

Üretim Sektöründe Veri Madenciliği Uygulamaları: Literatür Taraması *Data Mining Applications In Manufacturing Industry: A Literature Review*

Ezgi DEMİR¹, Sait Erdal DİNÇER²

Başvuru Tarihi: 30.04.2020

Kabul Tarihi: 23.06.2020

Öz: Bu çalışma üretim sektöründe yapılan veri madenciliği çalışmalarının (tez, makale, bildiri) bir literatür taraması yapmayı amaçlamaktadır. Üretim sektöründe veri madenciliği uygulamaları üretim süresi tahmini, ürün geliştirme, kalite optimizasyonu, imalat sektörü, kar marjı fazla ürün üretme, üretim sürecinde raf yönetimi, üretim planlama, tedarikçi seçimi gibi birçok alanda kullanılmaktadır. Veri madenciliği uygulamaları son yıllarda literatürde aktif olarak kullanılmaktadır. Bu çalışma veri madenciliği metodolojisine değil, üretim alanında veri madenciliği teknikleri, yazılımları ve kullanılan algoritmalara yoğunlaşmaktadır. Çalışma kapsamında 2010 sonrası üretim alanında veri madenciliği tezleri, makaleleri, bildirimleri baz alınmıştır. Bu çalışmanın yapılmasının literatüre katkısı veri madenciliğinin eğitim, sağlık, bilişim, ticaret, hayvancılık, duygudurum analizi, bankacılık, pazarlama, çevre, tarım gibi alanlarda sıklıkla kullanılırken, üretim alanında daha az kullanılmasının sebebini açıklamaya çalışmaktır. 2017 ve 2018 yıllarında yapılan yoksis veritabanına kayıtlı 69 adet veri madenciliği tezlerinden sadece 2 tanesi üretim sektöründe yapılmıştır. Bu çalışma ile üretim sektöründe kullanılan veri madenciliği teknikleri ifade edilerek bu alanda çalışma yapmak isteyenlere referans olması amaçlanmaktadır.

Anahtar Kelimeler: Veri Madenciliği, Üretim, Veri Madenciliği Teknikleri

Abstract: The paper aims to make a literature review of data mining studies (thesis, article, paper) in manufacturing sector. Data mining applications in manufacturing sector are used in many fields such as production time estimation, product development, quality optimization, manufacturing sector, production of surplus margin, shelf management in production process, production planning, supplier selection. Data mining applications have been used actively in the literature in recent years. This study focuses on data mining techniques, software and algorithms used in the production area, not on data mining methodology. Within the scope of the study, theses, articles and papers in the field of production after 2010 were taken as basis. The contribution of this work to the literature is to explain why data mining is frequently used in education, health, informatics, trade, animal husbandry, emotion-situation analysis, banking, marketing,

¹ Arş. Gör., Piri Reis Üniversitesi, Yönetim Bilişim Sistemleri Bölümü, Marmara Üniversitesi-Yöneylem Araştırması Ana Bilim Dalı Doktora Öğrencisi, Tuzla/İSTANBUL, edemir@pirireis.edu.tr, ORCID : 0000-0002-7335-5094.

² Prof. Dr., Marmara Üniversitesi, Yöneylem Araştırması Ana Bilim Dalı, Göztepe/İSTANBUL, edincer@marmara.edu.tr, ORCID: 0000-0002-8310-1418.

environment, agriculture, and to explain why it is used less in production. Only two of the 69 data mining theses registered in the 2017 and 2018 database were conducted in the manufacturing sector. With this study, the data mining techniques used in the manufacturing sector are stated and it is aimed to be a reference for those who want to study in this field.

Keywords: Data Mining, Manufacturing, Data Mining Techniques

1. GİRİŞ

Artan talep ve endüstrileşme, şirketler için üretim planlaması alanında seri üretime geçme noktasında zorluklar yaratmıştır. Diğer taraftan endüstri 4.0 ve dijital teknoloji araçları gibi bilgi ve teknoloji çağı fırsatlarıyla seri üretime geçilmiştir. Piyasalarda, müşterilerin ihtiyaç analizi ve talep analizi, şirketlerin ürünleri depolarında bulundurma veya tedarik etme süreleri, şirket yatırımcılarının ve finans departmanlarının kar marjını yükseltmek ve maliyetleri düşürme çabası üretim planlama ve kontrol noktasında entegrasyonu zorlaştırmıştır. Rekabet ortamının artmasıyla pazarda yer edinmeye ve yerini korumaya çalışan şirketler, hem dinamik hem ekonomik olarak üretim sistemlerini daha verimli ve etkin bir şekilde planlamalıdır.

