EasyChair Preprint

№ 2968

Estimating Shopping Center Visitor Numbers Based On Various Environmental Indicators

Cagatay Ozdemir, Sezi Çevik Onar and Selami Bagriyanik

March 15, 2020

Estimating Shopping Center Visitor Numbers Based on Various Environmental Indicators

Cagatay Ozdemir[1], Sezi Cevik Onar[2] Selami Bagriyanik[3]

[1] Digital Business Solutions, Turkcell, İstanbul, Turkey
cagatay.ozdemir@turkcell.com.tr
ozdemircag@itu.edu.tr
[2] Industrial Engineering Department, ITU, İstanbul, Turkey
cevikse@itu.edu.tr
[3] Turkcell Technology, Turkcell, İstanbul, Turkey
selami.bagriyanik@turkcell.com.tr

Abstract. The value of data as a result of the rapid increase of data production is gaining importance in recent years both in Turkey and globally. As data gains importance, data mining also changes and develops. With the help of data mining, companies have started to determine their customer management strategies based on data models. The literature review in this field shows that many data models have been studied in the field of customer management. When a more detailed literature review is made, it is observed that the number of sources where demand estimation and location analysis applied together with the machine learning algorithms is very low. When the studies are analyzed on the basis of the sector, it is observed that the studies made for shopping centers are scarce. Within the scope of this study, a new model has been developed by combining location analysis and demand forecasting models that will estimate the number of customers for shopping malls in order to overcome this deficiency in the literature. This model was strengthened with estimation algorithms and tested to generalize this model to all shopping malls. In this study conducted through a large-scale technology and communications services provider company, it was found that environmental factors such as temperature, precipitation market variables and traffic density had a significant effect on the number of customers going to shopping centers.

Keywords: Shopping Mall, Customer Strategy, Predictive Analysis, Machine Learning, Location Analysis, Demand Model, Regression

1 Introduction

One of the important economic indicators for countries is the number of shopping centers. With the increase in the number of shopping centers, competition is increasing and as a result of this, shopping centers have started to differentiate their services and campaings in order to stand out from the competition [1]. The increasing number of shopping malls causes many shopping malls to be unable to reach the desired number of

customers and this situation constitutes a major obstacle for the shopping malls to reach the expected revenues [2]. Since the beginning of the investment, the main target of the shopping centers is to establish the shopping mall in the right location, to attract customers to this location and to reach expected revenue [3] & [4]. The most important issue in order to reach this expected revenue is to estimate the number of customers and attracting the customer to the location of shopping malls [5]. One of the most important strategies to stand out from the competition in shopping malls and to generate the expected revenue is the customer management strategy [6]. In order to achieve customer management, companies have developed many strategies. Identifying the target audience, determining the needs of these audience and reaching out these audience through the right communication channels are the most critical issues for shopping centers [7]. As a result of the digital transformation, the increase in data diversity and volume has led to customer data management strategies being data-driven.

In recent years, with the increasing importance of data in all sectors, the value of the data has started to increase for the shopping mall sector, as well [6] & [8]. Shopping centers have started to use this data, which is produced and stored, in order to estimate the number of customers and expected revenue for following days [9]. When the literature study was carried out specifically for shopping centers, it was found that there were few publications on demand forecast of customer numbers and there were almost no publications where demand forecast of customer numbers was combined with location analysis. With this study, the use of customer numbers estimation, which is a critical issue for shopping centers and which is lacking in the literature, is used together with location analysis.The location analysis was carried out together with the large-scale technology and communications services provider company. The number of customers on a daily basis for the selected shopping centers was reached through the signaling data at the locations [10].

