

Müşteri Memnuniyetinin İncelenmesi(Investigation of Customer Satisfaction)

Tuğçe AKIN

1. Süleyman Demirel Üniversitesi, Bilgisayar Mühendisliği Bölümü

Özet

Günümüzde hızla gelişmekte olan teknoloji ile birlikte GSM operatörlerinin sürekliliği, müşteri memnuniyeti ile doğrudan ilişkilidir. Teknolojideki hızlı değişimle birlikte sayısı artan GSM operatörleri rekabeti arttırmıştır. Müşterilerin sahip olduğu özellikler, sonuca giden yolda, bize yardım etmektedir. Çoğu şirket müşterilerini kaybetmemek için, müşterilerin evli olup olmadıkları, bazı servisleri kullanıp kullanmadıkları, ödeme yaparken hangi yöntemi tercih ettikleri, ne kadar ödeme yaptıkları gibi özelliklerin verisini tutmaktadır. Veri bilimi ile müşterilerin memnuniyetini tahmin etmek daha kolaylaşmaktadır. Bu makalede müşterilerin özellikleri incelenmiştir ve müşteri memnuniyeti tahmin edilecektir. Müşteri kayıpları görülerek veri bilimi ile müşteri kayıplarını iyileştirme programı yapılmaktadır.

Anahtar Kelimeler:classification, machine learning, data mining, data science

Abstract

Nowaday continuity of GSM operators, take part in service sector which grows up rapidly, is directly related to customer satisfaction.Increasing number of GSM operators with rapid change in technology has increased the competition between operators.The features that customers have are helping us on the way to the end.Many companies keep data in order to lose customers, whether they are married, using some services, choosing which method to pay, and how much they are paying.It is easier to estimate the customer satisfaction with the data.In this article, the characteristics of the customers are examined and customer satisfaction is estimated. Customer losses are observed and data loss and customer loss improvement programs are being implemented.

Keywords: classification, machine learning, data mining, data science

1. Giriş

Gelişen teknoloji ile kullanım alanlarının artmasıyla beraber her geçen gün verilerin miktarı da artmaktadır. Veri miktarının artışı ile birlikte verilerin analizi de zorlaşmaktadır. Araştırmalara göre 2 senede bir tüm dünyadaki veriler ikiye katlanmaktadır. Son yıllarda verileri analiz edip anlamlandırmak popüler olmuştur. Verileri analiz edip, bu verilerden anlamlı bilgiler çıkarırken istatistik, yapay zeka, makine öğrenimi, veri madenciliği, iş zekası gibi konulardan yararlanılmaktadır. Veri madenciliği yöntemleri pek çok algoritma içermektedir. Bankacılık, finans, pazarlama alanları gibi pek çok alanda veri madenciliğine ihtiyaç vardır. Buradaki amaç verilerden anlamlı sonuçlar çıkararak sektöre fayda sağlamaktır. Veri madenciliği, elindeki verileri bir modele koyup gelecekle ilgili tahmin etme sanatıdır. Sonuç kriterlerini değerlendirirken başarı oranlarına bakılır. Veri madenciliğinin kullanıldığı

alanlardan birisi de telefon operatörleridir. Veri madenciliği daha önce hiç akla gelmeyen düşünülmemiş şeylerin tahminini de yapmaktadır. Günümüzde operatörler çok fazla olduğu için öne çıkabilmek adına kampanyalar düzenlenmektedir. Bu kampanyalara verideki özelliklerden yola çıkarak karar verilmektedir. Örneğin belli bir yaşın üzerinde çocuk sahibi insanların ihtiyacı daha geniş olabilmektedir. Bu kişilerin özelliklerine bakarak en çok hangi kampanyaya ihtiyacı olduğu karar verilip, bu kişilere sunulmaktadır. Şirketler bu özelliklere tek tek baktığında anlamlı bir veri çıkaramazlar. Veri madenciliği sayesinde hangi özellikler birbirini etkiler, hangi özellikler birlikte kullanıldıklarında müşteri kaybını önler gibi sonuçlar çıkmaktadır.[1]



