



ÇELİK MALZEME SATIN ALMA SÜREÇLERİNDE VERİ MADENCİLİĞİ VE MAKİNE ÖĞRENMESİ UYGULAMALARI

Seray MİRASÇI^{1*}, Aslı AKSOY²

¹ LUT University, School of Business and Management, Lappeenranta, Finland

² Bursa Uludağ Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği, Bursa, Türkiye

Anahtar Kelimeler

*Veri Madenciliği,
Makine Öğrenmesi,
Stratejik Satın Alma,
Kümeleme Analizi.*

Öz

Firmaların varlıklarını sürdürebilmeleri için, belli karlılık hedeflerini tutturmaları gerekmektedir. Firmalarda karlılık hedeflerine doğrudan etki eden faaliyetlerden biri de satın almadır. Değişen dünya koşullarında satın alma süreçlerinin kritik malzeme grupları için çevik ve stratejik olması gerekmektedir. Çelik malzeme ürün grubunda stratejik satın alma kararlarının verilmesi ve karlılığın artırılması için veri madenciliği ve makine öğrenmesi yöntemleri ortaya konmuştur. Veri seti içerisinde tüketim, genişlik, uzunluk, kesim tipi, kalınlık, kalite ve ürün grubu gibi öznelilikler bulunmaktadır. Veri setinde bulunan gürültülü veriler tespit edilerek veri madenciliği teknikleri ile temizlenmiştir. Temizlenen veri seti makine öğrenmesi tekniklerinden kümeleme analizlerinden hiyerarşik kümeleme ve K-ortalama yöntemleri kullanılarak analiz edilmiştir. Bu analizde ideal küme sayısı ve bulunan ideal küme sayısının doğrulaması yapılmış olup stratejik açıdan en önemli proje ortaya konmuştur. Seçilen projede yer alan malzeme detayları teknik olarak incelenip, tüketim, kalınlık ve çelik malzemenin haddeleme tipi dikkate alınarak karlılık getirmesi beklenen satın alma stratejileri ortaya konmuştur. Bu çalışma, çelik malzeme ürün grubunda odaklanılan konu ve sektöre özgü bir perspektif sunmaktadır. Veri madenciliği tekniklerinin kullanıldığı ve makine öğrenmesi teknikleriyle veri setindeki ilişkilerin belirlendiği bu çalışma, satın alma kararlarının stratejik boyutunu vurgulayan özgün bir niteliğe sahiptir. Bu çalışmada önerilen analizler ile satın alma süreçlerinde, çalışan kaynaklı hataların satın alma stratejileri geliştirme süreçlerindeki etkileri azaltılmış, satın alma çalışanlarının uzun zaman harcayarak yapacağı analizler, veri analizi ve makine öğrenmesi gibi yöntemler ile gerçekleştirilmiştir.

DATA MINING AND MACHINE LEARNING APPLICATIONS IN STEEL MATERIALS PURCHASING

Keywords

*Data Mining,
Machine Learning,
Strategic Purchasing,
Clustering Analyses.*

Abstract

Companies have to meet certain profitability targets to ensure their survival. Purchasing is one of the key activities that directly affect company profitability. Purchasing processes for critical material groups need to be agile and strategic in a changing world. This study presents data mining and machine learning methods for making strategic purchasing decisions and increasing profitability in the steel materials product group. The dataset used in the study includes steel materials features such as consumption, width, length, cut type, thickness, quality and product group. Data mining techniques were used to identify and clean up noisy data in the dataset. Hierarchical clustering and K-means clustering methods were used to analyze the cleaned data set. The cleaned dataset was analyzed using hierarchical clustering and k-means clustering methods and the ideal number of clusters was determined and validated to identify the most strategically important project. The material details of the selected project have been technically analyzed and the purchasing strategies that are expected to bring profitability have been proposed, considering the consumption, thickness and rolling type of the steel material. This study provides an industry-specific perspective on the focused topic and sector in the steel materials product group. By using data mining and machine learning techniques to identify relationships in the steel materials product dataset, the study

provides a different perspective that emphasizes the strategic dimension of purchasing decisions. The analyses proposed in this study have reduced the impact of human error in purchasing processes on the development of purchasing strategies, and time-consuming analysis that would normally be carried out by purchasing staff has been replaced by methods such as data mining and machine learning.

Alıntı / Cite

Mirasçı, S., Aksoy, A. (2023). Çelik Malzeme Satın Alma Süreçlerinde Veri Madenciliği ve Makine Öğrenmesi Uygulamaları, Mühendislik Bilimleri ve Tasarım Dergisi, 11(3), 1174-1189.

Yazar Kimliği / Author ID (ORCID Number)

S. Mirasçı, 0000-0003-4654-6474
A. Aksoy, 0000-0002-2971-2701

Makale Süreci / Article Process

| | |
|---|------------|
| Başvuru Tarihi / Submission Date | 20.12.2022 |
| Revizyon Tarihi / Revision Date | 05.07.2023 |
| Kabul Tarihi / Accepted Date | 03.08.2023 |
| Yayın Tarihi / Published Date | 28.09.2023 |

*İlgili Yazar / Corresponding author: seraymir@hotmail.com, +905343640918

DATA MINING AND MACHINE LEARNING APPLICATIONS IN STEEL MATERIALS PURCHASING

Seray MİRASÇI^{1*}, Aslı AKSOY²

¹LUT University, School of Business and Management, Lappeenranta, Finland

²Bursa Uludağ Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği, Bursa, Türkiye

Highlights

- This study used data mining and machine learning methods to make strategic purchasing decisions and increase profitability in the steel materials product group.
- The study analyzed a dataset of steel materials characteristics using hierarchical clustering and k-means clustering methods to identify the most strategically important project.
- A technical analysis of the material details of the selected project was carried out in order to propose purchasing strategies that were likely to provide profitability.
- This study reduces the impact of human error in the purchasing process by using data mining and machine learning techniques, and replaces time-consuming analysis typically performed by the buyer with methods such as data mining and machine learning, providing a strategic perspective on purchasing decisions in the steel materials industry.

The paper aims to use data mining and machine learning techniques to make strategic purchasing decisions and increase profitability in the steel materials product group, by analyzing a dataset, applying clustering methods, performing technical analysis, and proposing profitable purchasing strategies.

Design/methodology/approach

Various methods are used to achieve the research objectives. First, data mining techniques are used to cleanse and analyse the purchasing data. Then, the steel material properties dataset was analysed using hierarchical clustering and k-means clustering methods to facilitate data analysis and identify the most strategically important project. Technical analysis of the material details of the selected project was also carried out in order to propose purchasing strategies that are likely to be profitable, taking into account factors such as the consumption of steel material, its thickness and the type of rolling. This study uses data mining and machine learning techniques to increase the efficiency and accuracy of the purchasing process, demonstrating that human error in purchasing process analysis can be reduced and that time-consuming analysis, typically performed by the buyer, can be performed quickly and effectively.

Findings

The results of the study show that the use of data mining and machine learning techniques in purchasing processes can lead to the identification of strategically important projects, improved data analysis and the development of profitable purchasing strategies. The methods used in the study can reduce human error in procurement processes, improve decision making and contribute to cost optimisation and supplier management. The study highlights the practical value of data-driven approaches into strategic purchasing decisions and emphasizes their potential to improve overall business performance.

Research limitations/implications

The limitations of the research include its reliance on a particular set of data and the potential for bias in the collection and analysis of the data. Future studies should aim to address these limitations and extend the scope of analysis to gain a more comprehensive understanding of strategic purchasing decisions.

* İlgili Yazar / Corresponding author: seraymir@hotmail.com, +905343640918

Practical implications

The research findings highlight the potential for data mining and machine learning techniques to improve purchasing practices and decision making. Practical implications include the need to incorporate these techniques into the procurement workflow to improve strategic sourcing and increase profitability. This research highlights the importance of adopting data-driven approaches to purchasing and suggests that companies should invest in the necessary tools and expertise to use these techniques effectively.

Social Implications

This research can have a positive impact on society by improving purchasing efficiency and promoting data-driven decision making. It has the potential to influence public attitudes towards the value of advanced techniques in procurement and encourage sustainable sourcing practices.

Originality

This study brings a new perspective by harnessing the power of data mining and machine learning in the area of strategic purchasing and profitability for steel materials. It offers valuable insights and innovative strategies for purchasing professionals, while also captivating the interest of researchers and practitioners in the field. With its proposed approach, the paper opens up new possibilities and potential for smarter strategy development in the ever-evolving landscape of purchasing and supply chain management.

1. Giriş (Introduction)

İşletmelerin hayatta kalabilmesi için kritik öneme sahip faaliyetlerden bir tanesi de satın alma faaliyetleridir. Satın alma faaliyetleri, tedarikçi ve müşteri arasında bir köprü görevi görmekle birlikte, rekabet, maliyet, kalite, zaman planlaması gibi konuları da kapsamaktadır. Değişen dünya koşullarında, satın alma fonksiyonlarının çevik ve doğru kararlarla yönetilmesi gerekmektedir. Aksi halde istenmeyen maliyetlere, üretim fonksiyonlarında verimlilik kayıplarına veya üretim duruşlarına yol açabilir. Bir işletmede, satın alma yapılan ürün çeşitleri arasında, kritik öneme sahip birçok malzeme vardır. Önemli grupta yer alan malzemelerden biri de çelik malzemelerdir. Çelik malzemeler, otomotiv, beyaz eşya, havacılık, madencilik gibi birçok sektörde farklı amaçlarla kullanılabilir. Çelik malzemeler, büyük bobinler halinde üretilip sonrasında istenilen ebatlarda kesilerek ilgili süreçlerden geçtikten sonra talep edilen formlara dönüştürülmektedir. Ancak karmaşık ve çok operasyonlu bir süreç olması sebebiyle doğru yönetilmesi gerekmektedir.

