

KREDİ KARTI DOLANDIRICILIK TESPİTİNİN MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE TAHMİN EDİLMESİ

Melike SELİMOĞLU*, Atınç YILMAZ**

ÖZET

Kredi kartları, dünya üzerindeki yaygınlığı ve güçlü altyapısı sayesinde çok kısa bir sürede insanların hayatlarına girmiş ve korkusuzca kullanabilecekleri ödeme aracı haline gelmiştir. Kredi kartlarının sayılarının her geçen gün takip edilmesi zor bir hal alması ve bunun paralelinde işlem hacminin hızla büyümesi, bu pazardan haksız kazanç elde etmek isteyen dolandırıcıları ortaya çıkarmıştır. Günümüzde herhangi bir insanın kredi kartı bilgilerini elde etmenin oldukça kolay hale gelmesi kredi kartı dolandırıcılarının işini kolaylaştırmaktadır. Gelişen teknoloji sayesinde, gerçekleşen hesap hareketleri değişikliğinde zaman ve harcamaların analiz edilmesi ile kötü amaçla elde edilen verilerin kullanıldığı analiz edebilir. Kaggle veritabanından elde edilen Kredi Kartı Dolandırıcılık Teşhis veri seti kullanılarak Çok Katmanlı Yapay Sinir Ağı ve Naive Bayes yöntemleri ile modelleme yapılmıştır. Bu çalışmada ki amaç kişilerin kredi kartlarını kullanma zaman aralıklarını analiz ederek yapılan işlemin farklı kişi tarafından yapıldığını tespit etmektir. En yüksek başarı oranı çok katmanlı yapay sinir ağı ile %99,943 elde edilmiştir. Naive Bayes yöntemi ile %98,207 başarı elde edilmiştir.

Anahtar Kelimeler: Kredi Kartı dolandırıcılık tespiti, Yapay Sinir Ağı, Naive Bayes, Makine öğrenmesi

*Makalenin Gönderim Tarihi: 03/02/2021 Makalenin Kabul Tarihi: 23/02/2021 Makale Türü: Araştırma

DOI: 10.20854/bujse.873804

*Sorumlu yazar: Beykent Üniversitesi Mühendislik-Mimarlık Fakültesi Bilgisayar Mühendisliği Bölümü, İstanbul. melikeselimoglu@gmail.com

**Beykent Üniversitesi Mühendislik-Mimarlık Fakültesi Bilgisayar Mühendisliği Bölümü, İstanbul. atincyilmaz@beykent.edu.tr (ORCID ID: 0000-0003-0038-7519)

1. GİRİŞ

Kredi kartları, bankaların hizmette bulunduğu kişilere verdiği, alışveriş noktalarında kullanılan POST cihazları sayesinde ödeme yapılabilmektedir. Aynı zamanda ATM'lerden nakit para çekilebilmektedir. Kredi kartları insanların hayatlarını yapılan harcamaların taksitlendirilmesi konusunda da kolaylaştırmaktadır. Bu sayede insanlar tek seferde ödemek yerine belirli ay sayısına bölerek aylık harcama miktarını azaltmaktadır. [1]

Kredi kartları, dünya üzerindeki yaygınlığı ve güçlü altyapısı sayesinde çok kısa sürede insanların kolayca ve sıkça kullanabilecekleri bir ödeme aracı haline gelmiştir. Bu duruma ek olarak kredi kartlarının kullanım alanlarının artması ve insanların tercih etme nedenlerinin çoğalması ile birçok sorunda beraberinde olmuştur.

İnsanların bu kadar çok kullandıkları kredi kartı, bilgilerinin başka insanların eline geçmesi ve kötü amaçlı kullanılması oluşan sorunların en önemlisidir. Kredi Kartı dolandırıcılıkları, var olan kartın birebir şekilde yeni bir karta kopyalanması ile olabileceği gibi yine var olan kart üzerindeki bilgilerin e-ticaret sitelerinden çalınarak kartın sahibi gibi kullanılması veya içerisinden para aktarılması gibide olabilir.

Dünya üzerindeki bütün ülkeler kredi kartı dolandırıcılığı konusunda büyük finansal kayıplar vermektedir. Bu nedenle çalışmada kredi kartı hareketlerinden elde edilen veriler kullanılarak belirli analizler yapılmakta ve bu analiz sonucunda kredi kartı dolandırıcılığının önüne geçilmek hedeflenmektedir.

