Literatür Araştırması

Yapay Zeka ile Kredi Kartı Fraud Analizi

181180014 - Büşra Bayındır

181180005 - Emre Can Ant

C181112034 - Erencan Tezel

BM495 BİTİRME PROJESİ

Öğr.Gör.Dr. MUHAMMET ÜNAL

Kelime sayısı: 2703

İÇİNDEKİLER

1. Giriş

Dolandırıcılık, teknolojinin ve dünya çapında iletişimin ilerlemesi ile büyük ölçüde artmaktadır. Dolandırıcılık iki yolla önlenebilir: önleme (prevention) ve algılama (detection). Önleme, bir koruma katmanı görevi görerek dolandırıcıların saldırılarını önler. Algılama, önleme aşaması başarısız olduğunda gerçekleşir. Bu nedenle algılama, şüpheli bir kredi kartı işlemi tetiklenir tetiklenmez belirlenmesine ve sistem yöneticisinin uyarılmasına yardımcı olur. Kredi kartı dolandırıcılığı birkaç kategoriye ayrılabilir. Esas olarak bir dizi işlemde tanımlanabilen iki tür dolandırıcılık mevcuttur. Bunlar; kartsız (CNP) dolandırıcılık ve kartlı (CP) dolandırıcılıktır.
Makine öğrenimi, bu tür metodolojilerin yerini alan ve insanlar için kolay kolay mümkün olmayan büyük veri kümeleri üzerinde çalışabilen bir çözüm sunar. Makine öğrenimi teknikleri iki ana kategoriye ayrılır; denetimli (supervised) öğrenme ve denetimsiz (unsupervised) öğrenme. Dolandırıcılık tespiti her iki şekilde de yapılabilir ve sadece veri setine göre ne zaman kullanılacağına karar verilebilir. Denetimli öğrenme, anormalliklerin önceden sınıflandırılmasını gerektirir. Denetimsiz öğrenmede bu işi model kendi öğrenir. Son birkaç yılda, kredi kartı sahtekarlığını tespit etmek için bir çok denetimli ve denetimsiz algoritma kullanılmıştır.
Bu araştırmamızda kredi kartı dolandırıcılığında daha önce yapılmış projelerde hangi algoritmaların kullanıldığını, bu algoritmaları kullanma sonucunda ne kadar başarı elde edildiğine bakılmıştır. Elde edilen sonuçlar doğrultusunda yapacağımız projede hangi algoritma/algoritmalara yer vereceğimize karar verilmiştir.

2. İlgili Çalışmalar

Geçmişteki araştırmalarda, çeşitli tespit modellerinin analizi ile dolandırıcılık tespiti ile ilgili birçok problem bulunmuştur.

2.1. Genel Tarama

Bazı makalelerde gerçek hayat verilerinin eksikliğinden büyük bir sorun olarak bahsetmişlerdir. Veri hassasiyeti ve gizlilik sorunları nedeniyle gerçek hayat verileri eksiktir [3][8]. Bazı makalelerde ise dengesiz verileri ve verilerin çarpık dağılımını incelemişlerdir. Bu çarpıklığın nedeni, işlemlerin bulunduğu veri setlerinde dolandırıcılık olmayanlara kıyasla oldukça az miktarda dolandırıcılıkla sınıflandırılmış verilerin bulunmasıdır [3][6].

Büyük veriyle uğraşırken veri madenciliği tekniklerinin yürütülmesinin zaman aldığı belirtilmektedir. Verilerin örtüşmesi, aynı verilerin bulunması, kredi kartı işlem verisetinin hazırlanmasındaki bir diğer önemli dezavantajdır [3].

Bazı makalelere göre de sorun, normal işlemlerin tam olarak sahte işlemler gibi göründüğü bazı senaryolar nedeniyle ortaya çıkmaktadır [2][6]. Kredi kartı işlem verileri göz önüne alındığında, özelliklerin çoğu kategorik değerlere sahiptir. Bu durumda hemen hemen tüm makine öğrenmesi algoritmaları kategorik değerleri desteklemez.

