NAAN MUDHALVAN

ARTIFICIAL INTELLIGENCE

PROJECT TITLE

SENTIMENT ANALYSIS FOR MARKETING

REG NO: 712321104010
NAME: PREMANANTH.P
DEPT: COMPUTER SCIENCE & ENGINEERING
YEAR & SEM: 3rd & 5th
COLLEGE NAME: PARK COLLEGE OF ENGINEERING AND TECHNOLOGY

PHASE 4

TITTLE: selecting a machine learning language algorithm,training the model,and evaluating its performance.

MACHINE LANGUGAE FOR SENTIMENT ANALYSIS:

The supervised machine learning technique best suits sentiment analysis because it can train large data sets and provide robust results. It is preferable to semi-supervised and unsupervised methods because it relies on data labeled manually by humans so includes fewer errors.

EVALUATING AND SELECTING THE MI:

Introduction:

Machine learning has revolutionized various industries by enabling computers to learn from data and make predictions or decisions without explicit programming. However, with the multitude of machine learning algorithms available, it can be challenging to determine which one is best suited for a particular task. In this article, we will delve into the process of evaluating and selecting machine learning algorithms, empowering you to make informed decisions that drive optimal results.

1. Understanding Machine Learning Algorithms:

Before diving into the evaluation process, it is crucial to comprehend the different types of machine learning algorithms. **Supervised learning** algorithms learn from labelled training data to make predictions or classifications. **Unsupervised learning** algorithms identify patterns and structures in unlabeled data. **Semi-supervised learning** algorithms leverage a combination of labelled and unlabeled data for training.
Finally, **reinforcement learning** algorithms learn by interacting with an environment to maximize rewards.

2. Defining the Evaluation Criteria:

To assess the performance of machine learning algorithms, it is essential to establish evaluation criteria. These criteria typically include **accuracy, precision, recall, F1-score, training time, model complexity,** and **interpretability.** Accuracy represents the overall correctness of predictions, while precision and recall measure the algorithm's ability to minimize false positives and false negatives. The F1 score combines precision and recall, providing a balanced performance measure.

3. Preparing the Dataset:

A crucial step in algorithm evaluation is preparing a suitable dataset. The dataset should be **representative** of the problem at hand, **diverse** to capture various scenarios, and **sufficiently large** to ensure robustness. It is also crucial to **split** the dataset into training, validation, and testing sets. The training set is used to train the model, the validation set helps tune hyperparameters, and the testing set assesses the final performance.

4. Implementing the Algorithms:

To evaluate machine learning algorithms, you must implement them using a programming language or a machine learning framework. Popular choices include Python with libraries such as scikit-learn or TensorFlow. Implementing the algorithms involves **data preprocessing, feature engineering,** and **model training**. Data preprocessing includes handling missing values, scaling features, and encoding categorical variables. Feature engineering involves selecting relevant features or creating new ones to enhance model performance.

5. Performance Evaluation Techniques:

Several techniques are available to evaluate the performance of machine learning algorithms. **Cross-validation** divides the dataset into multiple subsets, allowing for more robust performance estimation. **Confusion matrices** provide a detailed breakdown of predicted and actual classes. **Learning curves** help identify underfitting or overfitting by plotting training and validation performance against the number of training examples. **Receiver Operating Characteristic (ROC) curves** measure the trade-off between a true positive rate and a false positive rate.

6. Comparing and Selecting Algorithms:

Once you have evaluated multiple algorithms, it is time to compare their performances and select the most appropriate one. Consider the evaluation criteria established earlier and weigh the strengths and weaknesses of each algorithm. Selecting an algorithm that strikes the right balance between performance, computational complexity, and interpretability is crucial. Remember that the best algorithm for one task may not be the best for another.

7. Iterative Process and Model Tuning:

Machine learning is an iterative process, and often, it is necessary to fine-tune the models to achieve optimal results. This involves adjusting hyperparameters, such as learning rate, regularization, or depth of decision trees. **Grid search** and **random search** are common techniques used to explore different hyperparameter combinations and find the optimal set. Regularization techniques like **L1 and L2 regularization** can be applied to prevent overfitting.

Conclusion:

Evaluating and selecting machine learning algorithms is a crucial step in building successful predictive models. By understanding the types of algorithms, defining evaluation criteria, preparing the dataset, implementing the algorithms, and employing appropriate evaluation techniques, you can make informed decisions. Remember that the iterative process of model tuning and fine-tuning is essential to achieve optimal results. By following these guidelines, you can leverage the power of machine learning algorithms to drive accurate predictions and unlock valuable insights in your domain.