Üretim planlama ve programlama konusunda ürün yöneticilerine önemli bir görev düşmektedir. İşin en zorlu kısmı mevcut ürünlere hangi zaman aralığında ve hangi miktarlarda ihtiyaç duyulacağını belirlemektir. Buna bağlı olarak üretime etki eden faktörleri de göz önüne alarak üretilen ürünlerin tedarikçilerden-tüketicilere, sürecin etkili ve doğru bir şekilde yönetilmesi gerekmektedir. Üretim sürecinde planlama şemasına bakılacak olursa Şekil 1.de yer alan bir süreç beklenmektedir. Üretim sürecinin etkili ve planlı bir şekilde gerçekleşebilmesi için süreçlere Şekil 2. de görüldüğü üzere talep analizi, maliyet analizi, duyarlılık analizi gibi etkin denetim mekanizmaları eklenmelidir. İlerleyen teknolojiler, artan sanayileşme sonucunda verilerin artması sonucunda analizler zorlaşmış ve istatistiksel metotlar yetersiz kalmıştır. Bu noktada veri madenciliği ve uygulamaları geliştirilmiştir.

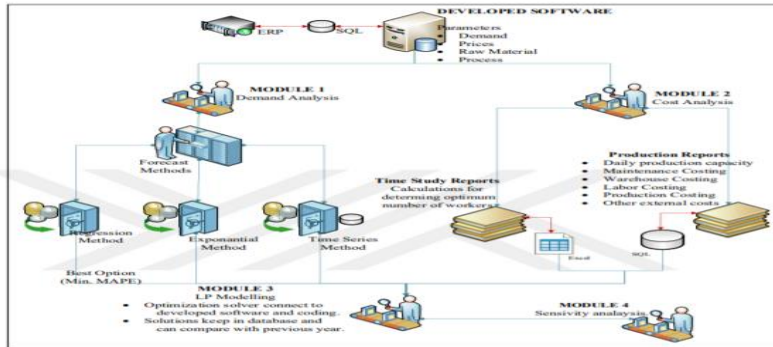
Literatürde üretimin içerisinde yer alan her alanda planlama, imalat, satış, kalite kontrol, hata tespiti, hatalı ürünlerin ayıklanması noktasında çalışmalar bulunmaktadır. Literatür çalışması; ulusal çalışmalar alanında, yöksen tez verileri (583 tez), ulusal makaleler, kitaplarda yer alan uygulama örneklerini içermektedir. Uluslararası çalışmalar ise uluslararası alanda yapılan yabancı tezler, uluslararası makaleler ve bazı teknoloji-bilişim şirketlerinin üretim alanında yaptıkları çalışmaları içermektedir. Öncelikle “veri madenciliği” sorgulaması yapılmış, bu sorgulamadan gelen veriler doğrultusunda üretim planlama sürecini kapsayan süreçler ele alınmıştır. Teyit etmek amacıyla üretim optimizasyon alanında yapılmış tezlerde başlık olarak olmasa da içerisinde veri madenciliği

barındırıyor olabilir düşüncesinden “üretim” le ilgili aramalar yapılmıştır. Bu araştırmalar sonucunda ulusal, uluslararası tezler ve makaleler bir bütün olarak ele alınmıştır.

Bu çalışmada veri madenciliği ve uygulamalarından bahsedilerek üretim sektöründe kullanılan veri madenciliği teknikleri ve algoritmaları ile ilgili literatür taraması verilmeye çalışılmıştır.



Şekil 1. Üretim Planlama Süreçleri (Topoyan, 2019)



Şekil 2. Üretim Sürecinde Optimizasyon (Topoyan, 2019)

2. VERİ MADENCİLİĞİ

Veri madenciliği büyük miktarlardaki verilerin araştırılarak anlamlı kalıplara ve kurallara göre keşfedilmesidir. Veri madenciliği ile ilgili birçok tanım yapılabilir. En temel tanımı çok miktardaki büyük verilerden öncesinde bilinmeyen verilerin çıkartılarak anlamlı verilere ulaşılmasıdır. Bu durum aynı zamanda verilerin belirli kalıplara göre karar verebilmek için işlenmesi anlamına da gelmektedir. Bir diğer ifade ise veri tabanında bilgi keşfi olarak tanımlanmaktadır.

