

## Perakende Satışlarda Çapraz Satış Davranışlarının Tespiti ve Analizi

Detection and Analysis of Cross-Selling Behaviors in Retail Sales

Eren Acar

Yeditepe Üniversitesi Yazılım Geliştirme Bölümü, [erenacar3747@gmail.com](mailto:erenacar3747@gmail.com)

### ÖZET

Çapraz satışın tespiti ve bu tespite dayalı bazı stratejiler, müşteri başına düşen hacmi ve müşteri yaşam boyu değeri gibi bazı istatistikleri yükselterek firmanın karlılığını arttıran olgulardan biridir. Çapraz satış fırsatları başta perakende satış yapan firmalar olmak üzere dijital satış yapan ticaret siteleri ve finansal yatırım operasyonları yöneten danışmanlık şirketlerinde sıklıkla yaşanmaktadır. Müşterilerin geçmiş davranışları model alınarak yapılan analizlerde gelecek stratejiler için etkili öngörüler ve aksiyonlar alınabilir. Bu makalede perakende satış yapan bir firmanın yaklaşık üç yıllık satış verisi çeşitli makine öğrenimi ve veri madenciliği teknikleri kullanılarak müşteri segmentasyonu ve ürünler arasındaki olası çapraz satış davranışları analiz edilmiştir. K-Means aracılığıyla müşteri ve ürün bazlı bir kümeleme yapıldıktan sonra her kümenin bazı favori ürünüyle ilişkili olduğu bazı ürünler tespit edilip çapraz satış ya da üst satış yaşanıp yaşanılmadığı incelenip, aynı zamanda kümelenen müşterilerin genel davranışları gözlemlenmiştir. Kümelemenin dağılımı ve küme sayısının doğru seçilip seçilmediğini test etmek için veri setindeki son altı aylık işlemlerle geri kalan işlemler ayrılıp iki veri setinin BIC, Silhouette ve WCSS skorları hesaplanıp karşılaştırılmıştır. Çalışmada çapraz satışları tespit etmek için kullanılan bir diğer yöntem ise birliktelik kural madenciliğidir. Bu yöntemle çalışan Apriori, FP-Growth ve Eclat algoritmalarından belli sınırlamalar ile oluşturulan modellerle bazı ürün çiftlerinin destek, güven ve kaldıraç değerleri hesaplanılarak veri setindeki çapraz satış olayları tespit edilmiş olup aynı yöntemle çalışan üç modelin arasındaki farklara değinilmiştir. Her modelin geliştirilme sürecinde CRISP-DM metodolojisinin ilgili kısımları kullanılmıştır. Ayrıca bu çalışmadaki yaklaşımlara alternatif olarak Optiwisdom tarafından geliştirilen Kolay.ai gibi yapay zekâ destekli analiz platformlarının bu çalışmadaki karşılığı ele alınıp bu tarz yapay zekâ motorlarının sunduğu dinamik öneri sistemlerinin gerçek zamanlı veri analizi üzerindeki etkisi vurgulanmıştır.

**Anahtar Kavramlar:** Çapraz Satış, Müşteri Segmentasyonu, Makine Öğrenimi, Birliktelik Kural Madenciliği, Veri Madenciliği

### ABSTRACT

Cross selling is one of the concepts that increases profitability by boosting statistics such as volume per customer and customer lifetime value. Cross sell opportunities often occur in retail companies, e-commerce websites and financial advisory firms managing investment operations. By analyzing past customer behaviors effective predictions and actions can be taken for future strategies. In this study around three years of sales data from a retail company were analyzed using various machine learning and data mining techniques to detect customer segmentation and potential cross sell behaviors between products. K-Means clustering was applied to group customers and products identifying certain products related to each cluster's favorite items and checking if cross sell or upsell events occurred. At the same time general behaviors of the clustered customers were observed. To test if the cluster distribution and number of clusters were correctly chosen the last six months of transactions were separated from rest of the dataset. BIC, Silhouette and WCSS scores were calculated and compared between the two datasets. Another method used to

detect cross sell events in this study is association rule mining. Using this method models were created with certain constraints based on Apriori, FP-Growth, and Eclat algorithms. The support, confidence and lift values of some product pairs were calculated to detect cross sell events in the dataset and the differences between three models were discussed. During development of each model, relevant parts of CRISP-DM methodology were applied. Additionally, as an alternative approach to this study, AI-powered analysis platforms like Kolay.ai developed by Optiwisdom were evaluated highlighting impact of dynamic recommendation systems on real time data analysis.

**Keywords:** Cross-Selling, Customer Segmentation, Machine Learning, Association Rule Mining, Data Mining

## 1 GİRİŞ

Çapraz satış, mevcut müşterilere satın alma davranışlarıyla uyumlu ek hizmet veya ürün sunarak satışları ve kârlılığını artırmak için kullanılan etkili bir yöntemdir. Bu yöntem, müşterilere halihazırda satın aldıkları ürün veya hizmetlerle uyumlu ek ürün veya hizmet önerme stratejisidir. Bu strateji, satış hacmini müşteri başına arttırmak için kullanılan araçlardan biridir [1]. Müşterilerin sepet değerini artırmanın yanı sıra, müşteri sadakatini ve yaşam boyu değerini de yükseltmeye yardımcı olur [2, 3].

Makine öğrenimi algoritmaları sayesinde, özellikle büyük veri setleri ve karmaşık müşteri davranışlarının analiz edilmesi gereken durumlarda, müşterilerin satın alma eğilimleri ve alışveriş alışkanlıkları gibi çeşitli veriler tespit edilebilir. Böylece işletmeler, müşterilere daha kişiselleştirilmiş ve ilgili öneriler sunabilir [4].

Makine öğrenimi uygulamalarında; müşterilerin geçmiş satın alma verileri, demografik bilgileri ve ödeme yöntemleri gibi çeşitli özellikler dikkate alınır [5]. Bu çalışmada, çapraz satışları tespit eden algoritmaların nasıl müşteri segmentasyonunu belirlediği, verileri nasıl incelediği ve bu verilerle çapraz satış ihtimali yüksek olan ürünleri nasıl grupladığı detaylıca incelenecektir.

## 2 LİTERATÜR ARAŞTIRMASI

Yapılan çalışmalarda makine öğrenimi modelleri, veri analizinden anlamlı bilgiler elde etmek ve firmaların kârlarını çeşitli pazarlama teknikleriyle artırmak için oldukça popüler hale gelmiştir [6]. Çapraz satış analizlerinde kullanılan algoritmalar, veri madenciliği ve makine öğrenimi teknikleriyle müşterilerin satın alma davranışlarını anlamak ve satış stratejilerini geliştirmek için kullanılmıştır. Literatürdeki yapılan çalışmaların ortak yönü, müşteri davranışlarının analizi ve bu analizlere bağlı olarak doğru algoritmaları tespitidir.

Çapraz satış, perakende sektörünün yanı sıra bankacılıkta da sıkça kullanılan etkili bir satış stratejisidir. 2017 yılında Özden Gür Ali ve ekibi tarafından yapılan çalışmada [11], yatırım envanterinin çapraz satışı hem müşterinin hem de bankanın çıkarlarını gözetken bir portföy optimizasyonu ile ele alınmıştır. Çalışmanın amacı, müşterinin portföyündeki risk-getiri dengesini iyileştirirken bankanın kârlılığını artırmak ve böylece her iki taraf için de kazan-kazan durumu yaratmaktır. Bu çapraz satış yaklaşımı, müşterinin mevcut portföyü üzerine inşa edilerek, yeni ürünlerin kabul edilme olasılığını tahmin eden lojistik regresyon modeline dayanmaktadır. Çalışmada oluşturulan modelin başarısını ölçmek için AUC (Area Under the Curve, Eğrinin Altındaki Alan) metriği, her ürün için ölçülmüştür. AUC değeri ürünler genelinde normalize edilmiş olarak yüzde oranlarıyla verilmiştir. "Current-Use" seti, müşterinin portföyündeki mevcut ürünleri gösteren 41 ikili değişkenden oluşturulmuştur ve bu veri setin AUC değeri %88 olarak ölçülmüştür. Tüm değişkenler kullanıldığında ise en yüksek ortalama AUC değeri %92 olarak kaydedilmiştir. Lojistik regresyon, müşteri kabul olasılığını tahmin etmede kullanılan yalın, hızlı anlaşılır bir yöntemdir. Ancak, bu modelin doğrusal varsayımlar üzerine kurulu olması, müşteri davranışlarının karmaşık yapısını tam anlamıyla yakalayamayabilir. Özellikle müşteri kabul oranının düşük olduğu durumlarda, veri

dengesizliği sorunu ortaya çıkabilir, bu da modelin kabul edilmeyen durumları daha iyi tahmin etmesine neden olabilir [12]. Literatürdeki benzer çalışmalar, çapraz satış genellikle bankanın kârını artırmaya odaklanır [13, 14], ancak Özden ve ekibinin çalışması hem bankanın kârlılığını hem de müşterinin çıkarlarını gözetken bir kazan-kazan stratejisi önermektedir. Bu yaklaşım, müşterilerin mevcut portföylerini dikkate alarak, teorik olarak optimal portföylere geçişte yaşanan zorlukları aşmayı hedeflemektedir. Müşterinin portföy getirilerini artırırken, risk seviyesini yükseltmeden bankanın kârını önemli ölçüde iyileştiren bu model, çapraz satış stratejisinde önemli bir boşluğu doldurmaktadır. Ayrıca, müşterinin mevcut portföyünden hareketle, kabul edebileceği ürünleri doğru şekilde tahmin ederek hem müşteri sadakatini artırmakta hem de bankanın uzun vadeli kârlılığını güvence altına almaktadır. Bu çapraz satış modeli, klasik yaklaşımlardan farklı olarak, müşterinin mevcut durumunu ve yatırım tercihlerini dikkate alırken, bankanın kârını artırma hedefini de göz önünde bulundurarak, sürdürülebilir ve uzun vadeli müşteri ilişkileri kurmayı amaçlamaktadır.

