





## Determining the factors that affect the production time in metal industry utilizing data mining methods

Kübra Işık<sup>1,2\*</sup> , Selda Kapan Ulusoy<sup>1</sup> 

<sup>1</sup>Erciyes University, Faculty of Engineering, Department of Industrial Engineering, 38039, Talas, Kayseri, Turkey

<sup>2</sup>Güven Engineering Machinery, 18. Cadde No:18, 38070 Melikgazi, Kayseri, Turkey

### Highlights:

- Determination of factors affecting production time with various classification algorithms
- Reduction of features with different feature selection methods
- Correlation between production time and features that affecting production time

### Keywords:

- Data mining
- Feature selection
- Rule-based classification
- Decision tree-based classification
- Factors affecting production time

### Article Info:

Research Article  
Received: 13.05.2020  
Accepted: 21.03.2021

### DOI:

10.17341/gazimmfd.736659

### Correspondence:

Author: Kübra Işık  
e-mail:  
kubraisik@gmm.com.tr  
phone: +90 352 207 6666 /  
32478

### Graphical/Tabular Abstract

Metrics	Algorithms						
	Random tree	Part	Rep tree	OneR	Jrip	LMT	Decision table
Accuracy rate	97,97	71,87	69,52	62,78	62,67	61,71	61,60
True positive rate	0,98	0,72	0,7	0,63	0,63	0,46	0,62
False positive rate	0,02	0,29	0,32	0,41	0,44	0,45	0,45
F-scale	0,98	0,7	0,67	0,57	0,54	0,52	0,52
ROC curve	0,99	0,85	0,84	0,61	0,61	0,63	0,64

**Table A.** Comparison of evaluation metrics of algorithms after feature selection

**Purpose** The purpose of this article is to show that data mining methods can be used for identifying the factors that affect the production time in manufacturing industry. Then, this information can be used to take the measures for reducing the production time for especially companies that use the make-to-order policy.

### Theory and Methods:

Rule-based and decision tree-based classification algorithms were used to build the data mining models for the identification of the factors. Based on data types, as rule based classification algorithms, part, decision table, jrip, oner; as decision-tree based classification algorithms, random tree, rep tree and lmt are chosen. Data preprocessing methods are applied to organize several production factors. To reduce the number of features several feature selection evaluators are applied to the data. This created several reduced data sets. For each data set and the full data set with all the features, models are built using the selected classification algorithms. Performance measures of the models are compared and the best model is selected accordingly. This methodology is applied to the production data of a machine manufacturing company which uses the make-to-order policy. Weka data mining program was used to build the models.

### Results:

As a result of the study, it was determined that the factors that affect the production time the most are ambient temperature, product quantity, production month, part name, operator name, name of the machine produced, and cnc machine dimensions. The algorithm obtained with the best evaluation metrics result by using classification algorithms with before and after feature selection in the data set is random tree. The decision tree obtained with the random tree were evaluated and recommendations were given to the company as decision support input according to the rules determined to be crucial. Part, decision table, jrip, oner algorithms which are rule-based classification algorithms; random tree, rep tree and lmt were used among decision tree-based classification algorithms. Performance evaluation metrics as accuracy, roc curve, true positive, false positive, f-score obtained before and after feature selection were compared.

### Conclusion:

The aim achieved by using 5 different feature selection evaluators and then 7 different classification algorithms, and the results obtained were compared. It is concluded that, the methods used are very successful in terms of classification evaluation metrics.



## Metal sektöründe üretim sürelerine etki eden faktörlerin veri madenciliği yöntemleriyle tespit edilmesi

Kübra Işık<sup>1,2\*</sup>, Selda Kapan Ulusoy<sup>1</sup>

<sup>1</sup>Erciyes Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, 38039 Talas Kayseri, Türkiye

<sup>2</sup>Güven Mühendislik Makine 18. Cadde No:18, 38070 Melikgazi Kayseri, Türkiye

### Ö N E Ç İ K A N L A R

- Çeşitli sınıflandırma algoritmaları ile üretim süresine etki eden faktörlerin belirlenmesi
- Birden fazla nitelik seçimi yöntemi ile niteliklerin indirgenmesi
- Üretim süresine etki eden faktörler ile üretim süresi arasındaki ilişki

### Makale Bilgileri

Araştırma Makalesi

Geliş: 13.05.2020

Kabul: 21.03.2021

DOI:

10.17341/gazimmfd.736659

### Anahtar Kelimeler:

Veri madenciliği, üretim süresine etki eden faktörler, kural tabanlı sınıflandırma, karar ağacı tabanlı sınıflandırma, nitelik seçimi

### ÖZ

Günümüzün küresel rekabet koşullarında hayatta kalabilmek için işletmeler üretimlerinde düşük teslim zamanı, düşük maliyet, yüksek kalite ve yüksek esnekliği hedeflemek zorundadırlar. Proje bazlı üretim yapan firmaların bu hedeflere ulaşabilmesi için siparişe dayalı üretim yöntemini tercih etmeleri gerekmektedir. Siparişe dayalı üretimde ürünün teslim tarihinde hazır olması büyük önem taşımaktadır. Teslim tarihlerinin azaltılması için üretim süresini etkileyen faktörlerin tespit edilmesi gerekmektedir. Üretim süresine etki eden faktörlerin tespit edilmesi, bu faktörler üzerinde yapılabilecek iyileştirmeleri öngörmeyi sağlayacaktır. Bu çalışmada üretim süresine etki eden faktörlerin veri madenciliği yöntemleri ile belirlenebileceği metal sektöründe üretim yapan bir firmaya uygulanarak gösterilmiştir. Bu faktörler araştırılırken veri madenciliğinden çeşitli sınıflandırma algoritmaları kullanılmıştır. Uygulama sonucunda en iyi sonuçlar random tree algoritması ile elde edilmiştir. Üretim süresine etki eden faktörler parça adı, makine adı, üretim ayı, ortalama sıcaklık, operatör adı, tezgâh boyutu, ürün miktarı olarak bulunmuştur. Uygulama sonucunda üretilen bilgiler ile işletmeye üretim süreçleri için iyileştirme tavsiyeleri verilmiştir. Çalışmada kullanılan ham veriler ek dosyada sunulmuştur.

## Determining the factors that affect the production time in metal industry utilizing data mining methods

### H I G H L I G H T S

- Determination of factors affecting production time with various classification algorithms
- Reduction of features with different feature selection methods
- Correlation between production time and features that affecting production time

### Article Info

Research Article

Received: 13.05.2020

Accepted: 21.03.2021

DOI:

10.17341/gazimmfd.736659

### Keywords:

Data mining, factors affecting production time, rule based classification, decision tree based classification, feature selection

### ABSTRACT

In order to survive in today's global competitive environment, companies must aim for low delivery time, low cost, high quality, and high flexibility in their production. Companies engaged in project-based production should prefer the order-based production method to achieve these goals. For order-based production method, it is very important that the product is ready at the delivery date. To reduce delivery dates, factors affecting production time should be determined. Determining the factors affecting the production time enables companies to plan the improvements that can be made on these factors. On an application, it is shown that data mining can be used to identify the factors affecting production times in metal industry. While investigating these factors, various classification algorithms were used. In the result best evaluation metrics were obtained with random tree algorithm. The features that best express the model used are part name, machine name, month of production, average temperature, operator name, machine size and product quantity. With the information produced, improvement recommendations that can be applied to the production processes are given to the company. The raw data set can be accessed as a supplementary file.

\*Sorumlu Yazar/Yazarlar / Corresponding Author/Authors : \*kubraisik@gmm.com.tr, skapan@erciyes.edu.tr /

Tel: +90 352 207 6666 / 32478

## 1. GİRİŞ (INTRODUCTION)

Veri madenciliği büyük boyutlardaki veriden anlamlı bilgi üretmek için kullanılan bir yöntemdir. Büyük ölçekte veri depolanan hemen hemen her alanda kullanılabileceği için uygulama alanları sürekli artmaktadır. İmalat sanayi de veri madenciliğinin kullanılmaya başlandığı yeni alanlardan birisidir. Birçok bileşenin oluşturduğu imalat sistemleri oldukça karmaşıktır ve kalite yönetimi, tedarikçi yönetimi, işlerin çizelgelenmesi, üretim süreçlerinin iyileştirilmesi gibi alanlarda sürekli çözülmeyi bekleyen problemler bulunmaktadır. Kurumsal işletmeler bu problemlerle ilgili önemli boyutta veriyi veri tabanlarında depolamaktadır, ancak bu veriler çoğunlukla analiz edilmemektedir.

İmalat sanayisinde verimliliğin artırılabilmesi için çözülecek problemlerden biri üretim süresine etki eden faktörlerin belirlenmesidir. Belirlenen faktörler üzerinde gerekli iyileştirmeler yapılarak üretim süresi azaltılabilir. Üretim süresinin azaltılması özellikle siparişe dayalı üretim yapan firmaların sipariş teslimlerindeki gecikmeleri en aza indirmesine yardım edecektir. Üretim sürelerine etki eden faktörlerin belirlenmesi problemi uygulamada üretim mühendislerinin ve makine operatörlerinin gözlemlerinden ve verilerin manuel analizinden (ortalama, standart sapma vb.) yola çıkarak çözülmektedir. Ancak üretim işlemini etkileyen birçok faktör olabilir ve bu faktörlerde birbirleri ile ilişkili olabilirler. İnsan algısı ve manuel hesaplamalarla gözlemlenemeyen bu ilişkiler veri madenciliği yöntemleri ile bilgiye dönüştürülerek üretim süresinin azaltılması için daha iyi çözümler elde edilebilir. Veri madenciliği imalat sanayi problemlerine nispeten yeni uygulanmaya başladığı için literatürde böyle bir çalışma bulunamamıştır. Bu çalışmanın amacı bu eksikliğe işaret ederek üretim süresini etkileyen faktörlerin belirlenmesinde veri madenciliğinin başarı ile uygulanabileceğini metal sektörden bir örnek üzerinde göstermektir.

