Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

A Mood-Based Movie Recommendation System Using Positive and Negative Affect Schedule and Machine Learning Algorithms

Anand Prakash Srivastava[1], Dr. Sakuldeep Singh[2]

[1]*Research Scholar, Computer Science & Engineering, Sanskriti University, India, anand.infotech@gmail.com*
[2]*Associate Professor, Computer Science & Engineering, Sanskriti University, India, sakuldeeps.cse@sanskriti.edu.in*

This study presents a movie recommendation system tailored to the user's mood, using the Positive and Negative Affect Schedule (PANAS) and cosine similarity for recommendations. The system categorizes movies into genres that evoke positive or negative emotions, matching the user's current emotional state. The user's mood is determined by their responses to a PANAS questionnaire, guiding the selection of the appropriate genre dataset. From this dataset, a random movie is chosen, and ten similar movies are recommended using cosine similarity. The system's effectiveness was evaluated using three machine learning algorithms: Naive Bayes (NB) (Gaussian, Multinomial, Bernoulli), Support Vector Machine (SVM) with linear and radial basis function (rbf) kernels, and Decision Tree (DT) with log loss, gini, and entropy criteria. Models were trained on datasets comprising 70%, 75%, and 80% of the available data, and their performance was assessed. Results indicated that the DT method, particularly with the gini criterion, achieved the highest accuracy, while the SVM also performed well. Naive Bayes showed the lowest accuracy. The DT algorithm's consistent and superior performance highlights its suitability for this recommendation system, whereas NB was less effective for this application.

Keywords: Movie recommendation system, PANAS, Mood-based recommendations, Genre categorization, User mood recommendation.

1. Introduction

Recommendation systems [1] are used in various sectors like e-commerce, retail banking and entertainment. The primary objective of these systems is to provide customers with recommendations derived from their continuously acquired and processed data. The most popular methods for recommendation system [2] are Content-based Filtering (CBF), Collaborative Filtering (CF) and Hybrid Filtering. With the use of CBF (Content based filtering) technique [3] which checks about the features of each item and suggests other items that have similar attributes. By exploring the correspondence between users and items, CF [4, 5] improves some drawbacks of CBF and provides suggestions. It uses the data of user's past selection and other likeminded users' choice to offer personalized recommendations. Many existing recommendation systems (RS) use a hybrid-filtering (HF) technique [6] that mixes the strengths of Content-Based filtering (CBF) and Collaborative Filtering (CF). A film rides on audiences feedback. Other users rely so much on these reviews while making their own choices. People tend to be more likely to choose a movie that has been well-received by the majority, rather than one that has been mostly disliked. Considering these reviews, excluding those that contain misleading information, also adds to the intricacy of decision-making. There is a potential solution to this problem through sentiment analysis. Utilizing Natural Language Processing (NLP), Sentiment Analysis [7] allows for the extraction of information from textual sources and the classification of statements, words, or documents as positive or negative.

Recognizing the distinctions between emotions and affect is crucial when addressing moods [8]. Psychological research distinguishes between emotions and moods, indicating that emotions are intense feelings aimed at a specific entity, whereas moods are milder affective states that do not necessitate a particular stimulus. [9] offers an extensive definition of affect, encompassing both emotions and moods. Numerous attempts were undertaken to categorize emotions into distinct dimensions [10]. In summary, it was determined that mood can be categorized into a positive affect or a negative affect dimension. Positive affect mostly pertains to an individual's feelings of enthusiasm, activity, and alertness [11]. Negative affect encompasses a broad spectrum of human suffering and adverse engagement, comprising various unwanted emotional states [11]. Each of these dimensions has been subdivided into a list of 10 items, with each item reflecting a distinct mood state. In assessing an individual's mood using the PANAS (Positive Affect-Negative Affect Scale), respondents assign a rating from 1 to 5 for each item, reflecting the degree of presence or intensity. The responses are compiled to yield a score that reflects the user's current emotional state. Considering that this calculated score neglects supplementary factors, such as interdependencies within the questionnaire, it is crucial to assess the significance of incorporating all individual replies [11, 12]. We evaluate our system using three machine learning algorithms namely, Naive Bayes (NB), Support Vector Machine (SVM), and Decision Tree (DT) across different data splits. In contrast to prior studies, we provide a comprehensive comparison with state-of-the-art methods, demonstrating the robustness and competitive performance of our approach. The DT algorithm, particularly with the gini criterion, consistently outperforms others, confirming its effectiveness for this application. The rest of the paper is structured as follows: Section 2 details the methodology, including the dataset, PANAS scale, and machine learning algorithms. Section 3 presents the experimental results and comparative analysis. Section 4 concludes the study with a discussion on findings and future work.