Olası çapraz satışların analizi ve tespiti için kümeleme algoritmaları da kullanılabilir. 2007 yılında Maria Teresa Salazar ve ekibinin yayınladığı bir makalede [15] müşteri verilerinin analizine dayalı çapraz ve yukarı satış fırsatlarını belirlemeye yönelik kapsamlı bir çalışma sunmaktadır. Müşterinin satın alma davranışlarını analiz etmek için kümeleme algoritmalarından faydalanılmıştır. K-Means algoritması, müşterileri sosyo-demografik ve tüketim davranışlarına göre dört farklı kümeye ayırmıştır. Kümelerin belirlenmesinde, MSE (Mean Squared Error, Ortalama Kareysel Hata) metriği kullanılmış ve dört kümelili çözümün en uygun olduğu kararlaştırılmıştır. Ayrıca, küme çözümünün doğruluğunu test etmek için ayırma analizi (discriminant analysis) kullanılmış ve %95'in üzerinde doğru sınıflandırma oranına ulaşılmıştır. Bu segmentasyonun sonucunda, her kümenin kendine özgü tüketim ve satın alma davranışları olduğu belirlenmiştir. Örneğin, bazı kümeler daha fazla ürün satın alırken, diğerleri daha az ürün almış; ayrıca, yaş ve finansal bilinç düzeyi gibi faktörlerin de satın alma kararları üzerinde etkili olduğu gözlemlenmiştir. K-Means algoritması, müşterilerin farklı yaşam döngüsü aşamalarında hangi ürünleri tercih ettiklerini anlamak ve çapraz satış fırsatlarını belirlemek için etkili bir araç olarak kullanılmıştır. Bu sayede, her bir segment için uygun çapraz satış stratejileri geliştirilebilmiştir. Oluşturulan gruptan sonra, her bir grubun ortak satın alma örüntüsü analiz edilmiş ve bu analizlerin doğrultusunda çapraz satış fırsatları "Satın Alma Ağaçları" (Purchase Trees) olarak adlandırılan görselleştirme tekniğiyle görselleştirilmiştir. Salazar ve ekibinin çalışmasında kullanılan K-Means algoritması, müşteri segmentasyonu için etkili bir yöntemdir ancak küme sayısının önceden belirlenmesi ve öklit mesafe metriği gibi sınırlamaları bulunmaktadır. Bu sınırlamalar, karmaşık ve doğrusal olmayan müşteri ilişkilerini tam olarak yansıtmayı zorlaştırabilir. İlgili çalışma, çapraz satış fırsatlarını belirlemenin yanı sıra, lojistik regresyon ve Cox regresyonu gibi teknikler kullanarak yeniden satın alma eğilimlerini de tahmin etmeyi hedeflemektedir. Bu sayede, müşterilerin hangi zaman diliminde hangi ürünleri satın alma olasılıklarının ne olduğunu ve ne zaman yeni bir satın alma yapacaklarını da tahmin etmeyi amaçlayıp analiz etmiştir. Bu analiz, olası çapraz satış davranışlarının zamanlamasını iyileştirmektedir.

Kümeleme algoritmaları gibi denetimsiz algoritmaların performansını ve oluşan kümelerin iç yapısını değerlendirmek için çeşitli metrikler mevcuttur. Rousseeuw tarafından bulunan ve önerilen silhouette yöntemi [7], oluşturulan kümelerin doğruluğunu değerlendirmek için kullanılan bir metriktir. Bu yöntem, her bir girdinin ait olduğu kümesine olan uyumunu ölçerek grafiksel bir yapı sunar ve doğal küme sayısının belirlenmesinde önemli bir rol oynar. Silhouette genişliklerinin kümeleme yapılarını değerlendirmek ve yanlış sınıflandırmaları belirlemek için etkili bir ölçüt olduğu tespit edilmiştir. Ayrıca, bu yöntem sadece K-Means veya K-Median gibi belirli kümeleme algoritmalarında kullanılmaz aynı zamanda farklı algoritmaların sonuçlarını karşılaştırmak veya iyileştirmek için de kullanılabilir. Silhouette yöntemini uyguladıktan sonra oluşan değer -1 ile +1 arasındadır. Bu değer +1'e yakın

ise o kümedeki ilgili girdi, kendisine komşu olan kümeye oldukça uzak olduğunu, sıfır ise girdinin kümesinin iki komşusuna olan uzaklığı karar sınırına çok yakın bir değerde olduğunu, negatif değerler ise bu girdinin yanlış bir kümeye ait olabileceğini göstermektedir.

Kümeleme algoritmalarında kullanılan diğer bir metrik ise WCSS (Within-Cluster Sum of Squares) yaklaşımıdır. Edwards ve Cavalli-Sforza'nın çalışmasında [8] detaylandırılan bu yöntem, çok boyutlu veri noktalarının uygun bir şekilde gruplandırılmasını sağlamak amacıyla tasarlanmıştır. WCSS bir küme içerisindeki noktaların kendi küme merkezlerine olan uzaklıklarının karelerinin toplamı olarak tanımlanır ve böylelikle küme içi yoğunluk ölçülebilir. Yöntemin temel amacı, kümeler arası ayrımı olabildiğince maksimize ederken küme içi benzerliği arttırmaktır. Literatürde bu yöntem biyolojik verilerde doğal gruplamanın tespiti ve sosyolojik çalışmalarda grupların analizi gibi çeşitli farklı alanlarda başarıyla uygulanıp kullanılmıştır. Özellikle K-Means gibi modern algoritmaların temelinde bulunan optimizasyon prensini, bu yöntemin sunduğu matematiksel altyapı ile paralellik gösterir. WCSS'nin önemi, veri noktalarının stabil şekilde bölünmesini sağlamak için etkili bir performans sunmasından kaynaklanmaktadır.

İstatistiksel model seçimi için yaygın olarak kullanılan bir diğer yöntem ise Bayes Bilgi Kriteridir (Bayesian Information Criterion, BIS). George Schwarz'ın 1978 yılında yaptığı çalışmada [9, 10] geliştirilen bu metrik, büyük veri setlerinde olasılık modelleri arasında seçim yapmak için olasılıkları asimptotik olarak yaklaştıran bir yöntem sunmaktadır. BIC'in popülerliği, farklı algoritmalarda kullanmakta ve hesaplama kolaylığından gelmektedir. Özellikle değişkenlerin ve parametrelerin yapılandırılmalarının fazla olduğu kompleks modeller için etkili seçim kriteridir. BIC, işlem süresi kısa, ekonomik modelleri tercih etmektedir. Aday modeller arasındaki en düşük BIC değerine sahip model seçilir. Bu özellik, modelin gereksiz karmaşıklıkla aşırı uyum sağlamasını engellemektedir.

Perakende sektöründeki satış stratejilerini geliştirmek ve çapraz satış analizi yapmak için birliktelik kural madenciliğine bağlı Eclat, Apriori ve FP-Growth gibi algoritmalar da kullanılabilir. 2018 yılında Kumar ve ekibinin yaptığı bir çalışmada [17] perakende satış yapan bir firmanın beş aylık satış verisini kullanarak müşteri davranışları üzerinden çapraz satış fırsatlarını belirlemek amacıyla birliktelik kural madenciliği yapılmıştır. Çalışmada sık sık birlikte satılan ürünleri ve bu ürünler arasındaki birlikteliği belirlemek amacıyla Apriori algoritması kullanılmıştır. Çalışmada kullanılan veri setinde toplamda 52.674 işlem ve 9.055 adet farklı ürün yer almaktadır. Elde edilen birliktelik kuralları, ürünlerin birlikte satın alma olasılıklarını belirlemede önemli rol almıştır. İlgili makalede birliktelik kurallı analizlerde kullanılan destek (support), güven (confidence) ve kaldırmaç (lift) gibi temel metriklerin sonuçları da gözlemlenmiştir. Destek metriği, belirli ürünlerin kaç defa birlikte satın alındığını incelerken güven metriği ise satın alınan bir ürün ile olası satın alınacak bir ürünün satın alma olasılığını belirler. Kaldırmaç metriği ise ürünler arasındaki ilişkinin gücünü değerlendirerek bir ürünün diğer bir ürünün satın alınma olasılığını ne kadar arttırdığını ölçen bir metriktir. İlgili çalışmada, destek metriği %0.0001 olarak hesaplanmıştır, bu değer 52.674 işlemden yalnızca 53 tanesinde belirli bir ürün kombinasyonunun tekrarlandığı gözlemlenmiştir. Bu durum, incelenen ürün kümelerinin oldukça kısıtlı ve spesifik kombinasyonlar olduğunu gösterir. Çalışmada güven değeri %25 olduğu tespit edilmiştir, bu da bir ürün grubunun başka bir ürün grubunda %25 olasılıkla bulunabileceğini ifade etmektedir. Çalışmada ayrıca kaldırmaç metriği kullanılarak ürünler arasındaki ilişki gücü değerlendirilmiştir. Kaldırmaç değeri 1'den büyük olan ilişkiler, iki ürün arasında güçlü bir ilişki olduğunu ve buna bağlı olarak alınan bir ürünün diğer bir ürün ile birlikte alınabilme olasılığının oldukça güçlü olduğunu ifade etmektedir. Çalışmanın bulguları, mağazanın kendi markalı ürünlerinin diğerlerinden daha fazla talep gördüğünü ve çapraz satış analizlerinin perakendecilere ürün yerleşimi ve indirim stratejileri gibi kararları vermede önemli katkılar sunduğunu göstermiştir. Sonuç olarak, bu tür analizlerin perakendecilerin satışlarını arttırmalarına ve

müşterilere daha iyi hizmet sunmalarına olanak tanıdığı vurgulanmıştır. Kumar ve ekibinin ilgili çalışmada kullandığı Apriori algoritmasının önemli avantajlarından biri, özellikle küçük ve orta ölçekli veri setlerinde uygulanabilir olmasıdır. Algoritma, sık öge kümelerini belirlemede etkin bir yöntem sunar ve ürünler arasındaki ilişkileri keşfetmek için güçlü bir çerçeve sağlar. Bununla birlikte, büyük veri setlerinde algoritmanın işlem süresi ve hesaplama maliyeti artabilir. Apriori algoritması, tüm potansiyel ürün kombinasyonlarını taradığı için, veri setinin boyutu büyüdükçe hesaplama maliyetleri yükselir. Bu da algoritmanın büyük ölçekli perakende verileri üzerinde verimsiz hale gelmesine neden olabilir. Ayrıca, nadir görülen ürün kombinasyonlarını yakalama kapasitesinin sınırlı olması, stratejik olarak önemli olabilecek çapraz satış fırsatlarının kaçırılmasına yol açabilir [18]. Bu tür sınırlamaların aşılabilmesi için FP-Growth ve Eclat gibi birliktelik kural madenciliğine bağlı alternatif algoritmalar kullanılabilir. FP-Growth algoritması, FP-tree (Frequent Pattern Tree, Sık Örüntü Ağacı) yapısını kullanarak sık öge kümelerini daha verimli bir şekilde bulabilme yeteneği ile dikkat çeker ancak oluşturulan bu ağaçların karmaşıklığı Apriori'ye göre daha fazla bellek tüketimine sebep olabilir. Bir diğer alternatif algoritma ise Eclat algoritmasıdır. Eclat, derin öncelikli arama (Depth First Search, Derin Öncelikli Arama) yönetimi kullanır. Böylelikle veri setindeki anlamları dikey ekseninde sıralar ve tespit eder. Bu yapı sayesinde verimliliği yüksek tutar ancak kurduğu çıkarımlar kısıtlı olabilmektedir [19].