Veri madenciliğinin imalat sanayindeki farklı uygulamalarına bakmak için derleme makaleleri ön plana çıkan çalışmalardır. Bu alanda geniş kapsamı ile öne çıkan ilk çalışma Harding vd. 'ne aittir [1]. Harding vd. makalelerinde veri madenciliğinin özellikle, üretim süreçleri, işleyiş, arıza tespiti, bakım, karar destekleri, kalite iyileştirme, müşteri ilişkileri ve mühendislik tasarımları konulu uygulamalarını derlemişlerdir. Makalede veri madenciliğinin üretim ile ilişkisi gösterilmek istenmiştir. Çalışmalardaki mevcut boşlukları belirterek üretim planlaması ve atölye kontrolü gibi alanlarda da veri madenciliği tekniklerinin kullanılabileceğini belirtmişlerdir. Gelecekteki araştırmaların mühendislik tasarımları, atölye kontrolü, çizelgeleme, kurumsal kaynak planlama, tedarik zinciri ile ilgili verilerin analiz edilmesine ve bunların kapasitelerini arttırmak için mevcut bilgi tabanlı sistemlerle bütünleşebilecek sistemlerin geliştirilmesine yönelmesini beklediklerini belirtmişlerdir. Choudary vd. [2], veri madenciliğinin imalat sanayisi uygulamalarında hangi tip bilgiye ulaşılmak istendiği sorusu üzerinden bir derleme çalışması yapmışlardır. Ulaşılmak istenen bilgi konsepti

tanımı, birliktelik kuralları, sınıflandırma, kümeleme ve tahmin olarak kategorize edilmiş ve makaleler bu bilgi başlıkları altında gruplandırılmıştır. Usuga Cadavid vd. [3] makalelerinde, Endüstri 4.0 çerçevesinde makine öğrenmesinin üretim planlama ve kontrol alanında uygulanması ile ilgili çalışmaları sistematik bir şekilde derlemişlerdir. Bu makalede amaç kullanılan metodların tanımlanmasına katkıda bulunmak ve yeni araştırma alanlarını belirlemektir. Yazarlar şu araştırma sorularına cevap aramışlardır; üretim planlama ve kontrol konusuna makine öğrenmesini uygulayabilmek için yapılması gereken aktiviteler, kullanılan makine öğrenmesi teknikleri, kullanılan veri kaynakları, çözülen problem tipleri nelerdir, Endüstri 4.0'ın hangi özellikleri kullanılmıştır. Derleme için 93 makale incelenmiş ve araştırma sorularının cevapları incelenen makalelerin yüzdeleri cinsinden verilmiştir. Benzer bir derleme Köksal vd. [4] tarafından veri madenciliğinin kalite iyileştirme problemlerine uygulaması üzerine yapılmıştır. Ele alınan kalite problemleri ürün/süreç kalitesinin tanımlanması, kalitenin tahmin edilmesi, kalitenin sınıflandırılması ve parametrelerin optimizasyonudur. Bu kapsamlı derleme çalışmalarında üretim süresine etki eden faktörlerin belirlenmesine ait bir çalışmaya rastlanmamıştır. Derleme çalışmalarına ek olarak veri madenciliğinin imalat sanayisi uygulamalarına literatürde farklı örnekler vardır.

Sim ve Chan [5], rulmanların seçimi için bilgi tabanlı bir sistem geliştirmişlerdir. Yeni ürünün zamansal verilerini bilgi tabanına göre ayarlayarak tasarım özelliklerini optimize etmek için üreticinin kataloğu tarafından desteklenen buluşsal bilgiyi kullanmışlardır. Romanowski ve Nagi [6], veri madenciliği bilgisinin yaşam döngüsü verilerinden tasarım sürecinin ilk aşamalarına geri bildirimini destekleyen bir tasarım sistemi önermişlerdir. Giess vd. [7], çeşitli denge ve titreşim testleri arasındaki ilişkileri belirlemek, ölçmek ve kritik alanları vurgulamak için bir gaz türbini rotoru imalat ve montaj veri tabanını çıkarmışlardır. İlk aşamada uygun soruşturma alanlarını belirlemek ve verilerle ilgili sorunları belirlemek için bir karar ağacı kullanmışlardır. Bir sonraki aşamada, verileri modellemek için bir sinir ağı kullanmışlardır. Gardener ve Bieker [8], yarı iletken üretimindeki verim problemini çözmek için karar ağacı algoritmaları ve sinir ağları uygulayarak yarı iletken üretiminde önemli tasarruflar sağlamıştır. Park ve Kim [9], takım aşınması, takım kırılması ve takım sapması gibi parça işleme hatalarını en aza indirmek için uygun değer üretim parametrelerini belirlemişlerdir. En uygun çözümleri belirlemek ve bilgi tabanlı uzman sistemler kullanarak çevrimiçi uyarlanabilir kontrol için cad sistemlerine, operasyonel araştırmalara ve hesaplama zekâsına dayanan farklı teknikleri incelemişlerdir. Liao vd. [10], diğer bilgi tabanlı sistemleri aletlerin ve ürünlerin kalitesinin şartlarının yorumlanması için geliştirilmiştir.

Skormin vd. [11], çevresel ve operasyonel koşulların tarihsel verilerine dayanan donanım hatası olasılığının doğru değerlendirilmesi ve tahmini için veri madenciliğini uygulamışlardır. Bilgilendirici alt alanın düşük boyutta

belirlenmesi için bir yaklaşım geliştirmiş ve ardından verileri modellemek için bir karar ağacı kullanmışlardır. Chen vd. [12], yarı iletken üretiminde kusur tespiti için birleştirme kuralları oluşturup, arızalı makineyi belirlemek için farklı makineler ile arızalı kombinasyon arasındaki ilişkiyi belirlemişlerdir. Batanov vd. [13], bilgi temelli bakım sistemlerini araştırmış ve expert-mm adı verilen ve geçmiş arıza verileri üzerinde çalışan ve uygun bir önleyici bakım takvimi için öneriler sunan bir prototip sistemi geliştirmiştir. Çetin [14], bir üretim işletmesinde, üretilen ürünlerin uygunsuz olarak ayrılmasının nedenlerini belirleyerek bu nedenlerin analizi ile uygunsuz ürünlerin sayısını azaltıcı stratejiler geliştirilmesi üzerine çalışma yapmıştır. Tapkan ve Özmen [15], bir iplik üretim tesisinde veri madenciliği yöntemlerinden biri olan sınıflandırma ile kural çıkarımı gerçekleştirmişlerdir. Sınıflandırma öncesinde iplik kalitesine etki eden nitelikler belirlenmiş, Taguchi deneysel tasarım yöntemi ile etkin nitelikleri tespit ederek nitelik seçimi yapmışlardır. Kural çıkarımı aşaması ise hem yanlış sınıflandırma hataları sayısını en küçüklemeyi amaçlayan maliyete-duyarsız, hem de beklenen yanlış sınıflandırma maliyetini en küçüklemeyi amaçlayan maliyete-duyarlı sınıflandırma şeklinde uygulanmıştır. Elde edilen kurallar, kaliteli iplik üretimi için işletmeye yol gösterici özellikte olmuştur.

Bilekdemir [16], makinelerin üretim süresinin tahmini üzerine bir çalışma yapmıştır, bu çalışmayı yaparken veri madenciliği tekniklerinden biri olan karar ağacı C4.5'i algoritmasını kullanmıştır. Türkoğlu vd. [17], soğuk dövme makinelerinden alınan verileri analiz eden veri madenciliği algoritmalarını karşılaştırmışlardır. Elde edilen sonuçlara göre soğuk dövme makinelerinde gözlemlenen çalışma duraksamalarının sayısının azaltılmasında veri madenciliği yöntemlerinin umut verici yöntemler olduğunu belirtmişlerdir. Türker vd. [18], çalışmalarında atölye tipi üretim yapan ve gerçek zamanlı üretim verilerinin kaydedildiği bir işletme kurgusu için yeni siparişlere göre iş merkezlerine dinamik olarak iş atanması problemini ele almıştır. Kurgulanan işletme yapısının benzetim modeli kurulmuş ve üretim verileri elde edilmiştir. Elde edilen veriler veri madenciliği sınıflandırma algoritmaları ile analiz edilerek gecikme potansiyeli olan siparişler belirlenmiş ve bu veriler bir uzman sisteme verilerek gecikmelerin en aza indirilmesi için iş atanması yeniden düzenlenmiş ve gecikme potansiyeli olan işler dış kaynak üretimine yönlendirilmiştir.

Chen vd. [19], çalışmalarında üretim hattının her bir sürecindeki verileri toplayarak ürünün kalite oranını etkileyen faktörleri belirlemek için veri madenciliği yöntemini kullanmışlardır. Kullandıkları veri madenciliği yöntemi, işletmelerin ürün üretim kalitesini ve ürün yeterlilik oranını iyileştirmeleri için yeni bir fikir sağlamış ve bu da işletmelerin daha mükemmel bir kalite yönetim sistemi gerçekleştirmelerine yardımcı olmuştur. Nkonyana vd. [20], büyük hacimli verilerde karşılaşılan problemleri ele almışlardır. Çelik sektörü için verilerde bulunan karmaşıklık azaltmak için veri madenciliği, yapay sinir ağları ve destek vektör makineleri yöntemlerini kullanmışlardır. Kullanılan

modellerin performanslarını değerlendirip rastgele orman yönteminin diğer makine öğrenmesi yöntemlerinden daha iyi performans gösterdiğini tespit etmişlerdir. Cheng vd. [21], çalışmalarında büyük veri çağında veri madenciliği tekniklerinin gelişimini gözden geçirmekte ve veri madenciliği tekniklerinin üretim yönetimindeki gelişmiş planlama ve çizelgeleme, kalite iyileştirme, hata analizi, hata teşhisi gibi uygulamalarını seçip analiz ederek tartışmışlardır. Sonuç olarak veri madenciliği tekniklerinin geleneksel üretim yönetimi yöntemlerine kıyasla daha verimli, doğru ve bağımsız olduklarını ortaya koymuşlardır.