Hayatımızın her adımında yaptığımız işler; akıllı telefonlar, akıllı uygulamalar, akıllı devlet uygulamaları, mail sistemleri, internet ve mobil tabanlı yazılımlar, coğrafi bilgi sistemleri sayesinde kayıt altına alınmaktadır. Bu durum gündelik hayatta Büyük Veri Yığınları' nın oluşmasına neden olmaktadır. 1992 yılında yayınlanan bir çalışmada William ve arkadaşları; Dünya üzerindeki veri

miktarının her yirmi ayda ikiye katlandığı tahmininde bulunmuşlardır (Frawley, Piatetsky-shapir, & Matheus, 1992). Bu durum daha 1992 yılında 2 katına çıkarken günümüzde veri yığınları üstel olarak ilerlemektedir. Bu veri yığınlarının belirli işlemlerden geçirilerek kullanılabilir bilgiye dönüştürülmesi sürecinin tamamı, 1989 yılında gerçekleştirilen KDD (Knowledge Discovery in Databases) çalışması ile 'Veri tabanından bilgi keşfi' olarak ilk kez hayatımıza girmiştir (Sağın, 2018). Veri madenciliği de, belirlenen veri kümesi üzerinde; sınıflama, kümeleme, birliktelik kuralları vb. yöntemler ve bu yöntemlere uygun algoritmaların kullanılarak çeşitli örüntülerin elde edildiği, VTBK (Veri Tabanından Bilgi Keşfi) sürecinin bir aşamasıdır. Veri Tabanında Bilgi Keşfi'ni tanımlamak amacıyla sınıflama, kümeleme, birliktelik kuralları vb. yöntemler ve bu yöntemlere uygun algoritmalar geliştirilmiştir. Veri madenciliği kavramı gelişmeden önce klasik istatistiksel yöntemler uygulanarak veriler arasında ilişki kurulmaktaydı, fakat verilerin yoğunluğu ve parametreleri arttıkça, büyük ölçekli veriler yaygınlaşmaya başlamış olup, klasik istatistiksel yöntemlerle büyük ölçekli veriler içerisinde bilgi yığınları yeterince açıklanamamaya başlamıştır. Klasik istatistiksel yöntemlerin bir diğer dezavantajı ise verilere ulaşmak için büyük kurum kuruluşlara (TUİK, OECD, WORLD BANK) ihtiyaç duyuluyor olması ve bu sürecin hem maliyetli hem de uzun bir süreç olmasıdır. Günümüzde veri yığınlarına mobil, internet, bilgisayar teknolojilerinin getirdiği uygulamalar ile daha az maliyetle ve daha az sürelerde kolaylıkla ulaşılabilmektedir.

Veri madenciliği içerisindeki bilgiler, büyük veri yığınları içerisinde saklı bir halde, tahmin ve deneyimlerle önceden bilinmemektedir. Ve bu verilerden veri madenciliği teknik ve algoritmaları ile süreçten geçtikten sonra işletmelerde, sağlık sektörü, eğitim sektörü, ekonomi, istatistik, hayvancılık, bilişim sektörü gibi alanlarda anlamlı ilişkiler çıkmaktadır. Bu ilişkiler, bu alanlarda hem ekonomik, hem kaliteli hem de teknolojik gelişmeler sağlamaktadır.

2.1. Veri Madenciliği Temel Kavramları

Veri, işlenmemiş her türlü anlamlı kelime, sayısal ifade demektir. Verilerin bilgisayarın anlayabileceği şekle dönüştürülüp kaydedilmesi, daha sonra çeşitli veri madenciliği yöntemleri uygulanmasıyla ulaşılan değer, bilgidir (Sağın, 2018).

Büyük ölçekteki veriler içerisinden çıkarılan, çok daha küçük ölçekli bilgi, karar vermede kullanılan önemli bir araçtır. Bu demektir ki büyük ölçekteki veri karar verme konusunda anlamlı değildir (Gürsaka, 2001).

Veri madenciliği ile ilgili çalışmalarda iteratürde geçen terimler: 'Data', 'Information', 'Knowledge' ve 'Wisdom' şeklinde sıralanabilir. Bu terimler temelinde 'bilgi' olarak yer almaktadır. Fakat bu

terimler arasında farklılıklar ve çeşitli dönüşümler bulunmaktadır. En önemlisi bu terimler arasında hiyerarşik bir bağ vardır (Şekil 3.)



