**Veri Madenciliği**

Proje Adı:

Öğrenci Başarı Durumu Tahmini

Emre Demir

15542023

# 1. Giriş

Günümüzde veri madenciliği eğitim kurumlarında hayati bir rol oynamaktadır ve

verilerden anlamlı bilgiler bulma amacıyla önemli araştırma alanları

büyük veri kümesinde saklanır. Eğitimsel veri madenciliği (EDM) çok önemli bir araştırma alanıdır.

Eğitimi geliştirmek için eğitim veri tabanından yararlı bilgileri tahmin etmeye yardımcı olur

Veri Madenciliği veya bilgi keşfi, yardımcı olduğu için önemi artan bir alan haline geldi.

Verileri farklı açılardan analiz etmede ve faydalı bilgilere özetlemede.

## Veri Madenciliği Nedir?

Veri Madenciliği, büyük veri kümelerinden bilgi çıkarmak olarak tanımlanır.

Veri madenciliğinin veriden bilgi madenciliği prosedürü olduğunu söylenebilir.

Veri Madenciliği, veri analizinde umut verici ve gelişen bir sınır olabilir.

Analiz sonucu birçok uygulamaya sahiptir. Veri Madenciliği aynı zamanda Bilgi olarak da adlandırılabilir.

Veri Madenciliği:

Bilgi teknolojisi, makine öğrenimi gibi alanları kapsayan multi disipliner bir alan;

İstatistik, örüntü tanıma, veri alma, sinir ağları, bilgi tabanlı sistemler, yapay

zeka ve veri görselleştirme alanlarını kapsamaktadır.

Veri madenciliği uygulaması eğitim sisteminde yaygın olarak görülmektedir.

Eğitsel Veri Madenciliği:

Eğitim alanında etkin bir şekilde uygulanabilecek gelişmekte olan bir alandır. Eğitici

Veri madenciliği, birliktelik kuralı madenciliği, sınıflandırma ve

kümeleme. ortaya çıkan bilgiler öğrencilerin yükselme oranlarını daha iyi anlamak için kullanılabilir,

Öğrencilerin kalıcılık oranı, öğrencilerin geçiş oranı ve öğrencilerin başarısı veri madenciliği sistemi kullanarak

öğrencilerin performans gelişimini ölçmek için kullanılabilir.

Öğrencilerin veri setlerini doğru bir şekilde sınıflandırmak ve analiz etmek için algoritmalar kullanılabilir.

Öğrencilerin akademik performansı, ebeveynlerin eğitimi, bulunduğu yer,

ekonomik durum, katılım, cinsiyet ve sonuca etki etmesi öngörülmüştür.

Projenin temel amacı, verileri incelemek ve analiz etmek için veri madenciliği metodolojilerini kullanmaktır.

Okul öğrencilerinin performansı. Veri madenciliği, verileri incelemek için kullanılabilecek birçok görev sağlar.

Öğrencilerin performansı ölçmek hedeflenmiştir.

Bilgi Keşfi Veri tabanı (KDD) nedir?

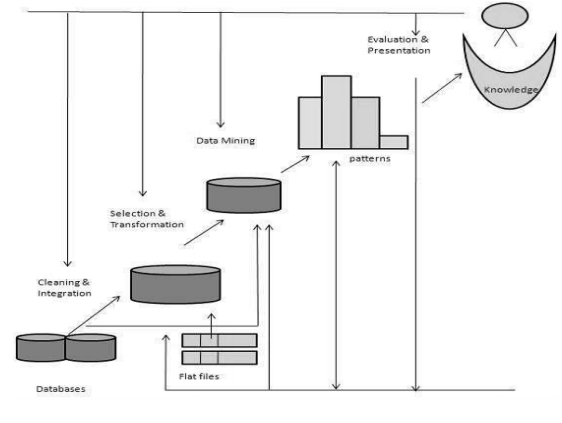
Veri tabanlarında bilgi keşfi (KDD), bir veri tabanından faydalı bilgiyi keşfetme sürecidir.

