TOKAT GAZİOSMANPAŞA ÜNİVERSİTESİ

2024

VERİ MADENCİLİĞİ DERSİ FİNAL PROJESİ

**BİR ÖĞRENCİ VERİTABANINDA VERİ MADENCİLİĞİ UYGULAMASI**

**Ad Soyad: Yeşim Gündüz**

**213908061**

**VERİ MADENCİLİĞİNE GİRİŞ**

[Şirket adresini yazın]

**İÇİNDEKİLER**

ÖZET………………………………………………………………………………………. 3

1.GİRİŞ……………………………………………………………………………………. 4

2. Random Forest……………………………………………………………… 5

2.1. ALGORİTMA………………………………………………………………………… 6

3. UYGULAMA…………………………………………………………………………… 6

3.1. VERİ SETİ…………………………………………………………………………….. 6

3.2. UYGULAMA YAZILIMI…………………………………………………………….. 7

3.4. VERİ MADENCİLİĞİ SÜRECİ……………………………………………………….7

3.4.1. VERİNİN HAZIRLANMASI……………………………………………………… .. 7

3.4.2. VERİ SEÇİMİ VE VERİ DÖNÜŞÜMÜ……………………………………………..8

3.4.3. VERİ MADENCİLİĞİ……………………………………………………………….. 8

3.4.4. SUNUM…………………………………………………………………………………9

4. SONUÇ…………………………………………………………………………………….10

KAYNAKLAR………………………………………………………………………………11

**ÖZET**

Veri madenciliği, değişken geniş veri setleri arasındaki önemli ilişkileri araştırmak için kullanılır.Veri madenciliği çoğunlukla, ticari uygulamalar için kullanılmaktadır.Bu çalışmada, Üniversite öğrencilerinin mevcut akademik durumlarını ve ruh sağlıklarını inceleyerek mental sağlığını ölçmek için çalışıldı.

*Anahtar Kelimeler : Veri Madenciliği,Mental Sağlık,Random Forest Algoritması*

**1. GİRİŞ**

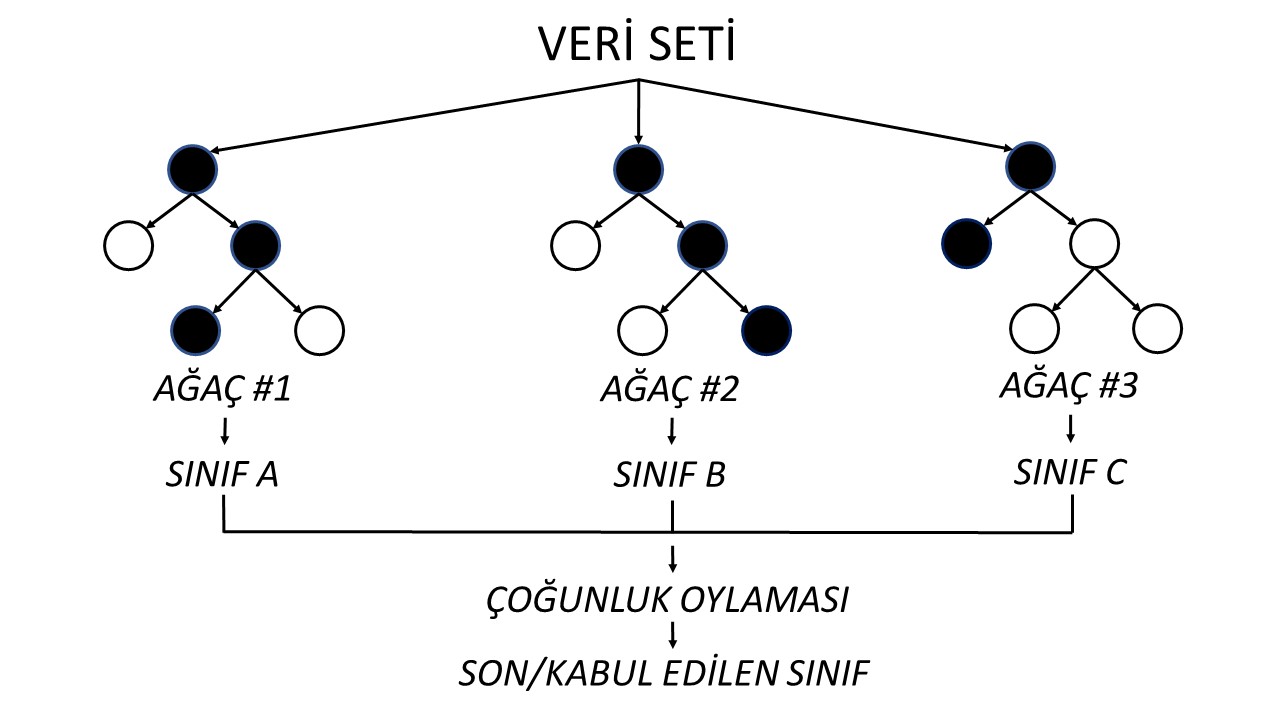
Kaggle’dan aldığım Veri setine göre öğrencilerin akademik durumlarını ve ruh sağlıklarını incelediğimizde çeşitli önemli bulgular ortaya çıkmaktadır. Öncelikle, öğrencilerin çoğunluğunun 1. sınıfta olduğunu yapmış olduğum grafiklerden yola çıkarak söyleyebiliriz. GPA (Not Ortalaması) dağılımına baktığımızda, öğrencilerin genellikle 3.00 ile 3.49 arasında bir not ortalamasına sahip olduklarını görmekteyiz.

