

# Fake Reviews Identifying Machine Learning Technology on Online Food Porters

L. Satya Naga Veni<sup>1</sup>; M. Jahnvi Padmapriya<sup>2</sup>; P. Srinu Vasarao<sup>3</sup>  
Swarnandhra College of Engineering & Technology

**Abstract:-** Wherever we go we will definitely look for food without any time limit. The food is an all-time emotion for everyone. Mainly if anyone visits the new place, he/she is excited to look for good food within less budget, with good management, maintenance and neatness.

Zomato is a digital platform that serves as a one-stop destination is ordering food from various restaurants. It users with an extensive data set of restaurants, menus, user reviews, and ratings, allowing them to browse. Zomato is such an interesting app which provides users with more comforts like how many users are approaching to the Zomato, ratings, reviews, restaurants data sets, fast hand delivery, to easily track the order by push up notifications or delivery boy phone number. Mainly Zomato has many facilities whatever food we need we can easily order within seconds. The Zomato has many different food items in different locations within the city/town but some times the restaurants are closed early. Mostly everyone looks for the best biriyani within their area because the biriyani is not a food but its an heartfelt emotion who really look for it.

Can use supervised machine learning algorithms as logistic regression, naive Bayes or support vector machines, classify the reviews as fake or genuine based on this feature. Every order served by the delivery boy has the charge for delivery. In the present modernized world, fame of food applications is expanding because of usefulness, view, book or request effectively by not very many snaps on telephone. Online reviews have become very easy to take users insight to get the good food in this app. has made it is solid and is currently present in 22countries with more than 1,000,000 eateries around the world and gets 1.25 million orders day to day.

**Keywords:-** Fake Review, Linear Regression, Logistic Regression, Naive Bayes, Machine Learning Algorithm.

## I. INTRODUCTION

Zomato is a notable Indian world wide Restaurant collection. and food organization spread out by Pankaj Chaddah and Deepinder Goyal in 2008. Zomato helps in giving menus, data as well as gives audit of theuser with respect to eateries and choices of food conveyance through accomplice cafés arranged in different urban areas. As per the report of 2019, Zomato is available in around 24 countries and in over in excess of 10,000 urban communities.

Zomato is one of the most far reaching and Easy to use applications where individuals can look for neighboring eateries and bistros, request food on the web and get it conveyed to their doorstep right away. Gross request esteem (GOV) — the complete worth, everything being equal, a rate increased 13.9% in its food conveyance business, as depended on August 2023. We have been really buckling down on further developing our evaluations calculation further Throughout the long term we have completed a few cleanups and changes to continue to remove counterfeit surveys, rates of sales and troublemakers from our framework. We do this to guarantee that we bring you 100 percent certifiable and dependable audits when you specially make your choice of which eatery from/eat at.

Zomato for the most part has two unique kinds of appraisals accessible, which are the conveyance evaluations and the café evaluations. The previous depends on how rapidly the food was conveyed, how well the conveyance specialist acted, and assuming they were following the security conventions. The more significant sort of rating which is pertinent to the eatery proprietors is the café evaluations. Clients can eat at the café somewhere in the range of 1 to 5 stars.

## II. LITERATURE SURVEY

- **Alexandridis, G., Varlamis, I., Korovesis, K., Caridakis, G., Tsantilas, P. (2021):** the volume of content made on interpersonal organizations builds, an ever increasing number of conclusions and Sentiments show that individuals share feelings diversely on issues. For this situation, profound and close to home examination is vital not exclusively to decide the size of the instant messages, (messages, tweets, and so on.). In spite of advances in message mining calculations, profound learning strategies, and message portrayal models, the consequences of these undertakings are generally excellent for a couple of undeniable level dialects.
- **Hemmatian, F., Sohrabi, M.K. (2019):** Thought mining is the most common way of separating human considerations and experiences from unstructured texts, which has turned into a valuable, intriguing and testing issue with the rise of online ads and numerous client surveys.
- **Huseyin, M.F. (2019):** Claims that internet based spam messages are offers of phony items and administrations. They are many times utilized as a contrivance to deceive perusers. It is significant yet challenging to Acknowledge this message. In this review, I attempt to take care of this

issue by utilizing various information mining. I investigate the advantages and disadvantages of information mining methods in distinguishing counterfeit remarks.

