**Çalışanların İşten Ayrılma Tahmini Projesi – TUBA AYDIN**

**1. Proje Konusu**

Bilişim sektöründe çalışan personelin işten ayrılıp ayrılmayacağını tahmin etmeyi amaçlayan bu projede, çalışanlara ait çeşitli özelliklere (memnuniyet, maaş, yaş vb.) bakılarak "attrition" sınıfı (Yes/No) tahmin edilmeye çalışılmıştır.

**2. Amaç**

Şirketlerin çalışan kayıplarını önceden tespit ederek önlem almalarını sağlamak amacıyla, çalışan memnuniyeti, iş pozisyonu, maaş gibi faktörleri inceleyerek çalışanların işten ayrılma ihtimalini sınıflandırmak.

**3. Kullanılan Veri Seti**

* **Veri Kaynağı**: IBM HR Analytics Employee Attrition & Performance
* **Erişim**: UCI Machine Learning Repository veya Kaggle
* **Hedef Etiket**: İşten Ayrıldı mı ? (Evet / Hayır)
* **Kullanılan Özellikler**:
  + Age
  + MonthlyIncome
  + JobRole
  + DistanceFromHome
  + JobSatisfaction
  + WorkLifeBalance
  + Education
  + MaritalStatus
  + ve diğer demografik/performans verileri

**4. Kullanılan Algoritmalar**

Projede aşağıdaki dört sınıflandırma algoritması uygulanmıştır:

1. Logistic Regression
2. Decision Tree
3. Random Forest
4. Naive Bayes

**5. Terim Açıklamaları**

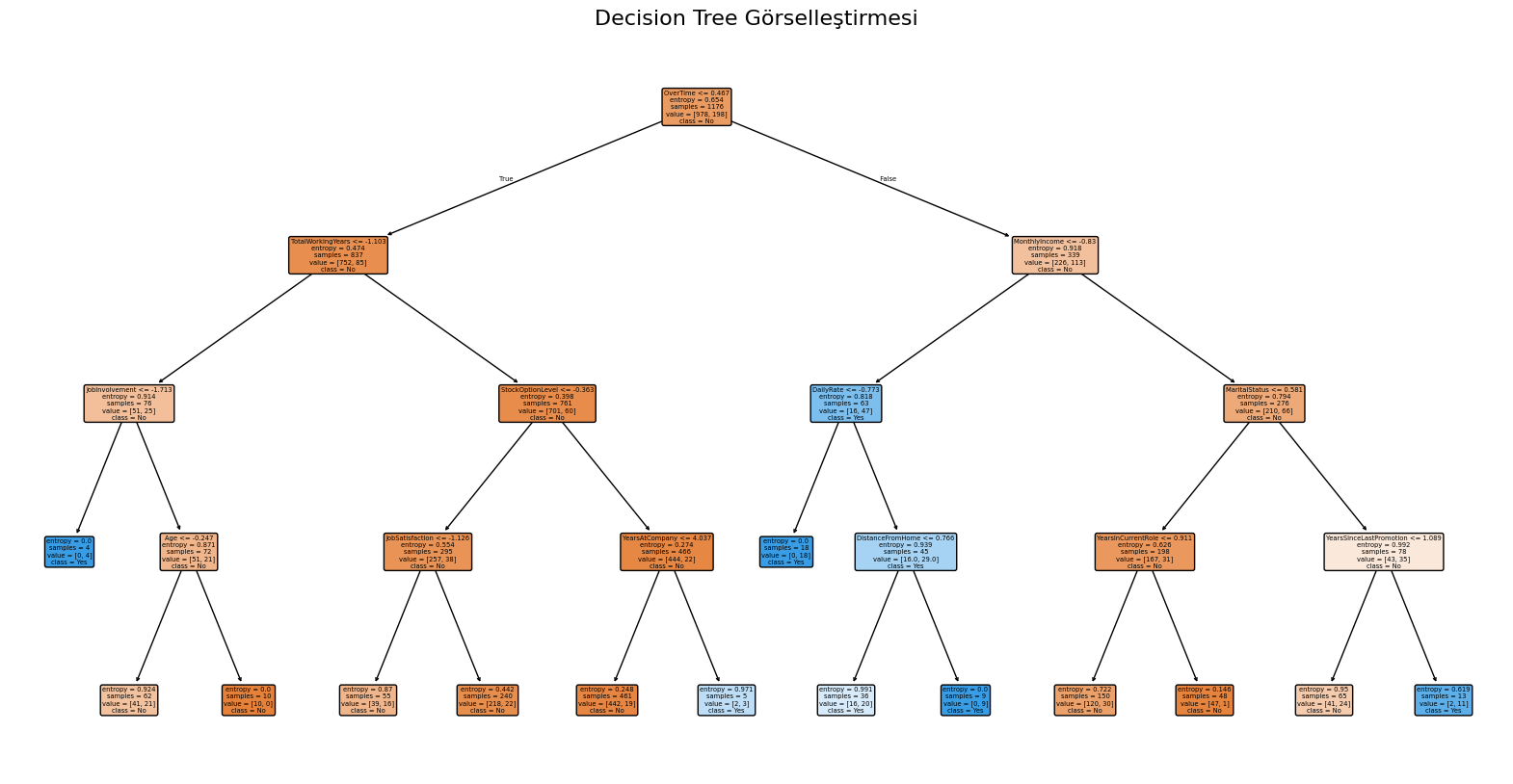
* **Accuracy**: Modelin doğru tahmin ettiği oran.
* **Precision**: Pozitif tahminlerin doğruluk oranı.
* **Recall (Sensitivity)**: Gerçek pozitiflerin ne kadarının doğru bulunduğunu gösterir.
* **F1-Score**: Precision ile Recall’un harmonik ortalaması.
* **Confusion Matrix**: Tahmin ve gerçek değerlerin karşılaştırıldığı matris.

