

Veri Madenciliği Üzerine Endüstriyel Bir Durum Çalışması

Buse Türkoğlu^{1*}, Murat Komesli², Mehmet Süleyman Ünlütürk²

¹ Norm Group, ² Yaşar University, * Corresponding author, buse.turkoglu@norm-fasteners.com.tr

Özet

Endüstri şirketleri, üretimdeki hatalarını ortadan kaldırmak ve pazardaki sürdürülebilirliklerini korumak için çözümler aramaktadırlar. Verimliliğin artması için tüm üretim süreçlerinin kesintisiz olması gerekmektedir. Bu nedenle, endüstriyel sistemlerde kullanılan makinelerin oluşturduğu verilerin bilgiye dönüştürülmesi şart olmuştur. Bu çalışmada, soğuk dövme makinelerinden gelen bilgileri analiz eden veri madenciliği algoritmaları karşılaştırılmaktadır. Elde edilen sonuçlara göre, soğuk dövme makinelerinin çalışma duraksamalarının sayısının azaltılmasında veri madenciliği yöntemleri umut verici yöntemler oldukları düşünülmektedir.

Anahtar kelimeler: Veri madenciliği, Endüstriyel sistemler, Kestirimci bakım, Soğuk dövme makineleri, Hata tahmin.

Title

An Industrial Case Study on Data Mining

Abstract

Nowadays, companies are looking for solutions to eliminate stops in production and maintain their sustainability in the market. In order to increase productivity, all production processes must be non-stop. To attain this goal, it is imperative to convert the industrial machine data into information. In this study, data mining algorithms which analyze the information obtained from cold forging machines are studied. According to the results, it is thought that this work can be a promising method in reducing the number of working stops in cold forging machines.

Keywords: Data mining, Industrial systems, Predictive maintenance, Cold forging machines, Error estimation.

Citation: Türkoğlu, B., Komesli, M., Ünlütürk, M. S. (2018, October) *Veri Madenciliği Üzerine Endüstriyel Bir Durum Çalışması*. Paper presented at the Fifth International Management Information Systems Conference.

Editor: H. Kemal İter, Ankara Yıldırım Beyazıt University, Turkey

Received: August 19, 2018, **Accepted:** October 18, 2018, **Published:** November 10, 2018

Copyright: © 2018 IMISC Türkoğlu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Giriş

Çalışmanın amacı, endüstriyel bir şirketteki soğuk dövme makinelerinde, kümeleme, sınıflandırma ve karar ağacı gibi veri madenciliği yöntemlerini kullanarak gerçek hayattaki veri setlerindeki makine duraksamalarını tahmin etmektir. Çalışmada 2014 ile 2017 yılları arasındaki veriler kullanılmıştır. Veriler öncelikle temizlenmiş, ilgili özellikler seçilmiş ve son olarak veri madenciliği algoritmaları kullanılarak işlenmiştir. Bu bildirinin amacı, makinelerin verimli çalışmasını sağlamak ve planlanmamış duraklamalar nedeniyle üretim sürecinin kesintiye uğramamasına yardımcı olmaktır. Böylelikle, ortaya çıkan ürünün kalitesi artacak ve müşteri şikâyetleri azalacaktır.

Literatür çalışmalarına bakıldığında Bastos, Lopes ve Pires (2012) yaptıkları çalışmada, benzer makinelerden veri toplayarak bir veritabanında saklamışlar ve veri madenciliği yöntemiyle hataları tahmin ederek verimliliği artırmaya çalışmışlardır. Bu bildiride, analiz edilecek verilerin türüne bağlı olarak farklı algoritmaların uygulanabileceği belirtilmektedir.

- Karar Ağaçları: Kararlar, kurallar kullanılarak ağaç yapısı şeklinde dallanır. Bu yöntem sayısal değerler için kullanılmaktadır.
- Yapay Sinir Ağları (YSA): Bu yapı insan beyninin öğrenme ve kendini düzeltme gibi süreçlerini taklit eder. Veri madenciliğinde en çok kullanılan yöntemdir.
- Genetik Algoritmalar: Bu teknik tüm veri türleri ile çalışabilir. Genellikle optimizasyon problemleri için kullanılmaktadır.
- Kümeleme: Bu tekniğe göre veri kümesindeki benzer veriler, ortak alan veya sınıfa göre gruplandırılırlar.
- Görselleştirme: Bu teknik, ortaya çıkan sonuçları kullanıcıya daha iyi anlatabilmek için kullanılan görselleştirme işlemidir.