Makine öğrenmesi algoritmalarının çoğu eğitim amacıyla tahmin etmekten çok daha fazla zaman aldığından, algılama algoritmaları seçiminden ve özellik seçiminden sahtekarlıkları tespit etmek bir zorluk oluşturmaktadır. Finansal dolandırıcılık tespitini etkileyen bir diğer önemli konu da özellik seçimidir. Dolandırıcılık tespitinin özelliklerini ve karakterlerini en çok tanımlayan özellikleri filtrelemeyi amaçlar [3][4]. Etkinlik, problem tanımına ve özelliklerine göre değişebilir, bu nedenle performans ölçüsünün iyi anlaşılması gereklidir [4].

Sahte kredi kartı işlem örnekleri, Gauss Karışım Modelleri (GMM) kullanılarak stratejilere ayrılmıştır. Burda sınıf dengesizliğini gidermek için sentetik azınlık over-sampling tekniği kullanılmıştır. Tahminlerin önemini öne çıkarmak için ekonomik değer duyarlılığı analizi kullanılmıştır. Sonuçlar, bir modeli yeniden eğitmek için minimum adımları kullanan pratik bir yöntemin, tipik olarak her turda yeniden eğiten bir sınıflandırıcı ile aynı işlevi görebileceğini kanıtlamıştır [7].

Peter ve çalışmayı beraber yürüttüğü ekip arkadaşları birkaç derin öğrenme algoritmasını etkinliklerine göre değerlendirmiştir.. Dört topoloji, Tekrarlayan Sinir Ağları (RNN), Kapılı Tekrarlayan Birimler (GRU), Uzun Kısa Süreli Bellek (LSTM) ve Yapay Sinir Ağları'dır (ANN). Projelerinde veri temizleme ve diğer veri hazırlama adımlarına ek olarak, alt örnekleme kullanarak sınıf dengesizliği ve ölçeklenebilirlik problemlerinin üstesinden gelmişlerdir. Hangi hiper parametrelerin modelin performansı üzerinde en yüksek etkiye sahip olduğunu keşfetmek için duyarlılık analizi yapılmıştır. Modelin performansının ağın boyutundan etkilendiğini keşfetmişlerdir. Ağ büyüdükçe daha iyi performans gösterdiği sonucuna varmışlardır. [5].

Kredi kartı verileri, sınıf dengesizliği olarak da bilinen çarpık dağılım sorununa sahiptir. Andrea ve çalışmayı beraber yürüttüğü ekip arkadaşlarına göre, projeleri, kavram kayması ve doğrulama gecikmesi gibi diğer konular da dahil olmak üzere sınıf dengesizliğini ele almaktadır. Ayrıca, kredi kartı sahtekarlığının tespitinde kullanılabilecek en alakalı performans matrisini de göstermişlerdir. Araştırmanın başarısı ayrıca resmi bir model ve 'doğrulama gecikmesi' ve bir 'uyarı ve geri bildirim' mekanizmasını ele almak için güçlü bir öğrenme stratejisi içermektedir. Deneylere göre, uyarıların kesinliğini en önemli önlem olarak ilan etmişlerdir [9]. Che ve çalışmayı beraber yürüttüğü ekip arkadaşları. kredi kartı dolandırıcılık tespitinde daha iyi doğruluk oranları elde etmek için AdaBoost ve çoğunluk oylama yöntemlerini kullanan on iki standart model ve hibrit yöntem kullanmışlardır [10]. Hem kıyaslama hem de gerçek dünya verileri kullanılarak değerlendirme yapmışlardır. Yöntemlerin güçlü yönleri ve sınırlamalarının bir özeti değerlendirilmiştir.. Performans ölçütü olarak Matthews Korelasyon Katsayısı (MCC) alınmıştır. Algoritmaların sağlamlığını değerlendirmek için verilere gürültü eklenmiştir.. Ayrıca, çoğunluk oylama yönteminin eklenen gürültüden etkilenmediğini kanıtlamışlardır.

2.2. Detaylı Tarama

Siddhartha Bhattacharyya 2011'de yayınladığı makalesinde [11] kredi kartı sahtekarlık analizini çeşitli yöntemleri karşılaştırarak gerçekleştirmiştir. Kullandığı veri seti yalnızca bir ülkeye ait olan yaklaşık 1.2 milyon kredi kartından yapılmış yaklaşık 50 milyon adet işleme sahiptir. 2006-2007 yılları arasındaki işlemleri kapsayan ve oldukça ayrıntılı sütunlara sahip olan bu veri seti bir bankadan direkt olarak temin edilmiştir. Yazar farklı durumlardaki performansları ölçmek amacıyla veri setini fraud oranı %15, %10, %5, %2 ve %0.5 olacak şekilde 5 parçaya bölmüş ve ayrı ayrı performans analizlerini yapmıştır. Logistic regression, support vector machines ve random forests algoritmaları kullanılmıştır. Sırasıyla %94.7, %93.8 ve %96.2 doğruluk oranları elde edilmiştir.

