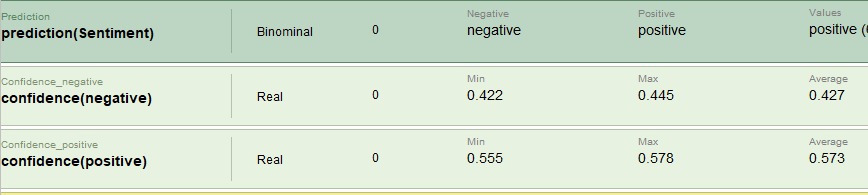
1. **Sentiment Analysis results**

The sentiment analysis was run on the first 6 pages of the ‘all recipes’ website to determine from each page the sentiment associated with the recipes on each page.

We observed positive sentiment for pages 2,3 and 5 has a similar returning a positive sentiment confidence of 0.576 and a negative confidence of 0.424; page 4 and 6 similarly returned a positive sentiment confidence of 0.578 and a negative sentiment confidence of 0.422 respectively. I also observed that only page 1 has a positive sentiment confidence of 0.555 and a negative confidence of 0.445.

The accuracy of the sentiment analysis is as shown below;





**2 The key advantages and disadvantages of a content-based recommender compared to a collaborative filtering approach.c=**

**A content based filtering** recommends other products similar to what the user wants

**Advantages** are it brings a unique suggestions for individual users which makes it easy to scale a large number of users.

The model does not require any data about any other users.

The model will simply capture specific interest of a user thereby recommending niche items that very few users are interested in.

**Disadvantages** are this model can only recommendations based existing interest of the user which makes it restricted to existing interest.

It requires a lot of domain knowledge.

The model can only be good as hand-engineered features.

**A Collaborative Filtering approach** provides suggestions based on common similarities between the users and things at the same time for solving some of the limitations on the content based filtering

**Advantages** is that the embedding are usually discovered automatically which means a domain expertise isn’t required.

It collects feedback from users on different items and uses it for recommendations.

It helps to discover new interest for users .

**Disadvantages** is that if an item isn’t observed during training the system would not be able to create a n embedding for a new query when the model’s forecast for a given pair is the product of their linked embedding.

It cannot handle fresh items.

**The Content base filtering is a better approach** for the academic abstract because it was easy to scale the 2 users and it doesn’t require data about any other users and it captures very specific interest of the users.

3a To improve the performance for NMAE error for KNN is 13.293 which was also similar to the Slope one , user item baseline , matrix factorization However using Random which gave a result of 13.788 which is higher than the KNN parameters but when I used Global Average I got the value of 1.922 which was also noticed to be the lowest error in the parameters .

3b

Sparsity means there is a shortage of an idea and this is a problem for collaborative filtering because the applicability of this is limited which may show that the transactional data are lacking or insufficient considering the total number of users with a very huge number of item which results in the fact there will be a problem with collaborative filtering because the similarity between the users cannot be defined and missing information makes evaluation unreliable. Meanwhile sparsity issue will tend to be less for content base recommenders because in this case we always make use of similarity metrics based on the interest of the users and items.

However, sparsity problems can be addressed using dimensional reduction because it will eliminate the irrelevant users or items to change to the user item matrix.