Makalenin geri kalanının planı şu şekildedir: İkinci bölümde kurulan modeller ve model kurulumu için kullanılan yöntemler anlatılmıştır. Üçüncü bölümde, modellerden elde edilen sonuçlar verilmiş ve modeller karşılaştırılmıştır. Dördüncü bölümde ise çalışmanın sonuçları tartışılmıştır.

## 2. VERİ MADENCİLİĞİ MODELİ (DATA MINING MODEL)

### 2.1. Yöntem (Method)

Veri madenciliği, büyük miktardaki verilerden gizli kalmış örüntüleri ve korelasyonları belirleyerek yararlı bilgilerin keşfedilmesi bilimidir [22, 23]. Veri madenciliğinin amacı, geçmiş verilerin analizi ile gelecekteki kararların tahminine yönelik karar verme modelleri oluşturmaktır [24].

Veri Madenciliği sürecinde birbirini izleyen birkaç aşama bulunmaktadır. Bunlar; uygulama alanının ve amacının belirlenmesi, uygun veri kümesinin seçimi, gürültülü ve tutarsız verilerin çıkarıldığı veri temizleme ve ön işleme, dönüştürme, akıllı yöntemler aracılığıyla büyük miktarda veriden anlamlı bilgilerin çıkarılması, çıkarılan örüntülerden yararlı olanların belirlenmesi için değerlendirme ve son aşama olan elde edilen bilginin görüntülenmesi ve bilgi gösterimi yöntemleri kullanılarak kullanıcıya sunulmasıdır. Bu aşamalar Şekil 1'de gösterilmektedir.

Veri madenciliğinde kullanılan modeller, tahmin edici ve tanımlayıcı olmak üzere iki ana başlık altında toplanır [26]. Tahmin edici modellerde, sonuçları bilinen verilerden hareket edilerek bir model geliştirilmesi ve kurulan bu modelden yararlanılarak sonuçları bilinmeyen veri kümeleri için sonuç değerlerin tahmin edilmesi amaçlanmaktadır. Tanımlayıcı modellerde ise veri kümesindeki örüntülerin bulunması amaçlanmaktadır [27].

Üretim süresine etki eden faktörlerin belirlenmesi probleminin çözümü için sınıflandırma modelleri seçilmiştir. Modeller Weka 3.9.4 programından faydalanarak kurulmuştur. Weka programı açık kaynak kodlu ve Java dilinde geliştirilmiş bir makine öğrenmesi ve veri madenciliği programıdır. Bünyesinde veri ön işleme, sınıflandırma, kümeleme, regresyon, birliktelik, nitelik seçimi ve görselleştirme fonksiyonları bulunmaktadır. Niteliklere hem nominal hem de nümerik değerler atanabilmektedir. Weka'da 15 sınıflandırma algoritması

içerisinden kategorik sınıf niteliği ile çalışabilen karar ağacı tabanlı ve kural tabanlı 13 algoritma bulunmaktadır. Çıktı niteliği kategorik olarak düzenlendiği için çalışmanın ilk aşamasında bu 13 algoritma kullanılmış ve en iyi sonucu veren 7 algoritma ile çalışmaya devam edilmiştir. Üretim süresine etki eden faktörler belirlenirken 5 farklı nitelik seçme yöntemi kullanılmıştır. Yöntemlere göre oluşturulan ve farklı nitelikleri içeren veri kümelerine, seçilen algoritmalar uygulanarak veri madenciliği modelleri Weka programı yardımıyla oluşturulmuştur. Oluşturulan modellerin performans ölçütleri karşılaştırılarak en iyi model belirlenmiştir. En iyi modele göre üretilen kurallar incelenerek üretim sürecini iyileştirici tavsiyeler hazırlanmıştır. İzleyen bölümlerde problem çözümünün her bir adımın detayları anlatılmıştır.

## 2.2. Veri Kümesi (Data Set)

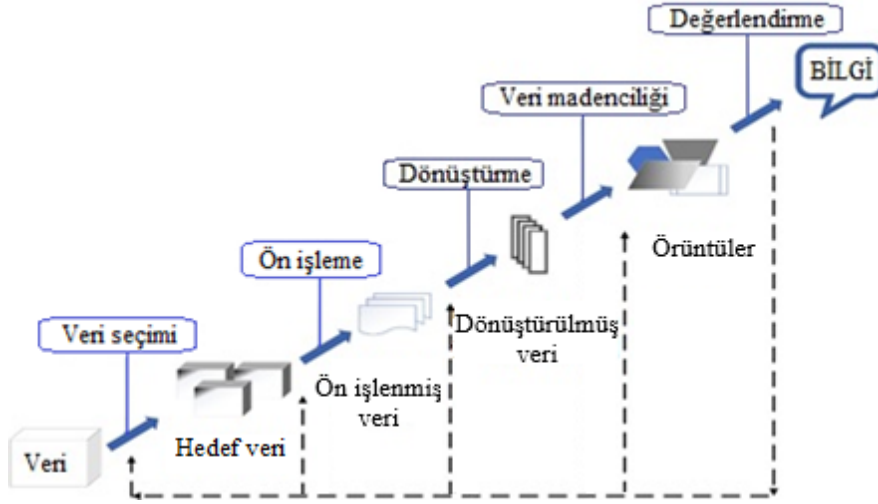
Yukarıda belirtildiği gibi çalışmanın amacı üretim süresine etki eden faktörlerin veri madenciliği yöntemi kullanılarak belirlenebileceğini göstermektir. Bu amaçla uygulama yeri olarak siparişe göre üretim yöntemini kullanan ve makine üretimi yapan bir işletme seçilmiştir. Makine parçalarının işlendiği CNC tezgâhları üretimin en önemli kısmını oluşturduğu için bu tezgâhlara ait üretim verileri analiz

edilerek üretim süresine etki eden faktörlerin belirlenmesine karar verilmiştir. Hali hazırda işletmede üretim süresinin en aza indirilmesi, tecrübe ile doğru tezgâha, yeterli eğitim düzeyine sahip, doğru operatör ataması ile gerçekleştirilmeye çalışılmaktadır.

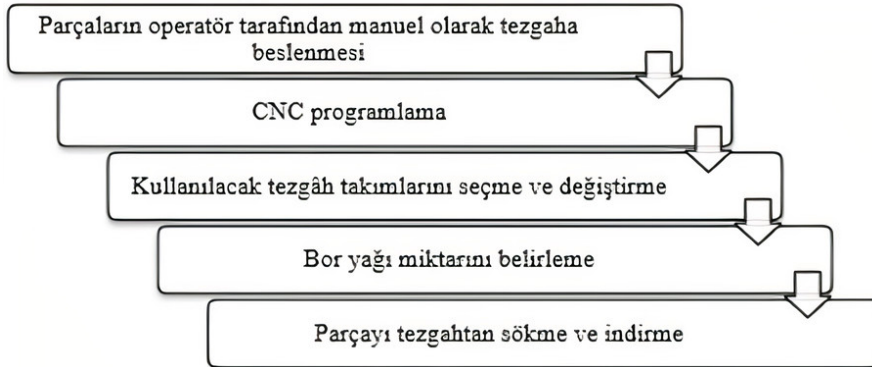
Firmada üç eksenle farklı boyutlara sahip 5 adet tezgâh bulunmaktadır. Üretime ait iş akış şeması Şekil 2'de yer almaktadır.

Tezgâhlarda üretim operatör tarafından manuel olarak tezgâha besleme, teknik resme göre CNC programlama, kullanılacak tezgâh takımlarını seçme ve değiştirme, bor yağ miktarını belirleme, parçayı tezgâhtan sökme ve indirme operasyonlarını izleyerek gerçekleştirilmektedir.

İşletmenin veri tabanından elde edilen ham veriler CNC tezgâhlarının üretimi ile ilgili 23 değişkenin değerlerinden oluşmaktadır. Veri tabanından alınan ham veriler çalışmanın ek dosyasında verilmiştir. Şekil 2'de görüldüğü gibi parçalar tezgâha manuel olarak yüklenmektedir. Bu aşama emek yoğun bir işlem olduğu için ortam sıcaklığının parça besleme hızına etki edebileceği düşünülmüş ve veri tabanından alınan ham verilere ortalama sıcaklık değişkeni de eklenerek toplamda 23 değişken ve 935 örnekten oluşan bir veri kümesi



Şekil 1. Veri madenciliği süreçleri (Data mining processes) [25]



Şekil 2. İş akış şeması (Work flow diagram)

elde edilmiştir. Sınıf niteliği olarak üretim süresi seçilmiştir. Geriye kalan 22 değişkenden veri ön işleme ile elde edilen 16 değişken nitelik olarak seçilmiştir. Seçilen niteliklerle ilgili bilgiler Tablo 1’de verilmiştir. Tablo 1’de gösterilen niteliklerin seçimi bir sonraki bölümde anlatılmıştır.