- **Elmurngi, E., Gherbi, A. (2018):** Opinion Examination (SA), applied as of late, has transformed into a text investigation of premium because of its great monetary outcomes One of the alluring subjects is the means by which to dissect positive and negative remarks from pundits. Moreover, surveys in regards to clients' requirements can be isolated into positive or negative audits, and clients can utilize these surveys to pick items. Check

### III. METHODOLOGY

This study aims to create a regression algorithm model using the Google Collaboratory tool and Jupyter Notebook. The steps of the modeling process are shown in Figure. The first step is to obtain the datasets from the Kaggle website. the Exploratory Data Analysis (EDA) is performed. To apply machine learning with the linear regression algorithms for some sample data is required. The table below shows some data about different Zomato reviews and ratings identifying based on their specifications. The data is sourced from Kaggle.com

1	rating	review
2	0	5 nice
3	1	5 best biryani , so supportive staff of outlet , personalize my order on call as I say. full Paisa vasool
4	2	4 delivery boy was very decent and supportive.δŸ'ŒδŸ'
5	3	1 worst biryani i have tasted in my life, half of the biryani is in dustbin.
6	4	5 all food is good and tasty . will order again and lots more to explore in bawarchi's menu.
7	5	5 shandar zabardast zindabad .. good going bawarchi keep it up
8	6	5 overall good experience . will order again
9	7	1 good but cold of food
10	8	5 tasty food . homestyle tasty prepration.
11	9	5 loved it . 5/5 .
12	10	4 full paisa vasool , taste me best Bawarchi aur Everest . . .
13	11	5 good taste , authentic punjabi style matar paneer . feels good to have such food in surat .
14	12	5 very tasty authentic food ..
15	13	5 superb tasty asli dilli wala flavour
16	14	5 yummiest manchurian fried rice i have ever had in city . authentic chinese taste. must order once
17	15	5 veey tasty. . order again nd again
18	16	1 i have ordered Manchurian fried rice but instead I got only simple fried rice. Feeling like I am cheated by bawarchi restaurant.
19	17	1 very bad not properly cooked

Fig 1: Zomato Reviews and Ratings Data

This data can cleaned and exported using machine learning techniques and that mostly stable for linear regression algorithms. The data flow can be represented as follows.

