Branch Efficiency and Location Forecasting: Application of Ziraat Bank

İlker Met[1], Güven Tunalı[2], Ayfer Erkoç[3], Sinan Tanrıkulu[4] and M. Özgür Dolgun[5]

Abstract

İt is important to if you can't measure you can't manage it. With this respect to get qualified information we need a method to big data. The size of Bank and operational data make classical productivity measurement methods impractical. For this reason, the productivity is measured by using data mining approaches. In this project; an analytical solution that enables efficient and productive use of centralized management and sources as well as the automation of location based reporting has been established in order to provide support for branching strategies.

1   Introduction

In recent years through the development of information systems and technology; public institutions and organizations, businesses and other organizations have been collecting various types of data depending on their nature and purposes. However, these data are stored in databases as a meaningless data stack unless they are processed [1]. The companies should analyze the collected data effectively in order to determine strategic road maps compatible with business processes.

---

[1]Dr.,Group Director in Ziraat Bank Enterprise Architecture Group Directorate.
[2]Service Network Manager in Ziraat Bank Enterprise Architecture Group Directorate.
[3]Service Network Officer in Ziraat Bank Enterprise Architecture Group Directorate.
[4]Service Network Officer in Ziraat Bank Enterprise Architecture Group Directorate.
[5]Dr.,Manager in Datamind Company, Lectures at Hacettepe University.

According to the prevailing view, the better future is mapped, the easier it is to apply the road maps and achieve the goal [2].The development of appropriate software and the motivation of companies to convert the collected data into knowledge have made it necessary to reveal available and interesting relationships, associations and patterns within the data by processing the collected data. Today, many organizations have not started to process their data with methods that will obtain advantageous and useful information regarding customer qualities and customers' buying patterns. Being successful and sustaining this success in the competitive market for the institutions which are rich in terms of raw data however poor in terms of qualified knowledge. Data mining is the assistant of all institutions that have begun to see the importance of data collection as a concept, and that cannot provide the highest level of benefit from past inquiries [3],[4] .

Each system has its own goals. These goals are often expressed by performance indicators such as high productivity, efficiency, profit maximization, cost minimization, service satisfaction, growth, and reputation [5]. Competition experienced in the banking sector and the fact that the banking sector undertakes the financial intermediation function, which is different from other economic sectors, determine the distribution of resources, and force the banks to use their resources in the most efficient way [6]. Efficiency and productivity analyses are crucial management tools in determining the extent to which inputs are used in the process of obtaining the desired output of the banks [7]. Along with the increasing competition conditions, measurement of branch productivity in the Banking sector becomes prominent. In the banking sector, which has a dynamic and fragile structure, monitoring and forecasting of branch efficiency has become one of the primary parameters due to the impact of digitalization emerging and the increasing influence of location selection on profitability in recent years.

"If you can't measure it, you can't manage it". This statement also explains why digital data explosion is so important in recent times. To put it all in simple terms, managers can measure efficiency owing to big data, so they can make more decisions about business problems and thereby improve performance. Decisions made based on data are more accurate than classical approach. The use of big data makes it possible for managers to make evidence-based decisions rather than heuristic decisions [8].

In the Turkish banking system, a significant portion of banking services and products are offered through branch channels. Despite the rapid development of alternative distribution channels, branches are the most actively used channel due to the reasons such as use of internet is still not widespread (Internet usage rate is 91.61% in the UK, 87.36% in the US and 51.1% in Turkey), cash utilization ratio is relatively high, rate of financial literacy is low compared to rate of developed countries, and the legal regulations. In Turkey, there are 19.8 banks per 100.000 persons while this rate is 25.2 in the UK and 32.3 in the US.

For the reasons mentioned above, branch distribution network optimization in Turkey has a critical proposition in terms of profitability and efficiency. Ziraat Bank is the bank that needs the most optimization in this regard because it is the

commercial bank with the largest branch network in Turkey (1.789 branches-2016) and the largest domestic ATM network (6.869 ATMs-2016).

Along with the increase in data size and the discovery of new methods, alternative methods have been developed used to measure productivity. In this study, some data mining methods which are among alternative methods are used to measure productivity. Various geographical and economic data taken from TURKSTAT and the Bank data are combined within the Bank and integrated into the map information system and used as decision support system in branching strategies. Regression analysis, artificial neural networks, decision trees and clustering methods are used for this purpose.

