



Bankacılıkta Müşteri Verilerini Anlama: CRISP-DM Yaklaşımı ve Makine Öğrenimi Uygulamaları

Sezai TUNCA, Alanya Üniversitesi, İktisadi, İdari ve Sosyal Bilimler Fakültesi, sezai.tunca@alanyauniversity.edu.tr, 0000-0001-9404-9005

ÖZ

Bu çalışma, günümüzde işletmeler için müşteri bilgilerinin önemini vurgulamaktadır. Özellikle finansal işlemlerle uğraşan bankaların bu bilgileri işleyip analiz ederek yönetim stratejilerinde kullanmasını ele alarak, veri analizinde CRISP-DM (Cross-Industry Standard Process for Data Mining) yaklaşımını önermektedir. Moro ve diğerleri tarafından sağlanan anonimleştirilmiş ve amaca uygun revize edilmiş veri seti analiz edilerek, banka müşterilerinin vadeli mevduata abone olma olasılığını etkileyen faktörler incelenmiş ve pazarlama stratejilerine yönelik veri destekli öneriler sunulmuştur. Çalışma, müşteri davranışını etkileyen değişkenleri belirlemek için Lojistik Regresyon, Karar Ağaçları (Decision Tree), Rastgele Orman (Random Forest), Destek Vektör Makineleri (SVM) ve XGBoost gibi çeşitli makine öğrenimi tekniklerini kullanarak, CRISP-DM metodolojisini altı aşamada yapmıştır. En iyi performansı gösteren Lojistik Regresyon modeli olmuştur. Bulgularda “emekli, öğrenci, yüksek eğitim seviyesi, önceki kampanya başarısı ve mart ayı gibi değişkenler”, müşterinin vadeli mevduat aboneliğine daha yatkın olduğunu göstermektedir. “Konut kredisi ve kredi borcu” gibi finansal yükümlülükler müşterinin vadeli mevduat aboneliği yapma olasılığını azaltmaktadır. Önerilerden bazıları arasında anlamlı değişkenlere (ör. meslek, eğitim seviyesi, medeni durum) odaklanarak stratejiler geliştirilmesinin yanı sıra eğitim seviyesi yüksek bireyleri hedefleyen pazarlama stratejileri geliştirilebilir. Finansal durumu zorlayıcı olan bireylerin hedef olay üzerinde olumsuz etkileri olduğundan, bu gruplara yönelik uygun öneriler sunulabilir. Çalışmanın sınırlamaları arasında veri setinin güncelliği ve sınıf dengesizliğini gidermek için kullanılan aşırı örnekleme tekniği yer almaktadır. Gelecekte yapılacak araştırmalarda farklı makine öğrenimi modelleri ve daha geniş veri setleri kullanılarak kapsamlı analizler gerçekleştirilebilir.

Anahtar Kelimeler : Bankacılık Analitiği, CRISP-DM Metodolojisi, Veri Analizi, Vadeli Mevduat Aboneliği

Understanding Customer Data in Banking: CRISP-DM Approach and Machine Learning Applications

ABSTRACT

This study highlights the importance of customer information for businesses today and addresses how banks, especially those engaged in financial transactions, process and analyse this information and use it in their management strategies and propose the CRISP-DM



(Cross-Industry Standard Process for Data Mining) approach to data analysis. Using the anonymised and purposefully revised dataset provided by Moro et al., the factors affecting bank customers' likelihood to subscribe to time deposits are examined and data-driven recommendations for marketing strategies are presented. The study applied the CRISP-DM methodology in six stages, using various machine learning techniques such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machines (SVM) and XGBoost to identify variables affecting customer behaviour. The Logistic Regression model performed the best. The findings show that 'variables such as retired, student, higher education level, previous campaign success, and month of March' indicate that the customer is more likely to subscribe to time deposit. Financial liabilities such as 'mortgage and loan debt'. The customer is less likely to subscribe to time deposits. Some of the suggestions include developing strategies by focusing on meaningful variables (e.g. occupation, education level, marital status) as well as developing marketing strategies targeting individuals with higher education levels. Since individuals with challenging financial situations have negative effects on the target event, appropriate recommendations can be made for these groups. Limitations of the study include the timeliness of the dataset and the oversampling technique used to overcome the class imbalance. In future research, comprehensive analyses can be performed using different machine learning models and larger data sets.

Keywords : Banking Analytics, CRISP-DM Methodology, Data Analysis, Term Deposit Subscription

GİRİŞ

Günümüzde müşteri bilgilerinden yararlanmak, işletmeler için oldukça önemli konulardan biridir. Özellikle finansal işlemlerle uğraşan bankalar söz konusu olduğunda, bu bilgilerin işlenmesi, analiz edilmesi ve çeşitli sonuçlara ulaşılması, yönetim kararları ve stratejilerinin önemli bir parçası haline gelmektedir. Bankaların müşterilerinin demografik verilerinden finansal alışkanlıklarını ve harcama eğilimlerinden kredi geçmişlerine kadar geniş bir veri kaynağına sahiptir. Bu verilerin doğru bir şekilde işlenmesi, analiz edilmesi ve anlamlı sonuçlara dönüştürülmesi, hem operasyonel verimliliği artırmak hem de stratejik karar alma süreçlerini desteklemek açısından kritik önem taşır (Moro et al., 2014). Bu verilerin analiziyle bankalar, müşterilerini daha iyi tanıma, kişiselleştirilmiş ürün yada hizmet sunma imkanı kazanır (Fawcett & Provost, 2013). Dahası bankalar yeni ürünlerin, kampanyalar ve operasyonel amaçlarını bu verilerin analiz sonuçlarına göre şekillendirebilir (Shmueli et al., 2017). Bu verilerin analizi için çeşitli yöntemlerden kullanılır. CRISP-DM (Cross-Industry Standard Process for Data Mining) yaklaşımı bu yöntemlerden birisidir. Analiz süreçlerinin daha sistematik bir şekilde yürütülmesi için kullanılan CRISP-DM yaklaşımı, bankaların veri işleme ve analiz süreçlerinde etkin bir rehberlik sunar. Veri analizi aşamalarında boşluk bırakmayan bu yaklaşım adım adım sonuçlara ulaşılmasını kolaylaştırır. Böylece müşteri bilgilerinin etkin bir şekilde kullanılmasıyla veri odaklı kararların alınmasının yolu açılır. Bu sistematik çerçeve bankaların yalnızca kısa vadeli operasyonel hedeflerine ulaşmasını değil, aynı zamanda uzun vadede müşteri ilişkilerini güçlendirmesine imkân verir. Bu çalışma, bu

