

Bankacılıkta Bireysel Müşteriler için Finansman Ürünlerinin Makine Öğrenmesi Algoritmaları ile Önerilmesi

¹Bulut Karadağ, ^{*1}Oğuzhan Bölükbaş, ^{*1}Mahmut Ahmet Ünal
¹Vakıf Katılım Bankası Ar-Ge Merkezi – Türkiye

Özet

Banka müşterileri talep ettikleri ürün karşılığında bankaya kısa, orta veya uzun vadeli borçlanarak finansman ürünü kullanmaktadır. Bankaların müşterilerine sundukları bu finansman ürünleri bankaların temel gelirlerini oluşturmaktadır. Eğer banka doğru finansman ürünlerini müşterilerine doğru sistemle önerebilirse gelirlerini yükseltme ihtimali daha da artacaktır. Kullanılan algoritma tabanlı öneri sistemlerinde yalnızca en yüksek ihtimale sahip ürün müşteriye önerilmektedir fakat bu durumda müşterinin kullanması olası en yüksek ihtimale sahip ikinci ve üçüncü ürün gözden kaçırılmaktadır. Eğer bu ihtimaller birbirine daha da yakınsa ve müşteriye sadece en yüksek ihtimale sahip ürün önerilip ikinci ve üçüncü üründen bahsedilmezse, ilk ürünü reddeden müşterinin diğer ürünleri alma ihtimali yüksekken bu tamamen gözden kaçırılmış olacaktır. Yapılan bu çalışmada, Vakıf Katılım Bankası müşterilerinin anonimleştirilen verileri kullanılarak gözetimli makine öğrenmesi algoritmalarından olan Naïve Bayes, rastgele orman sınıflandırıcı (Random Forest Classifier) ve k-en yakın komşu (KNN) algoritmaları ile geliştirilen modeller eğitilecektir ve eğitilen modeller ile en yüksek ihtimale sahip üç ürünün çıktığı olarak alınması sağlanacaktır. Modellerin sonuçları kıyaslanıp en yüksek doğruluğa sahip model kullanılacaktır. Eğitilen makine öğrenmesi modelleri ile bankanın sattığı finansman ürünlerinin satış miktarlarının, dolayısıyla banka gelirinin artması beklenmektedir.

Anahtar kelimeler: Perakende Bankacılık, Finansal Ürün Önerme, Ürün Tahmini, Naïve Bayes, Rastgele Orman, K-En Yakın Komşu

Abstract

Bank customers could use financing products in return for the product which they request by borrowing the bank for short, medium or long term. These financing products offer by banks to their customers constitutes the basic income of their own. If the bank could offer the right financing products to its customers with the proper system, the probability of rolling up its revenues would increase even more. The most likely product is recommended to the customer in the algorithm-based suggestion systems which is used. However, the second and third product with high probability of use by the customer is not recommended in this case. If these possibilities are even closer to each other, and if the customer is only offered the most likely product without recommended the second and the third product, the customer who rejects the first product is likely to use the other products which would be completely missed. In this study, models developed by Naïve Bayes, Random Forest Classifier and k-nearest neighbour (KNN) algorithms, which are supervised machine learning algorithms, would be trained using the anonymized data of Vakıf Participation Bank customers and with these trained models, the three most likely products would be output. The model with the highest accuracy rate would be used by comparing the outputs of these models. It is expected that the sales amount of the financing products sold by the bank would increase with the machine learning models therefore the bank income would increase.

Key words: Financial product proposition, Product forecasting, Retail banking

1. Giriş

Günümüz bankacılığında teknolojik gelişmelerin artışına paralel olarak hizmet kalitesi de artmıştır. Bankalar müşterilerine ATM, mobil uygulama ve internet şubesi kanalları ile hızlı ve güvenilir bir şekilde bankacılık hizmetlerini sunmaktadır. Giderek artan rekabetçi ortamda müşteriler; bankalar için artık daha da vazgeçilmezdir. Kendi bankaları ile çalışmayı bırakıp başka bankalara gitmek istemezler. Bunun için bankalar daha iyi hizmet sunma gayretiyle daha fazla müşteri kazanmak, hâlihazırdaki müşterilerini de kaybetmemek için çalışırlar.