Data mining, CRISP-DM process, clustering methods, classification and regression methods are mentioned in Section 2. Data mining processes in efficiency applications, ratio analysis, parametric and nonparametric methods are mentioned in Section 3. Application, value-based segmentation, potential value-based segmentation / predictive method, profiling phase and generating efficiency scores are mentioned in Section 4. Conclusion and discussion are mentioned in Section 5.

2   Data Mining

Data mining is the process of discovering new relationships, patterns, and trends by examining large data sets with the help of statistical, mathematical or pattern identification techniques [9]

This section is focus on CRISP-DM (Cross-Industry Standard Process for Data Mining), a methodology used to solve data mining problems. Besides, the classification models and clustering methods used in the data mining is briefly explained in this section.

2.1   CRISP-DM Process

There is a need for a standard methodology in data mining because of the difficulties in working with different and noisy data sources due to data mining structure that hosts many disciplines and the diversity of procedures with tasks in different application areas and data sizes. Answering the following steps before the project will be useful for the success and planning of the project [1].
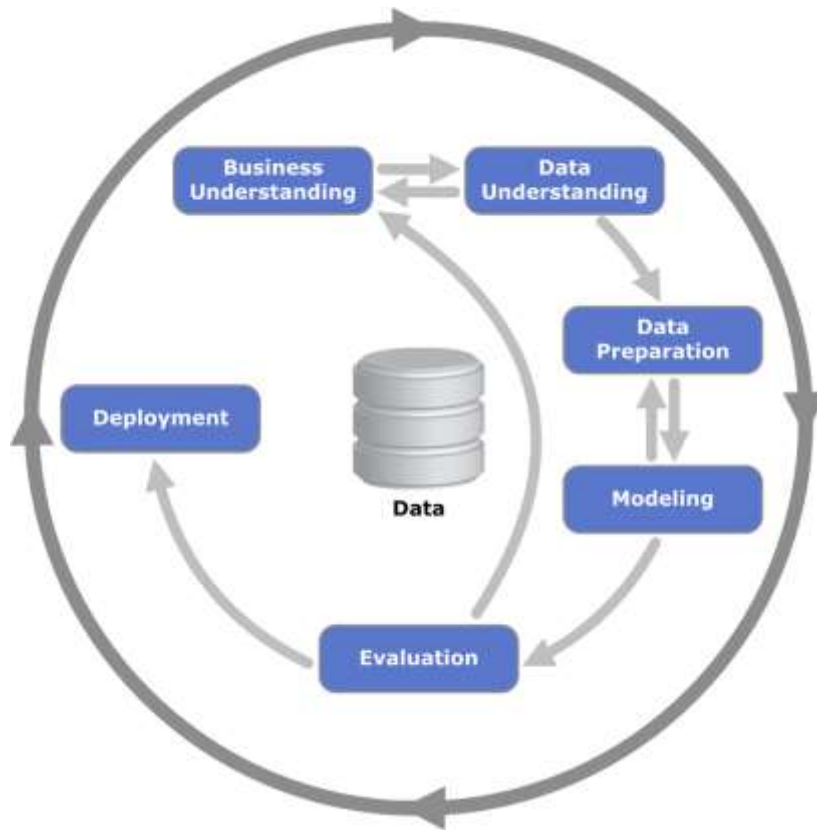
Figure.1 Crisp-Dm Methodology

In the study, CRISP-DM methodology is used, respectively;

1. Business Understanding: It is aimed to make profitability and productivity estimations for the new branches to be opened by measuring the value and potential values of the branches as Business understanding.

2. Data Understanding: Data sources, variables, data types that are thought to be used in the modeling phase are examined at this stage.

3. Data Preparation: If data included the outliers or not was examined through graphical analysis, Anomaly Detection algorithm; point of interest (POI) and variables considered to be important were added to the created project DataMart during data analysis phase.

4. Modeling: In this phase, the model that best expresses the branch productivity is chosen by using statistical evaluation criteria.

5. Evaluation: It was evaluated how much (%) of the generated model predicted the profitability and productivity of the branch.

6. Deployment: Analytical models and results are integrated into the map information system and transformed into a reporting and analytical tool that could be benefited from in branching strategies.