Şekil 1. Veri Bilimi

2.Verit Madenciliği Aşamaları Ve Yöntemleri

Veri Madenciliği, CRISP-DM (Cross Industry Standard Processing – for Data Mining, Endüstriler arası standart işleme – veri madenciliği için) adı verilen yonteme göre beş temel adımdan oluşur:

- Problemi Tanımak
- Veriyi Anlamak
- Veriyi Hazırlamak
- Model Oluşturmak
- Başarı Değerlendirme
- Üretime Geçme

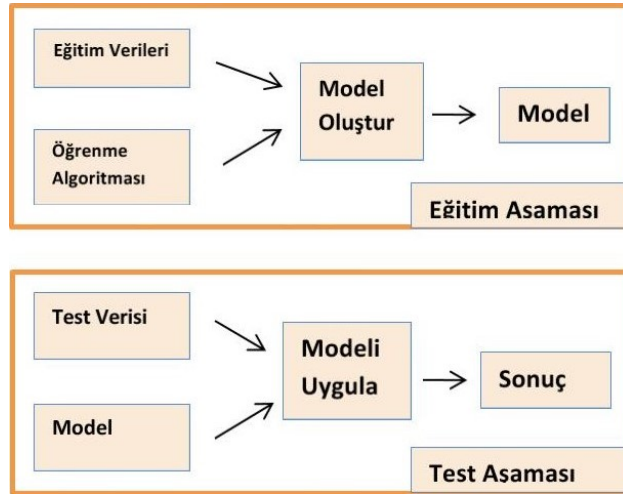
Birçoğu istatistiksel olmak üzere bu yöntemler bünyelerinde çok sayıda algoritma barındırmaktadırlar.

2.1. Sınıflandırma (Classification)

Var olan verilerin yardımı ile geleceğin tahmin edilmesinde, veri madenciliği teknikleri kullanılmaktadır. Bu tekniklerden en yaygın olanları sınıflandırma ve regresyon modelleridir. Bu modeller arasındaki temel fark tahmin edilmeye çalışılan bağımlı değişkenin kategorik veya sürekli olmasıdır. Bir kişinin eğitim süresine bakılarak ilk, orta

veya lise grubuna atanması kategorik bir gösterime örnektir. Fakat çok terimli regresyon gibi tekniklerin de kategorik verileri tahmin etmede kullanılması bu iki farklı konuyu birbirine yakınlaştırarak, tekniklerin aynı olması durumuna yol açar.[2] Sınıflandırma modelinde kullanılan başlıca teknikler:

- Karar Ağaçları (Decision Trees)
- K-En Yakın Komsu (K-Nearest Neighbor)
- Random Forest
- Naive-Bayes
- Lojistik Regresyon (Logistic Regression)



Şekil 2. Sınıflandırma Modeli

3. Veri

Bu çalışmada kaggle.com dan 7043 satır olan bir veri seti alınmıştır. Bu veri setinde 21 adet kolon bulunmaktadır. Bu veriler ile 7043 müşterinin özellikleri incelenmekte olup müşteri memnuniyeti tahmin edilecektir. Tahmin etmede kullanılan tekniklerin yanı sıra bazı veriler dışarıda bırakılacaktır. Gürültüyü düzeltme, silme veya görmezden gelme, özel değerlerle ve anlamlarıyla nasıl ilgilenilmesi gerektiğini kararlaştırma gruplama seviyesi, eksik değerler, aykırı gözlemlerin tespiti ile veri temizlenmektedir.

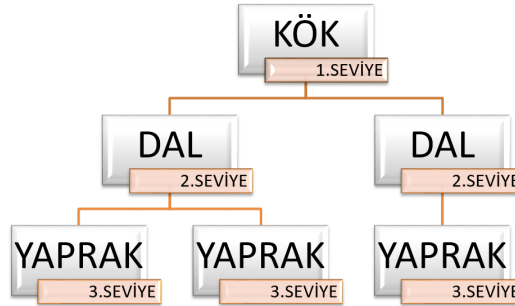
Bütün veri madenciliği araçları için biraz veri hazırlama işlemi yapmak gerekir. Buradaki ana amaç veri setlerindeki müşterilerin özelliklerinin veri madenciliği araçları tarafından en iyi şekilde analiz edilmesini sağlamak için veriyi bir takım dönüşüm işlemlerinden geçirmektir. Veri hazırlama işlemi bittikten sonra hatalı tahmin oranlarının, hazırlanma işleminden önceki orana daha düşmüş olması gerekir.[3]