Bu çalışmada, çapraz satış fırsatlarını tespit etmek amacıyla kullanılan farklı tekniklerin ve algoritmalarının performansları karşılaştırılmıştır. K-Means, müşterileri segmentlere ayırarak benzer satın alma davranışlarına sahip grupları belirlemek için kullanılmıştır. Apriori, Eclat ve FP-Growth algoritmaları, ürünler arasındaki birliktelik kurallarını tespit etmek amacıyla kullanılmıştır. Bu üç algoritma da ürünler arasındaki ilişkileri analiz ederek çapraz satış fırsatlarını belirlemeye yönelik güçlü araçlar sunmaktadır. Apriori, basit ve anlaşılır bir yöntem olmakla birlikte, büyük veri setlerinde işlem süresi ve verimlilik açısından sınırlamalarla karşılaşabilir. FP-Growth algoritması, sık öge kümelerini bulma sürecini daha verimli hale getirecek şekilde tasarlanmış olup, farklı veri yapılarında etkili sonuçlar elde edebilmektedir. Eclat algoritması ise, girdiler arasındaki ilişkiyi dikey olarak kurarak daha hızlı çalışan alternatif bir algoritmadır. Her üç algoritma da veri setlerinin büyüklüğüne ve yapısına bağlı olarak çeşitli durumlarda verimliliği artıp azalabilir [18, 19].

Sonuç olarak, bu çalışmada kullanılan K-Means, Eclat Apriori ve FP-Growth algoritmaları, çapraz satış fırsatlarını belirlemek ve müşteri davranışlarını analiz etmek için önemli araçlar sunmaktadır. Her bir algoritma, kendi avantajları ve sınırlamalarıyla değerlendirilmiş ve analiz amaçlarına göre seçilmiştir.

### 3 METODOLOJİNİN BELİRLENMESİ

Veri bilimi projelerinin planlı bir şekilde yürütülmesini sağlayan KDD, CROSP-DM ve SEMMA gibi metodolojiler bulunmaktadır. Bu tarz metodolojiler özellikle karmaşık veri analizlerinin başarılı bir şekilde yürütülebilmesi için bir izleni sunmaktadır. Çapraz satış analizleri gibi müşteri davranışlarına dayalı öngörülemez bulguların bulunduğu projelerde ise, esnek ve yeniden uyarlanabilir yapısıyla CRISP-DM (Cross-Industry Standard Process for Data Mining) metodolojisi öne çıkmaktadır [20]. CRISP-DM metodolojisi; iş anlayışı, veri anlayışı, veri hazırlığı, modelleme, değerlendirme ve canlıya almak üzere altı aşamadan oluşan, veri madenciliği projelerinde yaygın olarak kullanılan bir süreçtir. Her bir aşaması, çapraz satış analizine özel gereksinimlere göre şekillendirilebilmektedir ve süreç tamamlandığında, elde edilen sonuçlarla süreç yeniden başlatılarak sürekli iyileştirme sağlanabilmektedir. Bu özellikleri sayesinde, CRISP-DM çapraz satış projelerinde ihtiyaç duyulan esnekliği sunmakta ve süreci kontrollü bir şekilde yönetmeye olanak tanımaktadır.

### 3.1 İş Anlayışı

Çapraz satış analizinde, müşteri verileri makine öğrenimi ve veri madenciliği teknikleriyle incelenerek, çapraz satış potansiyeli taşıyan ürün veya hizmetler belirlenir. Bu amaçla, müşteri segmentasyonu ve sınıflandırma için K-Means algoritması kullanılarak benzer satın alım davranışlarına sahip gruplar oluşturulacaktır. Bu algoritmaların performansı; doğruluk, AUC, ROC eğrisi, hata oranı ve MSE gibi metriklerle ölçülecektir. Bunun yanı sıra, Eclat, Apriori ve FP-Growth algoritmalarıyla ürünler arasındaki birliktelik kurallarını analiz ederek birlikte satın alma olasılığı yüksek ürün grupları tespit edilecektir. Destek, güven ve kaldırmaç metrikleri bu analizde önemli rol oynamaktadır. Elde edilen sonuçlar doğrultusunda her müşteri segmenti için özel çapraz satış stratejileri geliştirilecektir. CRISP-DM metodolojisinin aşamaları izlenerek, veri hazırlığı, modelleme ve değerlendirme süreçleri düzenli olarak geliştirilecektir. Bu yaklaşım, müşteri başına düşen satışları artırmaya yönelik kişiselleştirilmiş stratejiler geliştirmeyi sağlar.

### 3.2 Veri Anlayışı

CRISP-DM metodolojisinin ikinci adımında veri setleri incelenip araştırmaya uygun olan veri seti seçilir. Bu araştırmada, üç yılı aşkın bir süreyi kapsayan ve kullanılabilirlik skoru 10/10 olan Kaggle'daki "Retail Data Set" adlı veri setinin "file\_out2.csv" çıktısı kullanılacaktır [21]. Veri setinde; müşteri numarası, ürün kodları, işlem numarası, satış tarihi, indirim miktarı, satılan ürünün adedi ve ürünün fiyatı gibi çeşitli bilgiler yer almaktadır. Veri temizleme ve analiz adımlarıyla, müşteri segmentasyonu ve çapraz satış potansiyeli belirlemek için gerekli olan veri yapısı oluşturulacaktır. Bu analiz, sonraki modelleme süreçleri için sağlam bir temel sağlayacaktır.

### 3.3 Veri Hazırlığı

Veri hazırlığında, veri setinin ön işleme süreçlerinden geçilerek çapraz satış tespiti ve analizi için uygun hale getirilir. Bu süreç, literatür taraması sonucunda birlenen algoritmaların veri setine uygulanmasını kolaylaştırmak için gerçekleştirilir. CRISP-DM metodolojisinin sağladığı esneklik sayesinde, ilerleyen aşamalarda ek veri işleme ihtiyacı doğarsa bu adıma geri dönülerek güncellemeler yapılabilir. Araştırma gereksinimlerinden elde edilen veriler doğrultusunda; veri temizliği, veri entegrasyonu ve gerekli durumlarda özellik mühendisliği uygulanacaktır. İlk adım olarak, eksik ya da tutarsız girdileri tespit edip veri setinden çıkarılması veya uygun yöntemlerle doldurulması sağlanacaktır. Ardından, müşteri numarası, ürün kodları ve işlem numaraları gibi bilgiler analiz sürecine uygun hale getirilecektir. Ardından işlem numarası üzerinden aynı işlemi oluşturan birden fazla ürün gruplandırılarak birlikte satış örüntüleri belirlenecektir. Bu ön işlemler, modelleme aşamasında veri setinin daha etkin bir şekilde kullanılmasını sağlayacaktır. Böylelikle müşteri geçmişi ve ürün özellikleri gibi veriler değerlendirilerek ürün bazlı çapraz satış tespiti yapılacaktır.

### 3.4 Modelleme

CRISP-DM'nin dördüncü aşaması olan modellemede, bir önceki aşamada ön işleme maruz kalmış veri seti ile literatür aşamasında incelenen algoritmalar kullanılarak modeller oluşturulur. İşlenmiş veriler, araştırmada kullanılan platforma yüklenir ve uygun kütüphanelere tanımlanır. Ardından ilgili veri seti, modelin eğitimi ve eğitilen modelin testi için iki ayrı veri setine ayrılır ve ayrı ayrı isimlendirilir. Eğitim için hazırlanan veri seti ile model kurulurken, test için hazırlanan veri seti ile oluşturulan modelin daha önce taramadığı veriler üzerinden çeşitli değerlendirme metrikleri kullanılarak performansı incelenir ve analizleri yapılır. Bu aşamanın uygulanması

sırasında herhangi bir eksiklikte önceki adım olan veri hazırlığı aşamasına geri dönülüp, eksikliğin üstesinden gelinir. Bu sayede modelin en doğru sonuçları üreteceği yapı bir döngü içerisinde hazırlanılır.

### 3.5 Değerlendirme

Değerlendirme aşamasında, oluşturulan modelin canlıya alınmadan önce incelemesi yapılır. Her algoritmanın ve buna bağlı olarak oluşturulan her modelin farklı metrikleri ve farklı değerlendirme süreçleri bulunmaktadır. Farklı algoritmaların farklı değerlendirme metriklerinin sonuçları elde edildikten sonra majör hatalar tespit edilirse metodolojisinin başına dönülüp aşamalar, ilgili hataları göz önünde tutularak tekrar uygulanır. Böylelikle sonuçların birbirleriyle karşılaştırılabildiği bir ortam oluşturmak mümkün hale gelmektedir. Ortaya çıkan sonuçlar çeşitli araçlar ve kütüphanelerle görselleştirilip son bir değerlendirme yapılır.

