T.C.
SAKARYA ÜNİVERSİTESİ
MÜHENDİSLİK FAKÜLTESİ

ENDÜSTRİ MÜHENDİSLİĞİ BÖLÜMÜ

BİTİRME ÇALIŞMASI

HAVAYOLU HİZMET KALİTESİNE AİT MÜŞTERİ MEMNUNİYETİ ANKET VERİLERİNİN VERİ MADENCİLİĞİ YAKLAŞIMLARI İLE SINIFLANDIRILMASI VE ANALİZİ

HAZIRLAYAN

Pınar KIZILTAN

Okancan ÖZDEMİR

DANIŞMAN

Dr. Öğr. Üyesi Berrin DENİZHAN

MAYIS 2020

İÇİNDEKİLER

TEŞEKKÜR 3

TABLOLAR LİSTESİ 4

ŞEKİLLER LİSTESİ 5

ÖZET 6

BÖLÜM 1.GİRİŞ 7

1.1.Müşteri Memnuniyeti 7
1.2.Literatür Araştırması 11
BÖLÜM 2.PROBLEMİN TANIMI 12

BÖLÜM 3.PROBLEMİN ÇÖZÜMÜ İÇİN TAKİP EDİLECEK AŞAMALAR 13

3.1.Verilerin İndirilmesi, İncelenmesi, Temizlenmesi Ve Düzenlenmesi 13
3.2. Eğitim Verisine Sınıflandırma Uygulamaları 13
3.2.1.K Nearest Neigborhood Yöntemi 14
3.2.2.Decision Tree Yöntemi 16
3.3.Sınıflandırılan Eğitim Seti Sonuçlarının İncelenmesi, Yorumlanması Ve Yöntemlerin Karşılaştırılması 22
3.4.Özelliklere Göre Eğitim Veri Setinin İncelenmesi Ve Son Düzenlemerin Yapılması 22
3.5.Test Verisine Sınıflandırma Uygulanması 22
3.6.Sınıflandırma Test Verisi Seti Sonuçlarının İncelenmesi Ve Yorumlanması 22

BÖLÜM 4.UYGULAMA 23

4.1.Verilerin İndirilmesi, Temizlenmesi Ve Düzenlenmesi 23
4.2.WEKA'da Eğitim Verilerinin Sınıflandırılması 25
4.2.1.KNN Algoritması İle Sınıflandırma 26
4.2.2.Decision Tree Algoritması İle Sınıflandırma 30
4.3.Sınıflandırılan Verilerin İncelenmesi, Uygulanan Yöntemlerin Karşılaştırılması Ve Uygun Modelin Seçimi 33
4.4.Özellikler Açısından Verilerin Analizi Ve Son Düzeltmeler 34
4.5.WEKA'da Test Verilerinin Sınıflandırılması 37
4.6.Sınıflandırılan Test Verilerinin İncelenmesi Ve Yorumlanması 40

BÖLÜM 5.İSTATİSTİKSEL ANALİZ 42

BÖLÜM 6.GERÇEKCİ KISITLAR VE KOŞULLAR ALTINDA DEĞERLENDİRME 45

BÖLÜM 7.SONUÇ 46

KAYNAKLAR 47

TEŞEKKÜR

Tasarım ve bitirme çalışmamız boyunca desteğini esirgemeyen, bilgi ve tecrübesini bizlerle paylaşarak, çalışmamıza yön veren ve bu çalışmayı gerçekleştirmemize olanak sağlayan, değerli tez danışmanımız **Sn. Dr. Öğretim Üyesi Berrin DENİZHAN'a** teşekkürlerimizi sunarız. Ayrıca maddi ve manevi her türlü desteği veren **Sevgili Ailelerimize** ve eğitim hayatımızda bu konuma gelmemizi sağlayan herkese teşekkür ederiz.

TABLOLAR LİSTESİ

Tablo 1. Yeni nesnenin var olan en yakın nesnelere öklid uzaklığı

Tablo 2. En yakın ilk 3 komşu ve sınıflar

Tablo 3. Örnek Veri Kümesi

Tablo 4. Örnek Veri Kümesinin 3 Gruba Ayrılması

Tablo 5. Sınıflandırılan Test Verileri

Tablo 6. Farklı Algoritmaların Başarı Oranları

Tablo 7. Farklı K Değerleri İçin Başarı Oranları

Tablo 8. Cinsiyet Verileri Çıkarılan ve Çıkarılmayan Başarı Oranları

Tablo 9. Cinsiyet ve Kalkış Varış Uygunluğu Verileri Çıkarılan ve Çıkarılmayan Başarı Oranları

Tablo 10. Cinsiyet, Kalkış Varış Uygunluğu ve Kalkış Gecikmesi Verileri Çıkarılan ve Çıkarılmayan Başarı Oranları

ŞEKİLLER LİSTESİ

Şekil 1. Müşteri Memnuniyetini Etkileyen Faktörler

Şekil 2. Müşteri memnuniyeti, Müşteri Sadakati ve Müşteri Değeri Oluşturma

Şekil 3. Veri kümesine yeni bir veri yüklenmesi

Şekil 4. Yeni eklenen verinin sınıflandırılması

Şekil 5. Karar Ağacının Alabileceği Olasılıklardan Biri

Şekil 6. Karar Ağacının Alabileceği Olasılıklardan Bir Diğer Örneği

Şekil 7. WEKA Programında Eğitim Verisini Görselleştirme

Şekil 8. WEKA Programında Cross-validation Yöntemi ile KNN Algoritması Kullanılarak Sınıflandırma İşlemi

Şekil 9. Farklı K Değerlerine Karşılık Gelen Farklı İsabetlilik Oranları

Şekil 10. Uygulama Sonucu Confusion (Karışıklık) Matrisi

Şekil 11. Confusion Matrisinin Taslağı

Şekil 12. Eğitim Verisinin KNN ile Doğruluk ve Hata Oranı

Şekil 13. Eğitim Verisinin KNN ile Detaylı Doğruluk Oranları

Şekil 14. WEKA'da Classify Sekmesi Bölümleri

Şekil 15. WEKA'da Cross-Validation Yöntemi ile Decision Tree Algoritması Sınıflandırma İşlemi

Şekil 16. Eğitim Setinden Oluşturulan Karar Ağacı Modeli(Önizleme)

Şekil 17. Uygulama Sonucu Confusion (Karışıklık) Matrisi

Şekil 18. Eğitim verisinin Decision Tree İle Doğruluk Oranları

Şekil 19. Eğitim Verisinin Decision Tree İle Detaylı Doğruluk Oranları

Şekil 20. Özelliklerin Önem Kıyaslama İşleminin Uygulaması

Şekil 21. Özelliklerin Etki Sıralaması

Şekil 22. Eğitim Verisinin Cinsiyet Verisi Çıkarılmış Halinin Decision Tree ile Doğruluk Oranları

Şekil 23. Eğitim Verisinin Cinsiyet ve Kalkış-Varış Zaman Uygunluğu Verileri Çıkarılmış Halinin Decision Tree ile Doğruluk Oranları

Şekil 24. Eğitim Verisinin Cinsiyet, Kalkış-Varış Zaman Uygunluğu ve Kalkış Gecikmesi Verileri Çıkarılmış Halinin Decision Tree ile Doğruluk Oranları

Şekil 25. WEKA'da İstenilen Verinin Eğitim Seti Olarak Kullanılması

Şekil 26. Eğitim Verisi Olarak Kullanılacak Seti Kendisini Sınıflandırmada Başarı Oranı

Şekil 27. WEKA'da İstenilen Verinin Test Seti Olarak Kullanılması

ÖZET

Günümüz havacılık sektöründe müşterilere sunulan hizmetlerin özellikleri, fiyatları ve teknolojileri arasındaki benzerliğin artması, firmalara bağlı müşteri kitlesinin oluşturulmasını zorlaştırmıştır. Bu bağlamda, müşterilerin havayolu firmalarına yönelik düşüncelerinin, inançlarının ve algılarının bilinmesi, kurum kimliğinin oluşturulmasını, olumlu kurum imajının ve müşteri sadakatinin yaratılmasını önemli bir konuma getirmiştir. Havayolu firmaları açısından rekabet avantajından yararlanarak, sürdürülebilir pazar payı ve başarı sağlamada, hizmet kalitesi ve müşteri sadakati kavramları ile güçlü bir kurum imajı oluşturmalarına bağlı hale gelmiştir. Bu kapsamda, araştırmanın amacı havayolu firmalarında, müşteri sadakatinin yaratılmasında kurum imajının ve algılanan hizmet kalitesinin etkisini değerlendirmektir. Araştırmada Amerika'da bir havayolu firmasının yaklaşık 100000 (eğitim veri seti) yolcusu için yaptıkları yolculuk sonrasında yolculuk detayları (yolculuk sınıfı, yolculuğun mesafesi, yiyecek içecek, wifi hizmetleri vs.) ve son olarak yolculuktan memnun kalıp kalmadıklarını belirttikleri bir memnuniyet anketi verileri kullanılmıştır. Ardından bu veriler, memnun kalıp kalmamaya dair bir ifade bulundurmayan ancak diğer yolculuk detaylarını barındıran yaklaşık 25000 farklı anket verisinin (test veri seti), en başarılı şekilde memnun kalma veya kalmama şeklinde sınıflandırılması için kullanılmıştır. Bu işlem gerçekleştirilirken uygun veri madenciliği sınıflandırma uygulamaları tespit edilmeye çalışılmış, tespit edilen 2 farklı algoritmanın(KNN ve Decision Tree) çalışma prensipleri aktarılmıştır. Aynı zamanda hangi anket sorularının memnuniyet kararını en çok etkilediği araştırılmış, başarı oranını düşüren anket soruları ve verileri tespit edilerek anket verileri tekrardan düzenlenmiştir. Uygulanan algoritmalar karşılaştırılmış, en yüksek doğruluk oranına sahip olan düzenlenmiş veriler ve algoritma kullanılarak test verileri sınıflandırılmıştır. Gerçekleştirilen işlemler ücretsiz olarak sunulan WEKA programı kullanılarak gerçekleştirilmiştir. Ayrıca gerçekleştirilen işlemlerin programda nasıl ve ne şekilde uygulandığı detayları olarak aktarılmış ve ekran görüntüleri ile desteklenerek aynı zamanda eğitici bir yaklaşım yakalanmaya çalışılmıştır. Araştırma sonuçlarının havayolu firmalarının algılanan hizmet kalitesini ve kurum imajını arttırmalarında ve müşteri sadakatini oluşturmalarında yapmaları gereken faaliyetlere yönelik oluşturulan öneriler doğrultusunda, yol gösterici olması beklenmektedir.

Anahtar Kelimeler: Havayolları, Müşteri Memnuniyeti, Veri Madenciliği, Sınıflandırma Yaklaşımları, WEKA

BÖLÜM 1. GİRİŞ

Günümüzde havayolu firmaları artan rekabet koşulları altında öncelikle müşterilerin istek ve beklentilerini anlayarak, müşterilerin ihtiyaçlarını karşılamaya yönelik en uygun hizmeti sunmaya odaklanmalıdırlar.

Hizmet kalitesi, şirket-müşteri alışverişinin müşteri tarafındaki faydaları ile ilgilidir. Havayolu endüstrisi, hizmet sektöründe önemli bir rol oynamaktadır. Yolcuları dünyanın her yerine ulaştırma yeteneği ile diğer sektörlere önemli ölçüde katkıda bulunmaktadır(Whau ve Kao, 2009).

Havayolu endüstrisi, küresel ekonomide önemli bir rol oynamaktadır. Eğlence ile turizm endüstrisinin hayati bir bileşenidir ve uluslararası işlerin yürütülmesi için vazgeçilmezdir. Endüstri, çevre etkilerine çok duyarlıdır. Endüstrinin genel karlılık problemlerini birkaç faktör oluşturmaktadır. Bunlardan ilki, ekonomik durgunluğun önde gelen bir göstergesi olan sektörün döngüsel doğasıdır. Ürüne olan talep bayram ve iş seyahatleri gibi ilgili faaliyetlere olan talebe bağlıdır. Ürün bozulabilir ve talep mevsime, haftanın gününe ve günün saatine göre değişir (Tiernan, 2008).

Son yıllarda, havayolu firmaları arasında hem fiyat serbestisi hem de yolcuların hizmet kalitesi bilincinin artması nedeniyle büyük bir rekabet yaşamaktadır. Günümüzün son derece rekabetçi hava taşımacılığı ortamında, havayolu firmaları sadece daha uygun güzergahlar oluşturmaya ve uçuşların sıklığını artırmaya değil, aynı zamanda daha fazla promosyon teşvikleri de getirmeye çalışmaktadır. Bununla birlikte, bu pazarlama stratejilerinin marjinal faydaları giderek azalmaktadır çünkü havayolu şirketlerinin çoğu aynı pazarlama stratejilerini uygulamaktadırlar. Dolayısıyla havayolu firmaları artık hizmet kalitesini nasıl geliştirmeleri gerektiğine odaklanmaktadırlar. Hizmet kalitesinin ne olduğunu anlamak, sürdürmek ve iyileştirmek, bugün havayollarının ana endişeleridir (Chou, 2011).

Günümüzün rekabetçi ortamında havayolu endüstrisinde yüksek kaliteli hizmet sunmak, bir pazarlama gereksinimi haline gelmiştir. Bununla birlikte küresel ekonomik gerileme ile, çoğu havayolu sadece hayatta kalmak için mücadele etmektedir. Maliyetleri ve hizmetleri olabildiğince azaltmaya zorlanmaktadırlar. Bu nedenle, havayolu firmaları yöneticilerinin müşterilerinin ne istediğini ve istemediklerini belirlemesi şarttır. Havayolu hizmet kalitesinde yolcuların yüksek hizmet algısını korurken, önemli hizmet kalemlerini tutmanın ve daha az önemli hizmet kalemleri için harcanan çabaları en aza indirmenin yolları, havayolu şirketlerinin ihtiyaç duyduğu yollardır (Liou, 2011).

1.1.Müşteri Memnuniyeti

"Müşteri her zaman haklıdır" sloganı, yüksek bir önceliği ve müşteri memnuniyetinin önemini vurgulamaktadır. Şirketler mevcut müşterileri korumanın yeni müşteriler edinmekten daha kârlı olduğunu bilmektedirler. Havayolu sektöründeki yoğun rekabet göz önüne alındığında, sadece müşteri memnuniyeti yeterli değildir. Daha ziyade, marka sadakatini başlatabilecek ve daha sonra rekabet etmede kolaylık sağlayabilecek müşterilerle daha uzun süreli ilişkiler kurulması gerekmektedir(Hussain,2015).

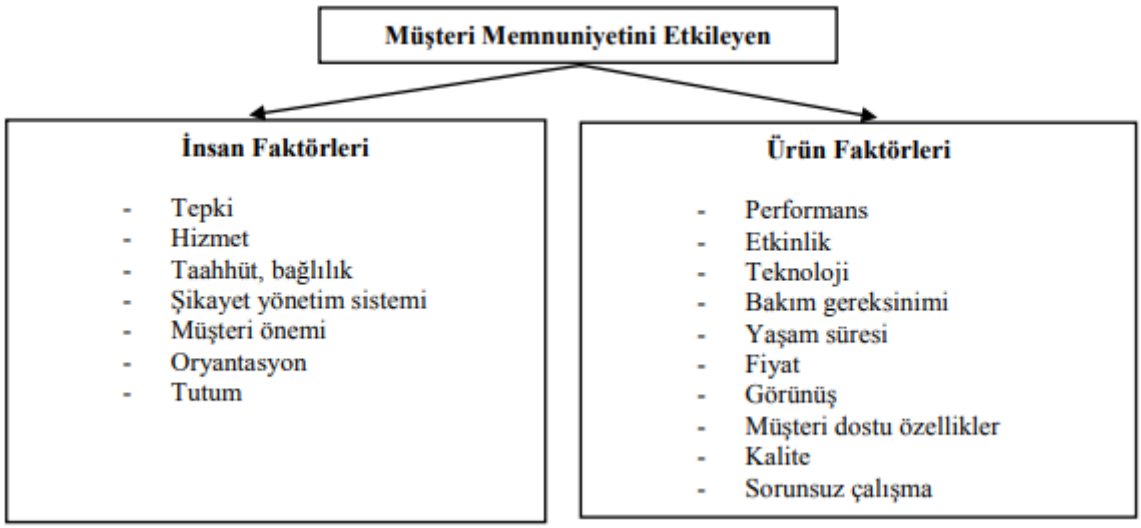
Müşteri memnuniyetinin en az iki farklı kavramsallaştırmasından bahsedilmektedir. Bunlar, işleme özgü ve kümülatif memnuniyettir. İşleme özgü memnuniyet, belirli bir ürün veya hizmet karşılaştırması hakkında belirli bir bilgi sağlayabilirken, kümülatif memnuniyet, firmanın geçmiş, şimdiki ve gelecekteki performansının daha temel bir göstergesidir (Anderson,1994). İşleme özgü bakış açısına göre müşteri memnuniyeti müşterinin son satın alma deneyimlerine dayanan değerlendirmedir. İşleme özgü bakış açısıyla kıyaslandığında, kümülatif perspektif, genel değerlendirmelerin altını çizerek, müşteri memnuniyetinin müşterinin son birkaç satın almasına değil bütün satın alma deneyimlerine bağlı olduğunu belirtmektedir (Kuo, 2009). Bir şirketin yüksek müşteri memnuniyeti seviyesinden birçok faydası bulunmaktadır. Yüksek müşteri memnuniyeti müşteri sadakatini arttıracak ve müşteri kaybını önleyecek, müşterilerin fiyat hassasiyetini azaltacak, başarısız pazarlama maliyetlerini düşürecek ve yeni müşteri kazandıracaktır. Müşteri sayısının artması nedeniyle işletme maliyetleri düşecek, reklamın etkinliği ve işletme itibarı artacaktır (Kim, 2004). Müşteri memnuniyeti genellikle geri alım niyeti ve müşteri sadakatinin önemli bir belirleyicisi olarak görülmektedir (Deng, 2010). Müşteri sadakati çok önemlidir çünkü sadık müşteriler bir firmaya birçok fayda sağlar. Bunlardan bazıları; sürekli bir kar akışının olması, pazarlama maliyetlerinin azaltması ve müşteri başına gelirin artmasıdır (Yi ve La, 2004). Sadık müşterilere hizmet verme maliyetleri daha azdır, Sadık müşteriler daha az fiyata duyarlıdır ve şirketle daha fazla zaman geçirir, favori markaları veya tedarikçileri hakkında olumlu tavsiyelerde bulunur (Kumar ve Shah, 2004).