1.1 Literatür Araştırması

E. Özbay tarafından 2007 yılında hazırladığı Finans Sektöründe Veri Madenciliği ile Dolandırıcılık Tespiti konulu tezde Saf Bayes Sınıflandırıcı algoritması kullanılmıştır. Çalışmanın sonunda %72,3077 doğruluk oranı elde edilmişken %27,6923 hata oranı elde edilmiştir. Veri setinde 196 adet veri bulunmaktadır. [2]

B. Adalı tarafından 2012 yılında hazırlanan Kredi Kartı Dolandırıcılığı Saptama Yöntemleri bitirme ödevinde Yapay Sinir Ağı yöntemi kullanarak yaptığı analizde %99,2 doğru sınıflandırma oranını elde etmiştir. [3]

S. Gültekin tarafından 2017 yılında hazırlanan Fiyat Tahmini Tezinde Doğrusal Regresyon algoritması ile Yapay Sinir Ağı algoritması kullanılarak 20 markanın her biri için ayrı değerlendirme yapılmaktadır. Bu değerlendirmelerin ortalamasını analiz ederek %93 doğruluk oranı elde edilmiştir. [4]

K. Soylu tarafından 2018 yılında hazırlanan Kredi Kartı Sahte İşlem Tespiti tezinde kredi kartı işlem veri kümesi üzerinde çalışılmıştır. Veri kümesinin %70'ini eğitim kümesi ile model eğitilmiş, %30'una karşılık gelen test kümesi ile model başarısı test edilmiştir. Bu verilerin ışığında, derin öğrenme modelinin sahte işlem tespit etme oranı %78,7'dir. Aynı veri kümesi üzerinde Rastgele orman modeli ile de çalışılarak, doğruluk hata oranını 0,00008 olarak tespit etmiştir. [5]

E. Özmen ve T. Özcan tarafından 2019 yılında yayımlanan makalede Finans Sektörü Dolandırıcılık Tespiti üzerine Regresyon Ağacı Yöntemi ve Melez Sınıflandırma kullanarak analiz etmişlerdir. Yapılan çalışmada elde edilen gözlemler sonucunda Matlab platformunun R2014a sürümü ile kullanılacak algoritmanın sonucunun iyi anlaşılabilmesi açısından CART ve GA-CART algoritmaları çözümlenmiştir. CART algoritmasında %64,28 doğruluk oranı verirken, GA-CART algoritmasında %87,95 doğruluk oranı elde edilmiştir. [6]

2. YÖNTEM VE MATERYALLER

Kredi Kartı Dolandırıcılık tespitinin yapılmasında yardımcı olacak veri seti ve kullanılan yöntemler bu bölümde anlatılmıştır.

2.1 Veri Seti

Çalışmada kullanılan veri seti Kaggle veritabanından elde edilmiştir. Veri setinin özellikleri ve açıklamaları aşağıda belirtilmiştir. [8]

Veri setleri, Avrupa kart sahipleri tarafından Eylül 2013'te kredi kartlarıyla yapılan işlemleri içermektedir. İlgili veri kümesi, iki gün içinde gerçekleşen ve 284.807 işlemden 492'sini dolandırıcılık yapılan işlemleri sunmaktadır. Pozitif sınıf (dolandırıcılık) tüm işlemlerin %0.172'sini oluşturmaktadır.

Yalnızca PCA dönüşümünün sonucu olan sayısal girdi değişkenlerini içerir (Şekil 1). Gizlilik sorunları nedeniyle orijinal özellikleri ve veriler hakkında daha fazla bilgi toplanamamıştır.

- V1, V2, ..., V28 PCA ile elde edilen ana bileşenlerdir.
- PCA ile dönüştürülmemiş özellikler 'Zaman' ve 'Miktar' dir.
- 'Zaman' özelliği, her işlem ile veri kümesindeki ilk işlem arasında geçen saniyeyi içermektedir.
- 'Miktar' özelliği işlem tutarıdır, bu özellik maliyete duyarlı öğrenme için kullanılır.
- 'Sınıf' özelliği yanıt değişkenidir ve dolandırıcılık durumunda 1, aksi takdirde 0 değerini alır.