Şekil 3. Veri Madenciliği Terimleri Hiyerarşisi (Bernstein, 2011)

Şekil 3.de gösterilen hiyerarşi ifade edilecek olunursa, örneğin yolda araçla belli bir hızla giderken kırmızı ışık yanması, yol, araç, trafik lambası, kırmızı, hız gibi birçok terimi barındıran veriler kısmı yer almaktadır. Bu kısım “VERİ(DATA)” olarak ifade edilir. Yol boyunca kırmızı, sarı, yeşil gibi ifadelerin trafik kuralı haline getirilmesi bir dönüşüm, enformasyondur. Bu kısım “ENFORMASYON(INFORMATION)” olarak ifade edilmektedir. Işıklara göre nasıl hareket edilmesi bir bilgidir. Bu kısım ise “BİLGİ(KNOWLEDGE)” olarak adlandırılır. Ve son olarak kırmızı ışıkta durma reaksiyonunun gösterilmesi bir bilgeliktir. Bu aşama ise “AKIL(WISDOM)” olarak ifade edilmektedir. Buradan şunu ifade edebiliriz. Piramitin yukarısına çıktıkça bilgi azalmakta fakat anlamlılık artmaktadır.

Veri madenciliği; elimizdeki büyük veri yığınları içerisinde gizlenmiş, önceden bilinmeyen, sezgi ve tecrübelerimizle tahmin edemediğimiz ayrıca öğrenilmesi durumunda işletmeye büyük yarar sağlayacak değerli bilginin gün yüzüne çıkarılması işlemidir (Özkan, 2016), (Silahtaroğlu, 2016).

Literatürde veri madenciliği ile ilgili çok sayıda tanım vardır. Sıralanacak olunursa;

Önemsiz olmayan bilgilerin, verilerde gizli, önceden bilinmeyen ve potansiyel olarak faydalı bilgilerle ifşa edilmesidir (Frawley, Piatetsky-shapir, & Matheus, 1992).

Büyük veri tabanlarından bilgi çıkarmak için veri madenciliği; makine öğrenmesi, veri tabanı, örüntü tanıma, görüntüleme tekniklerini ve istatistiğini birleştiren disiplinlerarası bir alandır. (Cabena, Hadjinian, Stadler, Verhees, & Zanasi, 1998).

Tek başına ham verinin sunamadığı bilgiyi ortaya çıkaran veri analizi süreci, veri madenciliği olarak ifade edilir (Jacobs, 1999).

Yüksek düzeyde tahmin edici anahtar parametrelerin binlerce olası parametreden izole edilebilmesi de veri madenciliği olarak ifade edilir (Kittler & Wang, 1999).

Hacim olarak geniş veri kümesinde saklı kalıpları ve trendleri keşfetme süreci veri madenciliği olarak tanımlanmaktadır (Özekes & Çamurcu, 2003).

Daha önce bilinmeyen, saklı, ifade edilebilir ve hacim olarak geniş veritabanlarından faydalı modeller sunan veritabanlarındaki bilgilerin keşfedilme süreci de veri madenciliği olarak ifade edilmektedir (Han & Kamber, 2006).

3. ÜRETİMDE VERİ MADENCİLİĞİ UYGULAMALARI

Üretimde veri madenciliği tezleri veri madenciliği kullanım alanları, kullanılan programlar, metot ve teknikler ve içerik sütunları olmak üzere aşağıdaki şekilde özetlenerek sıralanmıştır. İçerik olarak bahsedecek olunursa;

* Bir üretim işletmesinde soğuk press işlem yapan makinalar belli aralıklarla durmaktadır. Bu durma süresi işletmeye üretim gerçekleştirilmediğinden ekonomik olarak kayba neden olmaktadır. Bu sürenin azaltılması amacıyla makinaların optimize kullanım süreçlerinde tahminleme yapılmıştır. Bu tez çalışmasında veri madenciliği tekniklerinden sınıflandırma, karar ağacı ve ağırlıklı öğrenme algoritmaları kullanılmıştır. Bu metotlar Weka ve Rapidminer programlarında uygulanmıştır.

* Bir içecek üretimi yapan firmada, verilerin hacim, satış kanalı, teslimat lokasyonu, faaliyet tipi, ürün grup, marka ve zaman boyutuna göre oluşturulan veri küpü öncelikle analiz edilmiş ve veri madenciliğinde çıkabilecek sonuçlar kestirilmeye çalışılmıştır. Oluşturulan model sonucunda dağıtım yapılan müşterilerin faaliyetlerini sürdürdükleri faaliyet tipi, ürün grubu, ürün boyutu (hacmi) ve günlük alınan kasa miktarlarına göre kümelemenin önem kazandığı görülmüştür. İşletme, bölgesel olarak üretim yaptığı müşterilerinin 5 aylık işlem verileri üzerinde veri madenciliği uygulaması yapmıştır. Uygulamada Kümeleme Yöntemi ve Karar Ağaçları metotları kullanılmıştır. Firma müşterilerinin günlük işlem verilerini dikkate alarak firmanın dağıtım stratejilerini, müşteri gruplarına göre özel olarak belirleyebilmesi amacıyla kümeleme yöntemi kullanılarak müşteriler homojen gruplara ayrılmıştır. Karar ağaçları uygulanarak çok büyük hacimdeki veri içinde gizli kalmış ve firma için faydalı olabilecek bilgiler araştırılmıştır. İstatistiksel metotlar kullanılmıştır.