Müşteri memnuniyeti, herhangi bir hizmet endüstrisinin nihai hedefidir (Singh, 2015). Hizmet kalitesi, her bir olayda hizmetlerin bilişsel bir değerlendirmesidir, ancak memnuniyet, müşterilerin hizmetleri değerlendirilmesinde oluşan birikmiş bir etkidir (Saha ve Theingi, 2009). Havayolu hizmet kalitesi boyutlarından uçuş içi hizmetler, uçuş içi dijital hizmetler ve hava yolu arka ofis operasyonları, yolcuların memnuniyetinin önemli belirleyicileridir. Bu memnuniyet müşteri sadakatini ve havayolu şirketinin imajını etkilemektedir. Havayolu

sektöründe, yolcuların memnuniyeti hizmet kalitesini ölçmede ve hizmet sağlayıcılarla ilişkilerini sürdürme olasılıklarında önemli bir rol oynamaktadır (Ali , 2015).

Archana ve Subha'ya (2012) göre havayolu sektöründe müşteri memnuniyeti, uçuş öncesi hizmetleri, uçak içi hizmetler, bagaj taşıma, uçuş içi dijital hizmetler ve uçuş sonrası hizmetleri içeren çok boyutlu hizmet kalitesinden etkilenmektedir. Bu nedenle havayolu sektöründe müşteri memnuniyeti, sürdürülebilir bir iş ve müşterileri ile uzun süreli ilişki sağlamak için çok kritik bir unsurdur (Farooq, 2018). Bir hizmetin kalitesi, hizmet sağlayıcının tüketicilerin ihtiyaçlarını ve isteklerini tutarlı bir şekilde karşılama becerisine bağlıdır. Müşterilerin ihtiyaçları karşılandığında ve kaliteli hizmet sağlandığında, müşterilerin memnun olma olasılığı daha yüksektir. Yüksek bir müşteri memnuniyeti sağlamak için, bir hizmetin genel olarak müşteri memnuniyetinin bir öncüsü olduğu düşünülürse, bir kuruluş yüksek düzeyde hizmet sunmalıdır. Hizmet kalitesi müşteri memnuniyeti üzerinde önemli bir etkiye sahiptir (Hussain, 2016).

Müşteri memnuniyetini etkileyen faktörler Şekil 1'de görüleceği üzere iki bölümde sınıflandırabilir.



Şekil 1. Müşteri Memnuniyetini Etkileyen Faktörler (Bakırtaş, 2013)

Tatmin olmuş müşteri (Ada, 2011: 18);

i. Aldığı ürün miktarı artar.

ii. İşletmenin sattığı farklı ürünleri de alma eğilimindedir.

iii. İşletmenin ürünlerinden memnun olduğu için, bunu çevresi ile de paylaşarak işletmenin imajına olumlu yönde katkı sağlar.

iv. Başka işletmenin sattığı ürünlere karşı duyarlılığı azdır. İşletmeler günümüz rekabet koşullarında karlılıklarını sürdürebilmeleri için, iyi bir müşteri memnuniyeti programını tasarlaması, uyarlaması ve geliştirmesi gerekir.

Bunun için (Aktepe, 2009: 13);

i. Müşteriler şikâyetlerini zahmetsiz, ücretsiz ve hızlı bir şekilde işletmeye iletebilmelidir.

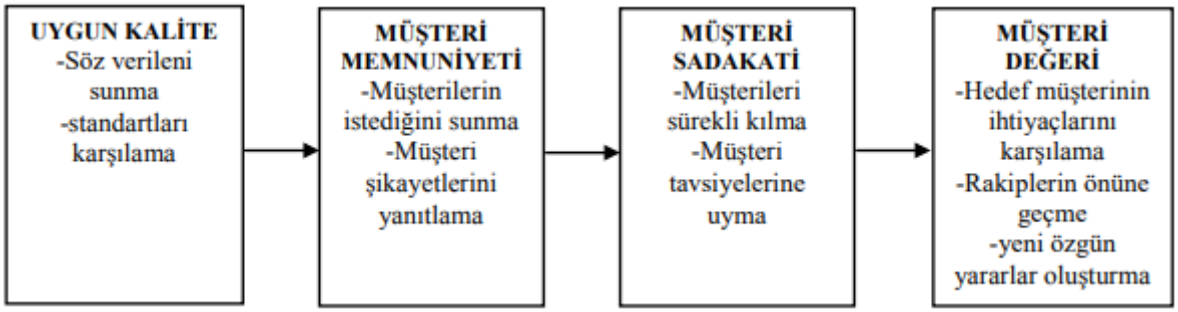
ii. Müşteri geribildirimlerinin alındığı, müşteri memnuniyeti araştırmaları işletme tarafından düzenli yapılmalıdır.

iii. İşletme üst yönetimi, müşteri şikâyetlerini takip etmelidir.

iv. Müşteri şikâyetleri hızlı bir şekilde cevaplanmalıdır.

v. Satış sonrası hizmetler etkinleştirilerek şikâyetler ortadan kaldırılmalıdır.

Sonuç olarak müşterilerin beklentilerinin iyi analiz edilerek, bu beklentilerin ya da fazlasının karşılanması müşteri tatmininin oluşmasını sağlar. Müşteri memnuniyeti, müşterinin satıcıya ve ürününe ilişkin düşüncesinin özetidir. Müşterinin bu memnuniyeti devam ettiği sürece de işletmeye olan güven duygusu artar ve işletmeye olan bağlılığında, sadakatinde bu duygu önemli bir rol oynar. Müşteri memnuniyeti, müşteri bağlılığı ve müşteri için değer oluşturma aşamaları Şekil 2'de gösterilmiştir.



Şekil 2. Müşteri memnuniyeti, Müşteri Sadakati ve Müşteri Değeri Oluşturma (Odabaşı, 2010)

Bu çalışmada havacılık sektöründe hizmet kalitesinin müşterilere yapılan anketler sonucu belirli analizler yaparak, müşteri memnuniyeti üzerindeki ilişkisi araştırılmıştır.Bu amaçla, veri madenciliğinde kullanılan ve birçok analizi sağlayabilen geniş çaplı Weka programı ile müşteriler tarafından yapılmış olan anketlerin analizi ile bazı sonuçlar elde edilmiştir.

1.2.LİTERATÜR ARAŞTIRMASI

Veri Madenciliği, geniş miktarda bilginin biriktirilmesi ve yararlı bilginin ayrıştırılarak taranması sürecini sağlamaktadır. Bu süreçte veri madenciliği litaratürde farklı uygulamalarla yer almıştır. Özellikle son 20 yıl verisindeki çalışmalar veri madenciliğinde uygulamarın ne yöne doğru ilerlediğini göstermektedir.

Köktürk ve arkadaşları 2009, temel olarak veri madenciliği algoritması üzerine araştırma yaparak veri madenciliği yöntemlerine genel bakış sağlayıp veri madenciliği hakkında bilgilendirme çalışması yapmışlardır(Köktürk, Ankaralı, Sümbüloğlu, 2009). Akpınar da yapmış olduğu çalışma ile veri madenciliği ve bilgi yönetimi arasındaki ilişkiyi inceleyerek bilgi yönetimini başarılı şekilde gerçekleştirip veri madenciliği uygulaması yapmıştır. (Akpınar, 2000)

Bilgin 2009, veri akışı diyagramları oluşturarak veri madenciliğini kullanıp yeni yazılım geliştirme ortamları oluşturmuştur. (Bilgin,2009)

Doğan 2004, yaptığı çalışmada veri madenciliği algoritmaları içerisinde sınıflandırmaları kullanarak başarılı veri analizi yapmıştır. (Doğan, 2004)

Dener ve Arkadaşları, 2009 açık kaynak kodlu veri madenciliği programlarından olan WEKA programını kullanarak bir uygulama geliştirmişlerdir (Dener, Dörterler, Orman, 2009).

Nair, 2019, veri madenciliğini kullanarak birliktelik kuralları algoritmaların otomotiv sektörü verileri üzerinde spmf ve weka programları ile performans analizini yapmıştır (Nair, 2019).

Arabacı, 2007 veri madenciliğinde Appriori, Tahminci Appriori ve Tertius algoritmalarının Weka ve Yale programlarına uyarlayarak bu iki programla karşılaştırarak sonuçlara ulaşmıştır. (Arabacı, 2007)

Taşkın'ın yapmış olduğu veri madenciliğinde kümeleme yaklaşımları ve kohonen ağları ile bir uyguama geliştirerek perakendecilik sektöründe bir iyileştirme çalışması gerçekleştirmiştir (Taşkın, 2010).

Bektaş, Veri madenciliği tekniklerinin kullanılarak koroner arter hastalığının varlığının belirlenmesi ve risk faktörlerinin kullanılmasıyla ve bir risk skor sisteminin oluşturulması çalışması gerçekleştirmiştir. (Bektaş, 2017)

Aktürk veri madenciliği çalışması yaparak en iyi yönetici modelini kurmuştur. (Aktürk, 2013)

Güntürkün'ün çalışması ile veri madenciliğinin kalite iyileştirmedeki uygulamalarını nasıl gerçekleştirdiğini ne gibi sonuçlar doğurduğunu anlatan geniş bir çalışma özeti bulunmaktadır. (Güntürkün,2007)

BÖLÜM 2. PROBLEMİN TANIMI

Havacılık sektöründe yaşanan yoğun rekabet ortamında, havayolu firmalarının pazarda varlıklarını sürdürmeleri, uzun dönemde yolcularını birer sadık müşteriye dönüştürmelerine bağlı hale gelmiştir. Müşteri sadakatinin yaratılmasında ise, kurum imajı ve algılanan hizmet kalitesi iki önemli birleşendir. Günümüzde ise, havayolu firmalarının fiyat odaklı yaklaşımla pazar payının arttırılmaya yönelik çalışmaları, orta ve uzun dönemde karlılığı negatif etki yaparak, kurum imajını olumsuz yönde etkilemektedir. Bu anlamda, fiyata göre üstün kaliteyle farklılaşmanın sağlanması, müşteri sadakatini yaratmada daha etkili bir unsurdur.

Havayolu taşımacılığında faaliyet gösteren işletmelerin pazarda yer alabilmesi veya pazar payını arttırabilmesi için, müşterilerin hizmet satın alma tercihlerine neden olan faktörlerin ve bu faktörlerin müşteri memnuniyetini belirleme derecelerinin öğrenilmesi çok büyük öneme sahiptir.

Hizmetler soyut nitelik taşıdıkları için, müşterilerin satın almadan önce hizmet kalitesini değerlendirmeleri zor olabilmektedir. Bu anlamda kalite de, tüketici tarafından algılanan performans düzeyi ya da hizmetin tüketiciyi tatmin etme düzeyine bağlı değişmektedir. Bu nedenle, hizmet kalitesi yerine "algılanan hizmet kalitesi" ifadesi dikkate alınarak anketler aracılığı ile öğrenmek hedeflenmiştir. Bu çalışmada, veri setleri ve veri bilimi çalışmalarına yönelik içerikler bulunduran Kaggle'dan, American Airlines havayolu şirketinin müşterilerine yaptığı bir müşteri memnuniyeti anket verileri incelenecektir. Yaklaşık **100.000** yolcunun anket yanıtlarının bulunduğu veride, müşteriler yaptıkları yolculukların sonucunda yolculuk mesafesini, inişte ve kalkışta belirtilen süreden sapma dakikalarını, bilet alırken ve bagaj verirken karşılaştığı hizmetin kalite seviyesine yönelik puanlama değerleri gibi veriler belirtip son olarak, genel olarak yaptıkları yolculuktan memnun kalıp kalmadıklarını belirtmişlerdir. Çalışmanın amacı bu 100.000 anket verisinin, ilk aşamada veriler arası korelasyonu anlamak üzere **eğitim** için kullanılıp yolcuların en sonda belirttiği memnun kalıp kalmadıklarına yönelik yargılarına etkisinin ne düzeyde olduğunun tespit edimesi, buna yönelik kullanılan sınıflandırma algoritmalarının tanıtılması ve bu algoritmaların başarı oranlarının incelenmesidir. İkinci aşamada ise **25.000 farklı** anket verisinin **test** için kullanılıp, memnun kalıp kalmama yanıtları bilinmeden, yalnızca anket verileri kullanılarak yolcuların memnun kalıp kalmadıklarını, bu veriler arasındaki bağıntılar kullanılarak sınıflandırmak, etiketlemek amaçlanmıştır. Çalışmada bahsedildiği üzere veri madenciliği yöntemlerinden sınıflandırma algoritmaları kullanılacaktır.

BÖLÜM 3. PROBLEMİN ÇÖZÜMÜ İÇİN TAKİP EDİLECEK AŞAMALAR

Çözüm Aşamaları	Yöntem	Kullanılan Program
1	Verilerin indirilmesi, temizlenmesi ve düzenlenmesi	
2	Eğitim verisine sınıflandırma uygulaması	WEKA
2.1	KNN Algoritma	
2.2	Decision Tree Algoritma	
3	Sınıflandırılan eğitim seti verilerinin incelenmesi ve uygulanan yöntemlerin karşılaştırılması	
4	Özelliklere göre eğitim verilerinin incelenmesi son düzenlemelerin yapılması	
5	Test verilerine sınıflandırma uygulanması	WEKA
6	Sınıflandırılan test verilerinin incelenmesi ve yorumlanması	

3.1. Verilerin İndirilmesi, İncelenmesi, Temizlenmesi ve Düzenlenmesi

Problemin ilk aşamasında, 125.000 anket verisi indirilecek, daha sonra bu verilerden müşteri memnuniyeti üzerinde etkisi hiç olmadığı veya etkisinin gözardı edilecek kadar düşük olduğu veriler silinecek, ayrıca veriler Türkçe diline çevrilecektir. Son olarak veri formatı, sınıflandırma işlemleri için kullanılacak olan WEKA programı için uygun hale getirilip kaydedilecektir.

3.2. Eğitim Verisine Sınıflandırma Uygulanması

İkinci aşamada uygun hale getirilen yaklaşık 100.000 adet etiketli(sınıflı, memnun kalıp kalmadığı bilinen) anket verisi üzerinde WEKA programında halihazırda var olan sınıflandırma algoritmalarından K Nearest Neighborhood ve Decision Tree yöntemleri uygulanacaktır.

K Nearest Neighborhood ve Decision Tree yöntemlerinin çalışma prensipleri bu kısımda aktarılacaktır.

3.2.1 K Nearest Neighborhood Yöntemi

Bu sınıflandırma yöntemi veriler arası koordinat uzaklıklarını kullanarak sınıflandırmaya dayanır. Algoritmada K değeri veriye ait olan en yakın, yani en benzer verilerin sayısıdır. Kullanıcı bu K değerini kendi belirleyebilir ve sınıflandırmayı bu K değeri kadar en yakın veri(komşu)değerlerine göre gerçekleştirebilir. K değeri sınıflandırabilmenin gerçekleştirilmesi için yalnızca 1, 3, 5 gibi tek sayılar olarak seçilmektedir.

Yeni bir değer geldiğinde K değerine mesafeler hesaplanır ve yeni değer bir kümeye ilave edilir. Mesafe hesaplama işleminde ise genellikle öklid uzaklığı kullanılırken, bunun yanında manhattan uzaklığı, minkowski uzaklığı gibi mesafe hesaplama yöntemleri de kullanılabilir. Aşağıda Öklid uzaklığı yönteminin adımları gösterilmektedir: (Uzun, 2020)

Bu algoritma beş adımdan oluşur:

Öncelikle K değeri belirlenir.

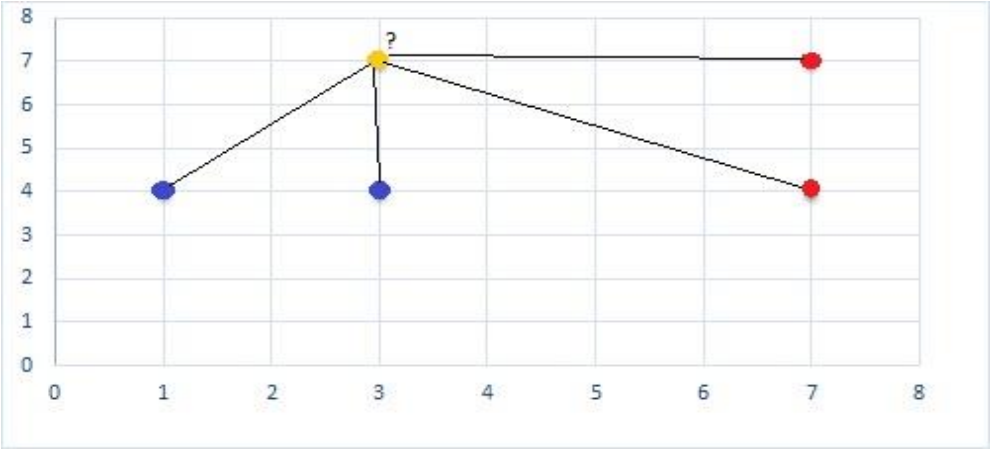
Diğer nesnelerden hedef nesneye olan öklit uzaklıkları hesaplanır.

