**M S Ramaiah Institute of Technology**

(An Autonomous Institute, Affiliated to VTU)

MSR Nagar, MSRIT post, Bangalore-54

A Dissertation Report on

**Sentimental Analysis : AMAZON FINE FOOD REVIEW**

Submitted by

**Ganesh M Hedge** **1MS12CS033**

**Charan Rao S** **1MS13CS406**

**Karthik M**   **1MS13CS407**

**Mithun R**  **1MS13CS414**

*in partial fulfillment for the award of the degree of*

*Bachelor of Engineering in Computer Science & Engineering*



**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**M.S.RAMAIAH INSTITUTE OF TECHNOLOGY**

**(Autonomous Institute, Affiliated to VTU)**

**BANGALORE-560054**

[www.msrit.edu](http://www.msrit.edu/), **May 2015**

**Acknowledgements**

First and foremost,my utmost Gratitude to our HOD, Dr. K G Srinivasa who gave us a chance to complete this project. He has been our inspiration as we overcome all the obstacles in the completion of this project work. He had kind concern and consideration regarding project work which helped us to successfully complete the project.

Secondly we would like to thank Sowmya B J, Assistant Professor, MSRIT and Chandrika Prasad, Assistant Professor who reviewed our project and gave us a valuable guidance which helped us in completing this project.

This work would not have been possible without the guidance and help of several individuals who in one way or another contributed their valuable assistance in preparation and completion of this study.

We would like to express sincere thanks to all the teaching and non-teaching faculty of CSE department and my dear friends who helped in all the ways while preparing the Report.

Abstract

Recommending products to consumers means not only understanding their tastes, but also understanding their level of experience. For example, it would be a mistake to recommend the iconic film Seven Samurai simply because a user enjoys other action movies, rather, we might conclude that they will eventually enjoy it—once

they are ready. The same is true for beers, wines, gourmet foods or any products where users have acquired tastes: the ‘best’ products may not be the most ‘accessible’. Thus our goal in this paper is to recommend products that a user will enjoy now, while acknowledging that their tastes may have changed over time, and may change again in the future. We model how tastes change due to the very act of consuming more products—in other words, as users become more experienced. We develop a latent factor recommendation

system that explicitly accounts for each user’s level

of experience. We find that such a model not only leads to better recommendations, but also allows us to study the role of user experience and expertise on a novel data set of Food reviews.

Contents

***Declaration i***

***Acknowledgments ii***

***Abstract iii***

***List of Figures***

***List of Tables***

1 **INTRODUCTION**

* 1. General Introduction………………. 6
  2. Statement of the Problem…………..
  3. Objectives of the project……………
  4. Project deliverables……………
  5. Current Scope………………………
  6. Future Scope……………………….

1. **PROJECT ORGANIZATION**
   1. Software Process Models
   2. Roles and Responsibilities
2. **LITERATURE SURVEY**

3.1….Introduction

3.2…Main Body

3.3 Conclusion of Survey

1. **SOFTWARE REQUIREMENT SPECIFICATIONS**

5.1 Product Overview

5.2 External Interface Requirements

5.2.1 User Interfaces

5.2.2 Hardware Interfaces

5.2.3 Software Interfaces

5.2.4 Communication Interfaces

5.3 Functional Requirements

5.3.1 Functional Requirement 1.1

:

5.3.n Functional Requirement 1.n

5.4 Software System Attributes

5.4.1 Reliability

5.4.2 Availability

5.4.3 Security

5.4.4 Portability

5.4.5 Maintainability

5.4.6 Performance

5.5 Performance Requirements

5.6 Database Requirement

5.7 Design Constraints

5.8 Other Requirements

1. **DESIGN**
   1. Introduction
   2. Architecture Design
   3. Graphical User Interface
   4. Class Diagram and Classes (represent Inheritance, Aggregation and Association)
   5. Sequence Diagram
   6. Data flow diagram
   7. Metric calculation
2. **IMPLEMENTATION**
   1. Tools Introduction
   2. Technology Introduction
   3. Overall view of the project in terms of implementation
   4. Explanation of Algorithm and how it is been implemented
   5. Information about the implementation of Modules
3. **TESTING**

**7.1** Results and Snapshots

1. **CONCLUSION & SCOPE FOR FUTURE WORK <No Page Numbers>**
2. **REFERENCES**
3. **INTRODUCTION**

1.1 General Introduction

The goal of this project is to develop a classifier which can predict sentiment of a text-fragment along a scaled range from one to five stars. So far, most of the major research on sentiment analysis

has been done to predict the polarity of text: positive or negative sentiment, but not subjective opinions along a multi-class continuum. We are able to find numerous kinds of topics and data sets on the web such as movie-ratings, book-reviews,twitter tweets, etc.   
Among those, we chose a data set of Amazon Food reviews which includes various sentiments of predefined aspects of these food review rated on a scale of 1 to 5 and also the review text.   
 Through these reviews, we will first analyze the properties of data set and explain the basic methods we used to select the features. Moreover we will suggest two novel approaches of extracting good features and illustrate the learning result compared to previous research done on similar domains.

1.2 **Statement of the Problem**

In order to predict how a user will respond to a product, we must understand the tastes of the user and the properties of the product. We must also understand how these properties change and evolve over time.   
High Reviews = G ood Food .

As an example, consider the Harry Potter film series: adults who enjoy the films for their special effects may no longer enjoy them in ten years, once their special effects are obsolete, children who enjoy the films today may simply outgrow them in ten years; future children who watch the films in ten years may not enjoy them, once Harry Potter has been supplanted by another wizard.