* Bir lojistik firmasında müşterilerden alınan siparişlere ait ana bilgilerin tutulduğu tablo ve müşterilerin talepleri doğrultusunda tutulan çizelgelerden faydalanılmıştır. Bu çalışmada iki boyutlu

uzayda talep ve termin tarihlerine uyum yüzdesine bakılarak kaybedilen ve çalışılmaya devam edilen müşterilere geometrik olarak denk gelen noktaları ayıran sınır değerin bulunması amaçlanmıştır. Bu modelde sınıflandırma ve karar ağaçları metotları uygulanmıştır.

* Bir üretim işletmesinde tedarikçi seçimi yapılmıştır. Veri setinde bulunan değişkenlerin içerisinde, tedarikçi seçimi için karar değişkeni olarak en önemli değişken “gecikme” olarak seçilmiştir. Her bir değişkenin gecikme ile olan ilişkisi kurulan modelde incelenmiş ve gecikmeyi etkileyen değişkenler tespit edilmiştir. Başarı düzeyi ölçülürken “Gecikme = 1” ,“Zamanında teslim = 0”, sırasıyla (Kalite Belgesi(ISO-TSE)- ARGE - Garanti Belgesi - Sektör Grubu - İdari Personel Sayısı - Teknik Personel Sayısı - Kuruluş Şekli - Firma Tipi – İl - Kalem Sayısı değişkenleri alınmıştır. Analizlerde, yapay sinir ağı, karar ağacı algoritmalarından C&R Tree, C5.0, Chad ve bu çalışmada kümeleme algoritmalarından K-Means, Kohonen ve Two Step algoritmaları kullanılmış, ancak bu algoritmalarından K-Means daha sağlıklı sonuç vermiştir. Bu analiz SPSS Clementine programıyla yapılmıştır.

* Bir üretim işletmesinde veri madenciliği ayrıntılı bir şekilde ele alınmıştır. Uygulama kısmında, bir üretim işletmesinde, üretilen ürünlerin uygunsuz olarak ayrılmasının nedenleri belirlenerek, bu nedenlerin analizi ile uygunsuz ürünlerin sayısını azaltıcı stratejiler geliştirilmesi hedeflenmektedir. Vardiya düzeni, üretim ayı, üretim günü, hata grupları(kişiden, makineden, makine ve kişi), üretim sıklığı, ambalaj içine konulan miktar, fabrika, makine arızası, ürün grubu, müşteri, çalışan kadro, profil türü, üretim türü (serim üretim ve seri üretim olmaması) olarak gruplara ayrılmıştır. Neden analizi için karar ağaçları ve yapay sinir ağları ile bir model geliştirilmiştir. SPSS Clementine 11.1 uygulaması kullanılmıştır.

* Bir halı imalatı yapan işletmede, halı imalat sürecini ve ürün kalitesini iyileştirme amaçlı veri madenciliği uygulamaları yapılmıştır. Ürün hata verileri ve makine verileri ayrı ayrı modellenmiştir. Uygunluk analizi ile değişken indirgeme için etkili bir yaklaşım kullanılmıştır. Halı hatalarının ve makineden kaynaklı sebeplerin bilgi kazancı tekniği ile standart bir noktaya taşınmıştır. Hata id, tarih, tezgah, posta, çalışan, kalite, desen, renk, en, boy, hata bölümü, müşteri, hata türü, hata sebebi, barkod sıra, barkod tarih gibi değişkenler kullanılmıştır. İstatistiksel yöntemler kullanılarak, ID3 algoritmasının uzantısı olarak geliştirilen C4 algoritması, modellemede karar ağacı yöntemi olarak C4.5 algoritması ve yapay sinir ağları (Radyal Temelli Fonksiyon Ağları) kullanılmıştır.

