**1.Giriş**

Üretimin dahi dijitalleştiği günümüz dünyasında, tüketim gittikçe artan bir ivme göstermektedir. İnsanların artık tek tıkla alışveriş yapabiliyor olmaları bu durumun başlıca sebebidir. Satın almanın giderek arttığı bu çağda, satışları tahmin edebilmek ve ona göre stok tutmak ya da üretim planlaması yapmak büyük önem kazanmıştır. Tahmin yapılmadan işletmelerin ne gelirleri ne de giderleri öngörülebilir. İşletmenin üretim planlama, satın alma, satış ve pazarlama gibi birçok departmanı bu tahminler üzerinden analiz yapmaktadır. Uzun ya da kısa vadede satış tahmininin doğru yapılamaması müşteri memnuniyetsizliği, para kaybı, ham madde ihtiyacı gibi birçok soruna yol açacaktır. Tahmin yapılmadan geleceğe dönük stratejik kararlar alınamayacağından işletmenin plan ve politikalar oluşturamamasına sebep olur. Bu durum sadece işletmenin kendisini etkilemekle kalmayıp tedarik zinciri elemanlarından üretici, perakendeci, tedarikçi ve müşteriye kadar birçok tarafın zarar görmesine sebep olur.

Hiçbir yöntem kullanılmadan büyük çaplı satışları tahmin edebilmek, işletmeler için imkânsız hale gelmiştir. Klasik yöntemler ise bu büyük veri yığınlarının oluşturduğu satışlara cevap verememekte ya da esneklikleri ve özel durumları hesaplayamamaktadır. Bu sebeple literatüre yeni kazandırılan yapay zekâ bileşenleri benzer sonuçlara cevap vermiştir. Yapay zekâ çağının getirdiği yeniliklerden olan makine öğrenmesi de birçok mühendislik uygulamasının getirdiği sorunlara hızlı şekilde cevap verebilecek bir alandır.

Bu çalışmada lineer regresyon, Ridge, Lasso, ElasticNet, K-En Yakın Komşu ve Rastgele Orman yöntemleri kullanılarak satış tahmini için model oluşturulmuş ve tahmin yöntemlerinde makine öğrenmesi algoritmalarının kullanılmasının gerekliliği üzerinde durulmuştur.

Çalışmanın ikinci ve üçüncü bölümünde sırasıyla satış tahmini ve makine öğrenmesi konuları ile ilgili literatür araştırması yapılmıştır. İzlenen metotun işlendiği üçüncü bölümde bir makine öğrenmesi modeli oluşturulurken izlenmesi ve dikkat edilmesi gereken adımlar keşifçi veri analizinden model optimizasyonuna kadar verilmiştir. Beşinci bölümde gerçek veri seti ile yapılan örnek bir satış tahmini çalışması yapılmış, sonuçlar altıncı bölümde işlenmiştir.

**2. Literatür Taraması: Satış Tahmininde Makine Öğrenmesi Uygulamaları**

Satış tahmin modelleri iki türe ayrılabilir: (i) zaman serisi yöntemleri ve (ii) derin öğrenme ve makine öğrenmesi algoritmalarını içeren yapay zeka yöntemleri.

XGBoost yöntemi, Avrupa eczane perakende şirketinin olası satışlarını tahmin etmek için Rastgele Orman ve Doğrusal Regresyon yöntemleriyle karşılaştırılmıştır. Sonuca göre XGBoost yönteminin diğerlerine göre çok daha başarılı bir şekilde çalıştığı görülmüştür. İleride yapılacak çalışmalar için, satış tahmininin yanı sıra şu 6 alanda da aynı eylemlerin gerçekleştirilebileceği önerilmiştir: reklam, öneriler, talep tahmini, müşteri bazlı fiyatlandırma, tatil / uzatılmış satış planlaması ve ürün sınıflandırması.

