iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VI **Month of publication:** June 2021

DOI: https://doi.org/10.22214/ijraset.2021.35308

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

Movie Review System using Sentiment Analysis and Social Networking Platforms (SNPs)

Ms. Dhiwar Pooja S[1], Prof. Nirmal S. S[2]
[1, 2]*M.E Computer Eigg. Dept, P.R.E.C. Loni, Ahmednagar, Maharashtra-413736., Savitribai Phule Pune University.*

Abstract: ***Today online social networking platforms SNPs have become an integral part of or life where we share a lot of information of all the things, we do in life from shopping to movie watching. With ever growing use of SNPs recommendation systems have emerged as a hot trend for applications in e-commerce and digital media. These recommendation systems are useful as well as misguiding. Today digital media use has increased tremendously with increase in internet speeds. But users do not get proper review of a movie and a user is lured to watch a substandard movie which he does not intended to do, thus costing a user time and money. So, there is a need of developing a movie review which will give correct reviews of a digital content like movies so he can only movies which he intends to do. So, we are studying various techniques authored by various authors and create a good movie review system of our own. The first technique we studied and intends to use is movie recommendation system using tweets. The second study is movie recommendation using similarity measures. The third study does find a public shamming using SNP. These techniques are useful and we propose to use some part of each in our new movie review framework by improving the techniques drawbacks. The new framework will be a combination of data from more than one SNP and using natural language processing and machine learning on the data. We are going to use two machine learning algorithms SVM and Naïve Bayes for this purpose. For natural language processing of SNP data, we are going to use OPEN-NLP. We intend to use SNPs such as Twitter and any other movie database like IMDB etc. for data on the movie. The movie will be classified in three classes bad, good and excellent. The results from each algorithm SVM and Naïve bayes will be analyzed for each SNP and try to give user a more accurate movie review by combining all the reviews together and classes accuracy and show overall prediction results with a rating. To get more accurate results for each movie we are going to create a dataset for each movie for demonstration and will not depend on a single combined dataset as keywords for each movie may be different. We are going to combine datasets for each movie from multiple SNPs.***

Keywords: SNP, Recommendation System, Movie Review, Machine Learning, OPEN-NLP, SVM, Naïve Bayes.

I. INTRODUCTION

Today internet and things on it has become an integral part of our life. With ever increasing use of internet, we face a problem of excessive information which can be good as well as misguiding. We will find information in abundance on SNPs like Twitter, Netflix, IMDB etc. which is impossible for a human to analyze by himself as it is in a large scale. Here recommendation and review systems come in to play where a user is recommended with digital material depending upon his previous choices. But these recommendation systems become un reliable and can unnecessarily over rate or under rate a movie thus costing a user time and mental stress to handle a bad movie by watching it. So, to avoid these two types of filters on SNP data can be used first is collaborative filtering and content-based filtering. If both used together and analyzed properly will create a new movie recommendation or review system. The process of both filters is demonstrated in the Fig.1.

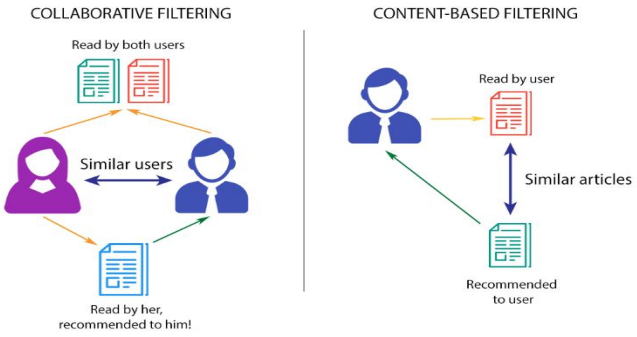


Fig.1. Existing process of SNP data filtering

©IJRASET: All Rights are Reserved

Thus, by studying process in Fig.1 a combined movie review system can be created where a user can be provided with a review in three categories like if the movie id bad, good or excellent. So, to combine these two filters we need to study various studies suggested by other authors in their research and implement and propose in our movie review and recommendation framework So, in other words the main objective of this paper is to:

A. Focus on giving a correct and more accurate review of a movie.
B. To help a user to avoid a user from traversing a large amount of movie data on various SNPs.
C. Study various sentiment analysis techniques and implement it for movie review system.
D. To get a more accurate result, a fusion frame work of machine learning and natural language processing techniques has to be used together.
E. More than one machine learning algorithms and more than one SNP has to be used together to get better results and analyzed them properly.
F. Evaluate and analyze the new movie review system framework and its strengths.