Literatürde incelenen makalelerde kullanılan veri madenciliği yöntemleri C5.0, Yapay Ağ, Apriori ve K-Means'dır. Bastos, Lopes ve Pires (2014), endüstriyel birimlerde bakım faaliyetleri modeli oluşturarak hataları tahmin etmeyi amaçlamışlardır. Veriler analiz edilirken, sonuçlar RapidMiner programı kullanılarak karşılaştırılmıştır. Araştırma dört aşamadan oluşmaktadır:

1. Problemi Anlamak: Karar destek sistemi, alarmlar, eylemler ve tahminler olarak üç bölüme ayrılmıştır. Bastos, Lopes ve Pires (2014) tarafından yayınlanan makale, üretimin verimliliğini ve sürekliliğini arttırmaya odaklanmaktadır.
2. Veri Hazırlığı: Tutsarsız, eksik ve kirli veriler çalışmalardan ayıklanarak daha verimli sonuçlar alınması sağlanmıştır. Veri nesnelerine ve kurallara sahip rastgele bir makine, performansı etkilemeden sonuçlara ulaşabilmek için seçilmiştir.
3. Veri Madenciliği: Toplam 529 kayıta 10 özellik veri kümelerinde işlenmiştir. Eğitim setinde 8 ay boyunca toplanan 330 kayıt bulunmaktadır. Test setinde ise, 3 ay boyunca tutulan 199 kayıt bulunmaktadır. İşlenen veri madenciliği modelinin güven aralığı 0.25 ve modelin doğruluk yüzdesi %71.3'tür.
4. Değerlendirme ve Yorumlama: Model test sonuçları incelendikten sonra, istenen güven ve doğruluğu sağlamıştır. Testler, örnekleme amacıyla bir makinede gerçekleştirilmiştir. Bununla birlikte, önerilen yaklaşım deneysel sonuçlara göre doğrulanmıştır.

Jean Nakamura'nın 2007 yılında yayınlanan makalesi, geçici titreşim veri sonuçlarına bakarak makinelerin başarısızlığını tahmin etmektedir. Endüstriyel makinelerde arızayı hesaplamının zorluğu, arızanın doğrusal olmamasıdır. Bu nedenle, makine arızalarını tahmin etmek için mevcut bir araç veya algoritma bulunmamaktadır.

Verilerin test edilmesi ve öğretilmesi için sekiz sayısal sütun kullanılmış, CaseID benzersiz değer olarak seçilmiştir. Veri hazırlama için ortak veriler gruplandırılmıştır. Veritabanının iki kopyası test ve eğitim için oluşturulmuş ve verilerin %17'si test için ayrılmıştır. Ekipman üreticileri, verimliliği artırma ve servis maliyetini azaltmada zorluklar yaşamaktadır. Tahmine dayalı bakım stratejisi, en uygun zamanı planlayarak bu sorunları çözmektedir. Bununla birlikte, tahmine dayalı bakım stratejisinde iki zorluk vardır. Makineler hakkında bilinen birkaç özellik ile büyük veri toplandığında bunları işlemek için makine gücü yoktur. Çözüm olarak veri madenciliği ve prognostik sistemler önerilmiştir. Bu yöntemdeki ilk adım, ilgili özellikleri en uygun düzeyde belirlemek için, veri madenciliği tekniği tarafından üretilen verileri incelemektir. İkinci adım, bu modeli öngörülen özellikler için kullanmak ve eylemi canlı verilerde gözlemlemektir.

Yöntem

Veri madenciliği, görüntü işleme, otomasyon, kontrol sistemleri mühendisliği, en iyileme vb. gibi birçok uygulamada kullanılmaktadır. Her araştırma alanı için algoritmalar aynı olmadığından, çeşitli veri madenciliği uygulamaları bulunmaktadır. Bu çalışmada, endüstriyel makinelerdeki duraksamaları önlemek için kullanılan veri madenciliği yöntemleri sunulmaktadır. Veri madenciliği yöntemleri iki şekilde incelenir: Öngörücü ve açıklayıcı. Veri madenciliğinde yaygın algoritmalar; sınıflandırma, tahmin, zaman serileri analizi, ilişkilendirme, kümeleme ve özetlemedir.

