

Project Proposal

Student Name: Samahir Bantan

Question/need:

*** What is the framing question of your analysis, or the purpose of the model/system you plan to build?**

Movie rating is one of the most characteristics that attract the viewer's attention when choosing a movie. So, the goal of this project is to implement a model based on machine learning algorithm to classify the reviews based on the writer's way of highlighting his feelings and choosing the words that made the comment positive or negative.

*** Who benefits from exploring this question or building this model/system?**

General people and movies industries

Data Description:

*** What dataset(s) do you plan to use, and how will you obtain the data?**

Dataset -from Kaggle- contains reviews of a movie from IMDB that are labeled as positive/negative reviews.

*** What is an individual sample/unit of analysis in this project? What characteristics/features do you expect to work with?**

- Dataset contains 50000 rows and 2 columns
- Review: contains comment about movie
- Sentiment: is the label column (positive\negative)

*** If modeling, what will you predict as your target?**

Predict the movie review as a positive or negative based on comment

Tools:

*** How do you intend to meet the tools requirement of the project?**

- Data cleaning (check for missing data, plots for outlier)
- NLP and feature extraction
- EDA – not adapting this method due to data is clear
- Visualize data using matplotlib or seaborn
- Use CNN (convolutional neural network) algorithm

*** Are you planning in advance to need or use additional tools beyond those required?**

For now, I'm not planning to use other tools, but I might use if necessary.