### 3.6 Canlıya Alma

Canlıya alma aşamasında, geliştirilen model iş süreçlerine entegre edilir. Bu model, perakende sektöründe faaliyet gösteren firmaların veri setinden yararlanarak müşterilerin çapraz satış davranışlarını tespitini ve analizini yapacaktır. Bu tespit ve analizler ile perakende sektöründe faaliyet gösteren firmalar, müşterilere uygun ürünler sunarak çapraz satış fırsatlarını en üst düzeye çıkmasını hedeflenmektedir. Ürün bazlı çapraz satış önerileri sayesinde firmalar, müşteri başına düşen satış hacmini ve kârlılığını arttırabilir.

## 4 ÇAPRAZ SATIŞ MODELİ GELİŞTİRME

Perakende sektöründe çapraz satış fırsatlarını tespit edebilecek modellerin farklı algoritmalarla geliştirme süreçleri CRISP-DM metodolojisiyle ele alınacaktır. Bu bölümde çalışmada yapılan veri hazırlığı, modellerin geliştirilmesi ve geliştirilen modellerin performanslarının değerlendirilmesi yapılmıştır. Problem daha somut anlaşılması için anlatım, çeşitli tablo ve grafiklerle desteklenmiştir.

### 4.1 Veri Seti ve Veri Hazırlığı

Çapraz satış tespiti yapılacak veri seti önceki bölümlerde bahsedilen Kaggle platformunda bulunan “Retail Data Set” isimli veri seti seçilmiştir. 8 sütun ve 29103 satırdan oluşan veri setinin sütunları, isimleri, açıklamaları ve veri tipleriyle beraber Tablo 1’de belirtilmiştir.

Tablo 1: Retail Veri Setinin “file\_out2.csv” Çıktısının Özellikleri

Sütun Adı	Açıklama	Veri Tipi
Unnamed: 0	Satır Numarası	int64
InvoiceID	İşlem Numarası	int64
Date	İşlemin Gerçekleştirildiği Tarih	object
ProductID	Ürün Numarası	int64
TotalSales	İşlemden Elde Edilen Gelir	float64
Discount	İşlemden Uygulanan İndirim	float64
CustomerID	Müşteri Numarası	int64
Quantity	İşlemden Bulunan Ürünün Miktarı	int64

Veri setinin keşfi ve tanımlamaları yapıldıktan sonra CRISP-DM metodolojisine bağlı olarak bu bölümde veri hazırlığı için analizler yapılmıştır. Veri hazırlığı aşamasında, yapılan incelemelerde bazı tutarsızlıklar tespit edilmiştir. Beklenen şartlarda “TotalSales”, “CustomerID”, “InvoiceID” ve “ProductID” adlı sütunların pozitif değerlerden oluşması ve “Discount” adlı sütunun sıfır ya da pozitif değerlerden olması beklenirken toplam 493

girdide aykırı değerler tespit edilmiştir. Bu durum, veri kalitesini etkileyebilecek bir anomali olarak değerlendirilmiştir. Veri setlerinde bulunan anomalilerin üstesinden gelmesinin birden fazla yolu vardır. Bu yollardan bir tanesi ve bu çalışmada kullanılacak olan, anomalilerin hepsini tespit edip veri setinden kaldırmaktır [22]. Anomaliye sahip girdileri veri setinden çıkarıldıktan sonra toplam 28610 satır elde edildi. Böylelikle veri seti, modellemeye hazır hale geldi. Bu aşamadan sonra veri seti, farklı değerlendirme ve çalışma prensiplerine sahip algoritmalara göre ayrı ayrı değerlendirilecektir.

## 4.2 Modelleme

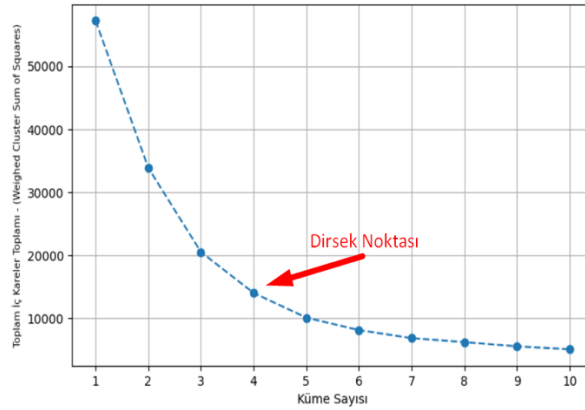
Literatürde kullanılan algoritmalar incelenmiş ve bu incelenen algoritmalarından veri setine uygun olanları seçilmiştir. Müşteri segmentasyonu ile çapraz satış tespitleri için K-Means algoritması, benzer satın alma davranışlarını tespit etmek için birliktelik kurallarını analiz etmek için Eclat, Apriori ve FP-Growth algoritmaları kullanılacaktır. K-Means’de oluşturulacak modelin performansını değerlendirmek için veri setinde bulunan son 6 aydaki işlemleri test, geri kalan işlemleri ise eğitim amacıyla ayırıp, BIC, Silhouette ve WCSS metrikleri karşılaştırılacaktır. Birliktelik kural çıkarımlı algoritmaları ise destek, güven ve kaldıraç metrikleriyle değerlendirilmiştir. Farklı parametrelere sahip algoritmaların modelleri alt başlıklarda tanımlanmış ve hazırlanmıştır. Oluşturulan her model değerlendirme kısmında kendi yöntemleriyle değerlendirme bölümünde değerlendirilecektir.

### 4.2.1 Müşteri Segmentasyonu ile Çapraz Satış Tespitleri ve Analizleri

Bu bölümde, K-Means algoritması ile müşteri segmentasyonu yapılmış ve müşteri gruplarının davranışları incelenerek çapraz satış fırsatlarının belirlenmesi hedeflenmiştir. Bu algoritma, gözetimsiz öğrenme sınıfına giren bir kümeleme algoritmasıdır. Bu nedenle, klasik makine öğrenimi modellerinde olduğu gibi bir eğitim aşamasına ihtiyaç duymaz. Algoritma, veri setindeki tüm girdileri doğrudan analiz ederek önceden belirlenmiş bir sayıya göre kümelere ayırır. Bu süreç, modelin bir hedef değışkene bağımlı olmadan çalışmasını sağlar [23].

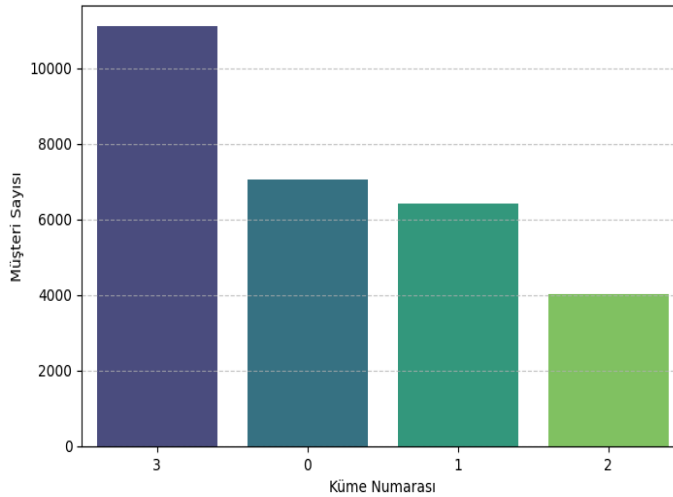
Çapraz satışları daha net görebilmek amacıyla kümeleme için sadece müşteri kimliğini taşıyan “CustomerID” ve ürün kimliğini taşıyan “ProductID” kullanılmıştır. Kümeleme işlemine geçmeden önce optimal küme sayısını tespit etmek gerekmektedir. Optimal küme sayısını belirlemek için çeşitli yöntemler mevcuttur. Bu çalışmada bu sayıyı belirlemek için direkt yöntemi kullanılmıştır [24].





Şekil 1: Dirsek Metodu ile Optimal Küme Sayısı Bulma

Dirsek metodunun sonucu Şekil 1'de belirtilmiştir. 1 ile 10 arasındaki tüm sayılar olası bir küme sayısı olarak değerlendirilmiştir ve kırılım, küme sayısı 4 iken gözlemlenmiştir. Bu yüzden küme sayısı 4 olarak tanımlanmıştır. Optimal küme sayısını belirledikten sonra modelin performansını ve oluşturulan kümelerin kalitesini ölçmek için Davies-Bouildin Endeksi (DBI) kullanılmıştır. DBI, kümelerin sıklığını ve kümeler arasındaki ayrışmayı belirler. DBI değeri 0'a yaklaştıkça seçilen küme sayısının doğruluğu, mutlak doğruya yaklaşır [25]. Küme sayısı dört iken DBI değeri 0,73 olarak hesaplanmıştır. Bu değer, kümelerin birbirinden yeterince ayrıldığını ve her kümenin de aynı zamanda kendi içinde sıkı bir yapı oluşturduğunu göstermektedir. Şekil 2'de küme sayısı dörtken oluşan kümelerin grafiği verilmiştir.



Şekil 2: K-Means Analizinden Sonra Oluşan Kümelerin Dağılımı

Şekil 2'de görüldüğü üzere kümelerdeki müşteri sayısı, birbirlerine yakın. Bu sebeple, Tablo 2'den yola çıkarak her kümenin hacmini, müşteri ve işlem sayısını tespit edilip genel yorumlar yapılabilir.