Sınıflandırma algoritmasında, sınıfa ilişkin veriler etiketleme olarak adlandırılır. Sınıflandırma algoritması, bu etiketleri öğrenerek sınıflandırma modeli oluşturur. Sonra, test verileri ile eğitim setine göre tahminlemek için işlenir. Bu çalışmada, verilerin sınıflandırılmasında bakım hatası kayıtları incelenerek "evet / hayır" olarak tahmin edilmiştir.

Tahminleme modellerinde, regresyon algortimaları kullanılır. Bu algoritmalar regresyon ve lojistik regresyon analizi olarak ikiye ayrılmaktadır. Lineer regresyon formülü (1)'de verilmiştir (Khan, 2007):

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon \quad (1)$$

Bu formüle göre Y_i değerinin sıfıra eşit olması istenir. Test veri setinde, öğrenilen doğrusal grafik modeline göre ilişki kurulmaktadır. Modeldeki hata gerçekleşen veri ile tahminlenen veri arasındaki farktır ve ϵ ile gösterilir.

Veriler arasında doğrusal ilişki olmadığı durumlarda ise, lojistik regresyon analizi kullanılır (Multiple Linear Regression Formula). Hata formülü ise, tahmin edilen verilerin koordinatları ile gerçek değerler arasındaki mesafe farkı olarak ölçülmektedir.

$$observed \rightarrow y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p + \epsilon \quad (2)$$

$$predicted \rightarrow y' = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p \quad (3)$$

$$error \rightarrow \epsilon = y - y' \quad (4)$$

Zaman serilerinde, geçmiş zamanın verileri alınarak hesaplanan değer eğilimleri göz önünde bulundurulmakta ve geleceğe yönelik tahmin süreci gerçekleştirilmektedir. Gelecekte verilerdeki olası senaryolar, veri kümesinin özellikleri dikkate alınarak benzetim yapılmaktadır. Bankacılık sektöründe kredi başvurularını kabul etme kararları veya yeni bir ürün için pazar araştırması, zaman serilerinde yapılabilir.

İlişkilendirme algoritmasında, iki değişken arasındaki korelasyona bakılarak anlamlı bir ilişki kurulmaya çalışılmaktadır. Böylece, nitelikler arasındaki ilişkiye göre anlamlı veriler oluşturulabilmektedir. Bu algoritmalar, genellikle satış ve pazarlama stratejilerinde kullanılır.

İlişkilendirme algoritması, iki formüle (5), (6) göre çalışır. $I = \{i_1, i_2, \dots, i_n\}$ veri seti için güven formülü aşağıdaki gibidir:

$$X \Rightarrow Y : (X \cap Y) / N \quad (5)$$

N: Veri kümesindeki toplam öge sayısı

Bu formülde her bir veri için veri kümesindeki tekrarlanma oranı hesaplanır. Daha sonra, ikili değerler arasındaki güven formülü hesaplanarak ilişkisel formül bulunur. İlişkisel formül aşağıdaki gibi gösterilir.

$$X \Rightarrow Y : support(X, Y) / support(X) \quad (6)$$

Böylece, yeni bir veri seti iletildiğinde, bağıntı modeline göre tahmin yapılır (Lai and Cerpa, 2001). Özetleme fonksiyonu genellikle metin madenciliğinde kullanılmaktadır. Sosyal ağlardaki metinlerin analizinde özetleme fonksiyonu ile birlikte gruplama, duygu analizi, fikir madenciliği çalışmaları ağırlıklı olarak yapılmaktadır. Özetleme fonksiyonunun önemli bir özelliği, büyük veri seti ile çalışmasıdır. Literatürdeki en yaygın gösterimde düğümler arasındaki ilişkiler için grafik teorisi kullanılmaktadır (Şeker, 2015). Makine üzerindeki arıza, bir veya daha fazla değişkenin etkilendiği sistemi durdurma veya duraksama olarak adlandırılır. Hataları tespit etme veya tahmin etmenin en yaygın yolu, yazılım ve donanımdan toplanan bilgileri analiz etmektir. Bilginin doğru analizi ve hatanın erken tespiti, sistemin sürekliliği

açısından kritik öneme sahiptir (Amooee, Minaei-Bidgoli, ve Bagheri-Dehnavi, 2011).