Training the model of sentiment analysis:

To train the sentiment classifier, convert the words to word vectors using the pretrained word embedding emb. First remove the words that do not appear in the word embedding emb.

```
idx = ~isVocabularyWord(emb,data.Word);
data(idx,:) = [];
```

Set aside 10% of the words at random for testing.

```
numWords = size(data,1);
cvp = cvpartition(numWords,'HoldOut',0.1);
dataTrain = data(training(cvp),:);
dataTest = data(test(cvp),:);
```

Convert the words in the training data to word vectors using word2vec.

```
wordsTrain = dataTrain.Word;
XTrain = word2vec(emb,wordsTrain);
YTrain = dataTrain.Label;
```

Train Sentiment Classifier

Train a support vector machine (SVM) classifier which classifies word vectors into positive and negative categories.

```
mdl = fitcsvm(XTrain,YTrain);
```

Test Classifier

Convert the words in the test data to word vectors using word2vec.

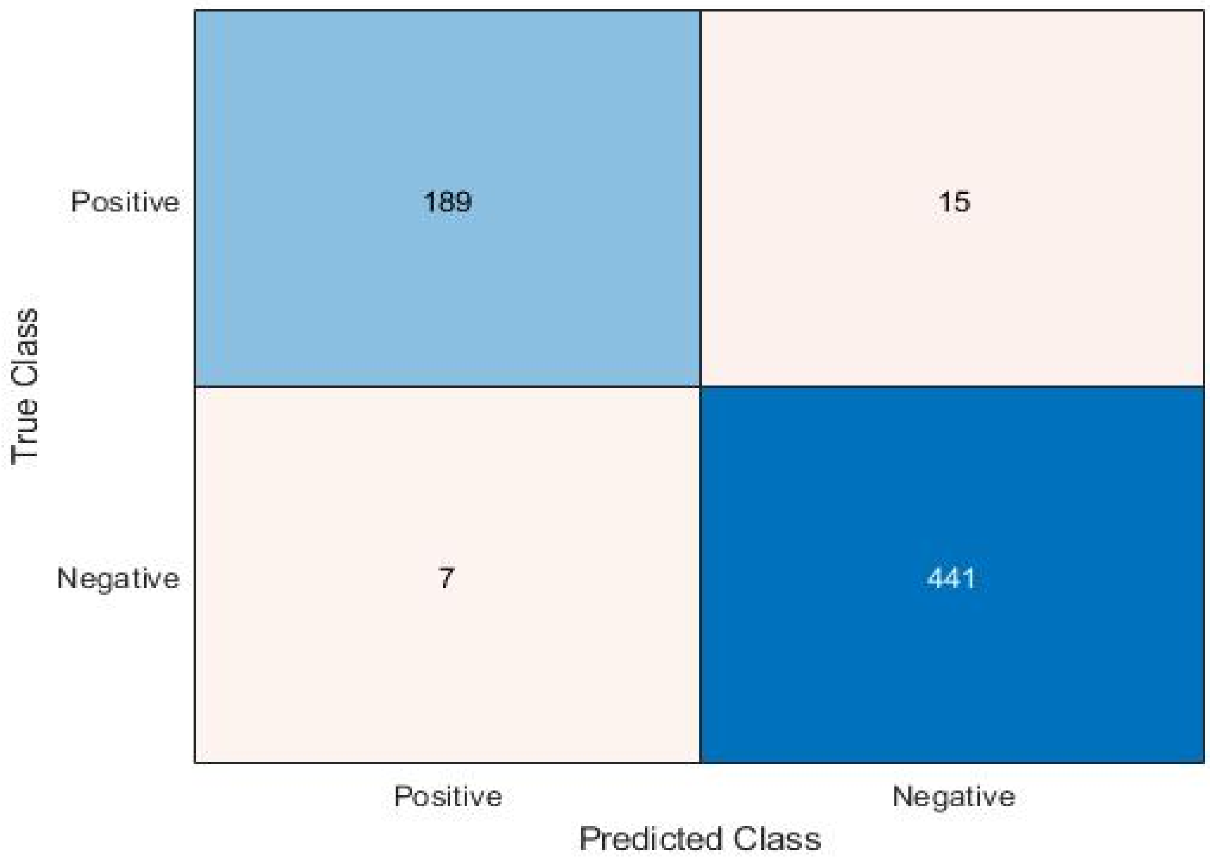
```
wordsTest = dataTest.Word;
XTest = word2vec(emb,wordsTest);
YTest = dataTest.Label;
```

Predict the sentiment labels of the test word vectors.

```
[YPred,scores] = predict(mdl,XTest);
```

Visualize the classification accuracy in a confusion matrix.