1.3 **Objectives of the project**

Classifying sentences/documents (e.g. reviews) based on the overall sentiments expressed i.e positive or negative and in which   
sentiment classification is are more important such that extraction of each words, and the words count are being processed in sentimental analysis.

Detect the target, source, or complex attitude types.However, it’s impossible to judge the actual **quality** of the raw materials that were used to make each specific batch of food available for purchase. So this drawback was considered to be overcome by our project which results the user to experience his taste for the product and his experience level so that he can recommend this product for other customer’s to try out it.

It Evaluates the degree of the writer expresses his food experienced reviews for customer to have an idea of it.

1.4  **Project Deliverables**

Project plan includes the following

A. Data Set

B. System Requirement Specifications

C. Prepared PPT which includes the following

1.Immediate notification of any major deficiencies detected during the review process

2. Discussion Document setting out findings, options and likely impacts/consequences of options.

3. Gathering and analysis of submissions.

1.5  **Current Scope**

An interesting finding of our work is that beginners and experts

have the same polarity in their opinions, but that experts give more

‘extreme’ ratings: they rate the top products more highly, and the

bottom products more harshly.

Recommending products to consumers means not only understanding their tastes, but also understanding their level of experience.

Good reviews equals to better food, such that the new user with no idea  
of the products will have the quick view of the review and based  
on the review the customer taste’s the product with greater satisfaction  
level and hence the customer gives positive, negative reviews.

1.6 **Future Scope**

we might conclude that we should simply recommend both groups of users the same products: nobody likes adjunct lagers, so what does it matter if

beginners dislike them less? The counter to this argument is that in

order to fully appreciate a product (by giving it the highest rating), a

user must first become an expert. Thus perhaps we should focus on

making a user an expert, rather than simply recommending what

they will like today. Would expert movie reviewers on Amazon also be ‘experts’ gourmet food reviewers? The answers to such questions could help us to determine which forms of personal development are more salient in online communities.

Finally, we believe that our models of expertise may be used to

facilitate expert discovery. We identified experts in review systems

primarily for the sake of predicting users’ ratings, but discovering

experts may be useful in its own right. This topic is known as expert

recommendation and has applications to product ranking and review summarization , among others. For instance, a user may wish to read reviews written by the most expert members of a community, or by members most similar in expertise to Themselves. In typical sentiment analysis, the task is simply to identify positive or negative reviews. This can be handled rather easily through bag of words techniques with uni grams and bi grams.

**2 . PROJECT ORGANIZATION**

**2.1 Software process models**

The parameter used here to classify the process models is based on defining what process elements the model covers.

The ***Activity-Oriented Models*** focus on defining the functions, activities and

information of the software process management, development and/or supporting Processes

The ***People-Oriented Models*** focus on defining the people involved in

the software process and their relationships

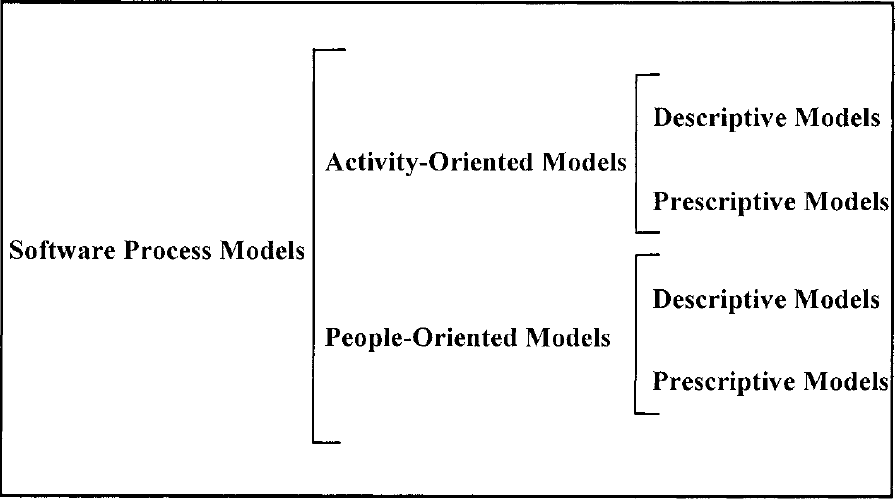
These two model categories are further divided into two alternatives depending on the goal for which the model was developed:

***Descriptive Models*** are mainly concerned with specifying a process now used within an organization or with representing a proposed process to be able to predict some process features Descriptive models answer the question, "How is software now developed?" Or how has software been developed?

***Prescriptive Models*** focus on defining how the process should be enacted.

Prescriptive models answer the question "How should software be developed?"

It should be noted that this review includes both models that are designed to simply model the process and models that pursue other objectives, such as process evaluation, improvement or prediction. These models are also considered, because, to achieve their specific objectives.



**Roles and Responsibilities**

* Ganesh Hedge - Maintaining the backend (Hadoop) and frontend. Technologies used are Hadoop, Mapreduce and Java.
* Karthik - Normalizing the dataset, stemming and implementing Naïve Bayes algorithm.
* Mithun - Collecting the dataset, performing the tests, documenting the details. Tehnologies used are Java, Hadoop and Mapreduce.
* Charan - Classifying the dataset as positive, negative and neutral using installation of hadoop .