yaklaşımı benimseyerek örnek bir veri seti üzerinden makine öğrenimi uygulamalarını kullanarak yapılmıştır. Çalışma örneği olan bankanın pazarlama departmanına veri odaklı stratejik öneriler sunmayı amaçlayan bu çalışma, veri seti olarak Moro ve arkadaşları (Moro et al., 2014) tarafından sağlanan, UCI web sitesinde bulunan anonimleştirilmiş verilerini kullanmaktadır. Amaçlara yönelik yapılan analizlerle müşterinin vadeli mevduat aboneliği yapma olasılığını etkileyen faktörleri anlamak ve bankanın pazarlama stratejilerine veri destekli öneriler sunmaktır. Bu bağlamda, veri setindeki ilişkileri keşfetmek ve müşteri davranışını etkileyen değişkenleri belirlemek ana hedef olarak belirlenmiş aynı zamanda farklı makine öğrenmesi yöntemlerini kullanarak bu yöntemlerin performansını karşılaştırılması yönetsel katkı sağlaması beklenmektedir. Çalışmanın ana sorunsalı, "*Hangi değişkenler müşterinin vadeli mevduat aboneliği yapma olasılığını etkileyebilir?*" sorusudur. Bu sorunun cevabı için R programlama dili kütüphaneleri kullanılarak analiz edilmiştir. Analizde kullanılan veri seti, banka müşterilerine ait çeşitli demografik ve finansal bilgileri içermektedir. Bu değişkenler arasında yaş, meslek, medeni durum, eğitim seviyesi, kredi durumu, bakiye, konut kredisi ve kişisel kredi durumu gibi özellikler bulunmaktadır. Hangi değişkenlerin analize dahil edileceğine ise test sonuçlarına göre karar verilmektedir. Ayrıca, müşteriye son kampanya kapsamında yapılan iletişim türü, kampanya ayı, haftanın günü ve kampanya süresi gibi faktörler de dikkate alınmıştır. Dahası çalışma müşterinin vadeli mevduat aboneliği yapma olasılığını etkileyen faktörleri belirlemek amacıyla çeşitli analiz teknikleri kullanılacaktır. Çalışmanın müşteri davranışını anlama konusunda veriye dayalı bilgiler elde edilerek, bankalara stratejik rehberlik sunması beklenmektedir.

Çalışmada, ilk olarak bankacılıkta veri madenciliği, makine öğrenimi ve uygulamaları, CRISP-DM metodolojisi hakkında literatür taraması yapılmış, daha sonra CRISP-DM aşamalarını içeren analizler yapılarak en son bulgular değerlendirilerek öneriler yapılmıştır.

1. LİTERATÜR TARAMASI

Bankacılık ve Veri Madenciliği: Bankacılık sektöründe veri madenciliği tekniklerinin kullanımı giderek artan bir eğilimdir. Bankacılık işlemleri sonucunda bankaların büyük miktarda veriye sahip olması ve bu verilerden anlamlı sonuçlar çıkarması, müşteri davranışlarını anlamada ve pazarlama stratejileri geliştirmede oldukça önemli konulardan birisidir. Bankaların pazarlama kampanyalarında veri madenciliği kullanarak müşteri eğilimlerini ve vadeli mevduat aboneliği gibi önemli parametreleri etkileyen faktörleri belirlemek için çalışmalar yapılmaktadır (Zaki et al., 2024; Zhuang et al., 2018). Bu tür veri odaklı çalışmalar, müşteri davranışlarını analiz etmede etkili sonuçlar vermektedir (Sandhya Kona, 2020). Veri madenciliği tekniklerinin bankacılıkta kullanımı, müşteri davranışlarını anlamakla sınırlı kalmamakta; aynı zamanda kredi risk değerlendirmesi ve dolandırıcılık tespiti gibi kritik işlevleri de kapsamaktadır. Shi ve arkadaşları (Shi et al., 2022), makine öğrenimi algoritmaları kullanarak kredi riski değerlendirmesinde veri madenciliğinin etkinliğini incelemiş ve bu algoritmaların kredi riski tespiti konusunda oldukça başarılı olduğunu göstermiştir. Aynı şekilde, dolandırıcılık tespiti için de veri madenciliği

algoritmaları, anormalliklerin belirlenmesinde önemli bir rol oynamaktadır (Noriega et al., 2023). Ayrıca son dönemlerde tüm alanlarda etkisini artıran üretken yapay zeka modellerinin, bankacılık çalışmalarında yeni bir ivme kazandırmıştır. Kumar ve arkadaşları (2024), üretken yapay zekanın finansal hizmetlerde müşteri analitiğini ve risk yöntemini üzerine etkisini incelemiştir. (Kumar et al., 2024). Başka bir çalışmada veri gizliliği ve güvenliği konuları, makine öğrenimi perspektifinden ele alınmış, müşteri verilerinin etik kullanımı ve düzenleyici çerçeveler üzerinde durulmuştur (Wang et al., 2024). Diğer taraftan gerçek zamanlı müşteri verilerinin pazarlama stratejilerinde uyarlanabilirliğini artırmadaki rolü de araştırmalarda önemli bir yer tutmuştur (Balamurugan, 2024). Bir diğer çalışmada bankacılıkta açıklanabilir yapay zekanın (XAI) uygulanabilirliği incelenmiş, bu modellerin müşteri davranış tahmini ve kredi risk analizi gibi alanlarda şeffaflığı artıran teknikler sunduğu belirtilmiştir (Rane et al., 2023). Ayrıca, müşteri kaybı tahmininde sınıf dengesizliği sorununa hibrit aşırı örnekleme yöntemleriyle çözümler sunulmuştur (Haddadi et al., 2024). Bu tür bankacılık uygulamalarında veri madenciliği tekniklerinin kullanımı, aynı zamanda müşteri segmentasyonu ve hedeflenmiş pazarlama stratejileri geliştirmede de etkili olmuştur (Smeureanu et al., 2013). Müşteri segmentasyonu, bankaların müşteri tabanını daha iyi anlamalarına ve bu sayede her bir segmente özel ürün ve hizmet teklifleri sunmalarına olanak sağlamaktadır (Osei et al., 2021). Makine öğrenimi teknikleri sayesinde bankalar, müşteri memnuniyetini arttırmak ve yeni müşteriler kazanmak için daha etkili stratejiler oluşturabilecektir.

Makine Öğrenimi Teknikleri ve Bankacılık Uygulamaları: Makine öğrenimi teknikleri, bankacılık sektöründe giderek daha fazla kullanılmaya başlamış ve bankaların çeşitli işlevlerini optimize etmelerinde önemli bir rol oynamıştır (Mahalakshmi et al., 2022). Bu teknikler, kredi risk değerlendirmesi, dolandırıcılık tespiti, müşteri segmentasyonu ve hedeflenmiş pazarlama kampanyaları gibi alanlarda yaygın olarak kullanılmaktadır (Nayak, 2024). Makine öğrenimi algoritmalarının kullanımı, veri setlerindeki karmaşık ilişkileri ortaya çıkarmakta ve bankaların stratejik karar alma süreçlerine katkı sağlamaktadır (Marqués et al., 2013). Karar Ağaçları (Decision Tree), Rastgele Orman (Random Forest), Destek Vektör Makineleri (SVM) ve XGBoost algoritmaları bankacılıkta sıkça kullanılan makine öğrenimi teknikleri arasındadır. Özellikle kredi riski değerlendirmesinde ve müşteri davranışının tahmin edilmesinde bu tekniklerin etkinliği kanıtlanmıştır (Bhatore et al., 2020; Shi et al., 2022).

Araştırma yöntemlerinin gelişmesiyle yeni çalışmalar, derin öğrenme tekniklerinin de bankacılıkta kullanımını incelemeye başlamıştır. Derin öğrenme (Goodfellow, 2016), karmaşık veri yapılarındaki özellikleri otomatik olarak öğrenebilmesi nedeniyle kredi riski, dolandırıcılık tespiti ve müşteri segmentasyonu gibi alanlarda etkili olmuştur (Mienye & Jere, 2024). Derin sinir ağları, büyük veri setleri ile çalışırken ve veri içindeki gizli ilişkileri ortaya çıkarmada oldukça başarılıdır (Samek et al., 2021). (Xiang et al., 2023), derin öğrenme

tekniklerinin dolandırıcılık tespiti için geleneksel makine öğrenimi tekniklerinden daha iyi performans gösterdiğini belirtmişlerdir.

