6COM2000 Advanced AI

NLP Mini Project Specification:

Sentiment Analysis of Customer Reviews

1 General Instructions

The aim of the mini project is to help you learn the module material. The following programming style is required:

(a) Your name and SRN must appear at the top of all files which you have created or modified.

(b) Write clear and understandable code. Improve the clarity of your code by using meaningful variable names, proper indentation and comments, vertical and horizontal spacing.

(c) Precede each function with comments indicating:

- what the function does, and
- what each parameter/variable is used for.

2 Deliverable

The project has to be submitted as a zipped file on the Canvas module page. The deliverable should include your implementation and a report containing (at least) the following:

- the problem statement including description of the dataset,
- an explanation of the approaches taken,
- performance of your model on labelled data, and
- a discussion of your results for both the training, test and validation data.

You will be implementing a software artefact and feedback will be in the form of a score identifying strengths in your work as follows:

Deliverable	**Maximum points**
The problem statement and description of the dataset	10
Building model and making prediction (or an NLP task of your choice)	90
Total	**100**

Building model and making prediction	**Maximum points**
Data preprocessing and building bag-of-words representation	15
Building a classifier	30
Prediction on unseen reviews	20
Report	25
Total	**90**

3 Assignment: Sentiment Analysis of Customer Reviews

Sentiment analysis is an important and well-known branch of Natural Language Processing. The main goal of this mini project is to use sentiment analysis to analyze segments of texts / tweets and deduce the sentiment within. You can either determine an NLP/sentiment analysis related task you would like to work on or consider (a variant of) the suggestion provided below.

You can choose either the business case provided below or a business case of your choice.

3.1 Business Case

A restaurant intends to build a binary review classification model for reviews received on their page. Binary here means either positive or negative, where positive reviews are appreciation and negative reviews are criticism.

The business intends to build an in-house customer support team to callback all customers who give negative feedback, and try to resolve their issues or give them discounts, so as to ensure they revisit. There is a data set of customer reviews for the restaurant along with positive/negative labels. Additionally, there are fresh customer reviews that can be used to generate labels. So, deliverables to the client would be:

1. a sentiment-based binary review classification model, and
2. labelling of fresh reviews as positives/negatives

3.2 Model Building

Your NLP mini project is expected to consist of implementing the steps descried below. The main aim is to have generated labels for fresh customer reviews that were shared with the restaurant. Your implementation will consist of two deliverables: first is on training a sentiment-based review classification model and second on generating labels for fresh customer reviews.

3.2.1 Dataset

Start with finding a dataset for your mini project. You can create it yourself (advisable) or use one of the open datasets such as Kaggle.

3.2.2 Data Preprocessing

Perform data preprocessing. The objective with this data cleaning exercise is to drop unwanted non-value-adding characters & words that would otherwise unnecessarily consume computational resources.

3.2.3 Bag-of-Words Representation

At this stage you need to transform your dataset into a bag-of-words representation.

3.2.4 Building a Naive Bayes Classifier

Build a Naive Bayes classifier for your sentiment analysis model.

3.2.5 Prediction on Unseen Reviews

Split up your data into training and testing and train your sentiment analysis model on the training set. Generate labels for fresh customer reviews.

3.2.6 Report

Write a short, 2-4 page report evaluating and summarising your process and results.