**3. Literature Survey**

3.1 **Introduction**

Sentiment Analysis is a *Natural Language Processing* and *Information Extraction* task that aims to obtain writer’s feelings expressed in positive or negative comments, questions and requests, by analyzing a large numbers of documents. Generally speaking, sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic of a document. In recent years, the exponential increase in the Internet usage and exchange of public opinion is the driving force behind Sentiment Analysis today. The Web is a huge repository of structured and unstructured data. The analysis of this data to extract latent public opinion and sentiment is a challenging task.

The analysis of sentiments may be document based where the sentiment in the entire document is summarized as positive, negative or objective. It can be sentence based where individual sentences, bearing sentiments, in the text are classified. SA can be phrase based where the phrases in a sentence are classified according to polarity.

3.2 **Main Body**

Growing availability of opinion rich resources like online review sites, blogs, social networking sites have made this “decision-making process” easier for us.

Hadoop supports the MapReduce model, which was introduced by Google as a method of solving a class of petascale problems with large clusters of inexpensive machines. The model is based on two distinct steps for an application:

• Map:

An initial ingestion and transformation step, in which individual input records can be processed in parallel.

• Reduce:

An aggregation or summarization step, in which all associated records, must be processed together by a single entity.

The core concept of MapReduce in Hadoop is that input may be split into logical chunks, and each chunk may be initially processed independently, by a map task. The results of these individual processing chunks can be physically partitioned into distinct sets, which are then sorted. Each sorted chunk is passed to a reduce task.

* SA broadly classifies the applications into the following categories.

a. Applications to Review-Related Websites - Movie Reviews, Product Reviews *etc.*

b. Applications as a Sub-Component Technology-

Detecting antagonistic, heated language in mails, spam detection, context sensitive information detection *etc.*

c. Applications in Business and Government Intelligence- Knowing Consumer attitudes and trends

d. Applications across Different Domains- Knowing public opinions for political leaders or their notions about rules and regulations in place *etc*

The Consumer's Perspective

While taking a decision it is very important for us to know the opinion of the people around us. Earlier this group used to be small, with a few trusted friends and family members. But, now with the advent

of Internet we see people expressing their opinions in blogs and forums. These are now actively read by people who seek an opinion about a particular entity (product, movie etc.). Thus, there is a plethora of

opinions available on the Internet.

From a consumers' point of view extracting opinions about a particular entity is very important. Trying to go through such a vast amount of information to understand the general opinion is impossible for

users just by the sheer volume of this data. Hence, the need of a system that differentiates between good reviews and bad reviews. Further, labeling these documents with their sentiment would provide a summary to the readers about the general opinion regarding an entity.

The Producer's Perspective

With the explosion of Web 2.0 platforms such as blogs, discussion forums, etc., consumers have at their disposal, a platform to share their brand experiences and opinions, positive or negative regarding any product or service.  
 According to the consumer voices can wield enormous influence in shaping the opinions of other consumers and, ultimately, their brand loyalties, their purchase decisions, and their own brand advocacy.

Since the consumers have started using the power of the Internet to expand their horizons, there has been a surge of review sites and blogs,  
 where users can perceive a product's or service's advantages and faults.   
These opinions thus shape the future of the product or the service.  
 The vendors need a system that can identify trends in customer reviews and use them to improve their product or service and also identify the requirements of the future.

The Societies' Perspective

Recently, certain events, which affected Government, have been triggered using the Internet. The social networks are being used to bring together people so as to organize mass gatherings and oppose oppression.

On the darker side, the social networks are being used to insinuate people against an ethnic group or class of people, which has resulted in a serious loss of life. Thus, there is a need for Sentiment Analysis systems that can identify such phenomena and curtail them if needed.

**Tasks based on Classification**

Identifying Subjectivity:

The basic question asked in Sentiment Analysis is whether a given piece of text contains any subjective content (opinions, emotions, etc.) or not. This task aims to tackle this problem of differentiating between subjective and objective content

Identifying discrete polarities:

Once the subjective part is determined, the next task is to determine if the content is positive or negative. This problem can be looked upon as a classification problem.

Identifying an ordinal value:

Some applications require not just the type of polarity but the intensity as well. For example, foods are typically rated on a 5 point scale. Thus, this task aims at identifying such an ordinal value.

**Tasks based on Levels of Sentiment Analysis**

Document level:

As the name suggests, document-level Sentiment Analysis tags individual documents with their sentiment.

The general approach here is to find the sentiment polarities of individual sentences or words and combine them together to find the polarity of the document. These techniques may involve complex linguistic phenomena like co-reference resolution, pragmatics, etc.  
  
Sentence or phrase level:

Sentence-level Sentiment Analysis deals with tagging individual sentences with their respective sentiment

polarities. The general approach that is followed is to find the sentiment orientation of individual words in the sentence/phrase and then to combine combine them to determine the sentiment of the whole sentence or phrase. Other approaches like considering discourse structure of the text, have also been considered.

Aspect level:

These methods not only concern themselves with tagging individual words with their sentiment but they also aim at identifying the entity towards which the sentiment is directed. These methods heavily use techniques like dependency parser and discourse structures.

**CONCLUSION**

This report discusses in details the various approaches to Sentiment Analysis, mainly Machine Learning and Cognitive approaches. It provides a detailed view of the different applications and potential challenges of Sentiment Analysis that makes it a difficult task.

