



## Star Ratings Prediction

Submitted by:

Harshitha K.S

# ACKNOWLEDGMENT

1. **“Automatic Movie Ratings Prediction using Machine Learning “ by Mladen Marovic, Marko Mihokovi C, Mladen Mik sa, Sini sa Pribil and Alan Tus University of Zagreb, Faculty of Electrical engineering and computing**

This paper reveals that Recommendation systems that model users and their interests are often used to improve various user services. Such systems are usually based on automatic prediction of user ratings of the items provided by the service. This paper presents an overview of some of the methods for automatic ratings prediction in the domain of movie ratings. The chosen methods are based on various approaches described in related papers. During the prediction process both the user and item features can be used. For the purpose of this paper, data was gathered from the publicly available movie database IMDb. The paper encompasses the implementation of the chosen methods and their evaluation using the gathered data. The results show an improvement in comparison to the chosen baseline methods.

2. **“Capturing the Stars: Predicting ratings for service and product reviews” by Narendra Gupta, Giuseppe DI Fabrizio and Patrick Haffne in the year 2017**

This paper tells that Bloggers, professional reviewers, and consumers continuously create opinion-rich web reviews about products and services, with the result that textual reviews are now abundant on the web and often convey a useful overall rating (number of stars). However, an overall rating cannot express the multiple or conflicting opinions that might be contained in the text, or explicitly rate the different aspects of the evaluated entity. This work addresses the task of automatically predicting ratings, for given aspects of a textual review, by assigning a numerical score to each evaluated aspect in the reviews. We handle this task as both a regression and a classification modeling problem and explore several combinations of syntactic and semantic features. Our results suggest that classification techniques perform better than ranking modelling when handling evaluative text.

### **3. “Movie Rating Prediction using Ensemble Learning Algorithm” by the Zahabiya Mhowwalal, A. Razia Sulthana, Sujala D.**

Over the last few decades, social media platforms have gained a lot of popularity. People of all ages, gender, and areas of life have their presence on at least one of the social platforms. The data that is generated on these platforms has been and is being used for better recommendations, marketing activities, forecasting, and predictions. Considering predictions, the movie industry worldwide produces a large number of movies per year. The success of these movies depends on various factors like budget, director, actor, etc. However, it has become a trend to predict the rating of the movie based on the data collected from social media related to the movie. This will help a number of businesses relying on the movie industry in making promotional and marketing decisions. In this report, the aim is to collect movie data from IMDB and its social media data from YouTube and Wikipedia and compare the performance of two machine learning algorithms – Random Forest and XGBoost – best known for their high accuracy with small datasets, but large feature set. The collection of data is done from multiple sources or APIs.

### **4. “The analysis and prediction of customer review rating using opinion mining” in the year 2017**

The customer review is important to improve service for company, which have both close opinion and open opinion. The open opinion means the comment as text which shows emotion and comment directly from customer. However, the company has many contents or group to evaluation themselves by rating and total rating for a type of services which there are many customer who needs to review. The problem is some customers given rating contrast with their comments. The other reviewers must read many comments and comprehensive the comments that are different from the rating. Therefore, this paper proposes the analysis and prediction rating from customer reviews who commented as open opinion using probability's classifier model. The classifier models are used case study of customer review's hotel in open comments for training data to classify comments as positive or negative called opinion mining. In addition, this classifier model has calculated probability that shows value of trend to give the rating using naive bayes techniques,

which gives correctly classifier to 94.37% compared with decision tree Techniques.

### **5.User-Personalized Review Rating Prediction Method Based on Review Text Content and User-Item Rating Matrix by Bingkun Wang , Bing Chen , Li Ma , and Gaiyun Zhou 1**

This paper reveals that With the explosive growth of product reviews, review rating prediction has become an important research topic which has a wide range of applications. The existing review rating prediction methods use a unified model to perform rating prediction on reviews published by different users, ignoring the differences of users within these reviews. Constructing a separate personalized model for each user to capture the user's personalized sentiment expression is an effective attempt to improve the performance of the review rating prediction. The user-personalized sentiment information can be obtained not only by the review text but also by the user-item rating matrix. Therefore, we propose a user-personalized review rating prediction method by integrating the review text and user-item rating matrix information. In our approach, each user has a personalized review rating prediction model, which is decomposed into two components, one part is based on review text and the other is based on user-item rating matrix. Through extensive experiments on Yelp and Douban datasets, we validate that our methods can significantly outperform the state-of-the-art methods

## INTRODUCTION

We have a client who has a website where people write different reviews for technical products. Now they are adding a new feature to their website i.e. The reviewer will have to add stars(rating) as well with the review. The rating is out 5 stars and it only has 5 options available 1 star, 2 stars, 3 stars, 4 stars, 5 stars. Now they want to predict ratings for the reviews which were written in the past and they don't have a rating. So, we have to build an application which can predict the rating by seeing the review.

With the explosive growth of product reviews, review rating prediction has become an important research topic which has a wide range of applications. The existing review rating prediction methods use a unified model to perform rating prediction on reviews published by different users, ignoring the differences of users within these reviews. Constructing a separate personalized model for each user to capture the user's personalized sentiment expression is an effective attempt to improve the performance of the review rating prediction. The user-personalized sentiment information can be obtained not only by the review text but also by the user-item rating matrix. Therefore, we propose a user-personalized review rating prediction method by integrating the review text and user-item rating matrix information. In our approach, each user has a personalized review rating prediction model, which is decomposed into two components, one part is based on review text and the other is based on user-item rating matrix.

## METHODOLOGY

Step1:- Firstly we have to scrape the data using different website. Here I am using the Amazon website to scrape the data regarding items like laptops, headphones, mobile etc

Step2:- Importing the required libraries:- here we are using the Numpy Library for data Analysis , Pandas for Scientific Computation, Selenium to drive the webdriver

Step 3:- Importing the text pre-processing package

Step 4:- Algorithm used:-

1. TFIDF[Term Frequency Inverse Document Frequency]:- it is a numeric statistic that is intended to reflect how important the model a word is to document in a collection of corpus. It is often used as weighing factor in searches of the information retrieval, text mining, user modelling. The TFIDF value increase proportionality to the number of words appears in the document and is the offset by the number of documents in the corpus that contain a word which helps to adjust the fact ha some word appear frequently in general. TF-IDF is a most popular weighting schemes today
2. Linear SVC:- The objective of a Linear SVC(support vector Classifier) is to fit to the data you provide, returning a best classifier is to fit the data you, provide, returning a best fit hyperplane that divides or categorizes , your data . From there, after getting the hyperplane, yo can then feed some features to your classifier to see what the “predicted” class is. This makes this specific algorithm rather suitable for our uses, though you can use this for many situations

Step5:- predicting the ratings using Classification report

## **CONCLUSION**

The accuracy obtained from the classification report is too low. Because due to the internet connect I can't able to scrape the more data. More the data we scrape from website, more the accuracy we get