International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 12, pp 132-136, December-2023

https://doi.org/10.47001/IRJIET/2023.712018

Predicto - Review Rate Predictor

[1]Sourav Chanda, [2]Abhishek Pandey, [3]Priyanka Mondal

[1]Department of Computer Science & Engineering, University of Calcutta, Kolkata, India

[2]Software Engineer Intern, Socielo Tech, Kolkata, India

[3]Department of Information Technology (Internet of Things), Maulana Abul Kalam Azad University of Technology, Kolkata, India

***Abstract* - This paper presents a new Chrome extension for predicting star ratings according to the customer's review. Predicto mainly deals with analyzing customer feedback to predict star ratings can provide valuable insights to both consumers and businesses. This research paper presents the development of a Chrome extension designed to predict star ratings based on customer reviews. Leveraging logistic regression as the predictive model, the extension employs natural language processing (NLP) techniques to extract pertinent features from textual feedback. The proposed Chrome extension capitalizes on web scraping capabilities to gather and preprocess customer reviews from diverse online sources. This research contributes to the field of sentiment analysis, customer feedback evaluation, and web scraping by presenting a practical implementation in the form of a user-friendly Chrome extension. The extension's utilization of logistic regression enhances its prediction capabilities and offers a valuable tool for enhancing the online shopping experience and review analysis.**

Keywords: Predicto, Sentiment Analysis, Logistic Regression, Tfidf Vectorizer, Chrome Extension.

I. INTRODUCTION

In the digital age, the proliferation of online commerce and services has revolutionized consumer behavior, turning customer reviews into pivotal resources that guide purchasing decisions. An individual's choice to engage with a product, service, or platform is frequently influenced by the collective sentiment expressed in reviews. As a result, accurately predicting star ratings based on textual customer feedback has emerged as a valuable pursuit, offering insights to both prospective buyers and businesses aiming to enhance customer experience. This research paper introduces a novel approach to predicting star ratings using a specialized Chrome extension. Leveraging the power of logistic regression as the predictive model, the extension combines natural language processing (NLP) techniques with web scraping capabilities to extract meaningful insights from customer reviews. The envisioned extension seeks to bridge the gap between consumer intent and business understanding, offering a tool that assists users in making informed decisions while providing businesses with valuable sentiment analysis.

This paper is to elucidate the development and implementation of the Chrome extension for star rating prediction. The extension's foundation rests on the utilization of logistic regression, a widely employed statistical technique in classification tasks. By capturing the nuanced relationships between textual features and star ratings, the extension aims to generate accurate predictions that empower users to gauge the quality of products or services.

II. METHODOLOGY

The methodology for logistic regression is a statistical method commonly used for predicting the probability of an event occurring, and it's particularly handy when the dependent variable is binary (having only two possible outcomes).

1. Define the Problem:

Identify the problem you want to solve and determine if logistic regression is the appropriate method.

2. Collect Data:

Gather data relevant to your problem. Ensure that your dependent variable is binary (0/1 or True/False), and collect independent variables that may influence the outcome.

3. Data Cleaning:

Handle missing data and outliers. Ensure your data is in a format suitable for analysis.

4. Explore Data:

Understand your data by using descriptive statistics and visualization techniques.

5. Split Data:

Divide your dataset into training and testing sets. The training set is used to train the model, and the testing set is used to evaluate its performance.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 12, pp 132-136, December-2023

https://doi.org/10.47001/IRJIET/2023.712018

6. Feature Scaling:

Standardize or normalize your independent variables to ensure they have a similar scale. This helps the algorithm converge faster.

7. Build the Logistic Regression Model:

Choose your independent variables and fit the logistic regression model to the training data. The logistic function is often used to model the probability.

8. Interpret Coefficients:

Examine the coefficients of the model to understand the impact of each independent variable on the log-odds of the dependent variable.

9. Make Predictions:

Use the trained model to predict outcomes on the testing set or new data.