Bu çalışmanın amacı, çelik malzeme satın alma süreçlerinde karlılık sağlayacak satın alma stratejilerini hızlı ve doğru şekilde ortaya koymaktır. Araştırmanın odaklandığı temel konular aşağıda belirtilmiştir:

- Satın alma verilerinin veri madenciliği teknikleri kullanılarak temizlenmesi ve analiz edilmesi, stratejik öneme sahip projelerin belirlenmesine nasıl katkı sağlar?
- Çelik malzeme satın alımında makine öğrenmesi yöntemleriyle stratejik öneme sahip ürün kümeleri nasıl belirlenebilir?
- Veri madenciliği ve makine öğrenmesi teknikleri kullanılarak oluşturulan stratejik kümeler, çelik malzeme satın alma stratejileri için nasıl değerlendirilebilir?
- Çelik malzeme satın alma süreçlerinde veri madenciliği ve makine öğrenmesi yöntemlerinin kullanılmasıyla elde edilen sonuçlar nelerdir ve nasıl stratejiler geliştirilebilir?

Bu kapsamda, otomotiv sektöründe yer alan bir firmaya ait çelik malzeme grubu için, doğru satın alma stratejilerini oluşturabilmek amacıyla bir uygulama yapılmıştır. Bu uygulama aşağıda belirtilen üç temel aşamadan oluşmaktadır.

- 1.Adım: Satın alma verilerinin incelenerek, gürültülü verilerin veri madenciliği teknikleri ile temizlenmesi
- 2.Adım: Makine öğrenmesi yöntemi ile stratejik öneme sahip ürün kümesinin belirlenmesi
- 3.Adım: Belirlenen küme için gerekli analizler yapılarak satın alma stratejilerinin oluşturulmasıdır.

Literatürde yapılan çalışmalar incelendiğinde seçilen konularda birtakım eksiklikler saptanmıştır. Satın alma için, literatürde müşterilerin satın alma davranışları ile ilgili pek çok çalışma mevcuttur. Örneğin, Köylüoğlu vd. (2018) yaptığı çalışmada akademisyenlerin araç satın alma davranışları üzerinde çalışmıştır. Bu çalışmada SPSS programını kullanarak, regresyon ve ANOVA gibi istatistiksel analiz ve testler kullanılmıştır. Kirgiz (2021) ise müşterilerin lüks otomobil satın alma kararlarını hangi faktörlerin etkilediği üzerinde korelasyon yöntemiyle çalışmıştır. Literatürde yer alan çalışmalarda müşterilerin satın alma davranışları, satın alma sürecini etkileyen karar kriterleri gibi konuların sıklıkla çalışıldığı fakat çalışma sonucunda belirlenen kriterlerin veya davranışların satın alma stratejilerine etkilerinin araştırılması konusunda yeterli araştırma olmadığı düşünülmektedir. Bu nedenle, bu çalışmada satın alma strateji geliştirme konusu da incelenmiştir.

Bu çalışmada ayrıca, satın alınan malzemelerde makine öğrenmesi teknikleriyle stratejik öneme sahip kümeler belirlenerek farklı satın alma stratejileri geliştirilmesi için de öneriler sunulmuş, literatürde yapılmış kümeleme çalışmalarından farklı olarak, oluşturulan kümeler analiz edilerek, satın alma stratejisi oluşturma çalışmaları için girdi olarak kullanılmıştır. Bu çalışma ile hedeflenen, günlük hayatta da sık karşılaşılan satın alma problemlerine, veri madenciliği, makine öğrenmesi yöntemleri ile çözüm oluşturmak ve karlı satın alma stratejileri oluşturmaktır. Çalışmanın ikinci bölümünde kaynak araştırması yapılmış olup, veri madenciliği ve makine öğrenmesi çalışmaları ile ilgili yapılan literatür araştırması sunulmuştur. Üçüncü bölümde önerilen yöntemler açıklanmış, dördüncü bölümde veri madenciliği ve makine öğrenmesi yöntemleri kullanılarak yapılan uygulama sonuçları yer almaktadır. Beşinci bölümde ise sonuç ve çıktılar ortaya konarak belirlenen stratejiler sunulmuştur.

2. Kaynak Araştırması (Literature Research)

Bu bölümde, çalışma kapsamında kullanılan veri madenciliği ve makine öğrenmesi konuları ile ilgili yapılmış çalışmalar açıklanmıştır.

2.1. Veri Madenciliği (Data Mining)

Veri madenciliği, tanınmayan, saklı veya faydalı verilerin veri tabanlarından belli kurallar çerçevesinde otomatik bir şekilde elde edilerek bilginin işlevsel hale getirildiği analiz sürecidir (Erbudak, 2022). Kameshwaran vd. (2014) veri madenciliğini, bir veri setinden, ihtiyaç duyulan bilginin doğru şekilde çıkarılarak, anlaşılır şekle dönüştürülme süreci olarak tanımlamaktadır.

Erpolat (2012) yaptığı çalışmada veri madenciliği yöntemlerinin kuvvetli ve zayıf yönlerini aşağıdaki gibi sunmaktadır:

Kuvvetli Yönler:

- Çok boyutlu ele alınabilecek sorunların çözümü için uygun bir araçtır.
- Karmaşık durumların çözümlenmesinde kullanılmak için uygundur
- Hem sayısal hem de kategorik değişkenlerin çözümünde kullanılabilir.

Zayıf Yönler:

- 0 ve 1 arasında giriş verisi olması gerekir. Yani veri setinin nümerik bir matris haline dönüştürülmesi gerekmektedir.
- Ürettikleri sonuçların açıklamasını sunamamaktadır.
- Sonuçlar farklı teknikler ile tekrarlanarak daha iyi sonuçlar ortaya konabilir.

Savaş vd. (2012) yaptıkları çalışmada, veri madenciliği çalışmalarında veri türlerine göre karşılaşılan problemleri aşağıdaki gibi belirtmektedir:

- Artık veri: Veri setindeki gereksiz ve ihtiyaç duyulmayan verilerdir.
- Belirsizlik: Hataların şiddetini belirtir.
- Boş değerler: Herhangi bir değere eşit olmayan ve niteliği olmayan verilerdir.
- Dinamik veri: Sürekli değişen verilerdir. Genellikle kurumsal hayattaki veriler bu gruba girmektedir.
- Eksik veri: Veri setinin içeriği gereğince veya hata ile doğan verilerdir. İstatistiksel analizlerde problem yaratma ihtimali yüksektir.
- Farklı tipteki veriler: Sadece sembolik veya kategorik veriler değil, tam sayılı, kesirli, coğrafi konum bilgileri gibi farklı tipte verilerin mevcut olmasıdır.
- Gürültülü değerler: Verinin toplanması veya aktarılması aşamasında doğan, istenmeyen ve analizleri yanlış etkileme riski bulunduran verilere gürültülü değerler denilmektedir.

Savaş vd. (2012) Türkiye’de veri madenciliği uygulamalarına 2000’li yıllardan sonra ilginin arttığını belirtmiş, mühendislik, bankacılık, eğitim, ticaret, sağlık gibi farklı sektörlerde veri madenciliği uygulamalarından faydalandığını ortaya koymuştur. Veri madenciliği çalışmalarında, önemli noktalardan biri veri setinin dağılımı hakkında bilgi sahibi olmaktır. Bir verinin normal dağılıma sahip olup olmadığını anlamak için geliştirilen birçok test vardır. Veri setinin normallik testinde çıkan sonuca göre doğru makine öğrenmesi yönteminin seçilmesi önem arz etmektedir. Jeong vd. (2016) yaptıkları çalışmada, verilerin normallik testini yaparken faydalanılabilecek birçok testi ortaya koymuştur. Bunlar, Kolmogorov-Smirnov testi, Lilliefors testi, Cramer-von Mises testi, Shapiro-Francia testi, Jarque-Bera testi, D’Agostinos testi, Doornik-Hansen testi, Enerji testi, Shapiro Wilk ve bu çalışmada kullanılan Anderson Darling normallik testidir.

Babaoğlu (2015) yaptığı çalışmada, veri madenciliği uygulama aşamalarını aşağıda belirtilen yedi ana adımda incelemiştir:

1. Problemin ve çalışma sonucunda ulaşılmak istenen bilginin tanımlanması ve veri setinin anlaşılması,
2. Veri setinin temizlenerek, istenmeyen, amaca hizmet etmeyen, tekrarlı veya uygunsuz verilerin elimine edilmesi,
3. Verilerin düzenli ve kontrollü bir şekilde bir araya getirilmesi,
4. İhtiyaç duyulan verilerin belirlenmesi ve seçilmesi,
5. Uygun veri madenciliği tekniği ile verilerin analizi,
6. Verilerin değerlendirilmesi,
7. İhtiyaç duyulan bilgilerin, verilerden sentezlenmesi.

