Research Inventy: International Journal of Engineering And Science
Vol.6, Issue 4 (April 2016), PP -91-94
Issn (e): 2278-4721, Issn (p):2319-6483, www.researchinventy.com

Restaurant Revenue Prediction using Machine Learning

Prof. Nataasha Raul, Yash Shah, Mehul Devganiya
Department of Computer Engineering Sardar Patel Institute of Technology Mumbai, India

***Abstract*:** Currently, making a decision about when and where to open new restaurant outlets is subjective in nature based on personal judgement and development teams' experience. This subjective data is difficult to extrapolate across geographies and cultures. Our supervised learning algorithm will construct complex features using simple features such as opening date for a restaurant, city that the restaurant is in, type of the restaurant (Food Court, Inline, Drive Thru, Mobile), Demographic data (population in any given area, age and gender distribution, development scales), Real estate data (front facade of the location, car park availability), and points of interest including schools, banks. Applying concepts of machine learning such as support vector machines and random forest on these parameters, it will predict the annual revenue of a new restaurant which would help food chains to determine the feasibility of a new outlet.
Keywords: *Machine Learning; Random Forest; SVM; Restaurant; Revenue; Prediction*

I. INTRODUCTION

New restaurant outlets incur huge time and capital investments to establish. When the new outlet fails to break even, the site closes within a short time and operating losses are incurred. Finding an algorithmic model to increase the return on investments in new restaurant sites would facilitate businesses to direct their investments in other important business areas, like innovation, and training for new employees. The problem can be defined as: design an automated approach to decide the task environment for new restaurant by applying concepts of Support Vector Machines, Gaussian Naive Bayes and Random Forest on certain parameters, it will predict the annual revenue of a new restaurant outlet which would help food chains to determine its feasibility. The primary objective of Restaurant Revenue Prediction using Machine Learning is to help restaurants make a more informed and optimal decision about opening new outlets. It aims to find an algorithmic model to increase the effectiveness of investments in new restaurant sites. One of the biggest features of the proposed application is that it aims to predict the revenue of new outlets of existing restaurant chains. Analytical prediction of data has proven more effective than by human judgement. Further, it can allow analysis and comparison of multiple new sites. Thus human errors can be avoided and operations can be performed faster than previous methods. Given a dataset with 37 obfuscated parameters, the algorithm will be trained on these parameters and no more. All in all, this revenue prediction system will compute an accurate forecast of a restaurant outlet's future revenues.

III. LITERATURE REVIEW

Multiple machine learning algorithms were studied for predicting the annual revenue of new restaurants. Research papers on Support Vector Machines and Stochastic Neighbor Embedding were referred. A research study on restaurant opportunities in India was read for better understanding.

A. Visualization and Interpretation of SVM Classifiers

This paper shows examples such as the data piling phenomenon for high-dimensional data improved understanding of SVM parameter tuning and SVM modelling of unbalanced data sets. This method has been used to explain the conditions for the effectiveness of Universal Learning. This method is used for multi-class problems, and to improve SVM model selection for unbalanced classification problems. In conclusion, this paper points out that all additional insights about interpretation and visualization of high dimensional SVM-based classifiers, have resulted from incorporating important properties of SVM models into a simple univariate graphical representation.[1]

B. Stochastic Neighbor Embedding

Stochastic Neighbor Embedding finds its application in mapping high-dimensional points into a low-dimensional space based on stochastic selection of similar neighbors. In self-organizing maps, the low-dimensional coordinates are fixed to a grid and the high-dimensional ends are free to move. But in SNE the high-dimensional coordinates are

