

## THE OVERVIEW OF DATA MINING APPLICATION ON CONSTRUCTION SECTOR AND INTERPRETATION OF ECONOMIC IMPACT\*

**Yazar / Author:** Yrd. Doç. Dr. Abdullah Emre KELEŞİ

### Abstract

The construction sector is a dynamic sector needs to work with different and compelling a large number of stakeholders. This dynamism leads to the acquisition, use and management of a large number of data. The data mining is one of the branches that help to get useful data from the data stacks. Data Mining can be defined, including outnumber large amounts of information the data access to meaningful data as a business purpose. Especially in recent years, data mining applications has been among the relevance topics such as the construction sector installed on a discipline that involves the amount and volume data. From this perspective in this study, was examined the researches which interesting in this subject in the construction industry. The obtained findings are intended basis for the following studies. At the same time, data mining methods of the application economic contribution in the construction sector were overviewed brought to the business.

**Key Words:** Data mining, data stacks, construction firms, construction industry, construction economy.

## İNŞAAT SEKTÖRÜNDE VERİ MADENCİLİĞİ UYGULAMALARINA GENEL BAKIŞ VE EKONOMİK ETKİLERİNİN YORUMLANMASI

### Özet

Üniversite-sanayi işbirliği, modern sanayinin çıkış noktası olan İngiliz sanayi devrimiyle birlikte oluşmaya başlamıştır. 1975-1985 döneminde 1970'lerden bu yana gittikçe yaygınlaşan üniversite-sanayi araştırma işbirliğinin önemi günümüzde bir kat daha artmıştır. Yeni bilimsel bilgiye duyulan ihtiyaç, sanayi ve akademik kurumlar arasındaki bağlantıyı artırmaktadır. Özellikle Küçük ve Orta Boy İşletmeler (KOBİ) içinde bilimsel ve teknik bilginin önemli ölçüde dağılımına ihtiyaç duyulmaktadır. Bu çalışmada dünyada üniversite-sanayi işbirliği önce ele alınacak, daha sonra Türkiye'de üniversite-sanayi işbirliğinin gelişimi incelenecektir.

**Anahtar Kelimeler:** Üniversiteler ve Sanayi, Bölgesel Kalkınma, Avrupa Birliği, Türkiye, Trakya Bölgesi

### 1. Giriş

İstatistik, bilgi teknolojileri, bankacılık sektörü, finans, tıp, biyoloji, genetik, kriminoloji, istihbarat, telekomünikasyon, eğitim bilimleri vb. hemen her alanda sıklıkla rastlanan “Veri Madenciliği metotlarıyla veri yığınlarından anlamlı ifadeler elde etme yöntemleri” Türk inşaat sektöründe henüz yeterince kullanılmayan yeni bir inceleme ve uygulama alanı olarak karşımıza çıkmaktadır. Ülkemiz sanayi ve hizmet sektörünün önemli bir parçası konumundaki inşaat sektöründeki bu anlamdaki eksikliklerin, bu gibi çalışmalardan olumlu sonuçlar alınması sağlanarak ve sonuçlarının sektör temsilcilerine sunulup farkındalık yaratılarak giderilebileceği düşünülmektedir.

\* Bu çalışma 26-27 Kasım 2016 tarihlerinde, İstanbul'da düzenlenen International Congress of Management Economy And Policy isimli uluslar arası bilimsel kongrede bildiri olarak sunulmuştur

<sup>1</sup> Adana Bilim ve Teknoloji Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, İnşaat Mühendisliği Bölümü. akeles@adanabtu.edu.tr