Thus, the rest of the paper is structured as follows:
1) Section II. explains literature survey which studies various techniques with their advantages and drawbacks.
2) Section III. explains the methodology i.e., mathematical model and algorithms to be used by the system.
3) Section IV. explains proposed system with block diagram or system architecture and working of the system.
4) Section V. shows the results and discussions and how it will be implemented.

II. LITERATURE SURVEY

This section describes the fundamentals of various techniques that can be used in designing a movie review system. It helps in understanding various ideas put forward by various technical papers published by various authors and how they put forth a more accurate and concrete techniques. Some of the ideas with technique and drawbacks are mentioned below:

In 2020 Kumar et al. [1] presented the paper focusses mainly on movie recommendation system using tweets It uses public dataset. This technique is quite good and covers all the things needed for a successful movie recommendation system. But main drawback of this system is that it concentrates public mixed dataset for all movies recommendation and does not create a dataset for each movie from multiple SNPs together to get more accurate results.

In 2020 Reddy et al. [2] presented the paper focusses mainly on movie recommendation system using collaborative filtering. It uses public movie-lens 100k dataset. This technique is quite good and covers all the things needed for a successful movie recommendation system. But main drawback of this system is that it concentrates public mixed dataset for all movies recommendation and does not create a dataset for each movie from multiple SNPs together to get more accurate results.

In 2019 Basak et al. [3] presented the paper focusses mainly on sentient analysis of tweets to detect public shamming of a person. This technique is quite good and can be used to do sentiment analysis of a movie tweets It uses twitter dataset. This technique is quite good and covers all the things needed for a successful movie recommendation system. But main drawback of this system is that it concentrates public mixed dataset for all movies recommendation and does not create a dataset for each movie from multiple SNPs together to get more accurate results.

In 2020 Tripathi et al. [4] presented the paper focusses mainly on analysing sentiments behind IMDB dataset. This technique is quite good and can be used to do sentiment analysis of a IMDB dataset to get a review. This technique is quite good and covers all the things needed for a successful movie recommendation system. But main drawback of this system is that it concentrates public mixed dataset for all movies recommendation and does not create a dataset for each movie from multiple SNPs together to get more accurate results.

In 2020 Moolthaisong et al. [5] presented the paper focusses mainly on emotion analysis and classification of movie reviews using TD-IDF and naïve bayes. This technique is quite good and can be used to do sentiment analysis of a movie reviews. This technique is quite good and covers all the things needed for a successful movie recommendation system. But main drawback of this system is that it concentrates public mixed dataset for all movies recommendation and does not create a dataset for each movie to get more accurate results. In 2020 Cahyanti et al. [6] presented the paper focusses mainly on sentiment analysis of movie reviews using feature extraction and SVM. It uses public IMDB dataset. This technique is quite good and covers all the things needed for a successful movie review system. But main drawback of this system is that it concentrates public mixed dataset for all movies recommendation and does not create a dataset for each movie from multiple SNPs together to get more accurate results.

©IJRASET: All Rights are Reserved

In 2020 Kapoor et al. [7] presented the paper focusses mainly on movie recommendation system using NLP. It uses public TMDB dataset. This technique is quite good and covers all the things needed for a successful movie review system. But main drawback of this system is that it concentrates public mixed dataset for all movies recommendation and does not create a dataset for each movie from multiple SNPs together to get more accurate results.

In 2020 Aishwarya et al. [8] presented the paper focusses mainly on review analysis of web series and movies using twitter data. This technique is quite good and can be used to do sentiment analysis of a review on a web series or a movie. It uses twitter dataset. This technique is quite good and covers all the things needed for a successful movie recommendation system. But main drawback of this system is that it concentrates public mixed dataset for all movies recommendation and does not create a dataset for each movie to get more accurate results.

III.METHODOLOGY

This section will study the mathematical conditions and algorithms to be used for designing an automatic and accurate movie review framework. These are explained as follows:

A. Mathematical Model

Our movie review technique can be explained in two sets with probability, success and failure conditions.

