UNIVERSITÀ DI MILANO-BICOCCA

Drug reviews: an NLP approach

By Greta Gravina and Niccolò Rocchi

Text Mining and Search, Final Project Presentation

Overview

- INTRODUCTION
- OBJECTIVES
- RELATED WORK
- DATA PREPARATION
- TOPIC MODELLING
- EVALUATION
- SENTIMENT ANALYSIS
- EVALUATION
- CONCLUSIONS

Intro duction

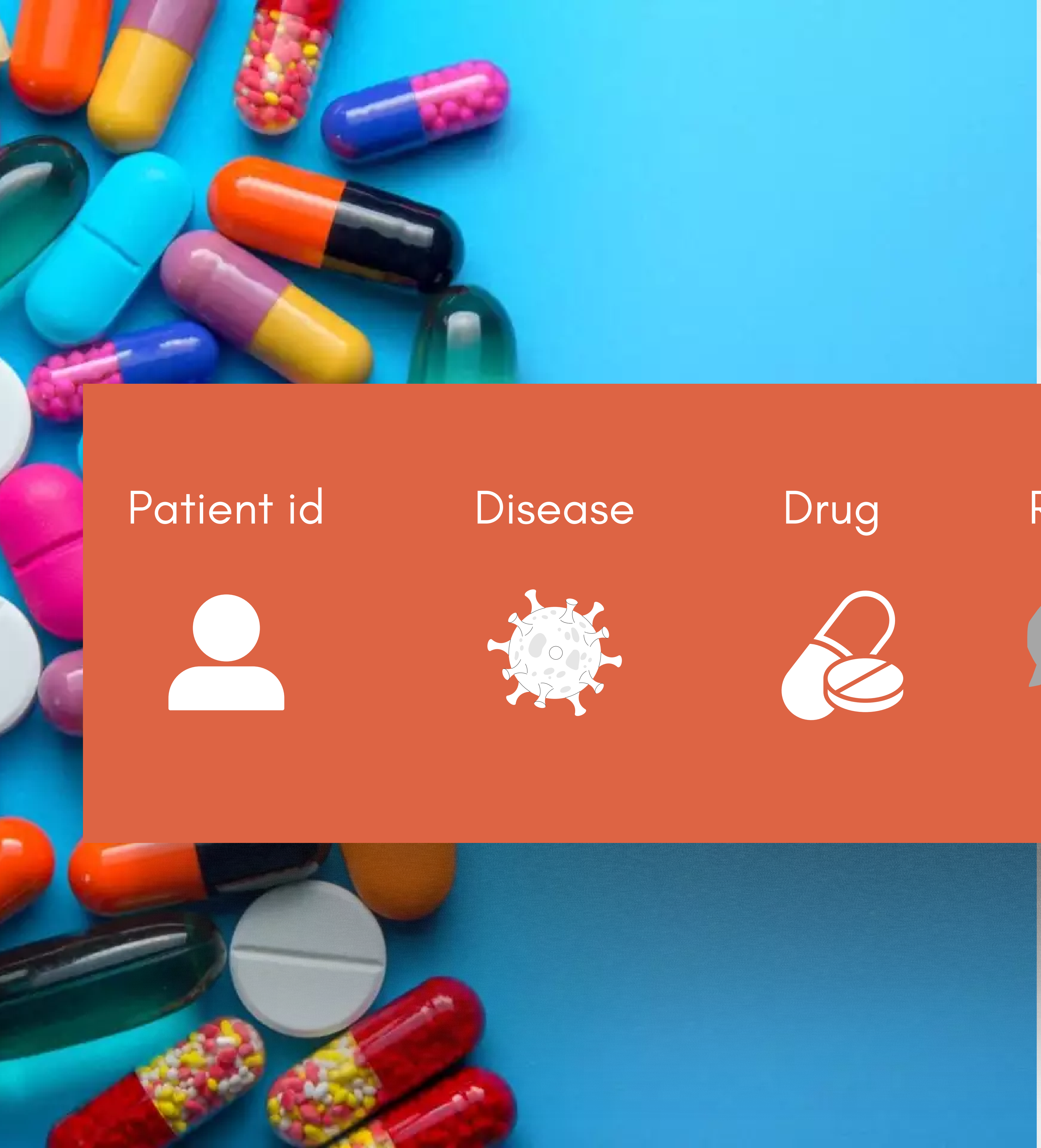
NLP aims to model human language

Central for companies:

- Recommender Systems
- Decision making
- Clients insights

In this work we mine the **Drug Review dataset** (Drugs.com) in order to discover patterns in patients opinions

Dataset structure



Patient id	Disease	Drug	Review	Rating	Date

Objectives

Research questions:

1) " What could drug companies do for patients? "

2) " How patients feel about drugs they take? "

01

Find useful patterns in reviews

that may help future decisions in companies and patients health

02

Evaluate discrimination

between good and negative opinions, accounting for data unbalancing

Approaches:

Topic Modeling: unsupervised learning to discover topics and their correlation with ratings

Sentiment Analysis: supervised classification task based to identify emotional states

Related work

He at al., 2020

"The Voice of Drug Consumers: Onoine Textual Review Analysis Using Structural Topic Model"

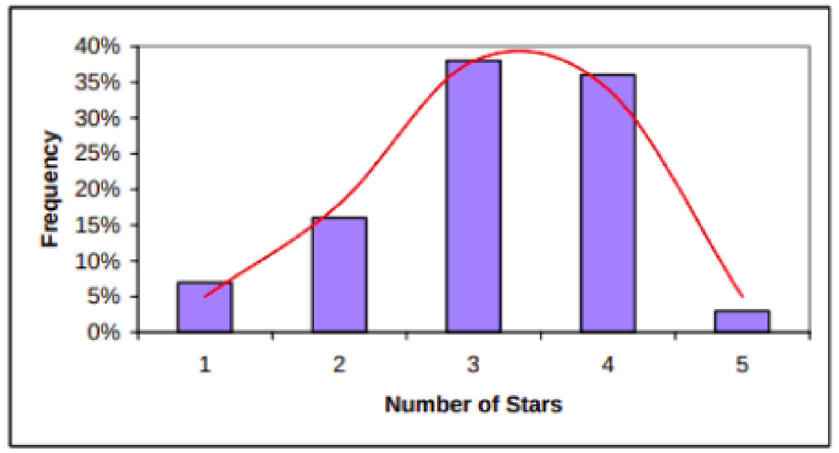
Uddin et al., 2022

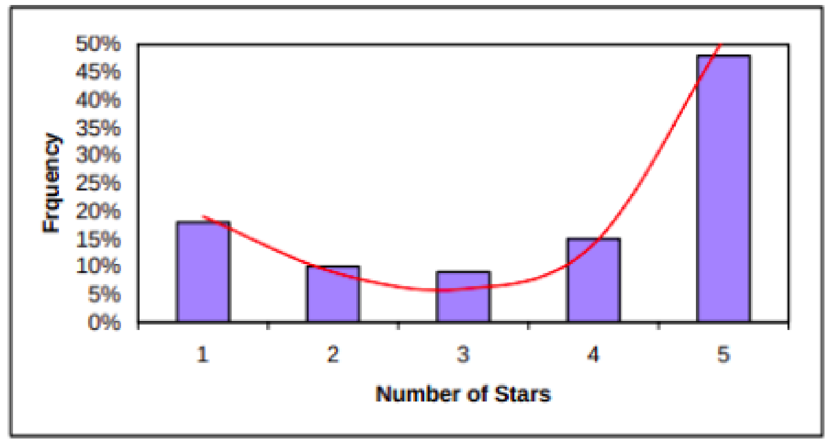
"Drug Sentiment Analysis using Machine Learning Classifiers"

Data preparation

Problem: J-shaped review distribution.