Teknolojik gelişmelerin büyük bir hızla devam ettiği günümüzde, inşaat sektörü de, yapım sürecinde pek çok yeniliğe açık hale gelmiştir. Yapı üretiminin merkezi olan inşaat işlerinde, sayıca çok fazla değişken bir arada ve etkileşim halinde bulunmakta ve farklı iş kombinasyonları ortaya çıkmaktadır. Bahsedilen bu dinamik üretim süreci yapı üretimini farklı riskleri bünyesinde barındıran tehlikeli işler sınıfına sokmaktadır. Bu durum inşaat yönetiminin rastgele davranışlardan ziyade; planlı, sistemli ve uyum içinde çalışılan bir ortam oluşturma ve sürdürme sanatı olması sonucunu beraberinde getirdiği söylenebilir. Yönetilmesi başlı başına bir ders konusu olan inşaatlara sektörel anlamda bakıldığında, ülkemizde çoğu işletmenin yönetiminin ilgili mühendis ya da mühendislerin kişisel tercihlerine bırakıldığı sonucu karşımıza çıkmaktadır. Bu durum, farkında olunmadan çok sayıda kaybı beraberinde getirmektedir. Yıllardır çoğu farkında bile olunmayan bu sorunların günümüz bilgi teknolojileri yardımıyla giderilebilmesi için bu çalışmada Veri Madenciliği metotları incelenmiş ve günümüze değin yapılan çalışmaların neler olduğuyla ilgili bir perspektif çıkarılmıştır. Ardından, gelecek çalışmalara dair öneriler geliştirilerek sunulmuştur.

## 2. Veri Madenciliğine Genel Bakış

Veri Madenciliği, fazla miktarda bilgi içeren veriden, belirlenen amaç doğrultusunda anlamlı veriye ulaşma işi olarak tanımlanabilir. Bir başka deyişle Veri Madenciliği, çok miktarda veriye dayanarak önceden keşfedilmemiş bilgileri ortaya çıkarma ve bunları kullanma süreci, yani büyük miktarda veri içinden, gelecekle ilgili tahmin yapılmasını sağlayan bağıntuların aranması işidir. Burada veriler arasındaki ilişkileri, kuralları ve özellikleri belirlemekten bilgisayar programları sorumludur. Amaç bir anlamda, daha önceden keşfedilmemiş veri örüntü ve desenlerini tespit edebilmektir (Dener vd., 2009).

Günümüzde karar verme sürecine ihtiyaç duyulan ve fazla miktarda veri bulunan birçok alanda Veri Madenciliği metotları yaygın olarak kullanılmaya başlamıştır. Çalışmanın giriş bölümünde de belirtildiği gibi; bankacılık, pazarlama, sigortacılık, borsa işlemleri, iletişim sektörü, tıp, mühendislik, kriminoloji, endüstri, istihbarat gibi pek çok alanda uygulamaları bulunmaktadır. Bu bilgi ışığında literatür incelendiğinde Veri Madenciliğinin en sık kullanıldığı alanların hacimsel olarak büyük ve çok sayıda veri içeren; tıp, biyoloji ve genetik bilimleri olduğu görülmektedir. Tüm bunlar ışığında, inşaat sektörü için bir değerlendirme yapmak gerekirse sektörle ilgili bilimsel çalışmalar bu alanda henüz gelişme aşamasındadır denilebilir. Sektörde yapılan çalışma ve aktif kullanılan program sayısı oldukça azdır.

Veri Madenciliği, çeşitli aşamaları bünyesinde barındıran bir süreç olarak görülmektedir. Veri yığınları arasında, istenilen yararlı veriyi belirlemekle birlikte, bilgi keşfi sürecinde örüntüleri ayırıştırarak süzmek ve bir sonraki adıma hazır hale getirmek de bu sürecin bir parçasıdır. Üzerinde inceleme yapılan verilerin özelliklerinin bilinmemesi durumunda ne kadar etkin olursa olsun hiç bir Veri Madenciliği algoritmasının fayda sağlaması olanaklı değildir. Bu nedenle, Veri Madenciliği çalışmalarına geçilmeden önce veri özelliklerinin detaylı analizi gerekmektedir.

Veri Madenciliği sürecinde izlenen adımlar aşağıdaki şekilde özetlenebilir:

- i. Problem tanımlanması,
- ii. Verilerin hazırlanması,
- iii. Modelin kurulması,

- iv. Modelin kullanılması,
- v. Modelin takibi.

