

# 1 Rapor-3

## 1.1 Veri Seti Hakkında Bilgi

Veri seti, çeşitli biyometrik ölçümler ve hastalık durumu (HastalikDurumu) bilgilerinden oluşmaktadır. Toplam 768 örnek (instances) ve 9 öznitelik (attributes) içerir. Bu öznitelikler şunlardır:

Pregnancies: Gebelik sayısı (numeric)

Glucose: Plazma glikoz konsantrasyonu (numeric)

BloodPressure: Diastolik kan basıncı (mm Hg) (numeric)

SkinThickness: Triceps deri kalınlığı (mm) (numeric)

Insulin: 2 saatlik serum insülin (mu U/ml) (numeric)

BMI: Vücut kitle indeksi ( $\text{weight in kg}/(\text{height in m})^2$ ) (numeric)

DiabetesPedigreeFunction: Soy geçmişi fonksiyonu (numeric)

Age: Yaş (years) (numeric)

HastalikDurumu: Diyabet hastalığı durumu (0: Hasta değil, 1: Hasta) (nominal)

## 1.2 Veri Setini Ayırma Yöntemi

Veri seti, %70 eğitim ve %30 test veri seti olarak ayrılmıştır. Bu ayırma işlemi rastgele bir şekilde yapılmıştır. Bunun nedeni, rastgele ayırmanın veri setindeki örneklerin homojen bir şekilde eğitim ve test setlerine dağıtılmasını sağlamasıdır. Bu sayede, modelin her iki set için de genelleme yapabilme yeteneği artırılmış olur.

Gerekli kodları projenin github linkinde bulabilirsiniz:

[https://github.com/islamkosker/diabetes\\_classification](https://github.com/islamkosker/diabetes_classification)

## 1.3 Kullanılan Algoritmalar Hakkında Bilgi

J48 (Decision Tree): J48, C4.5 algoritmasının bir uygulamasıdır ve karar ağaçları oluşturmak için kullanılır. Verileri öznitelik değerlerine göre dallandırır ve yapraklar sınıf etiketlerini temsil eder. Karar ağacı, sınıflandırma kararlarını görselleştirmek için oldukça kullanışlıdır.

SMO (Support Vector Machine): SMO, Support Vector Machine (SVM) algoritmasının bir uygulamasıdır. Doğrusal veya doğrusal olmayan bir ayırıcı hiper düzlem oluşturarak veri

noktalarını sınıflandırır. SMO, ikili sınıflandırma problemlerinde kullanılır ve veri setinde sınıflar arasındaki en iyi ayrımı bulmaya çalışır.

## 1.4 Elde Edilen Sonuçlar

J48 (Decision Tree)

Doğru Sınıflandırılan Örnekler: 167 (72.29%)

Yanlış Sınıflandırılan Örnekler: 64 (27.71%)

Kappa İstatistiği: 0.3524

Ortalama Mutlak Hata (MAE): 0.3295

Kök Ortalama Kare Hata (RMSE): 0.4647

Göreceli Mutlak Hata (RAE): 72.32%

Kök Göreceli Kare Hata (RRSE): 97.09%

SMO (Support Vector Machine)

Doğru Sınıflandırılan Örnekler: 169 (73.16%)

Yanlış Sınıflandırılan Örnekler: 62 (26.84%)

Kappa İstatistiği: 0.3652

Ortalama Mutlak Hata (MAE): 0.2684

Kök Ortalama Kare Hata (RMSE): 0.5181

Göreceli Mutlak Hata (RAE): 58.91%

Kök Göreceli Kare Hata (RRSE): 108.25%

### 4. Yorumlar ve Karşılaştırmalar

Doğruluk Oranı: SMO algoritması, J48 algoritmasına göre biraz daha yüksek doğruluk oranı (%73.16 vs. %72.29) göstermektedir.