	A	B	C	D	E	F	G	H	I	J	K	L
Time	"0"	"1"	"2"	"3"	"4"	"5"	"6"	"7"	"8"	"9"	"10"	"11"
0	1.359807133678	0.07781133098487	2.51634673796914	1.378152427441	0.31832076994218	0.46218777762292	0.2397	0.1193871133486	0.266597205961	0.1664801335321	0.44815407846011	0.06001794922214
1	1.35854815982	1.301651050743605	1.77320942831150	3.7977699304328	0.5013811318191	1.8004998079263	0.7914	1.0962711522087	0.185216008082891	1.7929331957872	0.86329127503645	0.031008796030821
2	1.15823309399232	0.8773879488451	1.48717046112	0.4030393095121	0.407193377116510	0.09921642688256	0.5929	2.04259688442454	0.86052404882885	1.1411094322215	0.1682507976031	0.420688677219
3	1.22967614507931	1.0103507049326	0.04530773899448	1.2026127367394	0.19180808597645	0.2720812289908	0.005	0.44209424238184	1.4796545047385	1.0743803783554	0.491399518790515	0.489140047043157
4	0.8942868208242	0.2815719627654	0.11339212772971	0.271261308804	0.699989595983	0.72181806127751	0.3709	9.031826175242575	1.11959137615661	1.0443665517316	0.22218726738206	0.4993080649927
5	10.449047814715	1.1761386253096	0.9185981281276	1.375666649994	1.0713813545323	0.6291521389734	1.423	100.3897821218095	0.6161904917847	0.474297870295924	0.09401862967152	0.20548381783881
6	10.124999842053	1.2218368901816	0.3839315182291	1.3248868766892	1.4851947377961	0.732301164566149	0.6894	11.069378878819	0.28772129331455	0.8286127266428	0.7125242916718	0.17898016248009
7	2.7018476709339	0.3177075668061	0.647101665665	0.7674714348885	1.3058484442462	0.807964682632	0.4225	12.1032154328381	0.040296145973447	1.2673208898919	1.2899146962552	0.73599713604008
8	14.54012576615825	5.4501478342064	1.1803046143652	1.73621880021095	3.04910587764025	1.76340557365201	1.55973	15.1492939769862	1.02934571888487	0.45479473734366	1.4380287991702	1.50543410186344
9	16.01094887780717	1.38193191300901	1.0022103968013	1.0141979216176	1.19120879454965	1.30910881872952	0.878	17.0196249699914852	0.284610209212	0.17147905415064	0.19294067740311	1.395657126894
10	18.115661682482428	0.50120087854101	0.0673003143663332	2.815692349128	0.428804194630708	0.0894735167245992	0.2	18.024749127893665	0.27766527353081	1.8547084217971	0.09292549857044	1.3849397897709
11												

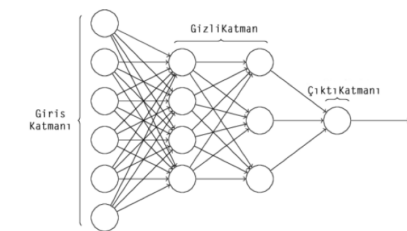
Şekil 1: Veri Seti Parçası

2.2 Yapay Sinir Ağ

Yapay sinir ağları, insan beyninin öğrenme kabiliyeti incelenerek geliştirilmiş bilgi işlem teknolojisi. Biyolojik nöronlar benzetimi ile ortaya çıkarılmış bir yapay zeka yöntemidir. Bilgisayarın öğrenme yetkisi ve eksik bilgi veya bilgisi olmadığı konu üzerinden yorum ve sonuç üretebilecek bir sistemdir. Yapay sinir ağları belirlenen sayıda yapay sinir hücrelerinin sistematik olarak birleşmesi sonucu ortaya çıkar.[9]