Hata algılama, anormal işlem davranışlarını tespit etmek olarak tanımlanır. Hata tespit teknikleri iki kategoriye ayrılır: Modele Dayalı Yaklaşımlar ve Veri Tabanlı Yaklaşımlar (Cuguer, Escofet, vd, 2016).

Model tabanlı yaklaşımlardan farklı olarak, veritabanı tabanlı algılama yaklaşımlarında, büyük miktarlarda süreç geçmiş gibi bir ön işlem bilgisi gerekmektedir. Veriler, sistemde nicel ve nitel olarak iki farklı şekilde tanımlanabilir. Nitelik bilgisi, uzman sistemler ve QTA (Kalitatif Eğilim Analizi) yöntemleri ile elde edilebilir. Nicel bilgiler ise, teknik, sınır ağları ve istatistiksel yöntemler ile öğrenilmektedir (Gertler, 1988).

Bulgular

Bu bölümde, hata tespiti için uygulanan farklı yaklaşımlar sunulmuş, karşılaştırılan algoritmalar arasında verimliliği en üst düzeye çıkaran ve en güvenilir sonucu veren algoritma belirlenmiştir. Karşılaştırma çalışmasının adımları aşağıdaki bölümlerde gösterilmiştir. Veri hazırlama aşaması, veri madenciliğinin en önemli aşamasıdır. Çalışmanın bu bölümünde veri seti, veri madenciliği algoritmaları için kullanılabilir forma dönüştürülmüştür. Bu nedenle, veri madenciliğinde kullanılacak veri seti, makinede üretilen ürünün karakteristik özellikleri ile başlamaktadır. Daha sonra, ulaşılan veri setindeki sapma değerleri belirlenmiş, verilerdeki boş değerler silinmiş, tutarsız değerler elle düzeltilerek temizlenmiş ve sabit değişkenler göz ardı edilmiştir. Her bir satır için bakım yapıp yapılmadığı bilgisi de son olarak eklenmiştir. Bazı algoritmalar, ".csv" uzantılı dosyaları, bazıları ".xls" veya ".xlsx" biçimindeki dosyaları işleme almaktadır. Bu nedenle, veri hazırlamanın son aşaması, verileri içeren dosyanın formatlanmasıdır. Elde edilen veri seti, Tablo 1'de sunulmuştur.

Tablo 1. Çalışmada kullanılan örnek veri.

DSINIFI	AOLCU	BOYM	OZELURUN	NUMUNE	QTY	HMDCINS	HMD_TAV	HMD_TIP	BAKIM
10.9	18	55	1	0	19000	2	0	1	0
10.9	18	55	1	0	19000	2	0	1	1
8.8	24	70	0	0	50000	8	0	1	0
8.8	24	70	0	0	50000	8	0	1	0
8.8	24	70	0	0	50000	8	0	1	0
8.8	24	70	0	0	50000	8	0	1	0
8.8	24	70	0	0	50000	8	0	1	0
8.8	24	80	0	0	40000	8	0	1	0
8.8	24	50	0	0	40000	8	0	1	0
8.8	24	50	0	0	40000	4	0	0	0
8.8	24	50	0	0	40000	8	0	1	0
8.8	24	50	0	0	40000	4	0	0	0
8.8	24	60	0	0	60000	8	0	1	1
8.8	24	60	0	0	60000	8	0	1	1
8.8	24	60	0	0	60000	8	0	1	0
8.8	24	60	0	0	60000	8	0	1	0
8.8	24	55	0	0	12000	8	0	1	0
8.8	24	55	0	0	12000	8	0	1	0
8.8	24	65	0	0	15000	8	0	1	0
8.8	24	65	0	0	15000	8	0	1	0
8.8	24	90	0	0	18000	8	0	1	0
8.8	24	90	0	0	18000	8	0	1	0
8.8	24	100	0	0	20000	8	0	1	0
10.9	20	60	1	0	10000	7	1	1	0
10.9	18	55	1	0	19000	2	0	1	0

