<u>*Araştırma Makalesi*</u>   <u>*Research Article*</u>

Financial Risk Estimation with KNN Classification Algorithm on Determined Financial Ratios

Oğuzcan Uludağ[1], Arif Gürsoy[2*]

[1] Ege Üniversitesi, Fen Fakültesi, Matematik Bölümü, İzmir, Türkiye (ORCID: 0000-0003-0516-0014), oguzcanuludag@gmail.com
[2*] Ege Üniversitesi, Fen Fakültesi, Matematik Bölümü, İzmir, Türkiye (ORCID: 0000-0002-0747-9806), arif.gursoy@ege.edu.tr

Abstract

Analyzing the financial situation of companies is very important today. Thanks to early analysis, companies can improve their financial situation and be saved from bankruptcy. Until now, many studies have been carried out on the financial analysis of companies. Data mining techniques are widely used in many different areas of the financial sector. In this study, data mining classification technique has been used to determine the financial risk status of companies. The data set has been created by using the balance sheets taken from the Kamu Aydınlatma Platformu (KAP) between 2013 and 2018 of different companies serving in the manufacturing industry. There are a total of 1027 records in the data set. On these records, the financial ratios determined for that year for each firm have been calculated and recorded in the database. The Springate model has been used to determine the financial success of the companies. Springate formula consists of 4 financial ratios. Springate score value is calculated by multiplying these ratios with certain coefficients. If the calculated score value is less than 0.862, the firm is considered unsuccessful, else the firm is considered successful. In this context, the Springate score has been calculated for each record and the companies have been labeled as successful or unsuccessful according to the score value. In this study, the KNN (K-Nearest Neighbor) algorithm has been used for classification. The KNN algorithm classifies the new element to be classified by looking at its proximity to its k neighbors. KNN algorithm needs learning data to classify unclassified data. In this study, cross validation method has been used for determining learning data and the data to be classified. According to the cross validation method, the data set is divided into n groups. For each group, the selected group creates the data to be classified. All other groups (n-1 groups) are used as learning data. Then the classification process is applied and the results are stored. After the classification process is completed for all groups, the statistical summary of the results is checked. In this study, the dataset was divided into 10 groups and the classification results have been evaluated. When the results have been examined, it has seen that the algorithm has make a successful classification with a rate of 88.42%. Thus, the financial risk estimation of the companies has made with the classification technique applied on the determined finance model.

Keywords: Data Mining, KNN Classification Algorithm, Financial Risk Analysis, Financial Mathematics.

Belirlenen Finansal Oranlar Üzerinde KNN Sınıflandırma Algoritması ile Finansal Risk Tahmini

Öz

Firmaların finansal durumlarının analiz edilmesi günümüzde oldukça önem taşımaktadır. Erken yapılan analizler sayesinde firmalar finans durumlarını düzeltebilir ve iflas etmekten kurtarılabilir. Bu zamana kadar firmaların finansal analizi üzerine birçok çalışma yapılmıştır. Veri madenciliği teknikleri finans sektörünün farklı birçok alanında yaygın olarak kullanılmaktadır. Bu çalışmada firmaların finansal risk durumlarının belirlenmesi için veri madenciliği sınıflandırma tekniği kullanılmıştır. Veri seti, imalat sanayi sektöründe hizmet veren farklı firmaların 2013 ile 2018 yılları arasında Kamu Aydınlatma Platformu (KAP)'ndan alınan bilanço tabloları kullanılarak oluşturulmuştur. Veri setimizde toplam 1027 kayıt bulunmaktadır. Bu kayıtlar üzerinde her bir firma için o yıla ait belirlenen finansal oranlar hesaplanarak veri tabanına kaydedilmiştir. Firmaların finansal başarılarının belirlenmesi için Springate modeli kullanılmıştır. Springate modeli firmaların gelecekteki iflasın erken tahminlenmesi için kullanılan bir modeldir. Springate

---

* Corresponding Author: arif.gursoy@ege.edu.tr

formülü 4 finansal orandan oluşur. Bu oranlar belli katsayılar ile çarpılarak Springate skor değeri hesaplanır. Hesaplanan skor değeri 0.862'den küçük ise firma başarısız büyük ise firma başarılı olarak kabul edilir. Bu kapsamda her bir kayıt için Springate skoru hesaplanmış ve skor değerine göre firmalar başarılı veya başarısız olarak etiketlenmiştir. Bu çalışmada sınıflandırma işlemi için KNN (K-En Yakın Komşu) algoritması kullanılmıştır. KNN algoritması, sınıflandırılacak yeni elemanı belirlenen k tane komşusuyla arasındaki yakınlığa bakarak sınıflandırır. KNN algoritması sınıflandırılmamış verileri sınıflandırmak için öğrenim verisine ihtiyaç duyar. Çalışmada öğrenim ve sınıflandırılacak verilerin belirlenmesi için çapraz sorgulama yöntemi kullanılmıştır. Çapraz sorgulama yöntemine göre veri seti n adet gruba ayrılır. Her bir grup için, seçilen grup sınıflandırılacak veriyi oluşturur. Diğer tüm gruplar (n-1 grup) öğrenim verisi olarak kullanılır. Daha sonra sınıflandırma işlemi uygulanarak ve sonuçlar saklanır. Bütün gruplar için sınıflandırma işlemleri bittikten sonra sonuçların istatistiksel özetine bakılır. Bu çalışmada veri kümesi 10 gruba ayrılarak sınıflandırma sonuçları değerlendirilmiştir. Sonuçlar incelendiğinde algoritmanın %88.42 oranla başarılı sınıflandırma yaptığı görülmüştür. Böylece firmaların finansal risk tahmini, belirlenen finans modeli üzerinde uygulanan sınıflandırma tekniği ile yapılmıştır.

