Food Industry Data Anlaysis Using R-Analytics

Pallavi R. Raut[1], Rasika P. Kurade[1], Shubham A. Peshane[1], Niraj O. Chafle[1], Rajesh Babu[2]

[1]*BE Students, Department of Computer Science Engineering, Tulsiramji Gaikwad-Patil College of Engineering & Technology, Nagpur, Maharashtra, India.*

[2]*Assistant Professor, Department of Computer Science Engineering, Tulsiramji Gaikwad-Patil College of Engineering & Technology, Nagpur, Maharashtra, India.*

Abstract— *This project intends to tap the power of classification techniques to predict the success of restaurant. Decision Tree is one of the classification technique which is used for this particular use case. But the logic used here is applicable to many existing small businesses and people intending to start a business. There are many ingredients in a food item served at a restaurant but the item is best when the right ingredients are used in right proportion and cooked to right temperature. A good chef is essential to a good restaurant but many successful chefs have failed to make a successful restaurant.*
Restaurant, unlike a food item, is a business and it will only succeed if its customers are happy. We don't need a psychic to predict what will make a customer happy, however, we can use our psychology/common sense to easily identify the essential things to run a restaurant. 99.9% of restaurant owners in the world have the basic understanding of how to run their business, yet, not all of them are successful. Which means, we need a psychic who can tell us what the customer is thinking and what is that one thing / group of things that will make him/her happy. This project intends to reveal exactly that, but, with the help of Decision Tree classification technique.
Keywords— *Internet of Things, Rstudio, Xampp, Windows 10, Datasets , Intel Processor i5.*

I.   Introduction

Restaurant anlaysis will thrive in future is an important concern for all shopkeepers. Knowing about long-term trends, shopkeepers can take corresponding actions in advance. For instance, if they know the store will come to a crisis in a matter of months, they could take steps to avoid the misfortune such as to make changes on the style of the store, or even consider choosing a new placement to minimize economic losses. Usually store owners make long-term decisions based on empirical judgement. Due to limited data sources and lack of analytic tools, it is traditionally a challenge to make data-driven decisions [1].

For example, in case of a restaurant, the food, the ambience, the service or even the discounts offered can often influence the user ratings. This information is not conceivable from rating alone, however, it is present in the reviews that users write. The classification of yelp restaurant reviews into one or more, "Food", "Service", "Ambience", "Deals/Discounts", and "Worthiness", categories is the problem in consideration. Inputs are the Help restaurant reviews and review ratings [2]. For example, in case of a restaurant, the food, the ambience, the service or even the discounts offered can often influence the user ratings. This information is not conceivable from rating alone, however, it is present in the reviews that users write. The classification of yelp restaurant reviews into one or more, "Food", "Service", "Ambience", "Deals/Discounts", and "Worthiness", categories is the problem in consideration. Inputs are the Help restaurant reviews and review ratings [3].

Restaurant business is unique because a successful restaurant becomes a tradition down the line and people like to visit their favorite restaurant even if they have moved into another suburb. Even in an extremely competitive market, good restaurants thrive. One reason is that there is lot of customization and personal touch that one can offer to its customer, in this line of business.

Data mining is a particular data analysis technique that focuses on modeling and knowledge discovery for predictive rather than purely descriptive purposes, while business intelligence covers data analysis that relies heavily on aggregation, focusing mainly on business information. In statistical   applications,   data   analysis can   be   divided into descriptive statistics, exploratory data analysis (EDA), and confirmatory data analysis (CDA). EDA focuses on discovering new features in the data while CDA focuses on confirming or falsifying existing hypotheses. Predictive analytics focuses on application of statistical models for predictive forecasting or   classification,   while text analytics applies statistical, linguistic, and structural techniques to extract and classify information from textual sources, a species of unstructured data. All of the above are varieties of data analysis.

Analysis refers to breaking a whole into its separate components for individual examination. Data

analysis is a process for obtaining raw data and converting it into information useful for decision-making by users. Data are collected and analyzed to answer questions, test hypotheses or disprove theories Analysis refers to breaking a whole into its separate components for individual examination. Data analysis is a process for obtaining raw data and converting it into information useful for decision-making by users. Data are collected and analyzed to answer questions, test hypotheses or disprove theories.

