Duvar İnşa Edilmesinde Verimliliği Etkileyen Faktörlerin Apriori Veri Madenciliği Yöntemi Kullanılarak Analizi

Abdullah Emre KELEŞ [1], Mümine KAYA [2]

[1] Adana Bilim ve Teknoloji Üniversitesi, İnşaat Mühendisliği Bölümü, Adana

[2] Adana Bilim ve Teknoloji Üniversitesi, Bilgisayar Mühendisliği Bölümü, Adana

aekeles@adanabtu.edu.tr, mkaya@ adanabtu.edu.tr

Özet: İnşaat sektörü Türkiye'de hızla gelişmekte olan bir sektördür. Duvar örülmesi ve işçiliği bu sektörün önemli parçalarından birini oluşturmaktadır. Bu çalışmada, Veri Madenciliği'nde Birliktelik Kuralı Çıkarım Algoritmalarından biri olan ve Weka veri madenciliği programı içerisinde bulunan Apriori Algoritması kullanılarak duvar işçiliği verimliliğini etkileyen faktörlerin belirlenmesi amaçlanmıştır. Bu çalışmanın amacını gerçekleştirebilmek için Türkiye'de bulunan duvar inşa eden firmalar ile sistematik bir zaman çalışması yapılmıştır. Duvar işçilerinin günlük verimlilik değerleri; ekip büyüklüğü, yaş ve tecrübeleri gibi faktörlerle ilişkili bilgilerle birlikte toplanıp Apriori Algoritması kullanılarak analiz edilmiştir.

Anahtar Sözcükler: Veri Madenciliği, Apriori, Birliktelik Kuralları Çıkarımı, Verimlilik.

The Analysis of the Factors Affecting the Productivity in the Wall Construction of the Using Apriori Data Mining Method

Abstract: The construction sector is a rapidly developing sector in Turkey. The building of the wall and the labor consists one of the important parts of the sector. In this study, it has been aimed to determine the factors affecting the productivity of brick crew members using Apriori Algorithm which is one of the Association Rule mining algorithms in Data Mining and is contained in Weka Data Mining Tool. To achieve the purpose of this study, a systematic time study was undertaken with firms which built wall in Turkey. Daily productivity values of brick crews were collected together with the information related with the factors like the crew size, age and experience of crew members and were analysed using Apriori Algorithm.

Keywords: Data Mining, Apriori, Association Rule Mining, Productivity.

1. Giriş

Yapılarda taş, tuğla, briket, kerpiç ve benzeri gereçlerle yapılan, yapının yanlarını dışa karşı koruyan, iç bölümlerini birbirinden ayıran düşey bölge elemanına duvar denilmektedir. Günümüzde genel olarak inşaat duvarlarının örülmesinde yük dayanımı olmayan tuğla, Ytong, Bims gibi yapı malzemeleri tercih edilmektedir.

Uygulanmasında dikkat, titizlik, tecrübe, bilgi birikimi ve estetiğin önemli yeri olan duvar örülmesi yapı inşasında kıymetli bir yere sahiptir. Duvarlar, tozlarından temizlenmiş, gerektiğinde ıslatılmış yapı malzemeleri ile yerinde, tamamen ipinde, şakülünde ve gönyesinde işlenmesi gereken, düşey derzlerin harç ile tamamen doldurulmasına dikkat edilerek yapılan, yüzeylerdeki fazla harçların malanın keskin kenarı ile kazınıp alınmasının önemli olduğu yapı elemanlarıdır.

Bu çalışmada; yukarıda bahsi geçen ve uygulanması titizlik gerektiren, ancak özenli işçilikle başarılı sonuçlara ulaşılabilecek duvar uygulamalarında; çalışan ekip sayısı, yaş kriteri ve tecrübe

dikkate alınarak veri madenciliği analizlerinden birliktelik kuralları çıkarımı analizi yapılmış ve verimlilik üzerindeki etkisi araştırılarak sonuçlar sunulmuştur.

2. Veri Madenciliği

Veri Madenciliği daha önceden bilinmeyen, geçerli ve uygulanabilir bilgilerin geniş veri kaynaklarından elde edilmesi işlemidir [1]. Bir başka tanıma göre ise veri madenciliği büyük ölçekli veriler arasından yararlı ve anlaşılır olanların bulunup ortaya çıkarılması işlemidir [2]. Veri madenciliği, büyük miktardaki veriden anlamlı kurallar keşfetme ve analiz etme anlamına da gelmektedir [3].

