



Data Platform
Information Technologies

Sayın kullanıcı2

2024-08-19 09:51:36 saatinde KNN Modeli üzerinde yapılan çalışmaya özel olarak hazırlanan rapor sizin için sunulmuştur. Data platform şirketimizi tercih ettiğiniz için de ayrıca teşekkür ederiz. Bu raporda, model eğitiminin başarıyla tamamlanmasının ardından oluşturulan grafikler detaylı bir şekilde incelenmiş ve analiz edilmiştir.

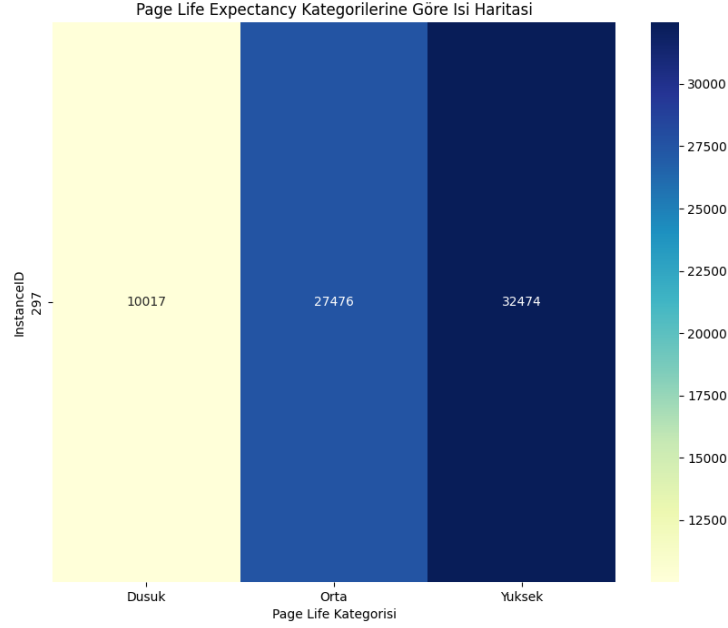
Seçtiğiniz KNN Modeli parametreleri arasında `n_neighbors:5` , `metric:euclidean` ve `weights:uniform` ,`sutun1:2` , `sutun2:7` , `test_size=0.2` ,`random_state=42` ve Hedef Etiketi `y = df['page life expectancy']` > 300 değerleri bulunmaktadır, bu parametrelerin sonuçları da raporda detaylı bir şekilde açıklanmıştır.

Focus on your business
We'll manage your databases

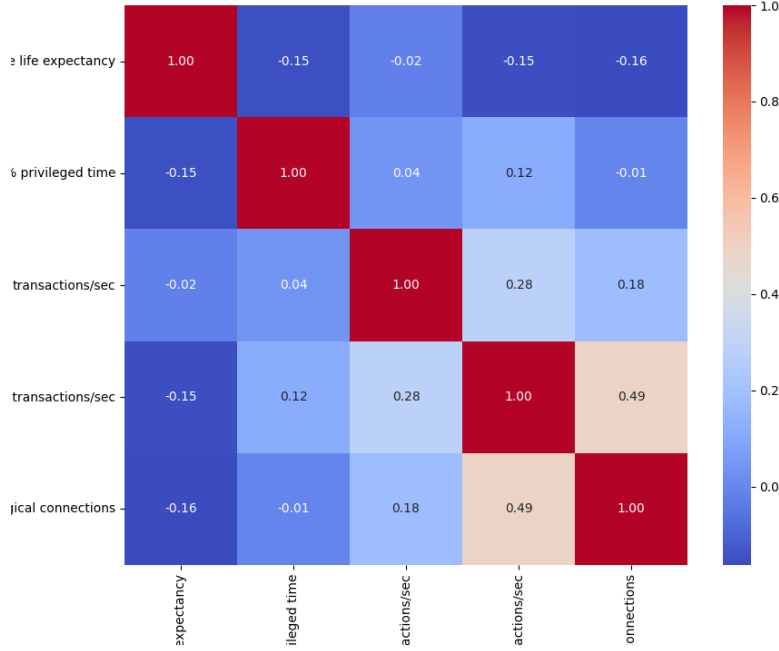
✉ info@datapatform.com.tr

📍 Data Platform Bilgi Sistemleri A.Ş. Bağlarbaşı Mah.
Irmak Sok. No: 15 D:2 K:2 Maltepe/İstanbul