Ruh sağlığı açısından değerlendirdiğimizde, öğrencilerin çoğunun depresyon, anksiyete ve panik atak gibi sorunlardan mağdur olmadığını görmekteyiz. Ancak, bazı öğrencilerin bu sorunlara sahip olduğunu ve hatta bazılarının hem depresyon hem de anksiyete yaşadığını gözlemliyoruz.

Akademik başarı ile ruh sağlığı arasında belirgin bir ilişki olmadığı görülmektedir. Bu da bize, akademik başarının ruh sağlığı üzerinde doğrudan bir etkiye sahip olmayabileceğini göstermektedir.

Sonuç olarak, veri setimizdeki öğrencilerin genellikle orta düzeyde akademik başarıya sahip olduğunu ve ruh sağlığı sorunlarının genel olarak düşük seviyede olduğunu söyleyebiliriz. Ancak, bireysel olarak bazı öğrencilerin ruh sağlığı açısından desteklenmeye ihtiyaçları olabileceği de göz önünde bulundurulmalıdır.

**2. Random Forest Sınıflandırıcısı**



***Random Forest*** (***Rastgele Orman***) algoritması; birden çok karar ağacı üzerinden her bir karar ağacını farklı bir gözlem örneği üzerinde eğiterek çeşitli modeller üretip, sınıflandırma oluşturmanızı sağlamaktadır.

Kullanım kolaylığı ve esnekliği; hem sınıflandırma hem de regresyon problemlerini ele aldığı için benimsenmesini ve kullanımının yaygınlaşmasını hızlandırdı.

Algoritmaya yönelik en beğenilen nokta ise; veri kümeniz üzerinde çeşitli modellerin oluşturulması ile kümenizi yeniden ve daha derin keşfetme imkanı sunmasıdır.

**2.1. ALGORİTMA**

Algoritma;

* Analiz edilecek veri seti hazırlanır,  
  (*Analiz edilecek küme oluşturulur, gerekli görülürse veri temizlemesi gerçekleştirilir.*)
* Algoritma her bir örnek için karar ağacı oluşturur ve her bir karar ağacının tahmini değer sonucu oluşur,
* Tahmin sonucu oluşan her değer için oylama gerçekleştirilir,  
  \*(*Sınıflandırma problemi için* ***Modu*** *(****Mode****), Regresyon problemi için* ***Ortalamayı*** *(****Mean****)*)
* Son olarak algoritma son tahmin için en çok oylanan değeri seçerek sonuç oluşturur.

adımları ile analiz gerçekleştirmektedir.

**3. UYGULAMA**

Üniversite öğrencilerinin mevcut akademik durumlarını ve ruh sağlıklarını inceleyerek mental sağlığını ölçmek için uygulama yapıldı.