Problemin tanımlanması; Veri Madenciliği çalışmalarında başarı elde edebilmenin en önemli şartı, projenin hangi amaç doğrultusunda yapılacağı ve elde edilecek sonuçların başarı düzeylerinin nasıl ölçüleceğinin tanımlanmasıdır.

Verilerin hazırlanması; Bu kısımdaki işlemler; “toplama”, “değer biçme”, “birleştirme ve temizleme”, “örnekleme seçimi” ve “dönüştürme” aşamalarından oluşmaktadır. Modelin kurulması; tanımlanan problem için en uygun modelin bulunabilmesi ancak çok sayıda modelin denenmesi ile mümkün olabilmektedir. Bu nedenle problemin tanımlanması aşamasından sonraki veri hazırlama ve model kurma aşamaları, en iyi olduğu düşünülen modele ulaşıncaya kadar yinelenen bir süreç olarak düşünülmektedir.

Modelin kullanılması; yukarıda açıklanan aşamalardan sonra kurulan ve geçerliliği kabul edilen model doğrudan bir uygulama olabileceği gibi, bir başka uygulamanın alt parçası olarak da kullanılabilir.

Modelin takibi; zaman içerisinde sistemlerin özelliklerinde ve dolayısıyla ürettikleri verilerde ortaya çıkan değişiklikler, kurulan modellerin sürekli olarak izlenmesini ve yeniden düzenlenmesini gerektirecektir. Bu aşama kurulan modelin izlenmesiyle beraber denetlenmesini de içermektedir.

### **2.1. Veri Madenciliği Metotlarının Özetlenmesi**

Temel anlamda Veri Madenciliği modelleri işlevlerine göre 3 ana grupta toplanabilir. Bunlar:

- i. Sınıflama (Classification) ve Regresyon (Regression),
- ii. Kümeleme (Clustering),
- iii. Birliklilik Kuralları (Association Rules).

Genellikle, “Sınıflama” ve “Regresyon” modelleri tahmin edici, “Kümeleme” ve “Birliklilik Kuralları” modelleri ise tanımlayıcı modeller olarak gruplandırılabilir.

### **3. Yöntem**

Araştırmada öncelikle veri madenciliği kavramı ve çalışma sistematığı detaylarıyla açıklanmış ardından inşaat işletmelerinde günümüze değin uygulanan örnekleri özetlenmiştir. Bundan sonra özellikle ülkemizde eksikliği hissedilen konuya dair durum analizi inşaat işletmelerinin yapısı ele alınarak yapılmıştır. Sektördeki uygulanma alanlarına vurgu yapılarak açıklanmıştır. Aynı zamanda veri madenciliği yöntemlerinin uygulanması sonucunda, verileri yönetmenin daha kolay ve anlaşılır olabileceği hesaba katılarak ilgili sistematığın inşaat işletmelerine getirebileceği ekonomik faydalardan bahsedilmiştir. Tüm bunlarla birlikte, ileriki çalışmalar için ipucu verebilecek görüşler ortaya atılmıştır.

### **4. Bulgular**

#### **4.1. İnşaat Sektöründeki Veri Madenciliği Çalışmalarının Özetlenmesi**

Yapım Sektöründe Veri Madenciliği konusu ile ilgili çalışmalar son yıllarda yaygınlaşmaya başlamıştır. Sektörel anlamda bakıldığında Veri Madenciliği uygulamaları, pek çok sayısal verinin elde edilebileceği inşaat işleri için verinin değerlendirilmesi açısından önemli bir yere sahip olacaktır. Günümüze değin yapılan çalışmalardan ülkemizde ve global çapta yapılanların bir kısmı aşağıda özetlenmiştir.

Subsymbolic Methods for Data Mining in Hydraulic Engineering adlı çalışmada, Yapay Sinir Ağları, Genetik Programlama ve Veri Madenciliği yöntemleri ile elde edilen verilerle deneysel verilerin sonuçları karşılaştırılmıştır. Veriler, farklı algoritma çalışmaları sonucunda gereksiz olabilirken, bazı gerçek bilgileri ise sağlayabilmektedir gibi sonuçlar elde edilmiştir (Minns, 2000).