We have seen the applications of machine learning techniques like Naïve Bayes, in SA and their potential drawbacks. As all of these are bag-of-words model, they do not capture context and do not analyze the discourse which is absolutely essential for SA. Also it achieves the difficult task of performing prediction over a continuum even though trained only on the extreme reviews. Thus machine learning models with a proper kernel that can capture the context will play an important role in SA.

Feature engineering, as in several Machine Learning and Natural Language Processing applications, plays a vital role in SA. We have seen the use of phrases as well as words as features. It has been seen that Adjectives as word features can capture majority of the sentiment.

The use of discourse analysis and tracking point of view are necessary for analyzing opinions in blogs, newspaper and articles where a third person narrates his/her views. We also discus some specific topics in Sentiment Analysis and the contemporary works in those areas.

**4**. **Software Requirement Specification**

4.1 Product Overview

Our project deals with counting the number of positive and negative reviews.  
The final output will be in pie chart for new user and for   
experienced user will be in bar chart

4.2 External Interface Requirements

4.2.1 User Interfaces:

The interface will be one of the most important stage of this project as it is simple

UI and easy to understand such that one can can easily build Graphical UI .

Controls which allow the user to interact with the application will be clear  
 and imply their functionality within the application.

The user interface shall be implemented using tools like Eclipse

4.2.2 Hardware Interfaces:

The application must run over the internet, all the hardware required

to connect internet will be hardware interface for the system ex: Modem.

The analysis doesn’t require any special hardware. It can run on a password protected personal laptop.

4.2.3 Software Interfaces:

1. The Application shall communicate with the Configurator to identify all the available components to configure the product.
2. Then it shall communicate with the Data Base to get the product specifications, offerings and promotions based on which reviews are configured.The software that are required for the analysis are Eclipse, Hadoop, EM editor(to open large files), Hive.

4.2.4 Communication Interfaces:

The system shall use the HTTP protocol for communication over the internet and for the intranet communication will be through TCP/IP protocol suite are required in order to make use of several functions and for references regarding the projects.

4.3 Functional Requirements

This includes the requirements that specify all the fundamental actions of the software system.

4.3.1 Retrieving Input :

The inputs are extracted from the huge datasets with following attributes

● User ID

● Product review

4.3.2 Data set pre-Processing:

Data set is pre-processed using eclipse IDE to eliminate stop word and stemming is done using porter’s algorithm.

4.3.3 Review Classification:

Reviews classification will be done to determine the nature of the Reviews based on the opinion of the users on the movies. The data analysis will provide a negative, neutral, or positive numeric value for the reviews.

4.3.4 Analysis of the data:

The data should be finally analyzed with values of the positive and negative reviews and then the movie should be categorized into good or bad.

4.4 Software System Attributes

4.4.1 Reliability:

The project’s main purpose is to categorize the data, so the will accurately divide the data into good and bad based in the input and it is very reliable to perform any function that is assigned to do.

4.4.2 Availability:

The project will be available when functioning on demand in any intervals of time. The functionality of the project will not depend on any external services such as internet or any other hardware support and it is proportion of time in a system.

4.4.3 Security:

The project will assure that security will monitor the risk, threats and never disclose any personal information of the users, and the analysis is only for personal reference so it very secure.

4.4.4 Portability:

This project is portable to any operating system and can be deployed as required

Such that no external mechanism is required and also it maintains the same level of security.

4.4.5 Maintainability:

The project is very precise and optimized it for execution so that storage is less for execution.

The code will be well documented. The whole project is modularized to ensure that maintenance is easy.

4.5 Performance Requirements:

The performance enhancement makes it easier to make decisions on whether the movie is good or bad for the viewers. The performance can enhanced by increasing the system properties like increasing the Ram etc, for better performance minimum 2GB of Ram is required and the system storage of more than 10GB is required and the so for better performance good hardware is required.

4.6 Database Requirement:

The database is required to store the huge datasets and Hadoop S3 is used for this purpose.

4.7 Design Constraints:

The design of the project is not so costly as all the software used are the open source software and thus the cost is low and speed of the project is very high as the parallel programming is used using HPC tool like Hadoop. Thus design is very optimized as supports high speed and produce almost accurate results.

4.8 Other Requirements:

The Internet Access to World Wide Web is required for references and information for the completion of the project.

**5. Design**

**5.1 Introduction**

* 1. Number of Modules

There are three modules in the design part

1. Sentiment Analysis module

2. Entity Analysis module

3. Filter Module

* 1. Modules description

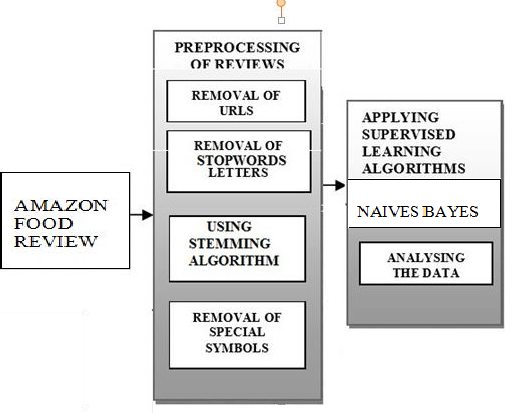
1. Sentiment Analysis Module performs opinion extraction, sentiment mining