Tiwari vd. (2018), çalışmasında 2010-2016 yılları arasında tedarik zinciri çalışmalarında veri madenciliği çalışmalarının önemine ve firmalar için rehber olarak kullanılabilirliğine vurgu yapmıştır. Veri madenciliği ile farklı sektörler için karmaşık veri setlerinin analizi, yorumlanması ve bilgilerin anlamlı hale getirilmesi sürecine değinmiştir. Safaei vd., (2018) çalışmasında tedarik zinciri ve satın alma maliyetlerinin azaltılması ve pazar payının artması için veri madenciliği tekniklerinin önemini vurgulamıştır. Sağlık satın almaları için yapılan bu çalışmada satın alma maliyetlerinin azaltılarak tedarikçi ilişkilerinin iyileştirildiği ortaya konmuştur. Qing vd., (2021) çalışmasında veri madenciliği yöntemlerinin kullanılmasının satın alma davranışları üzerinde belirleyici olduğunu ortaya koymuştur.

Ricci vd. (2011) veri madenciliğini üç ayrı adımda incelemiştir. Bu adımlar verinin hazırlanması, analizi ve yorumlanmasıdır. Verinin hazırlanması adımı, veriler arasındaki uzaklıkların ölçülmesi, örneklemin seçilmesi ve uzaklıkların azaltılması adımlarını kapsamaktadır. Analiz adımı, tahmin ve tanımlama olarak ikiye ayrılmaktadır. Tahmin için sınıflandırma yöntemleri kullanılmaktadır. Bu yöntemler k-en yakın komşu, karar ağaçları, Bayes ağları ve destek vektör makineleri olarak sayılabilir. Tanımlama için ise birliktelik analizleri kullanılmaktadır. Yorumlama adımı ise, veri madenciliğinden elde edilen çıktıların bilgiye dönüştüğü yorum aşamasını içermektedir. Veri madenciliği, çeşitli veri setlerinden faydalı bilgilerin elde edildiği bir analiz sürecidir. Veri madenciliği yöntemleri, karmaşık sorunların çözümünde etkili olmakta, literatürde farklı sektörlerde yapılan uygulamalar ile ilgili çalışmalar bulunmaktadır.

2.2. Makine Öğrenmesi (Machine Learning)

Makine öğrenmesi kısaca, bir bilgisayarın deneyimlere dayalı öğrenmesi olarak tanımlanabilir (Al-Omary vd., 2005). Russom (2011), veri yönetimini, farklı istatistiksel ve makine öğrenmesi yöntemlerinin veri setleri üzerinde uygulanması olarak yeni bir iş zekâsı uygulaması olarak tanımlar.



Şekil 1. Sık Tercih Edilen Makine Öğrenmesi Algoritmaları (Widely Used Machine Learning Algorithms)

Şekil 1.'de görüleceği üzere makine öğrenmesi kendi içinde, gözetimli ve gözetimsiz öğrenme olmak üzere ikiye ayrılmaktadır (Özgür, 2002).

Gözetimli öğrenme, bağımlı ve bağımsız değişken arasındaki ilişkiyi ortaya koymayı amaçlamaktadır. Bu yöntem için, bağımlı değişkenlerin değerlerinin biliniyor olması önem arz etmektedir. Gözetimli öğrenme kendi içinde lineer modeller ve ağaç tabanlı yöntemler olarak ikiye ayrılmaktadır. Lineer modellerin alt grubunda lineer

regresyon, ağaç tabanlı yöntemlerin altında ise karar ağaçları, rastgele orman yöntemi gibi örnekler yer almaktadır (Özgür, 2002).

Gözetimsiz öğrenmede tüm değişkenler ayrı ayrı ele alınmakta ve bağımlılık ilişkisi aranmamaktadır. Ancak gözetimli öğrenmeden farklı olarak, gözetimsiz öğrenmede ulaşılması gereken hedefler vardır. Bunlar, veri sayısının azaltılması, verilerin kümeleneceği gibi amaçlar olabilir. Değişkenlik içeren durumlarda, eğer yeterli sayıda veri varsa ve değişkenler biliniyorsa gözetimsiz öğrenme tercih edilebilir (Özgür, 2002). Gözetimsiz öğrenme, kümeleme analizleri ve birliktelik analizlerini içermektedir. Kümeleme analizi altında, K-ortalama ve hiyerarşik kümeleme sık kullanılan makine öğrenmesi algoritmalarına örnek verilebilirken, birliktelik analizleri içerisinde apriori algoritmaları örnek verilebilir.

Liu vd. (2000), gözetimli öğrenme tekniklerinden biri olan karar ağaçlarını, normal dağılıma sahip olan ve olmayan iki veri seti üzerinde uygulamış ve normal dağılıma sahip veri setindeki performans ve hassasiyetin daha iyi düzeyde olduğunu ortaya koymuştur. Erpolat (2012) çalışmasında, bir otomobil yetkili servisinde müşterilerin alışveriş verileri arasındaki bağlantıları apriori algoritması ile çalışmış ve hangi ürünlerin birlikte alındığını ortaya koymuştur. Saputra vd. (2019) çalışmasında kümeleme yöntemleri başlığı altında, K-ortalama yöntemini kullanarak ideal küme sayısını belirlemiş ve değerlendirmiştir. Bu çalışmada ideal küme sayısını belirlemek için dirsek yöntemi, Silhouette yöntemi ve Gap istatistik yönteminden faydalanmıştır. Çalışma sonucunda bu üç yöntemin aynı küme sayısını bulmasının bir zorunluluk olmadığı ve her yöntemin ayrı bir hesaplama mantığı ile küme sayısını belirlediğini belirtmiştir. Patel vd. (2022) çalışmasında, bir veri setinin optimal küme sayısının bulunması için K-ortalama yöntemini ve hiyerarşik kümeleme yöntemlerini karşılaştırmıştır. K-ortalama yöntemini için, dirsek, Silhouette ve Gap istatistik yöntemlerini kullanarak küme sayılarını belirlemiştir. Küme sayılarının yöntemlere göre değişebileceğinin önemini belirtmiştir. Küme sayısının geçerliliğini test edebilmek için ayrıca bağlantısallık, Dunn indeksi ve Silhouette indeksi yöntemleri ile test yapmıştır. Bu çalışma sonucunda, ilgili veri seti için en uygun kümeleme yöntemi, kümelemenin geçerliliğini test eden yöntem ve küme sayısını ortaya koymuştur. Zhao vd., (2017) çalışmasında paketleme birimine ait stoklama, satın alma, malzeme ve nakliye gibi maliyetlerini kümeleme yöntemleri ile azaltmıştır. Bu çalışmada k-ortalama, hiyerarşik kümeleme yöntemleri kullanılmıştır. Farklı makine öğrenmesi yöntemlerinin satın alma süreçlerinde uygulanması ile ilgili literatürde yapılan pek çok çalışma bulunmasına rağmen, makine öğrenmesi yöntemleri ile elde edilen sonuçların satın alma stratejisi geliştirilmesi ile entegre edildiği yeterli sayıda çalışma bulunmamaktadır.

3. Materyal ve Yöntem (Material and Method)

Bu çalışmada, bir otomotiv işletmesine ait 732 farklı çelik malzeme içeren veri setinde, doğru satın alma stratejilerini oluşturmak hedeflenmiş olup çalışmanın temel adımları aşağıda belirtilmiştir:

- Veri setinde gürültülü verilerin tanımlanması
- Veri madenciliği ile kümeleme analizi
- İdeal küme sayısının tespiti
- K-ortalama yöntemini ile maliyet tabanlı ürün grupları ve kümelerin oluşturulması
- Oluşturulan kümelerin değerlendirilmesi

Bu çalışmada R studio paket programı kullanılmış olup 1.4.1106 versiyonunda çalışmalar yapılmıştır.

Gerçek hayat verileri hiçbir zaman mükemmel değildir, bu sebeple analizler yapılırken gürültülü verilerin de dahil edilmesi; verilerin yorumlanması, oluşturulan modelleri ve alınan kararları olumsuz yönde etkileyebilirler. Gürültülü veriler, sınıflandırmayı da negatif yönde etkiler. Bu negatif etkiler, sınıfların sınırlarının bozulması, anlamsız küçük kümelerin oluşması veya önemli kümelerin örneklemelerinin değişmesi olarak sayılabilir. Gürültülü verileri kolayca temizlemek için birçok yöntem mevcuttur. Bunlardan bazıları aşağıdaki gibi sıralanabilir (Liebchen vd. 2007);

- Güçlü Öğrenme: Budama yöntemi ile karar ağaçlarındaki gürültüye sebep olan olasılıkları azaltmaya çalışır.
- Veri Cilalama: Amacı mevcut veri setinde gürültülü verileri azaltmaya çalışmaktır, ancak küçük veri setleri üzerinde çalışılması tavsiye edilir, zaman alan bir yöntemdir.
- Gürültü Filtreleri: Veri setinde gürültüye yol açan senaryolar biliniyorsa, bu senaryoların veri setinden filtrelenerek ayrılan yöntemdir.

Çalışılan satın alma veri setinde, strateji oluşturmak için çalışırken karşılaşılabilecek bazı problemler aşağıda sunulmaktadır. Burada bulunan problemler, istenmeyen sonuçlara ve kararlara yol açabilir:

- Çelik satın alma veri setinin gürültülü veriler içeriyor olması,
- Gürültülü verilerin, veri setinin yanlış analiz edilmesine sebep olma riski,
- Uygun makine öğrenmesi algoritmasının seçilememesi,
- Verilerin yanlış sınıflandırılması,
- Veri setine ait ideal küme sayısının bilinmemesi,
- Projenin stratejik öneminin sadece maliyet kriterine göre belirlenmesi,
- Çelik üretim süreçleri ve malzeme bilgisi alanında yetersiz bilgi,
- Yanlış zamanda, yanlış satın alma stratejisi oluşturulması.