* Otomotiv alanında önde gelen bir firmada optimizasyon çalışması yapılmıştır. Çalışma firmanın kesim bölümünde gerçekleşmiştir. Dikiş Hatası, Kesim Hatası, Yarım Parça, Kumaş Hatası, Seri Sonu Parça, Desende Kayma, Renk Farkı gibi hata değişkenleri belirlenmiştir. Kanonik korelasyon analizi ile girdi ve çıktı değişkenleri arasındaki ilişkiler yorumlanmış ve son olarak çoklu regresyon analizi ile

her bir çıktı için ayrı tahmin denklemleri oluşturulmuştur. Girdi değişkenleri için en uygun değerlerin atanması sonucuna, en ve boy için istenilen değerlere yüzde yüz oranında ulaşılrken, delik çapı için ise yüzde elli beş oranında bir iyileştirme sağlanmıştır. Gams programıyla hata kodları ve hata türleri tanımlanarak, istatistiksel metotlar ve çoklu regresyon analiziyle karşılaştırmalar yapılmıştır.

* Bu çalışmada üretim süresi tahmini yapılmıştır. Bu amaçla su sayacı üretim süreci incelenmiştir. Karar ağacı algoritmalarından, C4.5 algoritması kullanılarak WEKA programında analiz edilmiştir. Bu amaçla, veri madenciliği tekniklerinden birliktelik kuralları kullanılarak matematiksel programlama modeli oluşturulmuş ama büyük boyutlu problemlere uygulayabilmek amacıyla genetik, hibrit bir algoritma geliştirilmiştir. Uygulama olarak perakendecilik sektörü seçilmiştir. Çalışmada Ms-sql istatistiksel metotlar, birliktelik kuralı ve LINGO programı kullanılmıştır.

* Bu çalışmada, perakende sektöründe hizmet veren bir kuruma ait işlemsel bazda 2009 yılının ilk 9 ayını kapsayan veriler kullanılmıştır. Firmanın hedef kitesini, perakende satış yapan işletmeler ve catering işletmeler oluşturmaktadır. Verilerdeki parametreler; ürünler, ürün kodları, ürün ambalaj bilgileri, ürün ana marka kodları, alt markaya ait kodlar, kategori ve alt kategorilere ait bilgiler, ürün kategorisi ve müşteri bilgileri şeklinde sınıflanmıştır. Amaç kar bazlı yüksek müşterilerin gruplanmasını sağlamak ve bu faktörler üzerinde yoğunlaşmaktır. Veri manipülasyonu yöntemleri uygulanmış ve sonrasında EDA yöntemlerinden faydalanılarak, veri madenciliği modellemelerinden sınıflama ve regresyon modelleri, kümeleme ve birliktelik kuralları incelenmiştir. Aynı zamanda C&R Tree ve CHAID algoritmaları kullanılarak “kar tutarı yüksek” müşterilerin bu grupta toplanmasında hangi faktörlerin etkili olduğu belirlenmeye çalışılmıştır. Çalışmalar SPSS Clementine paket programıyla yapılmıştır.

*Üretim lojistik alanında karesel model, ürünlerin depo içerisine yerleşiminde faydalanılmış olan çok boyutlu ölçekleme analizi, sipariş toplama sırasındaki rotaların optimize edilmesinde kullanılan Gezgin satıcı problemi anlatılmıştır. Uygulamada, tek yönlü varyans analizi, birliktelik kuralları, Bağımsız örneklem t-testi, uygulamada kullanılan metotlar ve eşleştirilmiş örneklem t-testi, uygulamada kullanılan matematiksel modelin çözümünde karesel modelden ve istatistiksel metotlardan yararlanılmıştır.

*Üretim sektöründe iş yapan bir firmada kalite kontrol departmanında çalışma yapılmıştır. Kalite kontrol sürecinde ölçüm verileri çok fazladır ve verilerin hacmi arttıkça anlaşılabilirlik oranı azalmaktadır. Varyasyonlar kalitenin düşmanıdır ve her şeyde varyasyon bulunmaktadır. Bu tez çalışmasında veri madenciliği yöntemlerinden olan sınıflandırma algoritmaları ile kalite kontrol sürecinde bir karar destek sistemi uygulaması yapılmıştır. Bu çalışma hazırlanırken veri madenciliği

için Çapraz Endüstri Standart İşlem Modeli (CRISP) kullanılmıştır. NaiveBayes ve SMO algoritmaları C4.5 Random Forest Sınıflama Algoritmaları, R, Weka ve Minitab programları kullanılmıştır. Üretim sektöründe veri madenciliği ile ilgili ulusal ve uluslararası makaleler ise şu şekilde sıralanmıştır.

* Sinema sektöründe SVM (destek vektör makinaları) ve Latent Class algoritmalarıyla müşteri ürün derecelendirme yapılmıştır. İstatistiksel yöntemler, t-test ve f-test istatistikleriyle modellenmiştir.

