Sentiment Classification of Amazon Fine Food Reviews Using Deep Learning

Project submitted to the
SRM University – AP, Andhra Pradesh
for the partial fulfillment of the requirements to award the degree of

Bachelor of Technology
In
Computer Science and Engineering
School of Engineering and Sciences

Submitted by

VADLAMUDI SREENIDHI

AP22110010039

CSE-R

Under the Guidance of

Dr. Bala Venkateswarlu

SRM University–AP

Neerukonda, Mangalagiri, Guntur

Andhra Pradesh – 522 240

[December, 2025]

TABLE OF CONTENTS

ABSTRACT

This dataset contains thousands of reviews and rating scores for food products in Amazon, which is analyzed in this project using the Amazon Fine Food Reviews dataset. The target is for us to determine if the review is positive or negative in order to understand the consumers thoughts on different products. Using basic NLP, the text is processed-cleaned and converted into numbers. In order to establish strong baselines, the TF-IDF features are used to train several machine learning methods including Logistic Regression, Linear SVM and Multinomial Naive Bayes.

More powerful deep learning models such as BiLSTM(Bidirectional Long Short-Term Memory) and TextCNN(Text Convolutional Neural Network) are also applied for they can better capture word patterns context and the general meaning of the reviews. Exploratory analysis investigates such phenomena as review length, frequently used words in positive reviews and negative reviews, as well as the products with most reviews.

Every models performance is measured by Accuracy precision recall F1-score and Confusion Matrix. All in all, this project designs a scalable and reliable technique that can automatically detect the food reviews sentiment and provide the businesses with valuable suggestions to improve their products as well as the customer satisfactions.

INTRODUCTION

Customers reviews are significant in improving product quality, customer satisfication and business decisions over e-commerce sites e.g., Amazon. Since there are thousands of reviews about food products posted each day, reading and understanding customer opinions manually would be impossible and time-consuming. Sentiment analysis offers a way to automatically process these large quantities of text and provide insights as to whether consumers express positive or negative feelings about a product.

This project aims to develop an end-to-end sentiment classification pipeline based on the Amazon Fine Food Reviews dataset. The procedure involves cleaning and pre-processing the review text, missing value processing, and merging review summaries and full reviews into one document. The processed text is then converted into numerical form using NLP methodologies, which facilities the building of ML and DL models.

This project involves multiple models: TF-IDF–based classical machine learning techniques such as Logistic Regression, Linear SVM, and Multinomial Naive Bayes, together with more complex deep learning models like BiLSTM and TextCNN. Here the goal is to create a robust system that is capable of learning from patterns in customer feedback, accurately classifying new reviews, and providing actionable insights into trends in customer satisfaction.

OBJECTIVE

1. Explore and Understand the Review Data (EDA):

The first purpose of this project is to analyze the Amazon Fine Food Reviews dataset to figure out what the customer reviews define. This involves analyzing the distribution of the review scores (1 to 5), finding out whether the reviews are biased towards being positive or negative, and analyzing how long the reviews are and what kind of words people use examining the review texts. Some visualization (bar chart, histogram, word clouds) show some trends of the data. By identifying the most common words in positive and negative reviews, we understand how customers talk when they are satisfied and dissatisfied.

2. Build Machine Learning Models to Classify Review Sentiment:

Next, is to build models that predict whether a review is positive or negative using the text of the review. This includes preprocessing of the text samples, denoising, followed by transforming samples into numeric features using TF-IDF. Multiple machine learning algorithms, such as Logistic Regression, Linear SVM and Multinomial Naive Bayes are trained on these features to find the correlation between the different words/phrases and the sentiment of the reviews. The models are also tested on their performance using accuracy, precision, recall and f1-measure.

3. Model Performance Comparison & Best Model Selection:

The ultimate goal is to test various ML and DL models and find the best one that provides the best quality in terms of output. Deep learning models, such as BiLSTM and TextCNN are also evaluated due to their superior understanding of context and word order compared to simple models. Confusion matrices and metrics are used to study each model's strengths and weaknesses. So by testing on reviews that neither has seen before, this project determines which one generalizes the best, and is the best candidate for real-world usage.

SYSTEM DESIGN

The main components of the system are:

i. Data Loading and Cleaning

The procedure begins with the loading of the Amazon Fine Food Reviews dataset. Imputation for missing values, removing duplicates and cleaning the text in Summary and Review columns by removing special characters and converting everything to lower case.

ii. Text Preprocessing

Then basic NLP (tokenization, optional stopword removal, lemmatization) is also applied to the text. This processes and makes the reviews to be processed into numerical features so that they can be put into the model.

iii. Feature Extraction (TF-IDF or Tokenization)

TF-IDF is applied to convert the text into numerical vectors for machine learning models.

For deep learning model, each review is transformed to a sequence of numbers via tokenizer and all sequence are padded with same length.

iv. Data Splitting

Now, you split the dataset into train,validation and test so that the model can be trained unevaluated properly.

v. Machine Learning Model Training

The following algorithms are trained using the above TF-IDF features: Logistic Regression, Linear SVM, and Multinomial Naive Bayes. These provide fast and robust baselines to sentiment classification.

vi. Deep Learning Model Training

More advanced models, such as BiLSTM and TextCNN, are trained using the padded sequences. These models learn deeper meaning from the text, including context and word patterns.

vii. Model Evaluation

All models are evaluated using metrics such as accuracy, precision, recall, F1-score, and confusion matrices. This helps compare the models and understand how well each one performs.

viii. User Prediction Module

Finally, the system includes a simple prediction interface where users can enter a review index or type a new review to get an instant sentiment prediction—positive or negative.

