**Sentiment Analysis of Restaurant Reviews to Improve Customer Satisfaction**

**Abstract**

In the current era, social media networks, such as Twitter, Facebook, and Snapchat, have become popular means of communication between people to share ideas and emotions. The use of these internet tools has changed the way people communicate and the way customers interact with organizations. To serve customers better, it is imperative for businesses to pay attention to the impressions and experiences about their products and services that customers are making available through these communication tools. This Project investigated Sentiment analysis for build a model that predicts the sentiments of customer reviews regarding our restaurant.

1. **Dataset**

The dataset consists of 8588 instances with a large quantity of features since it is in textual form. You can find the dataset at the following link: https://github.com/hadyelsahar/large-arabic-sentiment-analysis-resouces/tree/master/datasets.

1. **Design**

Sentiment analysis is a powerful marketing technique that allows product managers to better understand customer attitudes and use that information in their marketing initiatives. The main objective is measuring the extent of customer satisfaction with the services provided through machine learning techniques and provide appropriate recommendations to improve performance and better serve the beneficiaries. In general, our system consists of three stages, as presented in previous Figure .

Diagram

Description automatically generated

Figure 1: Proposed approach phases

1. **Preprocessing**

Preprocessing is an important stage; it is used for preparing the data set for model construction. .**First** , normalization is applied to make characters more consistent. **Second**, tokenization splits the textual data into a sequence of tokens and removes spaces so that each word is separated by only one white space.to transform it into a suitable form for processing. **Third,** filteringis used to remove all non-alphabetic characters, especially signs frequently used on Twitter, such as (#) for hashtags and (@) for usernames. Moreover, stop word filtering and numbers and letters written in other languages must be removed. **Forth ,** ituses a stemming step to mitigate language complexities.

1. **Construct model(Algorithms)**

**Table

Description automatically generated**The collected data set is divided into a training set and a testing set. Machine learning algorithms use training data sets to construct a model used to classify new tweets while the testing set is used to evaluate the constructed model. In this project, I explored  the construction of several models to determine the sentiments of customer reviews in an automated way by using support vector machines (SVM) , Naïve Bayes (NB) and logistic regression(LR) with TF-IDF to extract words importance. Furthermore, deep learning models are investigated to show their performance with the AraVec pre-trained model and TF-IDF vectorization**. The next table** shows the performance of these algorithms in terms of accuracy, recall, and precision. **The results of SVM reflected a superior performance with 83.38% accuracy, compared to poorer performances by NB with 73.8%.** For deep learning methods, the best performance was achieved by LSTM with TF-IDF. However, their results are lower than traditional machine learning approaches. This may be because we used a small dataset.

Table 1: Accuracy, recall and precision of machine learning

1. **Analysis and visualization(Communication)**

To better grasp and utilize the data, it must be analyzed and visualized in a simplified manner to help managers draw conclusions and make the correct decisions for their products. First, we used a word cloud to show the most words that were present in each class, as in the following figure.

**Qr code

Description automatically generated**

Figure 2:Common words in positiva class

Second ,for each of machine learning algorithms ,we used heat map to illustrate confusion matrix as presented in next figure.

Chart, treemap chart

Description automatically generated

Figure 3:Heat map of SVM algorithm

**Third, for each of the deep learning algorithms, we used a accuracy during various epochs to illustrate the confusion matrix as presented in the next figure.**

**Chart, line chart

Description automatically generated**

**3. Tools**

In this project, we will use several well-known libraries in python such as: Pandas, NumPy, Sklearn, Matploitlib, Keras and NLTK.