Ibrahim Alowais 2012'de yayınladığı makalesinde [12] iki farklı yöntem kullanarak kredi kartı sahtekarlık analizi yapmaya çalışmıştır. Diğer makalelerden oldukça farklı olarak verisetini online anket yolu ile toplanmaya çalışılmıştır. Yazar 47 farklı kişiden 4924 adet işlem bilgisini temin ettiğini ve bu işlemlerin aynı 6 aylık dilime ait olduğunu söylemektedir. Bu işlemlerin 921 adedi yani verisetinin %18.7'sinin fraud olduğunu belirtmektedir. Online anket yoluyla toplanmış verilerden bu kadar yüksek bir oran elde edilmiş olması verisetinin güvenilirliğine gölge düşürmektedir. Random forest ve naive bayes algoritmalarını kullanarak analiz yapan yazar sırasıyla %89.48 ve %84.08 doğruluk oranlarını elde etmiştir. Diğer çalışmalara kıyasla düşük bir doğruluk oranı olması ayrıca dikkat çekmektedir.

Yiğit Kültür 2016 yılında yayınladığı makalesinde [13] tek bir yöntem yerine birden fazla yöntemin hibrit bir şekilde kullanılmasıyla maksimum doğruluk oranını elde etmeye çalışmıştır. Türkiye'nin önde gelen bankalarından birinden aldığı veriseti 105 ayrı kişiye ait 173 karttan yapılmış 152 bin işlemi kapsamaktadır. Verisetinin ülkemize ait olması ve halkımızın alışveriş tarzını yansıtması sebebiyle diğer makalelerden ayrılmaktadır. Birçok yöntem kullanılan çalışmada Decision Tree %95.19, Random Forest %95.81, Bayesian Network %96.92, Naive Bayes %94.10, Support Vector Machine %94.17, K* %91.73 doğruluk oranı sağlamıştır. Yazar bu yöntemleri çeşitli oranlarda kullanarak karar mekanizmaları oluşturmuştur. Bunlardan en iyisi %97.55 en kötüsü ise %86.65 doğruluk oranı vermiştir.

S. A. Oluwadare ve ekibinin 2017'de yayınladığı makalesinde [14] kredi kartlarında sahtekarlık analizi için karşılaştırmalı bir yöntem izlemiştir. Veriseti ULB Machine Learning Group'dan hazır olarak alınmıştır. Veriseti 2013 yılında avrupalı kişiler tarafından yapılmış 284 bin işlemi ihtiva etmektedir. Bunların %0.172'si fraud'tur. Çalışmada naive bayes %97.69, k-Nearest neighbour %0.9792 ve logistic regression %54.86 doğruluk sağlamıştır.

2017 yılında J. O. Awoyemi ve ekibi tarafından yapılan çalışmada [18] kredi kartı sahteciliği tespiti için Kaggle üzerindeki 284.808 işlem kaydı bulunan veri seti kullanılmıştır. Bu çalışmada fraud kategorisinde olan işlemlere over-sampling yapılırken normal işlemlerde under-sampling yapılıp 34:66 ve 10:90 oranlarında 2 ayrı veri seti daha oluşturulmuştur.

3 makine öğrenmesi algoritması kullanılmış olup oluşturulan 3 ayrı veri seti üzerinde modellenmiştir. Çalışmada Naive Bayes, K-Nearest Neighbor ve Logistic Regression classifierları kullanılmıştır.
3 modelin performansı accuracy, sensitivity ve Matthews Correlation Coefficient (MCC) temel alınarak değerlendirilmiştir.
Naive Bayes, k-NN ve Logistic Regression için performans analizi sırasıyla;

Un-sampled data için:

Accuracy:	0.9737	0.9691	**0.9824**
Sensitivity:	0.8072	0.8835	**0.9767**
MCC:	+0.1979	**+0.5903**	+0.2893

10:90 data dağılımı için:

Accuracy:	**0.9752**	0.9715	0.3639
Sensitivity:	0.8210	**0.8285**	0.7155
MCC:	+0.2080	**+0.8950**	+0.0077

34:66 data dağılımı için:

Accuracy:	0.9769	**0.9792**	0.5486
Sensitivity:	**0.9514**	0.9375	0.5833
MCC:	+0.9478	**+0.9535**	+0.1080

Layth Rafea Hazım'ın 2018'de yayınladığı yüksek lisans tezinde [15] da yine kredi kartı sahtekarlık analizi için kullanılan 4 farklı yöntemi aynı veriseti üzerinde test ederek avantaj ve dezavantajlarını karşılaştırmıştır. Yazar Awoyemi'nin de yayınında kullandığı ULB Machine Learning Group'un yayınlamış olduğu hazır verisetini kullanmıştır. Testler sonucunda naive bayesian algoritması %97.46, support vector machine %95.04, KNN %97.55, random forest %97.7 doğruluk oranı sağlamıştır.

2018 yılında S. D. Kavila tarafından yapılan çalışmada [17] Kaggle'da yer alan 284.808 işlem içeren veri seti kullanılmıştır. Veri setinde over sampling yapılıp ardından train ve test olmak üzere 2 ayrı veri setine bölünmüştür. Train set, ana veri setinin %70'i, test set ise %30'u olacak şekilde ayrılmıştır. Çalışmada kullanılan ve kıyaslanan makine öğrenmesi metotları; Logistic Regression, Decision Tree Algorithm ve Random Forest'tır.
Model sonuçları accuracy ((TP+TN) / (P+N)), error rate ((FP+FN) / (P+N)), sensitivity (TP/P) ve specificity (TN/N) performans ölçütleri ile değerlendirilmiştir.
Random Forest, Decision Tree ve Logistic Regression metotları için accuracy oranları sırasıyla %95.5, %94.3 ve %90.0 şeklinde sonuçlanmıştır.

2018 yılında yapılan çalışmada [16] 10 farklı makine öğrenmesi modeli birbirleriyle Accuracy, TPR, FPR, G-mean, Recall, Precision, Specificity and F1-Score değerleriyle karşılaştırılmıştır. Çıkarımda supervised modeller kullanılmış olup unsupervised tekniği kredi kartı işlemlerinin sınıflandırılmasında kullanılmıştır. Bu yöntem çalışma [19]'dan görülmüştür.

Unbalanced dataseti dengelemek için sampling metodu kullanılmıştır. Datasette 284.808 işlem bulunup bunların 492 tanesi fraud olarak sınıflandırılmıştır. %70 train seti oluştururken geri kalan %30 test setini oluşturmaktadır.
Karşılaştırılan makine öğrenmesi metotları: Stacking Classifier, Random Forest, XGB Classifier, KNN, Logistic Regression, Gradient Boosting, MLP Classifier, SVM, Decision Tree ve Naive Bayes.

Çalışmanın sonuçları Şekil 2.0'da gösterilmektedir.