2. Entity Analysis module extracts the attributes i.e positive or negative review using the modules

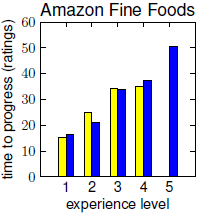
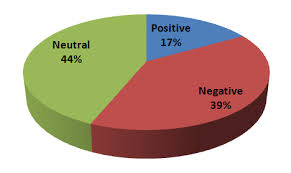
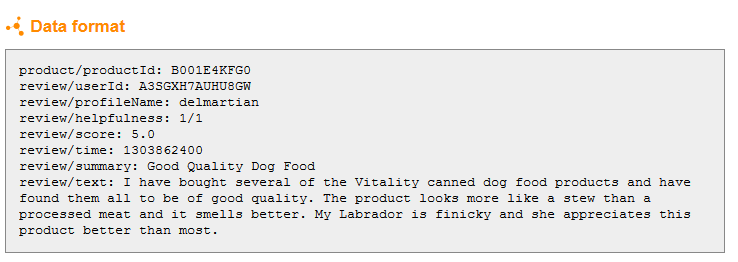
3. Filter module classifies the entities from the module for the sentiment opinion

**5.2 Architecture Design**

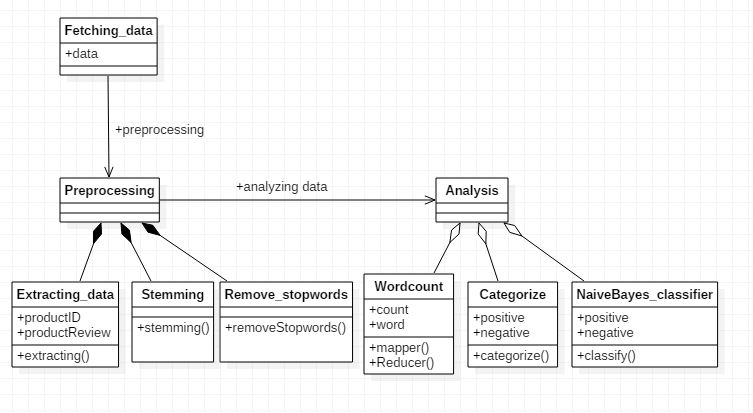
* 1. The architectural design is the design of the entire software system; it gives a high-level overview of the software system, it provides information on the decomposition of the system into modules (classes), dependencies between modules, hierarchy and partitioning of the software modules.



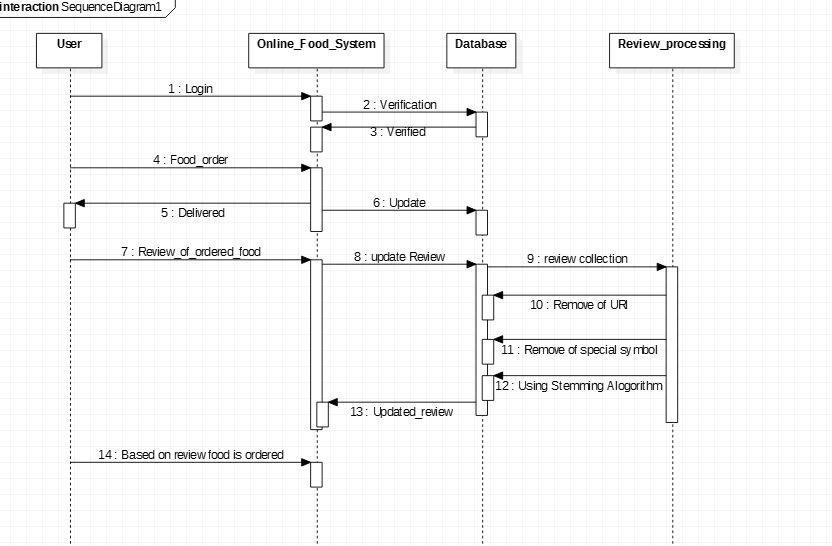
**5.3 GRAPHICAL USER INTERFACE**



**5.4 Class Diagram**



**5.5 Sequence Diagram**



1. **6 Metric calculation**

**Mean Squared Error (MSE)**

We also note that while we obtain significant benefits on Amazon data, the mean-squared-errors for this data set are by far the highest. One reason is that Amazon users use a full spectrum of ratings from 1 to 5 stars, Another

reason is that our Amazon data has many products and users

with only a few reviews, so that we cannot do much better than simply

modeling their bias terms. bias terms differ significantly between beginners and experts, so that modeling expertise proves extremely beneficial on such data.

**6. IMPLEMENTATION**

6.1 Tools Used: Hadoop, Eclipse, Java,

6.2 Technology Used: porter's algorithm, Map reduce , naïve bayes

6.3 Overall view of the project in terms of implementation

The Project is implemented as follows:

* Data Set Identification and Extraction
* Pre-processing of data set
  + Removing unwanted attributes
  + Eliminating stop words
  + Stemming ( Porter’s algorithm)
* Tweet classification
  + Positive
  + Negative
  + **Neutral**

**6.4 Algorithm Design**

In this we have used two major algorithms for processing of sentiment analysis

1. porter’s algorithm

**Stemming** is the term used in to describe the process for reducing inflected words to their base or [root](https://en.wikipedia.org/wiki/Root_(linguistics)) form

Step 1: Gets rid of plurals and -ed or -ing suffixes

Step 2: Turns terminal y to i when there is another vowel in the stem

Step 3: Maps double suffixes to single ones: -ization, -ational, etc.

Step 4: Deals with suffixes, -full, -ness etc.

Step 5: Takes off -ant, -ence, etc.