Internet of things is an innovation which will in general interface all the article on the planet to the internet. Applications are produced dependent on IoT empowered gadgets for monitoring and control in different spaces including modern procedures, home apparatuses. Inferential

statistics includes techniques to measure relationships between particular variables. For example, regression analysis may be used to model whether a change in advertising (independent variable X) explains the variation in sales (dependent variable Y). In mathematical terms, Y (sales) is a function of X (advertising). It may be described as $Y = aX + b + error$, where the model is designed such that a and b minimize the error when the model predicts Y for a given range of values of Analysts may attempt to build models that are descriptive of the data to simplify analysis and communicate results.

The most important aspect of any business is the ability to understand your customers. If you know your customers well, what they buy, when they buy, how they buy etc., half your job is done. When your business is small and your clientele is limited, it is easy to remember your customers at a personal level, often knowing what their favourite orders are. But, when your organization gains scale and it is difficult to even recognize them by face it is important not to lose track of that very attribute which helped you establish your business in the first place – your ability to understand your customers ! When you have more than one restaurant, the job becomes even more difficult. That is why you need to harness the power of data. But before you do that, you must devise mechanisms to collect the data on your customers. If you do not have a mechanism to collect data that identifies your repeat customers- don't lose heart. You can still do a lot by analyzing the POS(Point of Sale) data.

II.   Literature Review

We have dissected a few papers beneath.This paper [4] has genius represented By "The Use of Machine earning Algorithms in Recommender Systems:A Systematic Review Ivens Portugal, etal" The complexity of many existing data mining algorithms is exponential with respect to the number of dimensions. This paper [5] proposes Recommender systems use algorithms like PSO Algorithm. PSO is a stochastic, population-based evolutionary algorithm particularly suitable for solving multivariable optimization problems.

This paper [6] By "A Hybrid Restaurant Recommender Prema Dwivedi etal" Recommender systems use algorithms to provide users with product or service recommendations.Our end goal is to be able to consolidate the information using ASO Algorithm.We conduct comprehensive experiments on three different cities, and that the conclusions are quit consistent.

The ralylitics [7] proposes By "Recommendation System For Restaurants Anusha Jayasimhan, etal" This paper presents a systematic review of the literature that analyzes the use of machine learning algorithms in recommender systems.We consider the problem of classifying reviews by overall sentiment. Rather than predict a simple positive or negative evaluation with 72% accuracy.It is fairly straightforward more of thinking-to recommend restaurant to specific profile.

The ranalytics [5] proposes By "Classification of Recommender System : A review Shahab Saquib Sohail , Etal" International Journal of computer application (0975- 8887) Received 31 july 2017 , accepted 18 September 2017. The data we are using is user centered , more specifically, we quiry a user via Zomato's API. To build a recommendation to provide personalized restaurant Recommendations to users using PSO,ASO and various algorithm.
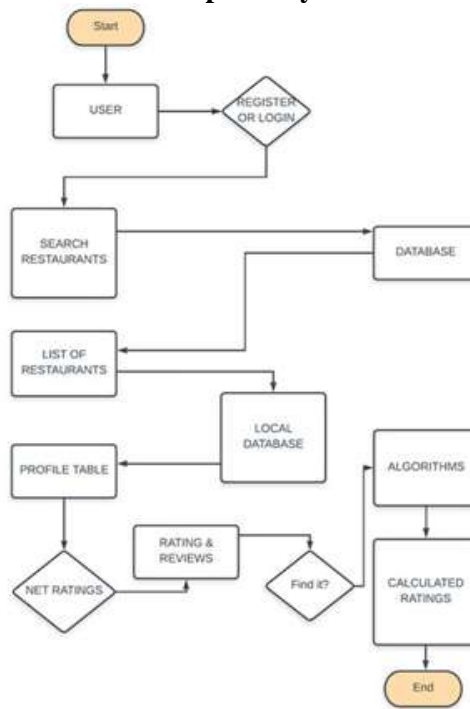
III. Proposed System

Figure 1. Block Diagram for Flow Chart

For instances ,if they know the store will come to crisis in a matter of months , they could take steps to avoid the miss fortune such as to make changes on the style of the store, or in consider choosing a new placement to minimize economic losses. Usually store owners make long term decisions based on empirical judgment .Due to limited data sources and lackof analytic tools , it is traditionally challenge to make data-driven decisions. With the development of information techniques, especially the growth of online location-based services a large amount of business related data can be collected through the internet.