**6. Sonuçlar ve Karşılaştırmalar**

**1. Logistic Regression**

* **Accuracy**: %89.45
* **Precision**: 0.70
* **F1 Score**:0.36
* **Recall**:0.47
* **Yorum**:
  + TN = 249: İşten ayrılmayanlar doğru tahmin edilmiş.
  + TP = 14: İşten ayrılacaklar doğru tahmin edilmiş.
  + FN = 25: Ayrılacak çalışanlar yanlışlıkla kalacak diye sınıflandırılmış.
  + FP = 6: Ayrılmayacak kişiler yanlışlıkla ayrılacak diye tahmin edilmiş.
  + **Model genel doğrulukta en iyilerden biri**. Ancak Recall = 0.36 olduğu için işten ayrılacakları bulma oranı düşüktür.

**2. Decision Tree**

* **Accuracy**: %84.69
* **Precision**: 0.20
* **F1 Score**:0.05
* **Recall**:0.08
* **Yorum**:
  + TN = 247, TP = 2, FN = 37, FP = 8
  + Pozitif sınıfın (işten ayrılanlar) tahmini oldukça başarısız.
  + Ancak **model anlaşılır ve görselleştirilebilir** olduğu için yorumlama kolaylığı sağlar.

**3. Random Forest**

* **Accuracy**: %87.75
* **Precision**: 0.80
* **F1 Score**:0.10
* **Recall**:0.18
* **Yorum**:
  + TN = 254, TP = 4, FN = 35, FP = 1
  + Ayrılmayan çalışanlar neredeyse tamamen doğru tahmin edilmiş.
  + Fakat **ayrılacak çalışanları tahmin etmede zayıf** (Recall = 0.10).
  + Dengesiz veri dağılımından etkilenmiş olabilir.

**4. Naive Bayes**

* **Accuracy**: %84.35
* **Precision**: 0.43
* **F1 Score**:0.59
* **Recall**:0.50
* **Yorum**:
  + TN = 225, TP = 23, FN = 16, FP = 30
  + Pozitif sınıfı tanıma oranı (Recall = 0.59) en yüksek modeldir.
  + **ekran görüntüsü, metin, dikdörtgen, renklilik içeren bir resim

    Yapay zeka tarafından oluşturulan içerik yanlış olabilir.Ayrılan çalışanları tahmin etmede en dengeli model**.

**7. Genel Değerlendirme**

* **En yüksek doğruluk oranı Logistic Regression (%89.45)** ile elde edilmiştir.
* Ancak **ayrılan çalışanları bulma oranı açısından (Recall)** en başarılı model Naive Bayes’tir.
* Random Forest ve Decision Tree modelleri, ayrılmayanları çok iyi tahmin ederken ayrılanları belirlemede yetersiz kalmıştır.
* **Veri setindeki dengesizlik (azınlık pozitif sınıf)** bazı algoritmaların performansını düşürmektedir.