2. Methodology

A. Dataset

Kaggle provided two datasets, namely tmdb 5000 movies and tmdb 5000 movies. Both csv files namely movies and credits [13] were used with each file containing 20 and 4 characteristics, respectively. Both dataset have been used for Movie Recommendation system. Movies dataset consists of features namely budget, genre, homepage, id, keywords, original language, original title, overview, popularity, production company, production countries, release date, revenue, runtime, spoken language, status, tagline, title, vote average and vote count. Credits dataset consists of features namely movie id, title, cast and crew.

	movie_id	title	cast	director	keywords	genres
0	19995	Avatar	[Sam Worthington, Zoe Saldana, Sigourney Weaver]	James Cameron	[culture clash, future, space war]	[Action, Adventure, Fantasy]
1	285	Pirates of the Caribbean: At World's End	[Johnny Depp, Orlando Bloom, Keira Knightley]	Gore Verbinski	[ocean, drug abuse, exotic island]	[Adventure, Fantasy, Action]
2	206647	Spectre	[Daniel Craig, Christoph Waltz, Léa Seydoux]	Sam Mendes	[spy, based on novel, secret agent]	[Action, Adventure, Crime]
3	49026	The Dark Knight Rises	[Christian Bale, Michael Caine, Gary Oldman]	Christopher Nolan	[dc comics, crime fighter, terrorist]	[Action, Crime, Drama]
4	49529	John Carter	[Taylor Kitsch, Lynn Collins, Samantha Morton]	Andrew Stanton	[based on novel, mars, medallion]	[Action, Adventure, Science Fiction]

Fig. 1: Merged Dataset

The two datasets used in movie recommendation are merged to form a single dataset keeping relevant columns shown in Figure 1. The columns kept under it include the movie ID, title, genre and tags shown in Figure 2.

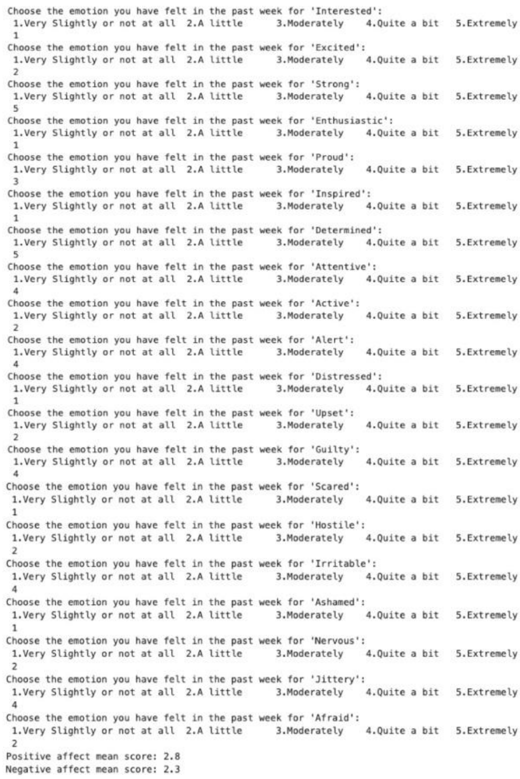
```
                       genres  movie_id                  title  \
0      [action, crime, drama]     49026  The Dark Knight Rises
1  [adventure, drama, action]       254              King Kong
2  [drama, romance, thriller]       597                Titanic
3     [action, drama, horror]     72190            World War Z
4            [drama, romance]     64682        The Great Gatsby

                                                tags
0  dccomics crimefighter terrorist christianbale ...
1  filmbusiness screenplay showbusiness naomiwatt...
2  shipwreck iceberg ship katewinslet leonardodic...
3  dystopia apocalypse zombie bradpitt mireilleen...
4  basedonnovel infidelity obsession leonardodica...
```

Fig. 2: Final Dataset

B. PANA Scale

The Positive and Negative Affect Schedule (PANAS) is a psychometric instrument designed to evaluate positive and negative affect. The study comprises 20 items that encompass a variety of emotions. Participants are requested to evaluate these issues using a 5-point Likert scale, from “Very Slightly or Not at All” to “Extremely”. Positive Affect (PA) include sensations of enthusiasm and vigor, whilst Negative Affect (NA) involves distress and unpleasant involvement. The PANAS is employed in clinical and community contexts to assess emotional states and their relationship with personality factors. Its applications include monitoring fluctuations in clients' moods over time [15]. The same has been utilized as a recommended approach to assess either the good or negative affect of a person based on their mood, as illustrated in Figure 3.



```
Choose the emotion you have felt in the past week for 'Interested':
 1.Very Slightly or not at all  2.A little      3.Moderately    4.Quite a bit   5.Extremely
 1
Choose the emotion you have felt in the past week for 'Excited':
 1.Very Slightly or not at all  2.A little      3.Moderately    4.Quite a bit   5.Extremely
 2
Choose the emotion you have felt in the past week for 'Strong':
 1.Very Slightly or not at all  2.A little      3.Moderately    4.Quite a bit   5.Extremely
 5
Choose the emotion you have felt in the past week for 'Enthusiastic':
 1.Very Slightly or not at all  2.A little      3.Moderately    4.Quite a bit   5.Extremely
 1
Choose the emotion you have felt in the past week for 'Proud':
 1.Very Slightly or not at all  2.A little      3.Moderately    4.Quite a bit   5.Extremely
 3
Choose the emotion you have felt in the past week for 'Inspired':
 1.Very Slightly or not at all  2.A little      3.Moderately    4.Quite a bit   5.Extremely
 1
Choose the emotion you have felt in the past week for 'Determined':
 1.Very Slightly or not at all  2.A little      3.Moderately    4.Quite a bit   5.Extremely
 5
Choose the emotion you have felt in the past week for 'Attentive':
 1.Very Slightly or not at all  2.A little      3.Moderately    4.Quite a bit   5.Extremely
 4
Choose the emotion you have felt in the past week for 'Active':
 1.Very Slightly or not at all  2.A little      3.Moderately    4.Quite a bit   5.Extremely
 2
Choose the emotion you have felt in the past week for 'Alert':
 1.Very Slightly or not at all  2.A little      3.Moderately    4.Quite a bit   5.Extremely
 4
Choose the emotion you have felt in the past week for 'Distressed':
 1.Very Slightly or not at all  2.A little      3.Moderately    4.Quite a bit   5.Extremely
 1
Choose the emotion you have felt in the past week for 'Upset':
 1.Very Slightly or not at all  2.A little      3.Moderately    4.Quite a bit   5.Extremely
 2
Choose the emotion you have felt in the past week for 'Guilty':
 1.Very Slightly or not at all  2.A little      3.Moderately    4.Quite a bit   5.Extremely
 4
Choose the emotion you have felt in the past week for 'Scared':
 1.Very Slightly or not at all  2.A little      3.Moderately    4.Quite a bit   5.Extremely
 1
Choose the emotion you have felt in the past week for 'Hostile':
 1.Very Slightly or not at all  2.A little      3.Moderately    4.Quite a bit   5.Extremely
 2
Choose the emotion you have felt in the past week for 'Irritable':
 1.Very Slightly or not at all  2.A little      3.Moderately    4.Quite a bit   5.Extremely
 4
Choose the emotion you have felt in the past week for 'Ashamed':
 1.Very Slightly or not at all  2.A little      3.Moderately    4.Quite a bit   5.Extremely
 1
Choose the emotion you have felt in the past week for 'Nervous':
 1.Very Slightly or not at all  2.A little      3.Moderately    4.Quite a bit   5.Extremely
 2
Choose the emotion you have felt in the past week for 'Jittery':
 1.Very Slightly or not at all  2.A little      3.Moderately    4.Quite a bit   5.Extremely
 4
Choose the emotion you have felt in the past week for 'Afraid':
 1.Very Slightly or not at all  2.A little      3.Moderately    4.Quite a bit   5.Extremely
 2
Positive affect mean score: 2.8
Negative affect mean score: 2.3
```