Tablo 2: Kümeler Bazında Harcama ve Ürün Çeşitliliği Analizi

Küme	Ortalama Toplam Harcama	Ortalama İndirim Miktarı	Toplam Müşteri Sayısı	Satılan Farklı Ürün Miktarı	Fatura Sayısı	Satılan Ürün Sayısı
0	3016	542	203	803	4992	34096
1	2415	434	183	681	4641	40546
2	3306	595	169	600	3261	22642
3	2036	366	235	843	6434	59943

Tablo 2'deki değere göre her küme için belli başlı karakteristik özellikler bulunmaktadır. 0. Küme ortalama toplam harcaması 3046. Bu durum, müşterilerin diğer kümelere göre orta seviyede harcama yapma eğiliminde olduğunu göstermektedir. Satılan toplam farklı ürün sayısının 803 olması, bu grubun ürün çeşitliliğine önem verdiğini ve farklı kategorilerden alışveriş yapmayı tercih ettiğini ortaya koymaktadır. Toplam 34096 ürün satın alarak, diğer kümelere göre daha düşük hacimli bir alışveriş profili sergilemektedir. Satılan ürün sayısının ve satılan toplam farklı ürün sayısının da fazla olması çapraz satış davranışlarının diğer gruplara göre daha az olabileceği düşünülebilir. Küme 1'de ise satılan farklı ürün sayısı, Küme 0'a göre daha az, toplam satılan ürün sayısı ise daha fazladır, bu da Küme 1'in çapraz satış davranışlarının Küme 0'a göre daha belirgin ve daha yoğun olduğu düşünülebilir. Küme 2, toplam müşteri sayısı 169 ile en düşük müşteri grubunu temsil etmekle birlikte, ortalama toplam harcama olarak 3306 ile en yüksek harcama düzeyine sahip olan bir gruptur. Bu kümenin en belirgin özelliği, daha az çeşit ile yüksek değerli ürünlere odaklandığını ortaya koymaktadır. Bu tür kümelerdeki çapraz satışlar, diğer kümelere göre daha benzersiz olabilir. Küme 3, toplam müşteri sayısı 235 ile en geniş müşteri grubunu temsil etmektedir. Bu müşterilerin ortalama toplam harcaması 2036 ile en düşük seviyededir. Ayrıca, bu kümede toplamda 843 farklı ürün ve 59943 adet ürün satılmıştır, bu da ürün çeşitliliği ve hacmi açısından oldukça aktif bir alışveriş profili sergilemektedir. Satılan toplam fatura sayısının 6434 olması, müşterilerin sık sık alışveriş yaptığını ve geniş bir ürün yelpazesine yöneldiğini göstermektedir. Özellikle müşteri ve satış hacminin bu denli yüksek olması, çapraz satışlar davranışlarının sıklıkla görülmesi beklenebilir.

Tablo 3: Kümelerde En Çok Satılan 5 Ürün

Küme	Ürün Kodu
0	[1756, 1830, 1158, 1128, 1347]
1	[192, 540, 543, 527, 885]
2	[1756, 1830, 1458, 1347, 1128]
3	[192, 543, 885, 540, 251]

Tablo 3'teki bilgiler göz önünde bulundurulduğunda her kümelere farklı ürün gruplarının ön plana çıktığı görülüyor. 1756, 1830, 1128 ve 1347 kodlu ürünler, Küme 0'da da Küme 2'de de en fazla satın alınan ürünlerdir. Benzer şekilde 192, 540 ve 543 kodlu ürünler ise, Küme 1 ve Küme 3'te en fazla satın alınan ürünlerdir. Bu tablodan yola çıkarak Küme 0 ve Küme 2 ile Küme 3 ve Küme 1'in benzer satın alma davranışlarına sahip olduğu düşünülebilir.

Kümelerdeki çapraz satışları tespit etmek için her kümede en çok satın alınan ürünler ve bu ürünlerin birlikte satın alınma sıklıkları analiz edildi. Tablo 4'te kümedeki müşterilerinin geçmiş satın alma verileri incelenerek, bir ürün alındığında sıklıkla birlikte satın alınan ürün çiftleri belirlenip gösterilmiştir.

Tablo 4: Kümelerdeki Çapraz Satış Özelliği Taşıyan Bazı Ürün Çiftleri

Küme	1. Ürün	2. Ürün
0	1128	1158
	1128	1167
	1128	1241
	1167	1241
	1341	1756
	1436	1756
	1458	1462
	1756	1830
	1756	1857
	1756	1894
1	189	192
	192	221
	521	525
	521	527
	525	527
	527	531
	531	532
	540	543
	543	545
	613	617
2	1128	1158
	1128	1167
	1128	1241
	1167	1241
	1347	1724
	1456	1462
	1458	1462
	1498	1830
	1502	1830
	1638	1830
3	98	192
	164	192
	189	192
	192	218
	192	221
	192	243
	192	247
	192	248
	192	251
	192	252

Tablo 4'te yer alan ürün çiftleri, belirli kümeler içinde birlikte satın alma eğilimi gösteren ürünleri ortaya koymaktadır. Her bir küme, benzer alışveriş davranışlarına veya ürün tercih kalıplarına sahip müşteri gruplarını temsil etmektedir. Küme 0, belirli ürünler etrafında yoğunlaşan çapraz satış eğilimleri ile dikkat çekmektedir. Özellikle 1128 numaralı ürünün, 1158, 1167 ve 1241 ile sık eşleşmesi, bu kümede 1128'in merkezi bir konuma sahip olduğunu göstermektedir. Benzer şekilde 1756 kodlu ürünün de farklı ürünlerle birlikte sıkça görünmesi, 1756 numaralı ürünün etrafında da önemli çapraz satış potansiyeli oluşturduğunu göstermektedir. Bu kümedeki

kritik birkaç ürün, diğer ürünlerle tamamlayıcı olduğunu göstermektedir. Küme 1 diğer kümelerle kıyasla daha dar bir ürünle çapraz satış davranışları yansıtmaktadır. 521, 525, 527, 531 ve 532 gibi ürünler aynı kümede sıklıkla birlikte geçmektedir. Bu kümeye dahil olan ürün çiftleri, muhtemelen benzer işlevsellik veya tamamlayıcı niteliklere sahip kısıtlı bir ürün grubu içinde çapraz satış fırsatlarının daha belirgin olduğunu öne çıkarmaktadır. Küme 2, Küme 0'daki benzer örüntüleri andırmaktadır fakat aynı zamanda farklı ürünler de bulundurarak daha geniş bir yelpaze sunmaktadır. 1128, 1158, 1167 ve 1241 gibi ürünlerin tekrarı, bu ürünlerin satışlarda kritik rol bir oynadığını göstermektedir. Bu küme, benzer ürün kombinasyonlarının farklı kümelerde de var olabileceğini ancak bu küme özelinde farklı ek ürünlerle de birlikte değerlendirilebileceğini ortaya koymaktadır. Küme 3, 192 numaralı ürünün merkezde olduğunu ve son derece geniş bir ürün yelpazesine çapraz satış ilişkisi kurduğunu göstermektedir. Bu küme, müşterilerin 192 kodlu ürünü satın aldıklarında pek çok farklı tamamlayıcı ürüne yönelebileceklerini göstermektedir.

K-Means ile yapılan analizler, çapraz satış fırsatlarını müşterileri kümelendirerek tespit eder. Her müşteri kümesine uygun çapraz satış fırsatların ve özel stratejilerin geliştirmesi adına önemli bir temel oluşturmaktadır. Ancak K-Means algoritması gözetimsiz bir algoritma olduğu için modelin yalnızca mevcut verileri kümelerle ayırmasına olanak tanır ve bu yüzden çapraz satışa sahip olmasa da çapraz satış potansiyeli taşıyan ürünlerin tespiti konusunda sınırlı kalabilir. Bu eksiklik, müşteri davranışlarının daha karmaşık ve doğrusal olmayan ilişkilerle modellenmesini zorlaştırır. Bunun önüne geçmek için farklı algoritmalar ve farklı metriklerin incelenmesi gerekebilir. Birlikte kural çıkarıma sahip algoritmalarla çapraz satış performansları, kümeleme algoritmalarından farklı ve daha başarılı performans gösterebilir. Böylece, çapraz satış tespitleri K-Means'in eksik kaldığı alanlarda daha derinlemesine ve etkili bir şekilde gerçekleştirilebilir.

#### 4.2.2 Apriori Algoritması ile Çapraz Satış Tespiti ve Analizi

Bu bölümde, Apriori algoritmasının perakende sektöründe çapraz satış tespitine yönelik uygulaması ve elde edilen sonuçların analizi sunulmaktadır. Bu analiz için Python programlama dili ve mlxtend kütüphanesi kullanılacaktır [26]. Bu kütüphanenin sağladığı Apriori implementasyonu ve bu implementasyonun çeşitli özellikleri kullanılarak ürünler arasındaki birliktelik kuralları tespit edilecektir. Bu yöntem, Apriori algoritmasının uygulamasını kolaylaştırırken elde edilen sonuçların detaylı bir şekilde analiz edilmesine de olanak tanır.

Veri ön işleme aşamasında, müşteri davranışlarını anlamak ve çapraz satış fırsatlarını belirlemek amacıyla ProductID değerleri, her müşteri için işlem bazında liste haline getirilmiştir. Satılan ürünlerin liste haline getirilmesi, algoritmanın birliktelik kurallarını oluşturabilmesi için gereklidir. Bu yapı ürün kombinasyonlarının destek değerlerini hesaplamayı, sık görülen ürün kümelerini belirlemeyi ve bu kümeler üzerinden güven ve kaldırma gibi metriklerle anlamlı birliktelik kuralları oluşturmaya olanak tanır. Listeleme algoritmanın ürünler arasındaki birliktelik kurallarını verimli bir şekilde analiz etmesini ve daha optimize sonuçlar üretmesini sağlar.

Tablo 5: Bazı CustomerID'ye Göre Listelenmiş Bazı Girdiler

CustomerID	ProductID
29	[1436, 1785, 1785, 1892, 1785, 151, 131, 1785, 1785, 1654, 859, 700, 140]
34	[225, 882, 1785, 1786, 107, 115, 1432, 1283, 1654, 1662, 569, 725, 1833, 1833]
239	[360, 464, 941, 1904, 464, 464, 1588, 20, 20, 464, 1830, 1609, 1609]
372	[1911, 1751, 54, 320, 362, 618, 1128, 1310, 1444, 1644, 885, 1727, 49]
477	[1246, 221, 217, 1679, 1894]

Apriori algoritması, çalışmada kullanılan veri setinde çapraz satış potansiyelini belirlemek amacıyla kullanılmıştır. Algoritmanın uygulanmasında kullanılan eşik değerleri, güçlü çapraz satış davranışlarının tespiti amacıyla belirlenmiştir. Destek değeri için %0,5, güven değeri için %80 ve kaldıraç değeri için 1.0 eşikleri belirlenmiştir. Bu eşik değerleri yalnızca sık satın alma davranışlarını yansıtan güçlü birliktelik kurallarını seçmek için kullanılmıştır.

