УДК 004.021
МРНТИ 20.51.23 DOI 10.55452/1998-6688-2021-18-3-36-41

PREDICTING FUTURE VISITORS IN THE RESTAURANT BUSINESS USING MACHINE LEARNING

BORTAN A.Y., BAISAKOV B.M.
Kazakh-British Technical university, 050000, Almaty, Kazakhstan

Abstract. *Restaurant owners must reliably assess restaurant customers in order to function effectively and productively to enhance the restaurant's service. It is important to have a successful forecast in order to prevent losses and boost service and market optimization. There are a variety of machine learning (ML) approaches that can be used to make these predictions, but each visitor is unique and will act in a unique way. As a result, we want to estimate how many guests a restaurant may expect in the future using big data and supervised training in this study. We used three different machine learning methods in a real dataset from supervised training to predict how many visitors a restaurant dataset "Recruit restaurant visitor forecasting" will receive: Neural Network, XGBoost and Random Forest regressor. The predicted values were compared to the real data after the simulation. Basically, algorithms used had mean errors of less than 9.5278, but the Random Forest regressor exceeded, with mean errors of 9.2902.*

Key words: *prediction, machine learning, big data, supervised training, dataset, XGBoost, Random Forest regressor, Neural Network.*

МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІН ҚОЛДАНА ОТЫРЫП МЕЙРАМХАНА БИЗНЕСІНДЕГІ СҰРАНЫСТЫ БОЛЖАУ

БОРТАН Ә.Е., БАЙСАКОВ Б.М.
Қазақстан-Британ техникалық университеті, 050000, Алматы, Қазақстан

Аңдатпа. *Экономикалық және өнімді жұмыс жасау, мейрамхана қызметін жақсарту үшін мейрамхана иелері мейрамхана клиенттерін дәл бағалауы керек. Жақсы болжам ысырапшылдыққа жол бермеу және қызмет көрсету мен бизнесті оңтайландыру үшін қажет. Бұл болжамдарда қолдануға болатын көптеген машиналық оқыту әдістері бар; дегенмен әр келуші сан түрлі. Сондықтан осы жұмыста үлкен деректерді және бақыланатын оқытуды қолдана отырып, мейрамхана болашақта қанша келуші күтетінін болжағымыз келеді. Бақыланатын оқытудың нақты деректер жиынтығында үш түрлі машиналық оқыту әдістері қолданылды. «Мейрамханаға келушілерді болжау» деректер жинағында қанша келуші бар екенін болжағымыз келеді: XGBoost, кездейсоқ орман регрессоры және нейрондық желі. Симуляциядан кейін болжамды мәндер нақты деректермен салыстырылды. Жалпы алғанда, қолданылған барлық алгоритмдер 9.5278-ден төмен орташа қателіктерге қол жеткізді, бірақ кездейсоқ орман регрессоры 9.2902 орташа қателіктерінен асып түсті.*

Түйінді сөздер: *болжам, машиналық оқыту, үлкен деректер, бақыланатын оқыту, деректер жиынтығы, XGBoost, кездейсоқ орман регрессоры, нейрондық желі.*

ПРОГНОЗИРОВАНИЕ СПРОСА В РЕСТОРАННОМ БИЗНЕСЕ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

БОРТАН А.Е., БАЙСАКОВ Б.М.
Казахстанско-Британский технический университет, 05000, Алматы, Казахстан

Аннотация. *Для экономичной и продуктивной работы, а также для улучшения обслуживания ресторана владельцам ресторана необходимо точно оценивать посетителей ресторана. Хороший прогноз необходим, чтобы избежать потерь и улучшить обслуживание и оптимизацию бизнеса. Есть много методов машинного обучения (МО), которые можно использовать в этих прогнозах;*

однако каждый посетитель индивидуален. Поэтому в этой работе, используя большие данные и контролируемое обучение, мы хотим предсказать, сколько посетителей ресторан может ожидать в будущем. Три различных метода машинного обучения были применены к реальному набору данных из контролируемого обучения, мы хотим предсказать, сколько посетителей в наборе данных «Прогнозирование посетителей ресторана»: XGBoost, регрессор случайного леса и нейронная сеть. После моделирования прогнозируемые значения сравнивались с реальными данными. В целом все применяемые алгоритмы достигли средних ошибок ниже 9,5278, но регрессор случайного леса превзошел их со средними ошибками 9,2902.