Fig. 3: PANA Scale

C. Cosine Similarity

Cosine similarity [16, 17] is a metric used to evaluate the similarity between two vectors. It neglects the magnitude of the vectors and focuses on calculating the cosine of the angle between them. In the context of movie recommendation systems, it is used to measure the similarity between users or movies based on ratings or other features. The similarity measure plays an important role in collaborative filtering techniques, which serve as the basis for numerous recommendation systems. Mathematically, the cosine similarity between two vectors C and D is defined as:

$$\text{Cosine Similarity (I, J)} = \frac{\sum_{i=1}^{n} I_i \cdot J_i}{\sqrt{\sum_{i=1}^{n} I_i^2} \cdot \sqrt{\sum_{i=1}^{n} J_i^2}} \qquad (1)$$

where, I ⋅ J is the dot product of vectors I and J. ‖I‖ and ‖J‖ are the magnitudes of vectors I and J, respectively. I_i and J_i are the components of vectors I and J at dimension i.

C. Support Vector Machine

The Support Vector Machine (SVM) [18, 19] is a supervised learning algorithm that was developed by Vladimir Vapnik in the 1990s. It is used for classification tasks and can also be applied to regression problems. The SVM algorithm operates by identifying the hyperplane that efficiently segregates data points belonging to distinct groups. In an n-dimensional space, the equation of a hyperplane can be expressed as:

$$a \times q + e = 0 \tag{2}$$

The weight vector is denoted by a, the input features are represented by q, and the bias term is denoted by e. The objective is to optimize the margin, which is defined as the ratio of 2 divided by the magnitude of vector a. This optimization is subject to the condition that the data points be accurately identified:

$$y_i(a \times q_i + e) \geq 1 \tag{3}$$

In situations where it is challenging to achieve a complete separation of classes due to noise or overlapping data points, SVM introduces the concept of a soft margin. This requires the incorporation of slack variables ξ_i to the optimization problem, which permits a certain degree of misclassification while also imposing a penalty through a regularization parameter C. The optimization objective changes:

$$\min \frac{1}{2}\|a\|^2 + C\sum_{i=1}^{n}(\xi_i) \tag{4}$$

$$\text{s.t}\begin{Bmatrix} y_i(a \times q_i + e) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{Bmatrix} \tag{5}$$

The linear kernel is the simplest type of kernel, where the decision boundary is a straight line or hyperplane in higher dimensions. It is given by

$$K(q_i, q_j) = q_i \times q_j \tag{6}$$

The radial basis function (rbf) kernel, also known as the Gaussian kernel, is a popular choice for non-linear data. It is defined as

$$K(q_i, q_j) = \exp(-\gamma\|q_i - q_j\|^2) \tag{7}$$

where γ is a parameter that determines the extent of the kernel's distribution.

D. Decision Tree

A Decision Tree (DT) [20–24] is a powerful tool in the field of machine learning, capable of handling both classification and regression tasks with ease. Log loss is a metric that evaluates the effectiveness of a classification model by considering the probability values it generates, ranging from 0 to 1. The objective of utilizing log loss is to reduce the disparity between the predicted probability and the true class label. The log loss for a binary classification problem is defined as follows:

$$\text{logloss} = -\frac{1}{k}\sum_{i=1}^{k}[y_i \log_2(q_i) + (1 - y_i)\log_2(1 - q_i)] \tag{8}$$

The gini Impurity quantifies the impurity level of a dataset by calculating the likelihood of randomly choosing an incorrect class, based on the distribution of classes within the dataset. The calculation is as follows:

$$\text{gini(Dataset)} = 1 - \sum_{i=1}^{k} q_i^2 \qquad (9)$$

where q_i is the ratio of instances belonging to class i in the dataset. The reduction in entropy or impurity after a dataset is split on an attribute is measured by Information Gain (IG). The calculation of the entropy for a dataset is as follows:

$$\text{Entropy(D)} = -\sum_{i=1}^{k} (q_i)\log_2(q_i) \qquad (10)$$

The calculation for determining the IG of an attribute A is as follows:

$$\text{IG(D, A)} = \text{Entropy(D)} - \sum_{v} \frac{D_v}{D} \text{Entropy}(D_v) \qquad (11)$$

where D_v is the subset of D where attribute A has value v.