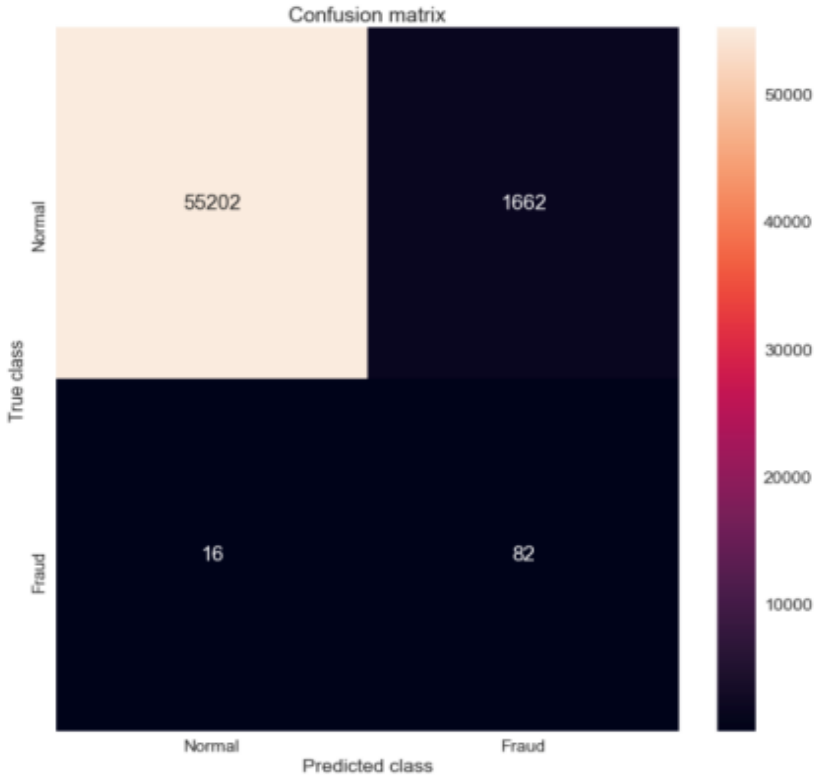
Model	Accuracy	Precision	Recall	TPR	FPR	F1- Score	G-Mean	Specificity
SC	0.95270	0.95	0.95	0.9387	0.0335	0.95	0.9524	0.9664
RF	0.94594	0.95	0.95	0.9251	0.0335	0.95	0.9455	0.9664
XGB Classifier	0.94594	0.95	0.95	0.93197	0.0402	0.95	0.9457	0.9597
KNN	0.94256	0.91	0.91	0.9183	0.0335	0.91	0.942	0.9664
LR	0.93918	0.94	0.94	0.93877	0.0604	0.94	0.9391	0.9395
GB	0.93581	0.94	0.94	0.9183	0.0335	0.94	0.942	0.9664
MLP Classifier	0.93243	0.93	0.93	0.9387	0.0738	0.93	0.9323	0.9261
SVM	0.93243	0.93	0.93	0.9183	0.536	0.93	0.9321	0.9463
Decision Tree	0.90878	0.91	0.91	0.9047	0.0872	0.91	0.9086	0.9127
Navies Bayes	0.90540	0.91	0.91	0.85714	0.04697	0.91	0.9037	0.953

Şekil 2.0. Classifierların performansları

2018'de Liu ve Apapan'ın yaptığı çalışmada kredi kartı sahtecilik analizinde deep learning modelleri kullanılmıştır. Daha önceki supervised çalışmalarla tespit edilememiş fraud işlemleri tespit etmek amacıyla geliştirilmiştir. Bunun için deep Auto-Encoder (AE) ve restricted Boltzmann Machine (RBM) geliştirilmiştir. Unsupervised öğrenme teknikleri kullanılmıştır. RBM iki katmandan oluşmuştur. AE, RBM ve H2O implementasyonu için Tensorflow kütüphanesi kullanılmıştır. German, Australian ve Europian datasetleri olmak üzere 3 datasett üzerinde ayrı ayrı çalışılmıştır.

Dataset Name	**No. of transactions**	**AUC's score based on AE**	**AUC's score based on RBM**
German Dataset	1000	0.4376	0.4562
Australian Dataset	690	0.5483	0.5238
European Dataset	284, 807	**0.9603**	**0.9505**

Şekil 2.1. AE ve RBM modellerin AUC skorları



Şekil 2.2. Europian dataset için confusion matrix

Nuwan Kuruwitaarachchii 2019 yılında yayınladığı makalesinde [1] ekip tarafından yürütülen çalışmada kullandıkları dataset toplam 917.781 kayıt içermekte, bunların 200'ü fraud (sahte) olduğu bilinmektedir. Çalışmada toplanan dataset, fraud örüntüsüne göre 4 parçaya ayrılmıştır;
1- Riskli MCC (Merchant Category Code) (MCC, işletmeyi sunduğu hizmet ve ürünlere göre sınıflandıran 4 haneli bir koddur.)
2- 100$'dan büyük işlemler
3- Riskli ISO Response kodlu işlemler
4- Bilinmeyen web adreslerdeki işlemler
Çalışmada Support Vector Machine (SVM), Naive Bayes, K-Nearest Neighbor ve Logistic Regression sınıflandırma ve makine öğrenmesi algoritmaları kullanılmıştır.
Çalışma gerçek zamanlı fraud tespiti için 3 ana yapı sunuyor; API Modülü, Fraud tespit sistemi ve Data Warehouse. API modülü gerçek zamanlı kredi kartı işlemleri verilerini fraud tespit sistemine olan iletiminde, data warehouse ile iletimin sağlanmasından sorumludur. Data Warehouse canlı işlemleri tutmak, makine öğrenmesi algoritmaları için gerekli verileri ve model sonucu çıkan tahminleri saklamak için kullanılmaktadır.
Çalışma sonucunda SVM,NB, K-NN ve LR modellerinin doğruluk oranları sırasıyla: %91, %83, %72 ve %74 çıkmıştır.