İnşaat sektörü son yıllarda hızla gelişen bir sektör olmuştur. Barınma, sosyal alanlar, iş hayatı gibi yaşamın her alanında ihtiyaç duyulan inşaat işleri özellikle son yıllarda sadece üretilen yapılar üzerinde inceleme yapılmasının yetersiz kaldığı, iş süreçlerinin tamamı üzerinde titizlikle durulması gereken bir hal almıştır. Maliyetlerin yüksek olması, harcama kalemlerinin çok sayıda ve önemli miktarlarda bulunması sebebiyle tüm iş süreçlerinde verimlilik ölçümü yapmanın, elde edilen verilerin üretime yansıtılmasının önemli olduğu aşikâr bir durumdur. Sektörde halen çok sıklıkla kullanılan yapı malzemelerinden oluşturulan duvarların yapım aşamasında, kayıplar meydana getiren dolayısıyla verimliliği düşüren faktörleri belirlemek kritik öneme haizdir.

Teknolojinin her sektörde kullanılmaya başlamasıyla ortaya çıkan büyük verilerden ilgili olanları elde etmedeki zorluk inşaat sektörü için de geçerli durumdadır. Anlamsız veri yığınlarından anlamlı ve işe yarar bilginin elde edilebilmesi için veri madenciliği yöntemlerinden yararlanılmaya başlanılmıştır.

Bu çalışmada, inşası uzun zaman alan, titiz işçilik ve birçok sanat kuralı gerektiren duvar uygulamalarında verimliliği etkileyebileceği düşünülen ekip sayısı, tecrübe ve yaş faktörlerini içeren ve Türkiye'deki bazı inşaat firmalarından elde edilen veriler üzerinden veri madenciliği yöntemlerinden biri olan birliktelik kurallarının çıkarılması amaçlanmıştır.

Çalışmanın materyal ve metot bölümünde, öncelikle toplanan veri seti hakkında kısaca bilgi verilmiştir. Daha sonra bu çalışmada kullanılacak veri madenciliği yazılımı olan Weka [4] programından kısaca bahsedilip, birliktelik kuralları ve birliktelik kuralı çıkarım algoritmalarından biri olan Apriori Algoritması [5] üzerinde durulmuştur. Apriori algoritması [5] incelenmiş, algoritma adımları ve parametreleri verilmiştir. Araştırma ve Tartışma bölümünde ise uygulama detaylandırılmıştır. Son olarak Sonuçlar ve Öneriler bölümünde uygulamanın sonuçlarına yer verilmiştir.

3. Materyal ve Metot

Türkiye'deki 68 adet duvar uygulaması yapan firmadan alınan verilere sistematik bir zaman çalışması yapılmıştır. Alınan bu verilere göre duvar işçiliğinde verimliliği etkileyen faktörleri belirleyip, bu faktörlerin verimliliğe ne derece etkisi olduğunu gözlemleyebilmek amacıyla veri madenciliği yöntemlerinden birliktelik kuralları çıkarımı yöntemi kullanılacaktır. Alınan bu verilerde duvar örme işçilerinin günlük verimlilik değerleri Tablo 1'de gösterildiği gibi ekip büyüklüğü, ekip üyelerinin yaşları ve tecrübeleri bilgileriyle birlikte toplanmıştır.

Toplanan veri daha sonra önişlemden geçirilmiştir. Bu ön işlem sırasında eksik veriler ve hatalı veriler veri setinden çıkarılmıştır. Ayrıca outlier denilen uç noktadaki veriler de veri setinden çıkarılarak verilerin daha tutarlı olması sağlanmıştır.

Tablo 1. Duvar Veri Seti

Ekip Sayısı	Yaş	Tecrübe	Verimlilik
2,00	18,55	4,11	0,46
4,00	19,75	0,67	0,64
5,00	25,00	6,80	0,42
10,00	29,20	2,80	0,94
8,00	33,13	3,00	0,43
7,00	33,75	0,67	1,40
3,00	34,00	4,00	0,22
6,00	32,67	3,00	0,96
…	…	…	…

Veri madenciliği için Weka [4], R [6], Keel [7], RapidMiner [8], MATLAB [9], KNIME [10] ve SPSS Clementine [11],…vs. gibi birçok yazılım bulunmaktadır [12]. Bu yazılımlardan kimisi ticari iken kimisi açık kaynak kodludur. Toplanan veri bu çalışmada Weka programı kullanılarak analiz edilecektir.