1) Data Preprocessing Module

Set (D)= {D0, D1, D2, D3, D4, D5}

D0∈D = Enter movie keyword.

D1∈D = Fetch movie reviews from various SNPs.

D2∈D = Remove stop words.

D3∈D = Perform lexicon analysis using OPEN-NLP.

D4∈D= Extract features.

D5∈D = View results.

2) Machine Learning Module

Set (M)= {M0, M1, M2, M3, D5}

M0∈M = Create training dataset using extracted features.

M1∈M = Pass a test movie reviews and create testing dataset.

M2∈M = Perform classification using training and testing dataset with SVM and Naïve Bayes algorithm.

M3∈M = Predict review in three classes good, bad and excellent.

D5∈M = View results.

So, by studying the sets we come to notice that elements are common in both modules and used in coordination in both sets so they be placed as

$$x \in D \cap M \text{ if } x \in D \text{ and } x \in M$$

Thus, the probability of intersection of elements in both modules can be given as

$$P(D \cap M) = P(D)+P(M)$$

So, intersection of common elements can be shown as

$$D \cap M = \{D5\}$$

The conditional probability of both modules using the same element can be shown as

$$P(D \mid M) = \frac{P(D \cap M)}{P(M)}$$

Thus, we conclude that our movie review framework's success and failure will depend upon the quality of SNP data we get, i.e., if the quality of SNP data is not good the SVM and Naïve Bayes classification results will not be proper and will not give accurate movie reviews, thus this is a case of failure, so our framework supports NP-Hard and not NP-Complete.

©IJRASET: All Rights are Reserved

B. Algorithms Used

Our movie review classification process can be explained using following algorithmic steps.

```
1:  procedure CLASSIFY
2:      urls[]=fetchURLDatabase
3:      urlsCount=urls.length()
4:      if urlsCount == 0 then
5:          status=addTrainUrl(newUrl)
6:          addTestUrl(newUrl)
7:          if status == True then
8:              Create deep learning model
9:              Train model
10:             saveStatus=Save Model()
11:             if saveStatus == True then
12:                 Load testing dataset
13:                 Apply model
14:                 classify URLs
15:             else
16:                 result ←Exit procedure
17:             end if
18:         else
19:             result ←Exit procedure
20:         end if
21:     else
22:         Repeat Step 2:
23:     end if
24:     View URL classification results.
25: end procedure
```

IV. PROPOSED SYSTEM

This This section is mainly divided in 4 main modules with other sub parts in them. The text that follows explains the modules with a block diagram or system architecture as shown in Fig.2. to illustrate them. The working of the framework is explained as:

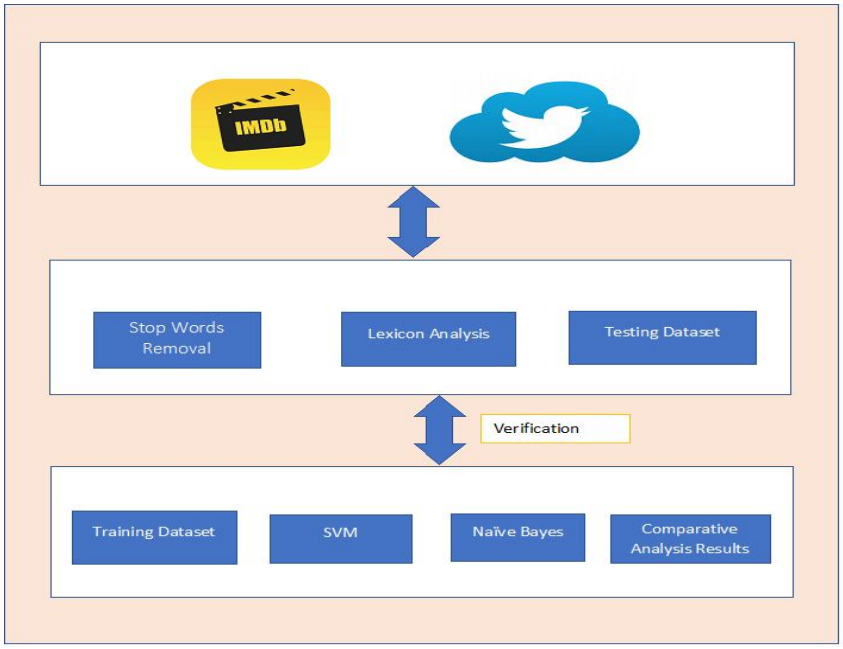


Fig.2. System Architecture Diagram.