Daha fazla müşteri kazanmanın yollarından biri müşteriye daha fazla ürün sunmaktan geçer. Sadece bireysel değil, ticari ve KOBİ ölçeğindeki firmalara sunulacak finansman ürünlerin çeşitliği de müşterilerin banka ile olan bağınu kuvvetlendirir. İnsanlar herhangi bir kredi ihtiyacı olduğunda, öncelikli olarak müşterisi olduğu banka üzerinden teminine çalışır. Eğer banka ürün çeşitliği ile birlikte iyi hizmet de sunabilirse müşteri memnuniyeti sağlanmış olur ve bunun da müşteri sadakatini olumlu yönde etkilediği söylenebilir [1].

Son yıllarda yapay zekâ konusunda yapılan çalışmaların artması finans sektörünü de olumlu yönde etkilemiştir ve bu alanda da yeni çalışmaların yapılmasına neden olmuştur. Bu çalışmalar genellikle bankaya ait risklerin ön tespiti ve önlenmesi, başarısızlıktan doğan maliyetleri azaltma konularında yapılmaktadır [2]. Özellikle bankalara ait risklerin erken uyarı sistemleri ile tespit edilmesinde, müşterilere ait kredi tahsis işlemlerinin yapılmasında ve mobil uygulama, internet şubesi ve ATM cihazları üzerinde oluşabilecek dolandırıcılıkların önlenmesi makine öğrenmesi kullanılmaktadır.

İlk olarak 1990'lı yılların ortalarında araştırmacıların bağımsız bir alan olarak ele aldığı ve üzerinde çalışmalar yaptığı öneri sistemleri [3], günümüzde birçok popüler çevrimiçi internet sistemi ile birlikte çeşitli amaçlarla etkin bir biçimde kullanılmaktadır.

Bir müşteri çalıştığı banka ile birçok finansman ürün ilişkisi kurmaktadır. Örneğin, araç kredisi kullanan bir müşteri daha sonra çek karne ürünü de kullanabilir. Bu sebepten dolayı bir müşterinin kullanabileceği potansiyel ürünleri önceden tespit edip sunmak hem müşteri memnuniyetini hem de banka karlılığını arttırmaktadır. Yapılan bu çalışmanın amacı henüz ürün kullanmamış müşterilere ve daha önce ürün kullanmış müşterilere kullanmaları en muhtemel üç ürünü doğru şekilde önermektir.

Çalışmaya konu olan veriler Vakıf Katılım Bankasına ait bireysel müşteri verilerinden örnek alınarak sağlandı. Bu veriler anonimleştirilip Kişisel Verileri Koruma Kanunu (KVKK) kapsamında çıkarıldı. Bu veri kümesinde bireysel müşterilerin kullandığı ürünlerin ve müşteri bilgilerini içeren 25,274 kayıt bulunmaktadır.

2. Önerilen Yöntem

Banka, istatistiksel hesaplama tabanlı algoritmalar kullanarak müşterisine finansman ürünü önerdiğinde çıktı tekil olmaktadır. Fakat müşteriler sadece bir finansman ürünü değil birden fazla

finansman ürünü de talep edebilmektedir veya kullanması en muhtemel olarak görülen ürünü değil de en muhtemel ikinci veya üçüncü ürünü de kullanabilmektedir. Müşteri bilgilerine göre birden çok finansman ürünü önerebilmesi açısından seçilen algoritmaların çıktı şekilleri değiştirildi. Örneğin Naïve Bayes algoritmasının tüm çıktı sınıflarına ait olasılıkları döndürmesi sağlanıp en yüksek olasılıktaki üç ürün alındı. Rastlantısal orman sınıflandırıcı algoritmasının yalnızca en çok oy alan ürünü değil onunla beraber oylamada ikinci ve üçüncü sıradaki çıktıları da dönmesi sağlandı. K – en yakın komşu algoritmasında ise en yakın üç komşuyu belirtecek şekilde çıktı vermesi sağlandı. En iyi sonuç veren algoritma seçilip oransal olarak en yüksek ihtimale sahip üç ürün çıktı olarak ürettirilip banka tarafından müşteriye sunulacaktır.

