

Türk Kamu İnşaat Projelerinde Yaşanan Uyuşmazlıklara Yönelik Bir Veri Madenciliği Yaklaşımı

İsmail C. YILMAZ,

Ar.Gör. İstanbul Arel Üniversitesi, İstanbul, cengizyilmaz@arel.edu.tr

Attila DİKBAŞ,

Prof. Dr. İstanbul Teknik Üniversitesi, İstanbul, dikbas@itu.edu.tr

ÖZET

Projelerin herhangi bir fazında yaşanan uyuşmazlıklar, taraflar arasında uzlaşma ile çözülemeyen, çözüm mercii olarak tahkime, üst kurullara ya da yargı yoluna başvurulur. Sonuç olarak bir hükme bağlanan uyuşmazlıklar yazılı bir doküman olarak kurul arşivlerinde kayıt altında tutulur. Bu yazılı dokümanlar, gelecekte uygulanacak inşaat projelerinde karşılaşılabilecek muhtemel ihtilafların uyuşmazlıklara yani daha büyük sorunlara yol açmadan önlenmesi açısından oldukça değerlidirler. Karşılaşılabileceği uyuşmazlıkları en az indirmek isteyen ya da karşılaştığı uyuşmazlık ile nasıl etkin bir şekilde mücadele edeceğini öğrenmek isteyen bir proje paydaşının bahsi geçen yazılı doküman arşivini tamamen taraması ve içeriklerinden kendisine yarayacak olanları seçebilmesi gerekir. Ancak bu; inşaat sektörünü tanımayan, teknik bilgi ve tecrübesi olmayan paydaşlar için son derece zor bir durumdur. Bundan dolayı, bu tarz dokümanların otomatik olarak sınıflandırılması oldukça kullanışlı ve faydalı olabilir. Bu çalışmada; inşaat projelerine ait dokümanların otomatik sınıflandırıldığı önceki çalışmalar ile bir kıyaslama çalışması gerçekleştirilmiştir. Bu çalışmalardan elde edilen en başarılı sonuçlarda kullanılan metin madenciliği teknikleri Yüksek Fen Kurulu'ndan alınan uyuşmazlık karar dokümanları üzerinde uygulanmış, test sonuçları tartışılmış ve yeni bir yaklaşım önerilmiştir.

Anahtar sözcükler: veri madenciliği, inşaat uyuşmazlıkları, kamu projeleri, metin madenciliği.

A Data Mining Approach For Disputes In Turkish Public Construction Projects

ABSTRACT When the disputes in public construction projects are not settled through negotiation participants may apply to arbitration, council or litigation. The disputes that have been adjudicated are recorded as a written matter in their archives. These documents are critical in terms of preventing future conflicts from turning into disputes. Project participants who want to minimize the risk of facing any disputes or who want to know how to tackle disputes should go through all documentation and look for similar cases. However it is a big challenge for participants who have little experience and knowledge about construction sector and its technical concepts. Accordingly it would be a useful solution to categorize such kind of documents automatically. This study presents a benchmarking with the previous studies in which the construction documentation is classified automatically in construction information technology. The most successful text mining techniques have been applied to documentation relating to public construction disputes that has been gathered from "Turkish Supreme Board of Infrastructure and Construction", the results are discussed and a new approach has been presented.

Keywords: data mining, construction disputes, public projects, text mining.