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
customerI	gender	SeniorCitz	Partner	Dependent	tenure	PhoneServ	MultipleLi	InternetSe	OnlineSec	OnlineBac	DevicePro	TechSupp	Streaming	StreamingI	Contract	PaperlessP	PaymentM	MonthlyCh	TotalChar	Churn
7590-VHVI	Female	0	Yes	No	1	No	No phone	DSL	No	Yes	No	No	No	No	Month-to	Yes	Electronic	29.85	29.85	No
5575-GNV	Male	0	No	No	34	Yes	No	DSL	Yes	No	Yes	No	No	One year	No	Mailed ch	56.95	1889.5	No	
3668-QPYE	Male	0	No	No	2	Yes	No	DSL	Yes	Yes	No	No	No	Month-to	Yes	Mailed ch	53.85	108.15	Yes	
7795-CFOI	Male	0	No	No	45	No	No phone	DSL	Yes	No	Yes	Yes	No	One year	No	Bank trans	42.3	1840.75	No	
9237-HQIT	Female	0	No	No	2	Yes	No	Fiber optic	No	No	No	No	No	Month-to	Yes	Electronic	70.7	151.65	Yes	
9305-CDSF	Female	0	No	No	8	Yes	Yes	Fiber optic	No	No	Yes	No	Yes	Month-to	Yes	Electronic	99.65	820.5	Yes	
1452-KIOV	Male	0	No	Yes	22	Yes	Yes	Fiber optic	No	Yes	No	Yes	No	Month-to	Yes	Credit car	89.1	1949.4	No	
6713-OKO	Female	0	No	No	10	No	No phone	DSL	Yes	No	No	No	No	Month-to	No	Mailed ch	29.75	301.9	No	
7892-POO	Female	0	Yes	No	28	Yes	Yes	Fiber optic	No	No	Yes	Yes	Yes	Month-to	Yes	Electronic	104.8	3046.05	Yes	
6388-TABK	Male	0	No	Yes	62	Yes	No	DSL	Yes	Yes	No	No	No	One year	No	Bank trans	56.15	3487.95	No	
9763-GRSF	Male	0	Yes	Yes	13	Yes	No	DSL	Yes	No	No	No	No	Month-to	Yes	Mailed ch	49.95	587.45	No	
7469-LKBC	Male	0	No	No	16	Yes	No	No	No interne	No interne	No interne	No interne	No interne	Two year	No	Credit car	18.95	326.8	No	
8091-TTVI	Male	0	Yes	No	58	Yes	Yes	Fiber optic	No	No	Yes	No	Yes	One year	No	Credit car	100.35	5681.1	No	
0280-XJGE	Male	0	No	No	49	Yes	Yes	Fiber optic	No	Yes	No	Yes	Yes	Month-to	Yes	Bank trans	103.7	5036.3	Yes	
5129-ILPIS	Female	0	No	No	25	Yes	No	Fiber optic	Yes	No	Yes	Yes	Yes	Month-to	Yes	Electronic	105.5	2686.05	No	
3655-SNQU	Female	0	Yes	Yes	69	Yes	Yes	Fiber optic	Yes	Yes	Yes	Yes	Yes	Two year	No	Credit car	113.25	7895.15	No	
8191-XWV	Female	0	No	No	52	Yes	No	No	No interne	No interne	No interne	No interne	No interne	One year	No	Mailed ch	20.65	1022.95	No	
9959-WOF	Male	0	No	Yes	71	Yes	Yes	Fiber optic	Yes	No	Yes	Yes	Yes	Two year	No	Bank trans	106.7	7382.25	No	
4190-MFLI	Female	0	Yes	Yes	10	Yes	No	DSL	No	No	Yes	No	No	Month-to	No	Credit car	55.2	528.35	Yes	
14183-MYFI	Female	0	No	No	21	Yes	No	Fiber optic	No	Yes	No	No	Yes	Month-to	Yes	Electronic	90.05	1862.9	No	
8779-QRD	Male	1	No	No	1	No	No phone	DSL	No	No	Yes	No	No	Yes	Month-to	Yes	Electronic	39.65	39.65	Yes
1680-VDC	Male	0	Yes	No	12	Yes	No	No	No interne	No interne	No interne	No interne	No interne	One year	No	Bank trans	19.48	202.25	No	
1066-JKSG	Male	0	No	No	1	Yes	No	No	No interne	No interne	No interne	No interne	No interne	Month-to	No	Mailed ch	20.15	20.15	Yes	
3638-WEAF	Female	0	Yes	No	58	Yes	Yes	DSL	No	Yes	No	Yes	No	Two year	Yes	Credit car	59.9	3505.1	No	
6322-HRPI	Male	0	Yes	Yes	49	Yes	No	DSL	Yes	No	Yes	No	No	Month-to	No	Credit car	59.6	2970.3	No	
6865-IJZF	Female	0	No	No	30	Yes	No	DSL	Yes	Yes	No	No	No	Month-to	Yes	Bank trans	55.3	1530.6	No	
6467-CHF	Male	0	Yes	Yes	47	Yes	Yes	Fiber optic	No	Yes	No	No	Yes	Month-to	Yes	Electronic	99.35	4749.15	Yes	
8665-UTDI	Male	0	Yes	Yes	1	No	No phone	DSL	No	Yes	No	No	No	Month-to	No	Electronic	30.2	30.2	Yes	
5248-YGJII	Male	0	Yes	No	72	Yes	Yes	DSL	Yes	Yes	Yes	Yes	Yes	Two year	Yes	Credit car	90.25	6369.45	No	
18773-HHU	Female	0	No	Yes	17	No	No	DSL	No	No	No	Yes	Yes	Month-to	Yes	Mailed ch	64.7	1093.1	Yes	
3841-HFEC	Female	1	Yes	No	71	Yes	Yes	Fiber optic	Yes	Yes	Yes	No	No	Two year	Yes	Credit car	96.35	6766.95	No	