2.2 Clustering Methods

Customer segmentation is one of the popular application of data mining with established customers. The purpose of segmentation is to fit products, services, and marketing messages to each segment [10]. Clustering is the task of segmenting a heterogeneous population into a number of more homogeneous subgroups or clusters. The records are grouped together on the basis of self-similarity. It is up to the user to determine what meaning, if any, to attach to the resulting clusters [1].

2.3. Classification and Regression Methods

In classification and regression methods, analyses are performed on data sets containing a target (dependent) variable and independent (predictive) variables. The goal of these methods is to construct a meaningful model in which the value of the dependent variable is the output of the model and in which the predictor variables are included as inputs. If the dependent variable is numeric, the problem is called regression, and if it is not numerical, it is called classification problem [1].

3 Data Mining Processes in Efficiency Applications

Efficiency is a measure of the extent to which resources are used economically and effectively.

Concepts that first come to mind in terms of productivity in banking may be listed as: Product diversity, scale and scope economies, ownership and market structure, mergers, acquisitions, entry of foreign banks into the market, privatization, environmental economic conditions, competitiveness, technological development, centralization of operations, restructuring of business processes, asset quality, capital adequacy, supervision effectiveness, transparency, alternative distribution channels, income/expense balance[11].

The basic approaches to efficiency measurement in the banking system are the production and the mediation approaches. In the production approach, outputs are measured based on the number of accounts and production costs include transaction costs. In the mediation approach, the output is measured based on currency and production costs include transaction and interest costs [12].

Three different techniques are used for efficiency measurement. These techniques are ratio analysis, parametric methods and nonparametric methods. The summary information on these techniques is given below.

3.1. Classification and Regression Methods

Ratio analysis is a method of measuring efficiency, which proportions a single output value to a single input value [13].

In general terms, it is a form of analysis that is used to determine the relationship

between financial variables and their own variables, which represents the ratio or percentage of the numerical relationship between numbers.

Ratio analysis is used more frequently than parametric methods and non-parametric methods. In this method, it is aimed to monitor the ratio generated by proportions of interested variables over the time. Although ratio analysis is preferred in applications because of its ease of use, it can be difficult to use when the number of variables increases and to understand the efficiency of the bank or branch by looking at a single ration.

3.2. Parametric Methods

Parametric methods that are used in the hypothesis testing of the parameters of the main mass, which are generally normal distribution, using the direct measurement values.When the relationship between variables is known, the values of the relevant variables may reach the optimum level if the influencing factors can be controlled as other variables can be estimated by looking at the value of a variable. In many applications of statistical estimation, the relationship between variables must be modeled using sample data to evaluate the relationship between groups of variables. By means of the model thus obtained, any future value of the variable selected as the dependent variable is estimated [14].

Assuming that the parametric methods are generally a regression equation that best prediction of the data set, observations that do not deviate from this regression equation are defined as efficient, while other observations are defined as inefficient. This method always assumes a random error.

3.3. Nonparametric Methods

Nonparametric methods try to measure the distance to the activity limit by using linear programming algorithms. These methods have some advantages over other methods. A clear advantage of a nonparametric test is that it can be reliably used when nothing is known about the main mass. For example, the sample volume is so small that the sampling distribution of statistics does not approach normal distribution. In this case a nonparametric technique is needed. Among these methods, the most common use is technical data envelop analysis [15].

4   Data Mining Processes in Efficiency Applications

In this study, the data obtained from branches of Ziraat Bank for the year 2015 are used. The data used are collected in two main titles, financial and non-financial data. The financial data of the branches (profit, number of customers, etc.) are kept daily in the banking software and the year-end figures are taken as basis. These data are derived from the banking software database with SQL queries. Non-financial data is provided internally (number of employees, branch age, etc.). Banking software is stored in excel lists which are not stored in the database. Non-bank non-financial data (population, POI, etc.) was obtained from the

relevant official institutions. Data mining programs have been used to match the supplied non-bank data to the branches.Information on the variables used is defined in Table 1.