Uzaklıklar sıralanır ve en minimum uzaklığa bağlı olarak en yakın komşular bulunur.

En yakın komşu kategorileri toplanır.

En uygun komşu kategorisi seçilir. (Uzun, 2020)

Örnek olarak aşağıda var olan 4 adet veri (mavi ve kırmızı renkte) bulunan koordinat düzlemine dökülmüş ifadede 5. verinin (sarı) hangi sınıfa ait olduğu tespit edilmeye çalışılsın. Nesne 1 ve nesne 2 mavi nesneleri(sınıfı) temsil ederken nesne 3 ve nesne 4 kırmızı nesneleri(sınıfı) temsil etmektedir.



Şekil 3. Veri kümesine yeni bir veri yüklenmesi (Uzun, 2020)

Yeni eklenen nesne ile diğer nesneler arasındaki öklid uzaklığı hesaplanır ve hesaplama sonuçları aşağıdaki Tablo 1 de gösterildiği gibidir.

$$Fark_e = \sqrt{(X_i - X_{yeni})^2 + (Y_i - Y_{yeni})^2}$$

Tablo 1. Yeni nesnenin var olan en yakın nesnelere öklid uzaklığı (Uzun, 2020)

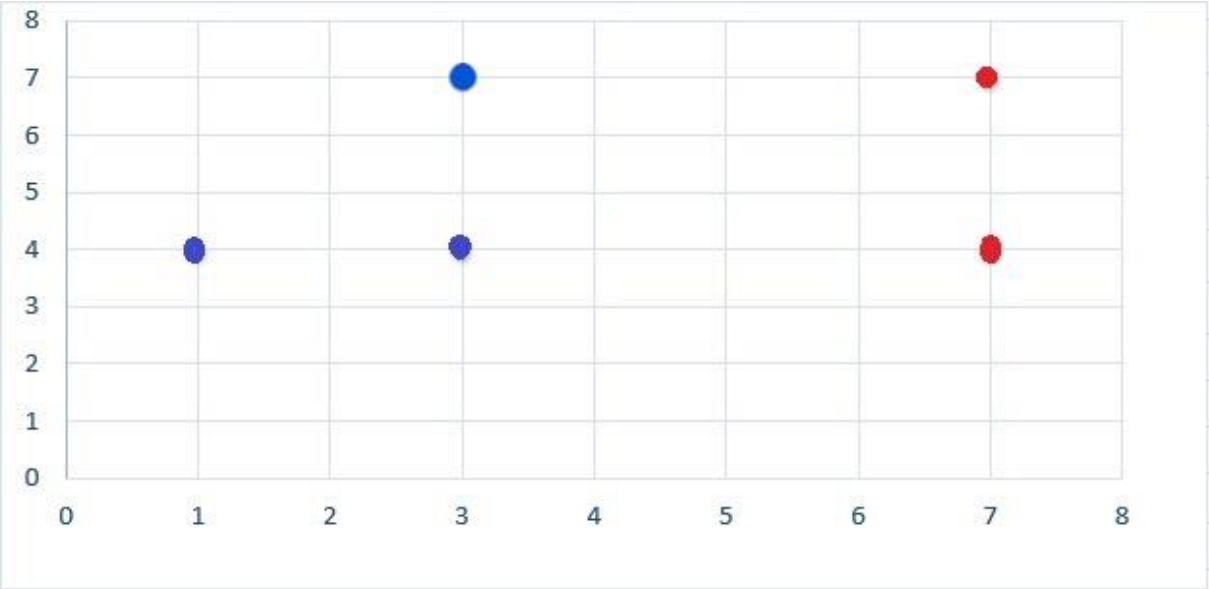
	Yeni Nesne ile Uzaklık
Nesne 1	4
Nesne 2	5
Nesne 3	3
Nesne 4	3.606

Daha sonra K değerine göre en yakın nesneler belirlendikten sonra sınıflandırma işlemi gerçekleştirilir. K değerinin 3 olarak alındığı varsayılırsa aşağıdaki gibi en yakın 3 nesneye göz atılıp, nesnelerin çoğunluğu hangi sınıfta ise, yeni nesne de o sınıfa atanır. (Uzun, 2020)

Tablo 2. En yakın ilk 3 komşu ve sınıfları (Uzun, 2020)

Sıralama	Sınıfları
Nesne 4	Sınıf 2
Nesne 3	Sınıf 2
Nesne 1	Sınıf 1
Nesne 2	Sınıf 1

Tablo 2. ile görüldüğü üzere en yakın ilk 3 komşudan ikisi(nesne 4 ve nesne 3) 2. sınıfa yani mavi renkli sınıfa aittir. Bu yüzden en yakın 3(k) komşu için yeni verinin sınıfı da 2. sınıf olarak atanır. K değeri 1 olarak belirlenseydi, en yakın komşunun(nesne 4) sınıfı, sınıf 2 olduğu için yeni değer bu durumda da 2. sınıfa atanırdı. (Uzun, 2020)



Şekil 4. Yeni eklenen verinin sınıflandırılması (Uzun, 2020)

Verileri koordinat düzlemine dökerken Şekil 4 ile gösterildiği gibi her özelliğin(attribute) düzleme döküldüğü unutulmamalıdır. Örneğin sadece iki özelliğin bulunduğu bir sınıflandırma işleminde özellikleri x ve y olarak düşünüp bunları kartezyen koordinat sisteminde düşünmek kolaydır. Ancak koordinat düzleminde özellik sayısı kadar parametre kullanılmaktadır. Bu çalışmada 21 özelliğin ve 1 sınıfın bulunduğu bir veri üzerinde çalışmaktadır. Bu durumda KNN yöntemi uygulanırken 21 farklı parametrenin bulunduğu çok boyutlu bir işlem gerçekleştirilmektedir. Eğitim verisinin büyük olması ve K değerinin uygun seçilmesi KNN açısından önemlidir. Yapılacak olan çalışmada eğitim seti sınıflandırırken farklı K değerleri için başarı oranı kontrol edilecektir. En yüksek başarı oranına sahip K değeri tespit edilip, veriler eğitim için kullanılırken bu K değeri kullanılacaktır.

3.2.2 Decision Tree Yöntemi

Sınıflandırma problemleri için yaygın kullanılan bir yöntemdir. Öğrenmiş sınıflandırma modeli ağaç şeklinde gösterilir ve karar ağacı (decision tree) olarak adlandırılır. Karar ağaçları akış şemalarına benzeyen yapılardır. Her bir nitelik bir düğüm tarafından temsil edilir. Dallar ve yapraklar ağaç yapısının elemanlarıdır. En son yapı yaprak en üst yapı kök ve bunların arasında kalan yapılar dal olarak isimlendirilir. C4.5 ağacı için WEKA programında J48 olarak açık kaynak kodlu bir versiyonunun yazılmış olduğunu söylemekte yarar vardır. Yukarıdan aşağıya bir yaklaşım kullanan algoritmalardan J48' i inceleyeceğiz. C4.5 algoritmasının bir üst versiyonudur. Karar ağacı algoritmaları, durumlar ve örnek kümeleri ile başlar. Olayların sınıflandırılmasında ağaç yapısı oluşturulur. Ağaçlar, düğüm ve yapraklardan oluşur.

Düğümden sonra dallanmalar başlar. Her bir durumun olasılığı dallanmada belirtilir. Karar ağacının her bir düğümü test verisi içerir. Test sonuçlarına göre sonraki dallanmanın hangisi olacağına karar verilir. Yapraklar da sınıf bilgisi içerir (Cihan ve Kalıpsız, 2015). Bu algoritma, sınıf bilgilerini kullanarak sınıflandırma yapar .C4.5 ağacında normalleştirme (normalization) kullanılır.

C4.5 ağacında entropi değerleri birer oran olarak tutulur. Ayrıca ağaç üzerinde erişim sıklıklarına göre alt ağaçların(subtree) farklı seviyelere taşınması da mümkündür.C4.5 ağacında budama(pruning) işlemi yapılmaktadır.

Decision Tree algoritmasının nasıl çalıştığını açıklamak gerekirse:

- Her adımda bütün özellikler kontrol edilir.
- Her özelliğin normalize edilmiş bilgi kazanımı (information gain) hesaplanır.
- En iyi bilgi kazanımını veren özellik karar ağacında karar olarak taşınır.
- Ardından bu yeni karar düğümünün altında bir alt liste oluşturularak alt karar ağacı inşa edilir.

Konuyu bir örnek üzerinden anlamaya çalışalım. Örnek olarak aşağıdaki Tablo 4 de bulunan veri kümesini (data set) ele alalım:

Tablo 3. Örnek Veri Kümesi

Özellik 1	Özellik 2	Özellik 3	Sınıf
Ali	70	Geçti	Sınıf1
Ali	90	Geçti	Sınıf2
Ali	85	Kaldı	Sınıf3
Ali	95	Kaldı	Sınıf4
Ali	70	Kaldı	Sınıf1
Evren	90	Geçti	Sınıf1
Evren	78	Kaldı	Sınıf1
Evren	65	Geçti	Sınıf1
Evren	75	Kaldı	Sınıf1
Şadi	80	Geçti	Sınıf2
Şadi	70	Geçti	Sınıf2
Şadi	80	Kaldı	Sınıf1
Şadi	80	Kaldı	Sınıf1
Şadi	96	Kaldı	Sınıf1

Öncelikle aşağıdaki formüllerle Bilgi Kazanımı(Gain Information) hesaplanır.

m: hedef değişkenin alabileceği değerlerin adedi

A: tahmin edici değişken

D: hedef değişken

v: tahmin edici değişkenin alabileceği değerlerin adedi

Entropi: Info(D) = $-\sum_{i=1}^{m}(pi \log_2 p\, i)$

A'nın Bilgisi: InfoA(D) = $-\sum_{j=1}^{v} \frac{|Dj|}{|D|}$ x Info(Dj)

Bilgi Kazanımı: Gain(A) = Info(D) - InfoA(Dj)

Bilgi kazanımı hesaplanırken, o anda veri kümesinde bulunan bütün veriler ve hesaplanması istenen belirli bir verinin üzerinden gidilir. Bu hesaplaması yapılacak olan belirli veriye örnekleme (misal, sampling) ismi verilir ve bütün veri kümesi üzerinden bu örneklemeye ait hesaplama yapılır.

Yukarıdaki formüle göre, 14 toplam satırdan 5 tanesi sınıf 2 ve 9 tanesinin sınıf 1 olduğunu dikkate alarak aşağıdaki eşitliği yazıyoruz. Önce bilgi değerlerini hesaplayacak sonra da kazanımı bulacağız:

$$\text{Bilgi(p)} = -9/14 * \log_2(9/14) - 5/14 * \log_2(5/14)$$
$$= 0.940 \text{ bit}$$

İlk bilgi değeri bütün parçanın hesaplandığı yani 14 satırın tamamının dikkate alındığı ve 9/14 ve 5/14 olarak iki ihtimalin hesaba katıldığı durumdur. Bu durum aynı zamanda entropi olarak da düşünülebilir.

İkinci bilgi hesabımızda özellik 1 kullanılacak. Buna göre veri kümemizin ilk 5 satırında Ali, sonraki 4 satırında Evren ve son 5 satırında Şadi özellikleri var. Buna göre Tablo 4 ile gösterildiği gibi tabloyu 3 parçaya bölersek :

Tablo 4. Örnek Veri Kümesinin 3 Gruba Ayrılması

Özellik 1	Özellik 2	Özellik 3	Sınıf
Ali	70	Geçti	Sınıf1
Ali	90	Geçti	Sınıf2
Ali	85	Kaldı	Sınıf3
Ali	95	Kaldı	Sınıf4
Ali	70	Kaldı	Sınıf1
Evren	90	Geçti	Sınıf1
Evren	78	Kaldı	Sınıf1
Evren	65	Geçti	Sınıf1
Evren	75	Kaldı	Sınıf1
Şadi	80	Geçti	Sınıf2
Şadi	70	Geçti	Sınıf2
Şadi	80	Kaldı	Sınıf1
Şadi	80	Kaldı	Sınıf1
Şadi	96	Kaldı	Sınıf1

Yukarıdaki yeni tabloya göre her özellik parçasının ayrı ayrı hesaplanarak denklemde yerine yazılması gerekir:

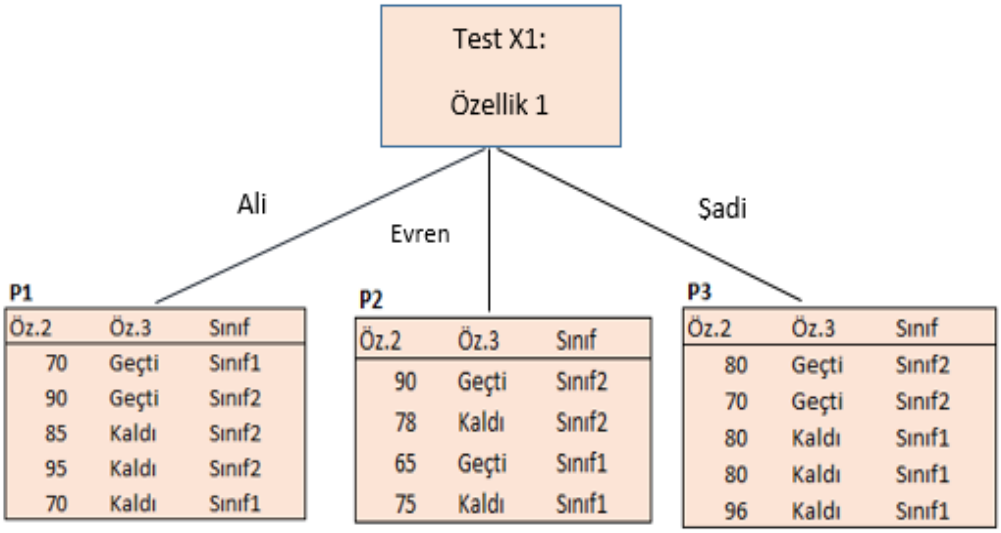
$$\text{Bilgi}_{x1}(T)=5/14(-2/5*\log_2(2/5)-(3/5)) + 4/14(-4/4*\log_2 4/4-0/4*\log_2(0/4))$$
$$+5/14(-3/5*\log_2(3/5)-2/5*\log_2(2/5))$$
$$=0.694 \text{ bit}$$

Yukarıdaki formülde mavi renkle belirtilen durum 1. özellik için (x1) 5, 4 ve 5 parçadan oluşan ve her parça için ayrı ayrı Sınıf1 ve Sınıf2 değerlerinin sayıldığı durumdur. Yani ilk 5 satırlık parçanın 2 satırı Sınıf1 ve 3 satırı Sınıf2 olduğu için 2/5 ve 3/5 şeklinde iki değer alınmıştır. Diğerleri de benzer şekilde hesaplanmıştır.

Son adımda bu iki değer arasındaki farkı hesaplayabiliriz:

$$\text{Kazanım}(x1)=0.940-0.694=0.246 \text{ bit}$$

Yukarıda bulunan bilgi kazanımı, bütün veri kümesindeki Özellik1 için bütün sınıflar arasındaki kazanımı göstermektedir. Bu değerleri kullanarak aslında veri kümemizi 3 parçaya bölmüş ve her birisi için bilgi kazanımının olası değerini hesaplamış oluyoruz:



Şekil 5. Karar Ağacının Alabileceği Olasılıklardan Biri

Yukarıdaki Şekil 5 ile gösterim karar ağacının alabileceği olasılıklardan birisidir ve ağacın özellik 1'de bulunan isimlere göre karar noktası oluşturulması halinde alacağı vaziyeti gösterir. Şimdi aynı hesabı, Özellik 3 için yapalım. Burada geçti / kaldı ihtimalleri bulunuyor dolayısıyla ağaç 2 dala ayrılabilir. Ancak biz önce hesabımızı yapalım.