Satın alma veri setinde görülecek gürültülü verilerin birçok kaynağı olabilir. Örneğin; satın alma çalışanın fiyat girişi esnasında yapacağı hatalar, dosyaların güncel olmaması, yanlış hesaplamalar vb. Bu çalışmada incelenen veri setinde gürültülü olarak tanımlanan durumlar aşağıda belirtilmiştir:

- Referansların projesinin belli olmaması: Bir çelik malzeme referansının hangi projede kullanılacağı veri setinde mutlaka bulunmalıdır. Bulunmaması halinde, proje maliyetleri belirlenirken hatalar ortaya çıkacaktır.
- Referansların seri üretimde olmaması: Seri üretim koşulları için yapılmak istenen bir analizde, seri ömrünü tamamlamış veya denemelerde kullanılan referansların yer almaması gerekir.
- Bir fiyat listesinde büyük sapmalar gösteren fiyatlar (beklenenden farklı seviyelerde fiyatlar) analiz edilmelidir. Bu fiyatlar özel bir malzemeye ait olabileceği gibi hatalı giriş sonucu ortaya çıkmış olabilir. Örneğin çalışmaya konu veri setinde satın alma çalışanın yanlışlıkla negatif değer girdiği fiyatlar mevcuttur.

Veri setinin normal dağılıp, dağılmadığını anlamak için yapılan Anderson-Darling testinin formülü Eşitlik 1'de verilmiştir:

$$AD = -n - \frac{1}{n} \sum (2i - 1) [\ln F(x_i) + \ln(1 - F(x_{n-i+1}))] \quad (1)$$

x_i : Veriler artan düzende sıralandığında seçilen i . sıradaki örnek
 $F(x_i)$: Belirtilen dağılım için kümülatif dağılım fonksiyonu (CDF)
 n : örneklem büyüklüğü
 AD : Anderson Darling

Normallik testi için oluşturulan hipotezler aşağıdaki gibidir;

H_0 : Veri seti normal dağılıma sahiptir.
 H_1 : Veri seti normal dağılıma sahip değildir.

Normallik testinde %95 güven aralığı için; $p < 0.05$ ise H_0 reddedilir ve veri seti normal dağılıma sahip değildir denir, $p > 0.05$ ise H_1 reddedilir ve veri seti normal dağılıma sahiptir denir (Jeong vd., 2016).

Satın alma çalışmalarında, veri seti Anderson Darling testi sonucunda normal dağılıma sahip değil ise makine öğrenmesi adımı için uygulanabilecek iki ayrı yöntem söz konusudur. İlki, normal dağılıma sahip olmayan veriyi, logaritmik dönüşüm, karekök dönüşümü vb. veri dönüştürme teknikleri ile normal dağılıma çevirmek ve sonrasında gözetimli öğrenme tekniklerini kullanmak veya normal dağılım gerektirmeyen gözetimsiz öğrenme tekniklerini kullanmaktır. Öztuna vd. (2006) çalışmasında, Kolmogorov-Smirnov, Pearson, Jarqua-Bera ve Shapiro-Wilk testlerini dağılımlar için hatalar ve güçlük açısından incelemiş olup normal olmayan dağılımlar için Shapiro-Wilk, Anderson Darling ve Jarqua-Bera testinin en güçlü sonuçları verdiğini belirtmiştir. Bu sebepten bu çalışmada Anderson Darling testi kullanılmıştır.

Bu çalışmadaki veri setinin normal dağılıma sahip olmadığı belirlenmiş, gözetimsiz öğrenme teknikleri uygulanmıştır. Gözetimsiz öğrenme tekniklerinin altında, en çok tercih edilen uygulamalardan biri olan K-ortalamlar yöntemi seçilmiştir.

1967 yılında MacQueen tarafından önerilen K- ortalamlar yöntemi büyük ve karmaşık verilerden basit ve anlaşılır veri kümelerine ulaşmayı amaçlamaktadır (Patel vd. 2022). Tekin (2018) çalışmasında K-ortalamlar yönteminin adımlarını başlangıç, tekrarlama, iyileştirme ve durma olarak aşağıdaki gibi belirtmiştir;

- Başlangıç aşamasında, her veri kendine en yakın rastgele bir küme merkezine atanır.
- Tekrarlama aşamasında ilgili küme merkezi, verilerin ortalaması olacak şekilde değiştirilir.
- İlgili verinin küme merkezine uzaklığı seçilen uzaklık metriği ile hesaplanır ve hata değeri belirlenir.

- İyileştirme aşamasında her veri yeniden kendine en yakın küme merkezine atanır.
- Hiçbir veri kümeler arası hareket yapamayacak hale geldiğinde durma işlemi gerçekleşir.

Verilerin gerekli sayıda kümelere ayrılması ve birbiriyle uyum içinde olması için doğru sayıda kümeye ayrılması gerekmektedir. Ancak veri setine göre ideal küme sayısı değişmektedir. Dolayısıyla her çalışmada, veri setine özgü ideal küme sayısı belirlenmelidir.

Küme sayısını belirlemek için üç ayrı yöntem mevcuttur: dirsek yöntemi, Silhouette yöntemi ve Gap istatistik yöntemi. Ancak her yöntem farklı bir küme sayısına işaret edebilir (Patel vd., 2022).

Dirsek yöntemi, veri gruplarının birbirine olan uzaklıklarının kareler toplamını hesaplama temeline dayanan bir yöntemdir. Veri kümelerinin birbirine benzerliklerini ölçmeyi ve benzer verileri bir arada kümelemeyi hedeflemektedir. Dirsek yöntemi grafikte dirseğe benzer bir kırılmanın olduğu yerde ideal kümelemenin olduğunu belirtmektedir. İdeal küme noktasından sonra, grafikte düşüşler daha yavaş gözlemlenmektedir (Cui, 2020; Patel vd., 2022). Dirsek yönteminin formülü Eşitlik 2'de verilmiştir.

$$\text{Dirsek Yöntemi} = \sum_{1.\text{küme}} \text{uzaklık} (P_i, C_1)^2 + \sum_{2.\text{küme}} \text{uzaklık} (P_i, C_2)^2 + \dots + \sum_{n.\text{küme}} \text{uzaklık} (P_i, C_N)^N \quad (2)$$

P_i : Seçilen kümedeki i . nokta

C_N : N. kümenin merkezi

N : Küme sayısı

Silhouette yöntemi, her değer için kümelerle arasındaki uzaklığa bakmaktadır ve oluşan her değer için (-1,1) arasında değerler üretmektedir. +1'e en yakın değer en uygun olarak belirlenmektedir. Silhouette yöntemi Eşitlik 3'de gösterildiği gibi hesaplanmaktadır (Wang vd., 2019; Patel vd., 2022):

$$s(i) = \frac{b_i - a_i}{\max(b_i - a_i)} \quad (3)$$

a_i = i . verinin aynı kümedeki diğer verilere ortalama uzaklığı

b_i = i . verinin en yakın komşu kümelere olan uzaklığı

$s(i)$ = Silhouette değeri

Gap istatistik yöntemi, kümelerin, gerçekleşen değerlerinin karelerinin toplamının, beklenen değerlerine göre logaritmik değerce farkını karşılamaktadır. Bu iki değer arasındaki Gap büyüklüğü, rastgele uniform dağılımdan ne kadar uzakta olduğu belirtmektedir. Gap ne kadar büyükse, sonuç o kadar anlamlıdır. Başka bir deyişle, Gap büyüklükleri farklı kümelere işaret etmektedir (Patel vd., 2022)

Yukarıda açıklanan yöntemler ile doğru küme sayısı belirlendikten sonra belirlenen küme sayılarının kontrolü ve doğrulamasının yapılması gerekmektedir. Özetle, bulunan kümelerin ne kadar doğru sonuçlar verdiği araştırılmalıdır. Bu doğrulama sürecinde üç ayrı yöntem takip edilmektedir (Shobha vd., 2017).

Bağlantı indeksi, verilerdeki yerel yoğunlukları gözlemlemekte ve veri gruplarını en yakın komşulara göre kümelere ayırmaktadır. Bu indeks, bir verinin kendisine en yakın komşu kümeye ne düzeyde yerleştiğini göstermektedir. Bağlantı indeksinden çıkan sonuçlar sıfır ve sonsuz arasında olup, minimum değer elde edilmesi gerekmektedir. Burada N gözlem sayısı, m toplam veri sayısı, L veriye en yakın küme sayısı, c gözlemlere ait ayrık kümeler olarak alınırsa, bağlantı indeksi Eşitlik 4'de gösterildiği gibi hesaplanabilir (Shobha vd., 2017):

$$\text{Conn}(C) = \sum_{i=1}^N \sum_{j=1}^L x_{i,m_j} \quad (4)$$

Silhouette indeksi, kümelere nesnelere bulunduğu yerlere göre kararlılıklarını göstermektedir. Silhouette indeksi, 1'e yakın ise iyi kümelenebilir, eğer -1'e yakın ise zayıf kümelenebilir verilerdir. Başka bir deyişle, Silhouette indeksi (-1,1) aralığında olup maksimize edilmesi gereken bir değerdir. Silhouette indeksi Eşitlik 3'de gösterildiği şekilde ölçülmektedir (Shobha vd., 2017).