Veri kümesinde dokuz özellik ve bir sonuç sütunu bulunmaktadır. DSINIFI, üretilen malzemenin dayanım (mukavemet) sınıfını belirleyen sayısal bir değerdir. Bu öznelikteki değerler şunlar olabilir: 8.8, 10.9 veya 12.9. AOLCU sütunu da 14 ve 24 arasında değişkenlik gösteren ve malzemenin anma ölçüsünü belirten sayısal bir değerdir. BOYM sütununda üretilen malzemenin boyunu gösteren başka bir sayısal değerdir ve 36 ila 297 arasındadır. OZELURUN sütunu, 0 veya 1 olarak değerlendirilen bir boolean değeridir. Bu sütun malzemenin standart ya da özel ürün olduğunu belirtir. Özel veya standart ürün

ayrımı malzeme üretilirken ortaya çıkacak verimlilik hesabı için önem arz etmektedir. Özel ürünlerin tolerans değerleri, özellikler ya da kullanılacak kalıpların ömürleri gibi teknik detayların daha hassas olması nedeniyle veri setine eklenmiştir. Üretici firmadan öğrenilen deneyimlere göre standart ürünlerde arıza oranı özel ürünlerden daha azdır. Bu nedenle OZELURUN sütununun arıza bakımını etkileyen kritik değerlerden biri olarak seçilmiştir. NUMUNE sütunu, numune olarak üretilen veya seri üretimdeki malzemeyi gösteren boolean bir değerdir. QUANTITY değeri, üretim miktarını gösteren sayısal bir değerdir. HMDCINS, HMD_TAVLAMA ve HMD_TIP kolonları, makinede üretilen ürünün hammadde hakkında bilgi vermektedir. BAKIM sütunu ise, hata kaydının değerini Boolean olarak tutmaktadır.

WEKA programı kullanılarak beş farklı model karşılaştırılmıştır. Bu algoritmalar, doğru sınıflandırılmış örneklerle ve RMSE (Root Mean Squared Error) formülüne göre karşılaştırılmıştır. Sonuç olarak, en verimli algoritma %77 doğrulukla J48 algoritması seçilmiştir.

Tablo 2. Karşılaştırılan modeller.

	Correctly Classified Instances	Root Mean Squared Error
Algoritmalar	Doğru Sınıflandırılan Örnekler	Etkin Hata Değeri
IB1 instance-based classifier	48.8722 %	0.6494
Naive Bayes Classifier	73.4962 %	0.4472
J48 pruned tree	77.4436 %	0.418
KStar Beta Version	60.1504 %	0.5199
Locally weighted learning	76.8797 %	0.4173

Veri sağlayan şirkette malzeme bazlı duraklamaların analiz edilmesi için J48 algoritması seçildiğinde ortaya çıkacak doğru tahminlemeler ile zamanında önlem alınabilir. Bu yöntem, zamansız duruşları azaltarak verimliliğin artmasını sağlamaktadır.

Tablo 3'de J48 algoritmasına ait sonuç matrisi gösterilmektedir. Ortaya çıkan sonuçlara göre J48 algoritması kullanılsaydı 28 arıza kaydının önceden yakalanabileceği görülmektedir. Aşağıdaki açıklamalar matristeki değerlerin karşılığıdır:

- True Positives (TP): Arıza var olarak tahmin edildi ve arıza kaydı var.
- True Negatives (TN): Arıza yok olarak tahmin edildi ve arıza kaydı yok.
- False Positives (FP): Arıza var olarak tahmin edildi ancak arıza kaydı yok. ("Tip I hata" olarak bilinmektedir.)
- False Negatives (FN): Arıza yok olarak tahmin edildi ancak arıza kaydı var. ("Tip II hata" olarak bilinmektedir.)
- n: Veri kümesindeki toplam öge sayısı.

Tablo 3. J48 sonuç matrisi.

n = 2028	Tahmin: Evet	Tahmin: Hayır
Gerçek: Evet	TP = 28	FN = 347
Gerçek: Hayır	FP = 119	TN = 1534

Rapidminer'de, kural modeli (Rule Model) aşağıda gösterilmiştir. İlk olarak, veriler karar ağacı modeline bir girdi olarak gönderilmiştir. Ardından, Naïve Bayes sınıflandırma işlevini kullanarak, verilerin sınıfı belirlenmiştir. Oluşturulan kurallar, indüksiyon yöntemi ile tahmin edilmiştir.