Step 6: Removes a final -e

1. Naïve Bayes algorithm

The Naive Bayes classifier is a simple probabilistic classifier which is based on Bayes theorem with strong and naïve independence assumptions. It is one of the most basic text classification techniques with various applications in email spam detection, personal email sorting, document categorization,

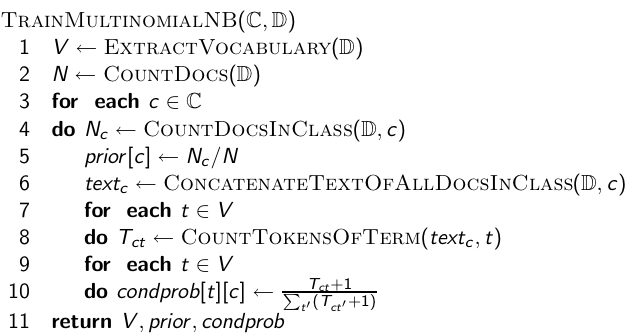
Bayes Theorem p(h/D)= p(D/h) P(h)/ p(D)

P(h) : Prior probability of hypothesis h

P(D) : Prior probability of training data D

P(h/D) : Probability of h given D

P(D/h) : Probability of D given h



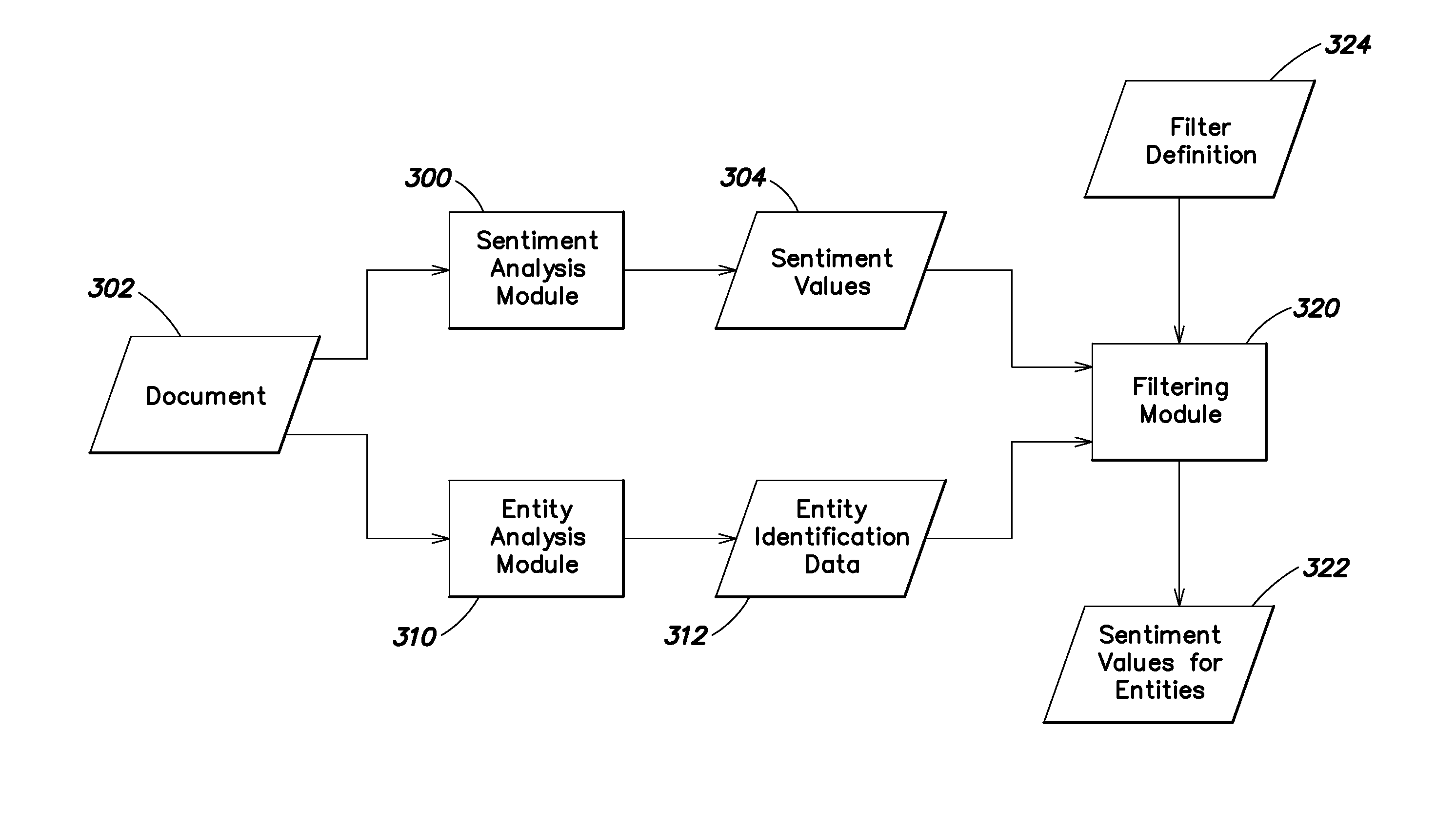
|  |  |  |  |
| --- | --- | --- | --- |
| **Boolean** | **Doc** | **Words** | **Class** |
| Training | 1 | Chinese Beijing | c |
|  | 2 | Chinese Shanghai | c |
|  | 3 | Chinese Macao | c |
|  | 4 | Tokyo Japan Chinese | j |
| Test | 5 | Chinese Tokyo Japan | ? |

1. **Handle Data**: Load the data from CSV file and split it into training and test datasets.
2. **Summarize Data**: summarize the properties in the training dataset so that we can calculate probabilities and make predictions.
3. **Make a Prediction**: Use the summaries of the dataset to generate a single prediction.
4. **Make Predictions**: Generate predictions given a test dataset and a summarized training dataset.
5. **Evaluate Accuracy**: Evaluate the accuracy of predictions made for a test dataset as the percentage correct out of all predictions made.
6. **Tie it Together**: Use all of the code elements to present a complete and standalone implementation of the Naive Bayes algorithm.