Table 1: Crisp-Dm Methodology

Variables	Explanation	Measurement Levels
Number of Employees	Branch employee numbers	Continuous
Profit and Revenue Variables	Last 1-year branch net profit and operating revenue	Continuous
Branch Age	Branches over 2 years old	Continuous
The Area of the Branch	Square of meter size of the Branch	Continuous
Number of Customers	Segment-based customer numbers	Continuous
POI	The important locations in the vicinity of 100, 250 and 500 meters	Continuous
Population	ABPRS (Address-Based Population Registration System) data received from TURKSTAT	Continuous
Number of Branches per Population	This is obtained by dividing the population of province to the number of branches.	Continuous
Center / District	The information if the Branch is located city center or in district is utilized. (Center = 1, District = 0)	Categorical
Region of the Branch	The information of the region to which the branch is dependent is used.	Categorical

In the variable selection phase, highly correlated variables are removed from the analysis.

About 40 independent variables are used in modeling phase, 10 of them are shown above.

In the modeling phase, correlation analysis, clustering, decision tree and regression methods are used. Segmentation and profiling phases are given below. In segmentation phases, value-based and potential value-based methods are used. In profiling phase, decision trees methods are used. Mathematical formulas such as transformation formulas are used in generating efficiency scores phases.

4.1 Value-based Segmentation

In the first phase, the branches are divided into 5 groups by their "Activity Revenue". Activity Revenue is the sum of the revenues of the branches from the banking activities K-Means and Two Step clustering methods are used in this study. Clustering methods can be evaluated according to Silhouette criterion which is varying between -1 and +1. It is desirable that this value is higher than 0.5, and it is assumed that this divides the clusters at a good level. Silhouette criteria for two step and k-means clustering are obtained for this data and given in Table 2.

Table2. Two Step and K-Means Clustering Results Comparison

	Two-Step	K-Means
Silhouette criteria	0.6	0.7

At the point of which method would be chosen, both the clustering results and the distributions are considered and the method which is compatible with the business information is decided. Distributions of the results obtained by the K-Means method (percentage of cluster sizes: 2.4%, 5.4%, 12.8%, 35.3% and 44.1%) are considered to be better than Two-Step (percentage of cluster sizes: 6.1%, 12.7%; 22.5%; 26.8% and 31.8%). Besides K-Means method is chosen because the Silhouette measure of k-means method is higher than two-step method. When the clusters are examined, it is found that they are separated from each other, which means they are homogeneous within themselves and heterogeneous among them.
In the continuation of the work, the forecasting estimation has been initiated.

4.2 Potential Value-based Segmentation / Predictive Method

In order to move to the section of potential value-based segmentation, it is necessary to obtain a forecasting model for the "Activity Revenue" variable which is used in the value segmentation and shows the values of the branches.
Forecasting methods can be used because the Activity Revenue is continuous;

- Linear Regression Analysis,
- Decision Trees – CHAID Analysis, (Chi-square Automatic Interaction Detector)
- Decision Trees – C&R Tree Analysis,(Classification and Regression Tree)
- Artificial Neural Networks Analysis Methods.

The above methods are used in this study. Because of the sectoral and analytical know-how, two separate models are developed.
In the evaluation of the results of the regression and decision trees, the R^2 measure is used with the separation of testing and training sub-sets. In the study, 95% of the data is used in the training phase of the model, and 5% of the data is tested for success. The models are used for the estimation of the target variable considered to be used in the potential value-based segmentation study and the results of the successes of these models are given as follows;

- Linear Regression Analysis yields the best result among the models obtained by using POI information only. So, the result of this model is used. The accuracy of this model (R^2 measure) is about 39%.
- CHAID method among the Decision Tree methods yields the best result among the models obtained by using number of customers only; the result of this model is used. The accuracy of this model (R^2 measure) is about 82%.
- The accuracy of meta-model obtained by using the information of both models (R^2 measure) is 83%.

After activity revenue is estimated by the meta-model established for each branch, the clustering phase is initiated.

In the K-means method, the silhouette criterion is obtained as 0.7 and for the two-way clustering method the silhouette criterion is 0.6. As a result, the k-means method is used.

Table 3. Meta-model Two-Step and K-Means Clustering Results Comparison

	Two-Step	**K-Means**
Silhouette criteria	0.7	0.7

At the point of which method would be chosen, both the clustering results and the distributions are considered and the method which is compatible with the business information is decided. It has been decided to choose k-means method by considering the distributions of the results obtained by the K-Means method (percentage of cluster sizes: %2,9; %8,4; %22,1; %30,1 and %36,6) better than Two-Step (percentage of cluster sizes: %2,9; %8,4; %18,1; %27,0 and %43,6) and using business information.

