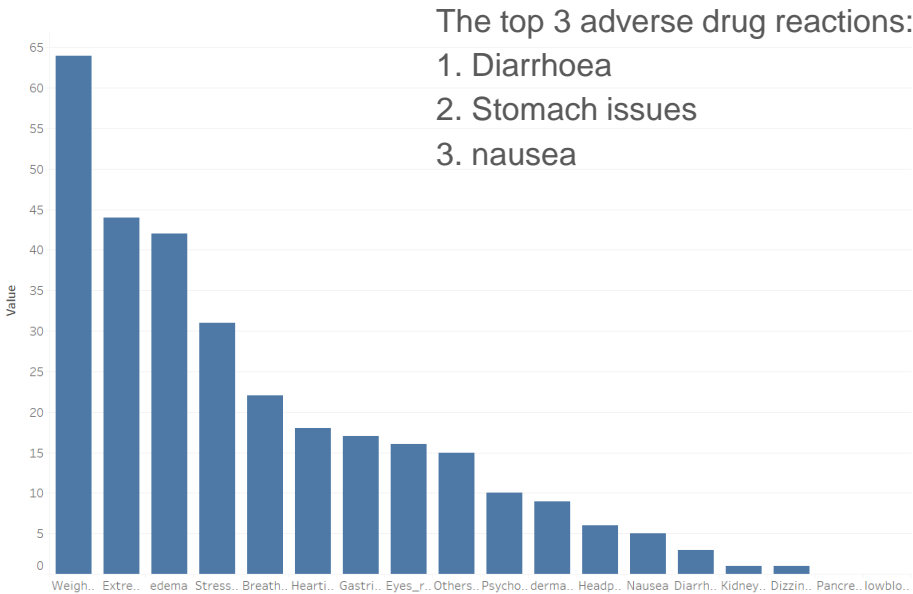
Rating analysis



# Results - facebook dataset analysis



# Result Comparison



# Conclusion and future work

- In this work, we have demonstrated that we can get different adverse drug reaction insights from social media analysis compared to Askapatient online health community dataset analysis.
- This will help doctors and health experts to answer important questions in the domain of knowledge discovery
- The work can be applied to any medicine to discover its side effects, associations and patterns, predictions and forecasting.
- Future work is to explore better visualization techniques that can provide better insights of the healthcare data.

Thank you!