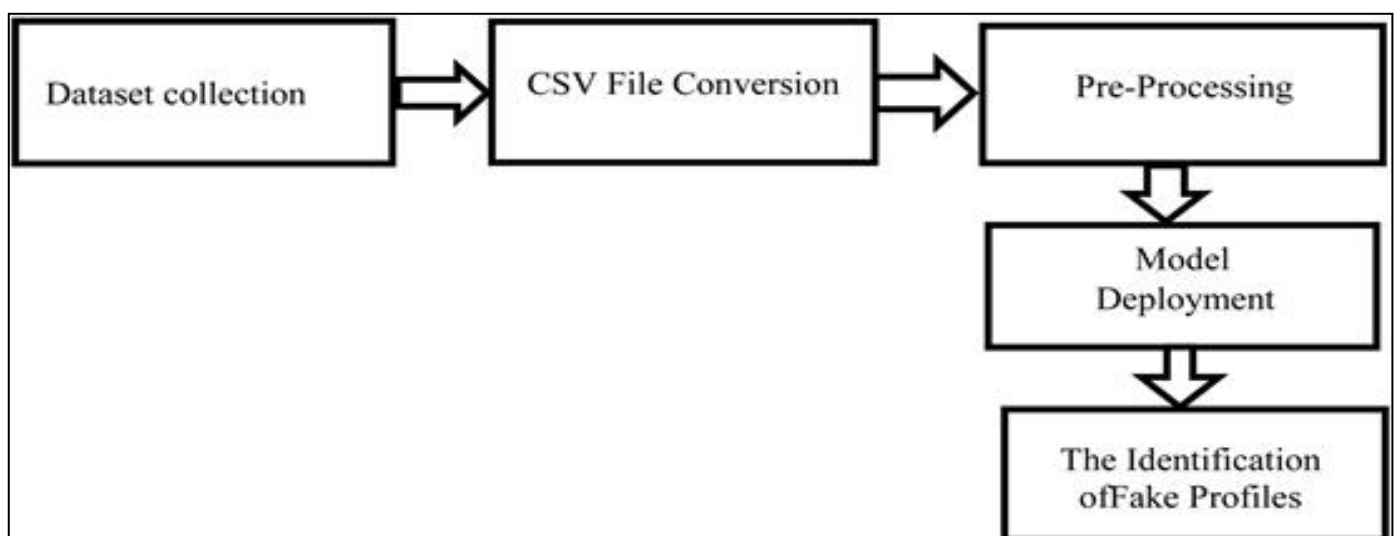


Fig 2: Process of Sentimental Analysis

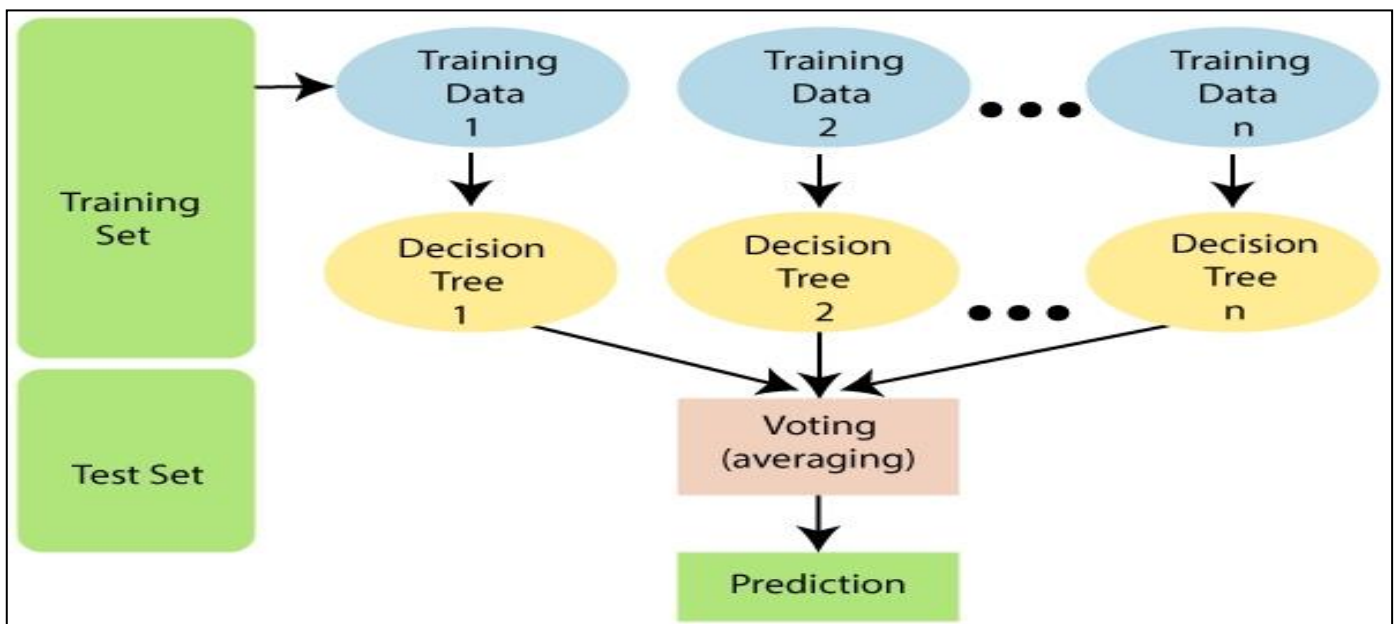


Fig 3: Decision Making Tree

A. Algorithm Used

Linear regression is also a machine learning algorithm, specifically a supervised machine learning algorithm that can best learn from datasets and contexts. It can be used to predictions on data.

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{[n(\sum x^2) - (\sum x)^2]}$$

$$b = \frac{[n(\sum xy) - (\sum x)(\sum y)]}{[n(\sum x^2) - (\sum x)^2]}$$

B. Linear Regression Algorithm Method

```

In [1]: import pandas as pd
import numpy as np
df=pd.read_csv("reviews.csv")
print(df)

   Unnamed: 0  rating  review
0           0       5      nice
1           1       5  best biryani , so supportive staff of outlet ,...
2           2       4  delivery boy was very decent and supportive. 🍽️ 👍
3           3       1  worst biryani i have tasted in my life, half o...
4           4       5  all food is good and tasty . will order again ...
...         ...     ...
5474        5474     5      complain
5475        5475     5  it took 1 hour to assign valvet and thn prepar...
5476        5476     5  took for an hour to prepare 3 khawsa, which in...
5477        5477     1  very very late, litttrally did time pass and it...
5478        5478     1  Taste was stale and they give only 5 pieces in...

[5479 rows x 3 columns]
  
```

➤ # Import Necessary Libraries

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix, roc_curve, auc
# Assuming your dataset is in a DataFrame named 'df'
# If it's in a CSV file, you can read it using: df = pd.read_csv('your_dataset.csv')
# Drop the 'Unnamed: 0' column if it's an index column
df = df.drop('Unnamed: 0', axis=1, errors='ignore')