When the clusters are examined, it is found that they are separated from each other that is homogeneous within themselves and heterogeneous among them.

Table 4: Matrix Display of Potential Value and Value Segmentation Results

Value_k-means		Potantial Value_k-means				
		1	2	3	4	5
1	Count	19	11	2	0	0
	Row %	59.375	34.375	6.250	0.000	0.000
	Column %	50.000	9.910	0.685	0.000	0.000
	Total %	1.436	0.831	0.151	0.000	0.000
2	Count	19	43	9	0	0
	Row %	26.761	60.563	12.676	0.000	0.000
	Column %	50.000	38.739	3.082	0.000	0.000
	Total %	1.436	3.250	0.680	0.000	0.000
3	Count	0	55	100	12	3
	Row %	0.000	32.353	58.824	7.059	1.765
	Column %	0.000	49.550	34.247	3.015	0.620
	Total %	0.000	4.157	7.559	0.907	0.227
4	Count	0	2	175	263	27
	Row %	0.000	0.428	37.473	56.317	5.782
	Column %	0.000	1.802	59.932	66.080	5.579
	Total %	0.000	0.151	13.228	19.879	2.041
5	Count	0	0	6	123	454
	Row %	0.000	0.000	1.029	21.098	77.873
	Column %	0.000	0.000	2.055	30.905	93.802
	Total %	0.000	0.000	0.454	9.297	34.316

In the matrix given with Table 4 represent value and potential value results. And the number 1 refers to the best and the number 5 refers to the worst segment. In this context:

- It has been determined that potential value-based segments of 19 branches whose value-based segment is 1st segment are also 1st segment.
- It has been determined that potential value-based segments of 2 branches whose value-based segment is 1st segment are also 3rd segment. It has been determined that the mentioned 2 branches have more valuable than their potentials.
- It has been determined that potential value-based segments of 3 branches whose value-based segment is 3rd segment are also 5th segment. It has been determined that the mentioned 3 branches have more valuable than their potentials.
- It has been determined that potential value-based segments of 2 branches whose value-based segment is 4th segment are also 2nd segment. It has been determined that the mentioned 2 branches have more valuable than their potentials.

4.3 Profiling Phase

Profiling phase has been done in order to provide a basis for estimated branch efficiencies and to facilitate evaluation. A total of 46 profiles related to the value segments of the branches have been created by using the center / district location, customer numbers and population per branch and POI variables through CHAID and C & R Tree algorithms. Some of the profiles created are as follows:

1. If the branches are located in city center and having more than 503 active entrepreneur customers then their value-based segments are ranked in the 1st and 2nd cluster.
2. If the branches are located in city center and having active entrepreneur customers between 87 and 503 then their value-based segments are ranked in the 3rd cluster.
3. If the branches having active entrepreneur customers lower than 87 then their value-based segments are ranked 4th and 5th cluster.
4. If the branches located in districts and having more than 657.762 per branch population and having more than 503 active entrepreneur customers then their value-based segments are ranked in the 2nd cluster.
5. If the branches located in districts and having more than 657.762 per branch population and having active entrepreneur customers lower than 503 then their value-based segments are ranked the 4th and 5th cluster.
6. If the branches located in city center and having more than 503 active entrepreneur customers and having more than 26 educational institutions then their value-based segments are ranked in the 1st cluster.

7. If the branches located in the districts and having more than 657.762 per branch population and having more than 24 financial institutions then their value-based segments are ranked in 2nd and 3rd cluster.

4.4 Generating Efficiency Scores

In the last part of the study, 13 efficiency scores for per branch and efficiency scores for per employee were established using financial variables in order to form the basis for branch strategies. Efficiency scores such as Efficiency-7 and Efficiency-11 are explained below.

Efficiency 7: All profit, income and sub-variables belonging to the customer are divided into the current employee numbers. Obtained scores are converted to 0-100 scale. Then the final score is calculated by using arithmetic average of these variables.

Efficiency 11: All profit, income and customer variables converted to 0-100 scale. Then the final score is calculated by using arithmetic average of these variables.

In the modeling stage, regression analysis and decision tree methods are used in order to forecast the branch efficiency scores by means of POI variables.