E. Naive Bayes

Naive Bayes (NB) [25–28] is a straightforward yet remarkably powerful probabilistic classifier that relies on Bayes' theorem. It is especially well-suited for classification tasks in the field of machine learning. Bayes' theorem is a fundamental concept that forms the basis of NB. It allows us to calculate the probability of a hypothesis based on the evidence we observe. Bayes' theorem can be expressed as: $P(C|D) = \frac{P(C|D).P(C)}{P(D)}$ (12)

The expression P(C|D) represents the posterior probability of class C given feature D. The expression P(D|C) represents the conditional probability of feature D given class C. The term P(C) refers to the initial probability of class C. The term P(D) refers to the initial probability of feature D. The "naive” component of NB stems from the assumption of independence. It presupposes that all characteristics are unrelated to one another, provided the category is known. Here is the streamlined model:

$$P(E|Q_1, Q_2, \ldots\ldots. Q_n) \alpha\, P(E) \times \prod_{i=1}^{k} P(Q_i|C) \qquad (13)$$

In this context, P(E) represents the initial probability of class E, whereas $P(Q_i|E)$ represents the probability of feature Qi given class E. Three primary categories of NB classifiers exist, each designed to handle distinct data types: Gaussian NB is designed for continuous data, it assumes that the features adhere to a normal distribution. It calculates the mean and variance for each feature and class, and then uses these parameters to determine the likelihood. The probability density function for a Gaussian distribution is

$$P(q_i|r) = \frac{1}{\sqrt{2\pi\sigma_r^2}} \exp\left(-\frac{(q_i - \mu_r)^2}{2\sigma_r^2}\right) \qquad (14)$$

Multinomial NB is well-suited for analyzing discrete data, particularly word counts in text classification. The likelihood is modelled using a multinomial distribution, which proves to be highly effective for document classification tasks where the features represent word frequencies. The probability of a feature vector given a class is

$$P(x|y) = \frac{N_y!}{x_1!\,x_2!\ldots\ldots x_n!}\left(\frac{\theta_{y_1}^{x_1}\theta_{y_2}^{x_2}\ldots.\theta_{y_n}^{x_n}}{(\sum_i x_i)!}\right) \qquad (15)$$

Bernoulli's contribution NB is specifically designed to handle binary or Boolean features. The assumption is that every feature conforms to a Bernoulli distribution, which means it can be used effectively for tasks such as binary text classification, where features indicate whether words are present or absent. The probability of a feature vector given a class is

$$P(x|y) = \prod_{i:x_i=1} \theta_{yi} \prod_{i:x_i=0} (1-\theta_{yi}) \qquad (16)$$

Figure 4 illustrates the workflow of a mood-based movie recommendation system. The process begins with data collection followed by data preprocessing tasks like feature selection and dataset merging. Next, the system performs mood detection, classifying the user's mood into either Good Mood or Bad Mood. Based on the detected mood, the system proceeds to genre classification and selects an appropriate genre. Movies are then compared using Cosine Similarity to recommend relevant films. The dataset is split into training and test sets, and machine learning algorithms (NB, SVM, and DT) are trained and evaluated for accuracy. The workflow concludes with model evaluation and accuracy calculation, ensuring personalized and efficient movie recommendations. Algorithm 1 outlines the step-by-step logic for implementing a mood-based movie recommendation system.

Step 1: Begin

Step 2: Collect Datasets: tmdb_movies, tmdb_credits

Step 3: Preprocess Data: Feature Selection and Merge Datasets

Step 4: Detect User Mood

if Mood == "Good":

Classify Genre → Good Mood Genre

else if Mood == "Bad":

Classify Genre → Bad Mood Genre

Step 5: Select Relevant Genre

Step 6: Calculate Cosine Similarity Between Movies

Step 7: Recommend Top-N Similar Movies

Step 8: Split Dataset: Training, Testing

Step 9: Train Machine Learning Models:

