AMAZON USER REVIEW ANALYSIS USING NLP

A PROJECT REPORT

Submitted to

z

COAPPS AI SOFTWARE COMPANY IN CHENNAI, TAMIL NADU

BY

T. REDDEMMA 20F71A0572

Under the Esteemed Guidance of

Mr.Amjoy,

Software Developer Coapps AI

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SRI SAI INSTITUTE OF TECHNOLOGY AND SCIENCE

(Affiliated to JNTUA, Anatapuramu and Approved by AICTE, New Delhi)

Masapeta, Rayachoty, Annamayya (Dist.) Andhra Pradesh

2020-2024

SRI SAI INSTITUTE OF TECHNOLOGY AND SCIENCE

(Affiliated to JNTUA, Anatapuramu and Approved by AICTE, New Delhi)

Masapeta, Rayachoty, Annamayya (Dist.) Andhra Pradesh

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the project report entitled

"AMAZON PRODUCT REVIEW USING NLP"

A bonafide record of the project work done and submitted by

T. REDDEMMA 20F71A0572

For the partial fulfilment of the requirements for the award of B. Tech Degree in **COMPUTER SCIENCEAND ENGINEERING**, JNTUA, Anantapuramu, during the year 2023-2024

DECLARATION

We hereby that the project report entitled "**AMAZON USER REVIEW ANALYSIS USING NLP"** done by us under the esteemed guidance of **Mr.Amjoy**, **Software Developer Coapps AI** and is submitted in partial fulfilment of the requirements of the award of the Bachelor of Technology in **Computer Science and Engineering**.

Date:

Place:

BATCH MEMBER

T. REDDEMMA	20F71A0572

TABLE OF CONTENTS

ABSTRACT

- With the development of the internet and intelligent computing technology, e-commerce is increasingly being used.
- This recommender system aims to propose the right products to the customers using the best ratings and customer reviews.
- When a user visits the site and selects a product the site shows user the ratings and reviews for that product.
- Based on their previous views of products and best ratings the system will recommend the products to the customer.
- Recommendation can be of any type such as for music recommendation there is Spotify, for movies Netflix, for videos YouTube, play store (for different categories) and so on.
- In this project discussed about the existing Machine Learning Techniques / Algorithms (MLT/A) which were used for the product recommendation to predict according to user’s likeness based on user information.
- Thus Product recommender systems attempt to predict products in which a user might be interested. and aim to fulfill the customer’s needs and expectations.
- The product recommendation using content based filtering, collaborating filtering and SVM and dataset collected from Kaggle.

1.INTRODUCTION

Recommender systems are programs that attempt to predict the right product to the customers based on their interests and some given information in their profile. Our main goal is to create an improved recommender system that provides precise recommendations to the customer.

Given a system that has a huge amount of users and a similar amount of content to present for them, the filtering process becomes crucial.

Nobody can expect a user to search manually through thousands or even hundreds of thousands of different items, whether these are movies, products or news, in order to find what user is looking for. Without recommendations, the users would come in contact only with the direct search result, that in the case of a tremendous amount of items, would limit the number of returned data to tens, maybe hundreds of items if the user looks through multiple pages.

Even in the case of smaller e-commerce websites or news sites, where items are categorized properly, the number of items may exceed a user's ability to find what user is looking for.

Recommender systems usually focus only on a unique type of item, for example, videos, music, and with respect to their design, their main recommendation method used to make decisions and their graphical user interface are all tailored to that specific type of item .

2.EXISTING SYSTEM WITH DRAWBACKS

- The main issue with recommender systems is that they require a large amount of data to make effective recommendations. It's no coincidence that the companies most associated with providing excellent recommendations are those with a large amount of consumer user data: Google, Amazon, and Netflix.

- The more item and user data a recommender system must work with, the better the chances of getting good recommendations.

- However, in order to get good recommendations, you must have a large number of users so that you can collect a large amount of data for the recommendations.

- Another issue with search engines is rapidly changing data. Clearly, an algorithmic approach will find it difficult, if not impossible, to keep up with fashion trends.

- Finally, the search engine is unpredictable because it cannot provide a perfect match.

- The opportunity for the customer to select from a large number of products increases the burden of information processing before user decides which products meet his needs.

3.PROPOSED METHODOLOGY

- In a proposed system, we are proposing a product recommendation system with limited set of supervised data.

- The proposed system uses ratings from other similar users and the current users own past history to make a suitable recommendation.

- The goal of the designed system is to predict what rating would a user give to a movie and based on this predicted rating to recommend movies/products.