3. Probabilistic classifiers

Probabilistic classifiers use mixture models for classification. The mixture model assumes that each class is a component of the mixture. Each mixture component is a generative model that provides the probability of sampling a particular term for that component. These kinds of classifiers are also called generative classifiers

6.5 **Implementation of Modules**



There are three modules in the Implementation part

1. Sentiment Analysis module

2. Entity Analysis module

3. Filter Module

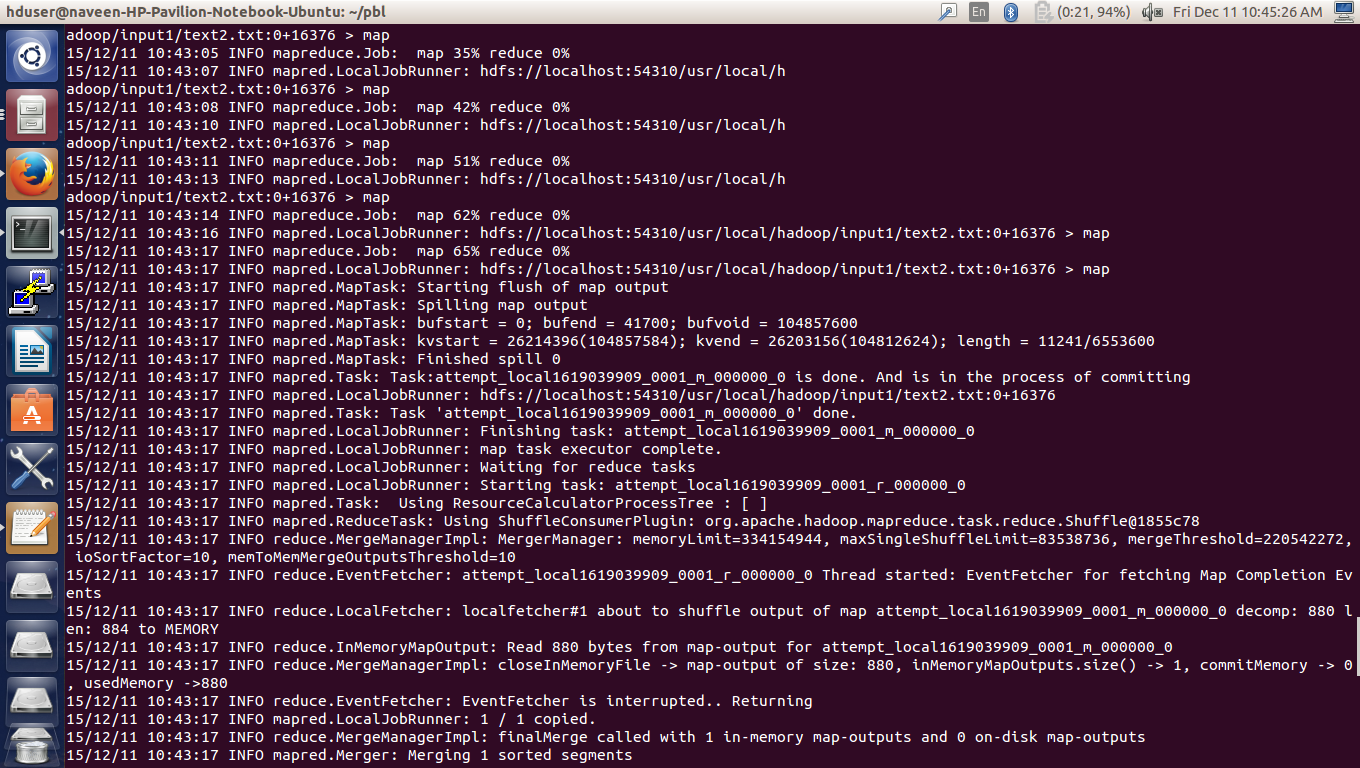
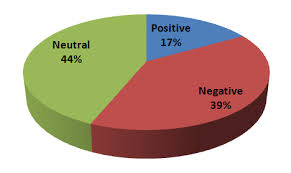
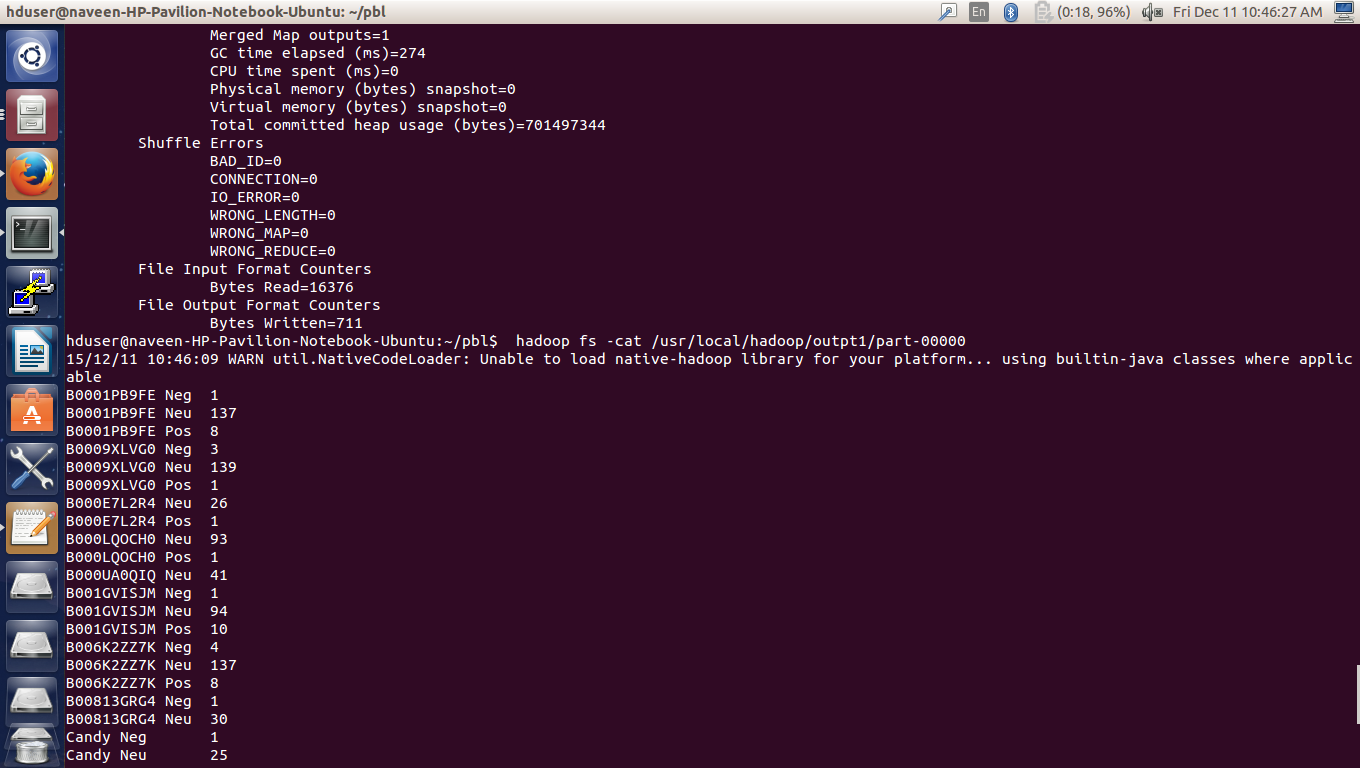
* 1. Modules description