1. GİRİŞ

Global ölçekte inşaat endüstrisindeki hızlı yükseliş beraberinde birçok uyuşmazlık ve hak talebi sorununu da getirdi. En hızlı yükseliş trendlerinden birine sahip olan Türkiye’ de Adli Sicil kayıtlarından alınan istatistiklere göre imar, kamulaştırma, yıkma ve buna bağlı işlere bağlı yıl içinde açılan ve bozularak gelen davalarda yıllık yaklaşık %12’lik bir artış söz konusudur [1-2]. Bu artıştan, ülkedeki inşaat faaliyetlerinin azımsanamayacak bir bölümünü oluşturan kamu projeleri de payını almaktadır. Dolayısıyla kamu projelerinde yaşanacak uyuşmazlıkların önüne geçebilecek her adım, sektör ve ülke için ciddi zaman ve maliyet tasarrufu anlamına gelebilir. Türkiye’ deki kamu projelerinde yaşanabilecek ihtilaflar için tahkim görevi görebilecek en önemli kurum Çevre ve Şehircilik Bakanlığı’na bağlı Yüksek Fen Kurulu’ dur ve kamu idareleri ile yükleniciler arasındaki anlaşmazlıkların çözümünde etkin rol oynamaktadır. Kamu işverenleri ile karşılaşabilecekleri uyuşmazlıklar ile ilgili Yüksek Fen Kurulu’nun genel eğilimini tahmin etmek isteyen kamu yüklenicileri öncelikli olarak bir vaka geri çağırım çalışması yapmak isterler. Ancak oldukça resmi bir dille yazılmış ve aynı zamanda teknik tabirler ile dolu uyuşmazlık dokümanlarını anlamak oldukça zahmetli bir uğraştır. Bu bakımından bu tarz dokümanların otomatik sınıflandırılması ve muhataplarına ulaştırılması oldukça önem taşır. Hukuki evrakların belli bir sistematikle sınıflandırılması ve emsal kararların bulunması işlemi çok eskilerden beri sürekli geliştirilmektedir ve bu tarz doküman arşivlerinin muazzam yığınlar haline gelmesi ile beraber son yıllarda bu işlemin otomatik bir hale getirilmesi çalışmaları hız kazanmıştır. Bu çalışmada; Türkiye’ deki kamu inşaat projelerinde yaşanan uyuşmazlıklarla ilgili Yüksek Fen Kurulu arşivlerinden özel izinle ilk etapta elde edilen belli sayıda doküman; mevcut bazı metin madenciliği teknikleri ile sınıflandırılmış, bu tekniklerin bu vaka için uygun olup olmadığı tartışılmış ve yine bu vaka için yeni bir model yaklaşım önerisi sunulmaya çalışılmıştır.

2. LİTERATÜR ANALİZİ

Bu çalışmanın gerçekleştirilmesinde kategorilerine ayrılabilen dokümanların metin madenciliği teknikleri ile sınıflandırılabilirliğini gösteren önceki çalışmalar etkili olmuştur. Bu tekniklerin inşaat dokümanlarının sınıflandırılmasında da yararlanılabileceği ayrıca yapılan literatür araştırması ile görülmüştür. Bununla ilgili; (Arditi ve diğ., 1998); yaptıkları çalışmada; inşaat işlerinde anlaşmazlıkların birçok kompleks ve birbirine bağlı faktörlerden meydana geldiğini bu faktörleri yorumlamanın zor olduğunu belirterek, aynı çalışmada son 12 yıllık Illionis İstinaf Mahkemesi kararlarını toplayıp, yapay sinir ağları teknolojisi kullanarak inşaat anlaşmazlıklarını tahmin etmeyi amaçlamışlardır. (Fan ve Li)(2011) çalışmalarında; Hong Kong’ ta inşaat işlerinde yaşanan kaza vakalarının raporları arasındaki benzerlikleri yine metin madenciliği teknikleri ile araştırmışlar ve bu şekilde kolaylıkla edinilen deneyimlerle yaşanan kaza oranlarında azalma yaşanabileceğini savunmuşlardır. (Caldas ve diğ.)(2002); inşaat enformasyon sistemlerinde metin dokümanlarının otomatik olarak sınıflandırılabilmesi için klasik makine öğrenmesi tekniklerinin kullanıldığı ve alınan sonuçlara göre bir model prototipinin önerildiği bir çalışma gerçekleştirmişlerdir. (Caldas ve diğ.)(2003), yaptıkları diğer çalışmada ise; inşaat proje dokümanlarının otomatik hiyerarşik şekilde sınıflandırılması için bir metodoloji tanımlamıştır. Bu çalışmalar; metin madenciliği