Veri toplama: Yaygın olarak kullanılan bu veri madenciliği tekniği, veri içeren bir süreçtir.

Hazırlama ve seçme, veri temizleme, veri setlerine ilişkin ön bilgileri birleştirme ve

gözlemlenen sonuçlardan doğru çözümleri yorumlama adımlarından oluşmaktadır.

İşte KDD'nin görsel üzerinde gösterimi :



1. Problem Tanımı

Veri madenciliği, bu alanda ortaya çıkan sorunları bulmak için eğitim alanında yaygın olarak kullanılmaktadır. Öğrenci performansı, çeşitli faktörlerin etkileyebileceği eğitim kurumlarında büyük bir endişe kaynağıdır.

Tahmin için gerekli üç bileşen şunlardır:

Öğrenci performansı, Veri madenciliği yöntemleri ve üçüncüsü ise veri madenciliği aracıdır. Bu Parametreler

Psikolojik, kişisel ve çevresel olabilir.

Bu faktörlerin öğrenci üzerindeki farklı etkilerini en aza indirerek enstitünün eğitim kalitesini

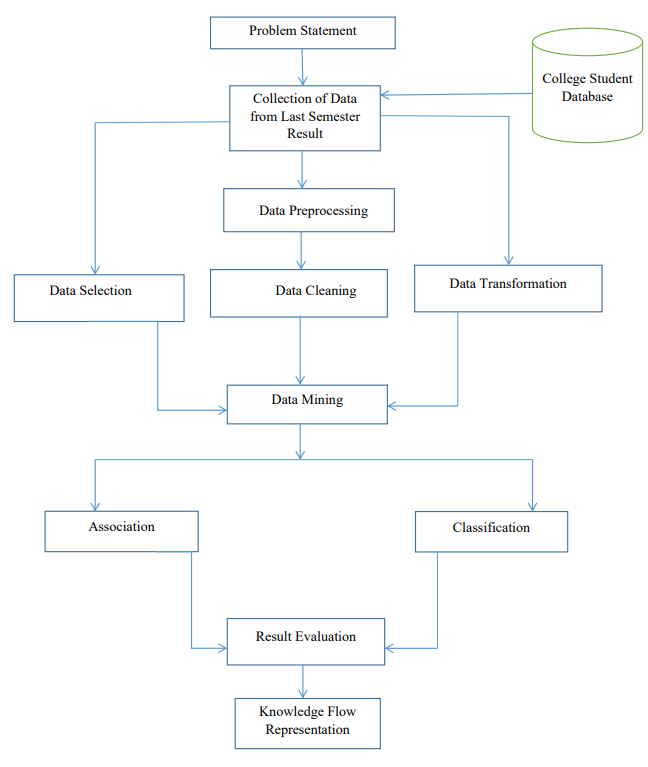
Öğrenci Performansının Tahmini Apriori uygulanarak yapılmaya çalışılmıştır.

Sınıflandırma teknikleri. Öğrenci verilerine veri madenciliği tekniklerini uygulayarak şunları yapabiliriz:

Öğrenci performansını tanımlayan bilgiyi elde edilmesi. Bu bilgi geliştirmeye yardımcı olacaktır.

Eğitim kalitesi, öğrenci performansını ve başarısızlık oranını azaltmaktır.

# İş Akış Şeması



# Gerekli Yazılım (WEKA)

Bu proje için WEKA adlı bir veri madenciliği yazılımı kullandık.

WEKA aracı şunları destekler:

Algoritmalar ve çok büyük veri kümeleri.

Görselleştirme araçları ve algoritmaları koleksiyonu. WEKA, altında yayınlanan açık kaynaklı bir yazılımdır.

GNU Genel Kamu Lisansı. Veri ön işleme, sınıflandırma, regresyon için araçlar içerir.