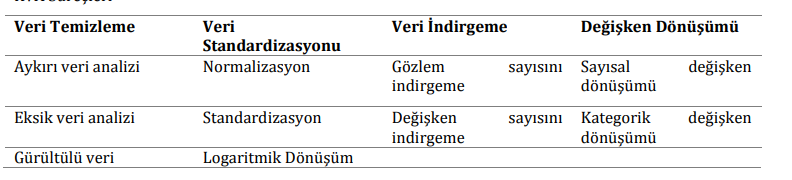
Microsoft Azure Machine Learning Studio platformunda Kaggle'dan Walmart satış verilerini kullanan satış tahmini çalışmasında, klasik zaman serisi yöntemleri makine öğrenmesi algoritmalarıyla karşılaştırılmıştır. Doğrusal, Bayes, Sinir Ağları, Karar Ağacı ve Artırılmış Karar Ağacı regresyonları kullanılmıştır. Doğrulama fonksiyonu olarak ortalama kare hata ve kök ortalama kare hata tercih edilmiştir. Karar Ağacı Regresyon modeli başlangıçta en düşük hatayı verirken, model tuning uygulandıktan sonra Artırılmış Karar Ağacı regresyonunun en düşük hatayı verdiği görülmüştür. Bu sonuç ise regresyon modellerinin klasik zaman serisi modellerinden daha iyi çalıştığının bir başka kanıtıdır.

**3. Yöntem**

Yöntem bölümünde keşifçi veri analizinden başlanarak model optimizasyona kadar uçtan uca makine öğrenmesi model kurma ve veriyi analiz etme süreci ifade edilmiştir.

**3.1 Keşifçi Veri Analizi**

Model KVA olmadan yapılırsa hata artar. Çünkü veri setinde bazı bozulmalar olabilir. Bunu önceden analiz etmek gerekmektedir. Verileri anlamak için öncelikle veri okuryazarlığı gereklidir.



**3.2. Makine Öğrenmesi Algoritmaları**

Bu bölümde makine öğrenmesi algoritmalarından doğrusal regresyon, Ridge, Lasso, ElasticNet, Rastgele Orman ve K-En Yakın Komşu algoritmaları ele alınmıştır.

**3.3 Model Performans Değerlendirme**

Model oluşturulduktan sonra performansı değerlendirilmelidir. Aksi takdirde oluşturulan modelin verimliliği analiz edilemez. Bu model hakkında yorum yapılmasını ve doğru sonuca ulaşılmasını engelleyen bir durumdur. Bunun için modelin önce Hold-out, K-katlı çapraz doğrulama, Leave One Out ve Bootstrap gibi yöntemlerle *doğrulanması* gerekir. Model doğrulandıktan *sonra hata fonksiyonları incelenmelidir*. Böylelikle eğitim seti ve veri setinden parçalanmış test seti için modelin hataları bulunacaktır. Daha sonra eğitim seti ile test seti arasındaki ilişkinin hatalar açısından incelenmesine olanak sağlar. *Yanlılık ve varyans durumları* incelenir ve modelde aşırı uyum veya yetersiz uyum olup olmadığı kontrol edilir.

**4. Ofis Malzemeleri Satışı Yapan Firma İçin Satış Tahmini**

Örnek çalışma, perakende ofis malzemeleri satışı yapan bir firmada gerçekleştirilmiştir.

Örnek çalışma, perakende ofis malzemeleri satışı yapan bir firmada gerçekleştirilmiştir. Alınan veri seti 01.01.2019 ile 31.08.2020 tarihlerini kapsamakta olup yirmi aylık veri içermekle birlikte 1470 satır ve 26 sütundan oluşmaktadır. Python programlama dili kullanılarak JupyterLab platformunda yazılan veri setine ilişkin bilgiler verilmiştir. İlgili veri setindeki 26 sütundaki eksik veri sayıları verilmiştir. Bu sütunlar 4 adet veri tipi içermektedir. Kategorik ve numerik değişkenler içeren bu veri setinde, eksik veri ve aykırı veriler veri tipine göre analiz edilmiştir.