Automated Classification of Construction Project Documents adlı çalışmada, proje belgelerinin önemli bir yüzdesinin metin biçiminde olduğundan, bu belgelerde yer alan bilgilere erişimin düzenlenmesi ve iyileştirilmesi için bilgi sınıflandırma şeması kullanılması gerektiği belirtilmiştir. Bu noktadan hareketle, bilgi alışverişini gerçekleştirmek için ortak bir çerçeve sunan şema oluşturulmuştur. Çalışma sonucunda bir belge sınıflandırma sistemi geliştirilmiştir (Caldas, vd., 2002).

Estimating Future Highway Construction Costs adlı çalışma, ABD'nin Louisiana eyaletinde gelecek yıllardaki otoyol inşaat maliyetlerini tahmin eden bir modelin geliştirilmesi amaçlanmıştır. Gelecek inşaat maliyetleri inşaat işçiliği, malzeme, ekipman ve beklenen sözleşme özellikleri ile fiyat tahminlere dayanan endeks değerleri açısından açıklanmıştır. (Wilmot ve Cheng, 2003).

Veri Madenciliği ve Çimento Sektöründe Bir Uygulama adlı çalışmada, çimento ürününün bir parametresi olan basma mukavemetini tahmin etmek için Veri Madenciliği tekniklerinden faydalanılması amaçlanmıştır. Çalışmada Portland kompozit çimentosunun basma mukavemetinin tahmini için gen denklem programlama ve yapay sinir ağları olarak adlandırılan Veri Madenciliği yöntemleri kullanılmıştır (Baykasoğlu, 2005).

Data Mining for Occupational Injuries in The Taiwan Construction Industry adlı çalışmada, 1999-2004 döneminde Tayvan inşaat sektöründe iş kazalarının özelliklerini tanımlamada kural madenciliği kullanılmıştır. İş kazaları sonucundaki ölümlerin ortaya çıkmasında diğer faktörlerle birlikte yağmur yağmasının önemli etkisinin olduğu belirtilmiştir (Liao ve Perng, 2008).

Applications of Artificial Intelligence and Data Mining Techniques in Soil Modeling adlı çalışmada; yapay sinir ağları, genetik programlama ve regresyon gibi yaygın kullanılan örüntü tanıma yöntemlerine ek olarak Veri Madenciliği sistemleri incelenmiş ve örüntü tanıma uygulamaları gerçekleştirilmiştir. Geleneksel metotların sonuçlarına yakın sonuçlar elde edilmiştir (Javadi ve Rezaia, 2009).

Construction Crew Productivity Prediction by Using Data Mining Methods adlı çalışmada ise, anket yoluyla inşaat işletmelerde çalışan seramik işçilerinden; ekip sayıları, ekipteki kişilerin o iş tecrübeleri ve yaş ortalamaları bilgileri dikkate alınarak verimlilik değerleri elde edilmiştir. Bu çalışmada ölçme teknikleri ile elde edilmiş olan inşaat sektöründeki seramik verileri üzerinden verimliliğin sınıflandırılarak kurallar çıkarılması vasıtasıyla seramik işlerinde yüksek verimliliğin girdilere göre nasıl sağlanabileceği üzerinde durulmuştur (Kaya vd., 2013).

İnşaat Mühendisliği'nde Bilişim Kavramı ve Veri Madenciliği Algoritmalarıyla Bir Sistemin Oluşturulması adlı çalışmada, İnşaat Sektöründe elde

edilen ve kullanılan bilgilerin fazlalığı nedeniyle elektronik ve takip edilebilir bir sistemin oluşturulması gerektiği belirtilmiştir. Kullanılacak en uygun malzeme ve yöntemin belirlenmesi için bir sistem oluşturulmuş ve ilgili sistemin geri besleme ile eğitilmesi amaçlanmıştır (Durap ve Doğan, 2014).