10. Evaluate the Model:

Assess the model's performance using metrics like accuracy, precision, recall, F1 score, and ROC-AUC.

11. Iterate and Fine-Tune:

If the model performance is not satisfactory, consider refining it by adjusting hyperparameters, adding or removing features, or exploring more advanced techniques.

12. Deploy the Model:

Once satisfied with the model's performance, deploy it for making predictions on new, unseen data.

Logistic Regression assumes a linear relationship between the independent variables and the log-odds of the dependent variable. If your data doesn't meet this assumption, you might need to explore more complex models.

II. Logistic Regression for Multiclass Classification

Logistic regression can be extended to handle multiclass classification problems through techniques like one-vs-all (OvA) or one-vs-one (OvO) approaches. Here's how each of these methods works:

1. One-vs-All (OvA):

- For k classes, train k separate binary logistic regression classifiers.
- Each classifier is trained to distinguish one class from the rest (positive class vs. all other classes).
- During prediction, the class with the highest predicted probability is chosen.

2. One-vs-One (OvO):

- For k classes, train $\frac{k.(k-1)}{2}$ binary logistic regression classifiers.
- Each classifier is trained to distinguish between two specific classes.
- During prediction, each classifier votes for a class, and the class with the most votes is chosen.

3. Softmax Function:

- In addition to OvA and OvO, another common approach for multiclass logistic regression is to use the softmax function. The softmax function generalizes the sigmoid function to handle multiple classes. The softmax function calculates the probability of each class and ensures thatthe sum of the probabilities across all classes is 1.
- $P(y = i \mid x) = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}}$

Where z_i is the linear combination of input features for class i, and k is the number of classes.

4. Training:

- The model is trained to optimize the likelihood of the observed data using a multiclass extension of the cross-entropy loss.

5. Decision Rule:

- During prediction, the class with the highest probability is selected as the final predicted class.

Logistic regression can be adapted for multiclass classification by either using multiple binary classifiers (OvA or OvO) or by directly extending the logistic function to handle multiple classes through the softmax function. OvA is simpler and more commonly used, while OvO is less common due to the increased number of classifiers. The choice between these methods often depends on the size of the dataset and computational considerations.

III. LITERATURE STUDY

Several research papers have reconnoitered Sentiment Analysis and Prediction of Online Reviews techniques. This section summarizes some of them.

Rating prediction is crucial for recommendation models, but existing methods based on historical review data often

ISSN (online): 2581-3048
Volume 7, Issue 12, pp 132-136, December-2023
https://doi.org/10.47001/IRJIET/2023.712018

struggle with unbalanced distributions and limited robustness due to insufficient data. To overcome these challenges, the versatile CID (causal inference debiasing) rating prediction model is proposed. CID accounts for the causal connection between review data and user ratings, reducing context bias and enhancing robustness. Experiments on four datasets confirm CID's ability to improve prediction accuracy and mitigate bias, making it a valuable asset in recommendation models [1].

The paper aims to enhance business analytics rating predictions by integrating performance features from customer feedback. The proposed dynamic model combines Latent Dirichlet Allocation topic membership values with sentiment analysis in an XGBoost model, achieving an 86% prediction accuracy. The results highlight practical implications for sentiment analysis in customer reviews, particularly in tasks with domain-specific adjustments. Shapley Additive Values are employed to evaluate the additive predictability of topic membership and sentiment-based methods using food delivery service reviews [2].

The paper delves into the role of recommendation systems in modern e-commerce, highlighting the rising importance of contextual information for personalized suggestions. Departing from traditional methods reliant on user reviews and rating history, the research introduces a machine learning-driven Review-Based Rating Prediction model, implemented on a Yelp dataset, aiming to elevate the relevance and accuracy of recommendations [3].