Elde edilen değerler, belirtilen metrikler dışında “antecedents” (öncüler) ve “consequents” (çıktılar) ölçütlerine göre sınıflandırılmıştır. Öncüler, sütununda bulunan ürünler, bir alışveriş sırasında hangi ürünlerin birlikte nasıl satın alındığını belirtirken, çıktılar sütununda bulunan ürünler, bu ilişki sonucunda ortaya çıkan ve birlikte alınan ürünleri temsil etmektedir. Destek metriği, öncülerin ve çıktılarının birlikte hangi sıklıkta satın alındığını ifade ederken güven metriği ise, öncü ürünlerin yer aldığı siparişlerde, çıktı olarak tanımlanan ürünlerin satın alındığı olasılığını açıklamaktadır. Kaldıraç değeri, öncü ürünlerin alınmasının çıktı ürünlerinin alınma olasılığını, alınmama durumuna kıyasla ne ölçüde artırdığını göstermektedir.

Tablo 6: Apriori Algoritmasıyla Oluşturulan Kuralların Bazı Sonuçları

Öncüler	Çıktılar	Destek	Güven	Kaldıraç
[885, 1158, 486]	[1128]	0.05	0.833	5.2083
[1830, 1158, 543]	[1128]	0.06	0.9375	5.8593
[882, 1830]	[1128, 885]	0.06	0.8108	5.8754
[1241, 1167]	[1128, 885]	0.08	0.8695	6.3011
[192, 1830, 1167, 885, 540]	[1128]	0.05	0.9259	0.8744

Belirtilen eşiklerde Apriori algoritması ile toplam 36796 adet çıkarım yapılmıştır. Tablo 6’da da örnek olarak belirtilen 1128 numaralı ürün, birçok satışta görülmektedir. Bu ürünün performansı, K-Means’te yapılan kümelemelerde de Küme 0 ve Küme 2’de öne çıkmaktadır. Küme 0’da 1128 kodlu ürün, daha geniş bir ürün çeşitliliği ve daha düşük hacimli alışveriş profiline sahip müşteriler için önemli bir tamamlayıcı ürün olmanın yanı sıra aynı zamanda Küme 2’de yüksek harcama yapan ve daha az çeşitli ürün tercih eden müşteri gruplarında öncelikli ürünlerden biri olarak tanımlanmaktadır. Bu sebeple 1128 kodlu ürün hem düşük hem de yüksek hacimli harcama yapan gruplar için stratejik bir değer taşımakta ve farklı segmentlerde farklı çapraz satış fırsatları sunmaktadır.

Apriori algoritmasından elde edilen sonuçlar, çapraz satış analizlerinde güçlü bir temel sunmaktadır ancak algoritmanın işlem süresi ve büyük veri setlerindeki yüksek hesaplama maliyetleri gibi sınırlamalar, daha verimli bir yöntem gerektirir. Bu sebeple benzer süreçler, literatür aşamasında da bahsedilen FP-Growth ve Eclat algoritmalarında da yapılabilmektedir. Bu sebeple benzer süreçler, literatür aşamasında da bahsedilen FP-Growth ve Eclat algoritmalarında da yapılabilmektedir.

#### 4.2.3 FP-Growth Algoritması ile Çapraz Satış Tespiti ve Analizi

Bu bölümde FP-Growth algoritması kullanılarak perakende sektöründe çapraz satış tespiti yapılması ve bu sürecin analizi, Apriori algoritmasının analiziyle karşılaştırılarak sunulmaktadır. Apriori algoritmasına alternatif olarak sık öge kümelerini daha hızlı ve aktif bir şekilde belirlemek için geliştirilmiştir. Literatür aşamasında da bahsedildiği gibi Apriori algoritmasının yüksek hesaplama maliyetlerine sahip olmasına karşın, FP-Growth algoritması, sık öge kümelerini oluşturmak için FP-Tree adı verilen özel bir veri yapısını kullanarak etkili bir çözüm sunar. Bu algoritmanın uygulanması için yine Python dili ve mlxtend kütüphanesi kullanılacaktır.

FP-Growth algoritmasını uygulamak için 4.2.2’de oluşturulan ve Tablo 5’te görselleştirilen, ürünlerin müşteri bazlı liste formatına dönüştürülmüş veri seti kullanılmıştır. Bu yapı, Apriori algoritmasıyla benzer şekilde

algoritmanın birliktelik kurallarını belirleyebilmesi için gereklidir. FP-Growth ile oluşturulan model, bu veri setinin tüm olası kombinasyonunu hesaplamak yerine çapraz satış yapan örüntülere odaklanarak işlem süresini düşürmüştür. İşlem süresi, Apriori'ye daha kısa olduğu için belirlenen eşik değerleri daha geniş seçilmiştir, bu sayede farklı örüntüler de tespit edilebilir. FP-Growth'ta kullanılacak modelde minimum eşik değeri 0.4, güven değeri minimum 0.6 ve kaldıraç değeri minimum 1,0 olarak belirlenmiştir. Kurallarda kullanılan eşik değerlerin daha geniş olması, az ama etkisi fazla olan çapraz satışları tespit etmek için etkili olabilmektedir.

Tablo 7: FP-Growth Algoritmasıyla Oluşturulan Kuralların Bazı Sonuçları

Öncüler	Çıktılar	Destek	Güven	Kaldıraç
[521, 531, 1919, 527]	[192]	0.05	0.8928	3.9160
[1919, 527, 165, 543]	[192]	0.05	0.9259	4.0610
[1756, 189, 246]	[192]	0.05	1.0000	4.3859
[249, 543]	[192]	0.06	0.9090	3.9872
[531, 540]	[192, 543]	0.07	0.8139	5.5750

FP-Growth ile ilgili kurallardan toplam 42987 çıkarım yapılmıştır. Tablo 7'de 192 numaralı ürünün çıktı olduğu bazı durumlarda oluşan kurallar görüntülenmektedir. Örnek olarak kullanılan 192 numaralı ürünün performansı, K-Means ile yapılan kümelerde de benzer sonuçlar vermekle birlikte kümelemede görünmeyen ancak FP-Growth ile alınan sonuçta olan farklı ürün çiftleri tespit edilmiştir, bu da FP-Growth algoritmasının, verimliliği sayesinde farklı ürün çiftleri tespit edebileceğini gösteriyor. Bunun yanı sıra, güven değerinin %81 ve %100 arasında olması, müşterilerin öncülerden birinin satın alındığı sırada, çok yüksek olasılıkla çıktı olan 192 numaralı ürün alabileceğini gösteriyor. Kaldıraç değerlerinin de 1'den yüksek olması, rastgelelikten çok, anlamlı çapraz satış olduğunu gösteriyor. İlgili tablodaki sonuçlarda, toplam işlemler ile ilgili ürün çiftlerinin satın alınma oranı olan destek metriği ise %7 ile %5 arasında, bu da çapraz satışların sıkı bir ilişkiye sahip olabileceğini göstermektedir.

Tablo 6'daki gibi Apriori ile yapılan eşleştirmelerin pek çoğu FP-Growth'la da benzer sonuçlarla görülmektedir. FP-Growth algoritmasının hesaplama verimliliği açısından daha avantajlı ve eşik sınırlarının daha geniş olması, bu veri setlerinde daha geniş çapraz satış örüntülerinin tespit edilmesini sağlamıştır ancak daha geniş satış örüntüleri daha güçsüz, az rastlanan ya da anlamsız, tesadüfen oluşan ürün çiftlerini de listelemeye potansiyelidir. FP-Growth, özellikle yüksek harcama yapan müşterilerden oluşan kümelerde tespit ettiği kurallar, ürün çeşitliliği açısından daha stratejik çapraz satış fırsatları sunmaktadır. Apriori ise, yapısı gereği tüm kombinasyonları hesaplayabildiği için benzer kümelerin öne çıktığı ortak ürün ile kümeler arasındaki etkileşimi ve çeşitliliği arttıracak kurallar sunma potansiyeline sahiptir.

Bir sonraki bölümde bir diğer birliktelik kural çıkarımı algoritması olan Eclat algoritması, FP-Growth ve Apriori algoritmaları ile karşılaştırmalı incelenecektir. Özellikle algoritmaların işlem süresi, bellek kullanımı ve tespit edilen kuralların çeşitliliği gibi metrikler üzerinden yapılan analizler, Eclat'ın avantaj ve dezavantajlarını daha net bir şekilde belirleyecektir. Bu karşılaştırma, Apriori'nin detaylı fakat yüksek işlem maliyetine sahip yapısıyla, FP-Growth'un verimlilik odaklı yaklaşımı arasında, Eclat'ın nasıl bir denge sağladığını anlamaya yardımcı olacaktır.

#### 4.2.4 Eclat Algoritması ile Çapraz Satış Tespiti ve Analizi

Eclat algoritması, FP-Growth ve Apriori algoritmaları gibi birliktelik kural çıkarımı yöntemiyle çalışan bir algoritma olmasına rağmen veri setini derin öncelikli arama ile dikey düzeyde işlediği için bu iki algoritmadan farklıdır. Bu işleme sayesinde, özellikle işlem sayısının ürün sayısından fazla olduğu veri setlerinde bellek tüketimini ve işlem süresini kısaltabilir. Bunun yanı sıra, Apriori'nin oluşturduğu yoğun kombinasyonlar veya FP-Growth'un kendine

has FP-Tree yapısına kıyasla daha az hesaplama gerektirebilir. Ancak Eclat'ta uygulanan eşik değerler diğer iki algoritmaya göre biraz daha yüksek olmalı çünkü Eclat, düşük eşiklerde çok fazla ve anlamsız ürün kombinasyonu üretip sonuçların anlamlılığını azaltabilir, bu yüzden Eclat ile kurulan modelin eşikleri, diğer iki algoritmanın eşiklerine göre biraz daha sınırlı tayin edilecektir. Bir diğer sınırlama ise, her ürün için anlamlı sonuçlar oluşturacak kombinasyonlar ya da minimum işlem sayısının veri setinde olmaması. Bu senaryonun yaşandığı durumlarda veri setini belli ürünler ya da fişler için sınırlandırmalar yapılabilir. Model, Python programlama dili ve pyECLAT kütüphanesi [16] kullanılarak yapılacaktır.