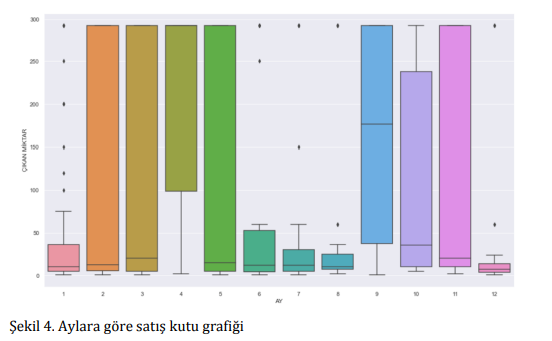
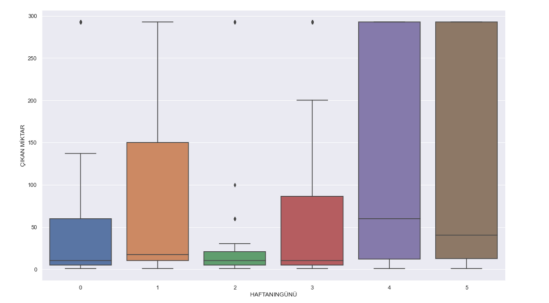
Keşifçi veri analizi: Verinin yapısını anlayabilmek adına keşifçi veri analizi yapılmıştır ve modelde kullanılmayacak olan 18 sütun düşürülmüş, kalan 8 sütun üzerinden model çalışılmıştır.

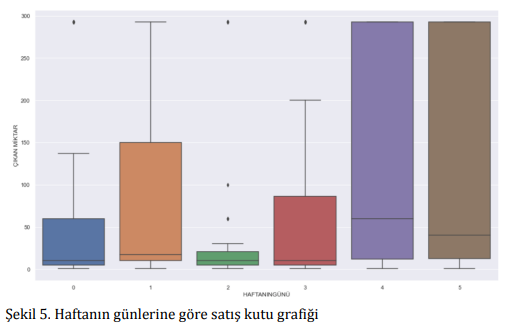
Eksik veri analizi: Eksik veriler analiz edildiğinde Şekil 3’te verildiği gibi FATURA SIRA NO’ya dair 16 eksik veriye rastlanmıştır. Eksik veriler sıra no içerdiğinden ortalama ya da herhangi bir yöntem ile verileri doldurmak mantıksız olacaktır. Bu sebeple ilgili veriler veri setinden çıkarılmıştır.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 4’teki kutu grafiği aylara göre satış miktarını gösterirken, Şekil 5 haftanın günlerine göre satış miktarını göstermektedir. Pazar günü satış yapılmadığından altı gün gösterilmiştir. Örneğin grafiğe göre cuma ve cumartesi günleri en çok satış yapılırken çarşamba günü yapılan satış en azdır. Bu da şirkete, alacağı aksiyon kararlarında fikir verecektir. Benzer durum aylara göre yapılan satışlarda da geçerlidir.



Modelin kurulması aşamasında ise öncelikle LinearRegression, Ridge, Lasso, ElasticNet, KNeighborsRegressor ve RandomForestRegressor kütüphaneleri indirilmiştir. ÇIKAN MİKTAR ve TARİH üzerinden sklearn.model\_selection() kütüphanesinden train\_test\_split() fonksiyonu çağırılarak model validasyonu gerçekleştirilmiş, modele en uygun eğitim ve test verileri oluşturulmuştur. Eğitim ve test verilerinin fonksiyon kullanılarak seçilmesinin sebebi, ilgili fonksiyonun veri setini okuyarak uygun şekilde bölmesinden ileri gelmektedir. x\_cv ve y\_cv olmak üzere oluşturulan eğitim verileri, x\_train ve y\_train olarak oluşturulan test setleri üzerinde test edilmiştir. Kendi içinde iki ayrı veri seti olarak gözükmesine rağmen ilgili veri setleri bağımlı ve bağımsız değişkenleri ifade etmesine göre ayrılmıştır. Burada y\_cv ve y\_train veri setleri sadece hedef değişken olan ANA DÖVİZ NET TUTAR sütununu içermektedir. Eğitim ve test verisi lin.fit(x\_train, y\_train) komutu ile fit edilmiş ve lin\_predict() komutu ile x\_cv veri seti tahmin edilmek için çağırılmıştır. İlgili algoritmaları içeren, yukarıda belirtilen kütüphaneler ve 𝑟 2 için iki dizi oluşturularak çağırılmış ve veri setleri bu haliyle girdi olarak algoritma kütüphanelerinde işlenmek üzere okutulmuştur. Modelden models.sort\_values(by="r\_2", ascending=False) komutu ile 𝑟 2 skorunu azalan şekilde vermesi istenmiş ve Şekil 6’daki 𝑟 2 skorları elde edilmiştir. Elde edilen skorlara göre en iyi tahmincinin Rastgele Orman yöntemi olduğu görülmüştür.