IMPLEMENTATION

Data collection and pre-processing:

- The project begins with importing the requisite packages and loading the Amazon Fine Food Reviews (Reviews.csv) review text, summary, product, user and rating score are included in the dataset.

- We replace any missing value in Summary or Text columns with an empty string and clean useless data by removing duplicates.

Text Cleaning:

The preprocessing aim is to transform the raw review text to clean and noise free text for future analysis.The cleaning instructions are as follows:

- make text all lowercase
- Remove numbers, special characters and keep only letters.
- Tokenize the text with Python/NLTK.
- Remove the stopword if you want.
- Do Lemmatization with WordNetLemmatizer.

Feature Extraction:

There are two types of features created for the models:

1. TF-IDF (for Machine Learning models)

- Transforms processed text to numerical vectors.
- Applying unigrams and bigrams with a vocabulary size of 5,000-10,000 words.
- It is Useful to Models Like Logistic Regression, SVM, and Naive Bayes to Identify Important Words.

2. Tokenization and Padding (for Deep Learning models)

- Convert words into integer sequences with the Keras Tokenizer.
- Each batch contains sequences in all batches, that is, each batch has 100 words, then padding all to the same length.

- Up to 30,000 vocabulary words.

Data Splitting:

The data set is partitioned into:

- 70% Training data
- 15% Validation set
- 15% Testing set
- It uses stratified splitting to keep the positive/negative balance consistent for all 3 sets.

Machine Learning Models:

Traditional machine learning models are trained with TF-IDF features:

- Logistic Regression
- Linear SVM (LinearSVC)
- Multinomial Naive Bayes

These are fast models, easy to train with, and yield reasonably strong baseline accuracy for sentiment classification.

Deep Learning Models:

Two deep learning approach are available for more sophisticated text comprehension:

1. BiLSTM (Bidirectional LSTM)

- Includes an Embedding layer
- Employs spatial dropout
- Contains a Bidirectional LSTM layer
- Followed by dense layers for final prediction
- Learns and models long-term word dependencies and overall context

2. TextCNN

- Uses an Embedding layer
- Has multiple CNN filters (size 3 and 4)
- Performs global max-pooling
- Ends with dense classification layers
- Identifies key phrases and sentiment indicators such as “not good” or “highly recommend”

Both deep learning architectures are trained with the binary cross-entropy loss and employ Early Stopping to avoid overfitting.

Model Evaluation:

All models are assessed on:

- Accuracy

- Precision
- Recall
- F1-score
- Confusion matrix

These measures give us insight into how each model performs in terms of accuracy and error.

Prediction Module:

The kit included a prediction interface which is as easy as 1-2-3:

- Users are allowed to input a review index from the test set
- The review is positive or negative, as predicted by the system

Exporting Results:

- All final output such as model scores, predictions and cleaned text are stored in CSV files.
- The tokenizer, TF-IDF vectorizer and trained models are serialized using pickle or Keras .h5 files which allows us to use them later without having the retrain.

RESULTS

The sentiment classification system had good performance for both ML and DL models. Among the traditional models, Linear SVM obtains the best baseline accuracy, and the deep learning models (BiLSTM and TextCNN) achieve even better results due to their ability to capture contextual information and salient sentiment terms. Accuracy, precision, recall, F1-score and the confusion matrices also demonstrated that the deep learning models were more trustworthy when evaluating on the unseen reviews. Exploratory data analysis also showed some useful trends (e.g. word distribution and average review length between positive and negative reviews). There is also a simplistic prediction functionality where users can input a review or select a sample index and get to know at the instant whether the sentiment is positive or negative, modifying the system from a purely academic perspective to an effective and practical tool for real world use.

```
FOOD REVIEW SENTIMENT CLASSIFIER – DEEP LEARNING MODELS

Available Models:
1) BiLSTM
2) TextCNN
3) Exit

Enter model number or name: 1

Example test indices: [468137, 202140, 495361, 222236, 140116, 459542, 494870, 368982, 18603, 495816]

Enter test set index to predict: 202140

--- Using BiLSTM ---
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 40ms/step

Prediction for Test Index: 202140
Predicted Sentiment : Positive
Actual Sentiment    : Positive

Prediction Score    : 0.9108

Review Text:

Good cookies and a fun wrapper These biscuits are very good with red wine. The wrapper can also be used for a party trick.
```

PERFORMANCE

- The machine learning models offer a powerful baseline of accuracy, with Linear SVM being the most accurate among them.
- Deep learning models such as BiLSTM and TextCNN obtained superior results as they captured context information and key sentiment patterns.
- Metrics such as accuracy, precision, recall, and F1 score indicated that deep learning models well-generalized to non-training reviews.

CONCLUSION

This makes us your one-stop solution for integrated sentiment analysis pipeline for amazon fine food reviews. It consists of data cleaning, text pre-processing, feature extraction, and training of machine learning and deep learning models. TF-IDF-based models were robust baselines, while BiLSTM and TextCNN were more accurate as they better captured not only the meaning of the words but also the order. The system also has a user friendly predict option, so that it is useful in real-time sentiment classification. In general, the project serves as a powerful and scalable framework to analyze customer opinions and identify product satisfaction trends, which can help consumers make purchase decisions.

FUTURE ENHANCEMENTS

- Train deep learning models for more epochs or with a larger embedding size to improve performance.
- Employ transformer-based models such as BERT to further enhance sentiment understanding.
- Integrate aspect-based sentiment analysis (eg taste, packaging, delivery).
- Develop a full-blown web application/REST API for greater accessibility.