* Bir gözlük üretimi yapan firmada 1 aylık verilerle üretimin kalitesi artırılması için ham madde, hassaslık ölçütleri, üretim ağı ve üretim sınıfı karakteristik özelliklerine dayanılarak 11.320 üretim verisi optimal kalite süreci için veri madenciliği ile gruplandırılmıştır. Karar ağaçları sınıflandırma metotlarıyla CHAID, ID3, C5.0 algoritmaları kullanılmıştır.

* Veri Madenciliğinde ürün sürecinde optimallik sınaması yapılmıştır. 3 ürün grubunun, 5 istasyondan geçerken istasyonlardaki makine sayısına göre geçirdiği işlemlere göre optimal üretim sürecinin veri madenciliği ile belirlenmesi sağlanmıştır.

* Sprey üretim sürecinde belli parametrelere göre üretim sürecinin veri madenciliğine göre analiz edilmiştir. Sınıflandırma algoritmaları ve karar ağaçları kullanılmıştır.

* Bir işletmede aynı işi yapan eski – yeni 2 makinenin kapasiteleri göz önüne alınarak veri madenciliği ile sınıflama yapılmıştır. Sınıflama algoritmaları, karar ağaçları ve Naive Bayes algoritması kullanılmıştır.

* Üretim sürecinde mevcut gecikmelerden veri madenciliği teknikleri ile gecikme tahmini yapılmıştır. Sınıflandırma algoritmaları ve lineer regresyon tekniği kullanılmıştır.

* Bir üretim işletmesinde, 3 farklı otomasyon sistemini veri madenciliği teknikleri ile karşılaştırılmıştır. İlk otomasyon sistemi periyodik çalıştırma, ikincisi otomatik çalıştırma, üçüncüsü geçici çalıştırma şeklindedir. Seri numarası, üretim hattı, makine id, üretimin başlama süresi ve bitiş süresi ve hatalı ürün parametrelerine göre veri madenciliği yapılmıştır. Veri madenciliği tekniklerinden, sınıflandırma algoritmaları grubundan CHRIS algoritmasına göre analiz edilmiştir.

* Bir üretim işletmesinde, üretim sürecinde üretim döngüsü tahminlemesi yapılmıştır. Yöneylem araştırması tekniklerinden şebeke modellerine uygun olarak en erken başlama, üretime girme ve en geç bitme süreleri tahmin edilmiştir. İstatistiksel metotlarla tahminleme yapılmıştır.

* Bir üretim işletmesinde ürün kalitesini geliştirmede, tarih, dizayn, renk, departman, id, çalışan, nitelik, ağırlık, yükseklik, boyut, hata sebebi, hata tipi, müşteri, barkod numarası, barkod tarihi gibi değişkenlere göre SPSS Clementine programında istatistiksel metotlarla sınıflandırma yapılmıştır.

* Araba üretim endüstrisinde yer alan bir firmada üretim süresi birliktelik kurallarıyla analiz edilmiştir. Bayesian ve markov zincirleri kullanılarak, zaman serileri analizi yapılmıştır.

* En son 2017 de yapılan bir çalışmada Standard Classification and Regression Trees (CART) • Boosting Tree (Boost) • Random Forest (RF) • Multivariate Adaptive Regression Splines (MARS) • Support Vector Machine (SVM) • K-Nearest Neighbour (KNN) • Multiple Regression (MR) • Neural Networks (Sinir ağları) methodları ele alınarak hangi metodun tahminlemede daha iyi sonuç verdiği araştırılmıştır. Bu noktada neural network(sinir ağları) istatistiksel sonuçlara göre önde çıkmıştır. Error rate 1 olarak çıkmıştır. Bu analiz istatistiksel metotlarla yapılmıştır.

4. SONUÇ

Üretim sektöründe veri madenciliği uygulamalarına bakıldığında, geleneksel olarak verilerin akışı tek taraflı işlemektedir. Günümüz üretim sistemlerinde ise verilerin akışı çift taraflı olarak yönlendirilmektedir.

Üretim sürecinde giriş işlemlerinden çıkış işlemlerine kadar yoğun bir süreç bulunmaktadır. Bu süreçte giren ürünlerin optimize edilmesinden, üretim sürecinin kısaltılmasından, üretim sürecinde hata paylarının azaltılmasından, paketlenme, tedarik, müşteri talebi gibi süreçler, büyük hacimli veri yığını oluşturmaktadır. Bu süreçlerin hepsinde veri madenciliği teknikleri kullanılabilir. Kullanılan algoritmalar, işletme açısından en az maliyetle ve en az girdiyle üretim sürecini tamamlama hedefi taşıırken, ürün açısından güncel, arz-talep-kalite üçlüsüne en uygun ürün üretimi hedefini taşır.