A. Movie Review data collection using SNPs

This module will be proposed to accumulate data needed for a review system to work on. This data can be collected by using application programming interface (API) of various SNP. The SNPs we are proposing to use are both collaborative and content-based SNPs like twitter, IMDB etc. The data from this SNPs will be collected using a movie keyword and stored as txt files on local drives for further use. Some API needs to use Apache spark for getting real time reviews.

©IJRASET: All Rights are Reserved

B. Preprocess Movie Review Data

This module will also be proposed to pre-process data collected from various SNP using API. First the data collected will be cleaned by using unnecessary keywords. Then cleaned data will further be cleaned by using stop words which are not used in text mining. Thus, after stop words lexicon analysis is done on remaining data using OPEN-NLP natural language processing toll which is open source. Thus, after lexicon analysis features are extracted from review data i.e., whether a word is a verb, adjective, noun etc. This feature will be used with machine learning in next module.

C. Apply Machine Learning

This module will be proposed to predict a movie review using machine learning. We are using SVM and naïve bayes algorithms as they have high accuracy on small datasets. It will not need a very big dataset as needed by other machine learning algorithms. Our datasets will be small as we are creating datasets for movie specific tweets. It will use the features extracted using OPEN-NLP. First a training dataset is created which will be used to train machine learning algorithms. The training dataset will have three classes bad, good and excellent. Then a test movie reviews data will be passed to get a movie prediction from those movie reviews. The test movie reviews will follow the pre-processing steps and create a testing dataset from it with same three classes as training dataset. Then both training and testing datasets will be given to Naïve Bayes and SVM algorithms which will predict the class in which the test movie reviews reside i.e., whether the movie is good, bad or excellent. This process of prediction will be used for both types of digital contents i.e., collaborative and content based. Then final comparison of machine learning algorithms with sperate SNP will be studied to improve prediction.

D. Dataset

We are going to create a training dataset for each movie and will not use a common dataset used for all movies. The dataset will be collected using SNPs as twitter, IMDB etc. the dataset will use IMDB movie review dataset which has two classes positive and negative. Then we will add movie specific keywords which will be labelled in two classes positive and negative. Thus, this dataset will be movie specific and will not be applied to all movies. This dataset will be used to train machine learning algorithms SVM and Naive Bayes. We are going to create a dataset using a single SNP and then using multiple SNPs. Using single SNP and multiple SNP dataset will give an insight on which SNP is more efficient as discussed in next chapter.

V. RESULTS AND DISCUSSION

Thus, to explain the above proposed system we have created 1 application using java and python. We are using 2 SNPs for fetching movie reviews i.e., IMDB and Twitter.



Fig.3. Fetch movie reviews from IMDB.

In Fig.3 it shows how movie reviews are fetched from IMDB. The same process will be done for Twitter also.