2.2.1 Çok Katmanlı Yapay Sinir Ağı

Çalışmada kullanılan algoritmaların biri çok katmanlı yapay sinir ağıdır. Çok katmanlı algılayıcılar (MLP) doğrusal olmayan problemlerin çözümlenmesi amacıyla ortaya çıkmıştır. Tek katmanlı algılayıcılar sadece doğrusal problemleri çözebilirler. Fakat MLP ağları doğrusal olmayan bağlantılar sayesinde doğrusal olmayan problemlerin de çözümünü sağlamaktadır. Birçok farklı çok katmanlı yapay sinir ağı modeli bulunmaktadır. Bu tür ağların bir kısmı denetimli, bir kısmı denetimsiz bir kısmı takviyeli öğrenme prensibi ile çalışmaktadır. MLP sınıflandırma problemlerinde daha iyi çalışmaktadır.[10]



Şekil 2: Çok Katmanlı Yapay Sinir Ağı

Şekil 2 de görülen mimarı çok katmanlı olup ilk katman giriş katmanıdır. Giriş katmanında kaç input var ise o kadar nöron olması gerekmektedir. Arada kalan katmanlar gizli katmanlardır. Gizli katmanlar bir veya birden fazla olabilirler. Katmanlarda bulunan nöronları birbirine bağlayan yapıya ağırlık denmektedir. Ağırlık, bağlantılı olduğu nöronun önem durumunu ortaya koyan katsayidir. Ara katman sayısına modelleyen kişi karar verir. Son katman ise çıktı katmanıdır.

Çalışmada ki modelde kullanılan yöntemlerden biri çok katmanlı yapay sinir ağıdır. Eğitim'de veri setinin %70 kullanılırken test içinde veri setinin %30'u kullanılmıştır. Maksimum iterasyon sayısı 100' dur. Kullanılan modelde 10 gizli katman bulunmakta iken; her gizli katman için 100'er nöron kullanılmıştır.

2.3 Naive Bayes

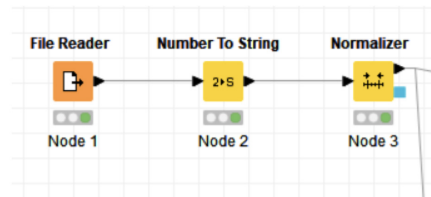
Naive Bayes algoritması olasılık hesaplarına göre belirlenmiş bir sınıflandırma hesaplaması ile sisteme verilen sınıfın yeni kategorisini tespit etmeyi hedeflemektedir.[11]

Naive Bayes yöntemi ulaşılmış istenen hedef etiket ile problemde uygulanan giriş parametreleri arasındaki ilişkiyi tahmin etmeye adapte olan sınıflandırma yöntemidir. Bu yöntem, bağımsız parametre ve bağımlı parametrelerin kombinasyonunun sıklığını hesaplayarak tahmin için bu olasılıkları kullanır.

Çalışmada ki modelde Naive Bayes yöntemi kullanılmıştır. Modelde eğitim için verinin %70'i kullanılmıştır. Eğitim setindeki veriler ile yapılan olasılık hesapları ile sisteme verilen test verileri tahmin için ayrılan %30'u kullanılarak verinin hangi sınıfta olacağını tahmini yapılmaya çalışılmıştır.

3. UYGULAMA

Çalışmada her iki yöntemin modellenmesi Knime üzerinden gerçekleştirilmiştir (Şekil 3).



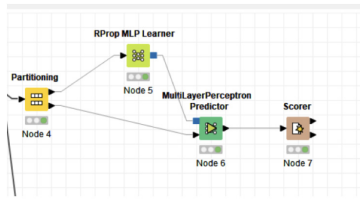
Şekil 3: Knime Modeli

Problem için modellenen çok katmanlı yapay sinir ağları yöntemi ve naive bayes yönteminde kullanılması için öncelikle veriler sayısallaştırılmıştır.

Daha sonrasında ise tüm veriler her iki model için daha verimli çalışma zamanı elde edilmesi için normalize edilmiştir (Şekil 4.)