2.1. Algoritmalar

2.1.1. Naïve Bayes

Önerilen yöntemlerden biri olan Naïve Bayes algoritmasının temeli Bayes teoremine dayanır. Bayes teoremi 1812 yılında Thomas Bayes tarafından bulunan koşullu olasılık hesaplama formülüdür. Bu teorem bir rastlantısal değişken için olasılık dağılımı içinde koşullu olasılıklar ile marjinal olasılıklar arasındaki ilişkiyi gösterir [4]. Bayes teoremine ait eşitlik, Denklem 1 de sunuldu.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

$P(A)$ = A olayının gerçekleşme olasılığı

$P(B)$ = B olayının gerçekleşme olasılığı

$P(A|B)$ = B olayı gerçekleştiğinde A olayının gerçekleşme olasılığı

$P(B|A)$ = A olayı gerçekleştiğinde B olayının gerçekleşme olasılığı

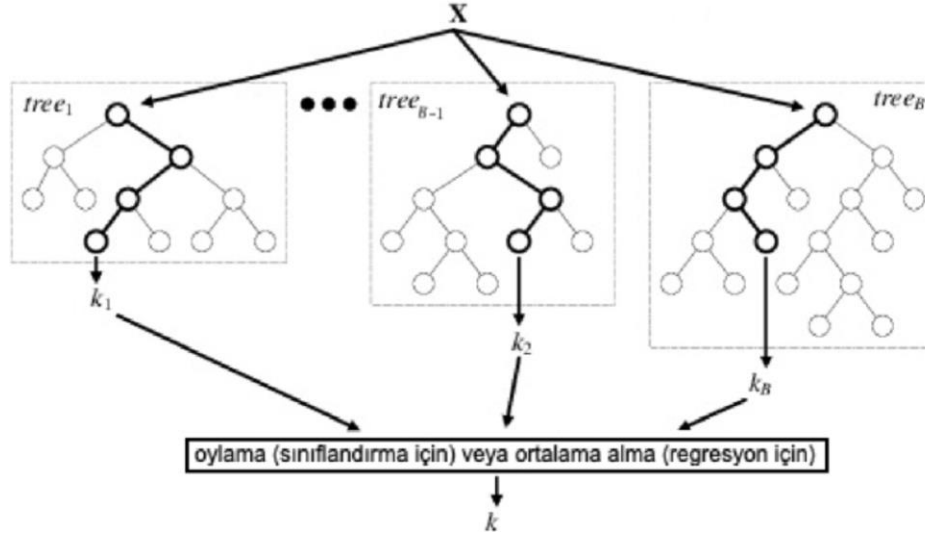
Her ürünün/örneğin n adet özneliğe sahip olduğu bir veri kümesi üzerinde çalışıldığı düşünüldüğünde ve hangi sınıfa ait olduğu bilinmeyen yeni bir ürün/örnek sınıflandırılmak istendiğinde, Bayes denklemi kullanılarak, örneğin/ürünün verilen m sınıfa ait olma olasılığı Denklem 2 yardımıyla hesaplanabilir.

$$P(y|x_1, \dots, x_n) = \frac{P(x_1|y)P(x_2|y)\dots P(x_n|y)P(y)}{P(x_1)P(x_2)\dots P(x_n)} \quad (2)$$

$P(y)$ ve $P(x_i|y)$ ifadelerinin değerleri veri kümesi kullanılarak hesaplanabilir. Bir örneğin/ürünün hangi sınıfa ait olduğunu tespit edebilmek için sistemde tanımlı olan her bir sınıf için bu değer hesaplanır. Daha sonra, en yüksek olasılığa sahip olan sınıf, o ürünün/örneğin ait olduğu sınıftır çıkarımında bulunulur [5].

2.1.2. Rastgele Orman Sınıflandırma

Rastgele orman veya rastgele karar ormanları eğitim aşamasında çok sayıda karar ağacı oluşturarak sınıflandırma, regresyon ve diğer görevler için, problemin tipine göre sınıf (sınıflandırma) veya sayı (regresyon) tahmini yapan bir toplu öğrenme yöntemidir. Rastgele karar ormanları ile karar ağaçlarının veri kümesine aşırı uyma (overfitting) problemlerinin önüne geçilmektedir. Şekil 1’de rastlantısal ormanların genel yapısı görülmektedir.



