**Restaurants Rating Prediction using Machine Learning  
Algorithms**

**Abstract:**

Restaurant Rating has become the most commonly used parameter for judging a restaurant for any individual. A lot of research has been done on different restaurants and the quality of food it serves. Rating of a restaurant depends on factors like reviews, area situated, average cost for two people, votes, cuisines and the type of restaurant. The main goal of this is to get insights on restaurants which people like visit and to identify the rating of the restaurant. With this article we study different predictive models like Support Vector Machine (SVM),Random forest and Linear Regression, XGBoost, Decision Tree and have  
achieved a score of 83% with ADA Boost.

**INTRODUCTION**

Zomato is the most reputed company in the field of food reviews.  
Founded in 2008, this company started in India and now is in 24  
different countries. Its is so big that the people now use it as a verb.  
“Did you know about this restaurant? Zomato it”. The rating is the  
most important feature of any restaurant as it is the first parameter  
that people look into while searching for a place to eat. It portrays the quality, hygiene and the environment of the place. Higher ratings lead to higher profit margins. Notations of the ratings usually are stars or numbers scaling between 1 and 5.  
Zomato has changed the way people browse through restaurants. It has helped customers find good places with respect to their dining budget. Different machine learning algorithms like SVM, Linear regression, Decision Tree, Random Forest can be used to predict the ratings of the restaurants.

**DATA SET DESCRIPTION**

This is a kaggle dataset.  
(https://www.kaggle.com/himanshupoddar/zomato-bangalore-  
restaurants).  
It Represents information of Restaurants in the City of Bangalore

**PreProcessing**

The Dataset contained 17 Attributes. Records with null values were dropped from ratings columns and were replaced in the other columns with a numerical value. Values in the ‘Rating’ column were changed. The ‘/5’ string was deleted. For eg. If the rating of a restaurant was 3.5/5, it was changed to 3.5.  Using LabelEncoding from sklearn library, encoding was

done on columns like book\_table,online\_order,rest\_type,listed\_in(city).

**Feature Selection**

We did not use any feature selection algorithms but eliminated some

columns due to available domain knowledge and thorough study of

the system.

Dropped columns mentioned below:

 URL

 Address

 Dish\_liked

 Phone

 Menu

 Review\_list

 Location

 Cuisine

Some of these columns may look like they are important but all of the

same information could be found in other columns with lesser

complexity.

The Columns being used are as follows:

 Name

 Online\_order

 Book\_table

 Votes

 Rest\_type

 Approx. cost of two people

 Listed\_in(type)

 Listed\_in(city)

**EXPLORATORY DATA ANALYSIS**

A lot of effort went into the EDA as it gives us a detailed

knowledge of our data.

Exploratory Data Analysis (EDA) is an approach/philosophy for

data analysis that employs a variety of techniques (mostly graphical) to maximize insight into a data set; uncover underlying structure; extract important variables; detect outliers and anomalies; test underlying assumptions.

**CONCLUSIONS**

This paper studies a number of features about existing restaurants of different areas in a city and analyses them to predict rating of the restaurant. This makes it an important aspect to be considered, before making a dining decision. Such analysis is essential part of planning before establishing a venture like that of a restaurant.  
Lot of researches have been made on factors which affect sales and market in restaurant industry. Various dine-scape factors have been analyzed to improve customer satisfaction levels.  
If the data for other cities is also collected, such predictions could be made for accurate.