# Data Preprocessing
# Assuming 'review' is your feature and 'rating' is your target variable
X = df['review'].astype(str) # Ensure 'review' column is treated as strings
y = df['rating']
# Import necessary libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix, roc_curve, auc
# Assuming your dataset is in a DataFrame named 'df'
# If it's in a CSV file, you can read it using: df = pd.read_csv('your_dataset.csv')
# Drop the 'Unnamed: 0' column if it's an index column
df = df.drop('Unnamed: 0', axis=1, errors='ignore')
# Data Preprocessing
# Assuming 'review' is your feature and 'rating' is your target variable
X = df['review'].astype(str) # Ensure 'review' column is treated as strings
y = df['rating']
# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
# Feature extraction using TF-IDF vectorizer
tfidf_vectorizer = TfidfVectorizer(max_features=5000) # You can adjust max_features as needed
X_train_tfidf = tfidf_vectorizer.fit_transform(X_train)
X_test_tfidf = tfidf_vectorizer.transform(X_test)
# Model building - Logistic Regression
model = LogisticRegression(random_state=42)
model.fit(X_train_tfidf, y_train)
# Model evaluation
y_pred = model.predict(X_test_tfidf) # Print accuracy and Classification report
print("Accuracy:", accuracy_score(y_test, y_pred)) print("Classification Report:\n", classification_report(y_test, y_pred))
```

```
Accuracy: 0.41332116788321166
Classification Report:
```

	precision	recall	f1-score	support
1	0.37	0.39	0.38	373
2	0.00	0.00	0.00	71
3	0.00	0.00	0.00	104
4	0.50	0.02	0.04	100
5	0.44	0.68	0.53	448
accuracy			0.41	1096
macro avg	0.26	0.22	0.19	1096
weighted avg	0.35	0.41	0.35	1096

➤ # Confusion Matrix