5 Conclusion

Efficiency is a measure on effective and efficient use of resources and a production-oriented phenomenon. Achieving effective levels of productivity for the businesses is a key success factor. For this reason, it is now an important goal to improve the present performance and to understand why the organization is ineffective. However, there are many different constraints that affect the success such as markets, financial resources, and economic indicators. For an effective productivity, banks need to optimally assess their scarce resources. In this study, the efficiency of Ziraat Bank branches has been evaluated by using the data mining methods and attempts have been made to measure the efficiency of the locations projected to open branch offices.

The results produced through the model have been integrated into the Ziraat GIS Application used by the bank. GIS distribution network developed by our bank and receiving Steve Awards is our most important decision support system developed used to ensure optimization. Ziraat GIS Application is a decision support system that can perform analysis and reporting by blending intra-bank and out-of-bank data on a map basis. The system uses important point information based on population, number of dwellings, development levels, education levels, number of rival bank branches and categories from out-of-bank sources as well as involving all detailed data sets belonging to the customers and branches within the bank. The model results integrated into the application form the basis of location based reporting.

After the study, the model results have been started to be used as a decision

support system for the annual branching plan prepared every year. Branch opening requests prioritized within the framework of requests from the field in the previous years have started to be evaluated according to model results along with the study. Thus, location assessment criteria have been standardized. Our analytical model was accepted at the 9th International Statistical Congress held in Antalya in 2015 and presented at the congress.

The model is updated every year and the results are re-evaluated. In case of need, inputs and outputs can be changed according to bank policies.

References

[1] Dolgun, M. Ö., Success of Data Mining Classification Methods; Comparison of Dependent Variable Prevelance, Sample Size and Independent Variables According to Relation Structure, Doctorate Bachelor Thesis, Hacettepe University Institute of Health ,Ankara,(2014).

[2] Met, İ., And Baktır, M. Ö. (2016), Probability of Probability and New Approaches in Strategic Management, Harward Business Review Turkey **p.97,** (2016).

[3] Han, J., Kamber, M., and Pei, J. ,Data Mining: Concepts and Techniques, The Morgan Kaufmann, **Third Edition**, 2011.

[4] Larose, D.,T., Data Mining Methods and Models, Wiley-Interscience, New Jersey, **322p**, 2006.

[5] Barutçugil, İ. (2002), İnformation Management. Kariyer Publishing, İstanbul, **p.13**, 2002.

[6] Atan, M., and Çatalban, G. K. ,Effectiveness of Banks' Efficiency and Capital Structure in Banking, (2005), **DOI: 10.3848/iif.2005.237.2776.**

[7] Bozdağ, E. G., Altan, M. S., and Bozdağ, A. E., Efficiency and Efficiency in the Banking System "An Application with Data Envelopment Analysis" Aksaray University Journal of İİBF, **Vol. 2, No:1,**(2010).

[8] McAfee, A., and Brynjolfsson, E. ,Big Data :The Management Revolution https://hbr.org/2012/10/big-data-the-management-revolution.(2012).

[9] Larose, D.,T., Discovering Knowledge in Data: An Introduction to Data Mining, Wiley Interscience, NewYork, **222**, 2004.

[10] Berry, M. J. A., and Linoff, G. S. Data Mining Techniques, Wiley Publishing, Inc., Indiana, USA, 2004 .

[11] Türkiye   Bankalar   Birliği   (Turkish   Bank   Assosiciation) https://www.tbb.org.tr/dosyalar/konferans_sunumlari/bankacilikta_verimlilik. pdf   (Date of Access:30.11.2016)

[12] Yıldırımoğlu, Suna Özlem, Comparing Turkish Banks and European Banks in terms of Competitiveness, (2005)

[13] Yeşilyurt C, Alan M. A, Measurement of 2002 Relative Efficiency of Science High School by Data Envelopment Analysis Method, Journal of Economics and Administrative Sciences, **Vol.4, No:2**, (2003), 2, 91-104.

[14] Okur, S., Comparative Analysis of Parametric and Nonparametric Linear Regression Analysis Methods, MS Thesis, (2009)

[15] Yaşa A.,Measuring Effectiveness and Data Envelopment Analysis Method in Banking Sector, Ankara University, Institute of Social Sciences, Graduate Thesis, Ankara, (2008)