Kural Modeli

```

if QUANTITY ≤ 30500 then No (1255 / 436)
if QUANTITY ≤ 68000 and OZELURUN ≤ 0.500 and
BOYM > 65 and AOLCU = 24 then No (41 / 15)
if QUANTITY ≤ 68000 and BOYM > 89.150 and
QUANTITY > 43000 and AOLCU = 16 then No (9 / 3)
if QUANTITY ≤ 53500 and AOLCU = 24 and HMDICINS >
2.500 then No (22 / 14)
if QUANTITY ≤ 68000 and BOYM ≤ 48.250 and BOYM >
40.250 and QUANTITY ≤ 60500 and AOLCU = 16
then No (17 / 10)
if HMD_TAVLAMA ≤ 0.500 and HMDICINS ≤ 3.500 and
AOLCU = 14 then No (2 / 1)
if QUANTITY ≤ 68000 and DSINIFI ≤ 9.850 and
HMDICINS > 2.500 and BOYM ≤ 150 and QUANTITY ≤
47500 and HMDICINS ≤ 11 and BOYM ≤ 140 and
AOLCU = 42871 then No (1 / 0)
if BOYM ≤ 62.500 and HMDICINS ≤ 7.500 and
QUANTITY ≤ 82500 and QUANTITY > 61500 and
AOLCU = 16 then No (6 / 5)
else No (89 / 82)
correct: 1442 out of 2008 training examples.

```

Tüm değerler, "kural modeli" olarak adlandırılan yukarıdaki karar ağacı modeline göre belirlenmiştir. Sonuç olarak, karar ağacı yapısında dallanma olmadığı için tüm sonuçlar "hayır" olarak elde edilir. Ancak, modelin başarısı %71.8 olup, 2008 kayıtlarından 1442 vakayı doğru bir şekilde sınıflandırmaktadır. Bu durumda, RapidMiner programı Weka'dan daha iyi sonuçlar elde edememiş, en iyi algoritma J48 modelinde %77'lik bir doğruluk elde edilmiştir. Elde edilen doğruluk düşüktür, ancak bu çalışma veri sağlayıcı şirket için önemlidir. Çünkü daha önce böyle bir çalışma yapılmamıştır ve seçilen veri madenciliği algoritması 2014 ve 2017 arasında kullanıldığında 28 başarısızlık yakalanmıştır. Daha iyi sonuçlar için karar ağacı modelinde parametrelerin değiştirilmesi gerekir.

Tartışma ve Sonuç

Bu çalışmanın amacı, veri madenciliği yöntemleri kullanarak soğuk dövme makinelerindeki duraksamaları tahmin etmektir. İlk olarak, veriler temizlenmiş, normalleştirilmiş ve sabit değişkenler çıkarılmıştır. Daha sonra, modeller Weka ve Rapidminer programları kullanılarak eğitilmiş ve test edilmiştir. Weka'da kullanılan J48 Decision Tree algoritmasının %77 doğruluk ve ~0.4 hata değeri ile en verimli sonucu verdiği gözlenmiştir. Rapidminer programı kullanılarak modellenen karar ağacında ise, %71.8'lik doğruluk sonucuna ulaşılmıştır. Bu nedenle, veri sağlayıcı şirket için Weka programında J48 algoritmasının kullanılması önerilmektedir. Veri sağlayıcı şirkette daha önce benzer bir çalışma yapılmadığından dolayı, karşılaştırılabilecek bir model bulunmamaktadır. Ancak, bu çalışma ile seçilen J48 modelinde, önlenebilecek 28 arızayı doğru bir şekilde yakalanmıştır.

Bu çalışmanın, gelecekte yapılacak ileri bir uygulaması olarak, tüm makine grupları veya tüm malzeme grupları için veri madenciliği fonksiyonları daha ayrıntılı bir şekilde eğitilmesi ve sonuçlarının web tabanlı bir yazılım uygulaması geliştirilerek yayınlanması önerilmektedir.

