**Competitor Analysis**

**Problem Statement:**

Given a accounting software review containing multiple aspects and varied opinions, the objective is to extract expressions or opinion about that particular aspect, and classify it as positive or negative.

**Objective:**

1.Extract all the aspect Category from the given review.

2. Extract opinion word that describe polarity of target aspect Category.

3. Classify extracted opinion word as positive or negative.

**WORK DONE:**

**Data Collection:**

1. Reviews were collected from different sources that were provided using python’s Beautiful soup Library.
2. Around 10000 reviews were collected (we will use 2000 Reviews for training the model).

**Data Exploration:(**for labelling purpose)

Task- what are the different categories/feature that one accounting software consist?

Implementation-

1. Aspect Categories were identified by exploring some websites.
2. From collected data, most common word used by reviewers were identified. (Most\_Common\_words\_used.ipynb)
3. From collected data, Mostly used nouns, adjective & verb were identified. (Count\_of\_noun\_adjective\_verb.ipynb)

From above implementation these aspect categories were identified:

1. Price/ Value for Money
2. Ease of use
3. Security
4. Customer support
5. UI/Graphic.
6. Functionality
7. Time taking
8. Accounting method- financial management
9. Storage
10. Updates- updates are provided or not
11. Auto update – can we customize this with our convenience or not, automation.
12. Inventory Management
13. Backup
14. Others
15. None – not talked about any specific feature but given overall polarity.

**Data Labelling:**

Data label using above mention 15 aspect categories & their polarity. Labelling of around 2000 reviews has been planned. Till now 800 reviews has been labelled, it will be completed around Wednesday afternoon.

|  |  |  |
| --- | --- | --- |
| id | Review | #aspect category # polarity |
| 1 | Auto updation of profit loss and balance sheet. Easy to use | #auto update &ease of use#1&1 |

**File name:** customer\_final\_dataset.xlsx

**FUTURE WORK:**

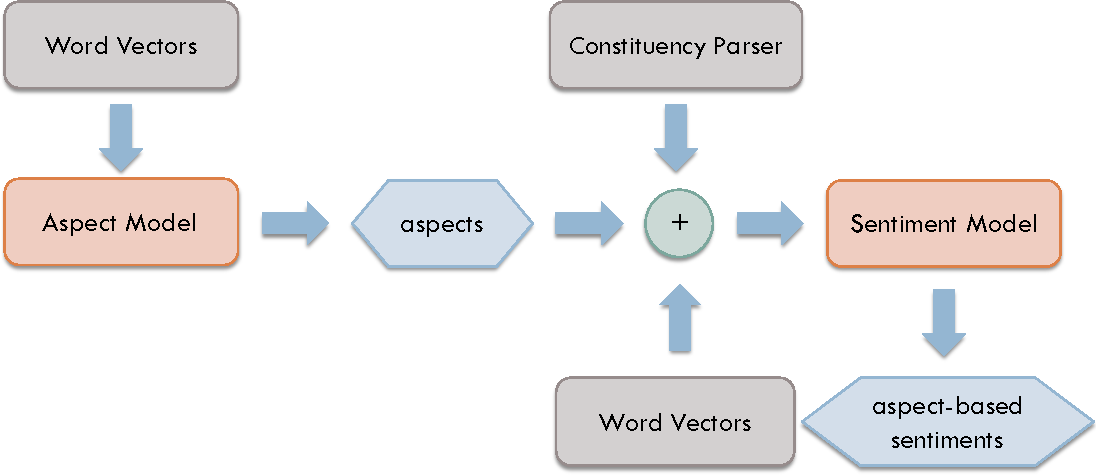
**Dataset:**

1. Data will be divided into 3 part – training data, validation data & test data.
2. For tuning the hyperparameter validation data will be used.
3. After training the model, it will be tested on test data (previously not used).

**Data Pre-processing:**

1. Standard text pre-processing technique will be used like- removal of stop words & punctuation, tokenization, lemmatization etc.

**Overall Architecture:**

****

(image source: reference1)

**Aspect Detection Model:**

For aspect category detection model, Bidirectional LSTM RNN will used as it can analyse input in both forward and backward direction simultaneously. Will represent reviews using word vectors. In order to prevent overfittingof our models, we use pre-trained word embeddings. Choice of word embedding is a hyperparameter. In word vector representation, each sentence is represented as a matrix Rn×d, where n is number of words in a sentence and d is the dimension of word embedding.

RNN will take word vectors as an input and outputs a probabilistic distribution over 15 aspects mention above. Every class having probabilistic distribution above threshold θ will treated as an aspect category. Here θ is hyperparameter.

**Evaluation of ADM:**

Still looking for best metrics to evaluate ADM. I have gone through some literature but didn’t find any relevant literature but they have not mention correct way that how they are going to evaluate the model.

Some of literature mention F1-score but have not mention the proper method.

**Aspect Based Sentiment Model:**

For aspect based sentiment analysis Model will use method mention in reference 1.

**Evaluation of ABSA:**

Reading some literature to find out best evaluation metrics.

**Reference for training the model:**

1. Deep Learning for Aspect Based Sentiment Analysis. (<https://cs224d.stanford.edu/reports/WangBo.pdf>)
2. Aspect Based financial Sentiment Analysis using Deep Learning. (<https://dl.acm.org/ft_gateway.cfm?id=3191827&type=pdf>)