Veri Madenciliği Kullanarak Beton Basınç Dayanımının Belirlenmesi adlı çalışmada, beton basınç dayanımının belirlenmesi için Veri Madenciliği yöntemleri kullanılarak modeller geliştirilmiştir. Çalışma sonucunda beton basınç dayanımının belirlenmesinde Veri Madenciliği yöntemleri ile gerçek değerlere yakın sonuçlar elde edilebildiği belirtilmiştir (Özel ve Topsakal, 2014).

Duvar İnşa Edilmesinde Verimliliği Etkileyen Faktörlerin Apriori Veri Madenciliği Yöntemi Kullanılarak Analizi adlı bir diğer çalışmada, inşaat sektöründe duvar ustası olarak çalışanlardan elde edilen demografik bilgilerle, çalışan verimliliğinin yükseltilmesi adına Birliktelik Kuralları çıkarılmıştır (Keleş ve Kaya, 2014).

İnşaat Projelerinde Şantiye Şeflerinin Liderliği ve Çalışan Motivasyonu İlişkisinin Veri Madenciliği ile Belirlenmesi adlı doktora çalışmasında, Türk inşaat sektöründe faaliyet gösteren inşaat işletmelerindeki şantiye şefi konumundaki liderlerle bu kişilerle birlikte çalışanlara çift taraflı bir anket sistematığı uygulanmış ve liderlik iki farklı açıdan belirlenmiştir. Ardından, veri madenciliği yöntemlerinden birliktelik kural çıkarım yöntemi ile şantiye şeflerinin liderlikleri ile birlikte çalıştığı aslarının motivasyonları arasındaki ilişkiler belirlenmiş ve sunulmuştur (Keleş, 2016).

## **5. Sonuçlar Ve Öneriler**

Bu çalışmayla elde edilen Türk inşaat sektöründeki veri madenciliği çalışmalarına dair sonuçlar ve gelecek çalışmalara fikir verebileceği düşünülen öneriler aşağıda sunulmuştur.

### **5.1. Sonuçlar**

Literatür incelendiğinde ülkemizde, “inşaat sektöründeki verilerin anlamlı ifadelerle dönüştürülmesini sağlayan veri madenciliği uygulamalarıyla örneklerinin azlığı dikkat çekmektedir. Oysaki çalışmada da vurgulandığı üzere, veri madenciliği uygulamaları ile inşaat sektörü gibi çok sayıda ham verinin elde edilebildiği/edilebileceği bir sektörde ilgili programların kullanımı vasıtasıyla daha yararlı sonuçlar elde edilebilecektir. Üretim yeri olan şantiye ve daha çok yönetim yeri olan ofis binalarında sistemli yürütülen çalışmaların sektör ve ülkemiz ekonomisine katkı sağlayacağı düşünülmektedir.

Bu çalışmanın bir diğer önemli sonucu, pek çok araştırmacı tarafından özellikle son yıllarda yaygın şekilde kullanılan veri madenciliği metot ve programlarının inşaat sektöründeki verilere uygulanabileceği ve yararlı sonuçlar elde edilebileceğidir. Amaç doğrultusunda şantiye ve ofislerden gelen bilgi ve verilerin bu şekilde sistematığe bağlanmasının, ardından sektör temsilcileriyle paylaşılıp kullanılmasının zaman ve iş gücü kayıplarını azaltmada etkili olabileceği gibi inşaat ekonomisine de faydaları olacağı düşünülmektedir.

### **5.2. Sonraki Çalışmalar İçin Öneriler**

İnşaat sektörünün çok sayıda ve çeşitli malzeme, işgücü, donanım, finansal kuruluş gibi yönetim araçlarıyla yönetilmesi durumu, birbirinden farklı ilişki ve