Kappa İstatistiği: SMO algoritmasının Kappa değeri, J48'e göre daha yüksek (0.3652 vs. 0.3524), bu da modelin rastgele sınıflandırmaya kıyasla daha iyi performans gösterdiğini belirtir.

Hatalar: SMO algoritması daha düşük bir ortalama mutlak hata (MAE) ve göreceli mutlak hata (RAE) değeri göstermektedir, bu da tahminlerin gerçek değerlere daha yakın olduğunu gösterir.

**Sınıflar Arası Performans:** Her iki algoritma da sınıf 0 için daha iyi performans gösterirken, sınıf 1 için performansları daha düşüktür. Bu dengesizlik, veri setinizde sınıf 0'ın daha fazla temsil edilmesinden kaynaklanıyor olabilir.

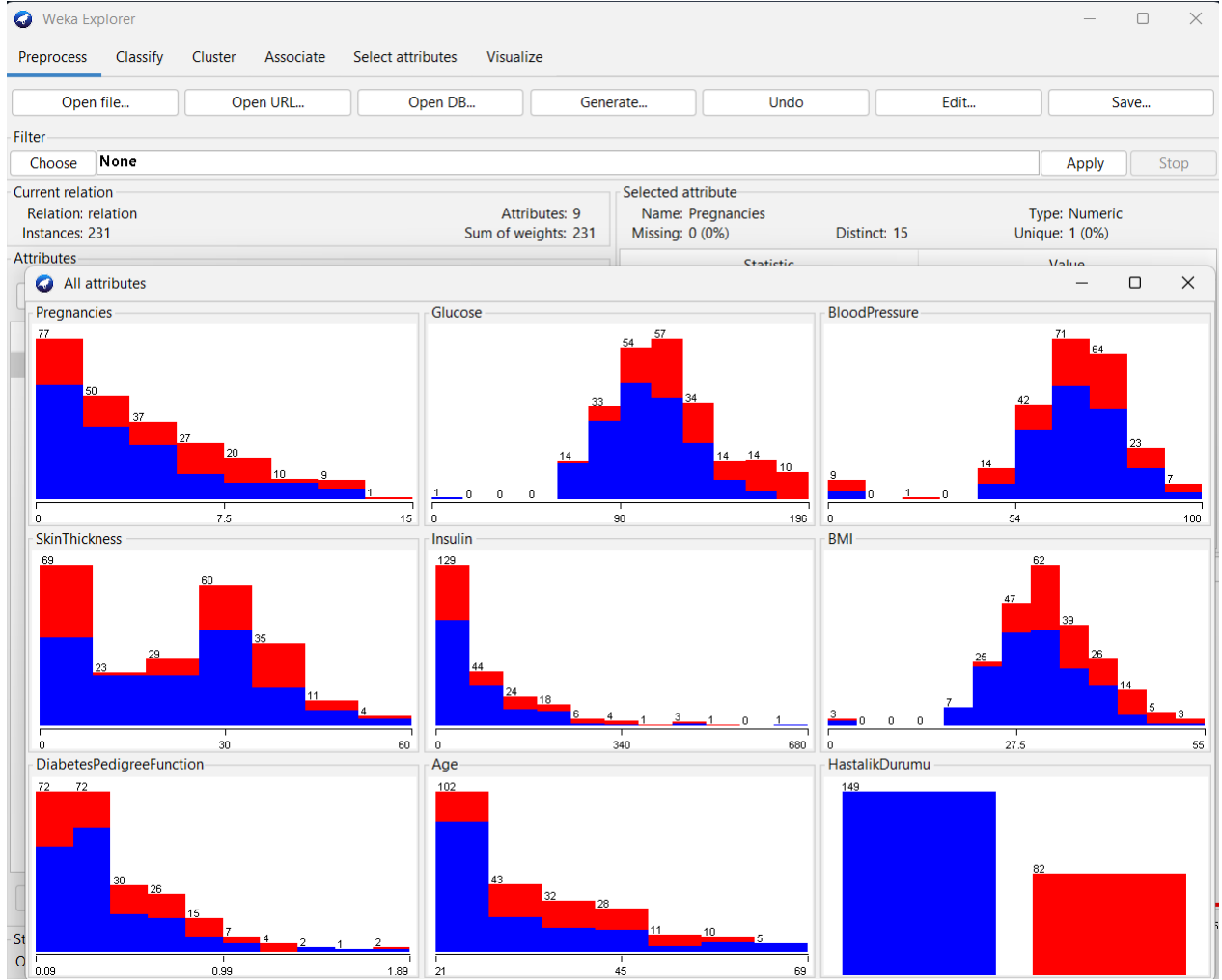
**Model Seçimi:** Genel olarak, SMO algoritması biraz daha yüksek doğruluk ve daha düşük hata oranları göstermektedir. Ancak, J48 algoritması karar ağaçlarının görselleştirilebilirliği sayesinde daha anlaşılır olabilir.