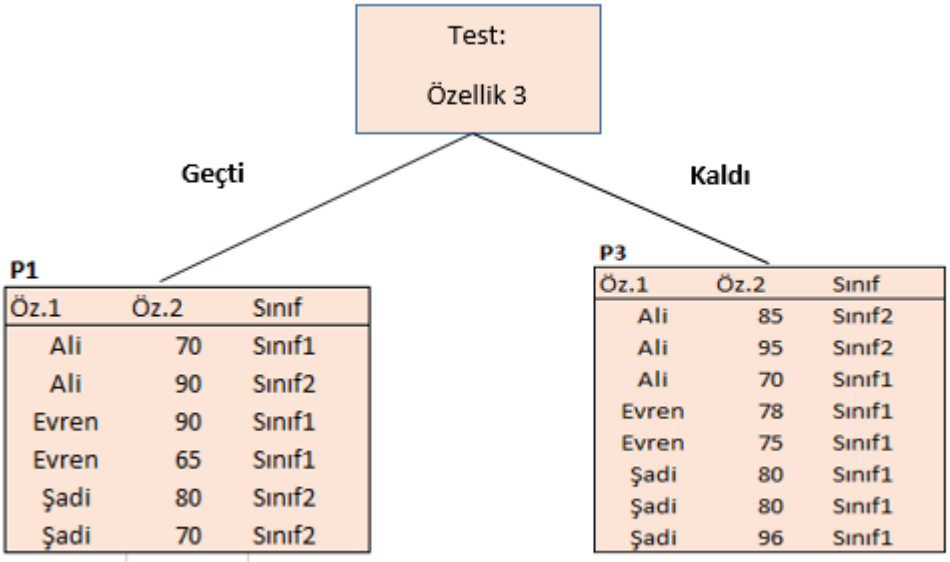
$$\text{Bilgi}(P)=-9/14*\log_2(9/14)-5/14*\log_2(5/14)=0.940 \text{ bit}$$

$$\text{Bilgi}_{\bar{O}3}(P)=6/14(-3/6*\log_2(3/6)-3/6*\log_2(3/6))$$

$$+8/148-6/8*\log_2(6/8)-2/8*\log_2(2/8))=0.892 \text{ bit}$$

$$\text{Kazanım}(\ddot{O}_3)= 0.940-0.892=0.048 \text{ bit}$$

Özellik 3 için 3/6 ve 6/8 olmak üzere iki parça bulunuyor ve her ikisi için de hesap yapılıp toplam bilgiden çıkarıldığında kazanım değeri olarak 0.048 bit bulunuyor. Bu değeri yorumlamadan önce ağacın şu anda hesaplanan halini göstermeye çalışalım:



Şekil 6. Karar Ağacının Alabileceği Olasılıklardan Bir Diğer Örneği

Yukarıdaki Şekil 6 da ağacın ilk karar noktasını Geçti / Kaldı ihtimalleri üzerine kuracak olursak bilgi kazanımı olarak 0.048 beklenmektedir.

Bu durumda C4.5 ağacı en yüksek kazanıma sahip olan değeri alacaktır. Bu değer Özellik 1 için isimler olduğundan karar ağacının bu adımda Özellik 1'e göre karar noktası eklemesi yerinde olacaktır.

Ardından diğer adımlar için benzer şekilde hesaplamalar yapılarak ağacın karar noktaları oluşturulmaya devam edecektir.

C4.5 Ağacının önemli bir diğer özelliği ise budama işlemidir. Esas olarak ağaçlarda iki tip budama yapılabilir. Birisi ön budama (preprunning) diğeri ise son budama (post prunning). C4.5 ağacı son budama (postprunning) yöntemini tercih etmektedir.

Hemen burada karar ağaçlarında (decision trees) ön budama ve son budamanın nasıl yapıldığından bahsedelim. Ön budama, genelde ağaç oluşturulurken bazı dalların oluşturulmaması yönündedir. Örneğin bazı dallar anlamsız olacağından veya hiç eleman içermeyeceğinden oluşturulmaz. Son budama ise ağacın bütün dallarını oluşturur ve sonra bazı şartlara göre budama yapar. Burada da eleman içermeyen dallar budanabileceği gibi, bazı durumlarda istatistiksel yaklaşımlar da kullanılabilir. Örneğin yukarıdaki veri kümemizi Özellik 2'yi kullanarak 10'luk dilimlere bölmek isteyelim. 100-90 arasında 90-80 ve 80-70 arasında verilerimiz olacak ancak 70'in altında verilerimiz olmayacak. Bunu verilere bakmadan anlayamayız. Şayet illaki ağaç oluşturulacaksa ve kural 0 ile 100 arasındaki notların 10'arlık

dilimler halinde bölünmesi ise bu dalların da ağaçta yer alması gerekir ancak hiç veri içermeyeceği için bu dallanmalar anlamsız olacak ve budanacaktır.

3.3. Sınıflandırılan Eğitim Seti Sonuçlarının İncelenmesi, Yorumlanması ve Yöntemlerin Karşılaştırılması

Bu aşamada eğitim verisi üzerinde gerçekleştirilen sınıflandırma algoritmalarının başarı oranları incelenecek, hangi etkenlerin sınıflandırma üzerinde etkisinin daha fazla, hangi etkenlerin daha az olduğu tespit edilmeye çalışılacaktır. Uygulanan algoritmalar kıyaslanacaktır. Bu aşama sayesinde etiketi bilinmeyen veriler üzerinde uygulanacak sınıflandırma işlemleri sonucunda verilerin yaklaşık olarak hangi oranda başarı ile etiketleneceği belli olacaktır. Test verisinin hangi algoritma kullanılarak sınıflandırılacağına karar verilecektir.

3.4. Özelliklere Göre Eğitim Veri Setinin İncelenmesi ve Son Düzenlemelerin Yapılması

Bu aşamada 3. aşamadan elde edilen bazı sonuçlara göre eğitim verisi üzerinde başarı oranını arttırabilecek son dokunuşlar yapılacaktır. Özelliklere göre veriler incelenecektir.

3.5. Test Verisine Sınıflandırma Uygulanması

Bu aşamada sınıfları bilinmeyen yaklaşık 25.000 anket verisi, karar verilen yöntem ve son dokunuşlardan sonraki eğitim verisi modelinden yararlanılarak sınıflandırılacaktır ve etiketi, sınıfı bulunmayan bu 25.000 veri sınıflandırılmış olacaktır.

3.6.Sınıflandırılan Test Verisi Seti Sonuçlarının İncelenmesi ve Yorumlanması

Bu aşamada test verilerinin, atanan sınıfları incelenecektir ve verilerin, verilen anket yanıtlarına göre gerçekleştirilen sınıflandırma işlemi yorumlanacaktır.

BÖLÜM 4. UYGULAMA

4.1. Verilerin İndirilmesi, Temizlenmesi ve Düzenlenmesi

İlk olarak havayolu şirketinin anketinde yer alan veriler indirilip düzenlenmiştir ve aşağıda ankette bulunan müşteri memnuniyetini tespit etmeye yönelik olarak oluşturulan özellikler(sorular) belirtilmiştir.

Cinsiyet: Yolcuların Cinsiyeti (Kadın, Erkek)

Müşteri Türü: Müşteri türü (Firmaya sadık olduğunu düşünen müşteri, firmaya sadık olduğunu düşünmeyen müşteri)

Yaş: Yolcuların yaşı

Seyahat Türü: Yolcu uçuşunun amacı (Kişisel Seyahat, İş Seyahati)

Sınıf: Yolcuların yaptıkları yolculuktaki seyahat sınıfı (Business, Economy, Economy Plus)

Uçuş mesafesi: Yapılan yolculuğun uçuş mesafesi

Aşağıdaki ifadeler için temizlik ifadesine kadar ankette uygun rakamlar seçilmektedir. (0-5 arası memnuniyet seviyesini belirtir. 5: en yüksek seviye, 0: en düşük seviye)

Uçak içi wifi hizmeti: Uçak içi wifi hizmetinin memnuniyet düzeyi

Kalkış / Varış zamanı uygunluğu: Kalkış / varış zamanının vaad edilen sürede olmasına yönelik memnuniyet seviyesi

Online rezervasyon kolaylığı: Online rezervasyon hizmetinin memnuniyet düzeyi

(Gate) Kapı konumu: Uçağın olduğu alana geçiş kapısının konumuna ait memnuniyet düzeyi

Yiyecek ve içecek: Yolculuk sırasındaki yiyecek ve içecek hizmetine ait memnuniyet düzeyi

Online biniş: Online biniş işlemlerine ait hizmetin memnuniyet düzeyi

Koltuk konforu: Uçak koltuk konforuna ait memnuniyet düzeyi

Uçak içi eğlence: Uçak içi eğlence hizmetinin memnuniyet düzeyi

On-board servis: On-board(Yolculuk sırasında) servis memnuniyet seviyesi

Bacak/Ayak yeri düzeyi: Bacakları ve ayaklar konulan bölgenin konforuna/genişliğine ait memnuniyet düzeyi

Bagaj yönetimi: Bagaj yönetimine ait memnuniyet düzeyi

Check-in hizmeti: Check-in işlem hizmetinin memnuniyet düzeyi

Temizlik: Temizlik durumuna ait memnuniyet düzeyi

Dakika Olarak Kalkış Gecikmesi: Dakika olarak kalkışta vaad edilen zaman ile gerçek zaman arasındaki gecikmesi

Dakika Olarak Varış Gecikmesi: Dakika olarak varış için vaad edilen zaman ile gerçek zaman arasındaki gecikmesi

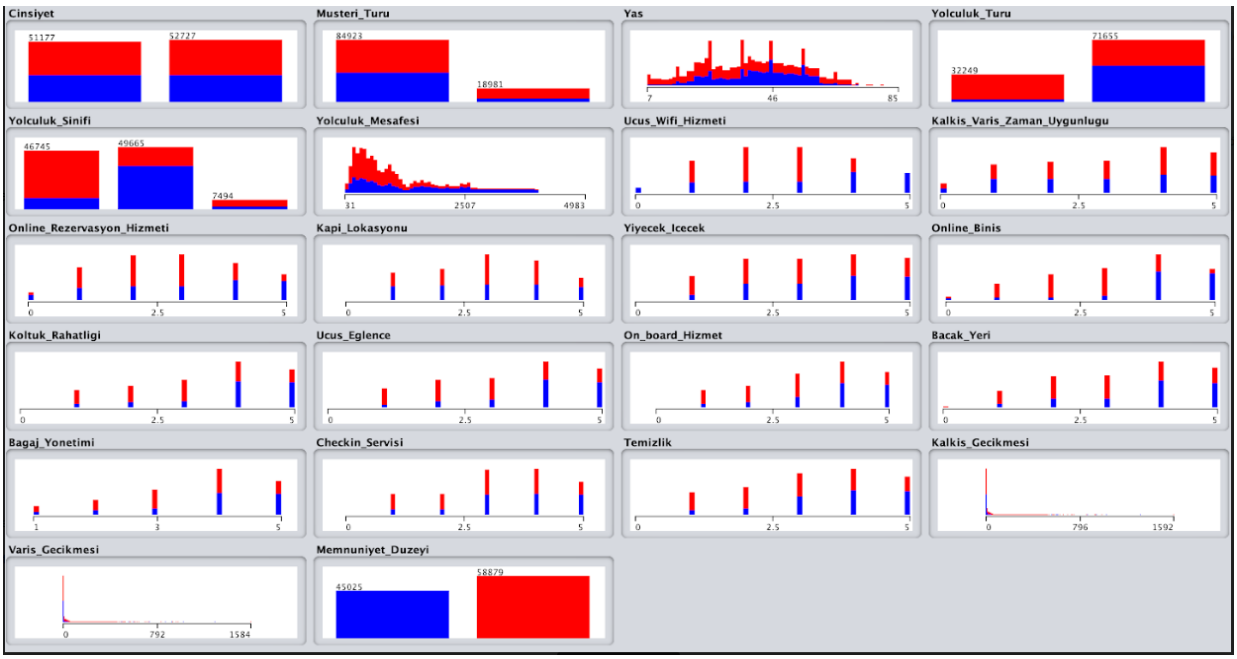
Memnuniyet: Uçuştan memnun kalıp kalmama durumu (Memnun - Nötr veya Memnuniyetsiz)
Verilerin WEKA programında uygun çalışması için veri dosyasında isimleri aşağıdaki şekilde oluşturulmuştur:
Cinsiyet: **Cinsiyet**
Müşteri Türü: **Musteri_Turu**
Yaş: **Yas**
Seyahat Türü: **Seyahat_Turu**
Sınıf: Yolculuk_Sinifi
Uçuş Mesafesi: Yolculuk_Mesafesi
Uçak içi Wifi Hizmeti: **Ucus_Wifi_Hizmeti**
Kalkış / Varış Zamanı Uygunluğu: **Kalkis_Varis_Zaman_Uygunlugu**
Online Rezervasyon Kolaylığı: **Online_Rezervasyon_Hizmeti**
(Gate) Kapı Konumu: **Kapi_Lokasyonu**
Yiyecek ve İçecek: **Yiyecek_Icecek**
Online Biniş: **Online_Binis**
Koltuk Konforu: **Koltuk_Konforu**
Uçak içi Eğlence: **Ucus_Eglence**
On-board Servis: **On_board_Hizmet**
Bacak/Ayak Yeri: **Bacak_Yeri**
Bagaj Yönetimi: **Bagaj_Yonetimi**
Check-in Hizmeti: **Checkin_Servisi**
Temizlik: **Temizlik**
Dakika Olarak Kalkış Gecikmesi: **Kalkis_Gecikmesi**
Dakika Olarak Varış Gecikmesi: **Varis_Gecikmesi**
Memnuniyet: Memnuniyet_Duzeyi

Veriler sınıflandırmaya hazır hale getirildikten sonra, ilk olarak 103.904 anket yanıtının bulunduğu eğitim için kullanılacak veri seti üzerinde sınıflandırma işlemi uygulanacaktır.
Bu sayede var olan etiketli verilerden yola çıkarak anket sorularındaki diğer 21 sorunun 22. sorunun yanıtının oluşumundaki etkisi tespit edilecektir. Kullanılacak olan sınıflandırma algoritmalarının bu veri seti için sınıflandırma işleminde başarı oranı elde edilecektir. Sınıflandırma işlemi için popüler sınıflandırma yöntemlerinden K Nearest Neighborhood ve Decision Tree algoritmaları kullanılacaktır.

4.2. WEKA'da Eğitim Verilerinin Sınıflandırılması

Çalışmada, sınıflandırma işlemi uygulanırken her iki kullanılacak olan algoritma için de literatürde K-Fold Cross-Validation denilen sınıflandırma biçimi işlemi uygulanacaktır. K değeri genellikle 10 olarak seçilmekle beraber bu çalışmada da bu şekilde uygulanacaktır.

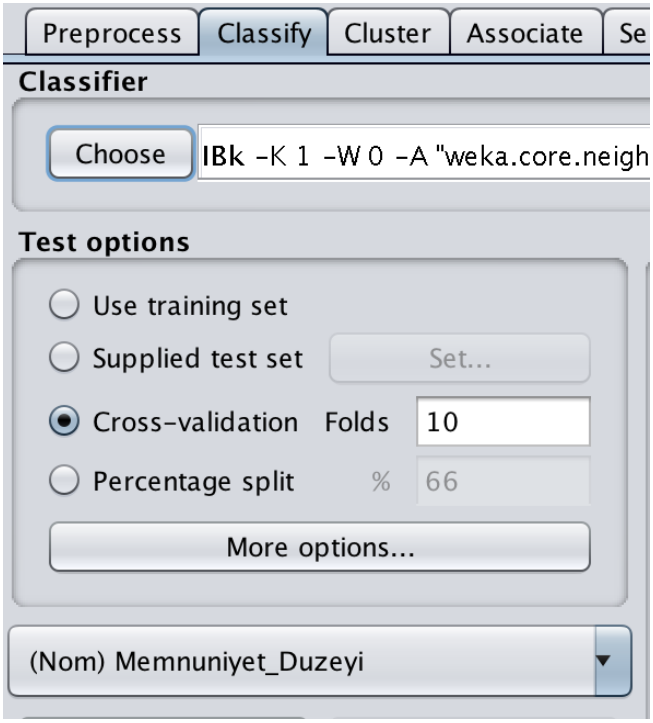
Cross Validation yönteminde eğitim için kullanılacak veri seti 10(k) parçaya bölünmekte ve veri setinin 10'da 1'i test işlemi için kullanılırken kalan 10'da 9'u da eğitim için kullanılmaktadır. İşlemin başarılı olmasını sağlayan özelliği ise bu 10 parçaya bölme işlemini 10 kere tekrarlaması ve her tekrarda 10'da 1'lik test bölümünü değiştirmesi, kaydırmasıdır. Bu sayede verinin tamamı hem test için hem de eğitim için kullanılmış olur. Yöntem WEKA'da seçilip kullanıldıktan sonra otomatik olarak bu gerçekleştirilen 10 farklı test ve eğitim verisi sınıflandırma sonucunun ortalamasını hesaplayıp, başarı oranını bu ortalama değer olarak oluşturmaktadır. Eğitim verisi programa yüklendikten sonra aşağıdaki şekilde her bir özellik için bilinen sınıf değerleri kırmızı ve mavi renkte, özelliğin kendi değeri de dikkate alınarak grafik haline getirilmiştir. Kullanılacak olan algoritmalar bu analizden yararlanır ve özelliklerin sınıflarına bağlı olarak kullanılan yönteme göre hesaplama desenleri(pattern) oluşturur. Kırmızı renkle gösterilen veriler yolculuklarından memnun kalan müşterileri gösterirken, mavi renk ile gösterilen veriler yolculuklarından memnun kalmayan veya nötr düşünen yolculara ait değerlerdir. Veri dosyasının sınıflandırılmış halinin görselleşmesi şekil 7 deki gibidir.