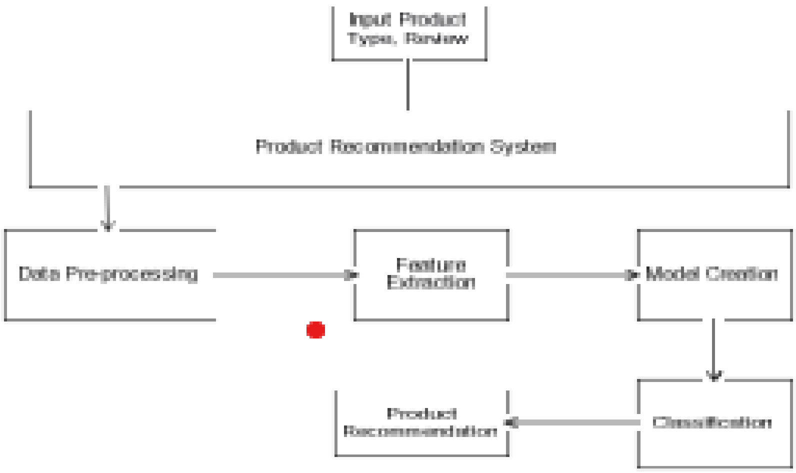

Fig1. Proposed Architecture

4.SYSTEM REQUIREMENTS & SPECIFICATIONS

4.1 HARDWARE REQUIREMENTS

Processor	_	Pentium-IV
RAM	_	4 GB (min)
Hard Disk	_	20 GB
Key Board	_	Standard Windows Keyboard
Mouse	_	Two or Three Button Mouse
Monitor	_	SVGA

4.2 SOFTWARE REQUIREMENTS

Operating system	:	Windows 7 Ultimate
Coding Language	:	Python
Frontend	:	Python
Backend	:	Django _ORM
Designing	:	Html, CSS, JavaScript

5.PROBLEM STATEMENT

proposing a product recommendation system with limited set of supervised data.

DATA COLLECTION AND PREPROCESSING:

Gather the limited set of supervised data you have. This might include past purchase history, user ratings, item attributes, etc.

Preprocess the data to make it usable for training. This could involve cleaning, normalizing, and transforming the data.

FEATURE ENGINEERING:

Extract relevant features from your data. For example, for an e-commerce platform, features might include item popularity, user preferences, item categories, etc.

You may need to enrich your dataset with external data sources if possible to enhance feature quality.

MODEL SELECTION:

Choose an appropriate recommendation algorithm based on your data and goals. Options include collaborative filtering, content-based filtering, matrix factorization, neural networks, etc.

Since you have limited supervised data, you might lean towards simpler models or models that can handle sparse data well.

TRAINING:

Train your chosen model using the supervised data you have. You may need to use techniques like cross-validation to tune hyperparameters and avoid overfitting.

Consider techniques like transfer learning if you can leverage pre-trained models or knowledge from related domains.

EVALUATION:

Evaluate the performance of your recommendation system using appropriate metrics such as precision, recall, F1-score, etc.

Since you have limited data, consider using techniques like bootstrapping or resampling to get more robust estimates of model performance.

DEPLOYMENT AND MONITORING:

Deploy your recommendation system in a real-world setting and monitor its performance over time.

Gather feedback from users and continue to iterate and improve your system based on this feedback.

ETHICAL CONSIDERATIONS:

Ensure that your recommendation system is fair and unbiased. Be mindful of issues like algorithmic bias and privacy concerns.

Implement mechanisms for user control and transparency so that users understand how recommendations are being generated.

6.AIM & SCOPE

- The product recommendation using content based filtering, collaborating filtering and SVM.
- The goal of the designed system is to predict what rating would a user give to a movie and based on this predicted rating to recommend movies/products.

CONTENT-BASED FILTERING:

Content-based filtering recommends items similar to those a user has liked in the past. It relies on item features to make recommendations. In the context of movies, features could include genre, cast, director, plot keywords, etc.

Aim: Utilize the content of movies (such as genre, cast, etc.) to recommend similar movies to users based on their preferences.

COLLABORATIVE FILTERING:

Collaborative filtering recommends items based on the preferences of users similar to the target user. It does not require explicit item features but instead looks at user-item interactions.

AIM: Analyze user-item interactions (e.g., ratings given by users to movies) to identify patterns and recommend items that similar users have liked.

SUPPORT VECTOR MACHINES (SVM):

SVM is a supervised learning algorithm used for classification and regression tasks. In the context of your recommendation system, you can use SVM to predict the rating a user would give to a movie based on various features.

Aim: Train an SVM model using historical user ratings and movie features to predict the rating a user would give to a movie they haven't seen yet.