**Özet ve Tavsiyeler;**

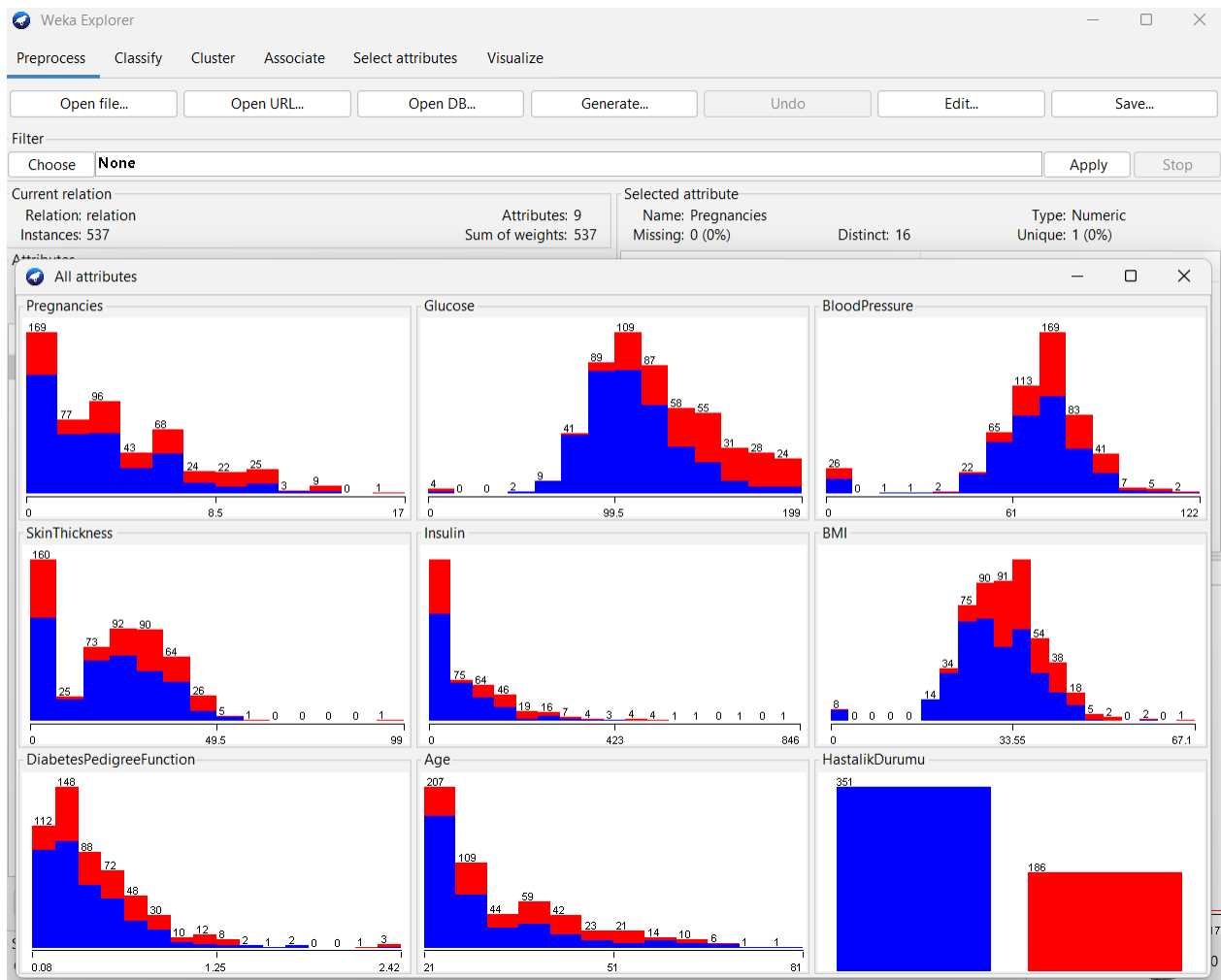
Her iki algoritmanın da güçlü ve zayıf yönleri vardır. Daha fazla veri ile model performansını artırabilirsiniz. Ayrıca, diğer makine öğrenmesi algoritmalarını deneyerek performansınızı daha da iyileştirebilirsiniz.