Şekil 7. WEKA Programında Eğitim Verisini Görselleştirme

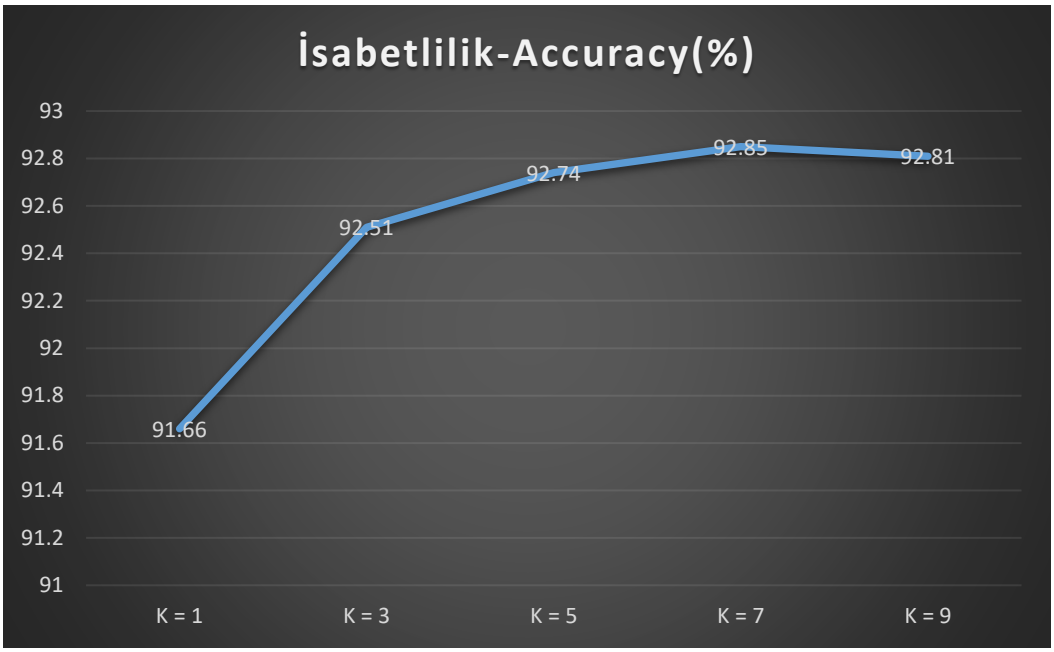
4.2.1 KNN Algoritması ile Sınıflandırma

WEKA programında KNN algoritması, veri yüklendikten sonra Şekil 8 de gösterilen explorer arayüzündeki classify sekmesinde, lazy bölümündeki IBk seçeneği seçilerek kullanılabilmektedir.



Şekil 8. WEKA Programında Cross-validation Yöntemi ile KNN Algoritması Kullanılarak Sınıflandırma İşlemi

Önceki kısımlarda bahsedildiği üzere KNN algoritması uygulanırken en yakın kaç komşuya bakılacağını yani k değerini kullanıcı belirlemektedir. Yapılan uygulamada k değeri için en uygun değerin hangisi olduğunu tespit etmek üzere, yöntem K değeri 1'den 9'a kadar olacak şekilde değiştirilerek uygulanmıştır. Aşağıda çeşitli k değerlerine karşılık gelen accurracy (isabetlilik) oranları yer almaktadır. Buna göre en isabetli oranı kullanmak için k değeri bu orana karşılık gelen değer olarak seçilip, eğitim verileri KNN algoritması ile sınıflandırılırken bu k değerine göre sınıflandırma gerçekleştirilecektir.



Şekil 9. Farklı K Değerlerine Karşılık Gelen Farklı İsabetlilik Oranları

Şekil 9 da da görülebildiği üzere sınıflandırma yöntemlerinin başarı oranları:

K = 1 için %91.66

K = 3 için %92.51

K = 5 için %92.74

K = 7 için %92.85

K = 9 için %92.81'dir. Bu sonuca göre test verisi sınıflandırılırken eğitim verisinin tamamı K = 7 değeri ile kullanılmalıdır.

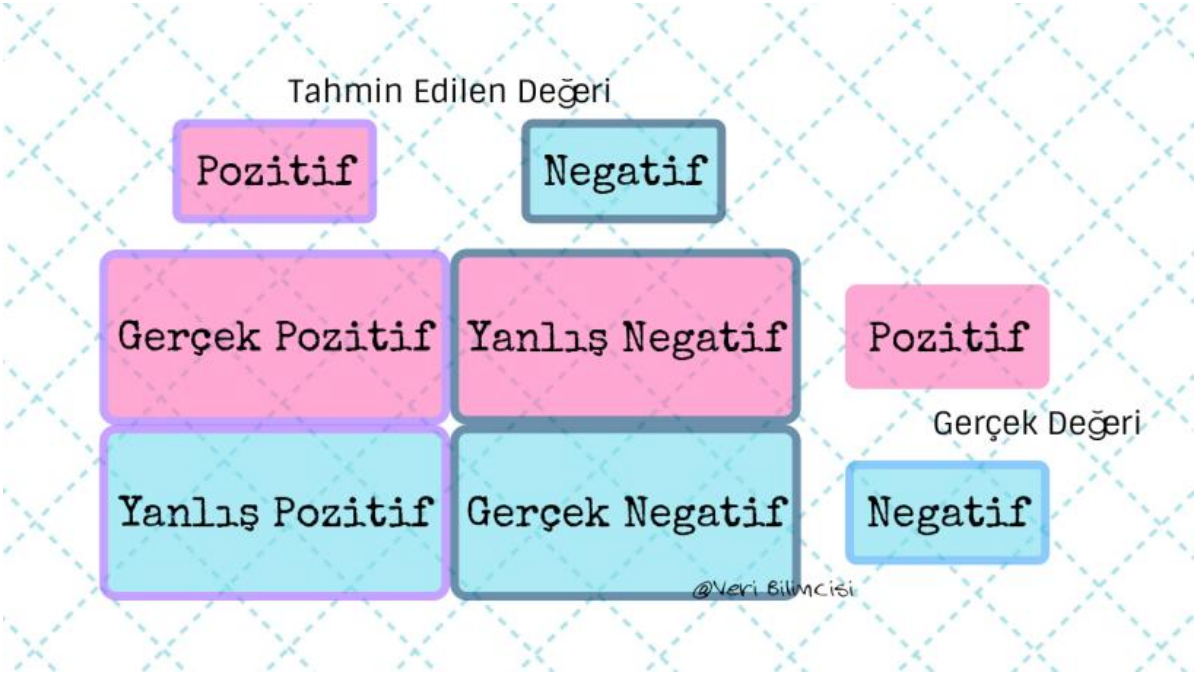
Sınıflandırma Sonuçları

KNN Algoritmasının K = 7 değeri için, eğitim verisinde kullanılmıştır ve sınıflandırma sonuçları aşağıdaki şekil 10 daki gibidir:

```
=== Confusion Matrix ===

     a     b   <-- classified as
 39678  5347 |     a = Memnun
  2077 56802 |     b = Notr_veya_Memnun_Degil
```

Şekil 10. Uygulama Sonucu Confusion (Karışıklık) Matrisi



Şekil 11. Confusion Matrisinin Taslağı (Beşer, 2018)

Bir karışıklık matrisi
olan Şekil 11, gerçek değerlerin bilinmekte olduğu bir dizi test verisi üzerinde, bir sınıflandırma modelinin performansını tanımlamak için sıklıkla kullanılan bir tablodur. (Beşer, 2018)

Şekilde görülebileceği üzere confusion matrisi 4 bölümden oluşmaktadır. Bu bölümler:

Gerçek pozitif (True Positive – TP) – A olarak tahmin edilen ve doğru
Gerçek negatif (True Negative – TN) – B olarak tahmin edilen doğru
Yanlış pozitif (False Positive – FP) – A olarak tahmin edilen ve yanlış
Yanlış negatif (False Negative – FN) – B olarak tahmin edilen ve yanlış olan ifadelerdir. (Beşer, 2018)

Program çıktısından görülebileceği üzere, uygulamada A yerine memnun olan müşteri, B yerine ise nötr veya memnun olmayan müşteriler karşılık gelmektedir. Buna göre değerler:

Gerçek Pozitif (TP) = 39678 (Gerçekte de tahminde de memnun olan müşteri sayısı)
Yanlış Pozitif (FP) = 2077 (Gerçekte memnun tahminde memnun olmayan müşteri sayısı)
Yanlış Negatif (FN) = 5347 (Gerçekte memnun olmayıp tahminde memnun olan müşteri sayısı)
Gerçek Negatif (TN) = 56802 (Gerçekte de tahminde de memnun olmayan müşteri sayısı) olarak bulunur.

Karışıklık matrisinden hesaplanan bazı oranlar vardır. TP, TN, FP ve FN'nin birbirleriyle ilişkisini gösteren bu terminolojiler sınıflandırma işleminin başarı oranı hakkında bilgiler verir. Şekil 12 deki KNN algoritmasının doğruluk ve hata oranları gösterilmiştir.



Şekil 12. Eğitim Verisinin KNN ile Doğruluk ve Hata Oranı

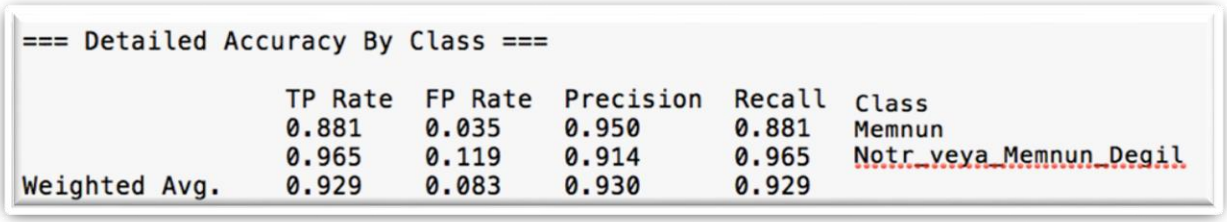
Doğruluk Oranı (Accuracy Rate): Genel olarak, sınıflayıcının ne sıklıkta doğru tahmin ettiğinin bir ölçüsüdür. (Beşer, 2018)

Hesaplanma şekli: (TP + TN) / TOPLAM

Yanlış Sınıflandırma Oranı (Misclassification Rate): Genel olarak, sınıflayıcının ne sıklıkta yanlış tahmin ettiğinin bir ölçüsüdür. Hata Oranı olarak da bilinir (Error Rate). (Beşer, 2018)

Hesaplanma şekli: (FP + FN) / TOPLAM

Buna göre Accuracy Rate %92.85 iken Error Rate %7.14 olarak hesaplanmıştır. Sınıflandırılan 103.904 verinin 96480'i doğru sınıflandırılmıştır. Şekil 13 ile oranların detaylandırılmış hali gösterilmiştir.



Şekil 13. Eğitim Verisinin KNN ile Detaylı Doğruluk Oranları

Gerçek Pozitif Değerlerin Oranı (True Positive Rate): Sınıflayıcının ne kadar gerçek pozitif değeri doğru tahmin ettiğinin bir ölçüsüdür. Hassasiyet, İsabet Oranı veya Hatırlama olarak da bilinir. (Sensitivity, Hit Rate or Recall) Mümkün olduğu kadar yüksek olmalıdır. (Beşer, 2018)
Hesaplanma şekli: TP / GERÇEK POZİTİFLER

Gerçek Negatif Değerlerin Oranı (True Negative Rate): Sınıflayıcının ne kadar gerçek negatif değeri doğru tahmin ettiğinin bir ölçüsüdür. Özgüllük veya Seçicilik olarak da bilinir. (Specificity or Selectivity) (Beşer, 2018)

Hesaplanma şekli: TN / GERÇEK NEGATİFLER

Program çıktısından görülebileceği üzere gerçekte memnun olup memnun olarak sınıflandırılan verilerin oranı %88 iken gerçekte memnun olmayıp tahminde de memnun olmayan olarak sınıflandırılan verilerin oranı ise %96.5'tur. Program memnun olmayan müşterileri daha iyi sınıflandırılmıştır.

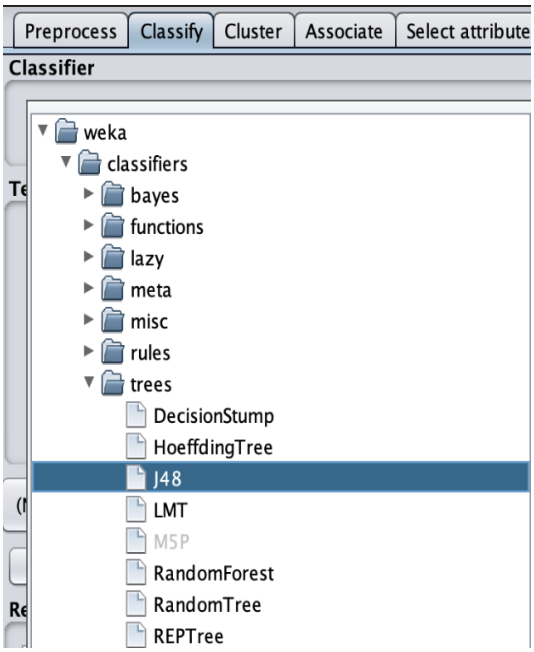
Hassasiyet (Precision): Tüm sınıflardan, doğru olarak ne kadar tahmin edildiğinin bir ölçüsüdür. Mümkün olduğu kadar yüksek olmalıdır. (Beşer, 2018)

Hesaplanma şekli: TP / (TP + FP) ve TN / (TN + FN)

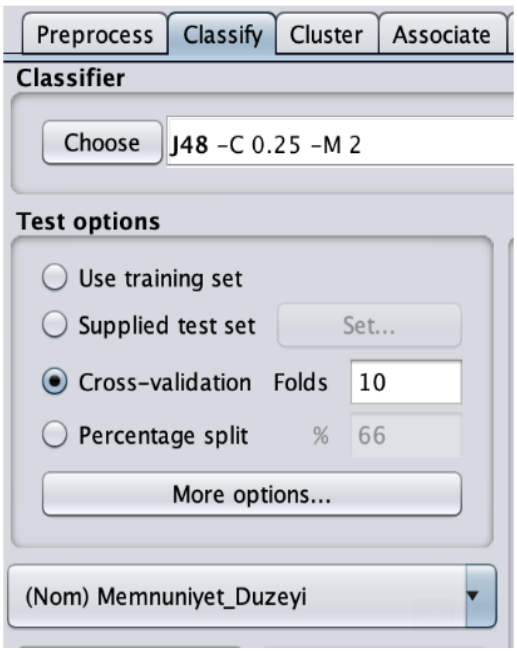
Program çıktısından memnun olma durumuna ait hassasiyet oranı %95 olarak hesaplanırken, memnun olmama durumunun tespit edilmesindeki hassasiyet oranı ise %91.4 olarak bulunmuştur.

4.2.2 Decision Tree Algoritması ile Sınıflandırma

WEKA programında Decision Tree algoritması içerisinde, sırasıyla şekil 14 ve şekil 15 deki adımlar izlenerek veri yüklendikten sonra explorer arayüzündeki classify sekmesinde, trees bölümündeki J48 seçeneği seçilerek kullanılmıştır.

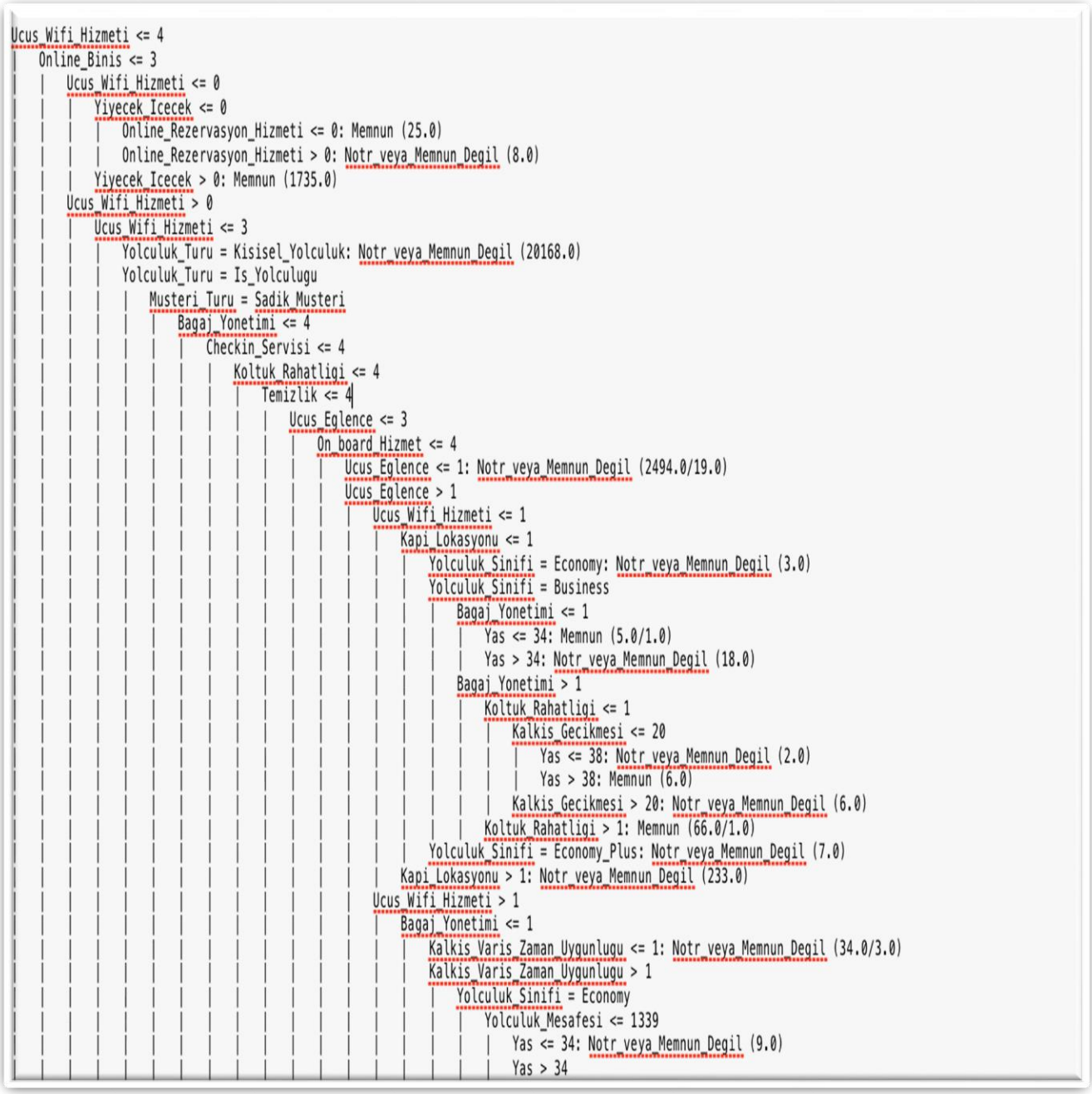


Şekil 14. WEKA'da Classify Sekmesi Bölümleri



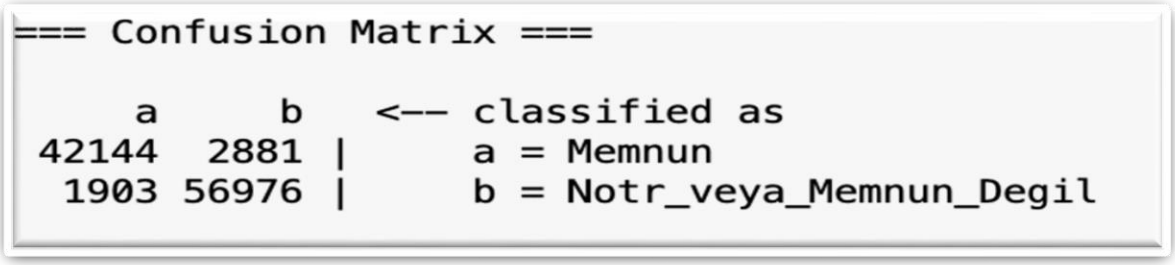
Şekil 15. WEKA'da Cross-Validation Yöntemi ile Decision Tree Algoritması Sınıflandırma İşlemi

Program aracılığı ile oluşturulan karar ağacı modeli, problemde kullanılan özellik sayısının çok fazla olması ve bu özelliklerin birbirleri ile etkileşimlerinin uygun hesaplanabilmesi amacı ile çok dallı ve uzun şekilde oluşmuştur. Karar ağacı modelinin başlangıcına ait bir önizleme aşağıdaki Şekil 16 da yer almaktadır.