Kümeleme, birliktelik kuralları ve görselleştirme.

Uygulama alanları, özellikle eğitim ve araştırma için kullanılmaktadır.

Explorer arayüzü, ana bileşenlere erişebileceğimiz birçok panele sahiptir.

Görselleştirme sekmesi, mevcut çalışma ilişkisinin 2 boyutlu bir çiziminin görselleştirilmesine izin verir,

Bu çalışmada birliktelik kurallarının oluşturulması için WEKA araç seti 3.8.1 kullanılmıştır.

Sonucun tahmini.

WEKA, birkaç standart veri madenciliği görevini, daha spesifik olarak veri ön işlemeyi, kümeleme, sınıflandırma, regresyon, görselleştirme ve özellik seçimine olanak sağlar.

Teknikler, verilerin tek bir düz dosya veya ilişki olarak mevcut olduğu varsayımına dayanır.

Burada her veri noktası sabit sayıda öznitelikle (normalde sayısal veya nominal) tanımlanır.

Öznitelikler, ancak diğer bazı öznitelik türleri de desteklenir). WEKA, SQL'e erişim sağlar.

Java Veritabanı Bağlantısı kullanan veritabanları ve bir veritabanı tarafından döndürülen sonucu işleyebilir.

Sorgular:

Çok ilişkisel veri madenciliği yapamaz, ancak dönüştürmek için ayrı bir yazılım vardır.

# Tasarım ve Uygulama

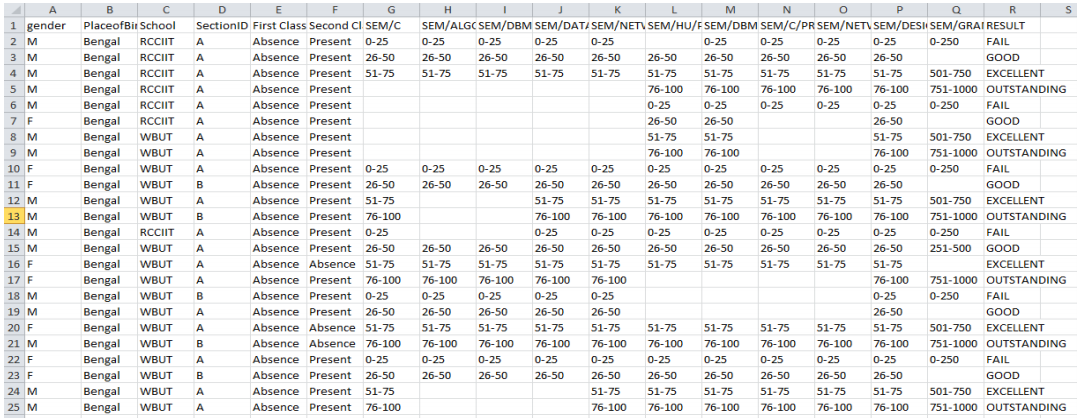
Aşağıdakiler, proje değerlendirmemizin adım adım sürecidir.

 Veri kümesi ve öznitelik seçimi- Aşağıdakileri içeren kukla bir veri kümesi topladık:

Son dönem öğrencilerinin sonuçları. Veri kümesi 507 örnek ve 18 öznitelik içerir.

Bazı eksik değerler de var. Veri dosyası "CSV" biçiminde veya "ARFF" biçiminde olmalıdır.

İşte 'CSV' formatındaki veri setimizin örneği.



Önişleme/ Veri Önişleme, bu projenin değerlendirilmesinin ilk adımıdır. Projemiz için yapacağımız WEKA Explorer ara yüzünü seçin. Burada kaynak veri dosyası yerel makineden seçilir.

Verileri Explorer'a yükledikten sonra farklı seçenekler seçerek verileri iyileştirebiliriz.