Makine öğrenimi tekniklerinin bankacılık uygulamalarındaki diğer bir önemli kullanım alanı da müşteri deneyimini iyileştirmeye yöneliktir (Abdulsalam & Tajudeen, 2024). Doğrusal olmayan makine öğrenimi algoritmaları, müşteri geri bildirimlerini analiz ederek bankaların müşteri memnuniyetini artırarak riskleri azaltacak stratejiler geliştirmelerine yardımcı olabilmektedir (Leo et al., 2019). Bu teknikler, müşteri geri bildirimlerinden elde edilen verilerle bankacılık hizmetlerini kişileştirerek, daha özelleşmiş hizmet sunulmasını sağlamaktadır. Bu teknikler, veriler arasındaki karmaşık ilişkileri daha iyi öğrenebilmekte ve bu sayede daha hassas tahminler yapabilmektedir (Hoang & Wiegratz, 2023). Bu da bankaların kredi riskini daha iyi değerlendirmelerine ve dolandırıcılık aktivitelerini daha etkin bir şekilde tespit etmelerine yardımcı olmaktadır.

Bankacılık Müşteri Davranışı, Pazarlama Stratejileri ve Makine Öğrenmesi

Müşteri davranışını anlamak ve bu davranışlara uygun pazarlama stratejileri geliştirmek, bankacılık sektöründe önemli bir rekabet avantajı sağladığı bilinmektedir. Müşteri verilerinin analiz edilmesi, bankaların müşterilerine kişileşmiş teklifler sunmaları ve müşteri memnuniyetini arttırmaları açısından kritik bir rol oynamaktadır (Kotler et al., 2016). Makine öğrenimi teknikleri, bu analizlerin etkin bir şekilde yapılmasına olanak tanımakta ve pazarlama stratejilerinin daha hedefli ve etkili olmasını sağlamaktadır (Ho et al., 2022).

Bankacılık müşteri segmentasyonu, bankaların müşteri tabanını daha iyi anlamalarına ve her bir segmente özel stratejiler geliştirmelerine yardımcı olan bir makine öğrenimi uygulamasıdır (Bach et al., 2013). Özellikle, kümeleme algoritmaları kullanılarak benzer özelliklere sahip müşteriler gruplandırılmakta ve bu gruplara uygun pazarlama kampanyaları düzenlenmektedir (Duarte et al., 2022). Bu sayede, bankalar müşteri memnuniyetini arttırarak, müşteri sadakatini sağlamak ve daha etkili pazarlama stratejileri oluşturabilmektedir. Müşteri davranışının tahmin edilmesinde kullanılan makine öğrenimi algoritmaları, Karar Ağaçları (Decision Tree), Rastgele Orman (Random Forest), Destek Vektör Makineleri (SVM) ve XGBoost gibi algoritmalarıdır. Bu algoritmalar, müşteri profiline göre belirli bir ürün veya hizmete yönelik ilgiyi tahmin etmekte etkili bir şekilde kullanılmaktadır. Bankalar bu teknikleri kullanarak müşterilerinin hangi finansal ürünlere daha fazla ilgi gösterdiğini belirleyebilir ve bu bilgiler doğrultusunda kişileşmiş pazarlama stratejileri geliştirilebilir (Probesto, 2024). Ayrıca Makine öğrenmesi müşteri kaybını (churn) önceden tahmin etmede de önemli bir rol oynamaktadır (Lalwani et al., 2022). Müşteri kaybını önceden belirlemek, bankaların kayıp yaşayabilecek müşterilere yönelik koruma stratejileri geliştirmelerine olanak tanır. Destek Vektör Makineleri (SVM), Lojistik Regresyon ve Derin Öğrenme teknikleri, müşteri kaybını önceden tahmin etmek için yaygın olarak kullanılmaktadır (Agarwal et al., 2022). Bu tahminlere belirli metodoloji ile yaklaşmak için CRISP-DM kullanılması, mevcut bankacılık verilerinin işlenmesi ve analiz edilmesinde sistematik bir yaklaşım getirmektedir.

CRISP-DM Metodolojisi ve Bankacılık

CRISP-DM metodolojisi, veri madenciliği projelerinde yaygın olarak kullanılan (Marbn et al., 2009) ve bankacılık sektöründe de etkili bir şekilde uygulanabilecek bir yaklaşımdır. Bu metodoloji, veri madenciliği projelerinin sistematik bir şekilde planlanması ve yönetilmesini sağlar (Eckerson et al., 2000). CRISP-DM metodolojisi, özellikle bankacılık gibi büyük miktarda veriyle çalışan endüstrilerde, veri madenciliği süreçlerinin daha verimli yürütülmesine yardımcı olur (Plotnikova et al., 2022). Bu çalışmada, bankaların müşteri davranışlarını daha iyi anlamaları ve doğrudan pazarlama kampanyalarının etkinliğini artırmaları konusunda bu yaklaşım yol gösterici olmuştur. Bu metodolojinin kullanılması, veri madenciliği projelerinin her aşamasında sistematik bir yaklaşım benimsenmesini sağlamakta ve bu da projelerin başarı oranını artırtıracağı beklenmektedir.

CRISP-DM altı temel aşamada gerçekleştirilen bir süreci kapsar. Bu aşamalar; iş anlayışı, veri anlayışı, veri hazırlığı, modelleme, değerlendirme ve yayılma şeklindedir. Çalışmanın veri setini analiz etmede yöntem olarak seçilmiştir.

2. METOD VE ANALİZLER

Araştırmada bankanın verileri, CRISP-DM yöntemi kullanılarak analiz edilmiştir. CRISP-DM yöntemi, veri madenciliği ve veri analizi projeleri için yaygın olarak kabul edilen endüstri standartlardan birisidir (Martínez-Plumed et al., 2021). Bu metodoloji, veri madenciliği projelerini planlama, uygulama ve değerlendirme aşamalarına ayırarak, projelerin sistematik ve etkili bir şekilde yönetilmesine katkı sağlar.

CRISP-DM metodolojisi genellikle aşağıdaki altı ana aşamayı içerir:

1. Problemin Tanımlanması (Business Understanding)
2. Verinin Anlaşılması (Data Understanding)
3. Veri Ön Hazırlığı (Data Preparation)
4. Modelleme (Modeling)
5. Değerlendirme (Evaluation)
6. Yayılma (Deployment)

Her bir aşama, analizlerin başarıyla tamamlanabilmesi için kritik öneme sahiptir. Analizlerde CRISP-DM metodolojisinin kullanılması, araştırmaya sistematik bir çerçeve sunmuş ve analizin her aşamasının gereksinimlere uygun olarak yürütülmesini sağlamıştır.

2.1. Problemin Tanımlanması

Portekizli bir bankanın doğrudan pazarlama kampanyalarını konu alan banka pazarlama veri seti amaca uygun düzenlenerek kullanılmıştır. Pazarlama kampanyaları telefon görüşmelerine dayanmaktadır ve ürüne (banka vadeli mevduat) abone olunup olunmayacağını belirlemek amacıyla müşterilerle birden fazla kez iletişim kurulmuştur. Bu bağlamda, problemimiz bir sınıflandırma problemidir. Sınıflandırma problemine uygun algoritmalar seçilmelidir.

2.2. Verilerin Anlaşılması

Mevcut veri setlerini incelemek ve anlamak için R programlama dilindeki kütüphaneler kullanılmıştır. Veri setinin özellikleri Tablo 1'de sunulmuştur. Veri seti, 7 adet sayısal ve 10 adet kategorik değişkenden oluşmaktadır. Veri matrisi, toplamda 45.211 gözlem ve 17 değişkene sahiptir.