Üretim sürecinde veri madenciliği tekniklerine bakıldığında, eski çalışmalar SQL tabanlı veri analizi uygulamaları geliştirirken, yeni çalışmalarda Weka, Rapidminer gibi paket programlar kullanılmaktadır. Ama Weka, Rapidminer gibi paket programlar durağan veriler için kullanılırken, dinamik tabanlı ve açık kaynak kodlu yazılımlar R, Python gibi programlarda gelişmeler mevcuttur. Veri madenciliği uygulamaları sanayi uygulamalarında eski çalışmalarda istatistiksel uygulamalı metotlar tercih edilmiştir. Fakat veri yığınlarının kontrolsüz artışı istatistiksel metotları yetersiz kılmış, kodlama tabanlı algoritmalara ihtiyaç duyulmuştur.

Üretim alanında hala ulusal ve uluslararası alanda yeterli düzeyde çalışma mevcut değildir. Üretim alanında veri madenciliği uygulamalarının az olmasının sebebi:

- Veri madenciliğinin üretim sürecinde tek bir geçerliliği ve mutlak bir doğruluk payı yoktur. Çünkü işletmelerin üretim hedeflerine göre parametreler değişmektedir.

- Üretim sektöründe çalışanların dinamik, öğrenmeye açık kişiler olması ve yeni geliştirilen veri madenciliği algoritmalarını uygulayabilmesi gerekmektedir.
- Üretim süreci tek giriş ve tek çıkış kaynaklarından çok karmaşık üretim süreçleri içermektedir. Bu durum üretimde veri madenciliği uygulamalarını geri plana itmektedir.
- Akademik anlamda üretimde veri madenciliği süreci uygulanırken akademik çalışmalar yapan kişilerin üretim sektöründe verilere ulaşması süreci zorlaştırmaktadır.

İlerleyen çalışmalarda üretim sektöründen, kapsamlı ve dinamik veri akışları alınarak, ürün hata gruplaması yapılacaktır. Hata gruplaması yapılırken, Python programlama dili ve Google colab uygulaması ile büyük boyutlu verilerle çalışılacaktır. Google colab uygulaması, gmail uzantılı hesapla kullanıcılara excel dosyalarında açamadığımız büyük boyutlu verilerle çalışma yapma imkanı tanımaktadır. Bu çalışmalar için ram desteği ve 8 saat boyunca çalışmanın depolanabileceği hafıza imkanı sağlamaktadır. Çalışmalarda AdaBoost, XGBoost gibi yeni algoritma teknikleri kullanılarak sonuçların iyileştirilmesi ve literatüre yenilik katılması hedeflenmektedir.

REFERANSLAR

- Bernstein, J. H. (2011). The Data-Information-Knowledge-Wisdom-Hierarchy and its Antithesis. *NASKO*, 68-75.
- Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., & Zanasi, A. (1998). *Discovering Data Mining From Concept to Implementation*. New Jersey: Prentice Hall.
- Frawley, J. W., Piatetsky-shapir, G., & Matheus, C. (1992). Knowledge Discovery in Databases: An Overview. *AI Magazine*, 57-70.
- Gürsakar, N. (2001). *Sosyal Bilimlerde Araştırma Yöntemleri*. Bursa: VİPAŞ.
- Han, J. & Kamber, M. (2006). *Data Mining: Concepts and Techniques*. USA: Morgan Kaufmann Publishers, Elsevier.
- Jacobs, P. (1999). Data Mining: What General Managers Need to Know. *Harvard Management Update*, 8-10.
- Kittler, R. & Wang, W. (1999). The Emerging Role for Data Mining. *Solid State Technology*, 45-58.
- Özekes, S., & Çamurcu, Y. (2003). Veri Madencilğinde Karar Ağaçları Yöntemi Uygulaması. *Bilgi Teknolojileri Kongresi II. Denizli: Pamukkale Üniversitesi*.

Özkan, Y. (2016). *Veri Madenciliği Yöntemleri*. İstanbul: Papatya Bilim.

Sağın, A. (2018). *Veri Madenciliği Algoritmaları ile Birliktelik Kurallarının Belirlenmesi: Perakende Sektöründe Bir Uygulama*. İstanbul, İstanbul Ticaret Üniversitesi.

Silahtaroglu, G. (2016). *Veri Madenciliği: Kavram ve Algoritmaları*. İstanbul: Papatya Bilim.

Topoyan, M. (2019, 08 03). <http://kisi.deu.edu.tr/mert.topoyan/dosyalar/bup.pdf> adresinden alındı