For Ex. People may post check-ins at some point of interest (POI) they are visiting; after consuming in a shop , they can write reviews on Yelp to show how they like the shop. Thus, there is a potential to exploit heterogeneous information to built automatic business intelligence tools for an enhancing the decision process.
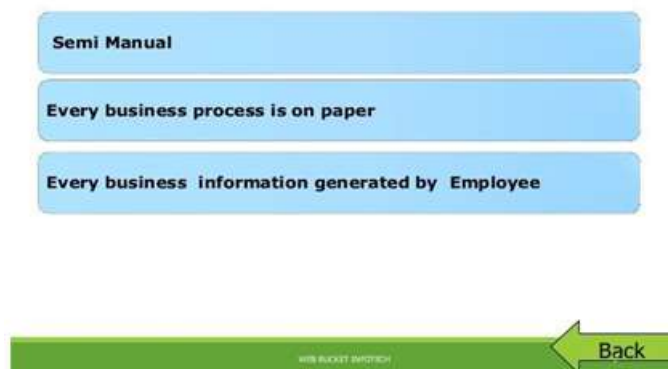
IV. Existing System

Figure 2. Existing System

The main data source we use in this paper is Dianping.com Dianping , known as "Yelp for china", is the largest consumers review site in china. It offers multi level knowledge through it's diverse functions such as reviews, check-ins and POI meta data. All the data we have crawled is publicly available on the website. Data

crawling process finished in april 2014. In the literature of churn analysis, a user is usually defined as a churner if he/she doesn't have any data during the last several periods of the data sets. We start by exploring the geographic fetures, which describe the location environment of the retailer store. The location and near by places play an important role in popularity of the shop.

Recent works have studied the economic impact of geographic and user mobility factors on the retailer store. As formulated in these works, geographic signs contain the types and density of nearby places, and user mobility includes transitions between venues or the incoming own of mobile users from distant areas. Inspired by them, we rest analyze these two types of features

V. Data Analysis

Depending upon the three parameters we are exploring data for restaurants, these parameters are rating, costing and ranking of the restaurants. Performing data analysis in several cities like Bangalore, Bhopal, Indore, Mumbai, Nagpur, Pune. In tire-1 we are using web scraping with R in city Bangalore, in tire-II Mumbai, in tire-III Bhopal, Indore and Nagpur. We are doing This Data analysis for restaurant data and their market in above different cities. Web scraping is an advanced task that not many people perform. Web scraping with R is, certainly, technical and advanced programming. An adequate understanding of R is essential for web scraping in this way.

To start with, R is a language for statistical computing and graphics. Statisticians and data miners use R a lot due to its evolving statistical software, and its focus on data analysis.One reason R is such a favorite among this set of people is the quality of plots which can be worked out, including mathematical symbols and formulae wherever required. R is wonderful because it offers a vast variety of functions and packages that can handle data mining tasks. rvest, RCrawler etc are R packages used for data collection processes.
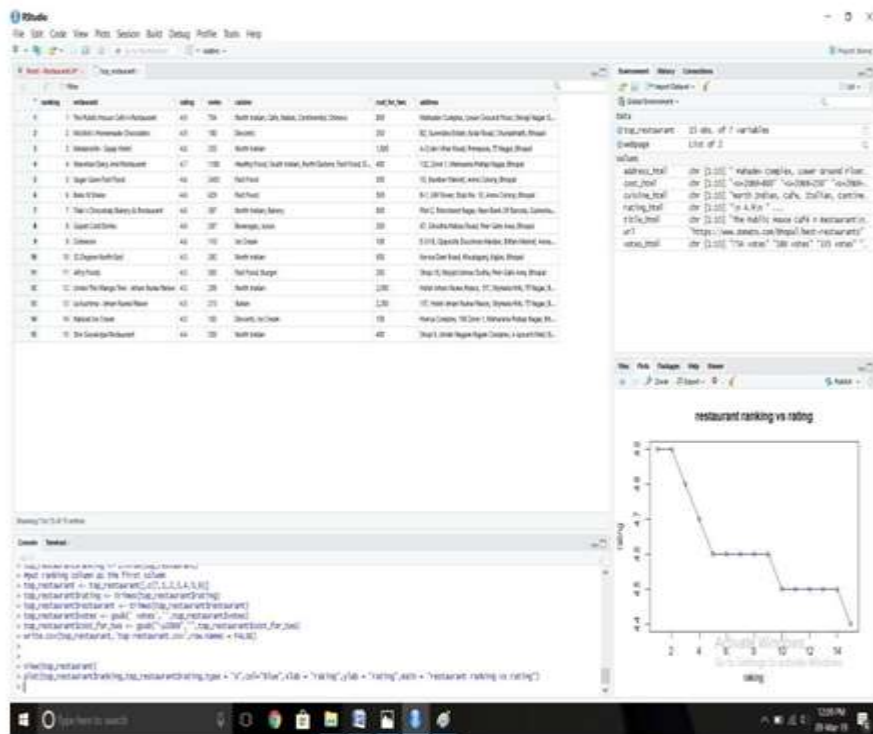