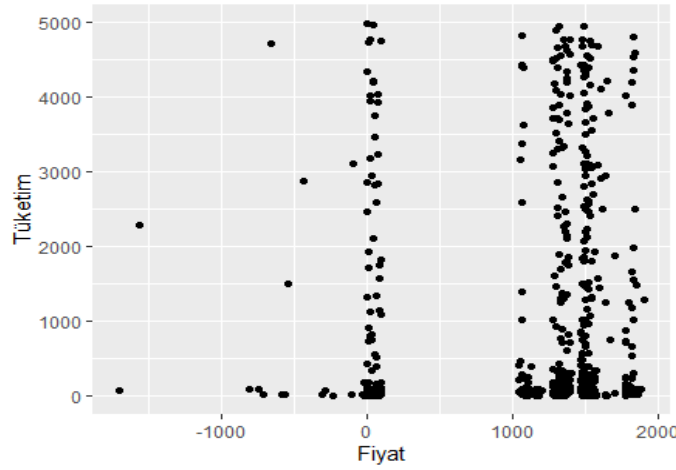
Dunn indeksi, aynı kümede olmayan en yakın iki verinin uzaklığının, aynı kümede olan birbirine en uzak iki veriye uzaklığına oranıdır. Dunn indeksi (0,1) aralığında olup, maksimize edilmesi gereken bir değerdir. Dunn indeksi Eşitlik 5'de gösterildiği gibi hesaplanabilir (Shobha vd., 2017):

$$D(c) = \frac{\min_{c_k, c_i \in c_k \neq c_i} (\min_{i \in c_k, j \in c_i, dist(i, j)})}{\max_{c_k} diam(c_k)} \quad (5)$$

c_i : i kümesinin benzerlik fonksiyonu
 c_k : k kümesinin benzerlik fonksiyonu
 $diam$: kümeler arası maksimum uzaklık

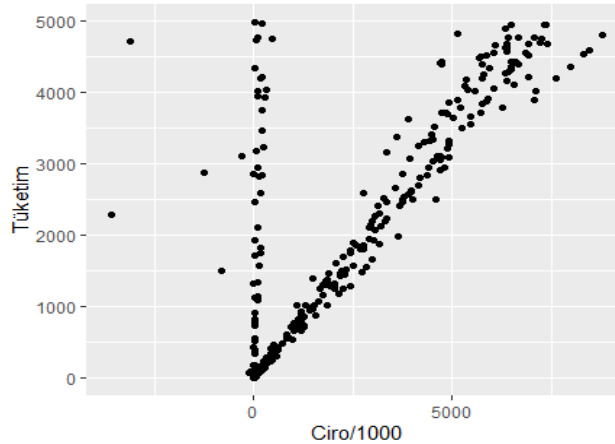
4. Deneysel Sonuçlar (Experimental Results)

Çalışmanın yapıldığı veri setindeki gürültülü verileri doğru yorumlamak ve temizleyebilmek için, ilk adımda veri setinin iyi tanınması ve analiz edilmesi gerekmektedir. Bu çalışmada R Studio "*library(dplyr)*", "*library(tidyverse)*", "*library(ggplot2)*", "*library(factoextra)*" kütüphaneleri kullanılmıştır.



Şekil 2. Gürültülü Veriler Barındıran Veri Setinin Fiyat-Tüketim Grafiği (Price- Consumption Graphic with Noisy Data)

İlk adımda, veri setinin anlaşılması için mevcut veri setinin grafiği oluşturulmuştur. Şekil 2’de verilerin fiyat tüketim grafiği yer almaktadır. Şekil 2’ye bakıldığında fiyatı sıfır ve negatifte bulunan birçok değer olduğu gözlemlenmektedir. Ancak bir satın alma operasyonunda, bir ürünün fiyatı sıfır veya negatif olmamalıdır. Eğer, veri seti bu şekilde gürültülü verileri barındırırken bir analiz yapılacak olursa, analiz sonuçlarının sağlıklı olmaması veya yanlış yönlendirmeye açık olması gibi sorunlarla karşılaşılabilir.



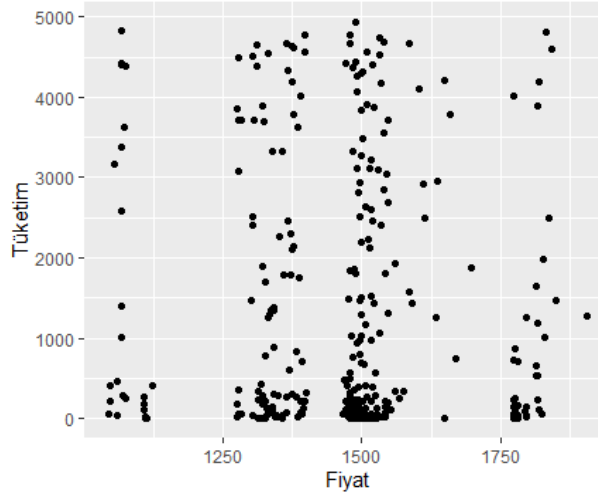
Şekil 3. Gürültülü Veriler Barındıran Veri Setinin Ciro-Tüketim Grafiği (Turnover - Consumption Graphic with Noisy Data)

Şekil 3’te ise verilerin ciro ve tüketim grafiği yer almaktadır. Şekil 2’de görülen sıfır ve negatif değer problemi Şekil 3’te yer alan grafikte de görülmekte, cirosu sıfır olan pek çok veri gözlenmektedir. Gürültülü verilerin temizlenmesi üç adımda yapılmıştır:

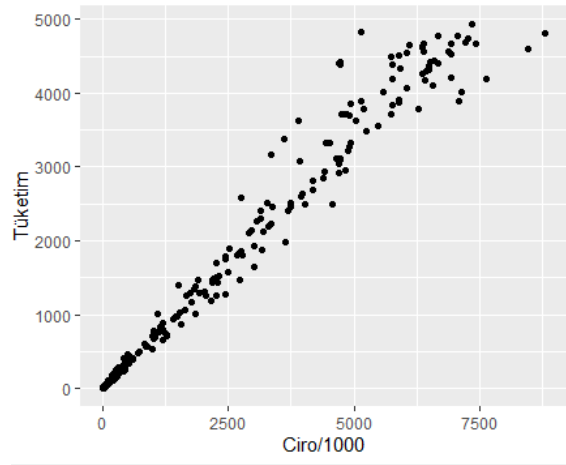
- Adım 1: Fiyatı negatif olmayan ve 500 €/ton’dan yüksek verilerin seçilmesi (500 €/ton temsili bir rakam olarak alınmıştır.)
- Adım 2: Sadece seri üretimde devam eden referansların seçilmesi

- Adım 3: Proje sütunu boş olan tüm satırların silinmesi

Bu adımlar sonucunda 732 satır veri, 414 satır veriye düşürülmüştür.



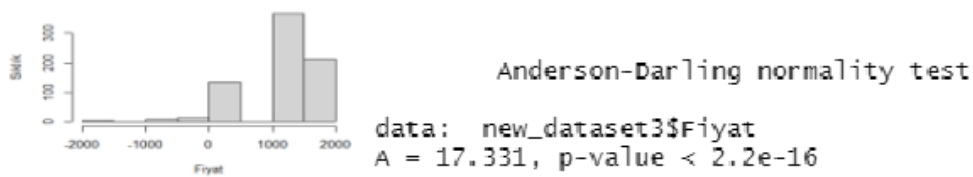
Şekil 4. Gürültülü Verilerin Temizlenmesinden Sonra Fiyat-Tüketim Grafiği (Price- Consumption Graphic without Noisy Data)



Şekil 5. Gürültülü Veriler Temizlenmesinden Sonra Ciro-Tüketim Grafiği (Turnover- Consumption Graphic without Noisy Data)

Şekil 4 ve Şekil 5'te görüldüğü üzere, gürültülü veri olarak adlandırılan fiyatı veya cirosu sıfır veya negatif olan veriler temizlenmiştir.

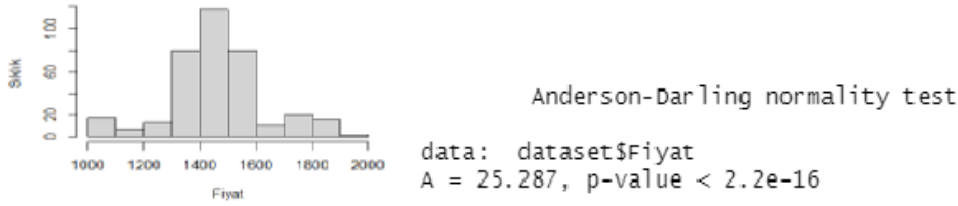
Veri temizliği tamamlandıktan sonra veri setinin dağılımının normal olup olmadığının tespit edilmesi gerekmektedir. Bunun için Anderson Darling testi yapılmıştır.