Weka [13], Waikato Environment for Knowledge Analysis kelimelerinin kısaltılmasıdır [6]. Yeni Zelanda Waikato Üniversitesi'nde GNU genel kamu lisansı altında geliştirilen Java tabanlı bir veri madenciliği ve makine öğrenmesi yazılımıdır. Veri kümeleri üzerinde önişleme, sınıflandırma, kümeleme, birliktelik kuralı madenciliği, özellik seçimi ve görselleştirme işlemlerini içerisinde bulundurmaktadır. Weka .arff (Attribute Relationship File Format) dosya formatı ile çalışmaktadır. Bu dosya formatı özel olarak tasarlanmış, metin yapısında tutulan bir dosya formatıdır.

Bu çalışma kapsamında kullanılacak veri seti için de öncelikle Tablo 2'de gösterildiği gibi bir .arff dosya formatı geliştirilmiştir. .arff dosya formatı her veri madenciliği metodu için farklıdır. Örneğin sınıflandırma ve kümelendirme algoritmaları için nümerik verilere daha çok ihtiyaç duyarken, birliktelik kuralları çıkarım algoritmaları için kategorik yani nominal verilere ihtiyaç duymaktadır.

Tablo 2. Duvar Veri Seti İçin .arff Dosya Formatı

Kategorizeleştirilmiş (Nominal) Arff Dosyası
@RELATION duvar
@ATTRIBUTE ekip {az,fazla}
@ATTRIBUTE yas {az,fazla}
@ATTRIBUTE tecrube {az,orta,fazla}
@ATTRIBUTE verimlilik {düşük,yüksek}
@DATA
az,genç,orta,düşük
fazla,genç,az,yüksek
fazla,genç,fazla,düşük
fazla,genç,orta,yüksek
fazla,yaşlı,orta,düşük
fazla,yaşlı,az,yüksek
az,yaşlı,orta,düşük
fazla,genç,orta,yüksek
…

Bu çalışmada birliktelik kuralları ve birliktelik kuralı çıkarım algoritmalarından biri olan Apriori algoritması ile çalışılacak olduğundan veri seti Tablo 3'teki gibi kategorik hale getirilmiştir.

Kategorilekleştirme işlemi yapılırken veri aralıklarını belirlemek için birçok yöntem denenmiştir. Yapılan testler sonucu ekip sayısı, yaş ve verimlilik nitelikleri için veri aralığını iki eşit sınıfa ayırmak, tecrübe niteliği için de veri aralığını üç eşit sınıfa ayırmak optimum sonucu vermektedir. Aralık bu testler sonucunda eşit-frekans yöntemi ile belirlenmiştir, yani her aralıkta eşit sayıda veri tutulmuştur. Ekip sayısı niteliği için 2-4 aralığı "az", 5-10 aralığı ise "fazla" olarak belirlenmiştir. Yaş

niteliği için 18,55-32,79 aralığı “genç”, 32,8-45,50 aralığı “yaşlı” olarak belirlenmiştir. Tecrübe niteliği için 0,50-2,53 aralığı “az”, 4,16-18 aralığı “fazla” olarak belirlenmiştir. Sınıf niteliğimiz olan verimlilik niteliği için ise 0,22-0,55 aralığı “düşük”, 0,56-1,4 aralığı ise “yüksek” olarak belirlenmiştir.

Tablo 3. Apriori Algoritması İçin Duvar Veri Setinin Kategorikleştirilmesi

Ekip Sayısı	Yaş	Tecrübe	Verimlilik
az	genç	orta	düşük
fazla	genç	az	yüksek
fazla	genç	fazla	düşük
fazla	genç	orta	yüksek
fazla	yaşlı	orta	düşük
fazla	yaşlı	az	yüksek
az	yaşlı	orta	düşük
fazla	genç	orta	yüksek
…	…	…	…

3.1 Apriori

Birliktelik kuralları, büyük veri kümeleri arasındaki birliktelik ilişkilerini bulur [14]. Çok tercih edilen ve yaygın olarak kullanılan birliktelik kuralları algoritması olan Apriori Algoritması [5], Agrawal ve Srikant [15] tarafından 1994 yılında geliştirilmiştir. Veri setinde içindeki genel eğilimi vurgulamak için kullanılan güvenilir ve kullanışlı bir algoritmadır.

Apriori algoritmasının çalışması sonucunda elde edilen her bir kural, destek ve güven kriterleri ile ifade edilir. Destek kriteri, öğeler arasındaki birlikteliğinin sıklığını ifade etmektedir. Güven kriteri ise bu birlikteliklerin doğruluğunu ifade etmektedir [16].