©IJRASET: All Rights are Reserved

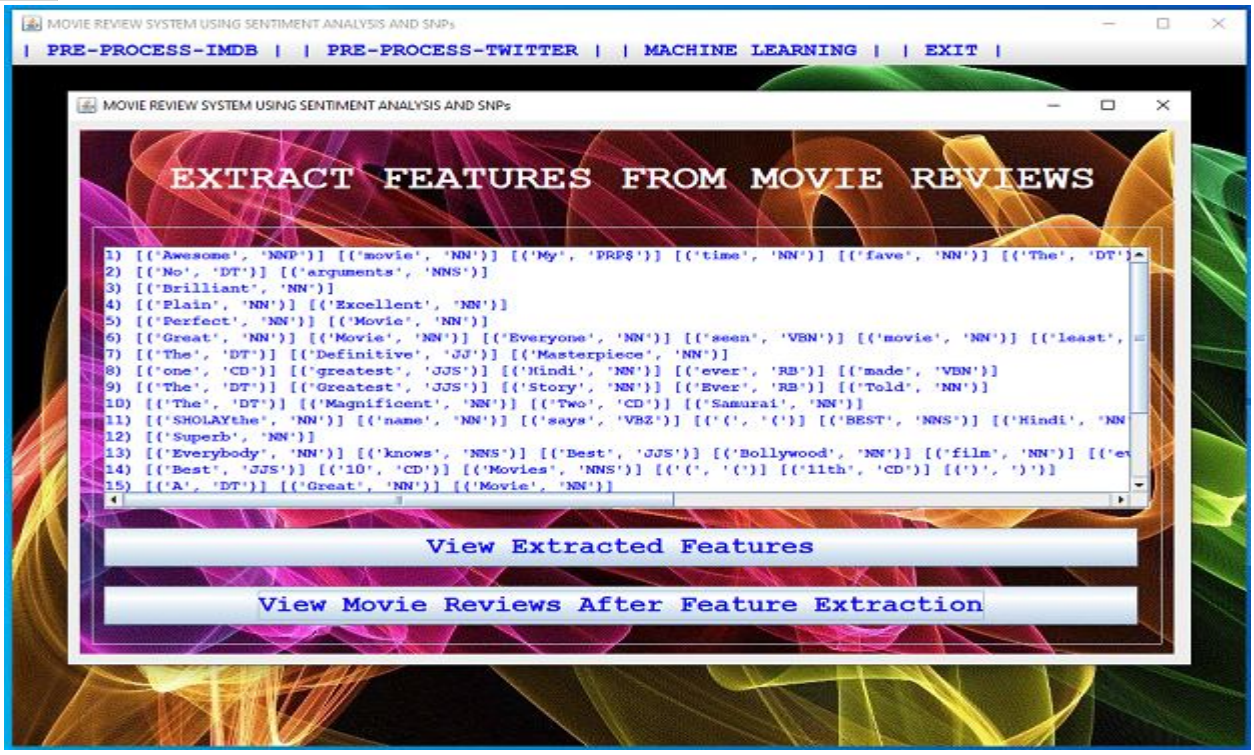


Fig.4. Sending skin diagnosis results.

In Fig.4 it shows how features are extracted from words using NLP tools. The screen is for IMDB but same process will be done for Twitter also.



Fig.5. Classification results of IMDB movie reviews.

In Fig.5 it shows how IMDB movie reviews are classified using SVM and Naïve Bayes algorithms.

©IJRASET: All Rights are Reserved



Sr-No.	Reviews	SVM	Naive Bayes
1)	delhi metro ddlj inspired rep...	Positive	Positive
2)	sorry played mainstream bol...	Negative	Negative
3)	idiot worth ever ddlj waste ti...	Negative	Negative
4)	delhi metro uses ddlj meme ...	Positive	Positive
5)	omfg this ddlj	Negative	Positive
6)	ddlj remains goat	Negative	Positive
7)	delhi metro uses ddlj meme ...	Positive	Positive
8)	nolan overrated srk best act...	Positive	Positive
9)	ddlj super duper swerve darl...	Negative	Negative
10)	still taken ddive ddlj think t...	Negative	Negative
11)	point film set foreign countr...	Negative	Negative
12)	want looking ddlj dubbing k...	Negative	Negative
13)	none watched movies u men...	Positive	Negative
14)	never watched ddlj	Negative	Negative
15)	ddlj dvd is delivered helpp	Positive	Positive
16)	omggg ddlj mbkd amp edser ...	Positive	Positive
17)	movies ka time hota hai shol...	Positive	Positive
18)	anjaani haan toh humne ka...	Positive	Positive
19)	tumhare maa baap ne ddlj d...	Positive	Positive
20)	ddlj always changed image i...	Positive	Positive

Fig.6. Classification results of Twitter movie reviews.

In Fig.6 it shows how Twitter movie reviews are classified using SVM and Naïve Bayes algorithms.

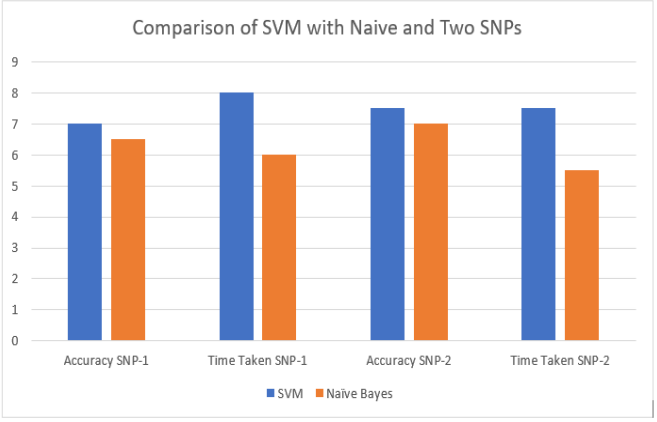


Fig.7. Comparison chart

Thus, the results in Fig.7. suggests that SVM has more accuracy in giving movie review than Naïve Bayes. But it also shows that the time to give result or process to complete movie review system is more for SVM than Naïve Bayes.