Naive Bayes, Support Vector Machine (SVM) and Decision Tree

Step 10: Evaluate Models Using Test Data

Step 11: Report Accuracy for Each Model

Step 13: End

Algorithm 1: Pseudocode of the mood-based movie recommendation system

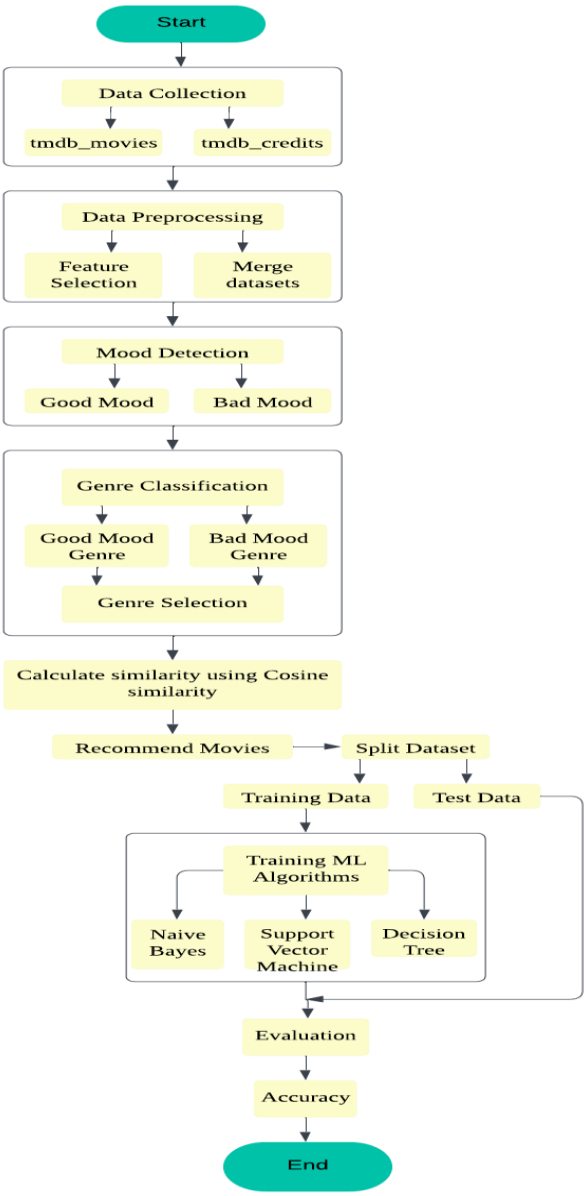


Fig. 4: Workflow diagram of the mood-based movie recommendation system

3. RESULT

Dataset is divided two datasets based on genre into good mood genre which consists of comedy, romance, drama, animation, fantasy, sci-fi, & music and bad mood genre consisting of drama, romance, documentaries, fantasy & music shown in Figure 5.

After the user has responded to all the questions in the questionnaire, positive effect means and negative effect mean is calculated.

```
Good Mood Genre
 ['Comedy', 'Romance', 'Drama', 'Animation', 'Fantasy', 'Sci-Fi', 'Music']
Bad Mood Genre
 ['Drama', 'Romance', 'Documentaries', 'Fantasy', 'Music']
```

Fig. 5: Good Genre and Bad Genre

Depending on the score good mood genre dataset or bad mood genre dataset is selected as shown in Fig. 6 and thereafter cosine similarity of a randomly selected genre-based dataset is calculated along with similarity score as shown in Figure 7.

	genres	movie_id	title	tags	score
0	[action, crime, drama]	49026	The Dark Knight Rises	dccomics crimefighter terrorist christianbale ...	2
1	[adventure, drama, action]	254	King Kong	filmbusiness screenplay showbusiness naomiwatt...	2
2	[drama, romance, thriller]	597	Titanic	shipwreck iceberg ship katewinslet leonardodic...	2
3	[action, drama, horror]	72190	World War Z	dystopia apocalypse zombie bradpitt mireilleen...	2
4	[drama, romance]	64682	The Great Gatsby	basedonnovel infidelity obsession leonardodica...	1
...	...	...	...	...	...
2251	[drama, foreign]	13898	The Circle	nargessmamizadeh maryiampalvinalmani mojganfa...	1
2252	[drama]	182291	On The Downlow	confession hazing gangmember tonysancho michae...	0
2253	[drama]	124606	Bang	gang audition policefake darlingnarita petergr...	0
2254	[sciencefiction, drama, thriller]	14337	Primer	distrust garage identitycrisis shanecarruth da...	2
2255	[comedy, drama, romance]	231617	Signed, Sealed, Delivered	date loveatfirstsight narration ericmabius kri...	2

Fig. 6: Movies data based on Drama genre

```
Cosine Similarity
 [[1.         0.02857697 0.00285389 ... 0.00246268 0.00229633 0.00276527]
 [0.02857697 1.         0.0026016  ... 0.00224498 0.00209333 0.00252082]
 [0.00285389 0.0026016  1.         ... 0.00227081 0.0181962  0.01830091]
 ...
 [0.00246268 0.00224498 0.00227081 ... 1.         0.00182717 0.0022003 ]
 [0.00229633 0.00209333 0.0181962  ... 0.00182717 1.         0.00205167]
 [0.00276527 0.00252082 0.01830091 ... 0.0022003  0.00205167 1.        ]]


Similarity Score
 [0.00289851 0.00264227 0.00267267 ... 0.00230631 0.00215052 0.00258969]
```

Fig. 7: Cosine similarity and Similarity score based on selected genre's dataset

Based on input, a movie title is randomly selected from the list of good mood genre dataset or bad mood genre dataset. After selecting a random movie, 10 similar movies are recommended along with their similarity score is provided to the user as shown in Figure 8.