### 2.2.1. Veri ön işleme (Preprocessing)

Veri ön işleme teknikleri temizleme, birleştirme, dönüştürme, indirgemeden oluşmaktadır [28]. Bu çalışmada ön işleme sürecinde öncelikle gereksiz ve boş olan bilgiler temizlenmiştir. Veri kümesindeki 23 nitelik temizleme işlemi ile 17 niteliğe indirgenmiştir. Nitelik seçimi yöntemleri uygulandıktan sonra ise veri kümesinde 7 nitelik ve üretim süresi de sınıf niteliği olmak üzere toplamda 8 nitelik kalmıştır. Tezgâh, tezgâh tipi ve tezgâh markası olarak iki niteliğe ayrılmıştır. Tezgâh boyutu, üç nitelikten bir niteliğe azaltılmıştır. Üretim süresi, eşit aralıklı üç farklı değer olacak şekilde kesiklendirilmiştir. Bu değerler 0-180, 181-360, 361-540 dakika arasındaki değerlerdir. Bu değer aralıklarının seçilme gerekçesi firmada üç tip üretim stratejisi uygulanmasıdır. Bunlar normal mesai, fazla mesai, fason üretimdir. Teslim süresinin kritik olduğu projelerde işletmenin stratejisi, üretim süresi düşük olan parçaların normal mesaide, üretim süresi fazla olan parçaların ise fazla mesaide veya fason üretimle üretilmesidir. Bu nedenle değerleri kesiklendirirken üretim talebinin kapasiteyi karşılayamaması durumunda, üretim süresi 0-180 dakika arasındaki parçalar normal mesai, 181-360 dakika arasında olanlar fazla mesai ve 361-540 dakika arasında olanlar ise

fason üretim stratejisiyle üretilmesi kararının verileceği düşünülmüştür.

### 2.2.2. Nitelik seçimi (Attribute selection)

Veri madenciliğinin önemli aşamalarından biri veri boyutunun azaltılması işlemidir. Veri boyutunun azaltılması kısaca veri kümelerinde gereksiz ve boş bilgilerin çıkarılması olarak tanımlanmaktadır. Veri boyutunun azaltılması için kullanılan yöntemlerin başında nitelik seçimi gelmektedir. Nitelik seçimi, orijinal veri kümesini en iyi şekilde temsil edebilecek alt küme seçimidir. Nitelik seçimi ile mevcut problem için en önemli nitelikler seçilerek ve veri kümesindeki nitelik sayısı azaltılarak veri boyutunu indirgeme amaçlanmaktadır [29].

Bu çalışmada veri kümesi, nitelik seçimi ile indirgenerek uygulanan sınıflandırma yöntemleri karşılaştırılmıştır. Kullanılan nitelik seçimi yöntemleri Weka programında bulunan yöntemlerdir. Bu yöntemlerin kısa tanımları aşağıda verilmiştir.

#### 2.2.2.1. Bilgi kazancı nitelik seçimi yöntemi (Information gain attribute eval)

Bilgi kazanımı, herhangi bir nitelik için sınıflandırma sonuçlarının ne kadar değer kazanabileceğini gösterir. 0 ile 1 arasındadır. 0 değeri, hiç bilgi kazanımı olmadığı anlamına gelirken 1 değeri ise bilgi kazanımının tam olduğunu ifade etmektedir [30].

**Tablo 1.** Kullanılan nitelik adları, tipleri ve değerleri (Attribute names, types and values that used)

Nitelik adı	Nitelik tipi	Değerler
Makine adı	Kategorik	Kablo boşaltıcı ve sarıcı, paletleme, konveyör, paketleme, otomatik depolama sistemi vb.
Parça adı	Kategorik	Bağlantı parçası, plaka, şase, mil vb.
Parça tipi	Kategorik	Dikdörtgen, silindirik
Operatör adı	Kategorik	AK, EC, HI, MD, SO, YY
Operatör tecrübe derecesi	Nümerik	2-14 yıl arasında değerler almaktadır
Operatör eğitim düzeyi	Kategorik	Lise, ön lisans
Tezgâh tipi	Kategorik	Dikey işleme, yatay işleme
Tezgâh markası	Kategorik	Mazak, Okuma, Taksan, Yüntes
Tezgâh yaşı	Nümerik	2-17 yıl arasında değerler almaktadır
Tezgâh boyutu	Kategorik	700x450x500, 500x550x500, 1600x3000x800, 1000x500x500, 550x550x550, cm
Ürün miktarı	Nümerik	1 - 90 adet arasında değerler almaktadır
Ortalama sıcaklık	Nümerik	-6 - 26 °C arasında değerler almaktadır
Üretim günü	Kategorik	1 - 31 arasında değerler almaktadır
Haftanın günü	Kategorik	Pazartesi - cumartesi arasında değerler almaktadır
Üretim ayı	Kategorik	1-12 arasında değerler almaktadır
En son bakım ile üretim arasındaki gün	Nümerik	0 - 173 gün arasında değerler almaktadır
Üretim süresi (sınıf niteliği)	Nümerik	4 - 540 dakika arasında değerler almaktadır

#### 2.2.2.2. CFS alt küme nitelik seçimi yöntemi (Classified feature selection subset eval)

CFS yöntemi nitelik alt kümesini, alt kümedeki her bir niteliğin sınıf niteliği ile olan korelasyonunun yüksek ve niteliklerin kendi aralarındaki korelasyonların düşük olması ilkesine dayanarak seçmektedir [31].

#### 2.2.2.3. Sınıflandırıcı nitelik seçimi yöntemi (Classifier subset eval)

Sınıflandırıcı, değerlendirme yaparken eğitim verilerindeki nitelik alt kümelerini veya ayrı bir test kümesini kullanır [32].

#### 2.2.2.4. Kazanç oranı nitelik seçimi yöntemi (Gain ratio attribute eval)

Kazanç oranı, bilgi kazancının çeşitli değerlere sahip özellikleri seçme eğiliminin önüne geçmek için kullanılmaktadır. Kazanç oranı 0-1 aralığında değer almaktadır. 1'e eşit olduğunda X bilgisi ile Y bilgisinin tamamen tahmin edilebildiğini, 0'a eşit olduğunda ise X ile Y arasında hiçbir ilişki olmadığını gösterir [29]. Sınıf niteliğine göre kazanç oranını ölçerek niteliğin değerini belirler [32].

#### 2.2.2.5. Korelasyon nitelik seçimi yöntemi (Correlation attribute eval)

Korelasyon katsayısı, iki rassal nitelik arasındaki doğrusal ilişkinin yönünü ve gücünü belirtir. Bu katsayı, -1 ile +1 arasında bir değer alır. Pozitif değerler direkt yönlü doğrusal ilişkiyi; negatif ise ters yönlü bir doğrusal ilişkiyi belirtir. Korelasyon katsayısı 0 ise söz konusu değişkenler arasında doğrusal bir ilişki yok demektir [33].

Korelasyon nitelik seçimi yöntemi, her bir nitelik ile sınıf niteliği arasındaki korelasyonu hesaplar. Hesaplama sonucunda pozitif ya da negatif korelasyonu (-1 ile 1) olan nitelikler seçilip düşük korelasyona sahip nitelikler çıkarılmalıdır [34].

#### 2.2.3. Sınıflandırma (Classification)

En yaygın veri madenciliği yöntemlerinin başında sınıflandırma gelmektedir [35]. Sınıflandırma genellikle farklı sınıflara ait verileri ayırarak tahminleme yapma işleminde kullanılmaktadır. Sınıflandırma, bir örneğin gerçek sınıfı ile sahip olduğu nitelikler arasında ilişki kurularak yapılabilir [36]. Veri madenciliğinde denetimli öğrenme kavramı, sınıflandırma ile bilinen veriler aracılığıyla bir sınıflandırma fonksiyonu öğretmek ya da bir sınıflandırma modeli kurmaktır. Bu fonksiyon ya da model, veri tabanındaki verileri niteliklere dönüştürür, bu sayede yeni veriler sınıf tahmininde kullanılabilir [37]. Sınıflandırma; karar ağacı tabanlı, kural tabanlı, yapay sinir ağları, bayes sınıflandırıcılar ve bayes ağları yöntemlerinden oluşmaktadır. Karar ağacı tabanlı yöntemler, sınıflandırma ve tahminleme açısından güçlü araçlardır. Bu yöntemlerin

çekici tarafı karar ağaçlarının kuralları temsil etmesidir. Böylelikle kuralları yorumlamak kolaylaşmaktadır. Kural tabanlı yöntemler ise bilinen veriler aracılığıyla belirlenen sınıf üzerinde kurallar oluşturur.

Bu çalışmada kullanılan sınıflandırma algoritmaları part, oneR, decision table, jrip, random tree, REPTree ve LMT'dir. Kullanılan algoritmaların tanımları aşağıda verilmiştir.

##### 2.2.3.1. Part (Part)

Part, C4.5 ve RIPPER algoritmalarının birleştirilmesiyle oluşan parçalı karar ağacı tabanlı bir algoritmadır. Budama yaparken modeli en çok kapsayan dal ile kural oluşturur [38].

##### 2.2.3.2. OneR (OneR)

OneR, Holte [39] tarafından önerilmiş bir nitelik seçimi yöntemidir. OneR algoritmasında, öncelikle eğitim veri kümesindeki her özellik için bir kural oluşturulmakta daha sonra oluşturulan her bir kural için sınıflandırma doğrulukları hesaplanmakta ve en az hatalı kurala ait nitelik seçilmektedir [40].

##### 2.2.3.3. Karar tablosu (Decision table)

Karar tablosu, karar ağaçlarından sayısal tahmin için kullanılan bir yöntemdir. Diğer karar ağacı tabanlı yaklaşımlara göre daha basit, daha az bilgi işlem yoğun bir algoritmadır. Böylece kuralların daha anlaşılır olma potansiyeline sahip bir sıralı eğer-o zaman kuralları kümesi oluşturmaktadır. Nitelik alt kümelerini en iyi ilk aramayı kullanarak değerlendirir. Her bir örnek için sınıfı belirlemek üzere en yakın komşu yöntemini kullanır [41].

##### 2.2.3.4. Jrip (Jrip)

Jrip, bir veri kümesinden en değerli nitelikleri seçerek kurallar oluşturur ve daha sonra özellikleri önem sırasına göre sıralar. Tekrarlanan artırımlı budama yöntemi ile hata azaltma yapmaktadır [42].