2 Background

Previous studies for shopping centers have generally been customer-based. Therefore, there is a need for a study that includes the estimation of the number of customers on a daily basis and examines the environmental factors affecting the number of customers. In order to perform this study, it is necessary to evaluate customer demand forecast analysis and location analysis together. As of 2010, the total number of studies including location analysis and customer demand forecast analysis for all sectors in the literature is 169. By narrowing this analysis further, the number of studies carried out for all sectors where location analysis, customer demand forecast analysis and machine learning algorithm are used together, decreases to 13. When we review these 13 studies on a sectoral basis, the number of resources for retail sales decreases to 6 and the number of resources in shopping centers decreases to 1. Table 1 and 2 shows the literature reviews details below. When we examine this research which uses location analysis, customer demand forecast and machine learning algorithm for shopping malls, customer faces are examined through camera images and their behavior is estimated [11].With this estimation, marketing actions are planned according to the customer's

next expected behavior. Since the number of resources in this area is limited and it cannot be determined what affects the estimation of the number of customers per day for shopping centers, and implementation studies have been started in this area.

This study has two main objectives. The first objective is to estimate the number of customers for the selected location by combining location analysis, customer demand forecasting analysis and machine learning algorithm that have not been studied thoroughly in the literature. The second objective is to contribute to the strategic plans and marketing actions of the shopping centers with the estimation of the number of customers and analysis of customer behavior according to the environmental factors. In this study, machine learning algorithms which can be used in customer demand forecasting model were tested. However, the data consists of daily customer and environmental information for the selected shopping center and location-based data sets are available instead of person-based data. In addition to these, all data types available are numerical. Considering all these limitations, since the classification structure cannot be established, artificial neural networks, support vector machines, naïve bayes and decision tree learning techniques can not be applied in this study. In the light of this information and the literature review, it was found that the most appropriate machine learning technique for the application was regression analysis.

Table 1. Customer Management Strategy Literature Review (Method Breakdown)

	Total Year	**2010-2019**		**2010-2019**		
Customer Strategy Models	**Total**	**Total**	**Machine Learning (ML)**	**Total**	**Machine Learning (ML)**	**ML + Regression**
Location Analysis	37.200	18.900	1.720	17.400	1.380	509
Demand Analysis	90.400	22.500	1.860	19.000	1.620	878
Location + Demand Analysis	313	243	19	**169**	**13**	**8**

Table 2. Customer Management Strategy Literature Review (Sectors Breakdown)

	2010-2019						
Customer Strategy Models with ML	**Retail**	**Banking**	**Telco**	**Mall**	**Production**	**Energy**	**Health**
Location Analysis	203	190	104	28	442	503	571
Demand Analysis	254	225	92	21	693	779	556
Location + Demand Analysis	6	1	1	1	6	7	5

3 Case Study

From the past to the present, people have been able to meet their shopping needs from the surrounding shops. The main purpose of the established shopping centers is to attract new customers and to estimate the number of customers. In this way, the customer-specific strategic plans to increase the revenue from the customer can be developed. In line with the main scope of the application, the number of individual customers on a daily basis will be determined based on the signaling at the shopping center location. With the obtained environmental data, the number of individual customers in the past period will be analyzed and the number of customers going to the selected shopping

center will be estimated when environmental conditions change in the following days. In this way, variables affecting the number of customers going to the shopping center, environmental factors making a significant impact and how the behaviors of the customers change will be revealed.

3.1 Data Sources

One of the biggest problems in the big data analysis is the fact that most of the data is unstructured and the available data is not clean, processable and analyzable. Although it was challenging to get the data within the scope of this study, the data was collected and made ready for analysis with data cleaning methods. Within the scope of this application, the number of individual customers on a daily basis for selected shopping malls was achieved by leveraging the analysis ability of the large-scale technology and communications services provider company's signaling data. In addition, the following data types have been accessed, cleaned and corrected for the analysis. Table 3 shows the variable categories and data types in detail.