The paper explores the impact of online reviews on consumer decisions from platforms like Amazon and Yelp. Addressing the difficulty of predicting ratings without star-level information, the study introduces a deep learning framework using bidirectional gated recurrent unit (Bi-GRU) models. This framework first predicts polarity and then utilizes it to predict review ratings from the textual content. Experiments on real-world datasets show significant enhancements in precision, recall, F1-score, and root mean square error compared to baseline approaches, validating the effectiveness of the proposed approach [4].

The study aims to leverage machine and deep learning models for predicting sentiment and ratings from tourist reviews in the context of the growing significance of the tourism industry. Utilizing models such as Naïve Bayes, Support Vector Machines (SVM), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and Bidirectional LSTM, the research classifies reviews into positive, negative, or neutral sentiments and assigns one to five-star ratings. Trained on data from TripAdvisor, the results demonstrate high accuracy, with the optimal model being a deep learning approach based on Bidirectional LSTM. The findings emphasize the efficiency and accuracy of deep learning models over traditional machine learning algorithms, offering practical implications for forecasting tourist arrivals, understanding tourist profiles, enhancing customer experience, and refining marketing strategies. The study contributes originality by developing and comparing various machine learning models on a substantial TripAdvisor hotel reviews dataset [5].

This paper addresses the prevalence of raw text data in the digital age, emphasizing its value for insights through natural language processing. Focused on Amazon.com customer reviews, the study employs a proposed framework to preprocess and analyze text, exploring how reviews contribute to star ratings. Using text-derived features in a multi-class classification task, the Logistic Regression Classifier emerges as the most effective in predicting ratings. Notably, factors such as review polarity and length are identified as pivotal in influencing rating outcomes [6].

The increasing prominence of online user reviews has spurred research on predicting review ratings in natural language processing. Conventional methods frequently neglect vital factors like user sentiments and product evaluations. A groundbreaking approach integrates user and product context into review texts, utilizing a three-part model encompassing global, user-specific, and product-specific review rating predictions. Experimental findings across four datasets underscore the considerable superiority of this approach over state-of-the-art baselines in review rating prediction [7].

Online reviews, reflecting users' firsthand experiences, are crucial for product information. E-commerce platforms facilitate consumer opinions through reviews. Sentiment analysis categorizes emotions in review text (positive, negative, or neutral) for information extraction, aiming to improve business operations. The study addresses the challenge of conflicting comments and ratings by introducing a classifier model that categorizes reviews based on the presence of ratings. It further predicts sentiments for unrated reviews using diverse classifiers [8].

Recommendation systems play a vital role in modern e-commerce applications, influencing targeted advertising, personalized marketing, and information retrieval. Recognizing the growing importance of contextual information, this study introduces a review-based rating approach that extracts contextual insights from user reviews. Unlike traditional systems relying on rating history, this approach utilizes Natural Language Processing (NLP) and information retrieval methods to compute a utility function for items based on textual reviews. Context inference is modeled

ISSN (online): 2581-3048
Volume 7, Issue 12, pp 132-136, December-2023
https://doi.org/10.47001/IRJIET/2023.712018

by assessing similarity between users' and items' review histories. Evaluations, using movie reviews as an example application, indicate that this system produces more accurate rating predictions compared to standard methods [9].

Today's information abundance, primarily in the form of opinions, demands effective mining strategies. Given the vast daily exchange on the Internet, sentiment analysis is crucial for deciphering this raw data. Opinion mining, especially through social media platforms like Instagram, Facebook, and Twitter, has become prominent. The hidden insights from this wealth of information find diverse applications in marketing, political polls, product reviews, market forecasting, and identifying detractors and promoters. The introduced sentiment rating system employs a comprehensive approach, encompassing search resolution, tokenization, classification, content identification, and probability extraction using a naive bayes classifier. It integrates sentiment dictionaries to enhance accuracy in determining the polarity of opinions [10].

IV. PROPOSED METHOD

Step 1: Collection of appropriate datasets.

Step 2: Encoding the text data.

Step 3: Vectorizing the data using TfidfVectorizer.