tekniklerinin inşaat enformasyon sisteminde kullanılabileceğini göstermiş ve bu çalışmada gerçekleştirilecek vaka çalışması için yaptığımız kıyaslama çalışmasına temel teşkil etmiştir. Her bir çalışmada, boyut azaltımı, özellik sayısı, vaka sayısı, terim ağırlığı, sınıflandırıcı şekli açısından benzer yöntemlerle testler yapılmıştır. Tablo 1’ de; bu üç çalışmada alınan en başarılı sınıflandırma sonuçları gösterilmiştir. Bu tabloda metin kategorizasyon çalışmalarında kullanılan 4 farklı parametre karşılaştırılmıştır. Bu parametreler vaka çalışmasının anlatıldığı Bölüm 3’de açıklanmıştır. Bölüm 3’de bu başarılı çalışmalar esas alınarak; bu vaka için bu sonuçlarda kullanılan parametreler ile hazırlanmış vaka çalışması aşamaları ve sonuçları yer alacaktır.

Tablo 1: İnşaat Enformasyon Teknolojisinde Metin Madenciliği Teknikleri Çalışmalarında En Başarılı Sonuçlar

	Vaka sayısı (a)	Boyut Azaltımı (b)	Terim Ağırlığı (c)	En iyi Sınıflandırıcı (d)
(Fan ve Li) (2011)	360	LSI	TF*IDF	KNN (Cos. sim.)
(Caldas ve Soibelman)(2003)	3030	No DR	TFC	SVM
(Caldas ve diğ.)(2002)	845	IG	TFC	SVM

- DR : Boyut Azaltımı
LSI : Saklı Anlam İndeksleme Metodu
IG : Bilgi Kazancı
TF*IDF: Terim Frekansı*Ters Doküman Frekansı Terim Ağırlığı
TFC : Normalize Edilmiş Ters Doküman Frekansı Terim Ağırlığı
KNN : k-En Yakın Komşuluk Algoritması
SVM : Destek Vektör Makineleri Aloritması

3. VAKA ÇALIŞMASI

3.1. Veri Dokümanlarının Hazırlanması ve Derlenmesi (a)

Bu çalışmada; metin madenciliği tekniklerini kullanabileceğimiz veri setleri, Türkiye’deki kamu inşaatlarında yaşanan uyuşmazlıklarda bir nevi tahkim görevi gören karar verici en üst kurum olan Yüksek Fen Kurulu’nun uyuşmazlık kararları olarak seçilmiştir. İlk etapta, Yüksek Fen Kurulu arşivlerinden basılı kopya şeklinde alınan 82 adet doküman, çalışmanın verimli bir şekilde yürütülebilmesi açısından belli tasniflerden geçirilmiştir. Bazı inşaat projelerinde işin doğası gereği uzlaşılammış birden fazla uyuşmazlık yaşanabilmektedir. Göze alınan bu kriterle; birden fazla uyuşmazlık konusu içeren dokümanlar değerlendirmeye alınmamıştır. Konu bütünlüğü açısından uyuşmazlık tipi tam olarak

kestirilemeyen ya da makinenin anlamlandıramayacağı örüntüler içerdiği düşünülen dokümanlar ve ayrıca; konusu genel olarak kategorilendirilemeyecek, bazı özel projelerin kendine has uyumsuzluklarını barındıran dokümanlar da elenmiştir. Böylece 49 adet doküman çalışmada kullanılmıştır. Basılı kopya halinde bulunan dokümanlar gerekli dönüşümler yapıldıktan sonra dijital kopya haline ve .txt formatına getirilmiştir. Daha sonra dokümanlar klasik olarak noktalama işaretlerinin kaldırılması, kelime bölümlenme, kelime gövdeleme, etkisiz kelimelerin çıkarılması gibi doğal dil işleme tekniklerine (NLP) tabi tutulmuştur. Bu çalışmada NLP işlemleri; Türkçe metinler için hazırlanmış paket bir yazılım olan "txt2arff" [3] kullanılarak gerçekleştirilmiştir. Dokümanların her birinin yalnızca bir uyumsuzluk tipini ifade edecek şekilde ayrılmasıyla, her dokümanı etiketleyeceğimiz uyumsuzluk konularının kategorilerine ayrılması işlemi sürecinde; tüm dokümanlar, konu, talepler ve verilen karar açısından analize tabi tutulmuştur. Böylece en sık yaşanan uyumsuzluk tipleri kategori adlarını oluşturacak şekilde belirlenmiş ve süreçte uzman görüşü olarak kamu inşaat sektör katılımcılarının esas profilini oluşturan kamu görevlisi kontrol mühendisleri, kamu işi yapan yüklenici temsilcileri de aktif olarak katkıda bulunmuştur.