Table 3. Variable List and Details

Category	Category of Variables	Variables	Metric	Data Type
1	Total Number of Customer	1	Numerical	Ouput
2	Daily Temperature	2	Numerical	Input
3	Avg. Daily Temperature	1	Numerical	Input
4	Daily Amount of Rainfall	1	Numerical	Input
5	Financial Market Data	5	Numerical	Input
6	Hourly Average Traffic Density	8	Numerical	Input
7	Daily Average Traffic Density	1	Numerical	Input
8	Internet Trend	4	Numerical	Input

3.2 Data Models for Predictive Analysis

In addition to the number of customers of the selected shopping center data in the past period, the environmental and financial data collected and the data set was enriched. By using machine learning algorithm on this data set, it was requested to estimate the number of customers or any output variable to be selected for the next day. Linear regression analysis was used as the machine learning algorithm since it was the most suitable model for our data set. The following formulas explain the multiple regression analysis and cost function (1)-(2).

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \in \quad (1)$$

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(\mathrm{h}\left(x^{(i)}\right) - y^{(i)}\right)^2 \quad (2)$$

This study is a predictive analysis since the next day of the number of customers are estimated. For this machine learning algorithm, 80% of the data set has been used for

training the model and 20% of the data has been used for testing the performance of the model. This model does not include only one regression analysis. Multiple regression models are generated to find the model that best predicts the number of customers or the selected output variable. Since each regression model contains different combinations of variables, it is aimed to find out which variable has the most significant effect by analysis. Therefore, the variables were chosen to be appropriate for the regression analysis. Studies involving this machine learning algorithm were performed on Python application. In this context, 7 regression analyzes were carried out using random values and the success rates were tested after the model setup. Adj R^2, Success Rate, VIF, Durbin- Watson (Auto Correlation) are used to evaluate the efficiency of prediction model whose definition are described as following formula (3)-(5).

$$R^2_{adj} = 1 - \left[\frac{(1-R^2)(n-1)}{n-k-1}\right] \qquad R^2 = 1 - \frac{SS_{RES}}{SS_{Tot}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2} \tag{3}$$

$$Success\ Rate = \text{Avg.}\left(\frac{1-|Prediction-Actual|}{\text{Actual}}\right) \tag{4}$$

$$VIF = \frac{1}{1-R_i^2} \qquad DW = \frac{\sum_{t=2}^{T}((e_t - e_{t-1})^2)}{\sum_{t=1}^{T} e_t^2} \tag{5}$$

Table 4 shows the model details below.

Table 4. KPI Comparision of Regression Models

Regression Models Name	Selected Variables	Adj R^2	Success Rate	Multi Collinearity	Auto Correlation
Model 1	Shopping Center Internet Trend, Daily Night Temperature Monthly CPI Change and Daily Average Traffic Density	95,70%	98,10%	No	Positive
Model 2	Shopping Center Internet Trend, Daily Night Temperature, Monthly CPI Change, Daily Average Traffic Density and E-Commerce Internet Trend	95,70%	97,90%	No	Positive
Model 3	Shopping Center Internet Trend, Daily Daily Gram Gold / TL Parity, Daily Sunrise Temperature, Daily Night Temperature and Daily BIST Parity	95,40%	90,50%	Yes	Positive
Model 4	Shopping Center Internet Trend, Daily Sunrise Temperature, Daily Amount of Rainfall and Daily Average Traffic Density	95,60%	77,50%	No	Positive
Model 5	Shopping Center Internet Trend, Daily Sunrise Temperature, Daily BIST Parity and Monthly CPI Change	95,60%	96,50%	Yes	Positive
Model 6	Shopping Center Internet Trend, Daily Sunrise Temperature, Daily BIST Parity and Daily Gram Gold / TL Parity	95,40%	97,30%	Yes	Positive
Model 7	Daily Average Temperature, Daily BIST Parity and Daily Average Traffic Density	95%	91,50%	Yes	Positive

3.3 Data Models Evaluation

Seven linear regression models prepared within the scope of the predictive analysis show us above, the VIF ratio is very high if at least two of the financial data and at least two of the temperature data are included in the same analysis. This means that the linearity between variables is high. As a result of this analysis, it was found that the most

appropriate method was to select one from the financial data and one from the temperature data. Under these circumstances, the first and second regression analyzes were the most effective and successful models in terms of test of statistical assumptions, R-adjusted, F Test and success rate. The first and second regression models show that the error value is at minimum and acceptable level when the variables normalized. According to this result, the best model variables (model 1) to estimate the number of customers are Shopping Mall Internet Trend, Daily Night Temperature, Monthly CPI (Customer Price Index) Change and Daily Average Traffic Density.