**3.1. VERİ SETİ**

RUH SAĞLIĞININ ÖĞRENCİLERİN CGPA VERİ SETİ ÜZERİNDEKİ ETKİLERİ ÜZERİNE İSTATİSTİKSEL BİR ARAŞTIRMA  
Bu Veri seti, Üniversite öğrencilerinin mevcut akademik durumlarını ve ruh sağlıklarını incelemek amacıyla Google formları kullanılarak yapılan bir anketle toplanmıştır.

A STATISTICAL RESEARCH ON THE EFFECTS OF MENTAL HEALTH ON STUDENTS’ CGPA dataset  
This Data set was collected by a survey conducted by Google forms from University student in order to examine their current academic situation and mental health.

**3.3. UYGULAMA YAZILIMI**

Uygulamada kullanılan programlama ortamı Anaconda'dır. Anaconda, veri analizi, bilimsel hesaplama ve makine öğrenimi gibi alanlarda kullanılan bir Python dağıtımıdır.

**3.4. VERİ MADENCİLİĞİ SÜRECİ**

Veri araştırması ve sunum süreci çeşitli basamaklardan oluşmuştur. Bu basamaklar verinin hazırlanması, veri seçimi ve dönüşümü, veri madenciliği ve sunumudur.

**3.4.1. VERİNİN HAZIRLANMASI**

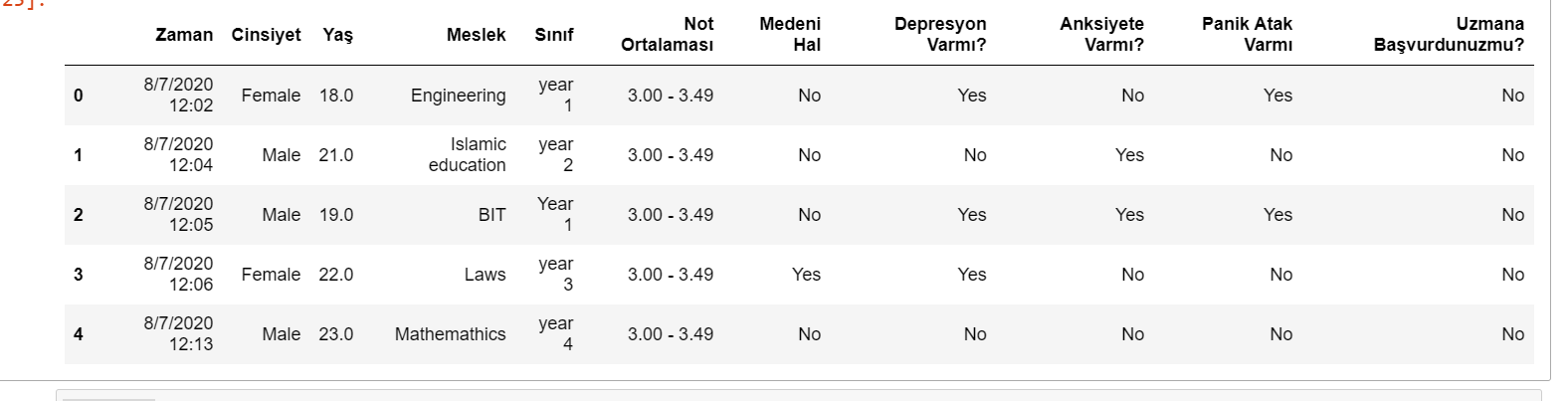
Bu projede, veri setindeki özelliklerin (yaş, cinsiyet, meslek, not ortalaması vb.) ve belirtilen hedef değişkenlerin (depresyon varlığı, panik atak varlığı vb.) algoritmaya beslenerek, bu değişkenlerin varlığını tahmin etmek için kullanıldığını göstermektedir. Hesaplama süreci, Random Forest algoritmasının içsel işleyişiyle ve veri setinin özellikleriyle ilişkilidir.