Şekil 1: Rastlantısal Orman Yapısı

Rastgele ormanlar sınıflandırma problemlerinde oylama sonucu olarak en yüksek oy alan sınıfı çıktı olarak döndürmektedir. Yapılan değişiklikler ile bu algoritma bahsedilen probleme uyarlanarak en yüksek oy alan üç ürünü döndürecek şekilde düzenlenmiştir.

2.1.3. K-En Yakın Komşu

KNN algoritması 1967 yılında T. M. Cover ve P. E. Hart tarafından önerilmiştir. Algoritma, sınıfları belli olan bir örnek kümesindeki verilerden yararlanılarak kullanılmaktadır. K-NN parametrik olmayan, tembel (lazy) bir öğrenme algoritmasıdır. Eager öğrenme aksine tembel öğrenmenin bir eğitim aşaması yoktur. Eğitim verilerini öğrenmez, veri kümesini “ezberler”. Bir tahmin yapmak istediğimizde, tüm veri setinde en yakın komşuları arayarak sonuca ulaşmaktadır.

K-en yakın komşu algoritmasının çalışmasında bir k değeri belirlenir. Bu k değeri; bakılacak eleman sayısını ifade etmektedir. Yeni bir girdi değeri için en yakın k kadar eleman alınarak gelen değer arasındaki uzaklık hesaplanır. Uzaklık hesaplamada genellikle öklid fonksiyonu kullanılır. Öklid fonksiyonuna alternatif olarak “Manhattan”, “Minkowski” ve “Hamming” uzaklık hesaplama fonksiyonları da kullanılabilir. Uzaklık hesaplandıktan sonra hesaplanan değerler sıralanır ve yeni girdi değeri kendisine en yakın sınıfa atanır.

K-en yakın komşu algoritması da diğer algoritmalar gibi düzenlenmiştir. Yeni bir girdi modele verildiğinde yalnızca en yakın sınıfı değil, en yakın üç sınıfı orantısal hesaplamalarıyla beraber sonuç olarak vermektedir.

2.2. Veri Toplama

Çalışmaya konu olan 25.274 satır veri Vakıf Katılım Bankasının bireysel müşterilerinin kayıtlarından anonimleştirilerek oluşturulmuştur. Şekil 2’de veri kümesinin sütunları ve ilk üç satırı gösterilmiştir. Girdi verileri “Yaş”, “Cinsiyet”, “Meslek”, “Eğitim Seviyesi”, “KKB (Kredi Kayıt Bürosu) Skoru”, “Ev Sahibi Mi?”, “Araç Sahibi Mi?”, “Çocuğu Var Mı?” ve “Şahıs Firması Sahibi Mi?” alanlarından oluşmaktadır. Bu verilerin seçilmesinin sebebi, bireysel otomatik kredi tahsis yöntemlerinde de bankalar tarafından benzer girdilerin kullanılmasındır [6].

	Yaş	Cinsiyet	Meslek	Eğitim Seviyesi	KKB Skoru	Ev Sahibi Mi?	Araç Sahibi mi?	Çocuğu Var Mı?	Şahıs Firması Sahibi Mi?	Ürün Adı
0	63	Erkek	Bankacı	Lisans	1361	Hayır	Evet	Hayır	Hayır	Tüketici-Gayrimenkul Kredisi
1	50	Erkek	Danışman	Doktora	1846	Hayır	Evet	Hayır	Hayır	Tüketici-Gayrimenkul Kredisi
2	45	Kadın	Bankacı	Lisans	1611	Hayır	Evet	Hayır	Hayır	Tüketici-Gayrimenkul Kredisi