Strengthening Data Models.

It was aimed to test the general spread and validity of the first and second regression analysis based on machine learning algorithm established for only one shopping center. For this purpose, these models were applied in new two shopping centers after the first shopping center. For this analysis, firstly the internet data for the second and third shopping centers and the number of customers of the previous period were collected. Other outsourcing data were taken from the data set at the first shopping center. Regression assumption tests for these analyzes were performed for these two new shopping mall data. When the first and second regression analyzes of the second and third shopping centers were examined, it was observed that the model was found to be significant by F test and was highly explanable with the high R-adjusted. Furthermore, when Durbin-watson and VIF values were examined in the table below, these values were found to be suitable for regression analysis. In the light of these results, when the success rates are considered, it is observed that both models achieve at least 90% success rate for both shopping malls. This rate represents a high success. In addition, when the error values are examined, it is observed that the error values are at minimum and acceptable levels. However, the success rate of the two models in the first shopping center is higher than the other two shopping centers and the error values are lower than the other two shopping centers. This situation stems from the fact that the model was established according to the first shopping center and was specially designed for the first shopping center. However, both the success of the regression tests and the high estimation rates show us that the first and second regression models can be widely spread. When the first and second regression models are examined, it is seen that the first regression model predicts better than the second regression model. Table 6 shows the strengthening data model details below. MSE and F-Test definition are described as following formula (6).

Table 6. Strengthening Details of First and Second Regression Model

	Model 1			Model 2		
KPI	**First Shopping Mall**	**Second Shopping Mall**	**Third Shopping Mall**	**First Shopping Mall**	**Second Shopping Mall**	**Third Shopping Mall**
Success Rate	98,10%	95,20%	93,70%	97,90%	94,90%	93,60%
Adj R2	95,70%	96,60%	97,10%	95,70%	96,70%	97,10%
MSE	0,82%	1,46%	1,63%	0,83%	1,45%	1,64%
F Test	Significant	Significant	Significant	Significant	Significant	Significant
Multi Collinearity	No	No	No	No	No	No
Auto Colleration	Positive	Positive	Positive	Positive	Positive	Positive

$$F = \frac{(SS_{Total} - SS_{Res})/(k-1)}{(SS_{Res})/(n-k)} \sim F_{k-1,n-k} \qquad MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{6}$$

4 Discussion

This study will benefit in the accurate estimation of the number of customers in the future, which is one of the most important problems in shopping centers. In this way, shopping centers will have the opportunity to reach their expected revenue by planning their operations and strategies correctly. The use of location analysis in the estimation of the number of customers for shopping malls which is lacking in the literature was applied together with this study and contributed to the literature. In addition to these, with this study, how the number of customers coming to the shopping center changes according to the environmental climate, financial and internet trends and the estimation of the next day based on these variables are examined. According to the results of the analysis, it was found that the variables mentioned in the case study section had a significant effect on the number of customers. Furthermore, the model was found to be successful when applied to other shopping centers. This indicates the potential for the generalizability of the model. This analysis was performed on a 217-day data set and the results were based on these data. However, the validity of this analysis may be reduced if the climate changes substantially, financial variables change in an unusual situation and change upwards or downwards dramatically, and if the internet trend is collected in a different format. For that reason, new variables, data sets and algorithms may be needed to strengthen these estimation models in the subsequent analyzes.