**3.4.2. VERİ SEÇİMİ VE VERİ DÖNÜŞÜMÜ**

Öncelikle, veri seti okunmuş ve eksik değerler içeren satırlar çıkarılmış (NaN değerler içeren satırlar).Daha sonra, 'Depresyon Varmı?' sütunu hedef değişken olarak belirlenmiş ve diğer sütunlar ise özellikler olarak kullanılmış.Kategorik sütunlardaki değerler LabelEncoder kullanılarak sayısal değerlere dönüştürülmüş. Bu, algoritmanın kategorik verilerle çalışmasını sağlar.Veri seti özellikler (X) ve hedef değişken (y) olarak ayrılmış. X, 'Depresyon Varmı?' sütunu hariç tüm sütunları içerirken, y sadece 'Depresyon Varmı?' sütununu içerir.Veri seti train\_test\_split işleviyle eğitim ve test setlerine ayrılmış (%80 eğitim, %20 test).Random Forest sınıflandırıcı modeli oluşturulmuş ve eğitim setiyle uyum sağlanmış.Model, test seti üzerinde değerlendirilmiş ve doğruluk skoru hesaplanmış.

**3.4.3. VERİ MADENCİLİĞİ**

Verilerimi test ve eğitim diye eğitip öğrenmesini sağladım bu sayede başka veriler geldiğinde belirli yüzde şeklinde doğru söyleyecektir.



X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

model = RandomForestClassifier(random\_state=42)

model.fit(X\_train\_imputed, y\_train)

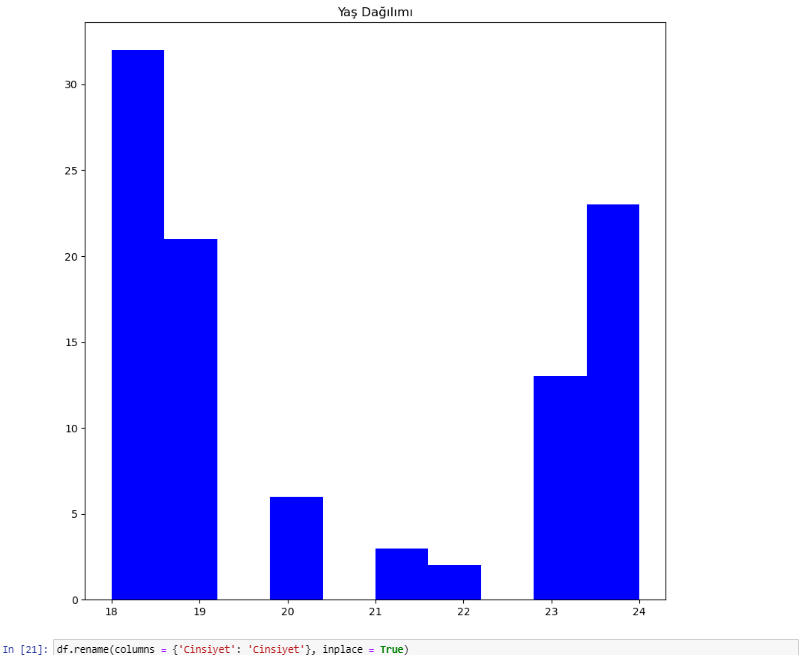
y\_pred = model.predict(X\_test\_imputed)

print("Eğitim seti doğruluk skoru:", model.score(X\_train\_imputed, y\_train))

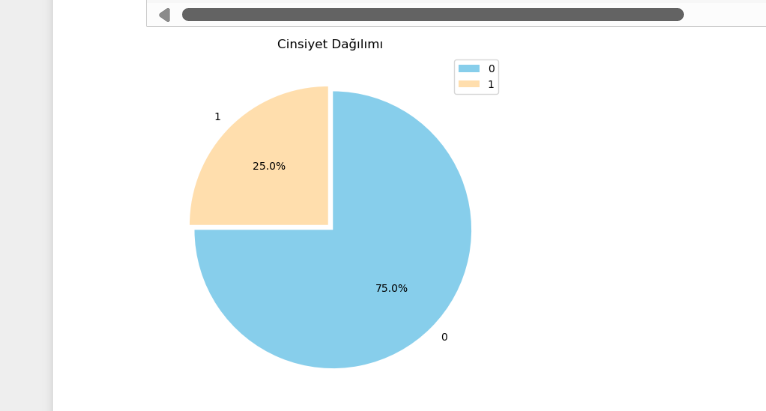
print("Test seti doğruluk skoru:", model.score(X\_test\_imputed, y\_test))