Step 4: Splitting the data into train and test dataset.

Step 5: Training the Logistic Regression model with the training dataset.

Step 6: Calculating the time taken to train the models.

Step 7: Testing the models with test datasets and generating the accuracy scores.

V. RESULTS AND DISCUSSIONS

By using of the estimators from logistic regression for the classification, we have got the result of the accuracy from the different dataset and also tested with our own data to the model in the form of text and got an accurate model.

Product	Source	#Reviews	Accuracy (%)
Watches	Amazon	960026	72
Furniture	Amazon	791514	72
Software	Amazon	341215	67
Digital Video Games	Amazon	144715	73
Digital Software	Amazon	101826	69
Personal Care Appliances	Amazon	85919	70

As you can see in the above diagram, the model has an accuracy of the validation. We also developed a chrome extension for our results.

VI. CONCLUSION

From the results above, we can infer that for our problem statement, Logistic Regression with grid search Model is best with the accuracy. In this project, we presented a natural language processing technique to do the analysis of the product reviews efficiently and compare their performances using different metrics and predict whether given review is positive or negative. Finding the polarity of reviews can be useful in a variety of situations. Intelligent systems can be built to give users with comprehensive reviews of products, services, and other items without requiring the user to read individual evaluations; instead, the user can make decisions based on the intelligent systems' results. The aim is to apply natural language processing techniques in the analysis of product reviews it is the same as in the other industriesin order to improve the business and optimize its marketing strategies, to reduce work. Also chrome extension help to show your accurate output result and predicting star ratings according to the customer's review.

REFERENCES

[1] Y. W, C. W, Jiangang Nan, "Rating Prediction Model Based on Causal Inference Debiasing Method in Recommendation," *Chinese Journal of Electronics*, vol. 32, no. 4, pp. 932 - 940, 2023.

[2] N. K, D. Z, K. S, Nan Yang, "Incorporating topic membership in review rating prediction from unstructured data: a gradient boosting approach," *Annals of Operations Research*, 2023.

[3] A. C, D. G, A. K, Maaz Bin Shahid, "Review Based Rating Prediction using Machine Learning Techniques," *nternational Conference on System Modeling & Advancement in Research Trends (SMART)*, pp. 118-122, 2022.

[4] A. S. G, Basem. H. A. Ahmed, "Review Rating Prediction Framework Using Deep Learning," *Journal of Ambient Intelligence and Humanized Computing*, 2022.

[5] M. B. B, Karlo Puh, "Predicting sentiment and rating of tourist reviews using machine learning," *Emerald Insight*, vol. 6, no. 3, pp. 1188-1204, 2022.

[6] T. B, Ankit Taparia, "Sentiment Analysis: Predicting Product Reviews' Ratings using Online Customer Reviews," *Social Science Research Network*, 2020.

[7] S. X, Y. H, X. L, Bingkun Wang, "Review Rating

ISSN (online): 2581-3048
Volume 7, Issue 12, pp 132-136, December-2023
https://doi.org/10.47001/IRJIET/2023.712018

Prediction Based on User Context and Product Context," *Multidisciplinary Digital Publishing Institute*, vol. 8, no. 10, pp. 01-13, 2018.

[8] L. I. S, Sasikala P, "Sentiment Analysis and Prediction of Online Reviews with Empty Ratings," *Research India Publications*, 2018.

[9] V. S. K, G. A, N. V, M. Pravallika Reddy, "Review-based Rating Prediction," *INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS*, vol. 6, no. 1, pp. 843-846, 2018.

[10] S. U, T. N, S. G, A. NithyaKalyani, "Rating prediction using textual reviews," *Journal of Physics: Conference Series*, 2018.

Citation of this Article:

Sourav Chanda, Abhishek Pandey, Priyanka Mondal, "Predicto - Review Rate Predictor" Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 7, Issue 12, pp 132-136, December 2023. Article DOI https://doi.org/10.47001/IRJIET/2023.712018

****