Şekil 3. Veriler

4. Random Forest Algoritması

Random Forest algoritması Breiman tarafından geliştirilmiştir. Bu algoritmaya göre sınıflandırmanın amacı farklı eğitim kümelerinde eğitilmiş ağaçların kararlarını birleştirmektir. Dağılımları aynı, farklı eğitim kümeleri gelişmiş özellikler kullanır. Karar ağaçlarını oluşturmadan önce bütün seviyedeki kararlar incelenir. Diğer ağaçlarda da aynı işlemler yapılır ve en fazla kullanılan nitelik seçilir. Bu nitelik ağaca dahil olur ve aynı işlemler diğer seviyeler için tekrar edilir. Değişkenlerin sayısı ve oluşacak ağaçların sayısı algoritma başlamadan önce belirlenir. Bunlar her düğümde kullanılır. Random Forest algoritması, Classification ve Regression Tree algoritmalarını kullanır. Ağaç üretmek için düğüm ve dallar algoritmaya göre oluşturulacaktır.[4]



Şekil 4. Karar Ağacının yapısı

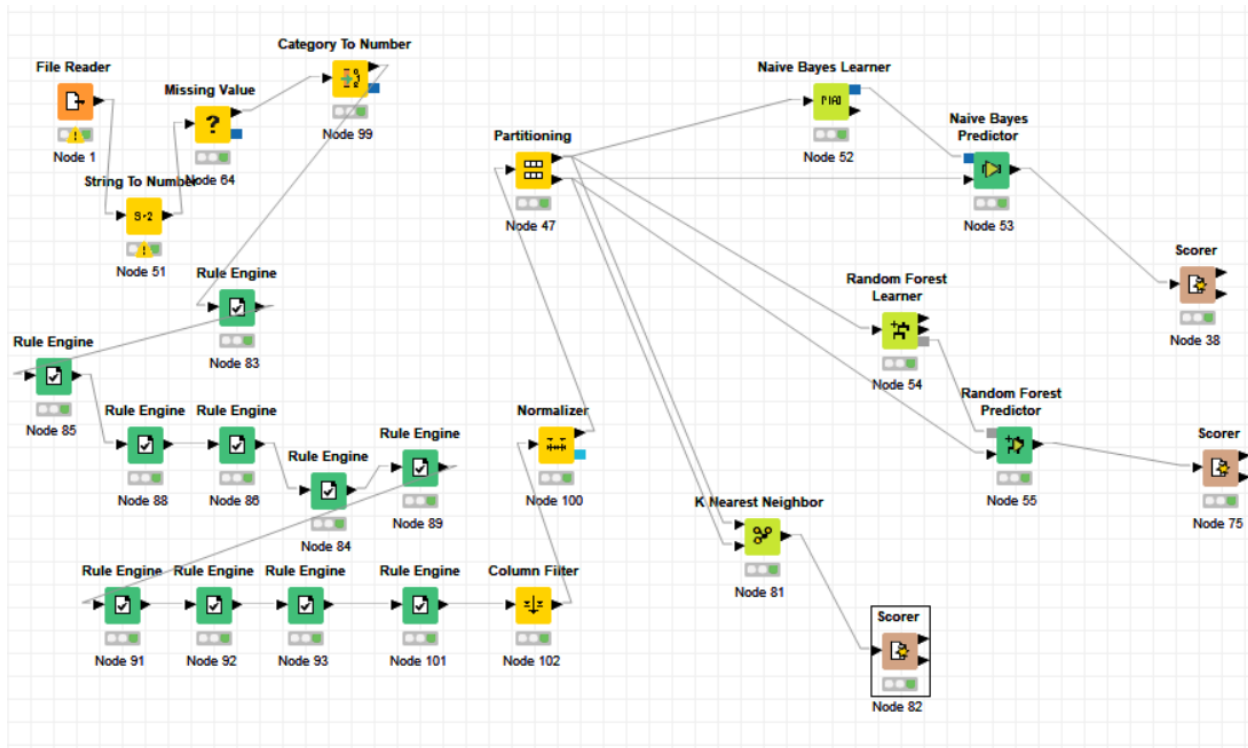
5. Kullanılan Programlar

5.1. Knime

Knime, Node Repository altında yer alan node'lar arasında ilişkilendirmeler yapılarak verinin işlenmesi, yorumlanması, görselleştirilmesi ve raporlanmasını sağlayan bir veri analiz platformudur. Bu çalışmada CRISP DM metodolojisi uygulanarak ilerlenilmiştir. Açık kaynak kodlu bir uygulamadır.[5]