Tablo 1: Veri setinin Özellikleri

age	job	marital	education	default	balance	housing	loan	contact
Min. :18.00	Length:45211	Length:45211	Length:45211	Length:45211	Min. : -8019	Length:45211	Length:45211	Length:45211
1st Qu.:33.00	Class :character	Class :character	Class :character	Class :character	1st Qu.: 72	Class :character	Class :character	Class :character
Median :39.00	Mode :character	Mode :character	Mode :character	Mode :character	Median : 448	Mode :character	Mode :character	Mode :character
Mean :40.94					Mean : 1362			
3rd Qu.:48.00					3rd Qu.: 1428			
Max. :95.00					Max. :102127			
day	month	duration	campaign	pdays	previous	poutcome	y	
Min. : 1.00	Length:45211	Min. : 0.0	Min. : 1.000	Min. : -1.0	Min. : 0.0000	Length:45211	Length:45211	
1st Qu.: 8.00	Class :character	1st Qu.: 103.0	1st Qu.: 1.000	1st Qu.: -1.0	1st Qu.: 0.0000	Class :character	Class :character	
Median :16.00	Mode :character	Median : 180.0	Median : 2.000	Median : -1.0	Median : 0.0000	Mode :character	Mode :character	
Mean :15.81		Mean : 258.2	Mean : 2.764	Mean : 40.2	Mean : 0.5803			
3rd Qu.:21.00		3rd Qu.: 319.0	3rd Qu.: 3.000	3rd Qu.: -1.0	3rd Qu.: 0.0000			
Max. :31.00		Max. :4918.0	Max. :63.000	Max. :871.0	Max. :275.0000			

2.3. Veriler Ön Hazırlığı

Bu aşamada, veri setleri temizlenir, dönüştürülür ve modelleme aşamasına uygun hale getirilir. Eksik veriler doldurulur, aykırı değerler ele alınır ve gerekli düzeltmeler yapılır. Veri hazırlığı aşaması, veri madenciliği projelerinin başarısında kritik rol oynar. Bu çalışmada da eksik değerlerin ve aykırı değerlerin kontrol edilmesi, kategorik değişkenlerin dönüştürülmesi gibi adımlar dikkatli bir şekilde uygulanmıştır. Veri hazırlığı sürecinin başarıyla tamamlanması, modelin doğruluğunu ve güvenilirliğini doğrudan etkilemektedir.

2.3.1. Eksik Değerlerin Kontrolü

Veri setindeki eksik değerlerin kontrol edilmesi, veri setinin doğruluğunu ve kullanılabilirliğini değerlendirmek açısından önemlidir. Eksik değerlerin olmaması, analiz ve modelleme süreçlerinin daha sağlıklı ve güvenilir olmasını sağlar. Veri seti kontrol edildiğinde eksik gözlem olmadığı belirlenmiştir. Bu sonuç model eğitimi ve analizler için potansiyel bir problemin olmadığını gösteren olumlu bir durumdur. Bu nedenle, veri setini kullanmaya devam edilip ileri analiz aşamalarına geçilmiştir.

2.3.2. Kategorik Sütunların Geçerlilik Kontrolü

Tüm kategorik sütunlardaki geçerlilik kontrolü, veri setinin temizliği açısından önemlidir. Geçersiz veya tutarsız değerlerin olmaması, model eğitimi ve analiz süreçlerinin güvenilir sonuçlar üretmesine katkı sağlar. Kategorik sütunlardaki eksik değer sayıları kontrol edildi ve eksik değer olmadığı görüldü. Bu kontrol adımlarının sonucunda veri setinin temiz ve model eğitimi için uygun olduğu teyit edilmiştir. Bu durum, sonuçların güvenilirliğini ve doğruluğunu artıracaktır.

2.3.3. Sayısal Değişkenlerin Aykırı Değer Analizi

Sayısal sütunların özet istatistikleri incelendiğinde, "balance" (bakiye), "duration" (süre), "campaign" (kampanya), "pdays" (önceki günler) gibi sütunlarda bazı uç ve aykırı değerlerin olduğu görülmüştür. Bakiye değişkeni için minimum değer -8019, ortalama ise 1428 olarak belirlenmiştir. Bu durum veri setinde olası aykırı değerlere işaret eder. Benzer şekilde, süre için minimum değer 0, ortalama ise 258.2'dir, bu da aykırı değerlerin varlığını gösterir. Aykırı değerler, Çeyrekler Arası Aralık (IQR) yöntemi kullanılarak veri setinden çıkarılmıştır. Bu yöntem, bir değişkenin değer dağılımını ölçerek aykırı değerleri belirler ve bu değerler veri setinden kaldırılmıştır.

2.3.4. Benzersiz Değerlerin Kontrolü ve Gereksiz Sütunların Kaldırılması

Aykırı değerler çıkarıldıktan sonra, tüm sütunlardaki benzersiz değerler kontrol edilmiştir. "duration" sütununun, bir model oluşturmaya katkı sağlamayan tek bir benzersiz değere sahip olduğu belirlenmiştir. Bu nedenle, bu sütun veri kümesinden çıkarılmıştır.

2.3.5. Kategorik ve Hedef Sütunun Dönüştürülmesi

Kategorik sütunları ve hedef sütunu karakter değişkenlerinden faktör değişkenlerine dönüştürmek, özellikle sınıflandırma modellerinin daha iyi çalışmasına yardımcı olur. Bu nedenle, kategorik sütunlar ve hedef sütun karakter değişkenlerinden faktör değişkenlerine dönüştürülmüştür.

2.3.6. Hedef Sütundaki Dengesiz Sınıflar

Veri setinde hedef sütun "y" olarak adlandırılmış ve iki sınıfı "evet" ve "hayır" şeklindedir. Ancak, "evet" sınıfında 5289 gözlem, "hayır" sınıfında ise 39.922 gözlem bulunmaktadır. Bu durum, dengesiz sınıflandırma problemine işaret eder, çünkü sınıflar arasındaki gözlem sayılarında belirgin bir fark vardır. Sınıf dengesizliği, modelin öğrenme mekanizmasına ve optimizasyon sürecine bağlı olarak bazı modellerde daha ciddi sorunlara yol açabilir. Örneğin, doğruluk (accuracy) gibi genel hata oranlarını minimize etmeye çalışan modeller, çoğunluk sınıfına odaklanarak azınlık sınıfı göz ardı edebilir (He & Garcia, 2009). Karar ağaçları ve destek vektör makineleri (SVM) gibi modeller, sınıf frekanslarındaki büyük

farklar nedeniyle sınıf sınırlarını azınlık sınıf aleyhine eğebilirken, sinir ağları gibi daha karmaşık modeller, azınlık sınıfa ait yeterli örnek olmadığında bu sınıfı öğrenmekte zorlanabilir (Chawla et al., 2002). Ayrıca, dengesiz sınıflar veri dağılımını bozarak modelin genelleme yeteneğini etkileyebilir ve performans metriklerinin yanıltıcı sonuçlar vermesine neden olabilir. Bu nedenle, sınıf dengesizliği yaşayan veri setlerinde, modelin duyarlılığına uygun yöntemler ve metrik odaklı yaklaşımlar tercih edilmelidir. Dengesiz sınıflandırma problemleri modelin eğitilmesi ve değerlendirilmesi üzerinde olumsuz etki yaratabilir. Bu nedenle, bu tür durumlarda bir dengesizlik giderme stratejisi uygulanmıştır. "rastgele aşırı örnekleme (Random oversampling)" tekniği kullanılarak az sayıdaki sınıfın gözlemleri artırılmıştır. Bu işlem, küçük sınıfın örnek sayısını artırarak modelin bu sınıfı daha iyi öğrenmesine ve dengesizlikten kaynaklanan yanlılıkları azaltmasına yardımcı olmaktadır.