5 Conclusion and Future Research

In this study, the model and variables that should be used in estimating the number of customers for the next day are examined with the help of the combined demand forecasting model and location model of customer management strategies which are strengthened with a machine learning algorithm. Regression analysis was found to be the most appropriate machine learning technique for the model. Within the scope of the analysis studies, the first and second regression models mentioned above are the best models to estimate the number of customers. Based on these results, the best environmental variables predicting the number of customers for the next day are Shopping Center Internet Trend, Daily Night Temperature, Monthly CPI Change, Daily Average Traffic Density. Based on these results, shopping centers can estimate the number of customers for the next day with a high success rate by looking at environmental variables. In order to put these environmental variables into their models, they need to follow the weather information of Istanbul and its environment, the traffic density information, daily / weekly / monthly change information of the finance market, daily change information of the internet trend of the shopping mall and e-commerce companies which is the competitor of the shopping mall follow the daily trend information of the internet trend. Shopping centers will be able to provide both cost optimization and marketing strategies by planning the next day with these models. Shopping centers will be able to

plan the number of security guards by estimating the number of customers arriving, will be able to draw up the plans of the number of stores and advisors in the shopping center and increase the number of employees when reaching the high customer number. In addition, Shopping centers will be able to plan and optimize energy costs and other costs according to these numbers. From the point of view of marketing strategies, it will be possible to create a good experience for new customers by estimating the number of future people by making special marketing studies for them, and also to ensure that customers remain loyal with existing marketing campaigns. Shopping centers will also have the chance to change their strategic plans according to market situation variability or traffic density variability.

In the future studies in order to strengthen this study, the number of customers in all shopping centers in the selected province will be examined, the scope of the analysis will be deepened and the data used will be enriched. At the same time, other potential machine learning algorithms will be applied and the success rate will be increased.

References

1. J. Chebat, M. Sirgy and S. Grzeskowiak, "How can shopping mall managment best capture mall image," Journal of Business Research, pp. 735-740, 2010.
2. A. Şekerkaya and E. Cengiz, "Kadın Tüketicilerin Alışveriş Merkezi Tercihlerinin Belirlenmesi Ve Bir Pilot Araştırması," Öneri , pp. 41-55, 2010.
3. K. E. Hedhli, J.-C. Chebat and M. J. Sirgy, "Shopping well-being at the mall: Construct, antecedents, and consequences," Journal of Business Research, p. 856–863, 2013.
4. E. W. L. Cheng, H. Li and L. Yu, "The analytic network process (ANP) approach to location selection: A shopping mall illustration," Construction Innovation, pp. 83-97, 2005.
5. S. H. Zolfani, M. H. Aghdaie, A. Derakhti, E. K. Zavadskas and M. H. M. Varzandeh, "Decision making on business issues with foresight perspective; an application of new hybrid MCDM model in shopping mall locating," Expert Systems with Applications, p. 7111–7121, 2013.
6. L. Wang, H. Fan and Y. Wang, "Sustainability Analysis and Market Demand Estimation in the Retail Industry through a Convolutional Neural Network," MDPI Sustainability, pp. 1-19, 2018.
7. C. A. Gauvreau, D. Kairy, B. Mazer, A. Guindon and G. L. Dorze, "Rehabilitation strategies enhancing participation in shopping malls for persons living with a disability," Disability and Rehabilitation, p. 917–925, 2018.
8. U. Fayyad, G. Piatetsky-Shapiro and S. P., "From Data Mining to Knowledge Discovery in Databases," AI Magazine , pp. 37-54, 1996.
9. S. Wesley, M. LeHew and A. G. Woodside, "Consumer decision-making styles and mall shopping behavior: Building theory using exploratory data analysis and the comparative method," Journal of Business Research, p. 535–548, 2006.
10. H. Eiselt and V. Marianov, Applications of Location Analysis, London: Springer, 2015.
11. V. Huotari, "Depth camera based customer behaviour analysis for retail," Master's thesis. Oulu: University of Oulu, 2015.