```
In [ ]: conf_matrix = confusion_matrix(y_test, y_pred)
plt.figure(figsize=(8, 6))
sns.heatmap(conf_matrix, annot=True, fmt='d', cmap='Blues', xticklabels=['1', '2', '3', '4', '5'], yticklabels=['1', '2', '3', '4'])
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title('Confusion Matrix')
plt.show()
```

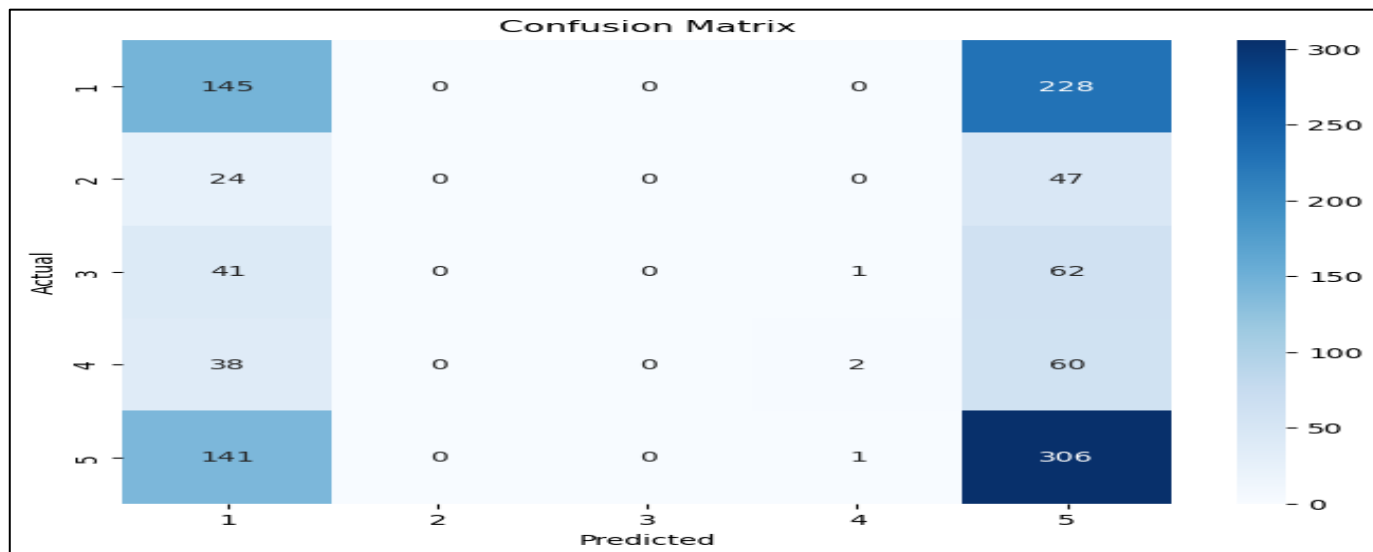


Fig 4: Confusion Matrix of Zomato

➤ #percentage Distribution of ratings