##### 2.2.3.5. Rastgele ağaç (Random tree)

Random tree (rastgele ağaç), birçok bireysel kuralı öğrenerek kural üreten topluluk öğrenme algoritmasıdır. Bir karar ağacı oluşturmak ve rastgele bir veri kümesi üretmek için bir torbalama fikri (bagging idea) kullanır. Standart bir karar ağacında her düğüm tüm değişkenler arasındaki en iyi bölünme kullanılarak bölünür. Random tree algoritmasında ise her bir düğüm o düğümde rastgele seçilen belirteçlerin alt kümesi arasında en iyisi kullanılarak bölünür. Bu sayede random tree yüksek doğruluk sonuçlarına ulaşır [43].

##### 2.2.3.6. Temelde azaltılmış hata budama ağacı (REPTree)

Temelde azaltılmış hata budama ağacı (REPTree), karar ağacı öğrenme yöntemidir. Varyansı azaltmaya dayalı bir karar ağacı oluşturmaktadır. Farklı yinelemelerde birden çok

ağaç oluşturarak üretilen ağaçlar arasından en iyisini seçmektedir. Ağacı budarken kullanılan ölçü ağaç tarafından yapılan tahminlerdeki ortalama hatanın karesidir. Ayırma kıstası olarak bilgi kazancı kullanan bir karar ağacı oluşturur ve azaltılmış hata budaması kullanarak budanan bir hızlı karar ağacı öğreticisidir [44].

#### 2.2.3.7. LMT (Logistic Model tree)

LMT (lojistik model ağacı), parçalı bir doğrusal regresyon modeli sağlamak için yapraklarında doğrusal regresyon modellerine sahip bir karar ağacıdır [45]. Sıradan karar ağaçlarında olduğu gibi LMT'nin özelliklerden biri üzerinde yapılan bir test her iç düğümle ilişkilendirilir. LMT için bazen tek bir yaprak en iyi genelleştirme performansına yol açar, bu da basit karar ağaçları için nadiren görülür [46].

#### 2.2.4. Sınıflandırma başarısını ölçme (Measuring classification performance)

Sınıflandırma modellerinde literatürde sık kullanılan değerlendirme ölçütleri doğruluk oranı, doğru pozitif, yanlış pozitif, f-ölçütü ve ROC eğrisidir [47-50]. Bu çalışmada da bu ölçütler kullanılarak modeller değerlendirilmiştir. Bu ölçütlerin tanımları aşağıda verilmiştir.

##### 2.2.4.1. Doğruluk oranı (Accuracy rate)

Modelin başarısının ölçülmesinde en çok kullanılan yöntem doğruluk oranıdır. Doğru sınıflandırılan örnek sayısının toplam örnek sayısına oranıdır. Doğruluk oranı, Eş. 1'deki formül kullanılarak hesaplanmaktadır [51].

$$\text{Doğruluk oranı} = \frac{DP+DN}{N} \quad (1)$$

Bu eşitlikte DP doğru pozitif, DN doğru negatif ve N örnek sayısını ifade eder.

##### 2.2.4.2. Doğru pozitif oranı (True positive rate)

Veri kümesindeki sınıfı ile aynı sınıfta tahmin edilen veri sayısının, veri kümesindeki sınıfı ile aynı sınıfta tahmin edilen ve veri kümesindeki sınıfı ile farklı sınıfta tahmin edilen veri sayısına oranlanması ile elde edilen gösterimdir. Doğru pozitif oranı, Eş. 2'deki formülden faydalanılarak hesaplanmaktadır [52].

$$\text{Doğru pozitif oranı} = \frac{DP}{DP+YN} \quad (2)$$

Bu eşitlikte DP doğru pozitif, YN yanlış negatif ifade eder.

##### 2.2.4.3. Yanlış pozitif oranı (False positive rate)

Veri kümesindeki sınıfı ile farklı sınıfta tahmin edilen veri sayısının, veri kümesindeki sınıfı ile farklı sınıfta tahmin edilen ve veri kümesindeki sınıfı ile aynı sınıfta tahmin edilen veri sayısına oranlanması ile elde edilen gösterimdir. Yanlış pozitif oranı, Eş. 3'teki formül kullanılarak hesaplanmaktadır [53].

1956

$$\text{Yanlış pozitif oranı} = \frac{YP}{YP+DN} \quad (3)$$

Bu eşitlikte YP yanlış pozitif, DN doğru negatif ifade eder.

##### 2.2.4.4. F-ölçütü (F-score)

Kesinlik ve hassasiyet ölçütleri, modelin başarısını değerlendirmek için tek başına yeterli değildir. Her iki ölçütü beraber değerlendirmek daha doğru sonuçlar verir. Bunun için f-ölçütü tanımlanmıştır. F-ölçütü, kesinlik ve hassasiyetin harmonik ortalamasıdır. F-ölçütünü hesaplamak için Eş. 4'teki formül kullanılmaktadır [54].

$$F\text{-ölçütü} = \frac{2KH}{K+H} \quad (4)$$

Bu eşitlikte K kesinliği, H hassasiyeti ifade eder.

##### 2.2.4.5. ROC eğrisi (ROC curve)

ROC eğrisi, hassasiyetin kesinliğe oranıdır. Doğru pozitiflerin, yanlış pozitiflere oranlanması ile elde edilir. ROC eğrisi, Eş. 5'teki formülden yararlanılarak hesaplanmaktadır [55].

$$ROC \text{ eğrisi} = \frac{H}{K} \quad (5)$$

### 3. MODELLER VE KARŞILAŞTIRILMALARI (MODELS AND THEIR COMPARISON)

Modelleme üç aşamada tamamlanmıştır. Bütün aşamalarda, literatürde önerildiği gibi verilerin %66'sı (2/3'ü) eğitim kümesi, %34'ü (1/3'ü) ise test kümesi olarak ayrılmıştır [15, 39, 45, 53].

**Birinci aşama:** Nitelik seçimi ile eleme yapılmadan 17 nitelik kullanarak Weka'da kategorik sınıf niteliği ile çalışabilen karar ağacı tabanlı ve kural tabanlı 13 algoritma denenmiş ve 13 model kurulmuştur. En iyi yedi modeli veren algoritma seçilmiştir. Veri kalabalığı olmaması açısından sadece en iyi ilk yedi modelin sonuçları makaleye koyulmuştur. İlk yedi modelin performans ölçütlerinin değerleri Tablo 2'de verilmiştir. Tablo 2'de görüldüğü gibi random tree algoritması bütün ölçütlere göre en iyi sonucu veren algoritmadır. Random tree algoritmasından sonraki en iyi algoritmanın ölçüt değerleri, bütün ölçütler için random tree algoritmasının ölçüt değerlerine göre önemli derecede geridedir. Bunun nedeni random tree algoritmasının bir topluluk öğrenme algoritması olmasıdır.

**İkinci aşama:** Nitelik sayısının daha aza indirgenip indirgenemeyeceğini belirlemek için veri temizleme aşamasında ön eleme ile elde edilen 17 niteliğe ikinci bölümde tanımları verilen nitelik seçimi yöntemleri uygulanmıştır. Tablo 3'te nitelik seçimi yöntemleri kullanılarak seçilen nitelikler görülmektedir. Bilgi kazancı ve korelasyon yöntemleri nitelikleri seçmeyi sıralamaktadır. Bu nedenle bu yöntemler için tabloda niteliklerin bilgi kazancı ve korelasyon değerleri verilmiştir. Bilgi kazancı ve korelasyon değerlerinin büyüktür küçüğe göre sıra değerleri parantez içerisinde yazılmıştır.



**Tablo 2.** Nitelik seçimi öncesi sınıflandırma algoritmalarında elde edilen performans ölçütlerinin karşılaştırılması  
(Comparison of evaluation metrics obtained in classification algorithms before feature selection)

Algoritmalar	Random tree	Part	REPTree	OneR	Jrip	LMT	Decision table
Performans ölçütleri							
Doğruluk oranı (%)	97,97	74,33	69,83	62,78	62,35	61,71	61,28
Doğru pozitif	0,98	0,74	0,70	0,63	0,62	0,62	0,61
Yanlış pozitif	0,02	0,24	0,31	0,41	0,44	0,45	0,46
F-ölçütü	0,98	0,67	0,67	0,57	0,54	0,52	0,52
ROC eğrisi	0,99	0,88	0,84	0,61	0,61	0,63	0,62

**Tablo 3.** Nitelik seçimi yöntemleri kullanılarak seçilen nitelikler (The attributes selected by the attribute selection methods)

Seçilen nitelikler	Nitelik seçimi Bilgi yöntemleri kazancı NSY	Korelasyon NSY	CFS alt kümesi NSY	Sınıflandırıcı NSY	Kazanç oranı NSY
En son bakım ile üretim arasındaki gün	0,009 (7)	0,010 (9)			
Haftanın günü	0,009 (7)	0,030 (7)			
Makine adı	0,216 (2)	0,040 (6)		X	X
Operatör adı	0,012 (5)	0,040 (6)			
Operatör eğitim düzeyi	0,004 (10)	0,040 (6)			
Operatör tecrübe derecesi	0,000	0,060 (4)			
Ortalama sıcaklık	0,163 (4)	0,160 (1)	X	X	X
Parça adı	0,326 (1)	0,020 (8)	X		X
Parça tipi	0,005 (9)	0,050 (5)			
Tezgâh boyutu	0,011 (6)	0,020 (8)			
Tezgâh markası	0,008 (8)	0,050 (5)			
Tezgâh tipi	0,001 (12)	0,040 (6)			
Tezgâh yaşı	0,000	0,060 (4)			
Üretim ayı	0,184 (3)	0,090 (2)	X	X	X
Üretim günü	0,000	0,030 (7)			
Ürün miktarı	0,002 (11)	0,070 (3)			

Tablo 3'e göre dört veri kümesi oluşturulmuştur; bilgi kazancı, CFS, sınıflandırıcı ve kazanç oranına dayanan veri kümeleri. Korelasyon nitelik seçimi yöntemi ile ayrıca bir veri kümesi oluşturulmamış, üretim süresi ile girdi nitelikleri arasındaki korelasyon değerleri, diğer yöntemlerde seçilen değişkenleri kontrol etmek için kullanılmıştır. Üretim süresi ile korelasyon değeri en yüksek olan ilk üç nitelik sırasıyla ortalama sıcaklık, üretim ayı ve ürün miktarıdır. Ortalama sıcaklığın üretim süresi ile en yüksek korelasyona sahip olmasının nedeninin, üretilecek ürünlerin CNC makinesine tek tek operatör tarafından manuel olarak beslenmesi olduğu düşünülmektedir. Üretim ayı ise iş yoğunluğuna etki ettiği için önemlidir. İşletmede iş yoğunluğu yaz aylarında nispeten daha az olmaktadır. Ürün miktarı niteliğinin korelasyon değerinin üçüncü sırada olmasının nedeni ise genelde ebatları büyük olan parçaların küçük miktarlarda üretilmesi ve üretim sürelerinin uzun olmasıdır.