Rastgele Orman yöntemine göre model hipertune edilmiştir. K-En Yakın Komşu algoritması için best\_params\_ fonksiyonu kullanılarak en iyi parametreler 2 olarak bulunmuştur. effective\_metric fonksiyonu ile veri setine uygun ölçümün Öklid olduğu belirlenmiştir. Bu işlemlere göre knn\_tuned adında bir dizi tanımlanmış ve bulunan veri setine ait özellikler içine yerleştirilmiş ve KNeighborsRegressor kütüphanesi tekrar çağırılarak x\_train ve y\_train veri setleri fit edilmiştir. Yapılan işlemlerden sonra 𝑟 2 skorunun 0,5302 olduğu görülmüştür. Rastgele Orman algoritması için ise max\_depth, max\_features ve n\_estimators parametreleri kullanılarak rf\_tuned adında bir dizi oluşturulmuştur. Veri seti ilgili parametreler için işleme koyulduktan sonra en iyi parametrelerin max\_depth için 9, max\_features için 5 ve n\_estimators için 200 olduğu bulunmuştur. Bu verilere göre RandomForestRegressor kütüphanesi ile algoritma tekrar çağırılmış ve 𝑟 2 skorunun 0,9726 olduğu hesaplanmıştır. Lineer regresyon ile 0,7213 ile elde edilen skor, Ridge, Lasso ve ElasticNet algoritmaları ile sırasıyla 0,7134, 0,7128, 0,7189 skorları elde edilmiştir. Yapılan analizin sonucunda 8,3 dakika ile en uzun süre Rastgele Orman algoritmasına ait olmasına rağmen en iyi sonucun elde edildiği gözlemlenmiştir. Model hipertune edildikten sonra elde edilen 𝑟 2 skorunun 0.9726 olduğu görülmüştür. Bu da Rastgele Orman yöntemiyle yapılan satış tahminin diğer yöntemlere göre iyi bir şekilde çalıştığının göstergesi niteliğindedir. Model oluşturulduktan sonra veriler istenilen tarih aralığı için tahmin yapmaya uygun haldedir.

**5. Sonuç**

Algoritmalar arasından en düşük hatayı veren algoritma Rastgele Orman algoritması olmuştur. Bu sebeple ilgili veri setinde belirtilen algoritmalar arasından Rastgele Orman algoritması kullanılarak tahminin yapılması uygun görülmüştür. Aylık veya haftalık olarak istenen verilerden tahmin yapılması mümkündür. Aynı zamanda ilgili işletmenin pazara yeni bir ürün koyması durumunda işletmenin verileri analiz edilmiş olduğundan yeni bir ürün için de tahmin yapılabilmesi mümkündür. Bu çalışmada geleneksel tahmin yöntemlerinde eksik veri ve aykırı verilerin analiz edilmemesinden doğan problemler aşılmaya çalışılmıştır. Satış tahmini ile ilgili literatürdeki yöntemlere kıyasla bu çalışmanın iki ana avantajı vardır. Öncelikle geçmişte diğer yöntemlere dayalı satış tahmini çalışmalarında karşılaşılan hız sorunu çözülmüştür. Bir firma yeni bir ürünü ilgili sektörün pazarına sürmek istediğinde benzer davranış gösteren ürünler dikkate alınarak satış tahmini kolaylaşacaktır. İkincisi, aykırı değerlere ve eksik değerlerden doğan tahmin eksiklikleri Keşifçi Veri Analizi ile çözülmüştür. Yapılan satış tahmini çalışması, farklı sektörlere de uygulanabilir olup diğer makine öğrenmesi algoritmaları da kullanılarak çalışmanın içeriğini genişletmek mümkündür. Bu çalışmanın sadece belirli alanlarda uygulanan makine öğrenmesi çalışmalarının, gerçek hayata adapte olması sebebiyle ilgili araştırmacılara rehber olması beklenmektedir. Gelecek çalışmalar birçok farklı sektörde gerçekleştirilebilir ve endüstrinin gerektirdiği diğer tahmin konularına uygulanabilir. Yöneticilerin tahminleri analiz etmesine izin verecek bir ara yüz geliştirmek de mümkün olacaktır. (Emine Nur NACAR, 2021)