Ключевые слова: *прогноз, машинное обучение, большие данные, контролируемое обучение, набор данных, XGBoost, регрессор случайного леса, нейронная сеть.*

Introduction

Contrary to different methods of forecasting visitors for different designs, such as national tourism, as well as the demand for hotel rooms for accommodation (for example, [1], [2] [3] [4] [5]) in the literature, restaurant managers have little idea about the number of possible visitors in the future making use of big data. Current work, as an example [4], was only justified by customer return visits. Note that in some tourist destinations the number of new customers may exceed the number of old ones. Therefore,a new method to estimate the total number of potential restaurant guests on a given day is being established. It is first of all budget optimization, namely, not wasting restaurant resources. That is, the effective distribution of products into dishes, the correct calculation for cash for delivery, for exchanges. The quality of service and labor productivity is the competent distribution of employee work schedules. Demand forecasting, detailed purchase of fresh products and other products.

The remainder of this paper is set out as follows: “Related work” - includes the work that has been done in relation to and visitor prediction. “Machine Learning” section - involves data on the principles and methods used in this study. “Environmental Setup” - consists of how the monitoring of the environment has been organized and describes the stage of data preparation. “Results” - section compares prediction methods and real results, presenting all related outputs from our experiments. Overall, “Conclusions and Future works” - ecaps the paper's key points while also adding proposals for future research.

Related work

In the service sector, estimating the number of potential visitors is important. In various fields, researchers have suggested various methods for forecasting the number of potential visitors based on big data. For example, the characteristics of a place can be used to make predictions, such as forecasting customers' return visits to a restaurant based on the restaurant's attributes [4]. Furthermore, regional influence is a key factor in forecasting potential visitors to Points Of Interest (POIs) [5]. For forecasting future numbers, several machine learning regression algorithms may find associations between variables and consequences. For instance, the Support Vector Machine (SVM) [6] is a commonly used regression tool (also called SVR). It can also be used to classify objects. It creates a hyperplane to display the training data patterns. To complete learning tasks on non-linear distributions of training data, users can adjust the kernel functions of SVM. The quality of features, on the other hand, has a significant impact on SVM results. SVM can have poor accuracy if the training data contains irrelevant features. Random Forests (RF) [7] is a decision-tree-based regression and classification system. Due to the non-linear existence of decision trees, RF can operate with both linear and non-linear data without any previous knowledge of linearity. The mean prediction is used to produce the final prediction, which is based on a variety of decision trees. Each tree is robust against irrelevant features since it contains a subset of all features. Deep Neural Networks [8] are another effective approach. It can model highly non-linear relationships in training data because it can use several layers of networks from input to output. However, it requires a considerable amount of computation, resulting in a significant

increase in the hardware cost and computation time required for users to complete a machine learning task. XGBoost [9], a new decision-tree-based system, was recently proposed. It is based on the Gradient Boosting concept. The aim of XGBoost is to provide high performance without requiring lengthy computations. It outperformed all other algorithms in a wide variety of real-world datasets.

Machine learning

Machine learning is a form of artificial intelligence that is categorized as a subfield of computer science. It can be used in a variety of areas and one of its benefits is its ability to solve problems that cannot be expressed by explicit algorithms. Some machine learning approaches are regression-based, and can be used to make potential predictions with the aim of predicting a numeric target. The best machine learning model can depend on a balance between expected error and system complexity. A complex model can produce a higher error than a simple model, depending on the database characteristics, as shown in Fig. 1.

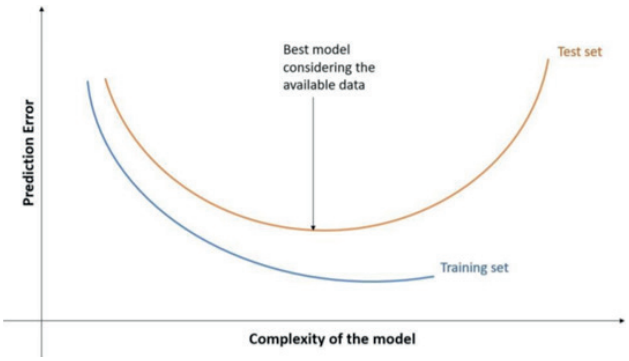

Figure 1. Model complexity versus prediction error.

Many of the approaches that will be discussed are regression-based. The methods used to create our proposed prediction model are described in the subsections below.

XGBoost is an optimized distributed gradient boosting library designed to be highly efficient, flexible and portable. It implements machine learning algorithms under the Gradient Boosting framework. XGBoost provides a parallel tree boosting (also known as GBDT, GBM) that solves many data science problems in a fast and accurate way. The same code runs on a major distributed environment (Hadoop, SGE, MPI) and can solve problems beyond billions of examples [10].

