Mini-Project

Due time (11 May 2024)

**Arabic Sentiment Analysis of Hotel Reviews**

# Objective:

o The goal of this project is to build and evaluate 3 models for sentiment analysis of movie reviews using Naive Bayes, Logistic Regression, and neural network. You may use python libraries: Numpy, pandas, sklearn, keras, NLTK

# Data:

* The dataset you will use for this project is a collection of text reviews and their corresponding sentiment labels (positive or negative). The review are in Arabic Language.

# Method:

* + - **Data preprocessing:**
      * Import the dataset and perform basic data cleaning and preprocessing. This includes removing stop words, punctuation, and stemming the text.
      * Split the dataset into training and test sets.

# Feature extraction:

* + - * Convert the text data into numerical features using techniques such as **bag of words** or **TF-IDF**.

# Model building:

* + - * Train a Naive Bayes classifier on the extracted features using a training algorithm provided by NLTK library.
      * Train a logistic regression model on the extracted features using a suitable training algorithm.
      * Train a simple neural network model with one hidden layer on the extracted features using a suitable training algorithm.

o

# Model evaluation:

* + - * Evaluate the performance of the trained Naive Bayes model on the test dataset using metrics such as accuracy, precision, recall, and F1 score.
      * Evaluate the performance of the trained logistic regression model on the test dataset using metrics such as accuracy, precision, recall, and F1 score.
      * Evaluate the performance of the trained neural network model on the test dataset using metrics such as accuracy, precision, recall, and F1 score.

# Model Comparison:

* + - * Compare the performance of the three models and discuss which one performed better and why.

# Deliverables:

* + - * A report detailing your approach and findings, including code and relevant visualizations.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Feat Ext. Method | Acc | Prec | Recall | F1 |
| NB | BoW | 0.803 | 0.806 | 0.803 | 0.802 |
| TFIDF | 0.811 | 0.817 | 0.811 | 0.8103 |
| LR | BoW | 0.819 | 0.8194 | 0.8195 | 0.8194 |
| TFIDF | 0.8295 | 0.8296 | 0.8296 | 0.8294 |
| NN | BoW | 0.8108 | 0.8114 | 0.811 | 0.8107 |
| TFIDF | 0.8065 | 0.8067 | 0.8066 | 0.8064 |

NN

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Epochs | Acc | Prec | Reca | F1 |  |
| 10 | 0.828 | 0.8285 | 0.82823 | 0.8279 |  |
| 15 | 0.811 | 0.8128 | 0.8114 | 0.8108 |  |
| 20 | 0.8106 | 0.8106 | 0.8106 | 0.8106 |  |
| 30 | 0.7935 | 0.7936 | 0.7936 | 0.7934 |  |