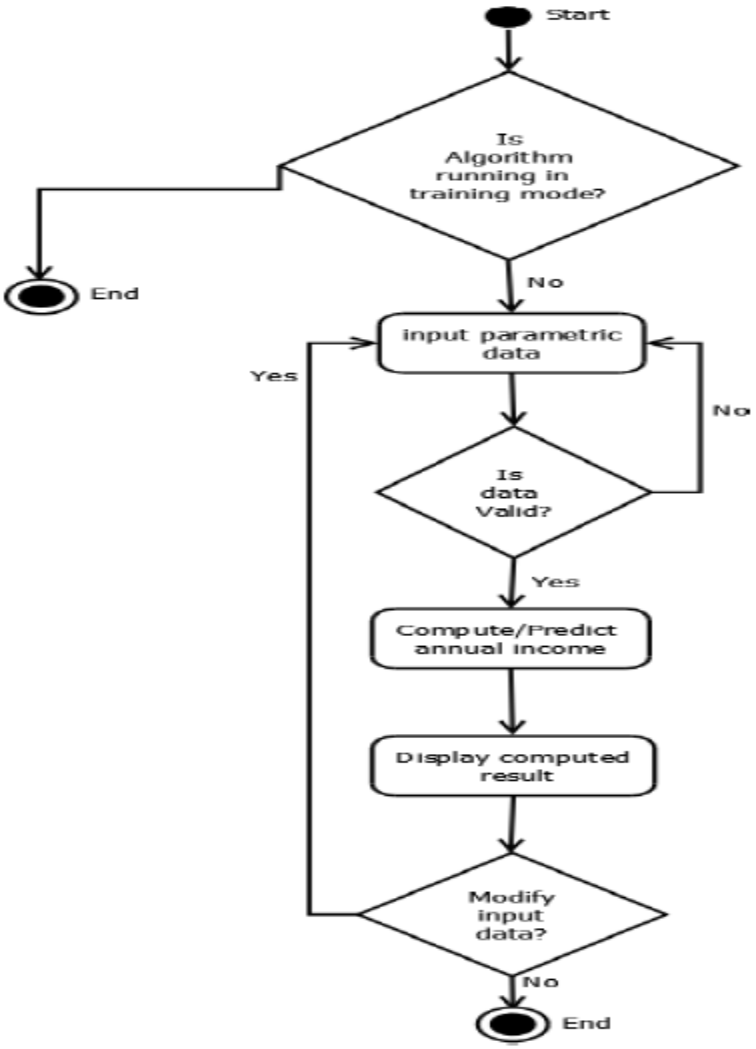
fixed to the data and the low-dimensional points move. The gradient of the SNE cost function has an appealing property in which the forces acting on yi to bring it closer to points it is under-selecting and further from points it is over-selecting as its neighbor. Since SNE has probabilistic formulation, it has the ability to be extended to mixtures in which ambiguous high-dimensional clusters can have several widely-separated points in the low-dimensional space. [2]

C. Restaurant Opportunities in India: Trends and Opportunities

As the costs of opening a restaurant and running it profitably continue to climb, restaurant owners need to be as certain as possible that the kind of operation they envision has a very good potential for success at a particular site. One way to find out is to conduct a feasibility study. A feasibility study approach involves gathering and analyzing a great deal of information, ranging from demographics to design, which helps the operator make a better informed decision about the potential success of a specific concept at a certain location. [3]

III. PROPOSED SYSTEM

A. System Flow



The system checks if the system is running in training mode. If not, it lets the user input parametric data. The data is checked for validation. In case of invalid data, the user is asked to reenter data. Otherwise our algorithm, computes the annual income and displays the predicted result.

B. Dataset/Features

The data set we obtained from kaggle is peculiar in many aspects. The data set consists of a training and test set with 137 and 100,000 samples respectively. This is interesting in itself since such a small training set is presented relative to the final test set. The 137 training samples provide actual revenue while the test set does not and expects the user to submit their predictions on the 100,000 testing examples. The 43 features provided are listed below:

 P-Variables (parameters from P1 to P37): Obfuscated variables from three categories: **demographic data,** which includes population, age, gender; **real estate data** which includes car parking availability and number of front facade; **commercial data** denotes points of interest like banks, schools, other public places etc. Each variable may contain a combination of the three categories or may be mutually exclusive

 ID: Restaurant ID

 Open Date: Date that the restaurant opened in the format MM/DD/YYYY

 City: name of the city in which the restaurant is located

 City Group: city group can be either *metro*, *big city* or *other*

 Type: Restaurant type where *FC* - Food Court *IL*- Inline *DT* - Drive through *MB* - Mobile

 Revenue: Annual revenue of a restaurant in a given year and is the target

As shown, the majority of the data fields are obfuscated variables without giving the statistician any prior knowledge of each one. [6]

IV. IMPLEMENTATION

The proposed system takes in the value of features from the user. The data fields such as opening date of the restaurant, restaurant type, city name, city type, number of front sides, car parking and points of interest are taken as shown in Fig. 1 below and generates the approximated Revenue depending upon inputs provided.