2020 yılında Ruttala Sailusha ve ekibi tarafından yapılan çalışmada [20] Random Forest ve Adaboost algoritmaları kullanılmıştır. Random Forest ve Adaboost bir ensemble metottur.

Adaboost algoritmasının çalışma biçimi kısaca şöyledir. Başlangıçta training datadan bir model oluşturulur. Ardından 2. model, ilk modelin hatalarını düzeltilmesi amaçlanarak oluşturulur. Bu tekrarlı işlemler maksimum model sayısına ulaşınca veya training dataseti doğru tahminlerle tamamladığında biter. Çalışmada Kaggle kredi kartı fraud dataseti kullanılmıştır. Çalışmanın sonuçları aşağıdaki görsellerdedir.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	93825
1	0.95	0.77	0.85	162
accuracy			1.00	93987
macro avg	0.97	0.89	0.93	93987
weighted avg	1.00	1.00	1.00	93987

Şekil 2.3. Random Forest için sonuçlar

```
Confusion Matrix on train data
[[190490      0]
 [     0    330]]

Confusion Matrix on test data
[[93818    37]
 [    7   125]]
```

Şekil 2.4. Random Forest için confusion matrix

```
Accuracy = 0.9990743400683073
             precision    recall  f1-score   support

          0  0.99938202 0.99969091 0.99953644     93825
          1  0.78195489 0.64197531 0.70508475       162
```

Şekil 2.5. Adaboost için sonuçlar

```
Confusion Matrix on train data
[[190464    120]
 [    26    210]]
Confusion Matrix on test data
[[93811    65]
 [   14    97]]
```

Şekil 2.6. Adaboost için confusion matrix

3. Tartışma ve Sonuçlar

Literatür araştırması sürecinde pek çok makale okunmuş ve fraud analizi için birçok yöntemin kullanıldığı görülmüştür. Karşılaştırmalı olarak yapılmış çalışmalardan aldığımız doğruluk oranları Tablo 1'de sunulmuştur.

Kaynak	**LR**	**SVM**	**NB**	**KNN**	**RF**	**DT**	**BN**	**K***	**GB**	**MLP**	**XGB**	**SC**
[1]	74	91	83	72	-	-	-	-	-	-	-	-
[20]	90	-	-	-	95,5	94,3	-	-	-	-	-	-
[21]	36,3	-	97,5	97,1	-	-	-	-	-	-	-	-
[14]	94,7	93,8	-	-	96,2	-	-	-	-	-	-	-
[15]	-	-	84	-	89,4	-	-	-	-	-	-	-
[13]	-	94,1	94,1	-	95,8	95,1	96,9	91,7	-	-	-	-
[17]	54,8	-	97,6	97,9	-	-	-	-	-	-	-	-
[18]	-	95	97,4	97,5	97,7	-	-	-	-	-	-	-
[16]	93,9	93i2	90,5	95,2	95,4	90,8	-	-	93,5	93,2	95,4	95,2
Ort.	73,9	93,4	92	91,9	95	93,4	96,9	91,7	93,5	93,2	95,4	95,2

Tablo 3.1. Yapılan çalışmalara göre algoritmaların doğruluk oranları

Tablo 3.1'den de anlaşılabileceği üzere kredi kartı fraud analizi için en çok support vector machine, naive bayes, knn, random forest ve decision tree algoritmaları tercih edilmiştir. Bunların arasından özellikle random forest algoritmasının yüksek doğruluk oranı dikkat çekmektedir. Random forest, ortalama %95 gibi yüksek bir doğruluk oranına sahip olmasına rağmen bizim hedeflediğimiz oranlara ulaşamamıştır. Ancak Yiğit Kültür'ün yayınladığı çalışmada [13] bu oranların artırılması için çeşitli algoritmaların hibritlenmesi yöntemi denenmiş ve çalışma sonuçlarından elde ettiğimiz veriler üzere bu hibritleme işleminin bizim istediğimiz başarı oranlarını sağladığı görülmüştür. Bundan yola çıkarak geliştireceğimiz modelin random forest, naive bayes, support vector machine, KNN ve decision tree

algoritmalarının bir hibriti olmasına karar verilmiştir. Bunun için iki aşamalı bir yöntem izlenecektir. Eğitilmiş ilk modelimizin elde ettiği sonuçlar tekrar bir algoritmaya sokularak daha doğru bir sonuç elde edilmeye çalışılacaktır.