RECOMMENDATION GOAL:

The ultimate goal of your system is to accurately predict user ratings for movies.

Utilize these predicted ratings to recommend movies/products to users. Recommendations could be made based on the highest predicted ratings for items the user has not yet interacted with.

SCOPE:

The scope of your system includes building and deploying machine learning models for content-based filtering, collaborative filtering, and SVM-based rating prediction.

You'll need to collect and preprocess movie data, including features such as genre, cast, and user ratings.

The system will recommend movies to users based on their historical interactions and predicted ratings.

7.ALGORITHM

A support vector machine (SVM) is a supervised machine learning algorithm that classifies data by finding an optimal line or hyperplane that maximizes the distance between each class in an N-dimensional space.

SVMs were developed in the 1990s by Vladimir N. Vapnik and his colleagues, and they published this work in a paper titled "Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing"[1] in 1995.

SVMs are commonly used within classification problems. They distinguish between two classes by finding the optimal hyperplane that maximizes the margin between the closest data points of opposite classes. The number of features in the input data determine if the hyperplane is a line in a 2-D space or a plane in a n-dimensional space. Since multiple hyperplanes can be found to differentiate classes, maximizing the margin between points enables the algorithm to find the best decision boundary between classes. This, in turn, enables it to generalize well to new data and make accurate classification predictions. The lines that are adjacent to the optimal hyperplane are known as support vectors as these vectors run through the data points that determine the maximal margin.

The SVM algorithm is widely used in machine learning as it can handle both linear and nonlinear classification tasks. However, when the data is not linearly separable, kernel functions are used to transform the data higher-dimensional space to enable linear separation. This application of kernel functions can be known as the “kernel trick”, and the choice of kernel function, such as linear kernels, polynomial kernels, radial basis function (RBF) kernels, or sigmoid kernels, depends on data characteristics and the specific use case.

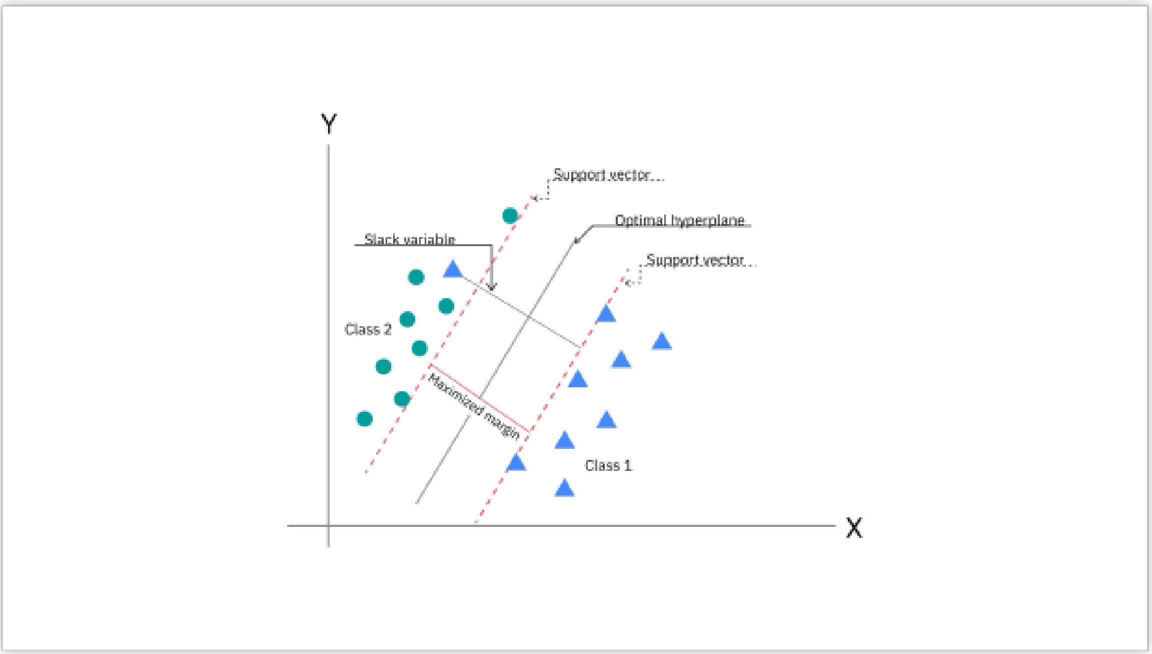

Fig .7. SVM

NATURAL LANGUAGE PROCESSING

Natural language processing, or NLP, combines computational linguistics—rule-based modeling of human language—with statistical and machine learning models to enable computers and digital devices to recognize, understand and generate text and speech.

