Team Members:

1. محمد عبد الله السعيد عبد المجيد
2. محمد عادل عبد الرحمن
3. عمر صلاح لطفي سليمان
4. امير حسني محمد صالح
5. مجدي عادل شحاتة علي
6. محمود خالد سيد محمد

Sentiment Analysis of movie reviews

Team ID:29

SENTIMENT ANALYSIS OF MOVIE REVIEWS

Introduction:

/\* Discuss in brief the project main idea and the objectives\*/

Determine whether the review about movie is good or bad

Methodology:

/\*Discuss the methodologies used in order to fulfil your objectives (i.e. The feature sets and the models implemented) \*/

1. **Data preprocessing**:

* Delete all symbols that does not affect the meaning.
* Convert all reviews to lowercase letters,
* Delete all stop words, tokenize, and stemmed all reviews.

1. **Features extraction**:

* Used Count Vectorizer to get numeric features can use it in training.

1. **Model training and testing**:

* Split data into 70% training and 30% testing
* Train it with linear SVC.

1. **Results visualization**:

* Plot accuracy of trained models.
* Plot Confusion Matrix for SVC Model

Data Set Summary:

Answer the following Questions:

1-What is the data set used?

Ans: **Sentiment polarity - movie review data.**

2- What is the summary of the dataset columns?

Ans: **Consists of 1000 reviews for each positive and negative Class.**

3- Visualize the dataset statistics

Ans:

Chart, bar chart

Description automatically generated

Figure 1. No. of samples in training and testing compared to size of dataBar chart

Description automatically generated with low confidence

Figure 2. No. of reviews in positive and negative class

Results:

/\*Use suitable graphs to visualize your models results\*/

Figure 3. Relation between Accuracy and No. of training samples

Chart

Description automatically generated

Figure 4. Relation between Trained models and their accuracy

Chart, treemap chart

Description automatically generated