Birinci veri kümesi bilgi kazancı nitelik seçme yöntemine göre oluşturulmuştur. Bilgi kazancı en yüksek olan ilk üç nitelik sırasıyla parça adı, makine adı ve üretim ayıdır. Parça adı ve makine adı işlenen parçanın ne incelikte işleneceğini belirlediği için önemlidir. Üretim ayının etkisi yukarıda açıklanmıştır. Bilgi kazancına göre hangi niteliklerin modele girmesi gerektiğine karar vermek için bir eşik değerinin belirlenmesi gerekmektedir. Ancak literatürde eşik değerinin

veri kümesinin özelliklerine göre değişeceği ifade edilmiştir [56, 57]. Bu nedenle modele girecek nitelikleri belirlemek için farklı eşik değerleri (0,2, 0,1, 0,01, 0,008, 0,003) denenerek modellerin performansı değerlendirilmiştir. Tablo 4'te denenen eşik değerleri, en iyi performans ölçütlerinin elde edildiği random tree algoritması kullanılarak karşılaştırılmıştır. Eşik değeri düşürüldükçe modele ek nitelikler girmiş ve performans ölçüt değerleri gittikçe azalan oranlarda iyileşmiştir. Ancak eşik değeri 0,008 den 0,003'e düşürüldüğünde performans ölçütlerinin hiçbirinin değişmediği gözlemlenmiştir. Bu da eşik değerinin 0,003'e düşmesi ile eklenen parça tipi ve operatör eğitim düzeyinin etkili nitelikler olmadığını göstermektedir. Eşik değeri 0,1'den büyük veya eşit değerler için sınıf niteliği ile korelasyonu en yüksek iki niteliğin bu veri kümelerinde bulunduğu görülmektedir. Tablo 4'de görüldüğü gibi eşik değeri 0,01 den 0,008 düşürüldüğünde 3 nitelik daha modele eklenmesine rağmen bütün ölçütlerdeki iyileşme oranı %1'in altındadır. Bu nedenle bilgi kazancı nitelik seçme yöntemine göre eşik değeri 0,01 alınarak 6 nitelikten oluşan (parça adı, makine adı, üretim ayı, ortalama sıcaklık, operatör adı ve tezgâh boyutu) veri kümesi bilgi kazancı veri kümesi olarak belirlenmiştir.

CFS, sınıflandırıcı ve kazanç oranı nitelik seçme yöntemlerine göre oluşturulan diğer üç veri kümesinde ortak

olan nitelikler ortalama sıcaklık ve üretim ayıdır. Bunlar üretim süresi ile korelasyonu sırasıyla en yüksek olan iki niteliklerdir. Tablo 3'e göre CFS veri kümesi ortalama sıcaklık, üretim ayı ve parça adı; sınıflandırıcı veri kümesi üretim ayı, makine adı ve ortalama sıcaklık; kazanç oranı veri kümesi ise ortalama sıcaklık, üretim ayı, parça adı ve makine adı niteliklerinden oluşmaktadır.

Nitelik seçimi yöntemleri ile uygulanan sınıflandırma algoritmalarında elde edilen performans ölçütleri Tablo 5 - Tablo 8'de verilmiştir. Tablolarda sonucu tire ile gösterilen ölçütler, o modellerde B sınıfına hiçbir atama olmadığı için hesaplanamamıştır ve tablolarda hesaplanamayan değerler için tire koyulmuştur. Bütün veri kümelerine ve bütün

ölçütlere göre en iyi modeli bilgi kazancı veri kümesi – random tree algoritması ikilisi vermiştir. Random tree bütün veri kümelerine göre de en iyi modeli veren algoritma olmuştur. Bunun nedeni random tree algoritmasının bir topluluk öğrenme algoritması olmasıdır.

*Üçüncü aşama:* Modellemenin üçüncü aşamasında üretim mühendisi ile yapılan görüşmede bilgi kazancı veri kümesinden (eşik değeri 0,01) oluşturulan modelin daha iyileştirilip iyileştirilemeyeceği üzerine düşünülmüştür. Görüşmede, büyük makine parçalarının az sayıda üretilmesi ve üretim sürelerinin daha uzun olması nedeni ile ürün miktarının üretim süresini etkileyebileceği bu nedenle modele katılması gerektiği belirlenmiştir. Bu gerekçe ürün

**Tablo 4.** Bilgi kazancı veri kümesinde çeşitli eşik değerleri için random tree performans ölçütleri  
(Random tree evaluation metrics for various threshold values in the info gain data set)

Eşik değerleri	0,2	0,1	0,01	0,008	0,003
Performans ölçütleri					
Doğruluk oranı (%)	74,87	88,24	96,04	96,90	96,90
Doğru pozitif	0,75	0,88	0,96	0,97	0,97
Yanlış pozitif	0,24	0,13	0,04	0,03	0,03
F-ölçütü	0,74	0,88	0,96	0,97	0,97
ROC eğrisi	0,90	0,98	0,99	0,99	0,99

**Tablo 5.** Bilgi kazancı nitelik seçimi yöntemi (eşik değer = 0,01) ile seçilen niteliklere göre sınıflandırma algoritmalarında performans ölçütlerinin karşılaştırılması

(Comparison of evaluation metrics in classification algorithms according to the features selected with info gain attribute eval (threshold = 0,01))

Algoritmalar	Random tree	Part	REPTree	OneR	Jrip	LMT	Decision table
Performans ölçütleri							
Doğruluk oranı (%)	96,04	70,91	68,98	62,78	62,57	60,96	61,60
Doğru pozitif	0,96	0,71	0,69	0,63	0,63	0,61	0,62
Yanlış pozitif	0,04	0,29	0,33	0,41	0,43	0,46	0,65
F-ölçütü	0,96	0,69	0,66	0,57	0,55	-	0,52
ROC eğrisi	0,99	0,84	0,83	0,61	0,61	0,59	0,64

**Tablo 6.** Kazanç oranı nitelik seçimi yöntemi ile seçilen niteliklere göre sınıflandırma algoritmalarında performans ölçütlerinin karşılaştırılması

(Comparison of evaluation metrics in classification algorithms according to the features selected with gain ratio attribute eval)

Algoritmalar	Random tree	Part	REPTree	OneR	Jrip	LMT	Decision table
Performans ölçütleri							
Doğruluk oranı (%)	88,24	68,77	69,63	62,78	61,82	60,96	61,18
Doğru pozitif	0,88	0,69	0,70	0,63	0,62	0,61	0,61
Yanlış pozitif	0,13	0,32	0,32	0,41	0,45	0,46	0,46
F-ölçütü	0,88	0,66	0,67	0,57	0,52	-	-
ROC eğrisi	0,98	0,81	0,84	0,61	0,60	0,59	0,60

**Tablo 7.** CFS nitelik seçimi yöntemi ile seçilen niteliklere göre sınıflandırma algoritmalarında performans ölçütlerinin karşılaştırılması

(Comparison of evaluation metrics in classification algorithms according to the features selected with CFS attribute eval)

Algoritmalar	Random tree	Part	REPTree	OneR	Jrip	LMT	Decision table
Performans ölçütleri							
Doğruluk oranı (%)	82,57	66,42	65,78	62,79	61,60	62,14	61,18
Doğru pozitif	0,83	0,66	0,66	0,63	0,62	0,62	0,61
Yanlış pozitif	0,18	0,36	0,38	0,41	0,46	0,44	0,46
F-ölçütü	0,82	0,63	0,61	0,57	-	0,52	-
ROC eğrisi	0,96	0,78	0,78	0,61	0,59	0,62	0,60

miktarının üretim süresi ile 3. sırada korelasyonu en yüksek nitelik olması sonucu ile de desteklenmektedir. Görüşme sonucuna göre ürün miktarı, bilgi kazancı eşik değeri 0,01 olan veri kümesine eklenerek 7 nitelikli yeni bir veri kümesi oluşturulmuş ve bu kümeye genişletilmiş bilgi kazancı veri kümesi adı verilmiştir. Genişletilmiş bilgi kazancı veri kümesine seçilen 7 algoritma uygulanarak 7 model oluşturulmuştur. Bu modellerin performans ölçüt değerleri Tablo 9'da verilmiştir. Tablo 9'a göre bütün ölçütlere için random tree algoritmasının en iyi sonucu verdiği görülmektedir. Random tree algoritmasından sonraki en iyi algoritmanın ölçüt değerleri, bütün ölçütler için random tree algoritmasının ölçüt değerlerine göre önemli derecede başarısızdır. Tablo 5'de verilen bilgi kazancı veri kümesi-random tree algoritması modeli performans ölçütleri ve Tablo 9'da verilen genişletilmiş bilgi kazancı veri kümesi-random tree algoritması modeli performans ölçütleri karşılaştırıldığında, ürün miktarının eklenmesi ile bütün performans ölçütlerinin iyileştiği gözlenmiştir. Ürün miktarının bilgi kazancı değeri küçük olmasına rağmen modele eklendiğinde model performans ölçütlerini iyileştirebilmesinin sebebi nitelik etkileşimleridir [57]. Tek başına önemli görülmeyen bir nitelik başka niteliklerle etkileşerek modelin başarısını artırabilmektedir. Burada yine random tree algoritmasının bir topluluk öğrenme algoritması olmasının önemi görülmektedir.