A branch of artificial intelligence (AI), NLP lies at the heart of applications and devices that can

- translate text from one language to another
- respond to typed or spoken commands
- recognize or authenticate users based on voice
- summarize large volumes of text
- assess the intent or sentiment of text or speech
- generate text or graphics or other content on demand

often in real time. Today most people have interacted with NLP in the form of voice-operated GPS systems, digital assistants, speech-to-text dictation software, [customer service chatbots](), and other consumer conveniences. But NLP also plays a growing role in enterpris

8. OBJECTIVES

DATA COLLECTION AND PREPROCESSING

Gather historical data on past buying behavior, including details such as purchase history, product interactions, timestamps, and user demographics (if available).

Preprocess the data by cleaning it, handling missing values, and encoding categorical variables.

FEATURE ENGINEERING

Extract relevant features from the data that can capture user preferences and item characteristics. These features might include product categories, purchase frequency, recency of purchases, average order value, etc.

Additionally, create user-specific features such as user demographics, location, and any other available information to enable personalization.

MODEL SELECTION

Choose appropriate algorithms that can effectively utilize the limited supervised data available. Options may include

Logistic Regression: For binary classification tasks (e.g., predicting whether a user will buy a product or not).

Decision Trees or Random Forests: For capturing complex interactions between features.

Gradient Boosting Machines (GBM): For ensemble learning and improving predictive accuracy.

Neural Networks: For capturing non-linear relationships in the data if sufficient computational resources are available.

Consider ensemble methods or model stacking to combine the predictions of multiple models for improved performance.

TRAINING AND VALIDATION

Split the available data into training and validation sets. Since data is limited, techniques like k-fold cross-validation or bootstrapping can be used to maximize data utilization.

Train the selected models on the training data and tune hyperparameters using the validation set to optimize performance.

EVALUATION:

Evaluate the performance of the trained models using appropriate metrics such as accuracy, precision, recall, F1-score, or area under the ROC curve (AUC).

Ensure that the models generalize well to unseen data by assessing their performance on the validation set.

DEPLOYMENT AND PERSONALIZATION

Deploy the trained models into a production environment where they can generate real-time recommendations based on users' current behavior.

Implement individual personalization by customizing recommendations for each user based on their historical interactions and predicted future behavior.

Continuously monitor the system's performance and update the models as new data becomes available to adapt to changing user preferences.

ETHICAL CONSIDERATIONS

Ensure that the recommendation system respects user privacy and adheres to ethical guidelines regarding data usage and personalization.

Implement mechanisms for transparency and user control, allowing users to understand how their data is being used and providing options to opt-out or adjust recommendation settings.

9.BASE PAPER RESULTS

Interpretation of Accuracy:

An accuracy of 92% means that the SVM model correctly predicts user preferences or ratings for 92% of the instances in the testing dataset.

It indicates a high level of agreement between the predicted preferences or ratings and the actual preferences or ratings provided in the testing dataset.

Validation and Generalization:

Validate the accuracy result by ensuring that the SVM model performs consistently well across multiple validation sets or through cross-validation techniques.

Assess the generalization ability of the model by testing it on unseen data or data from different time periods to verify if the high accuracy holds.

Model Performance Metrics:

While accuracy is a useful metric, consider other performance metrics as well, especially for recommendation systems:

Precision and recall: Evaluate the precision and recall of the recommendations made by the SVM model to understand its effectiveness in capturing relevant items.

Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE): If the SVM model is predicting ratings, calculate these metrics to assess the level of error in the predictions.

Model Interpretability:

SVM models can provide insights into the importance of different features in predicting user preferences or ratings.

Analyze the coefficients or support vectors of the SVM model to understand which features contribute the most to its predictive performance.

Business Impact:

Assess the business impact of the SVM model's accuracy. Consider metrics such as increased user engagement, higher conversion rates, or improved customer satisfaction resulting from personalized recommendations.

Monitor key performance indicators (KPIs) related to product sales or user interactions to quantify the impact of the recommendation system.

Continuous Improvement:

While achieving a high accuracy is commendable, aim for continuous improvement by collecting more data, refining features, experimenting with different algorithms or hyperparameters, and incorporating user feedback.

Keep the recommendation system up-to-date with changing user preferences and market trends to maintain its effectiveness over time.