To test the accuracy of Recommendation system, classification algorithms Gaussian NB, Multinomial NB, Bernoulli NB, SVM using linear kernel with value of regularization parameter C = 10 and rbf with the value of C = 10 and gamma = 0.05 and DT using criterion log loss, gini and entropy have been applied on 70%, 75% and 80% data as training dataset and the result has been formulated in Table 1. It can be seen that DT with criterion gini scored highest accuracy i.e. 91.87%, 98.40% and 97.34% for 70%, 75% and 80% training data respectively.

```
Recommendation for movie: Fireproof
Recommended movies with similarity scores:
Movie: Courageous, Similarity Score: 0.4026
Movie: Flywheel, Similarity Score: 0.3538
Movie: Facing the Giants, Similarity Score: 0.2514
Movie: Stone, Similarity Score: 0.1186
Movie: The Dead Girl, Similarity Score: 0.1157
Movie: Volver, Similarity Score: 0.1144
Movie: Desert Blue, Similarity Score: 0.1093
Movie: Far from Men, Similarity Score: 0.1083
Movie: Gangs of New York, Similarity Score: 0.1054
Movie: Poseidon, Similarity Score: 0.1047
```

Fig. 8: Recommended movies list

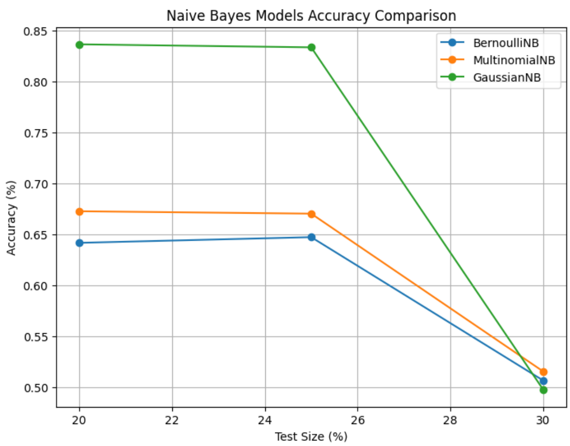
Algorithms	Training Sample		
	70%	75%	80%
Gaussian NB	49.78	83.33	83.63
Multinomial NB	51.55	67.02	67.26
Bernoulli NB	50.66	64.72	64.16
SVM (linear)	87.29	95.92	95.13
SVM (rbf)	81.38	94.32	93.58
Decision Tree (log_loss)	91.58	98.04	97.12
Decision Tree (gini)	91.87	98.40	97.34
Decision Tree (entropy)	91.87	98.04	97.34

Table 1: Comparison of Algorithms based on Accuracy (%)

NB has the accuracy range from 49.78% to 83.63%, SVM has the accuracy range from 81.38% to 95.92% and DT has the accuracy range from 91.58% to 98.40%. It can be stated that DT performed approximately same and is quite consistent for all three-training data while NB being the worst among three as shown in Figure 9. The receiver operating characteristic (ROC) curve for NB, SVM and DT is shown in Figure 10 which demonstrates that DT is the ideal choice for movie recommendation in every test scenario i.e. 20%, 25% and 30% having area under the curve (AUC) ranging from 0.99 to 1.00.

4. CONCLUSION

The objective of the paper was to develop a recommendation system that classifies movies into two unique genres, namely “good mood” and “bad mood”, based on their ability to impact a user’s emotional state. The happy mood genre comprises comedy, romance, drama, animation, fantasy, sci-fi, and music, whereas the bad mood genre consists of drama, romance, documentaries, fantasy, and music. This categorization guarantees that users are provided with film suggestions that are in line with their present emotional condition. Upon completion of a questionnaire, the responses provided by users are utilized to compute the average positive and negative impacts, so assessing their overall emotional state. Using these scores, the recommendation algorithm chooses either the dataset for the positive mood genre or the dataset for the negative mood genre. From the given dataset, a genre is selected at random, and then a movie title is chosen randomly from that genre. Afterwards, the system employs a recommendation algorithm based on cosine similarity to provide ten movies that are comparable to the chosen title. In order to evaluate the precision and efficiency of the recommendation system, several classification algorithms were utilized, including Gaussian NB, Multinomial NB, and Bernoulli NB, SVM with both linear and rbf kernels and DT with criteria such as log loss, gini, and entropy. The algorithms underwent testing on training datasets that consisted of 70%, 75%, and 80% of the entire data. The outcomes of these tests were then gathered and presented in a thorough table. The results showed that the DT method, specifically using the gini criterion, attained the maximum level of accuracy. It scored 91.87%, 98.40% and 97.34% when trained with 70%, 75% and 80% data. On the other hand, the NB algorithm showed the least accurate results, with accuracy ranging from 49.78% to 83.63%, indicating notable discrepancy. The SVM method demonstrated satisfactory performance, with accuracy ranging from 81.38% to 95.92%. The DT algorithm's consistent and strong performance on all training datasets indicates that it is the most dependable approach for this recommendation system. The fact that it can consistently achieve accuracy levels ranging from 91.58% to 98.40% demonstrates its resilience and appropriateness for the given task. However, the performance of NB suggests that it may not be suitable for this particular application, as it exhibits a broad range of accuracy and lower total scores. Same trends can be observed from the ROC curve of 70%, 75% and 80% of training data with having AUC 0.99 to 1.00 in all three cases. This is followed by SVM which showed almost similar performance in 75% and 80% category i.e. 1.00 but in 70% it scored 0.99 which got outperformed by DT. NB scored the least AUC score.