D_1	D_11	D_12	D_13	D_14	D_15	D_16	D_17	D_18	D_19	D_110	D_111	D_112
0	0.930	0.766	0.861	0.710	0.767	0.268	0.267	0.766	0.479	0.111	0.262	0.6
0	0.936	0.771	0.844	0.712	0.766	0.262	0.261	0.766	0.464	0.105	0.261	0.7
0	0.830	0.763	0.868	0.268	0.762	0.261	0.27	0.768	0.411	0.113	0.262	0.7
0	0.946	0.763	0.868	0.114	0.766	0.276	0.261	0.768	0.411	0.108	0.262	0.7
0	0.936	0.777	0.864	0.17	0.767	0.264	0.268	0.762	0.461	0.104	0.258	0.7
0	0.936	0.777	0.867	0.144	0.767	0.267	0.268	0.768	0.461	0.101	0.262	0.7
0	0.936	0.769	0.858	0.303	0.767	0.268	0.267	0.766	0.479	0.107	0.261	0.6
0	0.946	0.769	0.868	0.17	0.772	0.267	0.272	0.766	0.464	0.105	0.261	0.6
0	0.947	0.777	0.855	0.14	0.764	0.3	0.268	0.765	0.469	0.1	0.263	0.7
0	0.937	0.776	0.856	0.242	0.767	0.261	0.269	0.766	0.477	0.101	0.261	0.7
0	0.967	0.765	0.853	0.181	0.772	0.277	0.27	0.766	0.464	0.102	0.261	0.6
0	0.967	0.774	0.822	0.248	0.765	0.266	0.268	0.761	0.464	0.113	0.27	0.6
0	0.968	0.774	0.844	0.187	0.776	0.275	0.261	0.761	0.461	0.108	0.268	0.6
0	0.976	0.777	0.852	0.172	0.765	0.266	0.263	0.767	0.465	0.108	0.259	0.7
0	0.976	0.764	0.866	0.17	0.767	0.272	0.261	0.765	0.469	0.103	0.258	0.7
0	0.946	0.771	0.875	0.187	0.776	0.262	0.262	0.765	0.468	0.104	0.258	0.6
0	0.977	0.767	0.863	0.189	0.761	0.268	0.267	0.767	0.46	0.107	0.259	0.7
0	0.951	0.777	0.853	0.17	0.772	0.262	0.27	0.766	0.44	0.103	0.263	0.7
0	0.967	0.77	0.868	0.158	0.766	0.248	0.268	0.767	0.463	0.108	0.261	0.7
0	0.984	0.758	0.845	0.188	0.765	0.268	0.258	0.765	0.463	0.103	0.268	0.6
0	0.87	0.765	0.866	0.196	0.766	0.276	0.26	0.766	0.467	0.101	0.268	0.7

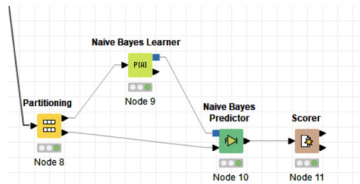
Şekil 4: Normalizer Tablosu Parçası



Şekil 5: Çok Katmanlı Yapay Sinir Ağı Modeli

Şekil 5'de gösterilen Çok Katmanlı Yapay Sinir Ağı modellemesinde eğitim setinde kullanılmı ve geriye kalan %30 veri test amacı için ayrılmıştır. Öğrenme sürecinde maksimum 100 iterasyon gerçekleştirilmiştir. Çok katmanlı yapay sinir ağı modelinde 10 gizli katman, ve her gizli katman içerisinde 100 nöron bulunmaktadır. Eğitim süreci tamamlandıktan sonra belirlenen parametreler üzerinden test setine uygulayarak sistem güvenilirliği test edilmiştir.

Şekil 6' de gösterilen Naive Bayes modellemesi için veri setinin %70'i eğitim için kullanılmı ve geriye kalan %30'u da test de kullanılmıdır. Naive Bayes öğrenimi için default olasılık 0.0001, minimum standart sapma 0,0001 almıştır. Daha sonra ise test seti üzerinden Naive Bayes modeli sınanmıştır.



Şekil 6: Naive Bayes Modeli

4. BULGULAR

Çok Katmanlı Yapay Sinir Ağı modellemesinde doğruluk oranı %99,943 elde edilirken hata oranı %0,057 olarak elde edilmiştir (Şekil 7).

Aynı problem Naive Bayes yöntemi ile modellendiğindeki doğruluk oranı %98,207 iken hata oranı %1,793 olarak tespit edilmiştir (Şekil 8).