Şekil 6. Gürültülü Veriler Temizlenmeden Önceki Duruma Ait Histogram Grafiği ve Anderson Darling Testi Sonuçları (Histogram Plot and Anderson Darling Test Results with Noisy Data)

Şekil 6'da gürültülü veriler temizlenmeden önceki durumun histogram grafiği ve Anderson Darling testi sonuçları

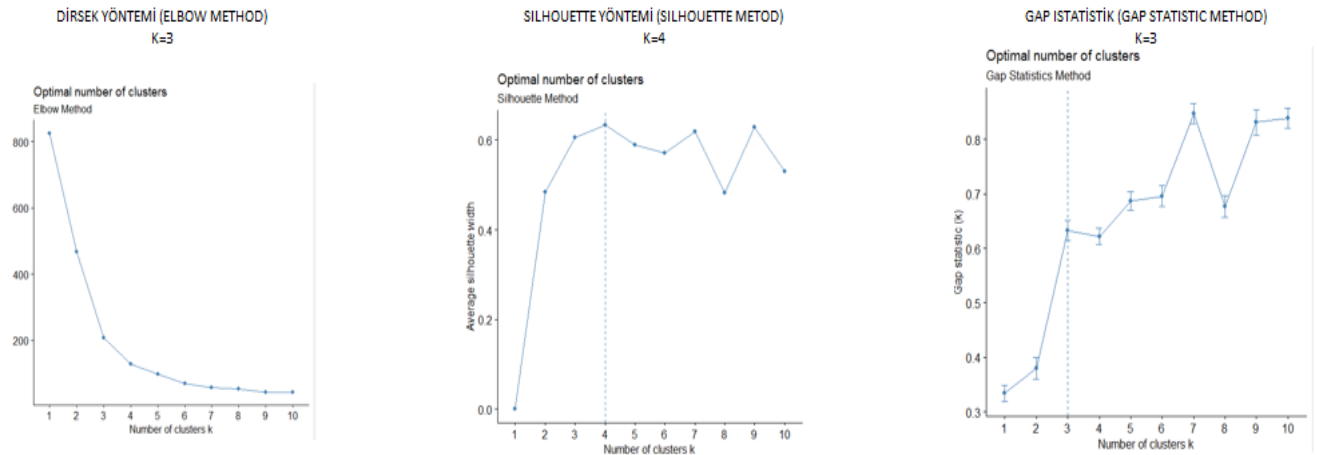
yer almaktadır. Şekil 6'da test sonuçlarına bakıldığında A değeri 17,331 bulunmuş olup $p < 0.05$ 'tir (p -değeri = $2.2e-16 \approx 2.2 \cdot 10^{-16}$). Dolayısıyla, H_0 hipotezi reddedilir. Anderson Darling testi, gürültülü veriler temizlenmeden önce verilerin normal dağılıma sahip olmadığını belirtmektedir.



Şekil 7. Gürültülü Veriler Temizlendikten Sonraki Duruma Ait Histogram Grafiği ve Anderson Darling Test Sonuçları (Histogram Graph and Anderson Darling Test Results without Noisy Data)

Gürültülü verileri temizledikten sonraki durum incelendiğinde, Şekil 7'de görüldüğü üzere, grafik normal dağılım grafiğine benzese de sadece grafik ile yorum yapılmamalıdır. Anderson Darling normalite testi, gürültülü veriler temizlendikten sonra A değerinde 17,331'den 25,287'ye yükselerek iyileşme sergilemiştir. Ancak $p < 0.05$ (p -değeri = $2.2e-16 \approx 2.2 \cdot 10^{-16}$) olduğu için H_0 reddedilir, veri seti normal dağılıma sahip değildir.

Veriler temizlendikten sonra, bir sonraki adımda stratejik öneme sahip kümelerin tespit edilmesi gerekmektedir. Ancak veri normal dağılmadığı için regresyon analizi gibi yöntemler uygulanmamış makine öğrenmesi yöntemleri tercih edilmiştir. Optimal küme sayısının bulunması için üç farklı yöntem uygulanmıştır. Bunlar dirsek yöntemi, Sillhoutte yöntemi ve Gap istatistik yöntemidir.



Şekil 8. Kümeleme Uygulamaları (Clustering Applications)

Şekil 8'de yer alan analizlere göre, veri setinin, dirsek yöntemine göre üç kümeye, Sillhoutte yöntemine göre dört kümeye, Gap istatistik yöntemine göre ise üç kümeye ayrılması önerilmiştir. Belirlenen küme sayısının doğrulaması için bağlantı indeksi, Sillhoutte indeksi ve Dunn indeksi kullanılmıştır.

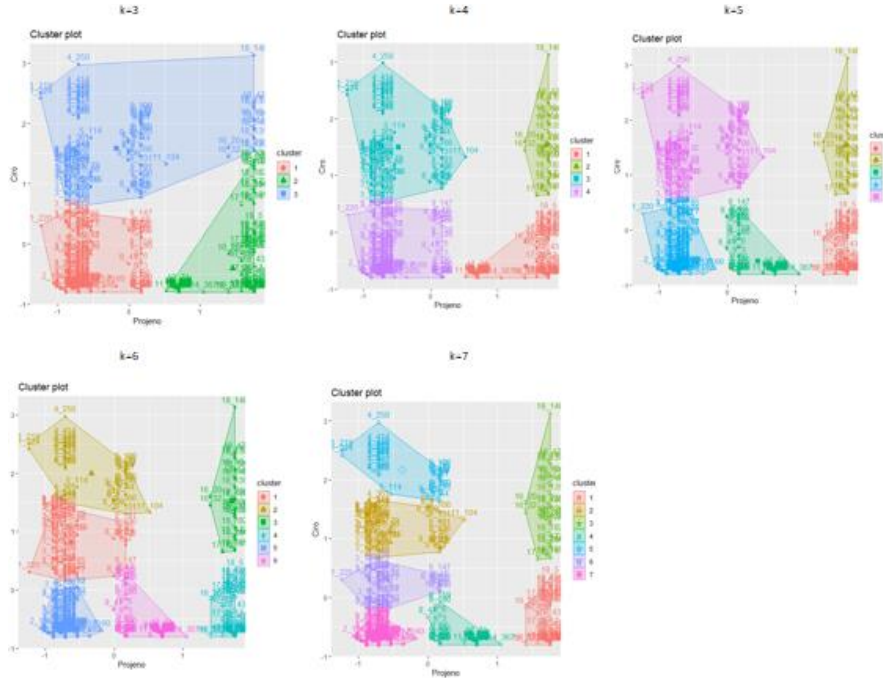
Tablo 1. Farklı Küme Sayıları için Validasyon Değerleri (Clustering Validation Indexes For Different Number of Clusters)

| Kümeleme Yöntemi | Validasyon Ölçüm Yöntemi | 3 Küme | 4 Küme | 5 Küme | 6 Küme | 7 Küme |
|----------------------------|--------------------------|--------|--------|---------|---------|---------|
| Hiyerarşik Kümeleme | Bağlantı İndeksi | 3,3778 | 8,3361 | 11,3901 | 12,0929 | 12,8119 |
| | Dunn İndeksi | 0,0773 | 0,0481 | 0,0481 | 0,0578 | 0,0578 |
| | Silhouette İndeksi | 0,4714 | 0,5836 | 0,5724 | 0,5449 | 0,6064 |
| K-Ortalamalar ile Kümeleme | Bağlantı İndeksi | 12,756 | 9,6663 | 15,9483 | 27,8802 | 34,0944 |
| | Dunn İndeksi | 0,0252 | 0,0391 | 0,0398 | 0,0479 | 0,0479 |
| | Silhouette İndeksi | 0,6063 | 0,6342 | 0,6039 | 0,5762 | 0,6215 |

Tablo 2. Optimal Küme Sayıları (Optimal Number of Clusters)

| Optimal Değerler | Değer | Yöntem | İdeal Küme Sayısı |
|--------------------|--------|----------------------------|-------------------|
| Bağlantı İndeksi | 3,3778 | Hiyerarşik Kümeleme | 3 |
| Dunn İndeksi | 0,0773 | Hiyerarşik Kümeleme | 3 |
| Silhouette İndeksi | 0,6342 | K-Ortalamalar ile Kümeleme | 4 |

Tablo 1’de kümeleme yöntemleri, optimal küme sayısının bulunması için farklı küme sayılarına göre hesaplanan indeks değerleri verilmiştir. Tablo 2’de görüleceği üzere hiyerarşik kümeleme yapılması durumunda bağlantı indeksi ve Dunn indeksi üç kümenin tercih edilmesi gerektiğini göstermektedir. Bu sonuç, Şekil 8’de gösterilen dirsek ve Gap istatistik yöntemi ile eşleşmektedir. Çalışmada tercih edilen K-ortalamalar yöntemi ile, Silhouette indeksine göre dört kümeye ayrılmalıdır. Bu sonuç Şekil 8’de gösterilen Silhouette yöntemi ile eşleşmektedir.

**Şekil 9.** Farklı Küme Sayıları için K-ortalamalar Yöntemi ile Kümeleme Çalışmaları (Clustering Applications with K-means Method for Different Number of Clusters)

Şekil 9’da veri setinin K-ortalamalar yöntemiyle farklı küme sayıları (3 kümeden 7 kümeye kadar) için kümeleme analizleri yapılmış ve belirlenen kümeler incelenmiştir. Analiz sonucunda incelenen veri setinin farklı kümelere

bölünebildiği, ancak Tablo 2'de yer alan Silhouette değerine göre en uygun küme sayısı olan dört kümeye bölünmesinin daha uygun olduğuna karar verilmiştir.

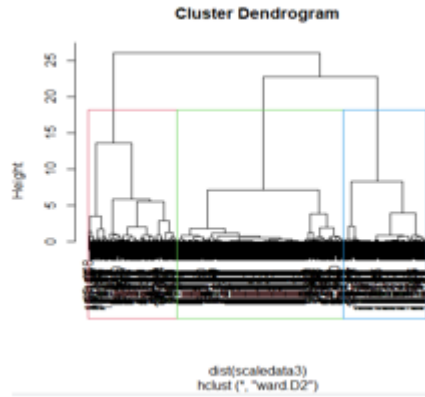
Veri seti dört kümeye bölündükten sonra incelenen 18 projeye ait referans numaralarının belirlenmiş kümelere göre dağılımları incelenmiştir.