Şekil 2: Çalışmada kullanılacak olan veri kümesinin ilk üç satırının gösterimi

Türkiye Cumhuriyeti vatandaşı olan her birey, girdi verilerinde de yer alan KKB skoruna sahiptir. KKB skoru bireylerin bankalar ile daha önceki ilişkileri değerlendirilerek hesaplanan bir değerdir. Finans kuruluşları müşterilerine ait risk, limit, çek, teminat vb. verilerini aylık olarak FTP (File Transfer Protokol) veya diğer çevrimiçi yollarla KKB’ye iletir [7]. KKB bu verileri işleyip skor hesabı yaparak hem bireysel müşterilerin hem finans kuruluşlarının çevrimiçi sorgu yapabilmesini sağlayan altyapıyı sunar. Bankalar ile daha önce ilişkisi bulunmayan bir kişinin KKB skoru varsayılan değer olan 0’dır. Bu skor en yüksek değer olarak 1900’e kadar yükselbilmektedir. Tablo 1’de KKB skor aralığının risk karşılıkları görülmektedir.

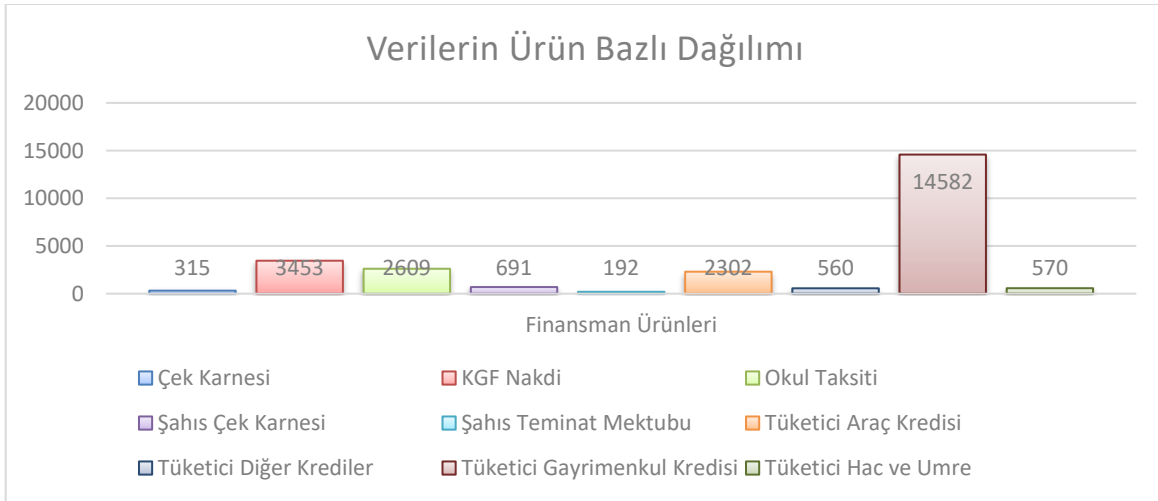
Tablo 1: Risk durumuna göre KKB skor tanımı

Skor Aralığı	Skor Tanımı
0-699	En Riskli
700-1099	Orta Riskli
1100-1499	Az Riskli
1500-1699	İyi
1700-1900	Çok İyi

Tabloda görüldüğü üzere KKB skoru yüksek olan kişilerin bankalar tarafından güvenilirliği “İyi” veya “Çok İyi” şeklinde nitelendirilir. Başvuruda bulunduğu finansman ürünlerinin banka tarafından değerlendirilip onaylanma ihtimali daha yüksektir. Bu skoru düşük olan kişiler ise “Orta Riskli” veya “En Riskli” müşteri statüsünde değerlendirilip bankalar tarafından başvurularında temkinli karşılanmaktadır.

Gözetimli makine öğrenmesi alt kümesini diğerlerinden ayıran özelliği; model eğitiminde kullanılacak veri kümesinde girdi verilerinin çıktı değerinin etiket bilgisi olarak veri kümesinde yer almasıdır. Bu çalışmada çıktı değerleri müşterinin kullandığı finansman ürününün ismidir. Bu ürünler “Tüketici-Gayrimenkul Kredisi”, “Tüketici-Diğer Krediler”, “Tüketici-Araç Kredisi”, “KGF Nakdi”, “Şahıs Çek Karnesi”, “Okul Taksit”, “Çek Karnesi”, “Şahıs Teminat Mektubu” ve “Tüketici-Hac ve Umre Kredisi”dir. Veri kümesinde bu finansman ürünlerini kullanan müşterilerin girdi bilgileri ile beraber etiket değeri olarak yer almıştır. Finansman ürünlerinin 9 tanesi ele alınmıştır, toplam müşteri talebi 10 adetten az olan ürünler ise çalışma dışında tutulmuştur.