File Reader operatörü ile veri içeri alınmalıdır ve configure edilmelidir. Browse'a tıklayıp veri seti seçilmelidir. Veriyi seçtikten sonra apply diyerek pencereden çıkılır. Daha sonra veri execute edilerek çalıştırılır. File table diyerek veri setimizin tamamını görebiliriz. Veriler doğru bir şekilde akıtılmış mı diye kontrol edilmelidir. Ayrıca Scatter Plot operatörü eklenerek istenilen kolonlar görselleştirilmektedir. X,Y kolonlarını seçerek hangi kolon birbirini ne kadar etkiliyor rahatça görülebilmektedir. Bir diğer görselleştirme operatörü Scatter Matrix'tir. Bu operatör çıktısında kolon sayısı artmakta ve daha detaylı görselleştirme yapmaktadır. Verilerde double görünmesi gereken kolonlar string olarak gözükmemektedir. O verileri düzeltmek için, String to Number operatörü kullanılmalıdır.

Bu çalışmanın model aşamasının bir kısmının Knime görseli Şekil 5'deki gibidir.



Şekil 5. Knime Modellemesi

Configure butonuna tıklanmaktadır ve açılan pencerede çevirilecek kolonlar seçilmektedir ve apply tıklanarak pencereden çıkılmaktadır. Daha sonra bu operatöre Missing Value operatörü bağlanmaktadır. Bu operatör eksik verileri düzenlemektedir. Configure butonuna tıklanır ve string ve number olan eksik verilerin yerine ne geleceği seçilmektedir. Burada Most Frequent Value string veriler için, mean değeri number olan veriler için seçilmektedir.