Kaynakça

- Accorsi, R., Manzini, R., Pascarella, P., Patella, M., & Sassi, S. (2017). Data mining and machine learning for condition-based maintenance. *ScienceDirect* (11), 1153 – 1161. Retrieved from <https://www.sciencedirect.com/science/article/pii/S235197891730447X>
- Amooee, G., Minaci-Bidgoli, B., & Bagheri-Dehnavi, M. (2011). A Comparison Between Data Mining Prediction Algorithms for Fault Detection (Case study: Ahanpishegan co.). *IJCSI International Journal of Computer Science Issues*, 8(3). Retrieved from <https://arxiv.org/abs/1201.6053>
- Bastos, P., Lopes, I., & L. Pires. (2014). Application of data mining in a maintenance system for failure prediction. In *Safety, Reliability and Risk Analysis: Beyond the Horizon*. London, UK.: Taylor & Francis Group. Retrieved from <http://hdl.handle.net/10198/8960>
- Bastos, P., Lopes, I., & Pires, L. (2012). A Maintenance Prediction System using Data Mining Techniques. *World Congress on Engineering* (pp. 1448-1453). London, U.K.: International Association of Engineers. Retrieved from <http://hdl.handle.net/10198/7298>
- Cuguero-Escofet, M., Quevedo, J., Alippi, C., Roveri, M., Puig, V., García, D., & Trovati, F. (2016). Model vs. Data based Approaches Applied to Fault Diagnosis in Drinking Water Transport Networks. *Journal of Hydroinformatics*, 18(5), 831-850. doi:<https://doi.org/10.2166/hydro.2016.218>
- Gertler, J. J. (1988). Survey of Model-Based Failure Detection and Isolation in Complex Plants. *IEEE Control Systems Magazine*, 8(6), 3 - 11. doi: 10.1109/37.9163
- Khan, I. (n.d.). How Linear Regression can also be used to do classification? Retrieved 2018, from <https://analyticsbuddhu.wordpress.com/2016/06/28/how-linear-regression-can-also-be-used-to-do-classification/>
- Koh, H. C., & Tan, G. (2005). Data Mining Applications in Healthcare. *Journal of Healthcare Information Management*, 19(2).
- Lai, K., & Cerpa, N. (2001). Support vs Confidence in Association Rule Algorithms. n Proceedings of the OPTIMA Conference. Chile: Universidad de Talca. Retrieved from https://www.researchgate.net/profile/Narciso_Cerpa/publication/233754781_Support_vs_Confidence_in_Association_Rule_Algorithms/links/0fcfd512a4907b8aca000000/Support-vs-Confidence-in-Association-Rule-Algorithms.pdf
- Multiple Linear Regression Formula. (n.d.). Retrieved from <http://www.saeedayad.com/mlr.htm>
- Nakamura, J. (2007). Predicting Time-to-Failure of Industrial Machines with Temporal Data Mining. University of Washington. Retrieved from <https://s3.amazonaws.com/academia.edu.documents/39357344/jnakamura.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1522678144&Signature=yQQYnOZsmNCNjcoGRSGgLi20blo%3D&response-content-disposition=inline%3B%20filename%3DJnakamura.pdf>
- Prytz, R. (2014). Machine learning methods for vehicle predictive maintenance using off-board and on-board data. In *Halmstad University Dissertations*. Halmstad University Press.
- Rachburee, N., Jantarat, S., & Punlumjeak, W. (2012). Time Series Analysis for Fail Spare Part Prediction: Case of ATM Maintenance. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, 9(2-6). Retrieved from <http://journal.utem.edu.my/index.php/jtec/article/view/2433>
- Rapidminer. (n.d.). Retrieved from <https://rapidminer.com>
- Seker, S. E. (2015). Çizge Teorisi (Graph Theory). *YBS Ansiklopedi*, 2(2), 17-29. Retrieved from http://ybsansiklopedi.com/wp-content/uploads/2015/05/cizge_teorisi.pdf
- Wang, J., Li, C., Han, S., Sarkar, S., & Zhou, X. (2017). Predictive maintenance based on event-log analysis: A case study. *IBM Journal of Research and Development*, 61(1), 11:121 - 11:132. doi:<https://doi.org/10.1147/JRD.2017.2648298>
- Zong, Y. (2017). Maintenance Cost and Residual Value Prediction of Heavy Construction Equipment.