Called "*Under reporting*" selection bias





Approach:

1. Filter out central reviews (5 to 8)
2. Matching drug names with their related drug class (scraping strategy)
3. Keep the classes with more than 1000 review (14 classes)
4. Balance positive - negative reviews, conditionally to classes

TOPIC MODELING

Pre-processing

Use of: Python Natural Language Toolkit (*nltk*) and *gensim* library
Operations: Lower case transformation, Tokenization, stop word removal, stemming with Porter Stemmer, creation of a dictionary with absolute frequencies, filter of common and rare words (thresholds: 5% and 55%, creation of BOW matrix

LDA and STM

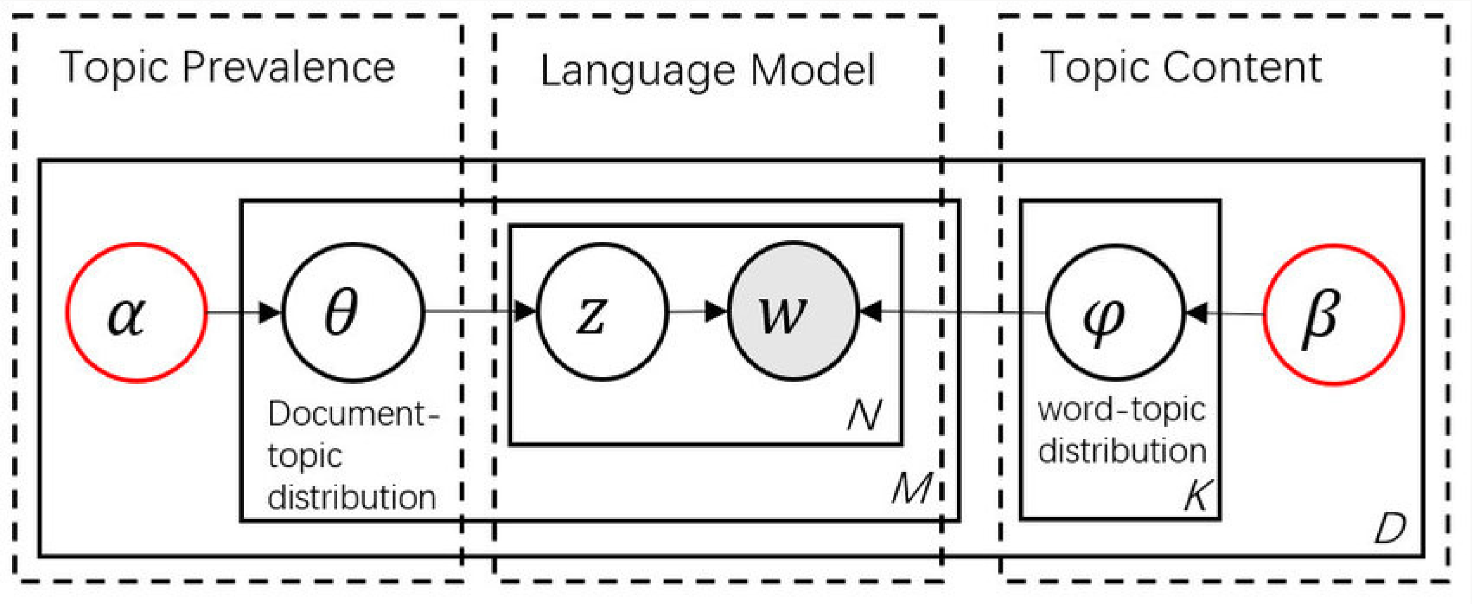
Use of: *gensim* Python library and *stm* R package.
LDA features: text
STM features: text, ratings, date, drug class