Decision Forest. Decision trees are non-parametric models that perform a sequence of simple tests for each instance, traversing a binary tree data structure until a leaf node (decision) is reached. The advantage of decision trees is that this method is efficient in both computation and memory usage during training and prediction. The Decision Forest model consists of an ensemble of decision trees. Each tree in a regression decision forest outputs a Gaussian distribution as a prediction. An aggregation is performed over the ensemble of trees to find a Gaussian distribution closest to the combined distribution for all trees in the model [11].

Neural Network Regression. Although neural networks are widely known for applications in deep learning and modeling complex problems, such as image recognition, they are easily adapted to regression problems. Any class of statistical models can be termed a neural network if they use adaptive weights and can approximate non-linear functions of their inputs. Thus, neural network regression is suited to problems where a more traditional regression model cannot fit a solution [12]. The layers of a neural network are made of nodes, the place where computation happens. A node combines input from the data with a set of coefficients, or weights, that either amplify or dampen that input, thereby assigning significance to inputs with regard to the task the algorithm is trying to learn. These input-weight products are summed and then the sum is passed through a node's so-called activation function, to determine whether and to what extent that signal should progress further through the network to affect the ultimate outcome [13].

Environment setup

Using a dataset in kaggle, which includes information about restaurants, historical visits and historical reservations, in order to train a dataset. The dataset was then imported, and the data was processed using the architecture we suggested in Fig. 2.

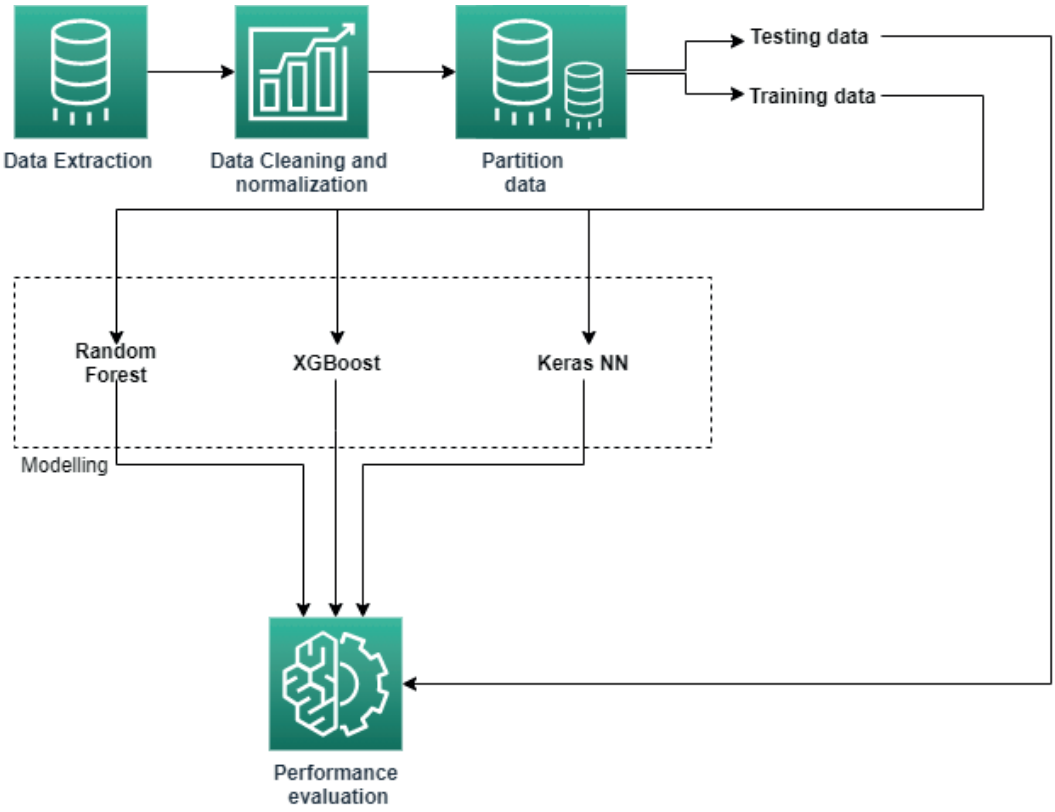

Figure 2. Process diagram.

Data Extraction. The first step in the process is data extraction. It is a compilation of all available data from sensors as well as weather data from external sources. Additionally, additional day classification features have been developed to boost algorithm decisions in later stages.