Apply diyerek çıkılmaktadır. Bu operatöre Category to Number bağlanmaktadır. Bu operatör Label encoder yapmaya yaramaktadır. Label encoder yapmak için Yes,No gibi iki özelliği olan kolonlar seçilmelidir. Configure edilerek açılan pencerede kolonlar seçilmektedir. Seçim bittikten sonra apply diyerek pencere kapatılmalıdır. Daha sonra bu operatöre ise Rule Engine operatörü bağlanmaktadır. Bu operatörün amacı one hot encoder yapmaktır. Configure diyerek açılan pencerede Expression bölümünde encoder yapmak istediğimiz kolonla ilgili kod yazılmalıdır. Her özellik için ayrı ayrı kolon oluşacaktır. Buradaki amaç kategorik verileri nümeriğe dönüştürmektir. Append kolon kısmından yeni bir kolon ismi vererek apply denir ve pencere kapatılır. Her bir kolon için Rule Engine operatörü eklenmektedir. Çünkü her operatör bir kolonu encoder etmektedir. Bu aşamada Column Filter operatörü eklenmektedir. Bazı kolonlar başarıyı düşürdüğü için filtreleme yapıldı. Bu işlemden sonra Normalizer operatörü çalışmaya eklenmektedir. Bu operatör ölçekleme yapmaya yaramaktadır. Ölçekleme farklı kategorideki verileri aynı dünyaya indirgemeye yaramaktadır. Buraya kadar veriler hazırlandı ve artık veri bölünebilir. Çalışmaya partitioning operatörü eklenmektedir. Configure diyerek Verimizi %60 eğitim %40 test olarak bölünür. Ok diyerek kapatılır. Bundan sonraki kısım için algoritmalar uygulanacaktır. Her bir algoritmanın başarıları görülecektir. K Nearest Neighbor operatörü eklenir. Bu operatöre Scorer operatörü eklendiğinde başarı oranını görmek için Confusion Matrix butonuna tıklanmalıdır. Bir başka operatör olan Naive Bayes Learner, Naive Bayes Predictor, Random Forest Learner, Random Forest Predictor operatörleri Partitioning operatörüne eklenir. Bu operatörlerin başarı oranları aşağıdaki tabloda görülmektedir.

Tablo 1.Knime Uygulamasında Operatörlerin Başarı Oranları

Random Forest Algoritması	Başarı Oranı %80.12
KNN Algoritması	Başarı Oranı %74.27
Naive Bayes Algoritması	Başarı Oranı %79.87
Decision Tree Learner	Başarı Oranı %83.36

5.2. Python Kodlaması

Burada veri dönüşümü yapılmaktadır. String veriler temizlenmektedir. String veriler temizlenmez ise doğru sonuçlar elde edilemez ve başarı düşer. Double olması gereken kolonlar içinde String ifadeler bulunmaktadır. Bu veriler temizlenerek daha başarılı sonuçlar elde edilebilir.

```
veri= veri[pd.to_numeric(veri['MonthlyCharges'], errors='coerce').notnull()]
```

```
veri['MonthlyCharges']=veri['MonthlyCharges'].astype('float')
```

Çalışmanın önemli adımlarından biri de verilerin eğitim ve test olarak bölünmesidir. Eğitim için %60 test için %40 olarak bölünmelidir. En sağlıklı sonucu almak için bu şekilde bölünmektedir. Eğitim ve test olarak bölmeden önce kullanabilmek için kütüphane import edilmelidir.[6]

```
x_train, x_test,y_train,y_test = train_test_split(a.iloc[:, :-1],a.iloc[:, -1:],test_size=0.33, random_state=0)
```

Veriler test ve eğitim olarak bölündükten sonra ölçekleme yapılmalıdır. Ölçekleme yaparken burada Standard Scaler kullanıldı. X train,fit tranform edildi, tekrar eğitildi. X test ise test için kullanıldı.

```
X_train = sc.fit_transform(x_train) X_test = sc.fit_transform(x_test)
```