verilerin elde edilmesi sonucunu doğurmaktadır. Günümüz yoğun rekabet ortamında bu verilerin doğru analizlerinin sağlanmasıyla sistemli hale dönüştürülmesinin işletmelerin yararına olacağı aşikârdır. Bu faydaların başında ise ekonomiklik gelmektedir. Şantiyeler, ofisler ve laboratuvarlardan elde edilen verilerin veri madenciliği ile yönetilmesi gereksiz kayıpları azaltırken; verimlilik, etkinlik, üretkenlik gibi araçları ise arttırabilecektir. Özellikle ülkemiz inşaat sektöründe faaliyet gösteren işletmelere ve çalışanlarına bu ve benzeri çalışmaların ulaşmasının önemli olduğu düşünülmektedir. Veri madenciliği programları ile sistematığe bağlanan verilerin yönetilmesinin ve işletmelerin geleceği öngörmelerinin daha kolay olacağı ele alınırsa, Türk inşaat sektörü yukarıda da sayılan faydaları sağlayabilmiş olacaktır. Bu doğrultuda işletmelerin öncelikle elindeki verilerden neye ulaşmak istediği belirlenmeli ardından bu amaç doğrultusunda sınıflandırma, kümeleme, regresyon, birliktelik kuralı çıkarımı gibi çalışmalarla işletme yararına sonuçlar elde edilmelidir. Sonraki çalışmalarda, bu verilerle inşaat ekonomisinden, şantiye yönetimine; deneysel çıkarımlardan pratik tüm inşaat uygulamalarına kadar veri madenciliğinin uygulanmasının sektöre olumlu yansıtacağı düşünülmektedir.

#### **Kaynakça**

- Baykasoğlu A., (2005). Veri Madenciliği ve Çimento Sektöründe Bir Uygulama. Akademik Bilişim Konferansları.
- Caldas C., Soibelman L., Han J., (2002). Automated Classification of Construction Project Documents., *J. Comput. Civ. Eng.*, 16(4): 234-243.
- Dener M., Dörterler M., Orman, A., (2009). Açık Kaynak Kodlu Veri Madenciliği Programları: Weka'da Örnek Uygulama. Akademik Bilişim Konferansları, AB 2014.
- Durap A., Doğan Y., (2014). İnşaat Mühendisliği'nde Bilişim Kavramı ve Veri Madenciliği Algoritmalarıyla Bir Sistemin Oluşturulması. Akademik Bilişim Konferansları, AB 2014.
- Javadi, A. A., Rezaia, M., (2009). Applications of Artificial Intelligence and Data Mining Techniques in Soil Modeling. *Geomechanics and Engineering, S&T Journals*, Techno-Press, 1(1): 53-74).
- Kaya M., Keleş A. E., Laptalı O., E., (2013). Construction Crew Productivity Prediction by Using Data Mining Methods, *Proceedings of the 4th World Conference on Learning, Teaching and Educational Leadership*, *Procedia - Social and Behavioral Sciences*, 141: 1249–1253.
- Keleş A. E., (2016). İnşaat Projelerinde Şantiye Şeflerinin Liderliği ve Çalışan Motivasyonu İlişkisinin Veri Madenciliği ile Belirlenmesi. Çukurova Üniversitesi Fen Bilimleri Enstitüsü Doktora Tezi, Adana.

Keleş A., E., Kaya M., (2014). Duvar İnşa Edilmesinde Verimliliği Etkileyen Faktörlerin Apriori Veri Madenciliği Yöntemi Kullanılarak Analizi. Akademik Bilişim Konferansları, AB 2014.

Liao C., W., Perng Y., H., (2008). Data Mining for Occupational Injuries in The Taiwan Construction Industry. *Safety Science*, 46(7): 1091-1012.

Minns A., W., (2000). Subsymbolic Methods for Data Mining in Hydraulic Engineering. *Journal of Hydroinformatics*, IWA Publishing.

Özel C., Topsakal A., (2014). Veri Madenciliği Kullanarak Beton Basınç Dayanımının Belirlenmesi. *Cumhuriyet Üniversitesi Fen Fakültesi Fen Bilimleri Dergisi*, 35:1.

Piramuthu S., (1998). Evaluating Feature Selection Methods For Learning in Data Mining Applications, Thirty-First Annual Hawai International Conference on System Sciences, IEEE Computer Society, Kohala Coast Hawaii USA, 294.

Wilmot C., G., Cheng P., E., (2003). Estimating Future Highway Construction Costs. *J. Constr. Eng. Manage*, 129 (3): 272-279.