Sınıf Birliktelik Kuralları (CAR) bu çalışmada genel birliktelik kuralları yerine kullanılmıştır. Birliktelik kurallarının özel bir alt kümesi üzerine odaklanarak veri madenciliğini gerçekleştiren ve Sınıf Birliktelik Kuralları (CAR) olarak adlandırılan bu özellik verilerin bütünleşmesi yani entegrasyonu için kullanılır [17]. Tüm olası kurallar kümesinden en ilginç kuralların seçimini sağlamak amacıyla bu çalışmada Sınıf Birliktelik Kuralları (CAR) kullanılmıştır.

4. Araştırma ve Tartışma

Weka programının Özellik Seçimi yöntemi bölümünde öncelikle ekip sayısı, yaş ve tecrübe nitelikleri arasından hangi niteliklerin daha etkili olduğu Ki-Kare [18] nitelik seçimi yöntemi kullanılarak test edilmiştir. Ki-Kare nitelik seçimi yöntemi oldukça yaygın kullanıma sahip bir yöntemdir. Ki-Kare ilgili sınıfa göre ki-kare istatistiğinin değerini hesaplayarak bir niteliğin değerini değerlendirir [18].

Şekil 1’de de gösterildiği üzere en etkili nitelik tecrübe olarak belirlenmiştir. Ardından yaş niteliğinin ikinci sırada etkili olduğu gözlenmiştir ve en son olarak da ekip sayısının geldiği belirlenmiştir.

```
=== Attribute Selection on all input data ===

Search Method:
        Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 4 verimlilik):
        Chi-squared Ranking Filter

Ranked attributes:
 4.8333  3 tecrube
 0.4615  2 yas
 0.0589  1 ekip

Selected attributes: 3,2,1 : 3
```

Şekil 1. Ki-Kare Nitelik Seçimi

Apriori Algoritması için Tablo 3’teki gibi kategorik verilerden oluşan veri seti kullanılmıştır. Güven değeri 0,60 olarak belirlenmiştir. Ayrıca minimum destek değeri de 0,1 olarak belirlenmiştir. Bu değerlere göre çıkan sonuçlar Tablo 4’te listelenmektedir:

Tablo 4. Apriori Algoritması'nın Duvar Veri Setine Uygulanması Üzerine Elde Edilen Sonuçlar

1	ekip=fazla yas=genc tecrube=orta 9 ==> verimlilik=yuksek 8	conf:(0.89)
2	ekip=az yas=genc 10 ==> verimlilik=dusuk 8	conf:(0.8)
3	ekip=fazla tecrube=orta 14 ==> verimlilik=yuksek 11	conf:(0.79)
4	yas=genc tecrube=az 14 ==> verimlilik=dusuk 11	conf:(0.79)
5	ekip= fazla tecrube=az 20 ==> verimlilik=dusuk 14	conf:(0.7)
6	yas=genc tecrube=orta 13 ==> verimlilik=yuksek 9	conf:(0.69)
7	tecrube=az 27 ==> verimlilik=dusuk 18 conf:(0	conf:(0.67)
8	ekip=az yas=yasli 15 ==> verimlilik=yuksek 10	conf:(0.67)
9	tecrube=orta 24 ==> verimlilik=yuksek 15	conf:(0.63)
10	yas=yasli tecrube=fazla 15 ==> verimlilik=yuksek 9	conf:(0.6)

Kurallardan da görüldüğü üzere, tıpkı nitelik seçiminde olduğu gibi kural çıkarımında da tecrübe niteliği önemli rol oynamaktadır.

Bu kurallara göre; duvar işçilerinin ekip sayısı fazla, yaşları genç ve tecrübeleri de orta düzeyde ise verimlilikleri yüksek olmaktadır. Ekip sayısı azsa ve çalışanların yaşları da genç ise verimlilikleri düşük olmaktadır. Ekip sayısı fazla iken tecrübeleri orta düzeyde ise verimlilikleri yüksek olmaktadır. Yaşları gençken tecrübeleri azsa ya da ekip sayıları fazla iken tecrübeleri azsa verimlilikleri de aynı oranda düşük olmaktadır. Genç çalışanların tecrübeleri orta düzeydeyse verimlilikleri de yüksek olmaktadır. Yaşlı olan çalışanların ise tecrübeleri de fazla iken verimlilikleri de yüksek olmaktadır. Ekip sayısı az olan çalışanlar yaşlıysa verimlilikleri de yüksek bulunmuştur. Sadece tecrübeye dayanarak da kural çıkarımı başarılı olmuştur. Tecrübeleri az olan çalışanların verimliliklerinin düşük olduğu gözlenirken, tecrübeleri orta düzeyde olan çalışanların verimliliklerinin yüksek olduğu gözlenmiştir.