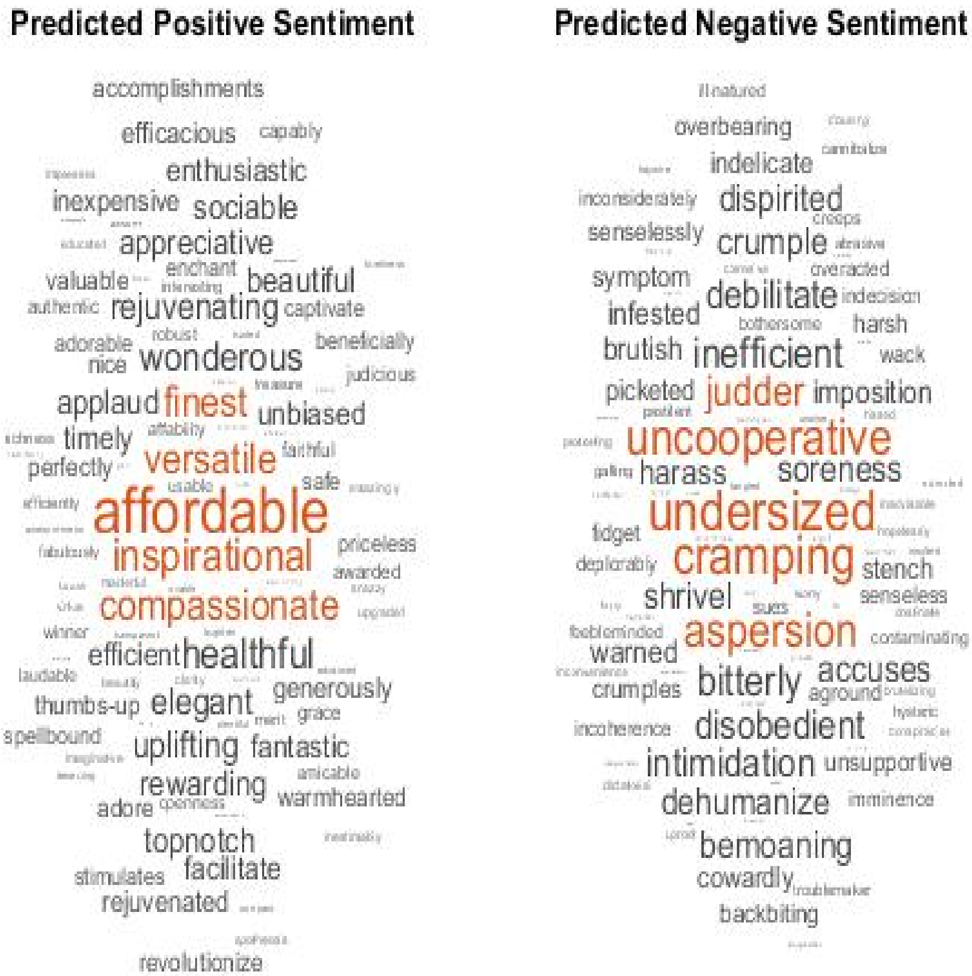
```
figure
confusionchart(YTest,YPred);
```



Visualize the classifications in word clouds. Plot the words with positive and negative sentiments in word clouds with word sizes corresponding to the prediction scores.

```
subplot(1,2,1)
idx = YPred == "Positive";
wordcloud(wordsTest(idx),scores(idx,1));
title("Predicted Positive Sentiment")

subplot(1,2,2)
wordcloud(wordsTest(~idx),scores(~idx,2));
title("Predicted Negative Sentiment")
```



Calculate Sentiment of Collections of Text

To calculate the sentiment of a piece of text, for example an update on social media, predict the sentiment score of each word in the text and take the mean sentiment score.

```
filename = "weekendUpdates.xlsx";
tbl = readtable(filename,'TextType','string');
textData = tbl.TextData;
textData(1:10)
```

```
ans = 10×1 string array
    "Happy anniversary! ❤ Next stop: Paris! ✈ #vacation"
    "Haha, BBQ on the beach, engage smug mode!    ❤    #vacation"
    "getting ready for Saturday night    #yum #weekend    "
    "Say it with me - I NEED A #VACATION!!! ☹  "
    "    Chilling    at home for the first time in ages…This is the life!    #weekend"
    "My last #weekend before the exam       ."
    "can't believe my #vacation is over    so unfair"
    "Can't wait for tennis this #weekend          "
    "I had so much fun!        Best trip EVER!        #vacation #weekend"
    "Hot weather and air con broke in car    #sweaty #roadtrip #vacation"
```