The general project contains 8 datasets and the data is taken from two servers: hpg and airReGI [14]. These sites are intended for users so that they can browse and reserve tables at suitable restaurants. Three datasets are associated with hpg and three with airREGI. And one dataset includes information about the date and another dataset is the location of the restaurants. Since the hpg datasets were slightly incorrect, of these listed datasets, we worked with airREGI and data info (Fig. 3). From the restaurant booking website AirREGI, we chose open large-scale real-world datasets.

Fig. 3. Dataset.

Data Cleaning and Normalization. The aim of this procedure is to ensure that the dataset is as homogeneous as possible. Data cleaning involves finding sections of data that are incomplete, incorrect, unreliable, or irrelevant, and then replacing, modifying, or deleting the dirty or coarse data. The aim of the normalization process is to convert the values of numeric columns in a dataset to a standard scale while preserving differences between variable values. By analyzing the data, we processed what dataset we need. The number of bookings is only a small fraction of the total number of visits (usually 10 times). And it is noticeable that the frequency begins with the beginning of the week. Attendance of the number of visitors is growing throughout the week.

Test and Training Dataset

After the data has been prepared and is ready to be processed, the dataset is randomly divided into two paths:

• 70% of data is sent to training models.

• 30% of the data is divided and used for test processing, with trained models being compared later.

Modelling

The previous stage's training dataset is used to train four different regression methods: XGBoost, Random Forest regressor and Neural Network.

Performance Evaluation

The predicted outcomes of the various models are compared to the real data. We'll use the mean absolute error and mean squared error formulas.

Results

Table 1. Prediction error

	XGBoost	Random Forest Regressor	keras NN
MAE	9.3452	9.2902	10.1669
Mean Error	172.59	179.77	226.14

Conclusion and future work

This research shows the use of different machine learning approaches in restaurant business, achieving mean errors of 9.5278 and 9.2902, respectively, in the best case scenario (boosted decision tree). Since each dataset pattern is various, different machine learning approaches should be used to find the right one for each task. We got a standard neural network model for predicting prospective restaurant visitors.

Need to improve:

- The number of visitors can be influenced by bad weather at the location of the restaurant.
- In terms of percentage of error, there is still a big effort to improve
- If a new restaurant is built next to an existing one, the number of potential visitors to the existing one will decrease.

REFERENCES

1. Xin Yang, Bing Pan, A. James, Evans, and Lv. Benfu, ”Forecasting Chinese tourist volume with search engine data”, Tourism Management, Vol 46, pp. 386-397, 2015.
2. Yang Yang, Bing Pan, and Haiyan Song, ”Predicting hotel demand using destination marketing organizations web traffic data”, Journal of Travel Research, Vol 53, no. 4, pp. 433-447, 2014.
3. Cho.Vincent, ”A comparison of three different approaches to tourist arrival forecasting”, Tourism management, Vol 24, no. 3, pp. 323-330, 2003.
4. Usep Suhud, and Arifin Wibowo, ”Predicting Customers Intention to Revisit A Vintage-Concept Restaurant”, Journal of Consumer Sciences, Vol 1, no. 2 (2016).
5. Shanshan Feng, Gao Cong, Bo An, and Yeow Meng Chee, ”POI2Vec: Geographical Latent Representation for Predicting Future Visitors”, In AAAI, pp. 102-108, 2017.
6. Corinna Cortes, and Vladimir Vapnik, ”Support-vector networks”, Machine learning, Vol 20, no. 3, pp. 273-297, 1995.
7. Leo Breiman, ”Random forests”, Machine learning, Vol 45, no. 1, pp. 5-32, 2001.
8. Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, ”Deep learning”, nature, Vol 521, no. 7553, pp. 436, 2015.

9. Tianqi Chen, and Carlos Guestrin, ”Xgboost: A scalable tree boosting system”, In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785-794. ACM, 2016.
10. XGBoost Documentation
11. https://xgboost.readthedocs.io/en/latest/
12. MICROSOFT,MachineLearningstudio.https://docs.microsoft.com/en-us/azure/machinelearning/studio-module-reference/decision-forest-regression. last accessed 2019/10/12.
13. MICROSOFT,MachineLearningstudio.https://docs.microsoft.com/en-us/azure/machinelearning/studio-module-reference/neural-network-regression. last accessed 2019/10/12
14. SKYMIND website. last accessed 2019/10/12.
15. AirREGI. https://air-regi.com/

Information about authors:

1. Bortan Aygerim Yesengeldikyzy – Master's student of the Faculty of Information Technology, Kazakh-British Technical University
Email: aigerim.bortan@gmail.com
2. Baisakov Beisenbek Miyatbekovich – Ph.D., professor, Kazakh-British Technical University
Email: beysenbek@gmail.com