Tablo 3. K-Ortalamlar Yöntemi ile Belirlenen Kümelerin Proje Bazlı Ayrılması (Project Based Classification of Clusters Determined by K-means Algorithm)

| | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P11 | P12 | P14 | P16 | P17 | P18 |
|--------|----|----|----|-----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|
| Küme_1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 25 | 1 | 3 | 7 | 54 |
| Küme_2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 22 |
| Küme_3 | 2 | 0 | 17 | 34 | 5 | 0 | 0 | 6 | 20 | 1 | 0 | 0 | 0 | 0 | 0 |
| Küme_4 | 1 | 1 | 19 | 149 | 16 | 2 | 1 | 2 | 22 | 0 | 0 | 0 | 0 | 0 | 0 |

Tablo 3'te K-ortalamlar yöntemine göre dört kümeye ayrılan veri seti içinde stratejik açıdan en önemli projenin dört numaralı proje olduğu tespit edilmiştir çünkü ciroya bağlı oluşturulan referansların dört numaralı kümede yığıldığı gözlemlenmiştir. Projeler arasında yer alan 10,13 ve 15 numaralı projelerde seride olan bir referans olmadığı için bu projedeki veriler gürültülü veri olarak işaretlenerek silinmiştir. Bu sebepten Tablo 3'de bu üç proje yer almamaktadır.

İncelenen veri seti, hiyerarşik kümeleme yöntemi ile de kümelendirilmiştir. Hiyerarşik kümeleme yöntemine göre Tablo 2'de yer alan bağlantı ve Dunn indeks değerlerine göre en uygun küme sayısı üç olarak önerildiği için, veri seti üç kümeye ayrılmış, incelenen 18 projeye ait referans numaralarının kümelere göre dağılımları incelenmiş ve hiyerarşik kümeleme yöntemine göre de stratejik açıdan en önemli projenin dört numaralı proje olduğu tespit edilmiştir.



Şekil 60. Hiyerarşik Kümeleme Dendrogramı (Hierarchical Clustering Dendrogram)

Tablo 4. Hiyerarşik Kümeleme Yöntemi ile Belirlenen Kümelerin Proje Bazlı Ayrılması (Project Based Classification of Clusters Determined by Hierarchical Clustering Algorithm)

| Küme Sayısı | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P11 | P12 | P14 | P16 | P17 | P18 |
|-------------|----|----|----|-----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|
| Küme_1 | 2 | 0 | 16 | 33 | 5 | 0 | 0 | 6 | 20 | 1 | 0 | 0 | 2 | 1 | 22 |
| Küme_2 | 1 | 1 | 20 | 150 | 16 | 2 | 1 | 2 | 11 | 0 | 0 | 0 | 0 | 0 | 0 |
| Küme_3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 1 | 25 | 1 | 3 | 7 | 54 |

Hiyerarşik kümeleme yöntemine göre kümelere ait dendrogram Şekil 10'da ve proje bazlı dağılım Tablo4'te verilmiştir.

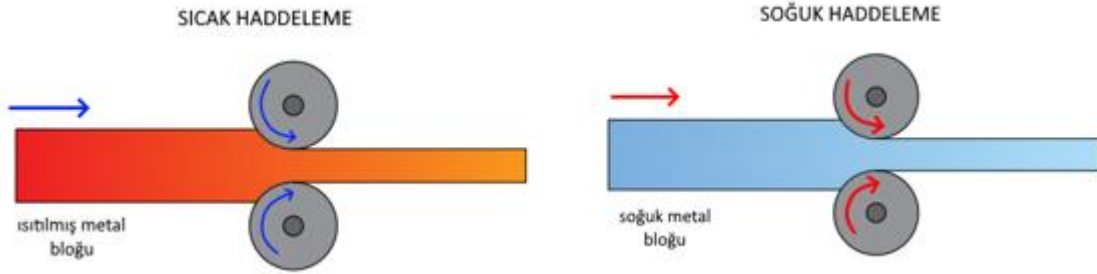
Yapılan kümeleme çalışmaları ile dört numaralı projenin farklı kümeleme yöntemleri olan hiyerarşik kümeleme ve K-ortalamlar kümeleme yöntemiyle en fazla stratejik öneme sahip proje olduğu ortaya konmuştur. Burada çalışandan kaynaklanabilecek hatalar ortadan kaldırılarak kümeleme yöntemleri ile hızlı ve hatasız çözüm

alınmıştır.

Stratejik açıdan en büyük öneme sahip dört numaralı proje ayrıntılı incelendiğinde ürün grubu, ortalama kalınlık ve tüketim verileri ortaya konduğunda farklı satın alma stratejileri aşağıdaki gibi oluşturulmuştur:

Dört numaralı projedeki çelik malzemelerin içerisinde üç ayrı ürün grubu söz konusudur,

- HR (Hot Rolled)-> Sıcak Haddelenmiş Çelik Malzemeler
- HDG (Hot Dip Galvanized)-> Sıcak Galvanizlenmiş Çelik Malzemeler
- CR (Cold Rolled)-> Soğuk Haddelenmiş Çelik Malzemeler



Şekil 11. Haddeme Operasyonları (Rolling Operations)

HR, HDG ve CR ürün grubu operasyonları Şekil 11’de gösterilmiştir. Sıcak haddelenmiş çelik malzemeler Şekil 11’deki sıcak haddeme operasyonundan geçmektedir. Soğuk haddelenmiş malzemeler ise, sıcak haddelendikten sonra, soğuk haddeme operasyonuna girmektedir. Sıcak galvanizli malzemeler ise, haddeme operasyonlarının haricinde galvaniz havuzlarına girerek galvanizleme operasyonu geçirmektedir. Üç ayrı ürün grubuna göre önerilen satın alma stratejileri aşağıda açıklanmıştır:

Strateji_1: HR ürün grubunda ortalama kalınlık projedeki çeliklerin ortalama kalınlığının yaklaşık %60 üzerindedir. Bu projede ortalama kalınlığın üzerinde kalan çelik malzemeler için gerekli teknik kontroller yapılarak kalınlık azaltma çalışması yapılabilir. Bu çalışma yapılırken, mevcut çeliklerin hangi fonksiyonda kullanılacağı ve dayanım, mukavemet gibi teknik detayları göz önünde bulundurulmalıdır. Çelik malzemede kalınlık azaltılabilirse, toplam maliyette de düşüş görülecektir.

Strateji_2: HDG malzemeler tüketim tonajı en yüksek ürün grupları arasında yer almaktadır. Galvanizleme, sıcak haddeme haricinde ekstra bir operasyon olduğu için, ekstra maliyet getirdiği bilinmektedir. Bu grupta yer alan malzemelerde galvanizleme operasyonu kaldırılarak, HR olarak beklentileri yerine getirip getirmediği araştırılmalıdır. Bunun haricinde HDG grubunda kalitelerde bir ortaklaştırma ile maliyet avantajı sağlanabilir.

Strateji_3: CR, HR’ye göre daha iyi yüzey kalitesine sahip olup hassas uygulamalar için uygundur. Ancak CR’de kesilme, kıvrılma veya kaynak operasyonu malzemedeki gerilimler serbest kalabilir ve malzemede kırılma veya çatlak gözlemlenebilir. HR’ye göre ekstra bir operasyon daha içerdiği için piyasadaki ekonomik dalgalanmalardan hızlı etkilenme ve fiyat artışına uğrama riski vardır. Malzemenin kullanıldığı yer, CR hassasiyeti gerektirmiyor ise CR ürün grubu tercih edilmemelidir.

5. Sonuç ve Tartışma (Results and Discussion)

Bu çalışmada bir otomotiv işletmesinde çelik malzeme satın alma süreçleri ile ilgili veri seti için veri madenciliği teknikleri ile gürültülü verilerin temizlenmesi, makine öğrenmesi yöntemleri ile kümelemenin yapılması, ideal küme sayısının belirlenmesi, kümelerin doğrulanması ve karlılık içeren satın alma stratejilerinin geliştirilmesine yönelik çalışmalar yapılmıştır.

Çelik malzeme satın alma veri setinde kritiklik arz eden referansların kolayca tespiti, satın alma süreçlerinde sıklıkla karşılaşılan gürültülü verilerin temizlenmesi ve önemli referansların kümelenecek gelecek dönemler için satın alma stratejilerinin oluşturulmasında fayda sağlamıştır. Satın alma çalışanlarının uzun zaman harcayarak yapacağı analizler, veri analizi ve makine öğrenmesi gibi yöntemler ile gerçekleştirilmiştir. Bu çalışmada önerilen analizler ile satın alma süreçlerinde çalışan kaynaklı hataların satın alma stratejileri geliştirme süreçlerindeki etkileri azaltılmıştır.