VI. CONCLUSION

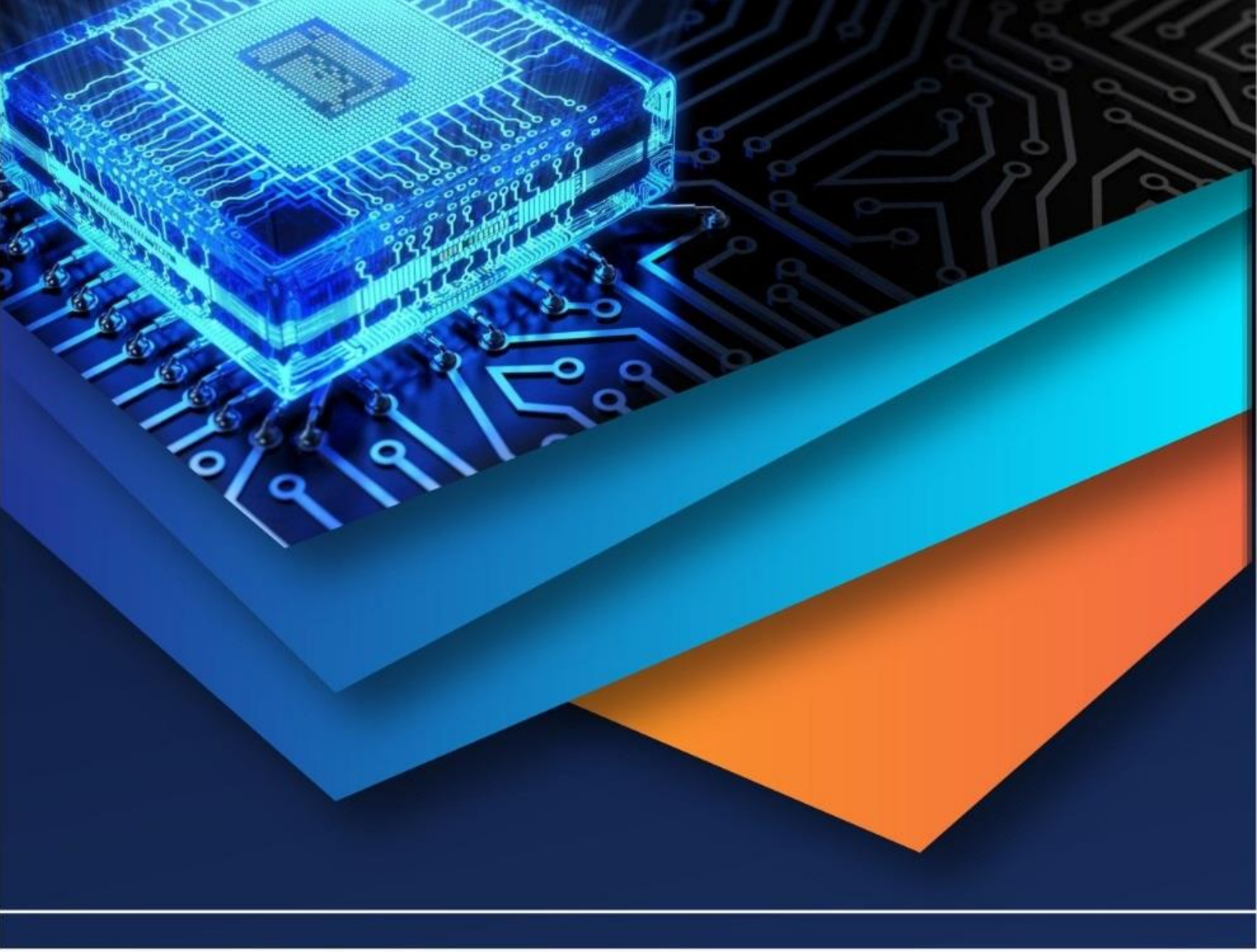
In this paper we have conclude to develop a novel approach of movie review recommendation system using data from multiple SNPs and using multiple machine learning algorithms. We have studied wide range of ideas [1][2][3][4][5][6][7][8] and taken in to consideration various techniques to create a movie recommendation system. To get the result more accurate we are going to use more than one machine learning algorithms i.e., Naïve bayes and SVM with feature set from OPEN-NLP natural language processing tool. We have studied and discussed why we are using both the algorithms in proposed system. We will try to combine and make use of Java and other programming languages needed as our programming language to implement the proposed framework. Thus, our project will decrease the efforts of a person to get correct movie review. Thus, we conclude that our proposed framework will be helpful in decreasing the trauma of a user by watching a bad movie by getting correct and accurate movie reviews.

