THE GOAL OF THE PROJECT

- <u>Explore</u> the hidden themes behind IMDB movie synopses, originating from Israel.
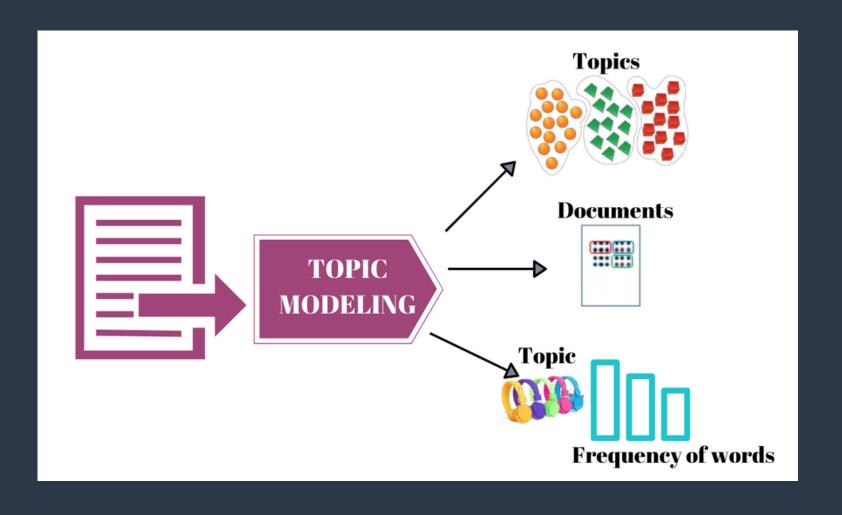- <u>Topic modeling</u> method application - discover topics from corpus.
- Gain some <u>insight</u> into movie topic trends.
- <u>Discover</u> interesting dynamics within the movies created by the Israeli film industry.

TOPIC MODELING

" ... topic model is a type of statistical model for discovering the abstract "topics" that occur in a collection of documents.
Topic modeling is a frequently used text-mining tool for discovery of hidden semantic structures in a text body.
Intuitively, given that a document is about a particular topic, one would expect particular words to appear in the document more or less frequently ..."

Wikipedia

WORK PLAN

IMDB and Wikidata/Wikipedia APIs will assist us in gathering movies synopses datasets

perform text preprocessing - removal of new lines, quotes etc.

use topic modeling LDA tool & evaluate

Visualize evaluated data

EXPECTED FINAL PRODUCT

- Website

Will enable users to access our findings and contain all the information needed to understand and use the project

- Grafics
  - "bag of words" of the topics statistics
  - Movies correlation
  - Statistics regarding the difference between the genre and the topic revealed by the model