Opinion mining on Movie reviews

**Abstract:**

Opinion mining also referred as sentiment analysis, an approach to NLP, which identifies the emotion behind the body of text. It also involves text mining to identify the keywords and the emotion behind the text. In this project, we are taking IMDB dataset of 50k movie reviews and they are labelled to positive and negative. Based on the reviews given by the customers over movies, we will use TF-IDF Classification and Multinomial Naïve bias Classification.

**Technologies Involved:**

1. **Cloud Tools:** AWS

2. **IDE:** Jupyter Notebook

3. **Mobile/client-side integration:** Web-App using Flask

**Learning Goals:**

* Knowledge on different types of techniques and difference between techniques for cleaning text data
* Knowledge of building web app using flask and ability to create the frond end
* Learning the deployment of model in aws platform

**Data:**

We are taking the dataset from Kaggle, and we have data of about 50k movie reviews and we must label them positive and negative. By using opinion mining, first the dataset should be preprocessed and then the attribute selection (bag of words), called initialization, then we should train our model with the algorithm to identify the emotion over the body of text.

**Team Members:**

|  |  |  |
| --- | --- | --- |
| **Name** | **Mail ID** | **Role** |
| Rajesh Nemani | [rajeshnemani@my.unt.edu](mailto:rajeshnemani@my.unt.edu) | WebApp Design, Deployment, Data Gathering |
| Tejeeswar Kommineni | [tejeeswarkommineni@my.unt.edu](mailto:tejeeswarkommineni@my.unt.edu) | Feature detection, Data Preparation, Training Model |
| Anuhyasiri Kondepati | [anuhyasirikondepati@my.unt.edu](mailto:anuhyasirikondepati@my.unt.edu) | Training model,  Word Cloud,  Documentation |
| Balaji Narravula | [balajinarravula@my.unt.edu](mailto:balajinarravula@my.unt.edu) | Documentation,  WebApp Design |

**Milestones:**

* Preprocessing the data.
* Building the model to classify the reviews
* Building web application and integrating the model
* Deploying Web ML model in AWS

**Exploratory and Extensible:**

The project can be extended to add one more label called neutral, and we can train the data to make neutral reviews also.

**GitHub Repo:**

<https://github.com/rnemani96/CSCE_5214_SDAI.git>

**Collaboration Proposal:**

**Meeting Times:** Online Meeting weekly on Wednesday, In Person in Wills Library on Saturday

**References:**

* AWS Documentation: [AWS Documentation](https://docs.aws.amazon.com/index.html)
* Kaggle Dataset: <https://www.kaggle.com/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews>
* WordCloud: <https://www.datacamp.com/community/tutorials/wordcloud-python>