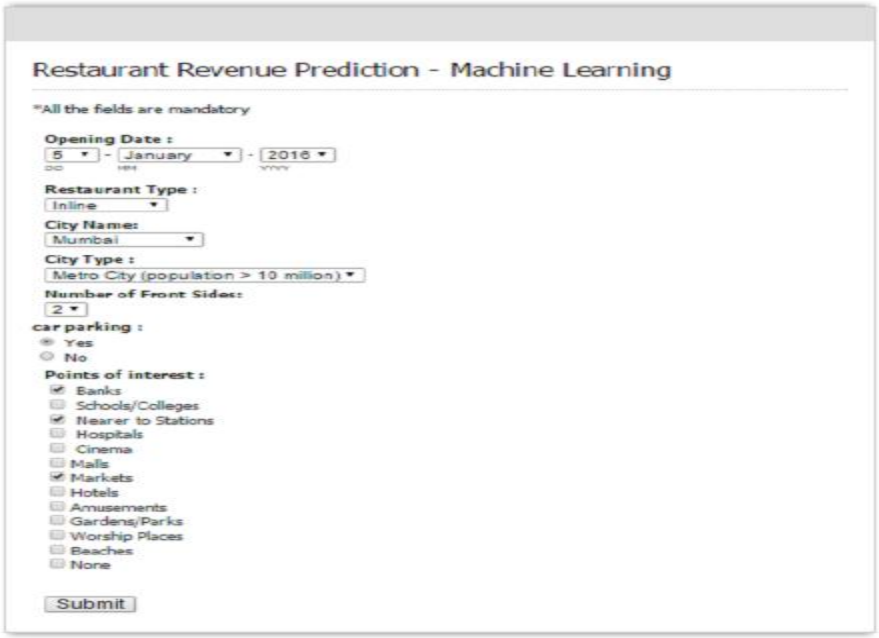


Fig. 1: Features accepting User Interface

The output shown in Fig. 2 is generated in correlation with the preceding input set of values. The output value indicates the annual revenue of the proposed restaurant site. The output is given in US dollars.

Fig. 2: Predicted Revenue

One of the models we are focusing on is random forest. Random forests are promising because they have desirable characteristics, such as running efficiently on large datasets and flexibility, having been used effectively for a variety of applications. A random forest is an ensemble of decision trees created using random variable selection and bootstrap aggregating (bagging). What this means is that first a group of decision trees are created. For each individual tree, a random sample with replacement of the training data is used for training. Also, at each node of the tree, the split is created by only looking at a random subset of the variables. The prediction is made by averaging the predictions of all the individual trees. Random forests can also provide an error estimate, called the out-of-bag error. This is computed by feeding the individual trees cases that they have not seen. The
training data that was not included in the bootstrap sample for a given tree is fed through that tree, and a prediction is made. The results of doing this for all trees are computed, and for each sample (row) an out of bag estimate is created by taking the mean of the results of all the trees.

V. RESULTS

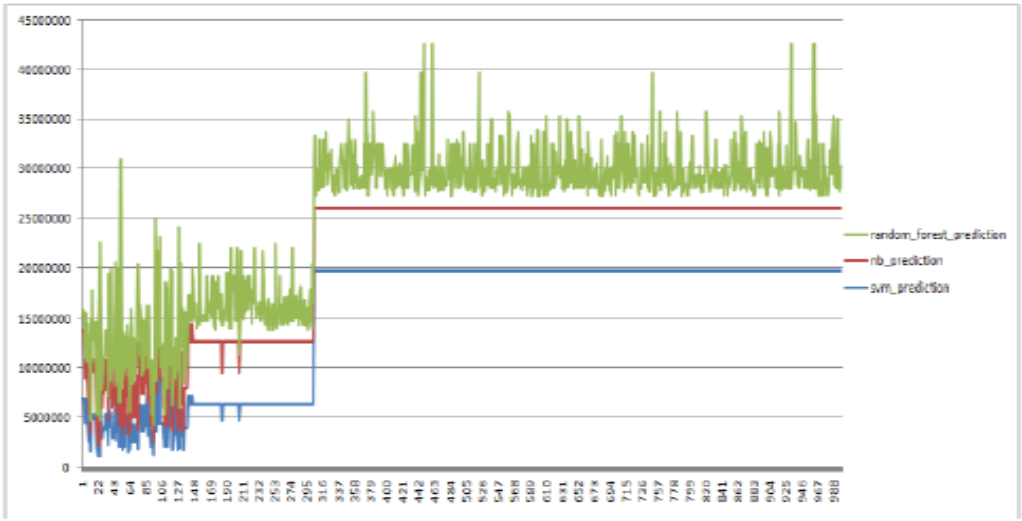


Fig. 3: Variation in results prediction using different algorithms

Fig. 3 indicates random forest (in green, uppermost line) to predicts the annual revenue as it produced a wider range of values whereas the predictions by Gaussian Naïve Bayes (in red, middle line) and SVM (in blue, lowermost line) generated a straight line over the first thousand test examples.

VI. CONCLUSION

Thus, the concept of prediction system for future revenues of new restaurant outlets can be developed as shown. Random forest and SVM were used to predict the annual revenue. Using this, a reference can be provided to aid human judgement and operational losses can be minimized for food chains.

REFERENCES

[1] SauptikDhar, Vladimir Cherkassky, "Vizualization and Interpretation of SVM Classifiers", Wiley Interdisciplinary Reviews

[2] Geoffrey Hinton, Sam Roweis, "Stochastic Neighbor Embedding", University of Toronto

[3] *"Restaurant Opportunities in India: Trends and Opportunities"*, http://www.hvs.com/Content/1336.pdf , 2004

[4] Wen-Chyuan Chiang, Jason C.H. Chen, XiaojingXu, "*An overview of research on revenue management : current issues and future research*, International Journal of Revenue Management", Vol. 1, 2007

[5] "Dataset : Restaurant Revenue Prediction", https://www.kaggle.com/c/restaurant-revenue-prediction