Ürünler ve kullanım miktarları Şekil 2’de gösterilmiştir. Veri kümesinde kayıp verilerin olmamasına dikkat edildi. Çalışmada kullanılan verilerdeki tüm alanlar tam ve eksiksizdir, kayıp veri durumu bulunmamaktadır.



Şekil 2: Verilerin ürün bazlı dağılımı

2.3. Veri Ön İşleme

Veri kümesinde sayısal ve kategorik veriler bir arada yer aldığından veri kümesi ön işleme tâbi tutularak yukarıda bahsedilen makine öğrenmesi modellerinin işleyebileceği hâle getirilmesi amaçlanmıştır. Veri setinde sayısal verilerin yer aldığı “Yaş” ve “KKB Skoru” sütunları veri kümeden ayrılmıştır. Kategorik veriler ise kukla değişkenler (dummy variables) ile farklı sütunlara bölünmüştür. 7 adet olan kategorik değişken sütunları bu işlem ile 79 adede yükselmiştir. Daha önce ayırdığımız iki adet kesikli sayılardan oluşan “Yaş” ve “KKB Skoru” sütunları ise normalleştirme (normalisation) işleminin ardından kategorik verilere eklenmiştir. Veri ön işleme sonucunda 81 sütuna sahip 25274 satır sayısal veriden oluşan veri kümesi elde edilmiştir.

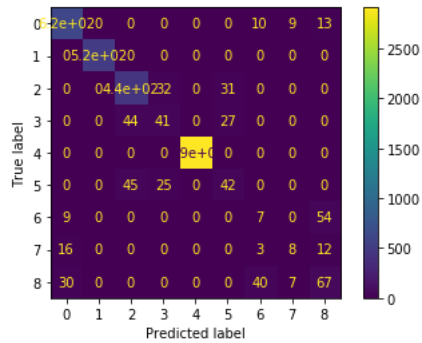
2.4. Modelleme

Bu çalışmada makine öğrenmesinin gözetimli öğrenme alt kümesinde yer alan sınıflandırma algoritmalarından olan Naïve Bayes (NB), Rastgele Orman Sınıflandırıcı (RFC) ve k-en yakın komşu (kNN) algoritmaları kullanılmıştır. Algoritmalar “sklearn” kütüphanesi yardımı ile Python programlama dili kullanılarak eğitilmiştir. Veri kümesi eğitim ve test olmak üzere iki kısma ayrılmıştır ve %70’lik kısım ile modeller eğitilmiştir. Eğitim sonucunda RFC ile 0.92, NB ile 0.78, KNN ile 0.71 doğruluk değerlerine ulaşılmıştır.

Seçilen RFC modeline 50 yaşında, erkek, mühendis, lisans mezunu, arabası ve çocuğu olan, 1600 KKB skoruna sahip, evi ve kendi şirketi olmayan bir müşterinin bilgileri test girdisi olarak verilmiştir. Modelin ürettiği sonuç ile bu özelliklerdeki bir müşterinin %77,63 oranında “Tüketici-Gayrimenkul Kredisi”, %16,19 oranında “Okul Taksit”, %6,18 oranında “Tüketici-Araç Kredisi” ürünü kullanabileceği sonucu elde edilmiştir. Bu çalışma şu an kullanılan sistemlerle gerçekleştirilseydi sonuç olarak sadece “Tüketici-Gayrimenkul Kredisi” önerisi elde edilecekti fakat eğitilen modeller ile en muhtemel ürünler oransal olarak elde edilmiş oldu.