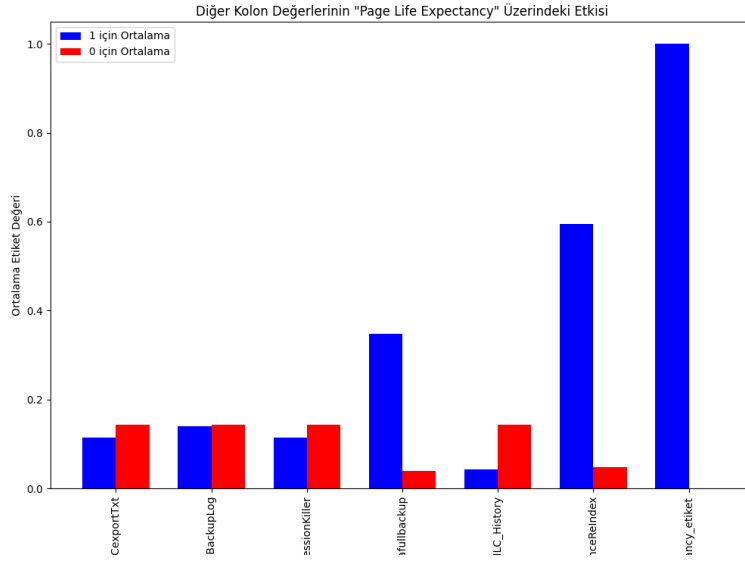
☎ 0532 320 22 39



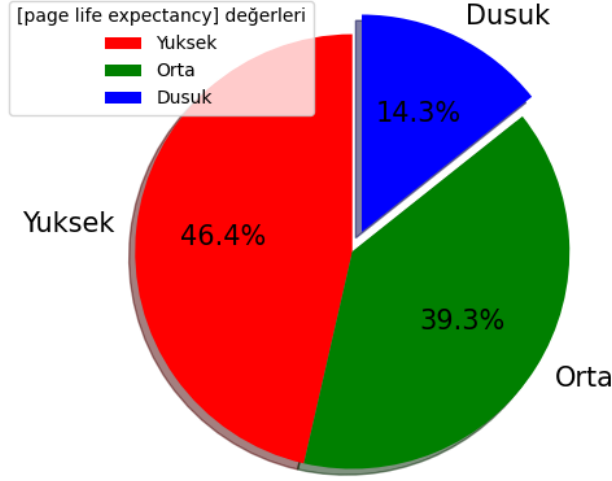
Chat-GPT tarafından üretilen analiz: "Veri setindeki gözlemler incelendiğinde, düşük page_life_category'ye sahip olan InstanceID 297'ye ait verilerin diğer kategorilere göre belirgin şekilde daha düşük olduğu gözlemlenmiştir. Orta ve yüksek kategorilerde ise sırasıyla daha yüksek değerlerin olduğu fark edilmektedir. Bu durum, InstanceID 297'nin page_life_category düzeyine göre değişen bir metriğe sahip olduğunu işaret etmektedir. Genel olarak, veri setinde page_life_category ile InstanceID arasındaki ilişkilerde belirgin bir örüntü bulunmaktadır."



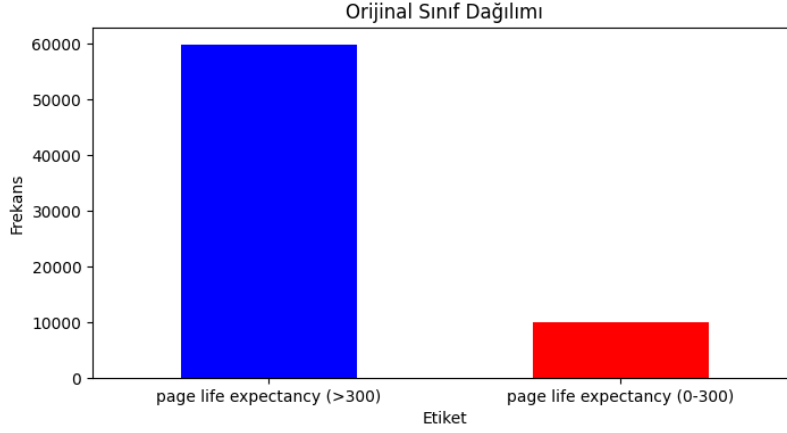
Chat-GPT tarafından üretilen analiz: "Veri setinde incelenen sütunlar arasında gözlemlenen korelasyonlara göre, 'page life expectancy' ile 'logical connections' arasında güçlü bir negatif (-0.162372) korelasyon olduğu görülmektedir. Diğer sütunlar arasında ise göreceli olarak daha zayıf veya nötr korelasyonlar bulunmaktadır. Özellikle 'write transactions/sec' ile 'logical connections' arasında güçlü bir pozitif (0.490174) korelasyon gözlenmektedir. Bu korelasyonlar, veri setindeki değişkenler arasındaki ilişkilerin incelenmesine ve değerlendirilmesine olanak sağlamaktadır."