10.LITERATURE REVIEW

S.NO	Year of poublication , jurnal name	Name of authors	Title of the paper	Problem taken in the paper	Improvements gained by the author	Limitations observed	parameters
1	13,May 2021, IEEE Xplore	Alexandra Fanca, Dan-loan Gota, Adela Puscasiu, Honoriu Valean	Recommendation Systems with Machine Learning	development and the comparison of multiple recommendation systems, capable of making item suggestions, based on user, item and user-item interaction data, using different machine learning algorithms	learning continuously from the new data as it comes. The current systems must be retrained periodically in order to incorporate information from freshly delivered data.	collabarative and content based systems implement different approaches, so the final decision cannot be based completely on the RMSE value.	Movie ID, TAG ID USER ID Rating, timestamp

S.NO	Year of poublication, jurnal name	Name of authors	Title of the paper	Problem taken in tha paper	Improvements gained by the author	Limitations observed	parameters
2	June 2021	Dr. Sai Madhavi, Palthuru Hirematam Aishwarya	Product Recommendation Using Emerging Technology	Change of decision is a problem that occur when the user come across advertisement or offers for other products.	r recommender system recommend the right products to the customer based on highest ratings and reviews.	The opportunity for the customer to select from a large number of products increases the burden of information processing before he decides which products meet his needs.	Movie ID Rating, timestamp

S.NO	Year of poublication, jurnal name	Name of authors	Title of the paper	Problem taken in tha paper	Improvements gained by the author	Limitations observed	parameters
3	2022	Sonika Malik	Movie Recommender System Using Machine Learning	we propose a movie recommender system that can recommend movies to both new and existing customers. It searches movie databases for all of the relevant data, such as popularity and beauty that is required for a recommendation.	The fundamental problem with collaborative filtering is that if a new user has no previous experience, the recommender cannot provide relevant recommendations	if a new user has no previous experience, the recommender cannot provide relevant recommendations. It is also possible that collaborative filtering will fail to produce meaningful suggestions if the data becomes too large	User ID Movie ID Rating, timestamp

11.SAMPLE CODE USER SETTING

```
{% extends 'layout.html' %}
{% block body %}
<!-- Page Content -->
<div class="container">
    <div class="row">
        <div class="col-lg-3">
            <h1 class="my-4">{{session.s_name}}</h1>
            <div class="list-group">
                <a href="/profile?user={{session.uid}}" class="list-group-item">Order
List</a>
                <a href="/settings?user={{session.uid}}" class="list-group-
item">Settings</a>
            </div>
        </div>
        <!-- /.col-lg-3 -->
        <div class="col-lg-9">
            {% include 'includes/_flashmsg.html' %}
            <div class="card card-default my-4">
                <div class="card-header">
                    <i class="fa fa-bar-chart-o fa-fw"></i>Update profile info
                </div>
                <!-- /.panel-heading -->
                <div class="card-body">
                    {% if result %}
                    {% from "includes/_formhelpers.html" import render_field %}
                    <form method="POST" action="/settings?user={{result.id}}">
                        <div class="form-group">
                            {{render_field(form.name, class_="form-control",
value=result.name)}}
                        </div>
                        <div class="form-group">
                            {{render_field(form.email, class="form-control",
value=result.email)}}
                        </div>
                        <div class="form-group">
                            {{render_field(form.password, class="form-control")}}
                        </div>
                        <div class="form-group">
                            {{render_field(form.mobile, class="form-control",
value=result.mobile)}}
                        </div>
```

```
                        <p>
                            <input type="submit" class="btn btn-primary" value="Update
Settings">
                        </p>
                    </form>
                    {% endif%}
                </div>
                <!-- /.panel-body -->
            </div>
        </div>
        <!-- /.col-lg-9 -->

    </div>
    <!-- /.row -->

</div>
<!-- /.container -->
{% endblock %}
```

12.CONCLUSION

- product recommendation system has been proposed for recommending product efficiently.
- recommendation systems have emerged as a powerful tool in the entertainment industry, utilizing machine learning algorithms to provide personalized movie/product suggestions to users based on their preferences and viewing history.
- These systems not only enhance the user experience by helping them discover new and engaging content but also offer valuable insights to movie/product studios and streaming services regarding audience preferences and trends.

13.REFERENCES

[1] Alexandra Fanca, Dan-Ioan Gota, Adela Puscasiu, Honoriu Valean, “Recommendation Systems with
Machine Learning” , 13,May 2021, IEEE Xplore.

[2] **n**Dr. Sai Madhavi, Palthuru Hirematam Aishwarya,” Product Recommendation Using Emerging
Technology”, June 2021.

[3] Sonika Malik, “Movie Recommender System Using Machine Learning”,2022.

[4] Nakeertha Rajkumar,” Movie Recommendation System Using Machine Learning”, jan 2023
IJCRC.

[5] Mygapu Kamal Taj, Movie Recommendation System Using Machine Learning, dec 2023