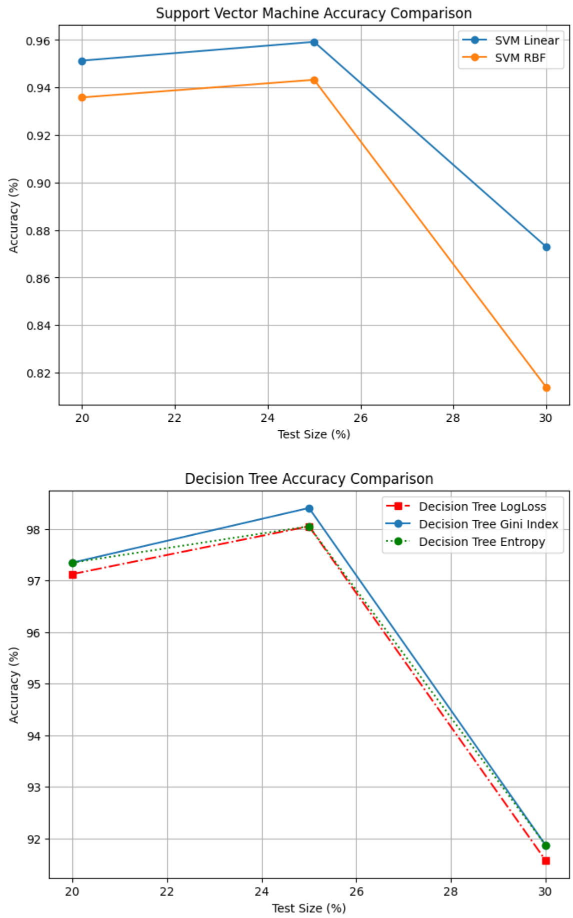


Fig. 9: Accuracy Comparison of Naive Bayes, SVM and Decision Tree

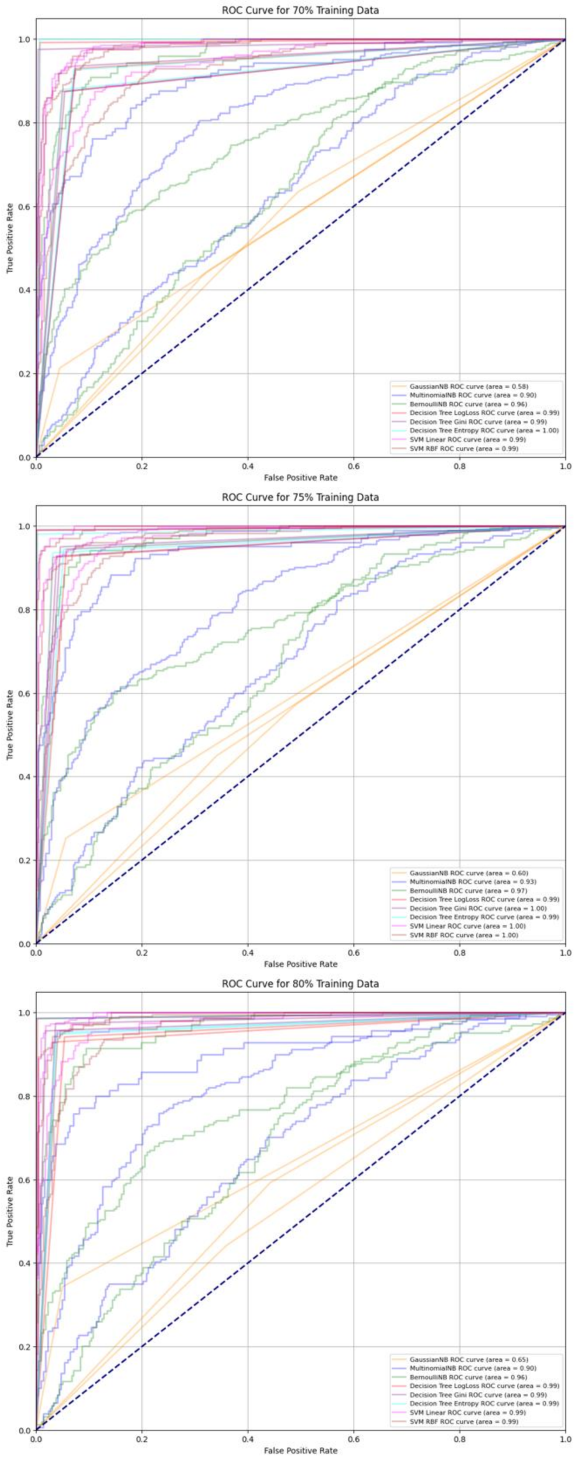


Fig. 10: ROC Curve for Naive Bayes, SVM and Decision Tree (Multiclass)