3. Bulgular ve Tartışma

Bu çalışma sadece bireysel müşteriler için gerçekleştirildi. Toplam 25,274 satır veri %70’i eğitim için %30 u test için kullanıldı. Modellerin kullanabileceği şekilde rakamsallaştırılan verilen ile eğitilen naive Bayes, rastlantısal orman sınıflandırıcı ve k-en yakın komşu algoritmalarından en yüksek doğruluk değeri veren model rastlantısal orman sınıflandırıcı modeli olup kullanıma alınmıştır. Rastlantısal orman sınıflandırıcı modelinin test sonuçlarına ait hata matrisi Şekil 3’te gösterilmiştir. Çalışmanın en önemli faydası müşteriye sadece bir kredi ürünü önermek yerine oransal olarak ürün çeşitliliği sunmaktadır.



Şekil 3: Rastlantısal orman sınıflandırıcı modeline ait hata matrisi

4. Sonuçlar ve Gelecekteki Tartışmalar

Her geçen gün ilerleyen katılım bankacılığında faizsiz finansal ürünlerin ve hizmetlerin geliştirilmesi ile sektörde işlem hacminin artırılması amaçlanmaktadır [8]. Bankaların bireysel müşterilere sunduğu ürün çeşitliliği KOBİ ve ticari müşterilere nazaran daha azdır ve müşterilerin finansman talepleri de genellikle belirli ürünler üzerinde yoğunlaşmıştır. Ancak bu çalışma KOBİ

ve ticari müşteriler için de gerçekleştirildiğinde daha fazla girdi verisine ve ürün çıktısına ihtiyaç duyulacaktır. Örneğin, ticari firmanın ortakları, firmanın ticari kredi notu, firmanın sektörü, firmanın alt sektörleri gibi daha fazla girdi verisi bulunmalıdır. Bu çalışmadan yola çıkılarak KOBİ ve ticari müşteriler için de benzer çalışma yapılabilir. Bireysel müşterilere göre KOBİ ve ticari firmalardan elde edilen kar daha yüksek olduğu için bu tip müşteriler için çalışma daha anlamlı olmaktadır.

Bu çalışmada örnek olması açısından KKB skoru tekil olarak ele alındı. Hâlbuki KKB'den daha detaylı verilere de ulaşılabilir. "En Kötü Ödeme Durumu", "Kanuni Takibe Düşmüş Mü," "Ödeme Performansı", "Kefil Durumu" Ve "Hane Halkı Bilgileri" bu verilere örnek gösterilebilir. Eklenebilecek bu yeni veri çeşitleri ile daha yüksek doğruluk değerlerine ulaşabilen makine öğrenmesi modellerinin eğitilebilmesi mümkündür.

5. Referanslar

- [1] Kuzu, B. (2018). The effect of service quality on customer satisfaction in mobile banking. (Unpublished master's thesis). Dogus University Institute of Social Sciences, Istanbul.
- [2] Gary Whalen, 1991. "A proportional hazards model of bank failure: an examination of its usefulness as an early warning tool," *Economic Review*, Federal Reserve Bank of Cleveland, vol. 27(Q I), pages 21-31.
- [3] D. N. D. O. B. M. a. T. D. Goldberg, "Using collaborative filtering to weave an information Tapestry," *Communications of the ACM*, vol. 35, no. 12, pp. 61-70, 1992.
- [4] J. Joyce, "Bayes Theorem," {Metaphysics Research Lab, Stanford University, The Stanford Encyclopedia of Philosophy, 2019.
- [5] E. YALÇIN, Naïve Bayes Sınıflandırıcı Tabanlı İkili-Veri Çoklu-Ölçütlü Öneri Sistemleri, Eskişehir: Eskişehir Anadolu Üniversitesi Fen Bilimleri Enstitüsü, 2016.
- [6] M. Çinko, "Kredi Kartı Değerlendirme Tekniklerinin Karşılaştırılması," *İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi*, vol. 5, no. 9, p. 149, 2006.
- [7] Risk Merkezi, "Krm Sadeleştirme," 27 4 2017. [Online]. Available: <https://www.riskmerkezi.org/tr/risk-merkezi/>.
- [8] Ş.Y. Yiğiter, S. S. Sarı, T. Karabulut, E. E. Başakın "Kira Sertifikası Fiyat Değerlerinin Makine Öğrenmesi Metodu ile Tahmini", *Uluslararası İslam Ekonomisi ve Finansı Araştırmaları Dergisi*, Kasım 2018, Cilt:4, Sayı: 3.