**3.4.4. SUNUM**

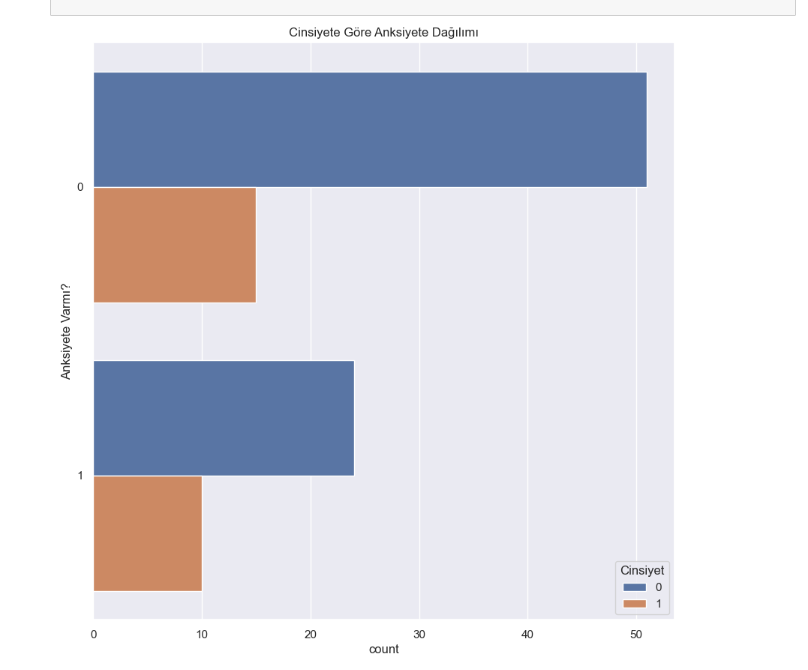
Veri madenciliği basamağının sonuçları bu basamakta verilir. Grafikler ve tablolar için Anaconda yazılım programını kullanılmıştır.



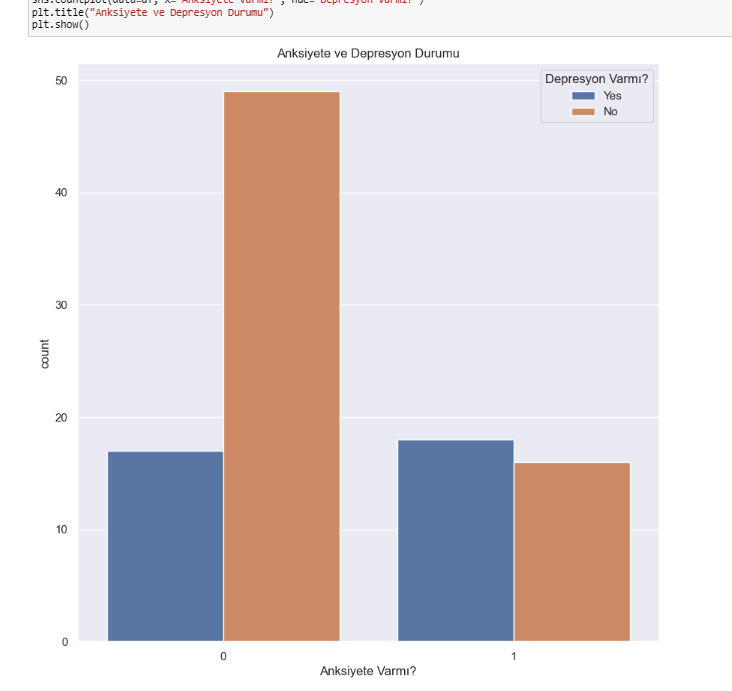
Bu tabloda öğrencilerin yaş dağılımları verilmiştir.18-19 yaş aralığı fazlayken 21-22 aralık azdır.