Anahtar Kelimeler: Veri Madenciliği, KNN Sınıflandırma Algoritması, Finansal Risk Analizi, Finansal Matematik.

1. Introduction

Financial distress is the inability of a firm to meet its debts. This distress leads to the company being close to bankruptcy, going into liquidation and even confiscating the company. Therefore estimating financial distress is crucial to the success of businesses [1]. Early detection of financial distress can minimize the effects of financial distress on the business. Additionally, they can reduce the company's risk of depreciation for investors and Lenders [2].

There has been a lot of research on financial distress so far. Beaver studied the estimation of financial failure using the discriminant analysis method [3]. Another approach to the prediction of financial failure is Altman's study which used the multiple discriminant analysis method [4]. There are many different approach studies such as the logistic regression used by Ohlson and the artificial neural network models used by Aktaş [5,6].

Data mining is one of the common techniques used in financial analysis of businesses. It will be possible to obtain useful information about the financial failure of companies when data mining techniques are applied to financial ratios of companies [7]. In this study, classification algorithm, a data mining technique, are applied to the determined financial ratios. The Springate model was used to determine the financial status of the companies. Then, the results have been analyzed by applying the KNN algorithm on the unlabeled data.

2. Material and Method

In this study the balance sheet tables taken from Kamu Aydınlatma Platformu (KAP) of different companies operating in the manufactıuring industry sector between 2013 and 2018 have been used as dataset. A total data consists of 1027 records. The Springate score of each record has been calculated to determine the financial status of the companies. Also determined financial ratios of each record has been calculated. Then these data has been merged. The KNN classification algorithm has been run with cross validation. Then the results have been analyzed.

2.1. Springate Model

Gordon L.V. Spingate developed the model uses the multivariate discriminative analysis. There are four basic rations in the Springate formula. According to the S value calculated in the formula, it is interpreted that the firms are successfull or unsuccessful.

S Score = (1.31X1) + (3.07X2) + (0.66X3) + (0.4X4)

X1= Net Working Capital/Total Assets

X2= Profit before Interest and Tax/Total Assets

X3= Profit before Tax/Short Term Debts

X4= Net Sales/Total Assets

According to calculated S value, If S < 0.862 it is predicted that the company will go bankrupt [8].

2.2. Financial Ratios Used

It is appropriate to use financial ratios in the analysis of a firm's financial position [9]. In this study, a total of 10 ratios that affect the financial status of companies have been used. The rations are shown in Table 1.

X1 measures the efficiency of the business in its activities. X2 shows that the company's debt load related to current profit. X3 gives the share of the net working capital required to continue the acitivites of the companies in the assets. X4 measures the liquidity of companies more precisely [12]. X5 shows the current financial strength of the company [12]. X6 shows the borrowing capacity of the company. The value of X7 is an indicator of how many times a year the firm consumes its stocks in production activities or converts it into sales revenue [13]. X8 shows how efficiently the capital invested in the business is used. X9 shows how many times the company sells its assets [14]. X10 allows us to make judgments about the company's investment and financing policies [13].

Table 1. Financial Ratios

Variable	Calculation
X1	*Earnings Before Interest and Tax / Net Sales*
X2	*Total Debt / Earnings Before Interest and Tax*
X3	*Net Working Capital / Total Assets*
X4	*(Current Assets – Stok) / Short-Term Liabilities*
X5	*Current Assets / Short-Term Liabilities*
X6	*Total Debt / Equity Capital*
X7	*Cost of Sales / Average Inventory*
X8	*Net Profit / Equity Capital*
X9	*Net Sales / Total Assets*
X10	*Net Profit / Net Sales*

2.3. KNN (K Nearest Neighborhood) Classification Algorithm

The purpose of classification is to determine which class the data belong to by looking at the similarities [10]. KNN is nonparametric controlled learning algorithm [11]. It is a widely used classification algorithm in data mining. KNN algorithm classifies the new element to be classified by looking at its proximity to its k neighbors. Euclidean, Manhattan and Minkowski distance functions commonly used for distance calculation. KNN alorithm steps are as follows:

1. Load the training data
2. Determine the value of k
3. For each record in training data
   a. Calculate distances
   b. Get the nearest k neighbors
4. Set label according to class member count
5. End

Figure 1 shows the steps of the KNN algorithm. In step 0, there are the green-painted points of Class A, the points of Class B painted in orange, the points of Class C painted in green, and the unclassified grey point. In step 1, the distance of the gray point to all other points is calculated. In step 2, the k nearest neighbors to the grey point are listed. For k equals 4, the class with the most members in step 3 is A, which is green. Then the grey point is labeled class A.