Veriler hazırlandı,eğitildi ve ölçeklendi. Bundan sonra en başarılı görülen Random Forest algoritması kullanıldı. Her zaman yapıldığı gibi önce kütüphaneler import edildi. Daha sonra X train ve X test dataframleri algoritma içinde kullanıldı.

```
rfc = RandomForestClassifier(n_estimators=10, criterion = 'entropy')
rfc.fit(X_train,y_train)
```

En son olarak X train, X test fit edildi yani modelleme yapıldı. Bundan sonra başarı ölçülecek ve bunun için de confusion matrix kullanıldı. Confusion matrix yerine başka yöntemlerle de başarıyı ölçebilmektedir. Bu bir yöntemdir.

```
y_pred = rfc.predict(X_test)
cm = confusion_matrix(y_test,y_pred)
```

Confusion matrix çıktısı bir matristir. Bu matriste çaprazlama yaparak topladığımızda doğru verileri ve yanlış verileri anlamamız mümkündür.

Tablo 2.Algoritma Başarı Oranları

Random Forest Algoritması	Başarı Oranı %86.36
KNN Algoritması	Başarı Oranı %85.32
NaiveBayes Algoritması	Başarı Oranı %80.39
Lojistik Regresyon Algoritması	Başarı Oranı %75.36
Decision Tree Learner	Başarı Oranı %83.38

6.Sonuç

Pythonda kodlamaya başlamadan önce ilk olarak kütüphaneler import edilir. Daha sonra veri seti içeri alınır. Veri seti ve kodlar aynı dosyanın içinde bulunmalıdır. Eğer eksik veriler var ise bununla ilgili düzeltmeler yapılmalıdır. Eksik verilerden sonraki işlem kategorik verileri nümerik verilere dönüştürmektir.

Label encoder ve one hot encoder yapılmalıdır. Burada kukla değişken tuzağına düşülmemelidir. Bundan sonra encoder yapılan özellikler kolon haline getirilmelidir. Çevirilen kolonlar birleştirilmelidir. Veriler hazır hale gelince, eğitim ve test olarak bölünmelidir.

Bölme işleminden sonra ölçekleme yapılmalıdır. Ölçekleme uygulandıktan sonra en başarılı sonucu gördüğümüz Random Forest algoritması uygulanmalıdır. Son olarak Confusion Matrix ile başarı oranı görülebilmektedir. Scatter Plot ile de görselleştirme yapılması mümkündür.

Çalışma sonuçları incelendiğinde Random Forest algoritmasının model testine ait %86.36 doğruluk derecesiyle en iyi sonucu ürettiği söylenebilir. Doğruluk ölçütü oldukça basit ve önemli bir kriterdir. Diğer algoritmaların da başarı oranları yakın görülmektedir. Random Forest algoritması yapısı bakımından Decision tree algoritmasına benzerdir. Birden fazla ağaçtan oluştuğu için avantajlıdır. Her bir yaprakta bir karar vardır denilebilir. En iyi kararı vererek algoritmanın başarısını arttırmaktadır.

Kaynakça

- [1]. Coşkun, Cengiz, and Abdullah Baykal. "Veri Madenciliğinde Sınıflandırma Algoritmalarının Bir Örnek Üzerinde Karşılaştırılması." *Akademik Bilişim 2011*
- [2]. Seker, Sadi Evren. "Öznitelik Mühendisliği (Feature Engineering)." *YBS Ansiklopedi*, (2017)
- [3]. <https://mertricks.com/2014/11/29/veri-madenciligi-3-veri-hazirlama-bolum-1/>
- [4]. Daş, Bihter, and İbrahim Türkoğlu. "DNA dizilimlerinin sınıflandırılmasında karar ağacı algoritmalarının karşılaştırılması." (2014).
- [5]. Seker, Sadi Evren, «Her Seviyeye Uygun Uçtan Uca Veri Bilimi, Knime ile,» UDEMY. <https://www.udemy.com/veribilimi/learn/v4/content>. (2018).
- [6]. Sengün, Cagdas. "Veri Madenciliği Örneği : Ev Fiyatları". *YBS Ansiklopedi*, (2018)