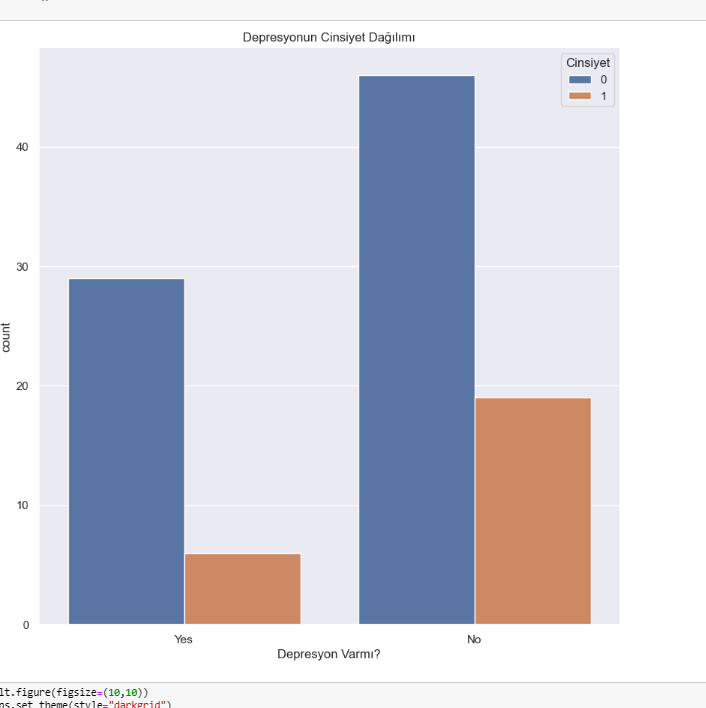
Bu grafikte erkeklerin %75 inden oluşurken kadınlar %25 kısmını oluşturuyor bu durumda çoğunluğu erkek olarak görüldü.



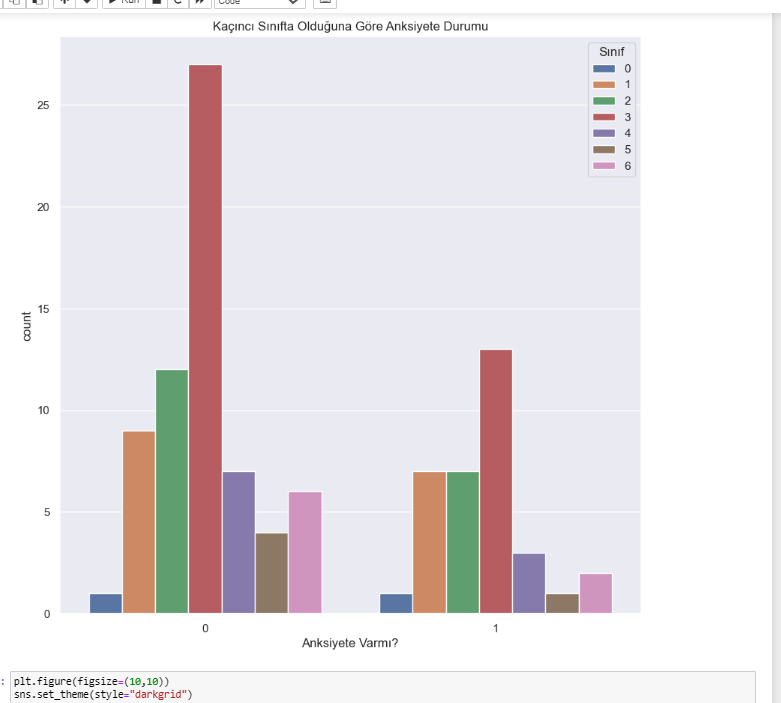
Bu grafikte cinsiyete göre anksiyete durumu incelenmiştir.Çoğunlukla erkek öğrencilerin anksiyete durumu vardır.