Figure 3. Data Analysis of Bhopal

review text can react a restaurant's operating status best. Comprehensive experiments show that integrating different predictors can lead to the best model, and it is common among different cities. In the future study, we are going to : (1) investigate more appropriate language models to extract better knowledge from review text; (2) design a unified model to incorporate heterogeneous learning algorithms so that the performance will not limited by a single learning algorithm such as GB-DT
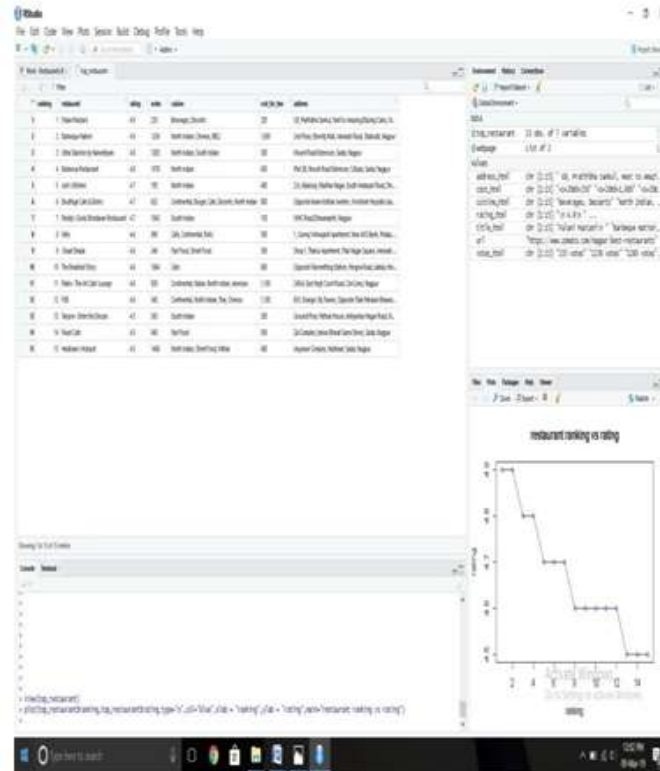
Figure 4. Data Analysis of Nagpur

VI.   Conclusions

This paper discusses the problem of restaurant survival prediction by modeling four perspectives: geographical metrics, user mobility, rating scores, and review text. We provide detailed analysis on each perspective separately and demonstrate its predictive power. We and that if used properly,

References

[1].   Ariyasriwatana, W., Buente, W., Oshiro, M., &Streveler, D. (2014). Categorizing health-related cues to action: using Yelp reviews of restaurants in Hawaii. New Review of Hypermedia and Multimedia, 20(4), 317-340.

[2].   Byers, J. W., Mitzenmacher, M., &Zervas, G. (2012, June). The grouponeffect on yelp ratings: a root cause analysis. In Proceedings of the 13th ACM conference on electronic commerce (pp. 248-265).ACM.

[3].   Hicks, A., Comp, S., Horovitz, J., Hovarter, M., Miki, M., & Bevan, J. L. (2012). Why people use Yelp.com: An exploration of uses and gratifications. Computers in Human Behavior, 28(6), 2274-2279.

[4].   Mukherjee, A., Venkataraman, V., Liu, B., & Glance, N. S. (2013, July).What yelp fake review filter might be doing?.In ICWSM IOT based Crop- field Monitoring and Irrigation Automation, Raja Lakshmi Mrs. S.Devi Maha Lakshmi

[5].   dos Santos, C. N., &Gatti, M. (2014). Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts.In COLING (pp. 69-78).Smart Drip Irrigation System using Raspberry pi and Arduino, Nikhil Agarwal, Smita Singhal.

[6].   Mullen, T., & Collier, N. (2004, July). Sentiment Analysis using Support Vector Machines with Diverse Information Sources.In EMNLP (Vol. 4, pp. 412-418).

[7].   Kiritchenko, S., Zhu, X., Cherry, C.,& Mohammad, S. (2014, August). NRC-Canada-2014: Detecting aspects and sentiment in customer reviews. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014) (pp. 437-442). Dublin, Ireland: Association for Computational Linguistics and Dublin City University.

[8].   Huang, J., Rogers, S., &Joo, E. (2014). Improving restaurants by extracting subtopics from yelp reviews.iConference 2014 (Social Media Expo).