Evaluation

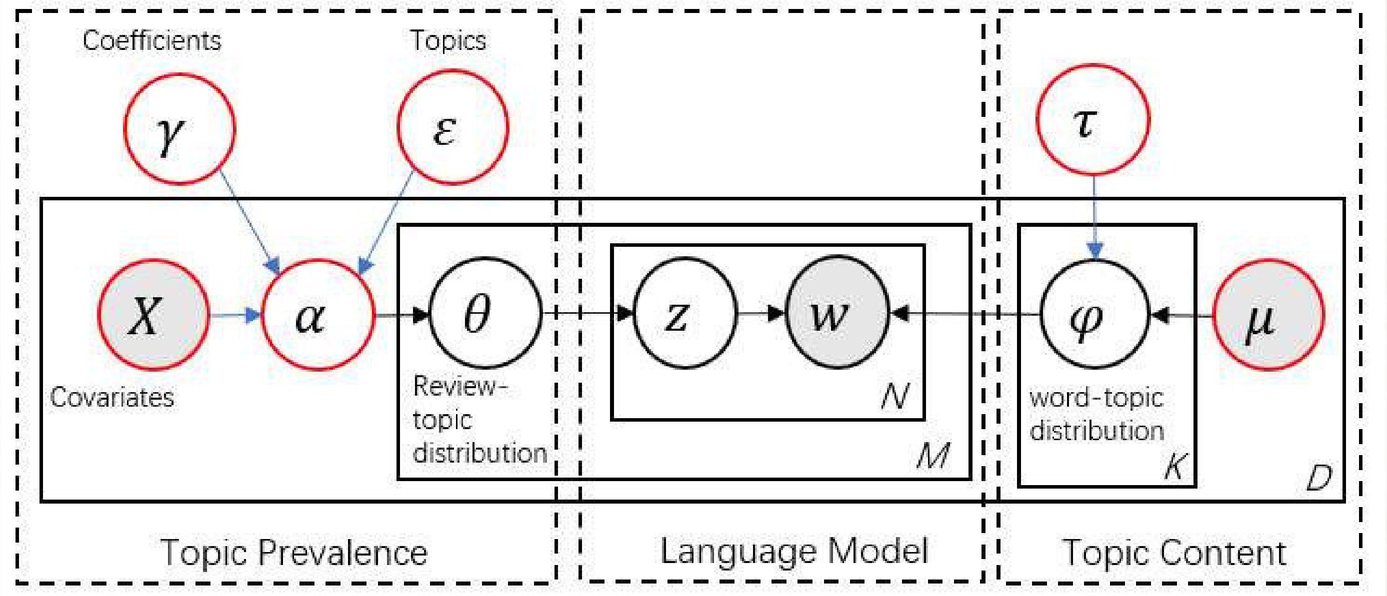
Different measures, both quantitive and qualitative. Importance of assessing information power and topic interpretability

A focus

LDA model



STM model



- Each model: trained with **different numbers** of a priori topics.
- Best model: choosen on the basis of the "best" coherence with **elbow method**
- *CV Coherence* [Syed et al., 2017] for LDA
- *Semantic Coherence* [Mimno et al., 2011] for STM

Evaluation

Measures:

- Coherence (quantitative)
- Division (qualitative for LDA, quantitative for STM)
- Topic interpretability (qualitative)
- Capability to answer to research questions (qualitative)