©IJRASET: All Rights are Reserved

REFERENCES

[1] Sudhanshu Kumar, Kanjar De, and Partha Pratim Roy., - Movie Recommendation System Using Sentiment Analysis from Microblogging Data, IEEE-2020.
[2] Muppana Mahesh Reddy, R. Sujithra and Dr. B. Surendiran, - Analysis of Movie Recommendation Systems with and without considering the low rated movies, IEE-2020
[3] Rajesh Basak, Shamik Sural, Niloy Ganguly, and Soumya K. Ghosh, - Online Public Shaming on Twitter: Detection, Analysis, and Mitigation, IEEE-2019.
[4] Sandesh Tripathi, Ritu Mehrotra, Vidushi Bansal and Shweta Upadhyay, -Analyzing Sentiment using IMDb Dataset, in IEEE-2020.
[5] Kamoltep Moolthaisong and Wararat Songpan, -Emotion Analysis and Classification of Movie Reviews Using Data Mining, in IEEE-2020.
[6] Fitri Eka Cahyanti, diwijaya and Said Al Faraby, - On The Feature Extraction For Sentiment Analysis of Movie Reviews Based on SVM, in IEEE-2020.
[7] Nimish Kapoor, Saurav Vishal and Krishnaveni K S, - Movie Recommendation System Using NLP Tools, in IEEE-2020.
[8] Aishwarya, Parth Wadhwa and Prabhishek Singh, - A New Sentiment Analysis based Application for Analyzing Reviews ofWeb Series and Movies of Different Genres, in IEEE-2020.
[9] Ahmed Al Marouf, Md. Kamrul Hasan, and Hasan Mahmud., - Comparative Analysis of Feature Selection Algorithms for Computational Personality Prediction from Social MedIA, IEEE-2020.
[10] O. Araque, I. Corcuera-Platas, J. F. Sánchez-Rada, and C. A. Iglesias, - Enhancing deep learning sentiment analysis with ensemble techniques in social applications, Expert Syst. Appl., vol. 77, pp. 236–246, Jul. 2017.
[11] E. Aslanian, M. Radmanesh, and M. Jalili, - Hybrid recommender systems based on content feature relationship, IEEE Trans. Ind. Informat., early access, Nov. 21, 2016, doi: 10.1109/TII.2016.2631138.
[12] Yasen, Mais, and Sara Tedmori. -Movies Reviews sentiment analysis and classification, IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT), 2019.
[13] Mishne, Gilad and Natalie Glance, (2006), -Predicting Movie Sales from Blogger Sentiment, AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs.
[14] Abinash Tripathy, Ankit Agrawal, Santanu Kumar Rath, (2016), -Classification of sentiment reviews using n-gram machine learning approach, Expert Systems with Applications, Vol. 57, PP. 117-126.
[15] Vivek Narayanan, Ishan Arora, Arjun Bhatia, (2013), -Fast and Accurate Sentiment Classification Using an Enhanced Naive Bayes Model, Intelligent Data Engineering and Automated Learning – IDEAL, Springer, Vol. 8206.
[16] Hossin, Mohammad, and M. N. Sulaiman. -A review on evaluation metrics for data classification evaluations, International Journal of Data Mining & Knowledge Management Process 5.2 (2015): 1.
[17] Kerstin Denecke, (2008), -Using SentiWordNet for multilingual sentiment analysis, IEEE 24th International Conference on Data Engineering Workshop, Cancun, PP 507-512

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)