## 2 EKLER

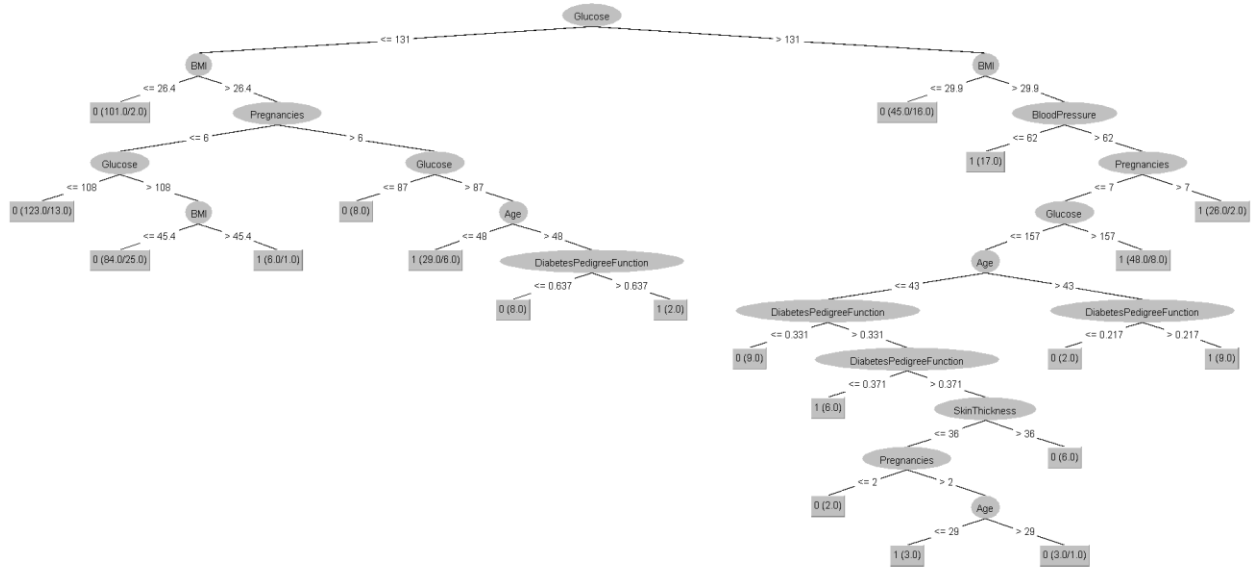
Çalışma kapsamında elde edilen ek bulgular.



EK-1 1: Test Veri Seti



EK-1 2: Train Veri Seti



EK-1 3: J48 Ağaç Yapısı