Example: assessment of division for LDA through a t-SNE chart

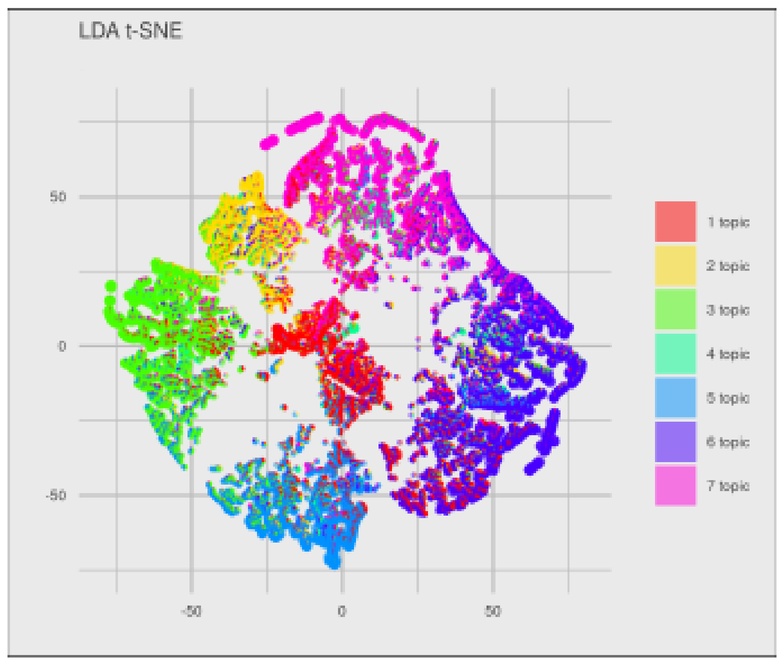


Figure 3: *t-SNE of LDA model*

Examples of topics and relations with positive/negative ratings:

LDA:

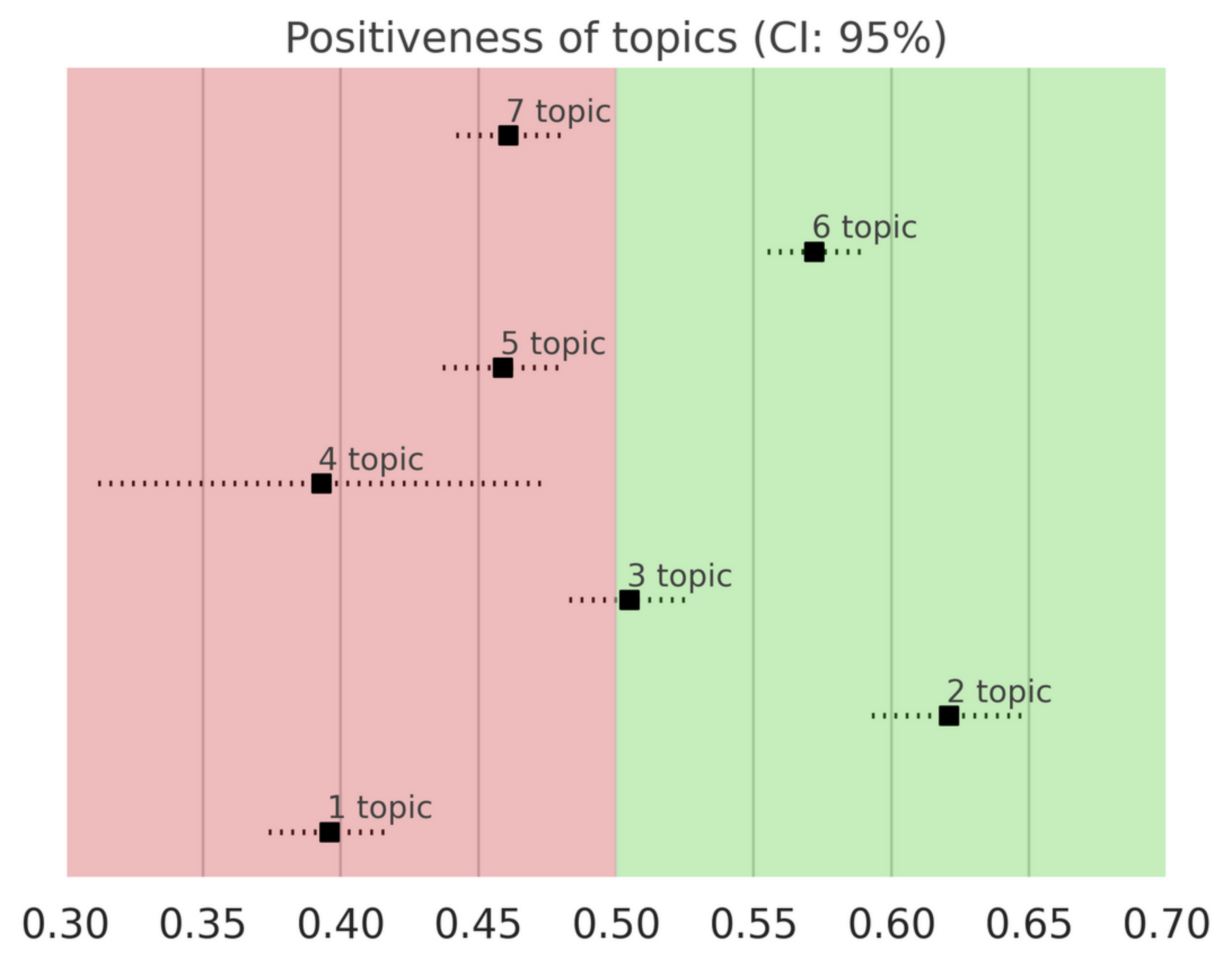
[2] Side effects: positively related with contraceptives but negatively with psychiatry
[5] Gaining weigth in antidepressants, negative ratings
[7] Economic cost of drugs. Negative ratings among contraceptives