4. Kaynaklar

[1] A. Thennakoon, C. Bhagyani, S. Premadasa, S. Mihiranga,N. Kuruwitaarachchi,"Real-time Credit Card Fraud Detection Using Machine Learning", 9th Int. Conf. on Cloud Comp., Data Sci. & Engineering, pp. 488-493, 2019.
[2] M. Zareapoor, S. K. Seeja K.R.ve M. Afshar Alam, "Analysis on Credit Card Fraud Detection Techniques: Based on Certain Design Criteria" Int. J. Comput. Appl., vol.52, no. 3, pp 35-42, 2012.
[3] David Robertson, "Investments Acquisitions - September 2016 Top Card Issuers in Asia-Pacific Card Fraud Losses Reach $21.84 Billion," Nilson Rep., no. 1096, 1090.
[4] J. West ve M. Bhattacharya, "An Inversitgation on Experimental Issues in Financial Fraud Mining", *Procedia Comput. Sci.*, vol. 80, pp. 1734-1744, 2016.
[5] A. Roy, J. Sun, R. Mahoney, L. Alonzi, S. Adams ve P. Beling, "Deep Learning Detecting Fraud in Credit Card Transactions", pp. 129-134, 2018.
[6] M. F. Zeager, A. Sridhar, N. Fogal, S. Adams, D. E. Brown ve P. A. Beling, "Adversarial learning in credit card fraud detection", *2017 Syst. Inf. Eng. Des. Symp.*, pp. 112-116, 2017.
[7] T. Cody, S. Adams, ve P. A. Beling, "A Utilitarian Approach to Adversarial Learning in Credit Card Fraud Detection", pp. 237-242, 2018.
[8] M. Rafalo, "Real-time fraud detection in credit card transactions", *Data Science Warsaw*, 2017.
[9] A. Dal Pozzolo, G. Boracchi, O. Caelen ve C. Alippi," Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy", *Ieee Trans. Neural Networks Learn. Syst.*, pp. 1-14, 2018.
[10] K. Randhawa, C. K. Loo, M. Seera, C. P. Lim ve A. K. Nandi, "Credit card fraud detection using AdaBoost and majority voting", *IEEE Access*, vol. XX, pp. 1-1, 2018.
[11] Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. *Decision support systems*, *50*(3), 602-613.
[12] Alowais, M. I., & Soon, L. K. (2012, June). Credit card fraud detection: Personalized or aggregated model. In *2012 Third FTRA International Conference on Mobile, Ubiquitous, and Intelligent Computing* (pp. 114-119). IEEE.
[13] Kültür, Y., & Çağlayan, M. U. (2017). Hybrid approaches for detecting credit card fraud. *Expert Systems*, *34*(2), e12191.
[14] Awoyemi, J. O., Adetunmbi, A. O., Oluwadare, S. A. (2017, October). Credit card fraud detection using machine learning techniques: A comparative analysis. In *2017 international conference on computing networking and informatics (ICCNI)* (pp. 1-9). IEEE.
[15] Hazım, L. R. (2018). *Four classification methods Naïve Bayesian, support vector machine, K-nearest neighbors and random forest are tested for credit card fraud detection* (Master's thesis, Altınbaş Üniversitesi).

[16] Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study
[17] S. D. Kavila, "Machine Learning For Credit Card Fraud Detection System", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, pp. 16819-16824, 2018
[18] J. O. Awoyemi, A. O. Adetunmbi and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis," 2017 International Conference on Computing Networking and Informatics (ICCNI), 2017, pp. 1-9
[19] M. Carminati, R. Caron, F. Maggi, I. Epifani, and S. Zanero, "BankSealer: A decision support system for online banking fraud analysis and investigation," Comput. Secur., vol. 53, pp. 175–186, 2015.
[20] R. Sailusha, V. Gnaneswar, R. Ramesh and G. R. Rao, "Credit Card Fraud Detection Using Machine Learning," *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2020, pp. 1264-1270,