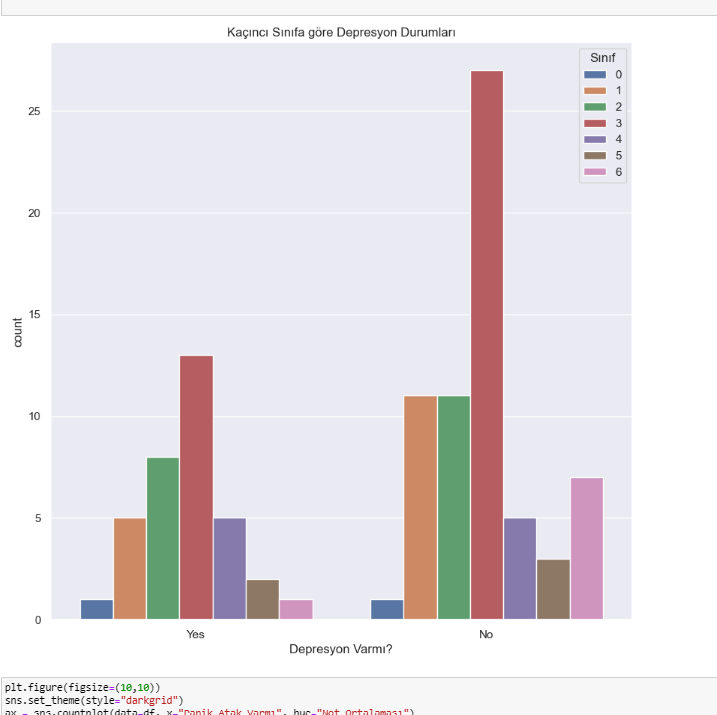
Bu grafikte hem anksiyete durumuna göre depresyon incelenmiştir.Anksiyetesi olanın çoğunlukla depresyonunda olduğu görülmüştür.



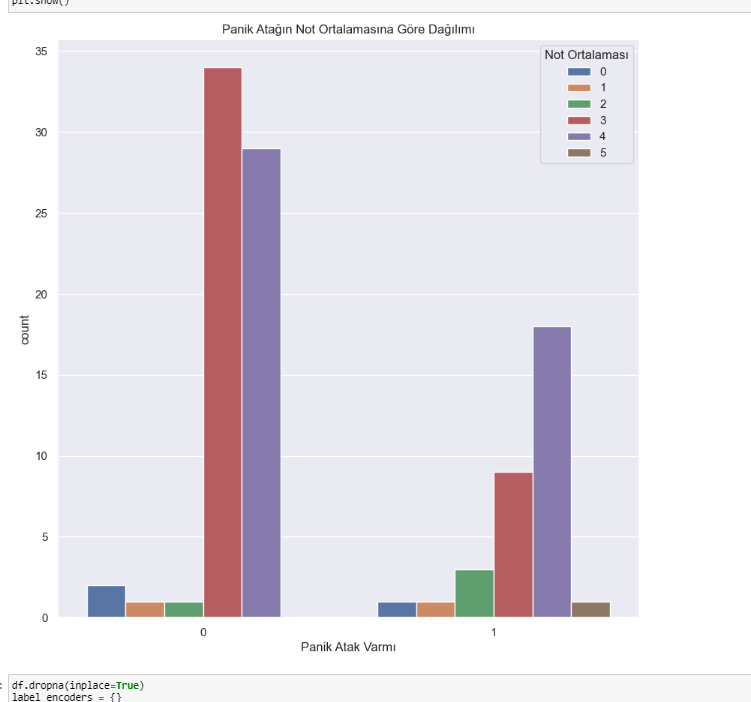
Bu grafikte cinsiyete göre depresyon dağılımı verilmiştir çoğunluğun erkek olduğu görülür.



Bu grafikte anksiyete durumunun kaçıncı sınıfta okuduğuna göre yapılmıştır.



Bu grafikte depresyon durumunun kaçıncı sınıfta okuduğuna göre yapılmıştır.



Bu grafikte panik atak durumunun kaçıncı sınıfta okuduğuna göre yapılmıştır.

**4. SONUÇ**

Bu çalışma eğitim alanında veri madenciliğinden faydalanmıştır. Veri madenciliği teknikleri Random Forest olarak analizleri kullanılmıştır. Veri madenciliği sürecinin basamakları uygulanmış ve detaylı olarak açıklanmıştır. Uygulama alanı bilinen veri madenciliği çalışmalarından farklı olarak eğitimde veri madenciliği tekniğinin kullanımı bize daha çeşitli ve önemli bulgular sağlayabilir ve bizi eğitimin kalitesinde artışa yönlendirebilir

**KAYNAKLAR**

**[1]**https://miracozturk.com/python-ile-siniflandirma-analizleri-rastgele-orman-random-forest-algoritmasi/

**[2]** https://www.kaggle.com/datasets/shariful07/student-mental-health