3.2. Boyut Azaltımı(b), Terim Ağırlıklandırma (c) ve Sınıflandırıcılar (d)

Dokümanlardan elde edilen terimler ile dokümanlar arasında oluşan vektör uzay matris görüntüsü Şekil.1' de verilmiştir. Terimlerin fazla büyüklükte olduğu metin madenciliği çalışmalarında özellik seçme ya da özellik çıkarma yöntemleri uygulanır. Özellik seçim yöntemleri dokümanlardan elde edilen özelliklerden en kaliteli ve bilgi içeren özellikleri seçmeye dayanırken, özellik çıkarımı ise; özellikleri dokümanlardan elde etmeden aralarında eş anlamlılık, eşseslilik, çok anlamlılık ilişkisi bulunan yeni terimlerden birleştirme yöntemine dayanır. Bunlardan özellik seçme yöntemleri sıklıkla uygulandığına göre; IG, CHI gibi yöntemlerdir. Özellik çıkarımında ise daha çok Saklı Anlam İndeksleme, Terim Kümeleme gibi yöntemler uygulanır. (Caldas ve diğ)(2002) çalışmasında; (Yang and Pedersen)(1997) ve (Joachims)(1998)'in çalışmalarına atfen IG(Bilgi Kazancı) yöntemini seçmiştir. Ancak (Caldas ve Soibelman)(2003) çalışmasında ise; bir önceki çalışmada seçtikleri IG yöntemine göre bir boyut azaltma yöntemi uygulamadan daha başarılı sonuç elde etmiştir. Onlar bu durumu sürpriz karşılamış ve (Joachims)(1998)' in belirttiğine dayanarak sadece az sayıda ilgisiz indekslenmiş terim bulunması ile açıklamışlardır. Bu sonuçlardan kıyasla, bu çalışmada herhangi bir özellik çıkarımı olmadan ya da LSI(Saklı Anlam İndeksleme) özellik çıkarımı ile daha başarılı sonuçlar alındığı düşünülerek boyut azaltmanın bu iki yöntemi ile bir farklılık yaratmak üzere ve Türk kamu inşaat sektörü terminolojisinin farklı yapısına vurgu yaparak özelliklerin el yordamıyla seçilmesi ya da oluşturulması ile elde edilen bir sözlük ile diğer boyut azaltımı yöntemleri karşılaştırılmıştır. Terim ağırlıklandırma yöntemlerinden ise; diğer çalışmalarda anıldığı şekilde, $tf^* tf-idf$ ve tfc kıyaslanmıştır.

$$N = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1j} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2j} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3j} \\ \dots & \dots & \dots & \dots & \dots \\ a_{i1} & a_{i2} & a_{i3} & \dots & a_{ij} \end{pmatrix}$$

i : Doküman sayısı
 j : Nitelikli kelime sayısı
 a_{ij} : Terim ağırlığı

Şekil 1: Dokümanlar ve Terimler Arasında Terim Ağırlıklandırma Matrisi

Metinler nihayetinde; kıyaslanan çalışmalardaki başarılı sınıflandırma algoritmaları ile (kNN,SVM) ve bunlara ek olarak Karar Ağacı Algoritması (DT) ve Naive Bayes (NB) gibi farklı algoritmalarla WEKA 3.6 programı ile sınıflandırma testlerine tabi tutulmuştur.