Confusion Matrix - 3:7 - Scorer

Class \ Pre...	0	1
0	85271	20
1	29	123

Correct classified: 85.294 Wrong classified: 49
Accuracy: 99,943 % Error: 0,057 %
Cohen's kappa (κ) 0,834

Şekil 7: Yapay Sinir Ağı Sonucu

Confusion Matrix - 3:... - Scorer

Class \ Pre...	0	1
0	83786	1507
1	25	125

Correct classified: 83.911 Wrong classified: 1.532
Accuracy: 98,207 % Error: 1,793 %
Cohen's kappa (κ) 0,138

Şekil 8: Naive Bayes Sonucu

Çok katmanlı yapay sinir ağı modelinde doğru pozitif sayısı 123 (80%), yanlış pozitif sayısı ise 29 (%20)'dur. Doğru negatif sayısı 85271 (%99) iken, yanlış negatif sayısı 20 (%1)'dir.

Naive bayes yöntemi modelinde ise doğru pozitif sayısı 125 (83%), yanlış pozitif sayısı ise 25 (%17)'dir. Doğru negatif sayısı 83786 (%98) iken, yanlış negatif sayısı 1507 (%2)'dir.

5. SONUÇ

Kaggle veritabanından elde edilen Kredi Kartı Dolandırıcılık Tespiti veri seti kullanılarak Knime platformu üzerinde bir analiz yapılmıştır. Yapılan analizde Çok Katmanlı Yapay Sinir Ağı ve Naive Bayes algoritmaları kullanılmıştır. Tablo 1' de iki yöntemin performans sonuçları gösterilmektedir.

Yöntem Adı	Doğruluk	Hata
Yapay Sinir Ağı	%99,943	%0,057
Naive Bayes	%98,207	%1,793

Tablo 1: Sonuçlar

Elde edilen sonuçlar ışığında iki algoritmanın çok iyi sonuçlar çıkarmasının yanı sıra bu çalışma için Yapay Sinir Ağı algoritmasının Naive Bayes algoritmasından daha başarılı olduğu tespit edilmiştir.

Bu çalışmada kullanılan veri setinin büyüklüğü doğrultusunda elde edilen doğruluk oranlarının kullanılabilirliği ve güvenilirliği literatürdeki benzer çalışmalardan daha yüksektir.

Bu çalışmayı geliştirmek için Veri Madenciliği'nde kullanılan diğer algoritmalar ile test edilebilir. Böylelikle çıkan oranlar karşılaştırılabilir ve Yapay Sinir Ağından daha iyi bir sonuç elde edilebilirliği kontrol edilebilir.

KAYNAKÇA

- [1] KAYA F. “Türkiye’de Kredi Kartı Uygulaması” (No.263), İstanbul: Türkiye Bankalar Birliđi, 2009
- [2] ÖZBAY E. , “Finans Sektöründe Veri Madenciliđi ile Dolandırıcılık Tespiti” (Doctoral Dissertation, Selçuk Üniversitesi Fen Bilimleri Enstitüsü), 2007
- [3] ADALI B. “Kredi Kartı Dolandırıcılığı Saptama Yöntemleri” (Dissertation, İstanbul Teknik Üniversitesi Bilgisayar ve Bilişim Fakültesi), Haziran 2012
- [4] GÜLTEKİN S. U. “Yapay Sinir Ađı ve Doğrusal Regresyon Yöntemleri İle Fiyat Tahmini” (Master’ thesis, Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü) Ağustos 2017
- [5] SOYLU K. “Kredi Kartı Sahte İşlem Tespiti” 2018
- [6] ÖZMEN E. P. , ÖZCAN T. , “Finans Sektöründe Dolandırıcılık Tespiti Üzerine Melez Sınıflandırma ve Regresyon Ađacı Uygulaması” Cilt5, Sayı2, Sayfa12-20, 2019
- [8] Credit Card Fraud Detection Dataset, <https://www.kaggle.com/mlg-ulb/creditcardfraud> (09.12.2020)
- [9] BİLGİN M. , YILMAZ A. , Makine Öğrenmesi, Papatya Bilim, 2018
- [10] ŞENGÜR D. , TEKİN A. , Öğrencilerin Mezuniyet Notlarının Veri Madenciliđi Metotları ile Tahmini, Bilişim Teknolojileri Dergisi, Sayı3, Cilt6, 2013
- [11] YILMAZ A. , R Programlama, Kodlab, 2020.