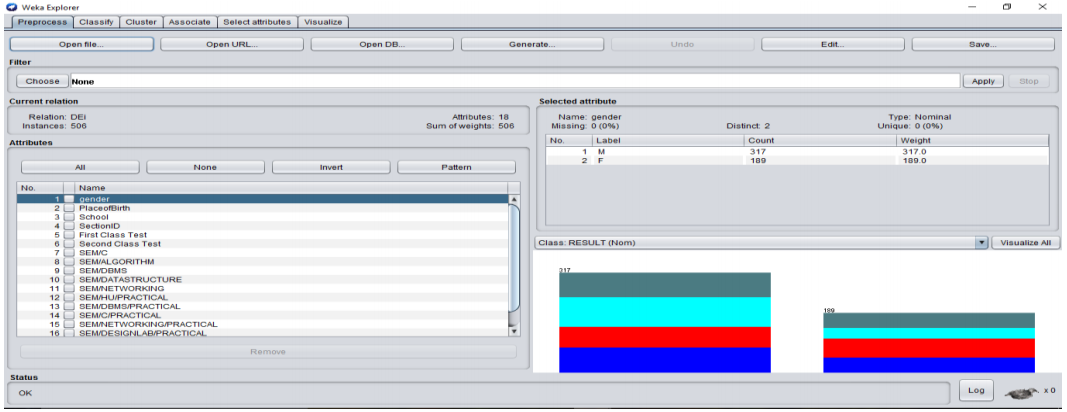
'Veri Temizleme olarak bilinen seçeneği seçebilir ve ihtiyacımıza göre öznitelikleri seçebilir veya kaldırabiliriz.

Aşağıdaki, veri setimizin ön işleme tabi tutulmuş halidir. Yukarıdaki ekranın sol tarafı gösterir

İlişki adı detayı, öznitelik sayısı ve kayıt sayısı. Sağ taraf verir

Öznitelik değerlerinin ayrıntıları, türü ve farklı değerlerin sayısı. Her birinin özellikleri

Öznitelik ekranın sağ alt kısmında görüntülenir.



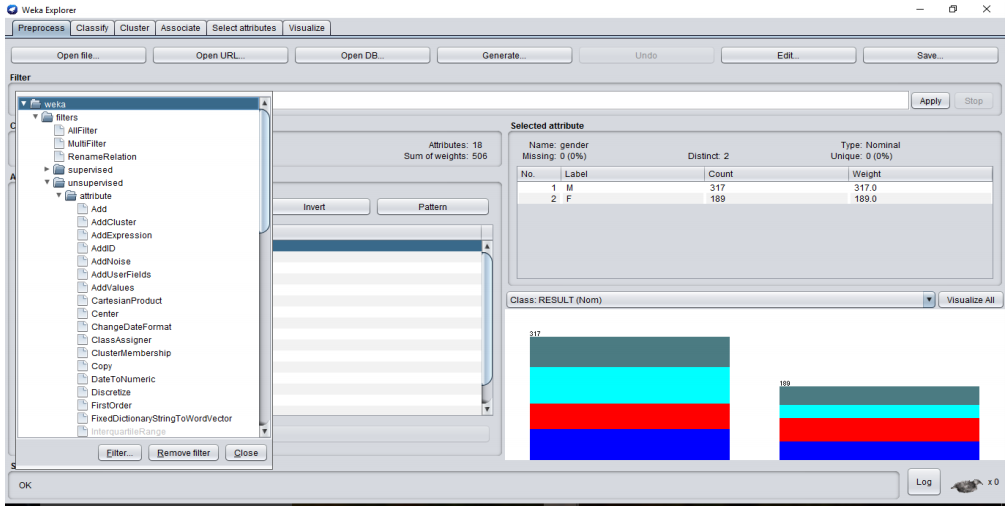
Filtreler -Ön işlem bölümü

Verileri çeşitli biçimlerde dönüştüren filtrelerin tanımlanmasına olanak tanır.