Şekil 16. Eğitim Setinden Oluşturulan Karar Ağacı Modeli(Önizleme)

Sınıflandırma sonuçları aşağıdaki Şekil 17 gibidir:



Şekil 17. Uygulama Sonucu Confusion (Karışıklık) Matrisi

Bu bölümde kullanılan farklı doğruluk ölçütlerinin, confusion matrisinin genel çalışma mantığı ve kullanılan formüller KNN Algoritması ile sınıflandırma bölümünde detaylı şekilde anlatılmıştır.

Program çıktısından görülebileceği üzere, uygulamada A yerine memnun olan müşteri, B yerine ise nötr veya memnun olmayan müşteriler karşılık gelmektedir. Buna göre değerler:

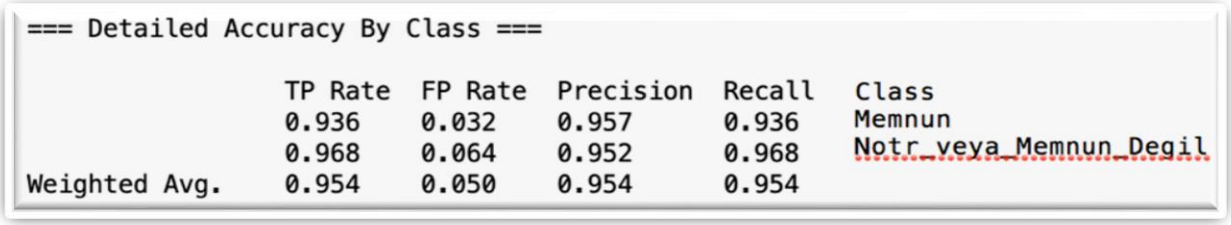
Gerçek Pozitif (TP) = 42144 (Gerçekte de tahminde de memnun olan müşteri sayısı)
Yanlış Pozitif (FP) =2881 (Gerçekte memnun tahminde memnun olmayan müşteri sayısı)
Yanlış Negatif (FN) =1903 (Gerçekte memnun olmayıp tahminde memnun olan müşteri sayısı)
Gerçek Negatif (TN) = 56976 (Gerçekte de tahminde de memnun olmayan müşteri sayısı) olarak bulunur.



Şekil 18. Eğitim verisinin Decision Tree İle Doğruluk Oranları

Doğruluk Oranı ve Yanlış Sınıflandırma oranı: Hesaplamaların sonucunda, yukarıda Şekil 18 ile gösterildiği gibi Accuracy Rate %95.3957 iken Error Rate %4.6043 olarak

hesaplanmıştır. Sınıflandırılan 103.904 verinin 99120'si doğru sınıflandırılmıştır. Aşağıdaki Şekil 19 ile oranlar detaylı gösterilmiştir.



```
=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   Class
                0.936    0.032    0.957      0.936    Memnun
                0.968    0.064    0.952      0.968    Notr_veya_Memnun_Degil
Weighted Avg.   0.954    0.050    0.954      0.954
```

Şekil 19. Eğitim Verisinin Decision Tree İle Detaylı Doğruluk Oranları

Gerçek Pozitif Değerlerin ve Gerçek Negatif Değerlerin Oranı(Recall): Program çıktısından alınan sonuca göre, gerçekte memnun olup memnun olarak sınıflandırılan verilerin oranı yaklaşık %94 iken gerçekte memnun olmayıp tahminde de memnun olmayan olarak sınıflandırılan verilerin oranı ise yaklaşık %97'dir. Program memnun olmayan müşterileri daha iyi sınıflandırılmıştır.

Hassasiyet(Precision): Program çıktısından memnun olma durumuna ait hassasiyet oranı %95.7 olarak hesaplanırken, memnun olmama durumunun tespit edilmesindeki hassasiyet oranı ise %95.2 olarak bulunmuştur.

4.3.Sınıflandırılan Verilerin İncelenmesi, Uygulanan Yöntemlerin Karşılaştırılması ve Uygun Modelin Seçimi

Bu aşama altında gösterilen; sınıflandırılan eğitim veri setlerinin incelenmesi ve başarı oranlarına ait detaylı bilgiler bir önceki aşamadaki, algoritmaların sınıflandırma için kullanıldıkları kendi bölümlerinde yer almaktadır. Test verilerinin sınıflandırılması için dikkate alınacak olan iki kıstas, algoritmaların verileri sınıflandırırken ne kadar hızlı oldukları ve verileri sınıflandırırken ne kadar başarılı oldukları olarak belirlenmiştir. Aşağıda bu iki kıstasa göre karşılaştırma ve yorumlar ifade edilmiştir.

Süre/Hız: İlk olarak modellerin kurulması ve çalıştırılması sırasında bu işlemlerin gerçekleşme süreleri ele alınacak olursa KNN algoritmasında test verisinin her birisi eğitim verisindeki bütün veriler ile tek tek kıyaslanmakta ve en yakın veri sayısını tespit edilerek n komşu sayısına göre atama gerçekleştirmekteydi. Bu işlem problemdeki gibi veri sayısı 100.000 civarında olan bir

işlem için gerçekleştirildiğinde hesaplama işlemi yaklaşık olarak 30 dakikaya yakın sürmüştür. Ayrıca bu işlem en uygun k sayısını belirlemek üzere 5 kez tekrarladığından dolayı sırf bu algoritma için hesaplama işlemi yaklaşık olarak 2.5 saat sürmüştür. Karar ağacı algoritmasında ise program başlangıçta eğitim setinden belirli bir model kurmakta ve her veriyi bu karar ağacı modelindeki şartlara bağlı olarak sınıflandırdığı için hesaplama işlemi yüksek boyutlu veri sayısına rağmen 10 dakikadan kısa sürmüştür. Ayrıca karar ağacı algoritmasında ideal sonuca ulaşmak için tek kez uygulama yapmak yeterlidir. Dolayısı ile karar ağacı yöntemini uygulamak süre kazanımı açısından daha avantajlıdır.

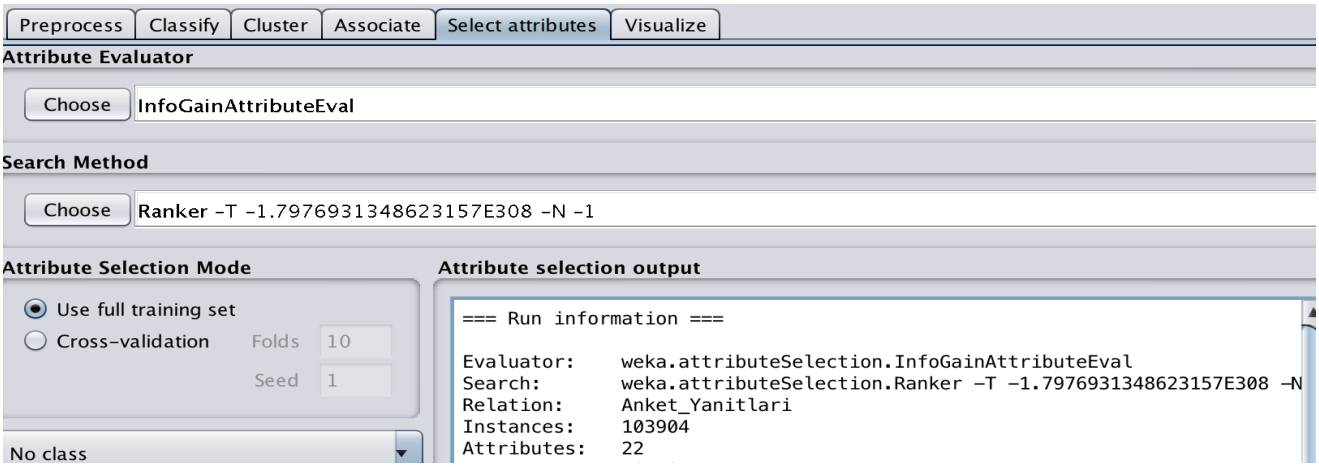
İsabetlilik/Doğruluk: Algoritmaların hesaplandığı kendi bölümlerinde isabetlilik oranları verilmiştir. Buna göre KNN algoritması ile toplam verilerin %92.85'i gerçek etiketin aynısı ile, %7.15'i ise gerçek etiketten farklı olarak etiketlenirken, Karar ağacı algoritması ile toplam verilerin %95.39'u gerçek etiketin aynısı ile, %4.60'ı gerçek etiketten farklı etiketlenmiştir. Toplam veri sayısı göze alındığında bu yüzdelik fark, 2640 adet daha fazla verinin doğru şekilde sınıflandırılması demektir. Toplam veri sayısı arttıkça bu oran farkı daha da anlam kazanacaktır. Bu aradaki fark hangi yöntemin seçilmesine dair fikir oluşturması açısından başlıca etkendir. Bu alanda da Decision Tree algoritması daha başarılı sonuçları vermiştir.

Buna göre bu karşılaştırılan veriler doğrultusunda test verisi için karar ağacı algoritmasının kullanılmasına karar verilmiştir.

4.4. Özellikler Açısından Verilerin Analizi ve Son Düzeltmeler

WEKA programında ayrıca kullanılan veriler ve bunların etiketlerine bağlı olarak, hangi özelliğin sonuca daha çok etki ettiği gözlemlenebilmektedir. Bu da organizasyonlar açısından oldukça büyük önem teşkil eder. Bu sayede veriler üzerinde etkisi olmayan özellikler yok sayılabilip genel doğruluk ve isabetlilik oranı arttırılabilmekte, bunun yanı sıra sonuç için en çok önem teşkil eden özelliklerin öğrenilmesi ile birlikte işletmede bu alanlara yönelik iyileştirme ve geliştirme çalışmalarının ne kadar daha önemli olduğu anlaşılabilir. Hangi özelliğe ne kadar ağırlık verilmesi gerektiği tespit edilebilir ve bunlara ait planlama eylemleri de önceden oluşturulabilir.

Aşağıda WEKA'da bu işlemin nasıl gerçekleştirildiği gösterilmektedir. Veriler uygulamada açıldıktan sonra ana işlem çubuğundan select attributes kısmı, bu kısımda da attribute evaluator ve search method kısımları aşağıdaki Şekil 20 deki gibi seçilmiştir.



Şekil 20. Özelliklerin Önem Kıyaslama İşleminin Uygulaması

Bu işlem sonucunda özellikler ve önem derecelerine göre sıralanmaları aşağıdaki şekildeki gibi oluşmaktadır.

```
Ranked attributes:
 0.304204   12 Online_Binis
 0.233285    7 Ucus_Wifi_Hizmeti
 0.192611    5 Yolculuk_Sinifi
 0.163941    4 Yolculuk_Turu
 0.135205   14 Ucus_Eglence
 0.113794   13 Koltuk_Rahatligi
 0.087614   16 Bacak_Yeri
 0.082517   15 On_board_Hizmet
 0.076315    6 Yolculuk_Mesafesi
 0.07477    19 Temizlik
 0.07333     9 Online_Rezervasyon_Hizmeti
 0.065752    3 Yas
 0.061506   17 Bagaj_Yonetimi
 0.04602    18 Checkin_Servisi
 0.037806   11 Yiyecek_Icecek
 0.026819    2 Musteri_Turu
 0.017336   10 Kapi_Lokasyonu
 0.009266   21 Varis_Gecikmesi
 0.004036   20 Kalkis_Gecikmesi
 0.003109    8 Kalkis_Varis_Zaman_Uygunlugu
 0.000108    1 Cinsiyet
```

Şekil 21. Özelliklerin Etki Sıralaması

Şekil 21 deki sonuçlara göre, bu havayolu şirketinin anket sonuçlarında online biniş işlemlerindeki servis kalitesi, uçuş sırasında sunulan wifi internet hizmetinin kalitesi ve yolcuların yaptıkları uçuştaki sınıfın farklılığı memnun olup olmama durumunda daha önemli faktörler iken, görülebileceği üzere yolcuların cinsiyetinin farklılığı memnun olup olmama durumunu hemen hemen hiç etkilememektedir. İşletme anketini buna yönelik tekrar düzenleyebilir, sonuç için bağımlı değişkenlerden daha çok önem teşkil eden alanlar üzerinde daha titiz çalışabilir, hizmetlerinin önem planlamalarını bu verilere göre oluşturabilir.
Bu veriler doğrultusunda eğitim verisindeki anket sorularından cinsiyet kısmına ait soruyu ve cevapları kaldırmak mantıklı görünmektedir. Bu sayede sonuca etki etmeyen bir soru anketten çıkartılacak ve müşterilerin zamanı boşa harcanmayacaktır. Aşağıda, önceden seçtiğimiz karar ağacı algoritması, verilerdeki cinsiyet bölümü çıkartılarak eğitim verisi üzerinde tekrar uygulanmıştır.



Şekil 22. Eğitim Verisinin Cinsiyet Verisi Çıkarılmış Halinin Decision Tree ile Doğruluk Oranları

Şekil 22 den de görülebileceği gibi, beklenildiği üzere eğitim setinden cinsiyet sorusuna yönelik veriler çıkarıldıktan sonra gerçekleştirilen sınıflandırma sonucunda algoritmanın oluşturduğu modelin doğruluk oranı azalmamıştır. Hatta doğruluk oranı %95.39'dan %95.51'e yükselmiştir. Bu yükselme oranının çok büyük farklar yaratmayacağı düşünülse de havayolu sektöründeki ekonomik pay büyüklüğü göz önüne alındığında, önceden bahsedildiği üzere her müşteriyi kazanabilmenin önemi oldukça büyüktür. Ayrıca veri sayısı arttıkça bu oranın öneminin artacağı aşikardır. Örneğin bu ufak düzeltme sayesinde gerçekleştirilen çalışmada 124 adet daha fazla müşterinin anket yanıtı doğru tespit edilmiştir.
Ardından başarı oranının daha fazla artıp atmayacağını tespit etmek üzere, cinsiyet ifadesinden bir sonraki en az etkiye sahip ifade olarak tespit edilen ifade olan kalkış varış zaman uygunluğu için de aynı işlem uygulanarak, verilerin son halinden(cinsiyetsiz) çıkartılmış ve yeni doğruluk oranı hesaplanmıştır. Aşağıdaki şekilde yeni doğruluk oranı görünebilmektedir.



Şekil 23. Eğitim Verisinin Cinsiyet ve Kalkış-Varış Zaman Uygunluğu Verileri Çıkarılmış Halinin Decision Tree ile Doğruluk Oranları

Şekil 23 ile görülebileceği üzere, bu işlem sonucunda da başarı oranı artmaya devam ederek %95.51'den %95.54'e yükselmiştir. Böylelikle son durumdan 30 adet daha fazla müşterinin anket yanıtı daha doğru tespit edilebilmiştir.

Ardından bu iki özellikten sonraki en düşük etkiye sahip olan kalkış gecikmesi özelliği için de aynı işlem yapılmıştır. Bu ifadenin de son durumdaki verilerden çıkarılması ile aşağıdaki şekilde görülebilecek olan doğruluk oranı elde edilmiştir.