```
In [ ]: plt.figure(figsize=(15,5))
df['rating'].value_counts().plot(kind='pie', autopct='%1.1f%%', fontsize=10, figsize=(10,8))
plt.show()
```

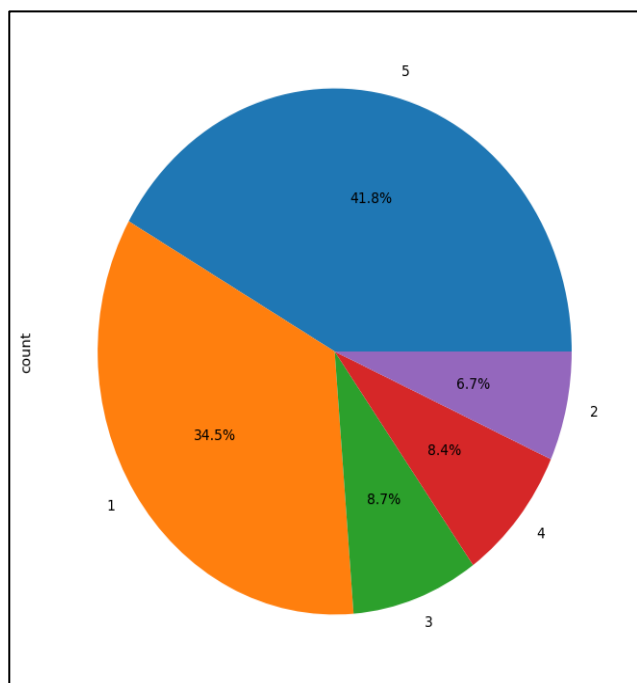


Fig 5: Percentage Distribution of Ratings

➤ df['rating'].hist()

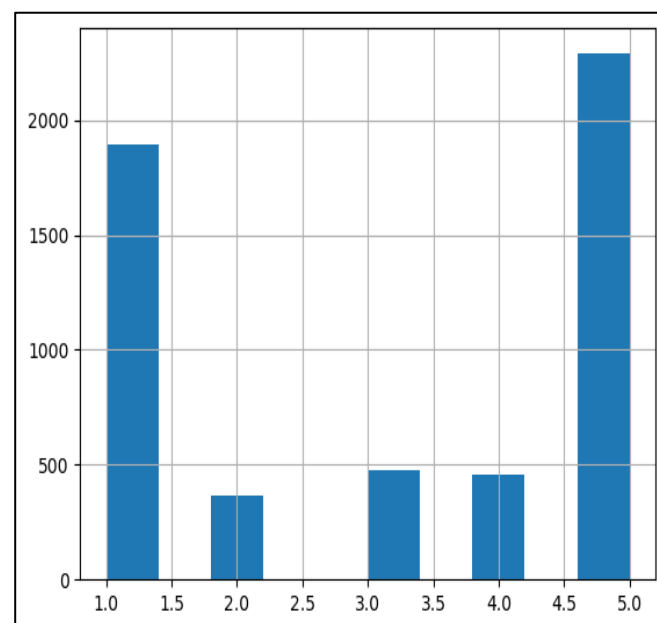


Fig 6: Comparison graph for all Zomato Ratings Algorithm

#### IV. CONCLUSION

In this paper, we showed the significance of audits and how they influence nearly everything connected with online information. Clearly surveys assume a vital part in individuals' choice. In this way, counterfeit surveys identification is a clear and continuous examination region. In this paper, an AI counterfeit surveys location approach is introduced. In the proposed approach, both the elements of the audits and the conduct highlights of the commentators are thought of. Various classifiers are executed in the created approach. Likewise, the outcomes show that considering the conduct highlights of the commentators increment the f-score by 3.80%. Not all commentators conduct highlights have been thought about in the ongoing work. It is profoundly expected that considering more social elements will upgrade the exhibition of the introduced counterfeit surveys discovery approach.

#### REFERENCES

- [1]. Alexandridis, G., Varlamis, I., Korovesis, K., Caridakis, G., Tsantilas, P. (2021). A survey on sentiment analysis and opinion mining in Greek social media. *Information*, 12(8): 331. <https://doi.org/10.3390/info12080331>
- [2]. Hemmatian, F., Sohrabi, M.K. (2019). A survey on classification techniques for opinion mining and sentiment analysis. *Artificial intelligence review*, 52(3): 1495-1545. <https://doi.org/10.1007/s10462-017-9599-6>
- [3]. Hossain, M.F. (2019). Fake review detection using data mining. Missouri State University. <https://bearworks.missouristate.edu/theses/3423>.
- [4]. Elmurngi, E., Gherbi, A. (2018). Fake reviews detection on movie reviews through sentiment analysis using supervised learning techniques. *International Journal on Advances in Systems and Measurements*, 11(1): 196-207.