Figure 1. KNN Steps

3. Results and Discussion

In this study, classification, which is a data mining technique, has been used in the financial situation analysis of companies operating manufacturing industry. Using the balance sheets of 179 companies, the financial ratios of each company used in the study have been calculated. The Springate score of each firm whose financial ratios have been calculated and their financial status has been labeled. Cross validation has been applied to group the dataset. The reason for using cross validation is to reach the most objective and accurate result possible by evaluating the results obtained for different partitionings. The dataset has been divided into 10 groups and the KNN algorithm has been run for each group. The results are as follows:

Table 2. KNN Result

accuracy: 88.42%	*true Fail*	*true Well*
Fail	*404*	*64*
Well	*55*	*504*
	88.02%	*88.73%*

As seen in Table 2, there are a total of 459 companies whose financial situation is fail. 404 of these records have been classified correctly and 55 of them have been incorrectly classified, resulting in an accuracy rate of 88.02%. For 568 companies with good financial status, 504 records have been classified correctly and 64 records incorrectly, an accuracy rate of 88.73% has been obtained. According to general table, 88.42% accuracy has been obtained. Classification accuracy ratio has been used for classification performance measurement.

All operations has been performed using RapidMiner program.

4. Conclusions and Recommendations

In this study, the financial risk estimation of the companies has been made using the classification method. KNN algorithm has been chosen for classification. Looking at the results, it is observed that the algorithm works with 88.42% success.

One of the factors affecting the success rate of the algorithm is the training dataset. In this study, the training dataset has been created by taking an equal number of random from both groups.

The results of this study showed that the classification technique can be applied in the financial situation analysis of companies and successful results will be obtained. In addition, it is clear that the improvements to be made on the model or algorithm will increase the classification success.

5. Acknowledge

References

[1] Ashraf, S., GS Félix, E., & Serrasqueiro, Z. (2019). Do traditional financial distress prediction models predict the early warning signs of financial distress?. *Journal of Risk and Financial Management*, *12*(2), 55.

[2] Selimoğlu, S., & Orhan, A. (2015). Finansal başarısızlığın oran analizi ve diskriminant analizi kullanılarak lçümlenmesi: BİST'de işlem gören dokuma, giyim eşyası ve deri işletmeleri üzerine bir araştırma. *Muhasebe ve Finansman Dergisi*, (66), 21-40.

[3] Beaver, W. H., (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 71-111.

[4] Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589-609.

[5] Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 109-131.

[6] Aktaş, R. (2003). Mali Başarısızlığın Öngörülmesi: İstatistiksel Yöntemler ve Yapay Sinir Ağı Karşılaştırılması. *Ankara Üniversitesi SBF Dergisi*, 58(04).

[7] Özkan, M., & Boran, L. (2014). Veri Madenciliğinin Finansal Kararlarda Kullanımı. *Çankırı Karatekin Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 4(1), 59-82.

[8] Kürklü, E., & Türk, Z, (2017). Financial failure estimate in bist companies with Altman (Z-score) and Springate (S-score) models. *Osmaniye Korkut Ata Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, *1*(1), 1-14.

[9] Altman, E. I. (2000). Predicting Financial Distress Of Companies: Revisiting ZScore and ZETA Models. Stern School of Business, New York University.

[10] Kılınç, D., Borandağ, E., Yücalar, F., Tunalı, V., Şimşek, M., & Özçift, A. (2016). KNN algoritması ve r dili ile metin madenciliği kullanılarak bilimsel makale tasnifi. *Marmara Fen Bilimleri Dergisi*, *28*(3), 89-94.

[11] Dasarathy, B. V. (1991). Nearest neighbor (NN) norms: NN pattern classification techniques. *IEEE Computer Society Tutorial*.

[12] Aktas, M. (2009). İstanbul Menkul Kıymetler Borsasında hisse senedi getirilileri ile ilişkili olan finansal oranların araştırılması. *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, *37*(2), 137-150.

[13] Yurdakul, M., & Yusuf, İ. Ç. (2003). Türk Otomotiv Firmalarının Performans Ölçümü ve Analizine Yönelik Topsıs Yöntemini Kullanan Bir Örnek Çalışma. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, *18*(1), 1-18.

[14] Karadeniz, E., & İskenderoğlu, Ö. (2011). İstanbul Menkul Kıymetler Borsası'nda işlem gören turizm işletmelerinin aktif karlılığını etkileyen değişkenlerin analizi. *Anatolia: Turizm Araştırmaları Dergisi*, *22*(1), 65-75.