Tablo 2 ve Tablo 9 incelendiğinde bütün performans ölçütleri arasındaki farkın %1'in altında olduğu görülmektedir. Nitelik seçimi sonrasında 17 nitelik yerine 7 nitelik model ifade edilebilmiştir.

Karar ağacının boyutu büyük olduğu için şekil halinde gösterilememektedir. Bulunan kurallardan iyileştirme faaliyetleri açısından en önemli görülenler Tablo 10'da gösterilmektedir. Tablo 10'da A sınıfı üretim süresi 0-180 dakika, B sınıfı 181-360 dakika ve C sınıfı ise 361-540 dakika arasında olan verileri ifade etmektedir. Tabloda bulunan ikinci kural üzerinden örnek verilecek olursa ortalama sıcaklık 18,5'tan küçük veya eşit, parça adı lama, makine adı boşaltıcı makinesi, operatör eğitim düzeyi ise ön lisansa eşit kuralını sağlayan veriler A sınıfına atanmakta yani 0-180 dakika arasında olmaktadır. Tek bir nitelik değerinde farklılaşan kuralları karşılaştırmalı olarak incelendiğinde 1. ve 2. kuraldan parça adının, 3. ve 4. kuraldan makine adının, 5. ve 6. kuraldan üretim ayının, 7. ve 8. kuraldan ortalama sıcaklığın, 9. ve 10. kuraldan operatör adının, 11. ve 12. kuraldan tezgâh boyutunun, 13. ve 14. kuraldan ürün miktarının etkili olduğu görülmektedir.

Random tree ile elde edilen karar ağacından niteliklerin üretim süresi üzerindeki etki sıralaması dallanma seviyesine göre ortalama sıcaklık, ürün miktarı, üretim ayı, parça adı, operatör adı, makine adı ve tezgâh boyutu olduğu gözlemlenmiştir. Bu niteliklerin üretim süresi üzerinde etkisini azaltmak için ortam sıcaklığı 18,5°C'tan dallandığı için ortam sıcaklığının bu seviyede sabit tutulması için tedbir alınmalıdır. Üretilen parçalarda kural örneklerinden de görülebileceği gibi parçayı üreten operatörün önemli olduğu görülmektedir. Operatör verimliliklerini artırmaya yönelik faaliyetler yapılmalıdır. Makinelerin kullanımı için CNC programlama dilinin iyi seviyede biliniyor olması gerekmektedir. Bu sebeple operatör önem arz etmektedir.

**Tablo 8.** Sınıflandırıcı nitelik seçimi yöntemi ile seçilen niteliklere göre sınıflandırma algoritmalarında performans ölçütlerinin karşılaştırılması

(Comparison of evaluation metrics in classification algorithms according to the features selected with classifier attribute eval)

Performans ölçütleri	Algoritmalar	Random tree	Part	REPTree	OneR	Jrip	LMT	Decision table
Doğruluk oranı (%)		74,33	67,17	62,99	61,60	61,60	60,96	61,18
Doğru pozitif		0,74	0,67	0,63	0,62	0,62	0,61	0,61
Yanlış pozitif		0,27	0,35	0,43	0,45	0,45	0,46	0,46
F-ölçütü		0,73	0,64	0,55	0,52	0,52	-	-
ROC eğrisi		0,89	0,75	0,68	0,59	0,60	0,59	0,60

**Tablo 9.** Genişletilmiş bilgi kazancı veri kümesi için algoritmaların performans ölçütlerinin karşılaştırılması

(Comparison of performance metrics in classification algorithms for extended information gain date set)

Performans ölçütleri	Algoritmalar	Random tree	Part	REPTree	OneR	Jrip	LMT	Decision table
Doğruluk oranı (%)		97,54	70,91	68,88	62,78	61,28	61,71	61,6
Doğru pozitif		0,98	0,71	0,69	0,63	0,61	0,62	0,62
Yanlış pozitif		0,03	0,29	0,33	0,41	0,46	0,45	0,45
F-ölçütü		0,98	0,69	0,66	0,57	-	0,52	0,52
ROC eğrisi		0,99	0,84	0,83	0,61	0,59	0,63	0,64

**Tablo 10.** Nitelik seçimi sonrası random tree aracılığıyla elde edilen kural örnekleri  
(Some of the rules obtained through the random tree after the feature selection)

Kural no	Kural	Atandığı sınıf
1	Parça adı = kama	A
2	Parça adı = çatal	C
3	Parça adı = kol; makine adı = asrs	A
4	Parça adı = kol; makine adı = paletleme	C
5	Operatör adı = yy; ortalama sıcaklık < 5,5; üretim ayı<6	A
6	Operatör adı = yy; ortalama sıcaklık < 5,5; üretim ayı>=6	B
7	Ortalama sıcaklık<18; ürün miktarı<5; operatör adı=md; makine adı=paletleme; ürün miktarı<3; tezgâh boyutu=1600x3000x800; ortalama sıcaklık<9	B
8	Ortalama sıcaklık<18; ürün miktarı<5; operatör adı=md; makine adı=paletleme; ürün miktarı<3; tezgâh boyutu=1600x3000x800; ortalama sıcaklık>=9	A
9	Ortalama sıcaklık>=18,5; üretim ayı>=8; ortalama sıcaklık <21; operatör adı = md	A
10	Ortalama sıcaklık>=18,5; üretim ayı>=8; ortalama sıcaklık <21; operatör adı = yy	C
11	Ortalama sıcaklık<18,5; ürün miktarı<5; operatör adı=md; makine adı=tokatlama; parça adı=mil; tezgâh boyutu=700x450x500	B
12	Ortalama sıcaklık<18,5; ürün miktarı<5; operatör adı=md; makine adı=tokatlama; parça adı=mil; tezgâh boyutu=500x550x500	A
13	Parça adı = şase; ortalama sıcaklık < 15,5; makine adı = transfer; ürün miktarı < 3	C
14	Parça adı = şase; ortalama sıcaklık < 15,5; makine adı = transfer; ürün miktarı >= 3	A

#### 4. SONUÇLAR (CONCLUSIONS)

Bu çalışmada makine üretimi yapan bir işletmenin üretim verileri kullanılarak üretim süresine etki eden faktörlerin veri madenciliği yöntemleri ile belirlenebileceği gösterilmiştir. Beş farklı nitelik seçimi değerlendiricisi ve yedi farklı sınıflandırma algoritması çalışılarak üç aşamada en iyi modele ulaşılmıştır. Nitelik seçimi değerlendiricileri bilgi kazancı, CFS, sınıflandırıcı, kazanç oranı; sınıflandırma algoritmaları ise part, oneR, jrip, decision table, random tree, REPTree, LMT'dir. Bu işlemler Weka programı aracılığıyla uygulanmıştır.

Oluşturulan farklı veri kümelerinin hepsi için en iyi model random tree algoritması ile elde edilmiştir. Bunun nedeni random tree algoritmasının topluluk öğrenme algoritması olmasıdır. Random tree birçok bireysel kuralı öğrenerek kural üretmektedir. Bireysel öğrencileri oluştururken torbalama metodunu kullanmaktadır yani her bir düğüm rastgele seçilen belirteçlerin alt kümesi arasında en iyisini kullanarak dallanır. En iyi modelin elde edildiği veri kümesi genişletilmiş bilgi kazancı veri kümesidir. Bunun nedeni nitelikler arası etkileşimin devreye girip modelin performansını artırmasıdır. Genişletilmiş bilgi kazancı veri kümesinde kullanılan girdi nitelikleri parça adı, makine adı, üretim ayı, ortalama sıcaklık, operatör adı, tezgâh boyutu ve ürün miktarıdır. Sınıf niteliği üretim süresidir. Nitelik seçimi sonrasında daha fazla nitelik kullanmak yerine nitelik seçimi yöntemi ile oluşturulan veri kümesi modeli ifade etmekte yeterli olmuştur. Nitelik seçimi öncesi ve sonrasında bulunan veri kümeleri için uygulanan modellere göre tüm ölçütler karşılaştırıldığında en iyi sonucun elde edildiği algoritma random tree olmuştur.

Çalışmada yeterli veri sağlanamadığı için ele alınamayan ancak üretim süresi üzerinde etki edeceği düşünülen faktörler vardır. Bunlar; parça hammaddesi, parça boyutu, parça işleme zorluğu ve hazırlık süresi verileridir. İşletme ile yapılan görüşmelerde bu verilerin de veri tabanına dâhil edilmesi önerilmiştir. Modelden elde edilen sonuçlara göre işletmede üretim planlaması yaparken sipariş yoğunluğuna bağlı olarak teslim sürelerini azaltmak için A sınıfına ait olan ürünleri normal mesai, B sınıfına ait olan ürünleri fazla mesai, C sınıfına ait olan ürünleri ise fason mesai ile üretilmesini tercih etmelidir.

Sonuç olarak yapılan çalışmada, üretim sektöründe faaliyet gösteren ve özellikle de siparişe üretim yöntemine göre çalışan firmalarda üretim sürelerinin iyileştirilmesi için veri madenciliğinin uygulanabilirliği ortaya konmuştur. Uygulamalar sonucunda elde edilen bilgiler işletmelerin üretim süreçlerinde ileriye yönelik kararlar alınırken karar destek girdisi olarak kullanılabilir.