STM:

[1] Panics attacks: it's less discussed in last years
[3] Side effects: hot topic in last years
[7] Induced sleep by anxiolytics

Evaluation

Example of assessment of the capability to answer to research questions (LDA)



Final Results:

- Both model were choosen based on "best" coherence
- Division of LDA (t-SNE) is satisfactory
- LDA obtains the best topic interpretability and informations
- STM is difficult to interpret: an R shiny app is indeed available
- STM topic prevalence and content don't give us additional insights

SENTIMENT ANALYSIS

Pre-processing

Use of: Python Natural Language Toolkit (*nltk*)
Operations: Lower case transformation, removing special characters, replacing two or more dots/two spaces, stop word removal, stemming with Snowball Stemmer, TextBlob for polarity score, new feature engineering, Label Encoder

Algorithms

Use of: Python library for ML
LightGBM: tree-based, pattern
Logistic Regression: predicting binary outcomes
Gaussian Naive Bayes: features, Gaussian distribution

Evaluation

Accuracy and ROC curves for evaluate the model performance on test set

Processing

Pre- processing steps to:

- Capture the complex nuances of language
- Provide a more accurate Aentiment Analysis

01

Cleaning reviews

- Cleaning reviews with specific pre-processing steps and use SnowBall stemmer

02

TextBlob

- Polarity Score [-1, 1] both for the cleaned reviews and the ones without removing stop words
- TextBlob vs VADER

03

Feature engineering

- Creating new features that are more relevant and informative to improve the accuracy of models.

04

Label Encoder

- Changing the categorical values of Drug Names and the Conditions into numerical values for ML modeling.

A focus

LightGBM

- A gradient boosting framework that uses treebased learning algorithms
- Number of estimators (10.000)
- Learning rate (0.10)
- Number of leaves (30)
- Classifier is used to predict the sentiment of new drug reviews based on the patterns