2.3.7. Eğitim ve Test Setlerine Bölme

Veri setini eğitim ve test setlerine bölmek, modelin eğitim verileri üzerinde öğrenmesi ve ardından bu öğrenilen bilgileri test verileri üzerinde değerlendirmesi için yaygın bir yöntemdir. Bu işlem, modelin gerçek dünya verileri üzerinde genelleme yeteneğini değerlendirmek amacıyla yapılır. Bu şekilde modelin genelleme performansını değerlendirme, aşırı uyum (overfitting) kontrolü ve performansın güvenilir olarak ölçülmesi sağlanır. Eğitim ve test setleri %80 ve %20 oranında ayrılmıştır; verilerin %80'i model eğitimi için, %20'si ise model test işlemleri için kullanılmıştır. Modelin performansını değerlendirmek için doğruluk (accuracy), kesinlik (precision), duyarlılık (recall), F1 skoru gibi denetimli öğrenme metrikleri kullanılmıştır. Bu metrikler, modelin sınıflandırma yeteneklerini nicel olarak ölçmektedir.

2.4. Modelleme

Bu aşamada, veri seti üzerinde çeşitli modelleme teknikleri uygulanmıştır. Amaç, belirli bir iş sorununa en uygun modeli seçmek ve eğitmektir. Verilerin dağılımı incelendiğinde doğrusal bir dağılıma sahip olmadığı görülmüştür. Veri setinin özellikleri sınıflandırmaya uygun olduğundan Lojistik Regresyon, Karar Ağaçları (Decision Tree), Rastgele Orman (Random Forest), Destek Vektör Makineleri (SVM) ve XGBoost analizleri yapılmıştır.

2.4.1. Lojistik Regresyon

Lojistik regresyon, bağımlı değişkenin kategorik veya ikili (binary) olduğu durumlarda kullanılan bir istatistiksel analiz yöntemidir. Bu yöntem, bir veya daha fazla bağımsız değişkenin bağımlı değişken üzerindeki etkisini incelemek için kullanılır. Bu çalışmada da, "y" değişkeni (*müşterinin vadeli mevduat aboneliği yapma olasılığı*) üzerinde değişkenlerin dağılımı incelenmiş ve sonuçlar Tablo 2'de sunulmuştur. Tablo 2'deki bulgular, bazı değişkenlerin bağımlı değişken üzerinde önemli bir etkisi olduğunu göstermektedir. Bu sonuçlar bir değerlendirme bölümünde Tablo 3'te detaylı olarak verilmiştir.

Modelin AIC (Akaike Bilgi Kriteri) değeri 17342 olarak hesaplanmıştır. Bu değer, modelin performansını değerlendirmek için diğer modellerle karşılaştırılacaktır.

Tablo 2: Lojistik Regresyon Sonuçları

Coefficients:	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-270000	0.205000	-13.222	< 2e-16
age	-0.000448	0.002470	-0.182	0.855864
jobblue-collar	-0.232000	0.081100	-2.861	0.004229
jobentrepreneur	-0.315000	0.139000	-2.267	0.023372
jobhousemaid	-0.366000	0.150000	-2.434	0.014953
jobmanagement	-0.130000	0.082300	-1.577	0.114764
jobretired	0.320000	0.109000	2.924	0.003452
jobself-employed	-0.218000	0.124000	-1.763	0.077956
jobservices	-0.176000	0.094400	-1.865	0.062231
jobstudent	0.318000	0.122000	2.601	0.0093
jobtechnician	-0.153000	0.077800	-1.962	0.049796
jobunemployed	-0.127000	0.126000	-1.011	0.312025
jobunknown	-0.256000	0.261000	-0.982	0.326268
maritalmarried	-0.170000	0.065900	-2.581	0.009845
maritalsingle	0.140000	0.075000	1.861	0.062762
educationsecondary	0.241000	0.072400	3.322	0.000893
educationtertiary	0.435000	0.084300	5.158	2.50E-07
educationunknown	0.316000	0.116000	2.721	0.006515
defaultyes	0.077200	0.179000	0.431	0.666312
balance	0.000012	0.000006	2.075	0.03799
housingyes	-0.690000	0.048900	-14.092	< 2e-16
loanyes	-0.403000	0.066500	-6.053	1.42E-09
contacttelephone	-0.085700	0.082900	-1.033	0.301394
contactunknown	-1.620.000	0.081400	-19.927	< 2e-16
day	0.012700	0.002790	Nis.57	4.87E-06
monthaug	-0.762000	0.088000	-8.658	< 2e-16
monthdec	0.670000	0.195000	3.434	0.000594
monthfeb	-0.197000	0.100000	-1.968	0.049011
monthjan	-1.380.000	0.138000	-10	< 2e-16
monthjul	-0.868000	0.085800	-10.111	< 2e-16
monthjun	0.458000	0.104000	4.395	1.11E-05
monthmar	1.530.000	0.134000	11.344	< 2e-16
monthmay	-0.447000	0.080600	-5.54	3.03E-08
monthnov	-0.914000	0.094200	-9.704	< 2e-16
monthoct	0.908000	0.120000	7.585	3.32E-14
monthsep	0.893000	0.132000	Haz.77	1.29E-11

pdays	0.000174	0.000344	0.506	0.613163
previous	0.007450	0.006340	1.177	0.239384
poutcomeother	0.209000	0.102000	2.061	0.039343
poutcomesuccess	2.340.000	0.092500	25.27	< 2e-16
poutcomeunknown	0.001240	0.105000	0.012	0.990605
Signif. codes: 0	***' 0.001	***' 0.01	**' 0.05	' 0.1
Null deviance:	26104 on	36167 degr	ees of f	reedom
Residual deviance:	17256 on	36125 degr	ees of f	reedom
AIC: 17342				

2.5. Değerlendirme (Evaluation)

En iyi performans gösteren birisi olan Lojistik Regresyon sonuçları (Tablo 2) değerlendirildiğinde; müşteri davranışlarını anlamak ve finansal ürünlerin benimsenmesine yönelik önemli içgörüler sunmaktadır. pozitif katsayısı olan değişkenler (Tablo 3), Emekli bireyler (jobretired) ve öğrenciler (jobstudent) gibi grupların vadeli mevduat aboneliğine daha yatkın olduğu görülmektedir. Bu bulgu, emekli bireylerin düzenli gelir kaynaklarına ve öğrencilerin geleceğe yönelik finansal planlama ihtiyaçlarına odaklanan literatürle uyumludur (Lusardi & Mitchell, 2014). Yüksek eğitim seviyesinin (educationsecondary ve educationtertiary) abonelik olasılığını artırması da finansal okuryazarlık düzeyinin artışıyla ilişkilendirilebilir (Bilge & Nur, 2023). Önceki kampanya başarısının (poutcomesuccess) müşteri davranışını önemli ölçüde etkilediği sonucuna ulaşılmıştır. Bu bulgu, müşteri ilişkileri yönetiminin ve geçmiş deneyimlerin gelecekteki ürün kabul oranları üzerindeki etkisini gösteren (Verhoef et al., 2003) tarafından desteklenmektedir. Benzer şekilde, mart ayında vadeli mevduat aboneliğine olan yüksek ilgi (monthmar), sezonluk etkiler ve vergi iadeleri gibi finansal davranışlarla açıklanabilir .