#### KAYNAKLAR (REFERENCES)

1. Harding J. A., Shahbaz M., Kusiak, A., Data mining in manufacturing: a review, *Asme The Journal of Manufacturing Science and Engineering*, 128 (4), 969–976, 2006.
2. Choudhary A. K., Harding, J. A., Tiwari, M. K., Data mining in manufacturing: a review based on the kind of knowledge, *Journal of Intelligent Manufacturing*, 20 (5), 501-521, 2009.
3. Usuga Cadavid J. P., Lamouri S., Grabot B., Pellerin R., Fortin A., Machine learning applied in production planning and control: a state-of-the-art in the era of

- industry 4.0., *Journal of Intelligent Manufacturing*, 31, 1531–1558, 2020.
4. Köksal G., Batmaz I., Testik M. C., A review of data mining applications for quality improvement in manufacturing industry, *Expert Systems with Applications*, 38 (10), 13448-13467, 2011.
  5. Sim S. K., Chan Y. W., A knowledge-based expert system for rolling element bearing selection in mechanical engineering design, *Artificial Intelligence Engineering*, 6 (3), 125–135, 1992.
  6. Romanowski C. J., Nagi R., *Data mining for design and manufacture: methods and applications*, Editör: Braha D., Kluwer Academic, Cilt 3, Dordrecht, 161–178, 2001.
  7. Giess, M. D., Culley, S. J., Shepherd, A., Informing design using data mining methods., *Asme International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, 207-215, 2002.
  8. Gardner M., Bieker J., Data mining solves tough semiconductor problems, *Proceedings of The Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Boston-ABD, 376–383, Ağustos 2000.
  9. Park K. S., Kim S. H., Artificial intelligence approaches to determination of cnc machining parameters in manufacturing: a review, *Artificial Intelligence Engineering*, 12, 127–134, 1998.
  10. Liao T. W., Chen J. H., Triantaphyllou E., Data mining applications in industrial engineering: a perspective, *International Conference on Computers and Industrial Engineering*, New Orleans-LA, 265–276, 1999.
  11. Skormin V. A., Gorodetski V. I., PopYack I. J., Data mining technology for failure of prognostic of avionics, *IEEE Transactions on Aerospace and Electronic Systems*, 38 (2), 388–403, 2002.
  12. Chen W. C., Tseng S. S., Wang, C. Y., A novel manufacturing defect detection method using data mining approach, *Lecture Notes in Artificial Intelligence*, 3029, 77–86, 2004.
  13. Batanov D., Nagarur N., Nitikhumkasem P., Expertmm: a knowledge based system for maintenance management, *Artificial Intelligence Engineering*, 8, 283–291, 1993.
  14. Çetin M., An application of data mining in a manufacturing industry, Yüksek Lisans tezi, Sakarya Üniversitesi, Fen Bilimleri Enstitüsü, Sakarya, 2009.
  15. Tapkan P. Z., Özmen T., Determining the yarn quality by feature selection and classification in a yarn production facility, *Pamukkale University Journal of Engineering Sciences*, 24 (4), 713-719, 2018.
  16. Bilekdemir G., Manufacturing lead time estimation using data mining techniques, Yüksek Lisans Tezi, Dokuz Eylül Üniversitesi, Sosyal Bilimler Enstitüsü, İzmir, 2010.
  17. Türkoğlu B., Komesli M., Ünlütürk M. S., An industrial case study on data mining, *International management information systems conference*, Ankara-Türkiye, 107-110, 26-28 Ekim 2018.
  18. Turker A. K., GÖLEÇ A., Aktepe A., Ersoz S., Ipek M., Cagil G., A real-time system design using data mining for estimation of delayed orders an application, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 35 (2), 709-724, 2020.
  19. Chen S., Li X., Liu R., Zeng S., Extension data mining method for improving product manufacturing quality, *Procedia Computer Science*, 162, 146-155, 2019.
  20. Nkonyana T., Sun Y., Twala B., Dogo E., Performance evaluation of data mining techniques in steel manufacturing industry, *Procedia Manufacturing*, 35, 623–628, 2019.
  21. Cheng Y., Chen K., Sun H., Zhang Y., Tao F., Data and knowledge mining with big data towards smart production, *Journal of Industrial Information Integration*, 9, 1-13, 2018.
  22. Emre İ. E., Selçukhan Erol Ç., Statistics or data mining for data analysis, *International Journal of Informatics Technologies*, 10 (2), 161-167, 2017.
  23. Gürbüz F., Özbakır L., Yapıcı H., Data mining application on component reports of an airline company in turkey, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 24 (1), 73-78, 2009.
  24. Koyuncugil A., Özgülbaş N., Data mining: using and applications in medicine and healthcare, *International Journal of Informatics Technologies*, 2 (2), 2009.
  25. Fayyad U., Piatetsky-Shapiro G., Smyth P., From data mining to knowledge discovery in databases, *AI Magazine*, 17 (3), 37-37, 1996.
  26. Çalış A., Kayapınar S., Çetinyokuş T., An application on computer and internet security with decision tree algorithms in data mining, *Journal of Industrial Engineering*, 253, 2-19, 2014.
  27. Kaya H., Köymen K., Data mining concept and application areas, *Eastern Anatolia Region Researches*, 159-164, 2008.
  28. Han J., M. Kamber, *Data mining: concepts and techniques*, Morgan kaufmann Publishers, 3, USA, 2001.
  29. Budak H., Feature selection methods and a new approach, *Süleyman Demirel University Journal of Natural and Applied Sciences*, 221, 21-31, 2018.
  30. Şeker S. E., <http://bilgisayarkavramlari.sadievrenseker.com/2012/11/13/information-gain-bilgi-kazanimi/>, Yayın tarihi Kasım 13 2012, Erişim tarihi Nisan 3, 2020.
  31. Karegowdal A. G., Manjunath A. S., Jayaram M. A., Comparative study of attribute selection using gain ratio and correlation based feature selection, *International Journal of Information Technology and Knowledge Management*, 2 (2), 271-277, July-December 2010.
  32. Hall M., <https://Weka.sourceforge.io/doc.dev/Weka/attributeSelection/ClassifierAttributeEval.html>, Erişim tarihi Nisan 3, 2020.
  33. Miles J., Banyard P., *Understanding and using statistics in psychology: a practical introduction*, Sage, 2007.
  34. Brownlee J., <http://spssistatistik.net/spss-korelasyon-analizi/>, Yayın tarihi Temmuz 13 2016, Erişim tarihi Nisan 3, 2020.

35. Max B., Principles of Data Mining, 1, Springer, London-UK, 2001.
36. Nisbet R., Elder J., Miner G., Handbook of statistical analysis and data mining applications, Elsevier, Burlington, 2009.
37. Yang D. P., Jin-Lin L., Ran L., Zhou C., Applications of data mining methods in the evaluation of client credibility, Applications of Data Mining in E-Business and Finance, Amsterdam, 35-43, 2008.
38. Frank E, Witten I.H., Generating accurate rule sets without global optimization, 15th International Conference on Machine Learning, Wisconsin-USA, 24-27 July 1998.
39. Holte R., Very simple classification rules perform well on most commonly used datasets, Machine Learning, 11, 63-91, 1993.
40. Novakovic J., Strbac P., Bulatovic D., Toward optimal feature selection using ranking methods and classification algorithms, Yugoslav Journal of Operations Research, 211, 119-135, 2011.
41. Kalmegh S., Comparative analysis of the Weka classifiers rules conjunctive rule and decision table on indian news dataset by using different test mode, International Journal of Engineering Science Invention, 73, 1-9, 2018.
42. Shahzad W., Asad S., Khan M. A., International Journal of Physical Sciences 818, 885-896, 2019.
43. Kalmegh S., Analysis of weka data mining algorithm reptime, Simple Cart and RandomTree for Classification of Indian News, International Journal of Innovative Science, Engineering & Technology, 22, 2015.
44. Landwehr N., Hall M., Frank E., Logistic model trees, Machine Learning, 591 (2), 161-205, 2005.
45. Provost F., Domingos P., Tree induction for probability based ranking, Machine Learning, 52 (3), 199-215, 2003.
46. Nizam H., Akın S.S., Sosyal medyada makine öğrenmesi ile duygu analizinde dengeli ve dengesiz veri setlerinin performanslarının karşılaştırılması, Türkiye'de İnternet Konferansı, İzmir-Türkiye, 2014.
47. Hossin M., Sulaiman M.N., A review on evaluation metrics for data classification evaluations, International Journal of Data Mining & Knowledge Management Process, 5 (2), 1-11, 2015.
48. Kumar G., Evaluation metrics for intrusion detection systems-a study, International Journal of Computer Science and Mobile Applications, 2 (11), 11-17, 2014.
49. Ferri C., Hernández-Orallo, J., Modroiu R., An experimental comparison of performance measures for classification, Pattern Recognition Letters, 30, 27-38, 2009.
50. Handelman G.S., Kuan Kok H., Chandra R.V., Razavi A.H., Huang S., Brooks M., Lee M.J., Asadi H., Evaluation metrics of machine learning methods, American Journal of Roentgenology, 212, 38-43, 2019.
51. Sebastiani F., Machine learning in automated text categorization, ACM Computing Surveys CSUR, 341, 1-47, 2002.
52. Göktepe A.B., Agar E., Lav A.H., Comparison of multilayer perceptron and adaptive neuro-fuzzy system on backcalculating the mechanical properties of flexible pavements, ARI The Bulletin of the Istanbul Technical University, 543, 65-77, 2004.
53. Çavuşoğlu Ü., Kaçar S., The performance analysis of data mining algorithms for anomaly detection, Academic Platform Journal of Engineering and Science, 72, 205-216, 2019.
54. Bilgin M., Performance analysis of classical machine learning methods in real data sets, Breast, 29, 683-688, 2017.
55. Kumar R., Indrayan A., Receiver operating characteristics roc curve for medical researchers, Indian Pediatrics, 48, 277-287, 2011.
56. Gnanambal S., Thangaraj M., Meenatchi V.T., Gayathri V., Classification algorithms with attribute selection: an evaluation study using Weka, Intelligent Journal Advanced Networking and Applications, 9 (6), 3640-3644, 2018.
57. Ang J., C., Mirzal A., Haron H., Nuzly H., Hamed A., Supervised unsupervised and semi-supervised feature selection: a review on gene selection, IEEE/ACM Transactions on Computational Biology and Bioinformatics, 13 (5), 971-989, 2016.