References

1. Shani, G., Gunawardana, A. (2011). Evaluating Recommendation Systems. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P. (eds) Recommender Systems Handbook. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-85820-3 8.
2. Raja Rajeswari, S., Naik, S., Srikant, S., Sai Prakash, M.K., Uday, P. (2019). Movie Recommendation System. In: Shetty, N., Patnaik, L., Nagaraj, H., Hamsavath, P., Nalini, N. (eds) Emerging Research in Computing, Information, Communication and Applications. Advances in Intelligent Systems and Computing, vol 882. Springer, Singapore. https://doi.org/10.1007/978-981-13-5953-8 28.
3. C.-S. M. Wu, D. Garg and U. Bhandary," Movie Recommendation System Using Collaborative Filtering," 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 2018, pp. 11-15, doi: 10.1109/ICSESS.2018.8663822.
4. N. Pradeep; K. K. Rao Mangalore; B. Rajpal; N. Prasad; R. Shastri. "Content based movie recommendation system". International Journal of Research in Industrial Engineering, 9, 4, 2020, 337-348. doi: 10.22105/riej.2020.259302.1156
5. Reddy, S., Nalluri, S., Kunisetti, S., Ashok, S., Venkatesh, B. (2019). Content Based Movie Recommendation System Using Genre Correlation. In: Satapathy, S., Bhateja, V., Das, S. (eds) Smart Intelligent Computing and Applications. Smart Innovation, Systems and Technologies, vol 105. Springer, Singapore. https://doi.org/10.1007/978-981-13-1927-342
6. Goyani, Mahesh; Chaurasiya, Neha. A Review of Movie Recommendation System: Limitations, Survey and Challenges. ELCVIA. Electronic letters on computer vision and image analysis, Vol. 19 N´ um. 3 (2020), p. 18-37. DOI 10.5565/rev/elcvia.1232 https://ddd.uab.cat/record/232276
7. S. Kumar, K. De and P. P. Roy," Movie Recommendation System Using Sentiment Analysis from Microblogging Data," in IEEE Transactions on Computational Social Systems, vol. 7, no. 4, pp. 915-923, Aug. 2020, doi: 10.1109/TCSS.2020.2993585.
8. T. Zhou, L. Chen and J. Shen," Movie Recommendation System Employing the User-Based CF in Cloud Computing," 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), Guangzhou, China, 2017, pp. 46-50, doi: 10.1109/CSE-EUC.2017.194
9. Robbins, Stephen, Timothy A. Judge, Bruce Millett, and Maree Boyle. Organizational behavior. Pearson Higher Education AU, 2013.
10. Arrow, Holly, Joseph E. McGrath, B. M. Staw, and L. L. Cummings. "Membership dynamics in groups at work: A theoretical framework." Research in organizational behavior 17 (1995): 373-373.
11. Watson, David, Lee Anna Clark, and Auke Tellegen." Development and validation of brief measures of positive and negative affect: the PANAS scales." Journal of personality and social psychology 54, no. 6 (1988): 1063.
12. Raghunathan, Rajagopal, and Michel Tuan Pham. "All negative moods are not equal: Motivational influences of anxiety and sadness on decision making." Organizational behavior and human decision processes 79, no. 1 (1999): 56-77.
13. https://www.kaggle.com/datasets/tmdb/tmdb-movie-metadata
14. Merz, Erin L., et al." Psychometric properties of Positive and Negative Affect Schedule (PANAS) original and short forms in an African American community sample." Journal of affective disorders 151.3 (2013): 942-949.
15. Magyar-Moe, Jeana L. Therapist's guide to positive psychological interventions. Academic press, 2009
16. Wen, H., Ding, G., Liu, C., Wang, J. Matrix Factorization Meets Cosine Sim ilarity: Addressing Sparsity Problem in Collaborative Filtering Recommender System. Web Technologies and Applications. APWeb 2014. Lecture Notes in Computer Science, vol 8709. Springer, Cham. https://doi.org/10.1007/978-3-319 11116-2 27
17. E. Bigdeli and Z. Bahmani," Comparing accuracy of cosine-based similarity and correlation-based similarity algorithms in tourism recommender systems," 2008 4th IEEE International Conference on Management of Innovation and Technology, Bangkok, Thailand, 2008, pp. 469-474, doi: 10.1109/ICMIT.2008.4654410.

18. Saunders, Craig, Mark O. Stitson, Jason Weston, Leon Bottou, and A. Smola. “Support vector machine-reference manual.” (1998).
19. Schoellkopf, Bernhard, and Alex Smola.” Support vector machines.” Encyclopedia of Biostatistics (1998): 978-0471975762.
20. Tu, P-L., and J-Y. Chung.” A new decision-tree classification algorithm for machine learning.” TAI’92-proceedings fourth international conference on tools with artificial intelligence. IEEE Computer Society, 1992.
21. Patel, Harsh H., and Purvi Prajapati.” Study and analysis of decision tree-based classification algorithms.” International Journal of Computer Sciences and Engineering 6.10 (2018): 74-78.
22. Anyanwu, Matthew N., and Sajjan G. Shiva.” Comparative analysis of serial decision tree classification algorithms.” International Journal of Computer Science and Security 3.3 (2009): 230-240.
23. Tangirala, Suryakanthi.” Evaluating the impact of GINI index and information gain on classification using decision tree classifier algorithm.” International Journal of Advanced Computer Science and Applications 11.2 (2020): 612-619.
24. Yang, Hang, and Simon Fong.” Improving the accuracy of incremental decision tree learning algorithm via loss function.” 2013 IEEE 16th International Conference on Computational Science and Engineering. IEEE, 2013.
25. Reddy, E. M. K., Gurrala, A., Hasitha, V. B., & Kumar, K. V. R. (2022). Intro duction to Naive Bayes and a review on its subtypes with applications. Bayesian reasoning and gaussian processes for machine learning applications, 1-14.
26. Kamel, Hajer, Dhahir Abdulah, and Jamal M. Al-Tuwaijari. “Cancer classification using gaussian naive bayes algorithm.” 2019 international engineering conference (IEC). IEEE, 2019.
27. Qiang, Guo.” An effective algorithm for improving the performance of Naïve Bayes for text classification.” 2010 Second international conference on computer research and development. IEEE, 2010.
28. Ashari, Ahmad, Iman Paryudi, and A. Min Tjoa.” Performance comparison between Naïve Bayes, decision tree and k-nearest neighbor in searching alternative design in an energy simulation tool.” International Journal of Advanced Computer Science and Applications (IJACSA) 4.11 (2013).