Sonuçlardaki negatif etkiler arasında (Tablo 3), mevcut finansal yükümlülüklerin (housingyes ve loanyes) vadeli mevduat aboneliğini azalttığı gözlemlenmiştir. Bu durum, borç ödeme yükümlülüklerinin tasarruf davranışlarını baskılayabileceğini savunan (Campbell & Cocco, 2003)'nun çalışması ile uyumludur. Mavi yaka çalışanların (jobblue-collar) abonelik olasılığının düşük olması, gelir seviyeleri ve harcama öncelikleriyle ilişkilendirilebilir (Çımrın et al., 2023). Temmuz ayında (monthjul) abonelik oranlarının düşüklüğü ise tatil harcamaları ve sezonluk harcama artışları ile ilişkilendirilebilir (Claessens & Laeven, 2004).

Bu lojistik regresyon analizinin bulguları, bankaların müşteri segmentasyonunu iyileştirmesi ve hedefli kampanyalar geliştirmesi için önemli bilgiler sunmaktadır. Yüksek eğitim seviyesine sahip veya önceki kampanyalardan olumlu sonuç alan gruplara odaklanmak, pazarlama stratejilerinin etkinliğini artırabilir (Özdemir, 2021; Demirel, 2024). Ayrıca, (Bumin, 2023) çalışmasına benzer mavi yaka çalışanlar ve finansal yükümlülüğü

yüksek bireyler gibi gruplara yönelik farklı ürünler veya finansal destek politikaları geliştirilmesi önerilmektedir.

Tablo 3: Analiz Sonuçlarının Değerlendirmesi

Değişken	Katsayı	p Değeri	Anlamlılık Seviyesi (***0.001,**0.01,*0.05)	Yorum
jobblue-collar	-0.232	0.004229	**	Mavi yaka işçilerin vadeli mevduat aboneliği yapma olasılığı düşüktür.
jobretired	0.320	0.003452	**	Emekli bireylerin vadeli mevduat aboneliği yapma olasılığı yüksektir.
jobstudent	0.318	0.0093	**	Öğrencilerin vadeli mevduat aboneliği yapma olasılığı yüksektir.
maritalmarried	-0.170	0.009845	**	Evli olmanın vadeli mevduat aboneliği yapma olasılığı üzerinde negatif etkisi vardır.
educationsecondary	0.241	0.000893	***	Ortaokul mezunlarının vadeli mevduat aboneliği yapma olasılığı yüksektir.
educationtertiary	0.435	2.50E-07	***	Üniversite mezunlarının vadeli mevduat aboneliği yapma olasılığı oldukça yüksektir.
housingyes	-0.690	< 2e-16	***	Konut kredisine sahip olmak, vadeli mevduat aboneliği yapma olasılığını önemli ölçüde azaltır.
loanyes	-0.403	1.42E-09	***	Kredi borcuna sahip olmak, vadeli mevduat aboneliği yapma olasılığını düşürmektedir.
contactunknown	-1.620	< 2e-16	***	İletişim türü bilinmiyorsa, vadeli mevduat aboneliği yapma olasılığı oldukça düşüktür.
monthmar	1.530	< 2e-16	***	Mart ayında vadeli mevduat aboneliği yapma olasılığı yüksektir.

monthjul	-0.868	< 2e-16	***	Temmuz ayında vadeli mevduat aboneliği yapma olasılığı düşüktür.
poutcomesuccess	2.340	< 2e-16	***	Önceki kampanyanın başarılı olması, vadeli mevduat aboneliği yapma olasılığını artırır.

2.5.1. Veri Setinin Diğer Makine Öğrenmesi Algoritmalarıyla Karşılaştırması

Veri setinin diğer modeller ile performansı karşılaştırıldığında (Tablo 4), Lojistik Regresyon, Random Forest ve XGBoost modeli için doğruluk (accuracy) değeri %88, Decision Tree ise %85,5 olarak tespit edilmiştir. Her bir modelde de AUC değeri %80'in üzerinde olduğundan, iyi uyumlu modeller olarak kabul edilmiştir. Modeller abonelik tahminlerinde, Class 0'da (abonelik yapmayanlar) oldukça başarılı performans sergilerken, Class 1 (abonelik yapanlar) için F1 skorları düşük kalmaktadır. Abonelik yapanların tekrar hesap açma ihtimalleri düşüktür.

Tablo 4: Model Performansları

Model	Accuracy	F1 Score (Class 0)	F1 Score (Class 1)
Logistic Regression	0.8879796527700984	0.94	0.32
Decision Tree	0.8551082100518141	0.92	0.47
Random Forest	0.8878687080333574	0.94	0.32
SVM	0.8789464863284238	0.93	0.25
XGBoost	0.8862173966065948	0.93	0.37

2.6. Yayılma

Bu aşamada analiz sonuçlarının ne anlama geldiği ve uygulayıcılara öneriler yer alır. Tablo 5'te bulguların sonuçları ve öneriler listelenmiştir. Analize konu olan bankanın bu sonuçları ve önerileri dikkate alarak mevcut durumun daha da ileriye götürebilecek stratejiler geliştirebilir.

Tablo 5: Sonuçların Yorumlanması ve Öneriler

Sonuçlar	Öneriler
Anlamli deęişkenler belirlendi.	Anlamli deęişkenlere (meslek, eğitim seviyesi, medeni durum) odaklanarak stratejiler geliştirilebilir.
Anlamsız deęişkenler tespit edildi.	Anlamli katkı yapmayan deęişkenler (yaş, defaultyes) modelden çıkarılabilir veya yeniden gözden geçirilebilir.
Eğitim seviyesi hedef olayı etkiliyor.	Eğitim seviyesi yüksek bireyleri hedefleyen pazarlama stratejileri geliştirilebilir.
Meslek gruplarının etkisi gözlemlendi.	Hedef kitlenin mesleki durumuna göre özelleştirilmiş kampanyalar hazırlanabilir.

Konut ve kredi borcunun negatif etkisi.	Finansal durumu zorlayıcı olan bireylerin hedef olay üzerinde olumsuz etkileri olduğundan, bu gruplara yönelik uygun öneriler sunulabilir.
Önceki kampanyanın başarısı önemli.	Önceki kampanyalarda başarı gösteren grupları belirleyip, bu grupları yeniden hedeflemek faydalı olabilir.