1. Sentiment Analysis Module performs opinion extraction, sentiment mining

2. Entity Analysis module extracts the attributes i.e positive or negative review using the modules

3. Filter module classifies the entities from the module for the sentiment opinion

**TESTING & SNAPSHOTS**



**Conclusion**

Users’ tastes and preferences change and evolve over time. Shifting

trends in the community, the arrival of new products, and even

changes in users’ social networks may influence their rating behavior.

At the same time, users’ tastes may change simply through the

act of consuming additional products, as they gain knowledge and

Experience.  
Existing models consider temporal effects at the level of products   
and communities, but neglect the personal development of users:   
users who rate products at the same time may have less in common  
 than users who rate products at different times, but who are at the same stage in their personal evolution.   
We developed models for such notions of user evolution, in order   
to assess which best captures the dynamics present in product rating   
data. We found that modeling users’ personal evolution, or ‘experience’,  
 not only helps us to discover ‘acquired tastes’ in product rating systems,  
 but more importantly, it allows us to discover when users acquire them.

**Future Work**

There are a number of things that may be done in the future to improve classification.

One important tool in improving sentiment analysis is subjectivity analysis. Within any body of text, there are two types of statements. Objective statements describe that which can be ob-served by other individual. These statements consist of (presumably) factual statements about what is and what has happened. Other statements can be said to be subjective.   
These are statements that cannot be directly observed by others and are thus referred to as private states. These include personal opinions and .  
Objective statements do not provide actual personal opinions and sentiments, and thus can safely be thrown out of the data set without adversely affecting the results of the analysis.

In the future we would like to implement something akin to the Good Grief algorithm which models the level of agreement within reviews. This provides a more accurate model than our current implementation which ignores these dependencies between different aspects.

For instance, a user may wish to read reviews written by the most expert members of a community, or by members most similar in expertise to Themselves. In typical sentiment analysis, the task is simply to identify positive or negative reviews. This can be handled rather easily through bag of words techniques with uni grams and bi grams.

**REFERENCES**

From Amateurs to Connoisseurs: Modeling the Evolution of User Expertise

through Online Reviews- Jure Leskovec, Julian McAuley.

Opinion Mining and Sentiment Analysis- **Bo Pang and Lillian Lee**

Sentiment Analysis: Mining Opinions, Sentiments, and Emotions - Bing Liu

G. Linden, B. Smith, and J. York. Amazon.com recommendations: Item-to-item collaborative filtering. IEEE Internet Computing, 2003.

N. Craswell, D. Hawking, A. Vercoustre, and P. Wilkins. P@noptic expert: Searching for experts not just for documents. Ausweb Poster Proceedings, 2001.

Mutilclass sentimental analysis food review - Moontae Lee

World Wide Web Internet