Apriori ve FP-Growth algoritmalarıyla kurulan modellerde çapraz satışları tespit etmek için kullanılan, her müşteri bazında liste halinde tutulan ürün kodları Eclat algoritmasında derin öncelikli arama yaptığı için tutulamaz. Veri önileme aşamasında, Eclat algoritmasına uygun bir veri hazırlamak için ProductID'deki bilgiler, dikey veri yapısına dönüştürerek bir matris oluşturulmuştur. Bu yapı aracılığıyla, ürünlerin kesişim kümeleri üzerinden sık öge kümelerinin tespit edilmesine ve bu kümelerden anlamlı birliktelik kuralları çıkarılmasına olanak tanımaktadır.

Tablo 8: ProductID'lerin Dikey Veri Yapısına Dönüştürülmüş Hali

0	1	2	3	4	5	6	7	8	9	...
1	1756	543	1793	151	152	673	1417	NaN	NaN	...
2	1502	521	540	843	192	219	224	164	59	...
3	524	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
4	580	559	189	669	679	252	221	924	199	...
5	900	1321	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...

Tablo 8'de yer alan yapıda satırlar, ilk satır hariç, bir müşterinin satın aldığı ürünleri, ilk sütun ise müşterileri temsil eder. Örneğin tablodaki 5. müşteri, 900 ve 1321 kodlu ürünü alırken, 3. müşteri sadece 524 numaralı ürünü almıştır. Bu yapı sayesinde, Eclat algoritmasında çapraz satış fırsatlarını daha etkili bir şekilde analiz etmek ve satış stratejilerini optimize etmek mümkün hale gelmiştir.

Eclat algoritmasının çalışma prensibini, ürünlerin birlikte satın alınma örüntülerini analiz etmek için kesişim kümelerine dayanır. Tablo 8'deki gibi dikey veri yapısında kesişim işlemlerini daha hızlı ve etkili hale getirmek için değerleri ikili forma dönüştürmek, ürünler arasındaki ilişkileri daha kolay ve anlaşılır analiz etme ve işlem süresini optimize etmek için kullanılabilir. Dikey veri yapısında her bir ürünün hangi işlemlerde yer aldığını matrisle listelemek yerine ikili bir matrisle her müşteri ve ürün için 1 veya 0 değerleriyle ürünün müşteri sepetinde olup olmadığı belirtilir. Tablo 8'deki veri seti ikili sisteme uyarlandığında satırlarda ürünlerin kimlikleri, sütunlarda ise müşterileri temsil eder. Eğer bir ürün, bir müşterinin işleminde bulunuyorsa karşılık gelen hücreye bir değeri atanırken aksi durumlarda ise sıfır değeri atanır.

Tablo 9: Dikey Veri Yapısının İkili Sisteme Örnek Entegrasyonu

	566	1280	578	1718	301	1228	50	1089	894	...
11	0	0	0	0	0	0	0	0	0	...
12	0	0	0	0	0	0	0	0	0	...
13	0	0	0	0	0	0	0	0	0	...
14	0	0	1	0	1	0	0	1	0	...
15	0	0	0	0	0	0	0	0	0	...
16	0	0	0	1	1	0	0	0	0	...
17	0	0	0	0	0	0	0	0	0	...

	566	1280	578	1718	301	1228	50	1089	894	...
18	0	0	0	0	0	0	0	0	0	...
19	0	0	0	0	0	0	0	0	1	...
20	0	0	0	0	0	0	0	0	0	...
21	1	0	0	0	0	0	0	0	0	...
22	0	0	0	0	0	0	0	0	0	...
23	0	0	0	0	0	0	0	0	0	...

Tablo 9, Tablo 8'in ikilik sisteme çevrilmiş halinin bir kısmını içeriyor. İlk satırlar, ürünlerin kimliklerini içerirken, ilk sütunlar müşterilerin kimliklerini içeriyor. Müşteri kimliklerin ve ürün kimliklerin kesiştiği hücrenin değeri bire eşit ise, müşteri ilgili ürünü sepetine eklediğini, sıfır ise müşteri ilgili ürünü sepetine ekmediğini gösteriyor. Örneğin Tablo 9'daki verilere göre, 16 numaralı müşteri 1718 ve 301 numaralı ürünleri satın almıştır.

Veri seti, Eclat modellemesi için uygun hale gelmiştir. Modeli oluştururken, minimum destek değeri %1, minimum ürün kombinasyonu 1 ve maksimum kombinasyonu 2 olarak belirlenmiştir. Bu sınırlamalar ve özelliklerin kurulan modelde toplam 24456 işlemden 3561'i tekil, 4863'ü ikili toplam 8424 adet geçerli kombinasyon tespit edilmiştir.

Tablo 10: Eclat Algoritmasıyla Oluşturulan Modelin Bazı Sonuçları

Kombinasyon	Destek	Güven	Kaldıraç
[1462, 1458]	0.44	0.91	3.3162
[1919]	0.17	0.90	4.0009
[1919, 540]	0.14	0.90	4,7771
[1347, 486]	0.21	0.94	2,2415

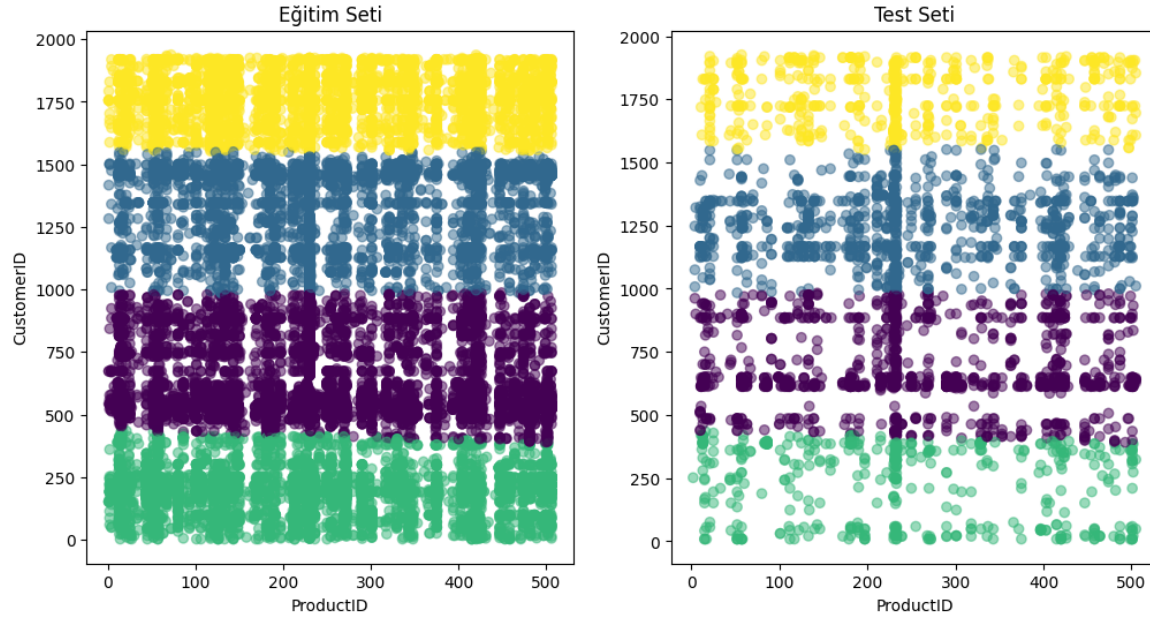
Tablo 10'da modelin bazı sonuçları gösterilmektedir. Eclat algoritmasıyla oluşturulan model, diğer iki modele göre daha farklı ve daha tutarsız sonuçlar verdiği gözlemlenmiştir. Ancak buna rağmen, yüksek destek ve güven değerlerine sahip kombinasyonlar ile çapraz satış ilişkilerini tespit etmekte ve müşteri davranışlarını daha iyi anlamak için değerli bilgiler sunabilmektedir.

### 4.3 Değerlendirme

Bu çalışmada, yaklaşık üç yıllık perakende satış verilerinden oluşan veri setinin analizi ve çapraz satış fırsatlarını tespit etmek amaçlanmıştır. Bu doğrultuda, K-Means, Apriori, FP-Growth ve Eclat algoritmalarından modeller oluşturulup performansları karşılaştırılmıştır. Modellerin sonuçları, farklı metrik ve performans göstergeleriyle karşılaştırılarak değerlendirilmiştir.

K-Means kümeleme algoritmasıyla veri seti, müşteri numaraları ve ürün numaralarına göre dört kümeye ayrılmıştır. Oluşturulan modelin performansını değerlendirmek amacıyla veri seti belirli bir zaman dilimine göre eğitim ve test setlerine ayrılmıştır. Bu bağlamda, veri setindeki son 6 aydaki işlemler test seti olarak kullanılırken, geri kalan işlemler ise eğitim seti olarak ayrılmıştır. Böylelikle eğitim setinde toplam 25,286 adet işlem bulunurken test setinde toplam 3,324 adet işlem tespit edilmiştir. Bu veri setlerinin Silhouette, WCSS ve BIC metrikleri ölçülüp karşılaştırılacaktır.





Şekil 3: Kümelenmiş Eğitim ve Test Setinin Dağılımı

Eğitim sürecinde veri seti üzerinde 4 küme oluşturulmuştur. Algoritmanın her bir adımdaki kümeler arası farkı en üst düzeye çıkaracak şekilde küme merkezleri optimize edilmiştir. Test seti üzerinde, eğitim sırasında elde edilen küme merkezleri sabit tutularak yeni verilerin bu merkezlere atanması sağlanmıştır.

Tablo 11: Test Setinin ve Eğitim Setinin Metrik Sonuçları

Veri Seti Çeşidi	BIC Skoru	Silhouette Skoru	WCSS Skoru
Test Seti	-864060517	0.43	856.66
Eğitim Seti	-145849370	0.47	8640605.86

Tablo 11'deki değerlere göre eğitim setinin performans metrikleri test metriklerine göre daha dengeli gözükmemektedir. Ancak test setinde belirgin bir performans düşüşü gözükmemektedir, bu da modelin aşırı uyum göstermediğini ve ideal düzeyde genelleme yapılabileceğini göstermektedir. BIC skorunun test setine göre daha yüksek olması, modelin test verisi üzerinde kısmen daha düşük bir uyum sağladığını fakat büyük bir bozulma olmadığını göstermektedir. Silhouette skoru test setinde 0.43 ile kabul edilebilir bir değerde kalmaktadır, bu ayrılan kümelerin belirgin özelliklere sahip olduğunu ifade etmektedir. Her iki veri seti farklı miktarlarda işlem barındırdıkları için WCSS'de Tablo 11'deki gibi bir fark oluşması beklenen bir sonuçtur. Ancak modelin analizlerinde herhangi bir anomali tespit edilirse bu skoru iyileştirecek düzenlemeler yapılabilir.