SONUÇ

Bu çalışma, bankacılık verilerini CRISP-DM metodolojisi kullanarak sistemli bir şekilde analiz etmiştir. Moro ve diğerleri tarafından sağlanan, Portekizli bir bankanın pazarlama departmanına veri odaklı stratejik öneriler sunma amacı güden anonimleştirilmiş veri seti, banka müşterilerinin vadeli mevduat aboneliği yapma olasılığını anlamak ve pazarlama stratejilerine veri destekli öneriler sunmak için kullanılmıştır. Veri madenciliği ve makine öğrenimi teorilerinin bir uygulaması olan CRISP-DM metodolojisi yaklaşımıyla da veri setinden anlamlı bilgiler çıkarma süreçlerine odaklanmıştır. Bu yaklaşım, bankaların müşteri verilerini analiz etme ve bu verilerden elde edilen içgörülerle pazarlama stratejileri ile entegre etme süreçlerinde, destekleyici olabileceğini göstermektedir. Veri setindeki ilişkileri keşfetmek ve müşteri davranışını etkileyen değişkenleri belirlemek hedeflenmiş; hangi değişkenlerin vadeli mevduat aboneliği yapma olasılığını etkileyebileceği sorusuna yanıt aranmıştır. Yöntem olarak CRISP-DM yaklaşımı kullanılmış ve çeşitli makine öğrenmesi algoritmalarıyla sonuçlar doğrulanmış ve performansları test edilmiştir. Lojistik Regresyon modeli, önemli değişkenlerle yüksek doğruluk sağlamış; diğer modeller de benzer performans sergilemiştir. Araştırma sorusuna yanıt olarak, müşteri bilgileri arasında özellikle yaş, finansal durum, önceki pazarlama kampanyalarına verilen tepkiler gibi değişkenlerin vadeli mevduat aboneliğini anlamlı bir şekilde etkilediği görülmüştür. Elde edilen bulgular, belirli demografik ve finansal değişkenlerin, müşterilerin vadeli mevduat aboneliği yapma olasılıkları üzerinde anlamlı etkileri olduğunu göstermektedir. Özellikle "emekli, öğrenci, yüksek eğitim seviyesi, önceki kampanya başarısı ve mart ayı" gibi değişkenler abonelik olasılığını artırırken; "konut kredisi ve kredi borcu" gibi finansal yükümlülükler bu olasılığı azaltmaktadır. Bulgular, anlamlı değişkenlere (meslek, eğitim seviyesi, medeni durum gibi) odaklanarak stratejiler geliştirilmesi gerektiğini ve eğitim seviyesi yüksek bireyleri hedefleyen pazarlama stratejilerinin daha etkili olabileceğini önermektedir. Ayrıca, finansal durumu zorlayıcı olan bireylerin vadeli mevduat aboneliği üzerindeki olumsuz etkileri göz önünde bulundurularak, bu gruplara yönelik uygun kişiselleştirilmiş öneriler sunulması gerekmektedir. Diğer taraftan her bir makine modeli, başarı kriterlerini karşılamaktadır. Bulgular ise müşteri bilgilerini anlamamanın önemini ortaya çıkarmaktadır. Özellikle finansal işlemlerle ilgilenen bankaların bu tür verileri analiz ederek, sonuçları yönetim stratejilerinde kullanmaları önerilebilir. Çalışmanın sınırlamaları arasında veri setinin güncelliği ve sınıf dengesizliğini gidermek için kullanılan aşırı örnekleme tekniği bulunmaktadır. Bu sınırlamaların etkisi, sonuçların genellenebilirliğini sınırlayabilir. Gelecekteki araştırmalarda farklı makine öğrenimi modelleri

(AutoML (Otomatik Makine Öğrenimi), Öngörücü Yapay Zeka Modelleri gibi) ve daha geniş veri setleri kullanılarak daha kapsamlı analizler yapılabilir. Bununla birlikte, teorik katkıların artırılması amacıyla müşteri davranışları ile bankacılık stratejileri arasındaki ilişkiyi açıklayan daha derinlemesine bir yaklaşım oluşturulması gerekebilir. Gelecekteki araştırmalar, farklı sektörlerdeki veri analizleri ve daha karmaşık makine öğrenimi teknikleri ile bu yaklaşımın geliştirilmesini sağlayabilir.

KAYNAKÇA

- Abdulsalam, T. A., & Tajudeen, R. B. (2024). Artificial intelligence (AI) in the banking industry: A review of service areas and customer service journeys in emerging economies. *Business & Management Compass*, 68(3), 19–43. <https://doi.org/10.56065/9hfvqrq20>
- Agarwal, V., Taware, S., Yadav, S., Gangodkar, D., Rao, A., & Srivastav, V. (2022). Customer—churn prediction using machine learning. *Proceedings of ICTACS*, 899. <https://doi.org/10.1109/ICTACS56270.2022.9988187>
- Bach, M. P., Juković, S., Dumičić, K., & Šarlija, N. (2013). Business client segmentation in banking using self-organizing maps. *South East European Journal of Economics and Business*, 8(2), 32–41. <https://doi.org/10.2478/jeb-2013-0007>
- Balamurugan, M. (2024). AI-driven adaptive content marketing: Automating strategy adjustments for enhanced consumer engagement. *International Journal For Multidisciplinary Research*, 6. <https://doi.org/10.36948/ijfmr.2024.v06i05.27940>
- Bhatore, S., Mohan, L., & Reddy, Y. R. (2020). Machine learning techniques for credit risk evaluation: A systematic literature review. *Journal of Banking and Financial Technology*, 4(1), 111–138. <https://doi.org/10.1007/s42786-020-00020-3>
- Bilge, C., & Nur, T. (2023). Finansal performansı etkileyen içsel faktörler: Mevduat ve katılım bankaları üzerine bir uygulama. *Ardahan Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 5(2), Article 2. <https://doi.org/10.58588/aru-jfeas.1292408>
- Bumin, M. (2023). Türk bankacılık sektöründe mevduat bankalarının karlılık performansını etkileyen faktörlerin panel regresyon analizi ile belirlenmesi. *Muhasebe ve Finansman Dergisi*, 100, Article 100. <https://doi.org/10.25095/mufad.1326939>
- Campbell, J. Y., & Cocco, J. F. (2003). Household risk management and optimal mortgage choice. *The Quarterly Journal of Economics*, 118(4), 1449–1494.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357. <https://doi.org/10.1613/jair.953>
- Çımrın, A. H., Kaya, İ., & Bahadır, H. (2023). Türkiye’de mavi yakalı çalışan olma hali üzerine bir analiz. *Sosyoloji Dergisi*, 45, Article 45. <https://doi.org/10.59572/sosder.1225697>
- Claessens, S., & Laeven, L. (2004). What drives bank competition? Some international evidence. *Journal of Money, Credit and Banking*, 36(3), 563–583.
- Demirel, S. (2024). Bankacılıkta dijitalleşmenin müşteri davranışları üzerine etkisi. Retrieved from <https://www.gazikitabevi.com.tr/bankacilikta-dijitallesmenin-musteri-davranislari-uzerine-etkisi>
- Duarte, V., Zuniga-Jara, S., & Contreras, S. (2022). Machine learning and marketing: A systematic literature review. *IEEE Access*, 10, 93273–93288. <https://doi.org/10.1109/ACCESS.2022.3202896>
- Eckerson, W. W., Hanlon, N., & Barquin, R. (2000). Director of education and research. 5(4).
- Fawcett, T., & Provost, F. (2013). *Data science for business*.