3.3. Sonuçlar , Tartışmalar ve Yeni Bir Yaklaşım

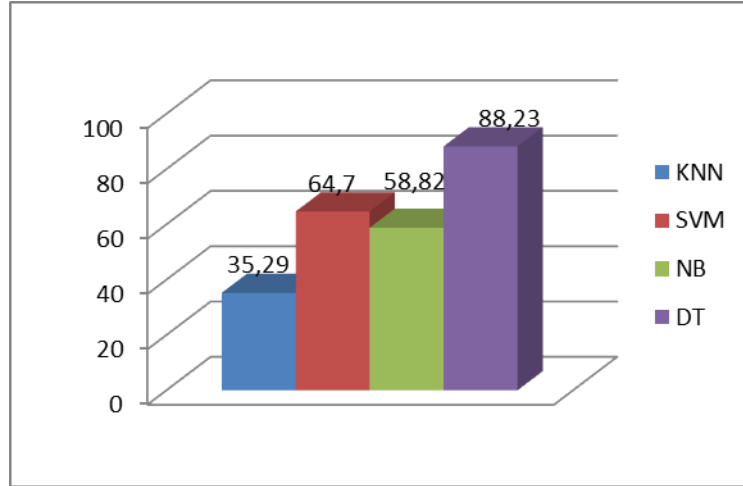
Farklı boyut azaltımı, terim ağırlığı, sınıflandırıcı kombinasyonlarıyla yapılan testlerden alınan en başarılı sonuç Tablo 2' de sunulmuştur. Buna göre; manuel özellik seçimi, tfc terim ağırlıklandırma ve DT sınıflandırıcı konfigürasyonu ile yapılan test %88.23 ile en başarılı sonucu vermiştir. Farklı sınıflandırıcılar ile yapılan test başarıları Tablo 3' te ve grafik görseli Şekil 2' de sunulmuştur.

Tablo 2: En başarılı konfigürasyon sonucu

Boyut Azaltımı	Vaka Sayısı	Terim Ağırlığı	Sınıflandırıcı
Manuel Seçim	49	TFC	DT

Tablo 3: Manuel seçim, TFC ağırlıklandırma yöntemi ve farklı sınıflandırıcılar ile yapılan test sonuçları

k-NN	SVM	NB	DT
35.29	64.70	58.82	88.23



Şekil 2: Manuel özellik seçim, TFC ağırlıklandırma yöntemi ile yapılan test sonuçları grafik görüntüsü

Tüm test sonuçları incelendiğinde; Türk kamu inşaat sektöründe yaşanan uyuşmazlıklar için hazırlanan raporların kategorilendirilmesinde diğer çalışmalardan farklı olarak bir sözcük listesi hazırlanıp bu liste üzerinden otomatik sınıflandırmayı sağlamak, yani manuel özellik seçim yöntemi, daha iyi sonuç vermiştir. Bu da resmi bir şekilde dokümanlarda geçen Türkçe inşaat terminolojisinin belli bir matematik algoritmaya dayanan boyut indirgeme yöntemine uzak kalması veya kendine has doğası ile açıklanabilir. Terim ağırlıklandırma yönteminde normalize edilmiş bir terim ağırlıklandırma yöntemi; (Caldas ve Soibelman)(2003)' in çalışmasında olduğu gibi daha efektif sonuç sağlamıştır. Sınıflandırıcılar arasında DT ' nin daha iyi sonuç vermesi sürpriz olarak karşılanabilir ancak yine; bu vaka çalışmasındaki farklı inşaat çevresinin ve tarzın etkili olduğu söylenebilir.

Bu çalışma ile beraber dört çalışma göz önüne alınarak yapılan değerlendirme ile tüm çalışmaların kabaca farklı sonuçlar elde ettiğini söylemek mümkündür. Buradan hareketle Türk kamu inşaat projelerinin farklı iklimi, sözleşmelerinin ise farklı yapısı göz önünde bulundurularak, daha iyi sonuçların alınması amacıyla ve bir sonraki çalışmamızda uygulayacağımız hiyerarşik doküman sınıflandırma ve vaka geri çağırım modelinin sağlıklı işleyebilmesi açısından, Türkçe inşaat uyuşmazlık dokümanlarında uygulanacak madencilik tekniklerinin bu çalışmada başarılı sonuçlar elde edilen boyut azaltma, terim ağırlıklandırma ve sınıflandırıcı tekniklerinin hibritlendiği bir yaklaşım düşünülebilir. Örneğin boyut azaltma yöntemlerinin farklı kombinasyonlarının denendiği testlerde bir tür özellik seçim yöntemi olan Information Gain (IG) de oldukça iyi sonuçlar vermiştir. Bu şekilde manuel seçim ve IG' nin beraber düşünüldüğü, terim ağırlıklandırılmasında normalizasyonun artırıldığı bir modelin verimli olabileceğini de düşünmek mümkündür. Bununla beraber klasik olarak terimler uzayında ağırlıklandırılan dokümanların benzerlikleri üzerinden yapılan sınıflandırmalardan farklı olarak uyuşmazlık kategorileri

