COMP472 - Reports

Thomas Backs[1], Marco Tropiano[2], and Earl Aromin Steven[3]

[1] 27554524 - thomasbacks@gmail.com
[2] 26789331 - tropiano.m@gmail.com
[3] 40004997 - earlaromin@gmail.com

1 Introduction

This project is completed within the COMP 472 course. It aims is to create an unsupervised machine learning program that will take some people feedback on various items they purchased and to classify it in two different group: Positive review or negative review. We used 80% of the data collected to train our machine, and the remaining 20% to test our machine. The analysis of each task of the project are shown in different section.

2 Task 1 Analysis

For the task 1, we created a function called `def train_nb(documents, labels)`, it takes two parament which are the `documents` and the `labels` to train our machine. We also import the `Counter` from the `collections` modules as well to help us to count the number of occurence of each word.

```
def train_nb(documents, labels):
neg_word_count = Counter()
pos_word_count = Counter()
neg_total_word = 0
pos_total_word = 0
# we now create our classification
classifier = list(zip(labels, documents))
for c in classifier:
    if c[0] == 'neg':
        neg_word_count.update(c[1])
    else:
        pos_word_count.update(c[1])

neg_total_word = sum(neg_word_count.values())
pos_total_word = sum(pos_word_count.values())
total_word = sum(neg_total_word, pos_total_word)
return neg_total_word, pos_total_word,
    total_word, neg_word_count, pos_word_count
```

3 Task 2 Analysis

4 Issues

5 What would you expand