- Goodfellow, I. (2016). *Deep learning*. MIT Press. Retrieved from [https://books.google.com/books?hl=en&lr=&id=omivDQAAQBAJ&oi=fnd&pg=PR5&dq=Goodfellow,+I.,+Bengio,+Y.,+%26+Courville,+A.,+\(2016\).+Deep+Learning,+MIT+Press.&ots=MOO5bmsHTY&sig=43Ua_77nITneAJaXBCXMRPWQYbU](https://books.google.com/books?hl=en&lr=&id=omivDQAAQBAJ&oi=fnd&pg=PR5&dq=Goodfellow,+I.,+Bengio,+Y.,+%26+Courville,+A.,+(2016).+Deep+Learning,+MIT+Press.&ots=MOO5bmsHTY&sig=43Ua_77nITneAJaXBCXMRPWQYbU)
- Haddadi, S. J., Farshidvard, A., Silva, F. dos S., dos Reis, J. C., & da Silva Reis, M. (2024). Customer churn prediction in imbalanced datasets with resampling methods: A comparative study. *Expert Systems with Applications*, 246, 123086. <https://doi.org/10.1016/j.eswa.2023.123086>
- He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284. <https://doi.org/10.1109/TKDE.2008.239>
- Ho, S. C., Wong, K. C., Yau, Y. K., & Yip, C. K. (2022). A machine learning approach for predicting bank customer behavior in the banking industry. In *Research Anthology on Machine Learning Techniques, Methods, and Applications* (pp. 1210–1232). IGI Global. <https://doi.org/10.4018/978-1-6684-6291-1.ch063>
- Hoang, D., & Wiegatz, K. (2023). Machine learning methods in finance: Recent applications and prospects. *European Financial Management*, 29(5), 1657–1701. <https://doi.org/10.1111/eufm.12408>
- Kotler, P., Keller, K. L., Brady, M., Goodman, M., & Hansen, T. (2016). *Marketing management* (3rd ed.). Pearson Higher Ed.
- Kumar, K., Kuhar, N., & Sharma, M. (2024). Artificial intelligence in the Indian banking system: A systematic literature review (SSRN Scholarly Paper No. 5088937). *Social Science Research Network*. <https://doi.org/10.2139/ssrn.5088937>
- Lalwani, P., Mishra, M. K., Chadha, J. S., & Sethi, P. (2022). Customer churn prediction system: A machine learning approach. *Computing*, 104(2), 271–294. <https://doi.org/10.1007/s00607-021-00908-y>
- Leo, M., Sharma, S., & Maddulety, K. (2019). Machine learning in banking risk management: A literature review. *Risks*, 7(1), 29. <https://doi.org/10.3390/risks7010029>
- Lusardi, A., & Mitchell, O. S. (2014). The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature*, 52(1), 5–44. <https://doi.org/10.1257/jel.52.1.5>
- Mahalakshmi, V., Kulkarni, N., Pradeep Kumar, K. V., Suresh Kumar, K., Nidhi Sree, D., & Durga, S. (2022). The role of implementing artificial intelligence and machine learning technologies in the financial services industry for creating competitive intelligence. *Materials Today: Proceedings*, 56, 2252–2255. <https://doi.org/10.1016/j.matpr.2021.11.577>
- Marbn, S., Mariscal, G., & Segovi, J. (2009). A data mining & knowledge discovery process model. In J. Ponce & A. Karahoc (Eds.), *Data mining and knowledge discovery in real life applications*. I-Tech Education and Publishing. <https://doi.org/10.5772/6438>
- Marqués, A. I., García, V., & Sánchez, J. S. (2013). A literature review on the application of evolutionary computing to credit scoring. *Journal of the Operational Research Society*, 64(9), 1384–1399. <https://doi.org/10.1057/jors.2012.145>
- Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernández-Orallo, J., Kull, M., Lachiche, N., Ramírez-Quintana, M. J., & Flach, P. (2021). CRISP-DM twenty years later: From data mining processes to data science trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 33(8), 3048–3061. <https://doi.org/10.1109/TKDE.2019.2962680>

- Mienye, I. D., & Jere, N. (2024). Deep learning for credit card fraud detection: A review of algorithms, challenges, and solutions. *IEEE Access*, 12, 96893–96910. <https://doi.org/10.1109/ACCESS.2024.3426955>
- Moro, S., Cortez, P., & Rita, P. (2014). A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, 62, 22–31. <https://doi.org/10.1016/j.dss.2014.03.001>
- Nayak, S. (2024). Enhancing credit risk assessment using machine learning: A case study on early payment risk prediction.
- Noriega, J. P., Rivera, L. A., & Herrera, J. A. (2023). Machine learning for credit risk prediction: A systematic literature review. *Data*, 8(11), 169. <https://doi.org/10.3390/data8110169>
- Osei, F., Ampomah, G., Kankam-Kwarteng, C., Opoku Bediako, D., & Mensah, R. (2021). Customer satisfaction analysis of banks: The role of market segmentation. *Science Journal of Business and Management*, 9(2), 126. <https://doi.org/10.11648/j.sjbm.20210902.19>
- Özdemir, G. A. (2021). Dijital bankacılıkta müşteri deneyiminin öncüllerinin ve ardıllarının analizi [Ph.D.]. Retrieved from <https://www.proquest.com/docview/2637686536/abstract/C61D402B481E4562PQ/1>
- Plotnikova, V., Dumas, M., & Milani, F. P. (2022). Applying the CRISP-DM data mining process in the financial services industry: Elicitation of adaptation requirements. *Data & Knowledge Engineering*, 139, 102013. <https://doi.org/10.1016/j.datak.2022.102013>
- Probesto. (2024). Etkili müşteri segmentasyonu: Yapay zekanın gücünü açığa çıkarma – Probesto. Retrieved from <https://www.probesto.com/tr/etkili-musteri-segmentasyonu-yapay-zekanin-gucunu-aciga-cikarma/>
- Rane, N., Choudhary, S., & Rane, J. (2023). Explainable artificial intelligence (XAI) approaches for transparency and accountability in financial decision-making (SSRN Scholarly Paper No. 4640316). *Social Science Research Network*. <https://doi.org/10.2139/ssrn.4640316>
- Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., & Müller, K.-R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE*, 109(3), 247–278. <https://doi.org/10.1109/JPROC.2021.3060483>
- Sandhya Kona, S. (2020). Customer segmentation and personalization in banking services: Investigating the use of big data analytics to segment banking customers based on their behavior, demographics, and preferences, and leveraging these insights to personalize banking services and marketing campaigns. *International Journal of Science and Research (IJSR)*, 9(8), 1566–1570. <https://doi.org/10.21275/SR24522131706>
- Shi, S., Tse, R., Luo, W., D'Addona, S., & Pau, G. (2022). Machine learning-driven credit risk: A systemic review. *Neural Computing and Applications*, 34(17), 14327–14339. <https://doi.org/10.1007/s00521-022-07472-2>
- Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R., & Jr, K. C. L. (2017). *Data mining for business analytics: Concepts, techniques, and applications in R*. John Wiley & Sons.
- Smeureanu, I., Ruxanda, G., & Badea, L. M. (2013). Customer segmentation in private banking sector using machine learning techniques. *Journal of Business Economics and Management*, 14(5), 923–939. <https://doi.org/10.3846/16111699.2012.749807>

- Verhoef, P., Donkers, B., Langerak, F., Leeflang, P., & Lemon, L. (2003). Understanding the effect of customer relationship management efforts on customer retention and customer share development. *Journal of Marketing*, 67, 30–45. <https://doi.org/10.1509/jmkg.67.4.30.18685>
- Wang, S., Asif, M., Shahzad, M. F., & Ashfaq, M. (2024). Data privacy and cybersecurity challenges in the digital transformation of the banking sector. *Computers & Security*, 147, 104051. <https://doi.org/10.1016/j.cose.2024.104051>
- Xiang, S., Zhu, M., Cheng, D., Li, E., Zhao, R., Ouyang, Y., Chen, L., & Zheng, Y. (2023). Semi-supervised credit card fraud detection via attribute-driven graph representation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(12), 14557–14565. <https://doi.org/10.1609/aaai.v37i12.26702>
- Zaki, A. M., Khodadadi, N., Lim, W. H., & Towfek, S. K. (2024). Predictive analytics and machine learning in direct marketing for anticipating bank term deposit subscriptions. *American Journal of Business and Operations Research*, 11(1), 79–88. <https://doi.org/10.54216/AJBOR.110110>
- Zhuang, Q. R., Yao, Y. W., & Liu, O. (2018). Application of data mining in term deposit marketing. *Hong Kong*.