uzayında terimleri ağırlıklandırarak, yeni dokümanların bu terim ağırlıkları üzerinden sınıflandırılmaya çalışılmasının efektif sonuçlar üretebileceği düşünülmektedir. Bir sonraki çalışmada sunacağımız otomasyon modelinin prototipi bu ilkelere dayanarak hazırlanmış ve aynı veri setinin genişletilmesiyle yapılan ön testlerinde daha başarılı sonuçlar elde edilmiştir.

4. SONUÇ

Bu çalışmada; inşaat dokümanlarında metin madenciliği tekniklerinin uygulandığı önceki üç çalışma esas alınarak bir kıyaslama ve buna dayanarak yeni bir model yaklaşımı aktarılmıştır. Kıyaslama kısmında diğer üç çalışmadan elde edilen farklı sonuçlar irdelenmiş ve alınan en başarılı sonuçların elde edildiği farklı parametreler altında benzer testler Türk kamu inşaat projelerinde yaşanan uyuşmazlıklar ile ilgili bir tahkim görevi gören Yüksek Fen Kurulu'ndan elde edilmiş karar dokümanlarına uygulanmış ve diğer çalışmalardan farklı sonuçlar elde edilmiştir. Bu durum; Türk kamu inşaat sözleşmelerinin farklı yapısına, Türkiye' deki inşaat projelerinin kendine özgü doğasına dayandırılmıştır. Buradan yola çıkarak; Türk kamu inşaat sektörünün uyuşmazlık dokümanlarına uygulanacak metin madenciliği tekniklerinin önceki çalışmalar ile bu çalışmayı harmanlayan parametrelerle hibrit bir model olması önerilmiştir. Bu şekilde tasarlanacak model ile nihai hedefimiz Türk Kamu İnşaat Projeleri yüklenici ve işverenlerin yapacakları vaka geri çağırım çalışmasını kolaylaştırmak ve Yüksek Fen Kurulu Kararlarını tahmin edebilmesini sağlayacak bir otomasyon modeli geliştirmektir.

KAYNAKLAR

- [1] Türkiye İstatistik Kurumu. (2012). *Dava konusuna göre idare mahkemelerinin iş durumu*, www.tuik.gov.tr
- [2] T.C. Adalet Bakanlığı Adli Sicil ve İstatistik Genel Müdürlüğü.(2012)., *Kayıtlar ve İstatistikler*, www.adlisicil.adalet.gov.tr
- [3] <http://www.kemik.yildiz.edu.tr/?id=29>
- Arditi D., Oksay F.E., Tokdemir B.O. (1998). Predicting the Outcome of Construction Litigation Using Neural Networks, *Computer-Aided Civil and Infrastructure Engineering* , 13 75–81
- Caldas C., Soibelman L.(2002). Automated Classification of Construction Project Documents, *Journal of Computing in Civil Engineering*
- Caldas C., Soibelman L. (2003). Automating Hierarchical Document Classification for Construction Management Information Systems , *Automation in Construction*
- Fan, H., Li H. (2012). Retrieving similar cases for alternative dispute resolution in construction accidents using text mining techniques, *Automation in Construction*
- Sebastiani, F. (1999). Machine learning in automated text categorization, *Tech. Rep.*, IEI-B4-31-1999
- Türkoğlu, F., Diri, B., Amasyalı, M.F. (2007). Author Attribution of Turkish Texts by Feature Mining, *Third International Conference on Intelligent Computing, ICIC 2007, Qingdao, China, 2007, LNCS Volume 4681/2007 Springer*