Bu çalışmada K-Means kümeleme algoritması dışında aynı zamanda birliktelik kural çıkarımı mantıyla çalışan Apriori, Eclat ve FP-Growth algoritmaları da kullanılmıştır. Bu algoritmalarla oluşturulan modellerde hesaplama maliyetini uygulanabilir kılmak amacıyla kendilerine özgü sınırlamalar uygulanmıştır.

Tablo 12: Algoritmalarla Göre Uygulanan Sınırlamalar

Algoritma Adı	Minimum Destek Değeri	Minimum Güven Değeri	Minimum Kaldıraç Değeri	Oluşan Kural Sayısı
Apriori	%0,5	%80	1.0	36.796
FP-Growth	%0,4	%60	1.0	42.987
Eclat	%0,1	%90	2.1	8.424

Apriori algoritması, birliktelik kuralları çıkarımı yaparak çalışmada kullanılan veri setinde 36.796 adet çıkarım yapmıştır. Bu çıkarımlar arasında, 1128 numaralı ürünün farklı müşteri gruplarında öne çıkması, algoritmanın açığa çıkardığı çapraz satış fırsatlarından birisidir.

FP-Growth algoritması, yalnızca Apriori'ye kıyasla daha hızlı sonuçlar ve işlem süresi üretmekle kalmamıştır aynı zamanda daha geniş değerlerle daha fazla kural çıkarımı yapmıştır. FP-Growth'un FP-Tree yapısı sayesinde sık görülen ürün kombinasyonları büyük veri setlerinde daha az maliyetli tespit edilmiştir. Elde edilen sonuçlar, Apriori ile büyük ölçekte uyum göstermiş olup daha geniş eşik değerleri sayesinde ek çapraz satış fırsatları tespit edebilmiştir. 192 numaralı ürünün farklı ürünlerle kombinasyonları kaldıraç değerleriyle anlamlı çapraz satış ilişkileri sunmuştur.

Eclat algoritması ise, dikey veri işleme ile derin öncelikli arama yöntemini kullanarak sık öge kümelerini belirlemiştir. Ancak düşük eşik değerlerinde anlamsız ürün anlamsız ürün kombinasyonu üretme eğilimi metriklerde sınırlamaya sebep olmuştur. Bu sebeple güven değeri %90 olarak tayin edilirken sadece 8.424 kural çıkarımı yapılmıştır. Çıkarılan kurallar, diğer algoritmalarla göre düşük olmasına karşın güven değerinin yüksek olması sebebiyle biraz daha gerçeğe yakın olabilir. Örneğin 1462 ve 1458 numaralı ürünler arasında kaldıraç değeri 3,31 olarak hesaplanmış olup bu kombinasyonun sık çapraz satış ilişkisi sunduğu belirlenmiştir.

## 5 SONUÇ

Perakende sektöründe müşteri davranışlarına dayalı çapraz satış fırsatlarının tespiti ve bu fırsatların ilgili işletmelere nasıl katma değer sunabileceği üzerine detaylı bir analiz gerçekleştirilmiştir. Çalışmada kullanılan veri seti, müşteri segmentasyonu ve ürünler arasındaki ilişki analizleri uygun çeşitle hazırlanarak çeşitli algoritmalarla modellenmiştir. Oluşturulan modellerin sağladığı bulgular vesilesiyle veri bilimi ekiplerinin çapraz satış tespiti için çeşitli veri madenciliği tekniklerinden yararlanabileceği vurgulanmaktadır. Fakat manuel olarak programlama dilleri ve algoritmalar kullanarak yapılan işlemlerin zaman ve işlem maliyetini azaltmak amacıyla yapay zekâ temelli uygulamaların önemi günden güne artmaktadır. Bu noktada Optiwisdom tarafından hazırlanan Kolay.ai gibi yapay zekâ çözümleri, çapraz satış analizi süreçlerini optimize edebilecek bir araç olarak öne çıkmaktadır [27]. Kolay.ai, otomatik veri analizi ve çapraz satış tespiti yapabilen gelişmiş bir yapay zekâ destekli bir platformdur. Bu platform aracılığıyla belli formatlı veri setlerinde; ürün performans analizi, müşteri segmentasyonu, ürünler arasındaki çapraz satış fırsatlarının tespiti gibi süreçleri hızlı ve kolay bir şekilde yapılabilmektedir. Böylece işletmeler, müşteri başına düşen ortalama harcamayı arttırıp daha etkili pazarlama stratejileri geliştirebilirler.

Sonuç itibarıyla, çapraz satış analizleri perakende sektöründe önemli bir stratejik araç olarak kullanılabilir. Bu analizlerin etkin şekilde uygulanması sadece satışları arttırmakla kalmaz aynı zamanda müşteri deneyimlerini de geliştirerek uzun vadeli müşteri yaşamı boyu değerini arttırabilir. Bu tür analizlerin yapay zekâ destekli dinamik öneri sistemlerine entegre edilmesi ve Kolay.ai gibi platformlarla gerçek zamanlı çalışması, özellikle KOBİ statüsündeki işletmelerin daha hızlı ve etkili kararlar almasına imkân sağlayacaktır.

## 6 REFERANSLAR

- [1] Kamakura, W. A. 2014. "Cross-selling: Offering the right product to the right customer at the right time." Profit Maximization through Customer Relationship Marketing, 41-58, Routledge.
- [2] Kamakura, W.A., 2012. "Sequential market basket analysis." Marketing Letters, 23, 505-516.
- [3] R. Yörük ve R. Eşmekaya, "Customer Lifetime Value (Müşteri Yaşamboyu Değeri)", YBS Ansiklopedi, cilt 5, no. 2, pp. 26-40, 2018.
- [4] Tariq Mahmood and Francesco Ricci. 2009. Improving recommender systems with adaptive conversational strategies. In Proceedings of the 20th ACM conference on Hypertext and hypermedia (HT '09). Association for Computing Machinery, New York, NY, USA, 73–82.
- [5] Sarwar, Badrul, George Karypis, Joseph Konstan, and John Riedl. "Analysis of recommendation algorithms for e-commerce." In Proceedings of the 2nd ACM Conference on Electronic Commerce, pp. 158-167. 2000.
- [6] Chen, H., Chiang, R.H. and Storey, V.C., 2012. Business intelligence and analytics: From big data to big impact. MIS quarterly, pp.1165-1188.
- [7] Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics, 20, 53-65.
- [8] Edwards, A. W., & Cavalli-Sforza, L. L. (1965). A method for cluster analysis. *Biometrics*, 362-375.
- [9] Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 461-464.
- [10] Neath, A. A., & Cavanaugh, J. E. (2012). The Bayesian information criterion: background, derivation, and applications. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(2), 199-203.
- [11] Ali, Özden Gür, Yalçın Akçay, Serdar Sayman, Emrah Yılmaz, and M. Hamdi Özçelik. "Cross-selling investment products with a Win-win perspective in portfolio optimization." *Operations Research* 65, no. 1 (2017): 55-74.
- [12] Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of clinical epidemiology*, 49(11), 1225-1231.
- [13] Lau, Kin-Nam, Haily Chow, and Connie Liu. "A database approach to cross selling in the banking industry: Practices, strategies and challenges." *Journal of Database Marketing & Customer Strategy Management* 11 (2004): 216-234.
- [14] Jarrar, Yasar F., and Andy Neely. "Cross-selling in the financial sector: customer profitability is key." *Journal of Targeting, Measurement and Analysis for Marketing* 10 (2002): 282-296.
- [15] Salazar, M. T., Harrison, T., & Ansell, J. (2007). An approach for the identification of cross-sell and up-sell opportunities using a financial services customer database. *Journal of Financial Services Marketing*, 12, 115-131.
- [16] Richard, J. pyECLAT <https://github.com/jeffrichardchemistry/pyECLAT>
- [17] Kumar, V. S., Renganathan, R., VijayaBanu, C., & Ramya, I. (2018). Consumer buying pattern analysis using apriori association rule. *Int. J. Pure Appl. Math*, 119(7), 2341-2349.
- [18] Heaton, J. (2016, March). Comparing dataset characteristics that favor the Apriori, Eclat or FP-Growth frequent itemset mining algorithms. In *SoutheastCon 2016* (pp. 1-7). IEEE.
- [19] Hipp, J., Güntzer, U., & Nakhaeizadeh, G. (2000). Algorithms for association rule mining—a general survey and comparison. *ACM sigkdd explorations newsletter*, 2(1), 58-64.
- [20] Azevedo, A., & Santos, M. F. (2008). KDD, SEMMA and CRISP-DM: a parallel overview. *IADS-DM*.
- [21] Sadi Evren SEKER. "Retail Data Set". Kaggle. 2023. <https://doi.org/10.34740/KAGGLE/DS/3067824>
- [22] G. Gürgen, "Customer Lifetime Value (Müşteri Yaşamboyu Değeri)", YBS Ansiklopedi, cilt 11, no. 2, pp. 56-72, 2023.
- [23] Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE access*, 8, 80716-80727.
- [24] Humaira, H., & Rasyidah, R. (2020, February). Determining the appropriate cluster number using elbow method for k-means algorithm. In *Proceedings of the 2nd Workshop on Multidisciplinary and Applications (WMA) 2018, 24-25 January 2018, Padang, Indonesia*.
- [25] Ros, F., Riad, R., & Guillaume, S. (2023). PDBI: A partitioning Davies-Bouldin index for clustering evaluation. *Neurocomputing*, 528, 178-199.
- [26] Raschka, S. (2018). MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. *Journal of open source software*, 3(24), 638.
- [27] Şeker, Ş. E. (2023). KOBİ'lere Özel Basit Yapay Zeka Çözümü: Kolay. *AI. Yönetim Bilişim Sistemleri Ansiklopedi*, 24.