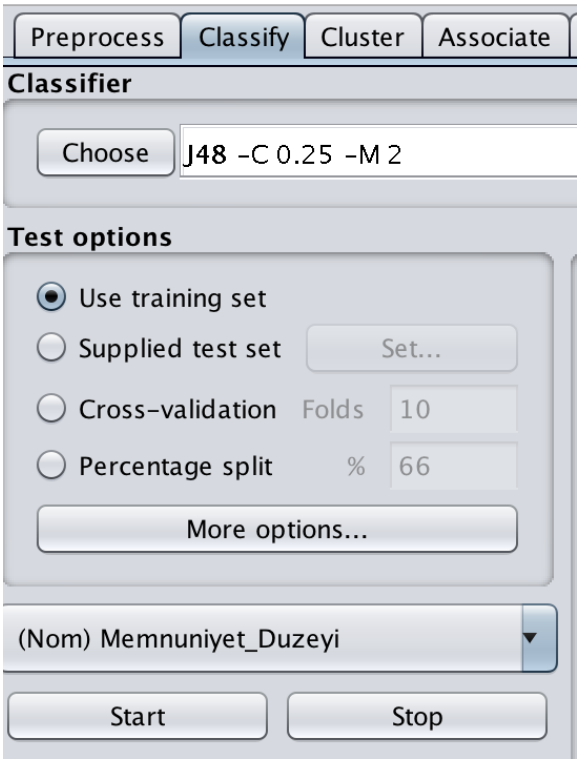
Şekil 24. Eğitim Verisinin Cinsiyet, Kalkış-Varış Zaman Uygunluğu ve Kalkış Gecikmesi Verileri Çıkarılmış Halinin Decision Tree ile Doğruluk Oranları

Bu işlem de tamamlandıktan sonra Şekil 24 ile gösterilen en sonki doğruluk oranı, bir önceki doğruluk oranından düşük çıkmıştır. Bu da artık çıkartılan verilerin başarı oranına arttırmadığı, aksine azalttığını göstermektedir. Dolayısı ile en yüksek başarı oranı ile ilerlemek için, çalışmaya bu çıkarma işlemini yapmadan, yalnızca cinsiyet ve kalkış-varış zaman uygunluğu verileri çıkartılarak devam edilecektir.

4.5. WEKA'da Test Verilerinin Sınıflandırılması

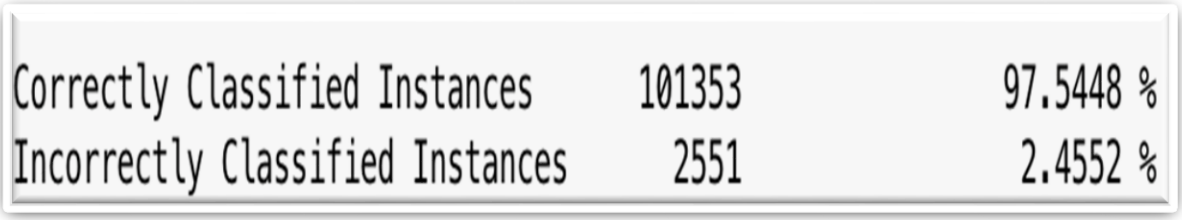
Bu aşamada yanıtları bilinmeyen 25976 anket verisi, seçilen karar ağacı yöntemi ve yapılan düzenlemelerin sonucunda elde edilen eğitim seti verilerinden oluşturulan sınıflandırma modeli ile sınıflandırılacaktır.

WEKA'da istenilen bir veri setinin, başka bir veri setinin eğitim seti olarak kullanılması aşağıdaki şekilde gerçekleştirilir. Veri seti yüklendikten sonra Şekil 25 de gösterilen classify sekmesinden, istenilen algoritma seçilir. Daha sonra "use training set" opsiyonu seçilerek çalıştırılır.



Şekil 25. WEKA'da İstenilen Verinin Eğitim Seti Olarak Kullanılması

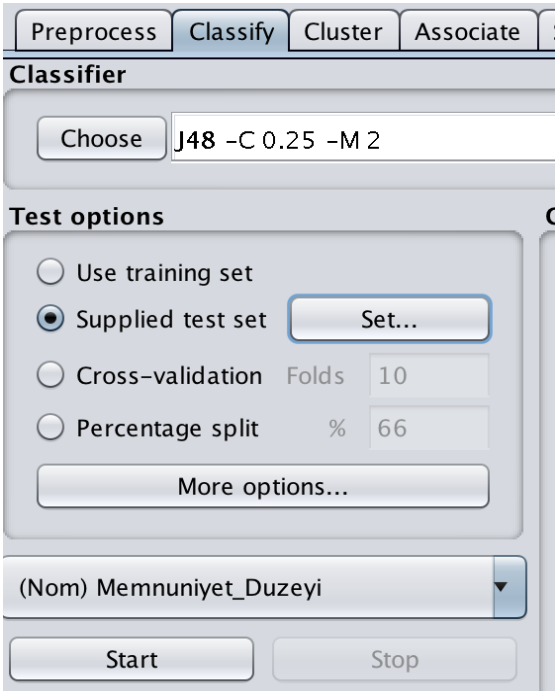
Bu işlemin ardından kullanılacak olan eğitim setinden sınıflandırma modeli oluşturulacaktır. Bu aşamada dikkate alınması gereken bir durum, eğitim seti olarak kullanılmak istenen bir veri setinin başarı oranı bu aşamadan daha önce gerçekleştirdiğimiz aşamalar ile ölçülmelidir. Çünkü bu aşamada gerçekleştirilen işlemde, verilerin hiçbirisi kendi içerisinde eğitim ve test parçaları olarak ayrılmadan, direkt olarak tek seferlik bir eğitim seti olarak işlenir. Bu da bir nevi veri seti içerisindeki farklı veya uç durumları ezberlemek gibi düşünülebilir. Bu yüzden bu işlemin sonucunda gösterilen doğruluk oranı, eğitim aşaması sırasında kendisinin de bire bir işlenerek yer aldığı verilerin sınıflandırılmasındaki başarı oranıdır. Cross Validation sisteminden hatırlanacağı üzere o işlemde başarı oranına bakılırken verilerin tamamı 10 parçaya bölünmekte ve bu parçaların 9'u eğitim için kullanılırken 1 tanesi test için kullanılmaktaydı ve başarı oranlarının ortalaması alınmaktaydı. Bu da bizim kontrol için kullandığımız başarı oranının veri setinin bilinmeyen verilere karşı ne derece başarılı tahminler yapabildiğini göstermektedir. Bu yüzden bir veri setini eğitim seti olarak kullanmaya karar vermeden önce kullanılan algoritma ve yöntemlerin o setin daha önceden barındırmadığı verileri ne derece başarılı tahmin edebildiği gözlemlenilip karar buna göre verilmelidir. Aşağıda Şekil 26 da karar ağacı algoritması seçilerek eğitim verilerinin tamamı eğitim için kullanıldığında başarı oranı görünmektedir.



Şekil 26. Eğitim Verisi Olarak Kullanılacak Seti Kendisini Sınıflandırmada Başarı Oranı

Tahmin edilebileceği üzere bu verilerin başarı oranı(%97.5) bizim hesapladığımız son başarı oranından(%95.54) daha yüksek çıkmıştır. Bunun sebebi bir önceki bölümde aktarılmıştır. Test verileri sınıflandırılırken beklenen başarı oranının bu orana değil, bizim hesapladığımız orana yakınsaması beklenmelidir.

Test verileri aşağıda bahsedilecek olan uygun formata getirildikten sonra WEKA'da açıldıktan sonra aşağıdaki şekildeki gibi seçilip sınıflandırılmaktadır. İstenen algoritma seçildikten sonra "supplied test set" opsiyonu seçilerek çalıştırılmalıdır. Bu aşamada dikkat edilmesi gereken bir husus, eğitim setindeki toplam özellik sayısı ve farklı özellik değer seçenekleri, test setindeki toplam özellik sayısı ve farklı özellik değer seçenekleriyle bire bir uyuşmalıdır. Örneğin yolcunun yolculuk sınıfına ait seçenekleri, eğitim setinde Business ve Economy iken, test setinde yalnızca Business olamaz, bu değerlerle uyuşmayan başka tanımlar olamaz veya bu değerleri bulundurmasına rağmen ek olarak başka bir tanım içeremez. Formatlar birbirine tamamen uyumlu olmalıdır. Test setinde eğitim setinden farklı olarak yalnızca anketlerdeki memnun olup olmama durumuna ait sonuç değeri yanıtsız olarak bulunmaktadır. Yani bu özellik tanımlıdır ancak özelliğin değerleri atanmamış (null) değerlerdir. WEKA'da yanıtı bulunmayan özellik değerleri "?" değeri olarak tanımlanmalıdır. Bu ifade yerine boşluk bırakmak veya herhangi başka bir ifade yazmak programın hata vermesine neden olacaktır.
Aşağıdaki Şekil 27 de Test Seti kullanılması için seçilecek alan gösterilmiştir.



Şekil 27. WEKA'da İstenilen Verinin Test Seti Olarak Kullanılması

4.6. Sınıflandırılan Test Verilerinin İncelenmesi ve Yorumlanması

Test verileri eğitim verisi kullanılarak sınıflandırılmıştır. Sınıflandırılan ifadeler düzenlenip Excel'e aktarılmıştır. Aşağıda önizleme olarak 30 adet memnun olup olmadığına dair yanıtı bulunmayan anket verisinin sınıflandırılmış hali hücrelerde gösterilmektedir. Verilerin sığdırılabilmesi ve okunabilmesi açısından özelliklere ait başlıklar küçültülmüştür. Şekilde kolonlarda bulunan başlıklar soldan sağa olarak şunlardır: Müşteri Türü, Yaş, Yolculuk Türü, Yolculuk Sınıfı, Yolculuk Mesafesi, Uçuş Wifi Hizmeti, Online Rezervasyon Hizmeti, Kapı Lokasyonu, Yiyecek-İçecek, Online Biniş, Koltuk Rahatlığı, Uçuş Eğlence, On board Hizmet, Bacak Yeri, Bagaj Yönetimi, Checkin Servisi, Temizlik, Kalkış Gecikmesi, Varış Gecikmesi,

Tahmin Güveni, Tahmini Sınıf/Yanıt

Tablo 5. Sınıflandırılan Test Verileri

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Musteri_Tur	Yas	Yolculuk_Tu	Yolculuk_	Yolculu	Ucus_	Onlir	Kapi	Yiy	On	Kol	Ucu	On_	Baca	Bag	Che	Tem	Kalki	Varis	Tahmin_Guv	Tahmini_Sinif
2	Sadik_Muste	52	Is_Yolculug	Economy	160	5	3	4	3	4	3	5	5	5	5	2	5	50	44	0.990596	Memnun
3	Sadik_Muste	36	Is_Yolculug	Business	2863	1	3	1	5	4	5	4	4	4	4	3	5	0	0	1	Memnun
4	Sadik_Olmay	20	Is_Yolculug	Economy	192	2	2	4	2	2	2	2	4	1	3	2	2	0	0	-0.973858	Notr_veya_Memnun_Degil
5	Sadik_Muste	44	Is_Yolculug	Business	3377	0	0	2	3	4	4	1	1	1	1	3	4	0	6	1	Memnun
6	Sadik_Muste	49	Is_Yolculug	Economy	1182	2	4	3	4	1	2	2	2	2	2	4	4	0	20	-0.928571	Notr_veya_Memnun_Degil
7	Sadik_Muste	16	Is_Yolculug	Economy	311	3	3	3	5	5	3	5	4	3	1	1	5	0	0	1	Memnun
8	Sadik_Muste	77	Is_Yolculug	Business	3987	5	5	5	3	5	5	5	5	5	5	4	3	0	0	0.99573	Memnun
9	Sadik_Muste	43	Is_Yolculug	Business	2556	2	2	2	4	4	5	4	4	4	4	5	3	77	65	1	Memnun
10	Sadik_Muste	47	Is_Yolculug	Economy	556	5	2	2	5	5	5	5	2	2	5	3	5	1	0	0.997768	Memnun
11	Sadik_Muste	46	Is_Yolculug	Business	1744	2	2	2	3	4	4	4	4	4	4	5	4	28	14	1	Memnun
12	Sadik_Muste	47	Is_Yolculug	Economy	1235	4	1	1	5	1	5	3	3	4	3	1	4	29	19	1	Memnun
13	Sadik_Muste	33	Is_Yolculug	Business	325	2	5	5	1	3	4	2	2	2	2	3	4	18	7	-0.928079	Notr_veya_Memnun_Degil
14	Sadik_Muste	46	Is_Yolculug	Business	1009	5	5	5	4	5	5	5	5	5	5	5	3	0	0	0.99573	Memnun
15	Sadik_Muste	60	Is_Yolculug	Business	451	1	4	1	5	5	4	5	5	5	5	3	5	117	113	1	Memnun
16	Sadik_Muste	52	Is_Yolculug	Business	925	2	2	2	5	5	4	4	4	4	4	3	5	10	0	1	Memnun
17	Sadik_Muste	50	Kisisel_Yolc	Economy	83	3	0	3	2	0	2	2	4	2	4	4	2	5	2	-1	Notr_veya_Memnun_Degil
18	Sadik_Muste	31	Is_Yolculug	Economy	728	2	5	5	2	2	2	2	4	3	3	4	2	2	0	-0.928079	Notr_veya_Memnun_Degil
19	Sadik_Muste	52	Kisisel_Yolc	Economy	1075	5	5	3	4	5	4	4	3	5	5	4	4	0	0	0.99573	Memnun
20	Sadik_Muste	43	Kisisel_Yolc	Economy	1927	3	3	1	4	4	5	5	5	3	5	4	3	0	0	-1	Notr_veya_Memnun_Degil
21	Sadik_Muste	50	Is_Yolculug	Business	3799	5	5	5	4	5	4	4	4	5	4	5	5	8	0	0.99573	Memnun
22	Sadik_Muste	60	Is_Yolculug	Business	612	4	4	4	2	4	5	5	5	5	5	5	5	21	49	0.995249	Memnun
23	Sadik_Muste	43	Kisisel_Yolc	Economy	1437	3	3	4	2	3	2	2	4	2	4	4	2	0	0	-1	Notr_veya_Memnun_Degil
24	Sadik_Muste	55	Kisisel_Yolc	Economy	302	1	4	3	4	4	4	4	1	3	2	4	4	0	0	-1	Notr_veya_Memnun_Degil
25	Sadik_Muste	25	Is_Yolculug	Business	1428	4	4	4	4	4	4	4	1	5	3	1	4	0	0	0.833333	Memnun
26	Sadik_Olmay	30	Is_Yolculug	Economy	528	4	5	3	2	5	2	2	3	2	3	4	2	0	0	-0.619048	Notr_veya_Memnun_Degil
27	Sadik_Muste	62	Kisisel_Yolc	Economy	710	3	3	4	2	3	2	2	3	5	5	4	2	0	0	-1	Notr_veya_Memnun_Degil
28	Sadik_Muste	24	Is_Yolculug	Business	3680	4	4	4	2	2	2	2	5	5	5	5	2	0	0	1	Memnun
29	Sadik_Muste	22	Is_Yolculug	Economy	1521	4	1	1	4	4	4	4	1	4	1	1	4	3	13	0.699187	Memnun
30	Sadik_Muste	44	Is_Yolculug	Business	1543	3	3	3	5	4	5	5	5	5	5	5	5	0	0	1	Memnun
31	Sadik_Muste	51	Is_Yolculug	Economy	235	4	3	3	4	4	4	4	3	2	3	3	4	0	0	0.25	Memnun

Tablo 5 de gösterildiği gibi WEKA'da test verilerinin sınıflandırma işlemi yapılırken program yaptığı her tahmin sonucu için **[-1, 1]** aralığında, güven seviyesi belirlemektedir. 0 ile 1 arasında yer alan güven değerlerine ait verilerin hepsi memnun olan müşteri sınıfına dahil edilmiştir. 0 ile -1 arasında yer alan güven değerlerinin hepsi ise memnun olmayan, nötr müşteri sınıfına dahil edilmiştir.

Burada güven değeri mutlak olarak bakıldığında 1 değerine ne kadar yakın ise yapılan tahmine ait güven değeri o kadar yüksektir. Yani kurulan model, etiketlediği sınıfa o kadar fazla güvenmektedir. Örneğin yukarıdaki önizlemede 2 sıra no'lu hücrede bulunan ilk müşterinin memnun olmasına yönelik tahmin değeri 0.99 güven aralığı ile ifade edilirken, 25 sıra no'lu hücrede bulunan müşterinin memnun olmasına yönelik tahmin değeri 0.83 güven aralığı ile ifade edilmiştir. Bu durumda ilk müşterinin memnun olma durumunun diğer müşteriye kıyasla daha fazla olduğu söylenebilmektedir. Aynı şekilde örneğin 17 sıra no'lu hücrede bulunan

müşterinin memnun olmama durumuna yönelik tahmin değeri -1 iken, 26 sıra no'lu hücrede bulunan müşterinin memnun olmama durumuna yönelik tahmin değeri -0.61'dir. Mutlak değerlere bakılacak olursa ifadeler sırası ile 1 ve 0.61 olacaktır. İlk müşterinin memnun olmadığın neredeyse kesin olarak söylenebilecekken diğer müşteri için bu kadar rahat söylenememektedir.

Tahmin güveni değeri, yapılan çalışmadaki özelliklik sayısı çok fazla olan komplike problemler için oldukça büyük önem taşımaktadır. Bu gibi durumlarda işletme için yalnızca memnun olup olmama etiketinin basılması ve bunun yanıtına ulaşılması yeterli olmayabilir. Bu iki yanıta da çok yakın durumda bulunan birçok anket verisi olması kaçınılmazdır. Bu sayede işletme güven değerlerine göre de analiz yapabilir ve daha başarılı sonuçlara ulaşabilir. Ayrıca sınıflandırılan test verilerinin excel ortamına rahatlıkla aktarılabilmesi de yine işletme için analiz çalışmalarında kolaylık sağlayan bir özelliktir. Bu sayede farklı attribute(özellik)lere yönelik sorgular yapılabilir. Spesifik değerler ve durumlar için müşterilerin memnuniyet seviyesi istenildiği şekilde gözlemlenebilir.