Ekip sayısının belirleyici nitelik olmasının da etkisiyle ekip sayısı düşük ise verimliliğin de düşük olduğu, fazla ise verimliliğin de fazla olduğu gözlemlenmiştir.

5. Sonuç ve Öneriler

Bu çalışma inşaat sektöründe duvar işçiliği verimliliğini etkileyen faktörleri veri madenciliği yöntemleri ile araştıran az sayıdaki çalışmadan biridir. Bu çalışmada, veri madenciliğinde güvenilir ve popüler olan birliktelik kuralı çıkarım algoritmalarından Apriori algoritmasının inşaat sektöründe kullanılmasıyla duvar işçilerinin verimliliklerinin analizi gerçekleştirilmiştir. Böylelikle verimlilikte kullanılabilecek en uygun kuralların çıkarımı sağlanmıştır.

Böylece elde edilen sonuçlar sadece inşaat uygulayıcıları için değil aynı zamanda yapı yönetiminde veri madenciliğinin uygulanmasıyla ilgili araştırmaları gerçekleştirecek olan gelecekteki araştırmacılar için de inşat işçileri ekiplerinin verimliliğini etkileyen önemli faktörleri belirlemede önem teşkil edecektir.

6. Teşekkür

Bu çalışma TÜBİTAK (Türkiye Bilimsel ve Teknolojik Araştırma Kurumu) tarafından desteklenen 106M055 araştırma projesinin bazı bulgularına dayanmaktadır.

7. Kaynaklar

[1] Silahtaroğlu, G., "Kavram ve Algoritmalarıyla Temel Veri Madenciliği", **Papatya Yayıncılık Eğitim,** İstanbul, (2008).

[2] Özkan, Y., "Veri Madenciliği Yöntemleri", **Papatya Yayıncılık Eğitim**, İstanbul, (2008).

[3] Linoff, G. S., Berry, M. J. A., "Data Mining Techniques : For Marketing, Sales, and Customer Support", **Wiley & Sons, Incorporated, John**, (1997).

[4] WEKA, http://www.cs.waikato.ac.nz/ml/weka/, (Erişim Tarihi: 2013).

[5] Han, J., Kamber, M., "Data Mining : concept and techniques", **CA: Academic Press,** (2001).

[6] R, http://www.r-project.org/, (Erişim Tarihi: 2013).

[7] KEEL, http://www.keel.es/, (Erişim Tarihi: 2013).

[8] RAPIDMINER, http://rapidminer.com/, (Erişim Tarihi: 2013).

[9] MATLAB, http://www.mathworks.com/products/matlab/, (Erişim Tarihi: 2013).

[10] KNIME, http://www.knime.org/, (Erişim Tarihi: 2013).

[11] SPSS CLEMENTINE, http://www-01.ibm.com/software/analytics/spss/, (Erişim Tarihi: 2013).

[12] Aziz, A. A., Ismail, N. H., Ahmad, F. "Mining Students' Academic Performance", **Journal of Theoretical and Applied Information Technology**, 5(3)3, 485-495, (2013).

[13] Garner, S.R., "Weka: The waikato environment for knowledge analysis", **Proc New Zealand Computer Science Research Students Conference**, 57-64, (1995).

[14] Thabtah, F., Cowling, P., Hammoud, S., "Improving rule sorting, predictive accuracy and training time in associative classification", **Expert Systems with Applications**, 31, 414–426, (2006).

[15] Agrawal, R., Srikant, R., "Fast Algorithms for Mining Association Rules in Large Databases", **Proceedings of the 20th International Conference on Very Large Databases (VLDB)**, 487-499, (1994).

[16] Güngör E., Yalçın N., Yurtay N., "Apriori Algoritması ile Teknik Seçmeli Ders Seçim Analizi", **Akademik Bilişim 2013 Konferansı Bildirileri**, (2013).

[17] Liu, B., Hsu, W., Ma, Y., "Integrating Classification and Association Rule Mining", **Proc. Fourth Int'l Conf. Knowledge Discovery and Data Mining (KDD '98),** 80–86, (1998).

[18] ChiSquaredAttributeEval, http://weka.sourceforge.net/doc.stable/weka/attributeSelection/ChiSquaredAttributeEval.html, (Erişim Tarihi: 2013).