Satın alma verilerinin veri madenciliği teknikleri kullanılarak temizlenmesi ve analiz edilmesi, stratejik öneme sahip projelerin belirlenmesine önemli katkılar sağlamaktadır. Bu teknikler, veri setlerindeki gürültüyü ve anlamsız verileri tespit ederek veri kalitesini artırmakta ve doğru analizler yapılmasını sağlamaktadır. Satın alma verilerinin analiz edilmesi, tedarikçi performansı, talep tahminleri, stok yönetimi gibi stratejik konuların da anlaşılmasına ve gelecekteki projelerin belirlenmesine yardımcı olmaktadır. Makine öğrenmesi yöntemleriyle çelik malzeme satın alınmasında stratejik öneme sahip ürün kümeleri belirlenebilmektedir. Bu yöntemler, büyük veri setlerindeki karmaşık örüntüleri tanımak ve tahminler yapmak için kullanılmaktadır. Belirlenen ürün kümeleri için uygun satın alma stratejileri oluşturulabilmesi için çeşitli analizler yapılabilmektedir. Fiyat analizleri, rekabet analizleri, tedarikçi performans analizleri gibi analizler, uygun satın alma stratejilerinin belirlenmesine yardımcı olmaktadır. Veri madenciliği ve makine öğrenmesi teknikleri kullanılarak oluşturulan stratejik kümeler, çelik malzeme satın alma stratejileri için değerlendirilebilmektedir. Bu kümelerin içerisindeki çelik malzemelerin tedarikçi çeşitliliği, kalitesi, talep tahminleri ve risk analizleri gibi faktörler değerlendirilerek stratejik kararlar alınabilir.

Veri madenciliği teknikleri ve analitik yeteneklerin kullanımı, satın alma çalışanlarının daha bilinçli ve veri odaklı kararlar almasını sağlar. Satın alma verilerinin temizlenmesi ve analiz edilmesi sayesinde, stratejik konulara ilişkin daha iyi bilgilere ulaşılabilir. Bu da satın alma çalışanlarının etkili stratejiler geliştirmelerine ve daha iyi sonuçlar elde etmelerine yardımcı olur.

Ayrıca, makine öğrenmesi yöntemleriyle belirlenen stratejik ürün kümeleri, satın alma çalışanlarına tedarikçi yönetimi konusunda önemli bir yol göstericidir. Tedarikçilerden satın alınan malzemelerin hangilerinin stratejik öneme sahip olduğu ve ciroda nasıl bir etkisi olduğu belirlenebilir. Bu bilgiler, satın alma çalışanlarının tedarikçi ilişkilerini güçlendirmelerine ve daha iyi anlaşmalar yapmalarına yardımcı olur. Benzer şekilde stratejik öneme sahip malzemeler tespit edildikten sonra, maliyet azaltılması için alternatif malzeme araştırılması gibi çalışmalar mühendislik birimleri ile birlikte çalışılabilir.

Bu çalışma aynı zamanda maliyet optimizasyonu açısından da satın alma çalışanlarına fayda sağlar. Veri madenciliği ve makine öğrenmesi teknikleriyle yapılan analizler, maliyetlerin daha iyi yönetilmesine yardımcı olabilir. Fiyat analizleri, rekabet analizleri ve tedarikçi performans analizleri gibi analizler, satın alma çalışanlarının maliyetleri optimize etmelerine ve bütçe tasarrufu sağlamalarına yardımcı olur.

Sonuç olarak, bu çalışmada sunulan veri madenciliği ve makine öğrenmesi teknikleri, satın alma çalışanlarının veri odaklı kararlar almasını, stratejik tedarikçi yönetimi yapmasını ve maliyet optimizasyonu sağlamasını desteklemektedir. Bu yöntemlerin kullanımıyla satın alma çalışanları, daha etkili ve verimli bir şekilde görevlerini yerine getirebilir, işletme performansını artırabilir ve rekabet avantajı elde edebilirler. Çalışma kapsamında geliştirilen modeller, bir otomotiv firmasına ait gerçek veriler kullanılarak geliştirildiğinden, satın alma stratejilerinin geliştirilmesi ile ilgili literatüre katkı sağlayacağı düşünülmektedir. Çalışmada elde edilen sonuçlar, malzeme fiyat tahminleme yöntemleri için girdi olarak kullanılabilir. Özellikle referans bazlı fiyat tahminleme araçlarının geliştirilmesi ve kullanılması için bu çalışmanın sonuçlarının kullanılabileceği düşünülmektedir. Bununla birlikte, gelecekteki çalışmalarda satın alma süreçlerine daha derinlemesine odaklanılması ve malzeme yönetimi stratejilerinin iyileştirilmesi önerilmektedir.

Çıkar Çatışması (Conflict of Interest)

Yazarlar tarafından herhangi bir çıkar çatışması beyan edilmemiştir. No conflict of interest was declared by the authors.

Kaynaklar (References)

- Al-Omary, A. Y., & Jamil, M. S. (2005). A new approach of clustering based machine-learning algorithm. *Knowledge Based Systems*, 19, 248-258.
- Asilkan, Ö. (2008). Veri Madenciliği Kullanılarak İkinci El Otomobil Pazarında Fiyat Tahmini [Akdeniz Üniversitesi].
- Babaoğlu, A. (2015). Veri Madenciliği Yöntemleri ve Bir Uygulama. Selçuk Üniversitesi. 117.
- Cui, M. (2020). Introduction to the K-Means Clustering Algorithm Based on the Elbow Method.
- Erbudak, A. E. (2022). Veri Madenciliği Ve Makine Öğrenimi İle Döviz Kuru Tahmini Uygulaması. 97.
- Erpolat, S. (2012). Otomobil Yetkili Servislerinde Birliktelik Kurallarının Belirlenmesinde Apriori ve FP-Growth Algoritmalarının Karşılaştırılması. *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, 12(1 (151-166)).
- Henderson, A. R. (2006). Testing experimental data for univariate normality. *Clinica Chimica Acta*, 366(1-2)
- Jeong, J., Park, S., & Lee, C. (2016). Comprehensive comparison of normality tests: Empirical study using many different types of data. *Journal of the Korean Data and Information Science Society*, 27(5), 1399-1412.
- Kameshwaran, & Malarvizhi. (2014). Survey on Clustering Techniques in Data Mining. 5, 5.

- Kirgiz, A. (2021). Lüks Otomotiv Sektöründe Satın Alma Kararını Etkileyen Faktörler. *R&S - Research Studies Anatolia Journal*.
- Köylüoğlu, A. S., Acar, Ö. E., & İnan, Ü. S. E. (2018). Tüketicilerin Otomobil Satın Alma Davranışlarına Etki Eden Faktörlerin Belirlenmesi: Akademisyenlere Yönelik Bir Uygulama. *Selçuk Üniversitesi Sosyal Bilimler Meslek Yüksekokulu Dergisi*, 21(2), 251-273.
- Liebchen, G., Twala, B., Shepperd, M., Cartwright, M., & Stephens, M. (2007). Filtering, Robust Filtering, Polishing: Techniques for Addressing Quality in Software Data. *First International Symposium on Empirical Software Engineering and Measurement (ESEM 2007)*, 99-106.
- Liu, B., Xia, Y., & Yu, P. S. (2000). Clustering Through Decision Tree Construction. *Conference Of Information And Knowledge Management*. 10.
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.
- Özgür, A. (2002). Supervised And Unsupervised Machine Learning Techniques For Text Document Categorization.
- Öztuna, D., Elhan, A. H., & Tüccar, E. (t.y.). Investigation of Four Different Normality Tests in Terms of Type 1 Error Rate and Power under Different Distributions. *Turk J Med Sci*.
- Patel, P., Sivaiah, B., & Patel, R. (2022). Approaches for finding Optimal Number of Clusters using K-Means and Agglomerative Hierarchical Clustering Techniques. *2022 International Conference on Intelligent Controller and Computing for Smart Power (ICICCSPP)*, 1-6.
- Russom, P. 2011. Big data analytics. TDWI Best Practices Report, 9/14/2011. <https://tdwi.org/research/2011/09/best-practices-report-q4-big-data-analytics.aspx?tc=page0&tc=assetpg&m=1>
- Saputra, D. M., Saputra, D., & Oswari, L. D. (2019). Effect of Distance Metrics in Determining K-Value in KMeans Clustering Using Elbow and Silhouette Method. *Advances in Intelligent Systems Research, Proceedings of the Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019)*(172).
- Safaei, A.S., Heidarpour, F., Paydar, M.M., 2018. Group purchasing organization design: a clustering approach. *Comput. Appl. Math.* 37, 2065–2093. <https://doi.org/10.1007/s40314-017-0439-8>
- Savaş, S., Topaloğlu, N., & Yılmaz, M. (2012). Veri Madenciliği Ve Türkiye'deki Uygulama Örnekleri. 23.
- Shobha, N., & Asha, T. (2017). Monitoring weather based meteorological data: Clustering approach for analysis. *2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, 75-81.
- Tiwari, S., H.M. Wee, and Y. Daryanto. 2018. Big data analytics in supply chain management between 2010 and 2016: Insights to industries. *Computers and Industrial Engineering* 115: 319-330. <https://doi:10.1016/j.cie.2017.11.017>
- Verlinden, B., J.R. Duflou, P. Collin, and D. Cattrysse. 2008. Cost estimation for sheet metal parts using multiple regression and artificial neural networks: A case study. *International Journal of Production Economics* 111(2): 1, 2008): 484–92. <https://doi:10.1016/j.ijpe.2007.02.004>
- Qing, H., Zhang, J., Fu, D., 2021. Data Mining Technology in Business Data Analysis. *J. Phys. Conf. Ser.* 1852, 022045. <https://doi.org/10.1088/1742-6596/1852/2/022045>
- Wang, X., & Xu, Y. (2019). An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index. *IOP Conference Series: Materials Science and Engineering*, 569(5), 052024.
- Zhao, C., Johnsson, M., He, M., 2017. Data Mining with Clustering Algorithms to Reduce Packaging Costs: A Case Study: Data mining approaches to reduce package costs: a case study. *Packag. Technol. Sci.* 30, 173–193. <https://doi.org/10.1002/pts.2286>