BÖLÜM 5. İSTATİSTİKSEL ANALİZ

Veri madenciliği, istatistik temelli çalışmaların uygulandığı bir daldır. Bu kısımda, çalışmada uygulanan yöntemler ve düzeltmelerin neden bu şekilde seçildiği, bu düzeltmeler sonucunda yakalanan başarı oranının diğer yöntem veya düzeltme işlemleri karşısında ne derece başarılı olduğu istatistiksel olarak özetlenecektir.

İlk olarak algoritmalar açısından kıyaslayacak olursak elimizde bulunan eğitim verisinin yapısına uygun olan, çalışmada kullanılmak üzere seçilen algoritmaların ve bazı diğer bilinen popüler sınıflandırma algoritmalarının eğitim verisi için başarı oranları aşağıdaki grafikte yer almaktadır.

Tablo 6. Farklı Algoritmaların Başarı Oranları

	KNN(k=1)	Decision Tree	Naive Bayes	OneR	Linear Regresyon	Lojistik Regresyon
Toplam Veri	103904	103904	103904	103904	103904	103904
Doğru Sınıflandırılan	95238	99120	88724	82124	90899	94229
Yanlış Sınıflandırılan	8666	4784	15180	21780	13005	9675
Doğruluk Oranı	91.66%	95.39%	85.39%	79.04%	87.48%	90.68%

Tablo 6 dan da görülebileceği üzere örnek olarak tek bir özellikten formül çıkarımı yaparak sınıflandırma işlemini uygulayan bir sınıflandırma algoritması olan OneR ile yalnızca %79 oranında başarı sağlanabilmiştir. Bu da verilerin karmaşık yapısında hakkında fikir vermekte ve daha gelişmiş ve özelleşmiş sınıflandırılma yaklaşımları ile çalışma yapılması gerektiğini özetler. OneR yaklaşımından daha gelişmiş olan ve bütün özelliklerin etkileşimini hesaba katan diğer yöntemlerden Naive Bayes, Lineer Regresyon ve Lojistik Regresyon ile daha başarılı sonuçlara ulaşılmıştır. Ancak görülebileceği üzere en yüksek başarı oranına sahip olan iki algoritma KNN ve Karar Ağacı olarak tespit edilmiştir. Dolayısı ile çalışmada kullanılmak üzere bu algoritmalar seçilmiştir.

Daha sonra bu yöntemlerden KNN algoritmasında en yakın komşu sayısını kullanıcın seçebildiği ve bu seçim sonucunda doğruluk oranının farklı değerler verebildiği incelenmişti. Şekil 9'dan da detaylı olarak incelenebileceği üzere en yakın komşu değerinin belirli değerler için verdiği doğruluk oranı aşağıdaki tabloda tekrar özetlenmiştir.

Tablo 7. Farklı K Değerleri İçin Başarı Oranları

K komşu değeri	1	3	5	7	9
Doğruluk Oranı	91.66%	92.51%	92.74%	92.85%	92.81%

K=7 için en yüksek doğruluk oranına %92.85 ile ulaşılsa da, bu değer hala Karar Ağacı yönteminin başarı oranından düşük olduğu için bu aşamada da KNN algoritması elenip, çalışmaya Karar Ağacı algoritması kullanılarak devam edilmiştir.

Ardından özellikler arası veriler incelenmiş ve eğitim veri setinde hangi özellik değerinin sınıflandırma üzerinde etkisinin ne kadar olduğu incelenmiştir. Buna göre cinsiyet özelliğinin sınıflandırmaya etkisinin diğer özelliklerden büyük bir farkla ayrışarak en az etkiye sahip olduğu tespit edilmiştir. Algoritmanın başarı oranı cinsiyet özelliğine ait veriler çıkarılarak tekrar hesaplanmış ve ilk durum ile kıyaslanmıştır. Aşağıdaki Tablo 8'de bu kıyas görülebilir. İşlem sonucunda tablodan da görülebileceği üzere cinsiyeti bulundurmayan anket verileri ile daha başarılı tahmin oranı elde edilmiştir. Dolayısı ile çalışmaya veriler bu şekilde kullanılarak ilerlenilmektedir.

Tablo 8. Cinsiyet Verileri Çıkarılan ve Çıkarılmayan Başarı Oranları

Cinsiyet	Bulunduran Veri	Bulundurmayan Veri
Doğruluk Oranı	95.39%	95.51%

Ardından, cinsiyetten sonraki en az etkiye sahip ifade kalkış-varış zaman uygunluğuna ait yanıtlar olarak tespit edilmiştir. Bu özelliğe ait veriler son durum(cinsiyetsiz) verilerinden çıkarılarak bu iki durum değerlerinin başarı oranı da kıyaslanmıştır. Aşağıdaki Tablo 9'da bu iki durum görülebilmektedir.

Tablo 9. Cinsiyet ve Kalkış Varış Uygunluğu Verileri Çıkarılan ve Çıkarılmayan Başarı Oranları

Kalkış-Varış Uygunluğu	Bulunduran Veri	Bulundurmayan Veri
Doğruluk Oranı	95.51%	95.54%

Tablodan görüleceği üzere, kalkış-varış zaman uygunluğu yanıtlarının da verilerden çıkarılması ile doğruluk oranı %95.51'den %95.54'e yükselmiştir. Bu yüzden çalışmaya bu verilerin de çıkarılarak devam edilmesine karar verilmiştir.

Son olarak bu iki özelliğin dışındaki, sonuca en az etki ifade olan kalkış gecikmesine yönelik olan özellik için de aynı işlem uygulanmıştır. Önceki son durum ile yeni son durumun başarı oranı kıyaslaması aşağıdaki tabloda görülebilmektedir.

Tablo 10. Cinsiyet, Kalkış Varış Uygunluğu ve Kalkış Gecikmesi Verileri Çıkarılan ve Çıkarılmayan Başarı Oranları

Kalkış Gecikmesi	Bulunduran Veri	Bulundurmayan Veri
Doğruluk Oranı	95.54%	95.45%

Tablo 10'dan görüleceği üzere, bu özelliğin çıkarılması ile başarı oranı son duruma göre azalmıştır. Dolayısı ile eğitim setinden kalkış gecikmesi yanıtlarına ait veriler çıkarılmamıştır.

Sonuç olarak çalışmaya eğitim setinden cinsiyet ve kalkış-varış zaman uygunluğu verileri çıkartılarak ve karar ağacı algoritması uygulanarak devam edilmiş, test verileri buna göre sınıflandırılmıştır.

BÖLÜM 6. GERÇEKCİ KISITLAR VE KOŞULLAR ALTINDA DEĞERLENDİRME

Uygulanan çalışma **ekonomik, sürdürebilirlik ve güvenlik** açısından incelendiğinde, veri madenciliği çalışmaları genel olarak bilgiyi işleme, yorumlama ve yeni bilgiler elde etme olarak düşünülürse, fiyat/performans olarak oldukça elverişlidir. Çünkü ortaya konan hizmetlerin/ürünlerin girdisi bilgidir. Araç olarak da bu bilgileri kullanma ve yorumlamaya yardımcı olan, aslında yine bilgi temin eden programlar ve hesaplayıcı donanımlar kullanılmaktadır. Çıktı olarak da bu bilgilerden, uzun vadeli olarak bakıldığında işletmeler için yüksek boyutlarda ekonomik kazanç sağlayan anlamlı bilgilere ulaşılmaktadır. Çalışmada kullanılan veri sayısının yüksek boyutlu olduğu söylenebilir. Bazı durumlarda çok daha yüksek boyutlu veriler ile çalışmak da gerekebilir. Bu durumda çalışmayı gerçekleştirilen personelin yeterince kalifiye olduğu varsayılır ise ve veri madenciliği gerçekleştirilebilen bazı kaliteli programların ücretsiz olarak sunulduğu düşünülürse, uygulamaların en önemli maliyeti işlemci (CPU) performansı yüksek bilgisayarlar kullanılmak zorunda olunmasıdır. Ancak bahsedildiği gibi devamlı olarak anlamlı bilgiler elde edip bu bilgileri etkili olarak kullanmak, uzun vadede hesaplama donanımlarının maliyetlerinin çok rahat bir şekilde göz ardı edilebileceğini gösterir. Bu tarz çalışmaların sürdürülebilirliği de veri madenciliği programlarının günümüzdeki seviyesi düşünüldüğünde kolaylıkla sağlanabilmektedir. Her ne kadar kullanılan verilerin miktarı ve anlamı sürekli değişebilse de, farklı veri tarzları ve miktarları için geliştirilen algoritmalar ve bu algoritmalar üzerinde yapılabilen otomatikleşmiş ayarlar, çoğu durumda yüksek oranda başarı ve doğruluk oranlarına ulaşılmasına imkan sağlamaktadır. Son olarak etik ve güven açısından bakıldığında, çalışmalarda firmalara ait çok önemli veriler kullanılabileceği için, bu verilerin güvenliğini sağlamak büyük önem taşımaktadır. Aynı şekilde veri madenciliği konusunda danışmanlık hizmeti veren işletmeler birlikte çalıştığı işletmelere, kullanacağı verilerin hiçbir şekilde başka personel veya işletmelere izinsiz şekilde satılmaması veya iletilmemesi konusunda güvence vermeli, bu tarz durumlar için karşılıklı güven sözleşmeleri kullanılmalıdır.

BÖLÜM 7. SONUÇ

Sonuç olarak genel olarak fazla sayıda farklı tanım (ifade / cevap vs.) bulunduran müşteri verilerinin işletmelerin yararına olacak şekilde nasıl ve ne şekildiği analiz edilebileceğine dair bir sınıflandırma çalışması gerçekleştirilmiştir.

Veri madenciliği yaklaşımlarının çok boyutlu verileri anlamlandırmada önemi bir havayolu firmasının anket verilerinden yararlanılarak ortaya konulmuştur. Bu sayede hem havayolu işletmeleri için örnek bir çalışma gerçekleştirilmiş, bu tarz analizlerin ve çalışmaların firmaya ne tür katkıları olduğu aktarılmış hem de benzer tarzda verilere sahip olan birçok farklı sektörden işletmeler için de örnek olabilecek bir analiz ve geliştirme uygulama örneği sunulmuştur.

Çalışma özelinde ise, kullanılan verileri en uygun şekilde sınıflandırabilmek için farklı sınıflandırma algoritmalarından KNN ve Decision Tree algoritmalarının diğer popüler başarılı sınıflandırma algoritmalarına kıyasla daha başarılı olduğu saptanmıştır. Bu iki algoritmaların işleme mantıkları da örneklerle desteklenerek okuyucuya aktarılmaya çalışılmıştır. Bu algoritmaların kıyaslanması sonucunda Decision Tree'nin bu veriler için en uygun algoritma olduğu ispatlanmış, ayrıca anket verilerindeki farklı sorulardan hangilerinin sonuç için en önemli olduğuna dair bir etki karşılaştırılması da gerçekleştirilip veriler en yüksek başarı oranı yakalanmak amacı ile yeniden düzenlenmiştir. Bu işlem bazı anket soruları ve yanıtlarının verilerden çıkarılması ile sağlanmıştır.

Gerçek hayatta karşılaşılabilecek sınıfı/etiketi olmayan veriler, gerçekleştirilen hesaplamalar sonucu yeniden düzenlenen veriler ve en uygun algoritma kullanılarak sınıflandırılmıştır.

Sınıflandırılan veriler excel formatında kaydedilip, bu sayede işletmelerin bu verileri farklı sorgulama formülleri ile detaylı olarak analiz edebilmesine de olanak sağlanmıştır.

Ek olarak çalışmada WEKA programında gerçekleştirilen işlemlerin büyük çoğunluğu için işleme ait komutlar ve ekran görüntülerine de yer verilmiştir. Bu sayede okuyucuların çalışmada uygulanan işlemleri teorik onlarak anlamasının yanında, bu program hakkında bilgi sahibi olmaları ve benzer çalışmaları yapabilecek temel seviyede fikir sahibi olmaları da amaçlanmıştır.

KAYNAKLAR

Ada, M.S. (2011). "İşletmelerde Müşteri İlişkileri Yönetimi Uygulamaları (MİY) ve Müşteri Bakış Açısıyla MİY Uygulamalarının Değerlendirilmesi" Yüksek Lisans Tezi, Namık Kemal Üniversitesi, Tekirdağ.

Aktepe, C., Baş, M. ve Tolon, M. (2009). Müşteri İlişkileri Yönetimi, Detay Yayıncılık, Ankara.

ALI, Faizan, Dey, Bidit Lal . ve Filieri, Raffaele (2015). An assessment of service quality and resulting. customer satisfaction in Pakistan International Airlines: Findings from foreigners and overseas Pakistani customers. International Journal of Quality & Reliability Management, 32(5), 486–502.

AMIN, Muslim., Isa, Zaidi ve Fontaine, Rodrigue (2013). Islamic banks: Contrasting the drivers of customer satisfaction on image, trust, and loyalty of Muslim and non-Muslim customers in Malaysia. International Journal of Bank Marketing, 31(2), 79–97.

ARCHANA, R. ve Subha, M. V. (2012). A study on service quality and passenger satisfaction on Indian airlines. International Journal of Multidisciplinary Research, 2(2), 50–63.

Bakırtaş, H., Yılmaz, A., Özmen, M. ve Barış, G. (2013). Müşteri İlişkileri Yönetimi, Anadolu Üniversitesi Yayınları, Eskişehir.

CHOU, Chien-Chang Liu, Li-Jen, Huang, Sue-Fen, Yih, Jeng-Ming ve Han, Tzeu-Chen (2011). An evaluation of airline service quality using the fuzzy weighted SERVQUAL method. Applied Soft Computing, 11(2), 2117–2128.

FAROOQ Muhammad Shoaib, Salam, Maimoona, Fayolle, Alain, Jaafar, Norizan ve Ayupp, Kartinaf (2018). Impact of service quality on customer satisfaction in Malaysia airlines: A PLS-SEM approach. Journal of Air Transport Management, 67, 169–180.

GURSOY, Dogan, Chen, Ming-Hsiang ve Kim, Hyun-Jeong (2005). The US airlines relative positioning based on attributes of service quality. Tourism management, 26(1), 57–67.

HUSSAIN, Rahim (2016). The mediating role of customer satisfaction: evidence from the airline industry. Asia Pacific Journal of Marketing and Logistics, 28(2), 234–255.

HUSSAIN, Rahim, Al Nasser, Amjad ve Hussain, Yomna K. (2015). Service quality and customer satisfaction of a UAE-based airline: An empirical investigation. Journal of Air Transport Management, 42, 167–175.

Odabaşı, Y. (2010). Kalıcı Başarı İçin Müşteri Hizmetleri, MediaCat Yayıncılık, İstanbul.

Okumuş, A., & Asil, H. (2007). Havayolu taşımacılığında yerli ve yabancı yolcuların memnuniyet düzeylerine göre beklentilerinin incelenmesi. Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 13, 152-175.

KIM, Koo Kim, Park, Myeong Cheol ve Jeong, Dong-Heon (2004). The effects of customer satisfaction and switching barrier on customer loyalty in Korean mobile telecommunication services. Telecommunications policy, 28(2), 145–159. KOUBAA, Yamen (2008). Country of origin, brand image.

perception, and brand image structure. Asia Pacific Journal of Marketing and Logistics, 20(2), 139–155. KUMAR, V. ve Shah, Denish (2004). Building and sustaining profitable customer loyalty for the 21st century. Journal of retailing, 80(4), 317–329.

KUO, Ying-Feng, Wu, Chi-Ming ve Deng, Wei-Jaw (2009). The relationships among service quality, perceived value, customer satisfaction, and post-purchase intention in mobile valueadded services. Computers in human behavior, 25(4), 887–896.

LIOU, James. J.H.., Hsu, Chao-Che, Yeh, Wen-Chien. ve Lin, Rong-Ho (2011). Using a modified grey relation method for improving airline service quality. Tourism Management, 32(6), 1381–1388.

SAHA, Gour C. ve Theingi. (2009). Service quality, satisfaction, and behavioural intentions: A study of low-cost airline carriers in Thailand. Managing Service Quality: An International Journal, 19(3), 350–372.

SINGH, Alok Kumar (2015). Modeling passengers' future behavioral intentions in airline industry using SEM. Journal of Advances in Management Research, 12(2), 107–127.

SUM VHAU, Vinh ve Kao, Yu-Ying (2009). Bridge over troubled water or long and winding road? Gap-5 in airline service quality performance measures. Managing Service Quality: An International Journal, 19(1), 106–134.

TIERNAN, Siobhan, Rhoades, Dawna L. ve Waguespack Jr, Blaise (2008). Airline service quality: Exploratory analysis of consumer perceptions and operational performance in the USA and EU. Managing Service Quality: An International Journal, 18(3), 212–224.

YI, Youjae ve La, Suna (2004). What influences the relationship between customer satisfaction and repurchase intention? Investigating the effects of adjusted expectations and customer loyalty. Psychology & Marketing, 21(5), 351–373.

Yayar, R., & Tekin, B. (2015). Türkiye'de ulaşım sektörünün gelişimi ve havayolu firma tercihinin belirleyicileri: İzmir ilinde bir uygulama. Ege Akademik Bakış, 15(4), 539-550.

https://erdincuzun.com/makine_ogrenmesi/k-nn-algoritmasi/(29.04.2020)

https://veribilimcisi.com/2017/07/18/karisiklik-matrisi-nedir/(01.05.2020)