Yollar. Filtre kutusu, gerekli filtreleri ayarlamak için kullanılır. esas olarak iki tane var

Filtreler-Denetimli ve Denetimsiz kategorileri. Burada denetimsiz seçeceğiz

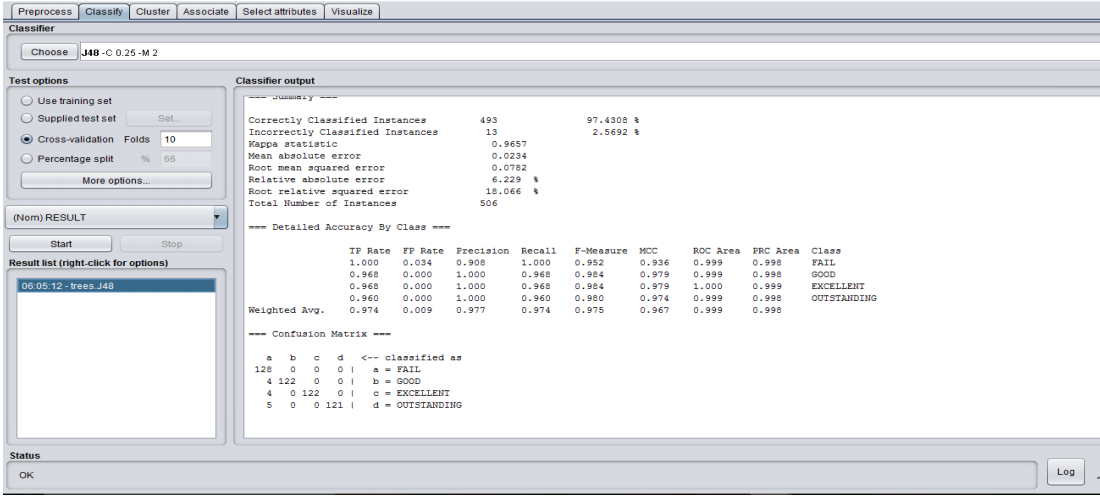
Kategori filtreleri. Veri kümesinin herhangi bir sayısal değer içermesi durumunda, gizlememiz gerekir.



Sınıflandırma - Nominal veya sayısal büyüklükleri tahmin etmek için WEKA'da sınıflandırıcılarımız var.

Tahmin amacımız için bir sınıflandırıcı seçmeliyiz. Bir standart seçtik;

sınıflandırma için J48 olarak adlandırılan sınıflandırıcı seçildi.



Yukarıdaki örnekten, J48'in doğruluğunu verdiği için iyi bir sınıflandırıcı olduğunu söyleyebiliriz.

%97,43 çünkü doğru sınıflandırılan örneklerin yüzdesine genellikle doğruluk veya

Örnek doğruluğu. Doğru ve yanlış sınıflandırılmış örnekler yüzdesini gösterir.

Doğru ve yanlış sınıflandırılmış test örnekleri. Ham sayılar gösterilir

Sınıf etiketlerini temsil eden a,b,c ve d ile karışıklık matrisi.

İşte sınıflandırıcı çıktısında bazı diğer faktörler:

 TP Oranı: gerçek pozitiflerin oranı (belirli bir sınıf olarak doğru şekilde sınıflandırılan örnekler)

 FP Oranı: yanlış pozitiflerin oranı (belirli bir sınıf olarak yanlış sınıflandırılan örnekler)

 Kesinlik: Gerçekten bir sınıfa ait örneklerin oranı, toplam örneklere bölünür.

 Geri Çağırma: Belirli bir sınıf olarak sınıflandırılan örneklerin oranı, o sınıftaki gerçek toplamla bölünür.

 F-Ölçü: 2 \* Kesinlik \* olarak hesaplanan kesinlik ve geri çağırma için birleşik bir ölçü : Geri Çağırma / (Hassas + Geri Çağırma)