Logistic Regression

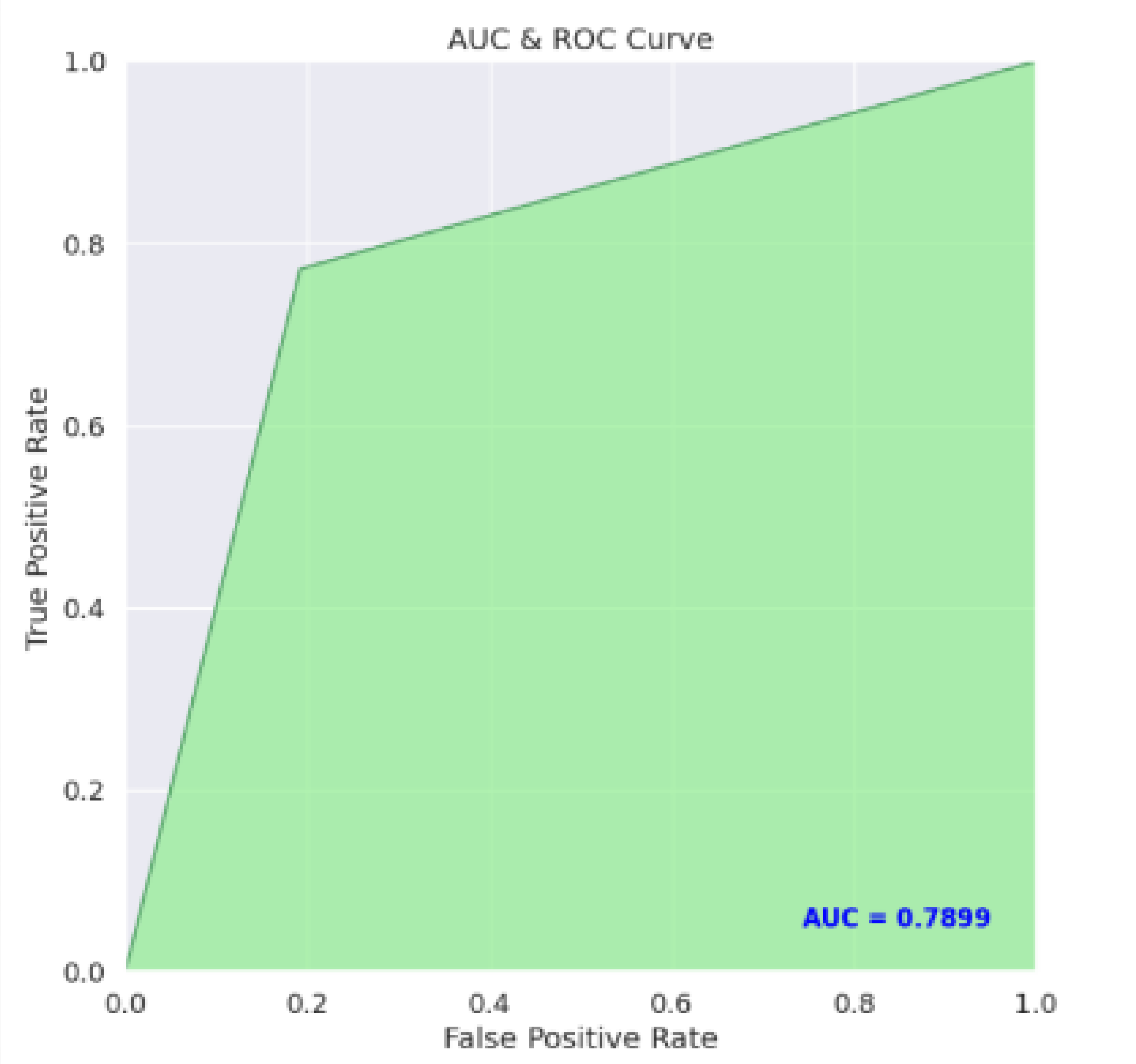
- Predicting binary outcomes [0,1]
- "liblinear" solver uses a modified version of the Coordinate Descent algorithm

Gaussian Naive Bayes

- Variant of the Naive Bayes algorithm
- Each feature follows a Gaussian distribution, and that the features are independent of one another

Evaluation

Example of AUC and ROC curve (LightGBM):



Final Results:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- AUC and ROC curves for performance evaluating of binary classifiers
- LightGBM. is the model with the highest accuracy value (0.8) and AUC nearest to 1.0 (0.8)

CONCLUSIONS

1

Topic Modelling allowed us to **discriminate topics** and also **discover patterns** with a **different granularity level**. LDA performs better in giving useful insights, while STM seems to give more general information.

2

Sentiment Analysis have revealed to **quantify the emotions** expressed by the features using a **polarity measure.** LightGBM seemed to be the best model in classifying new reviews using patients sentiment score.

Thank You