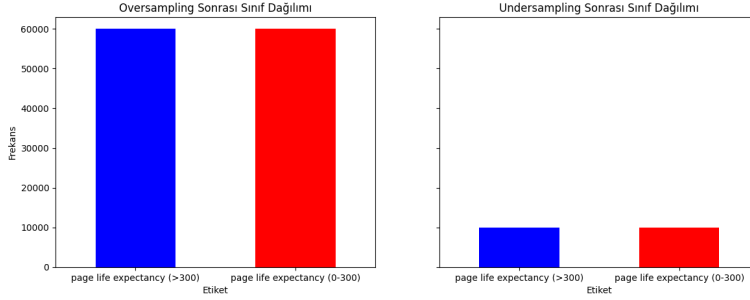
Chat-GPT tarafından üretilen analiz: 'Veri setinde belirli sütunların 'page life expectancy' sütunu üzerindeki etkileri incelendi. Ortalama değerlere göre, 'dbaHCDOCexportTxt' sütunu için 1 değeri, 'dbaBackupLog' sütunu için 1 değeri, 'dbaSessionKiller' sütunu için 1 değeri, 'dbafullbackup' sütunu için 0 değeri, 'OTS_Instrument_OHLC_History' sütunu için 1 değeri, ve 'dbaMaintenanceReIndex' sütunu için 0 değeri elde edilmiştir. Bu sonuçlara göre, 'dbaMaintenanceReIndex' sütununun 'page life expectancy' üzerinde en yüksek etkiye sahip olduğu görülmektedir. Diğer sütunlar arasında ise farklı düzeyde etkiler gözlemlenmiştir.'



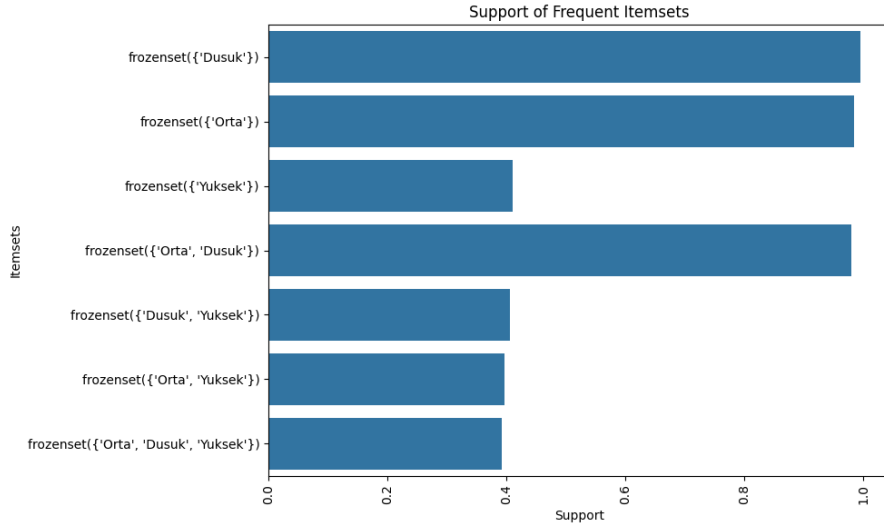
Chat-GPT tarafından üretilen analiz: "Veri setindeki sayfa yaşam süresi değerlerini kategorilere ayırarak analiz yaptım. Buna göre, 'Yüksek' kategorisinde 32,474 gözlem, 'Orta' kategorisinde 27,476 gözlem ve 'Düşük' kategorisinde 10,017 gözlem bulunmaktadır. Veri setinde genellikle yüksek sayfa yaşam süresi değerlerinin olduğu görülmektedir, orta seviyede olan gözlem sayısı da oldukça fazladır ancak düşük sayfa yaşam süresi değerlerine sahip gözlem sayısı daha azdır. Bu durum, sayfa yaşam süresi değerlerinin genel olarak yüksek olduğunu göstermektedir."



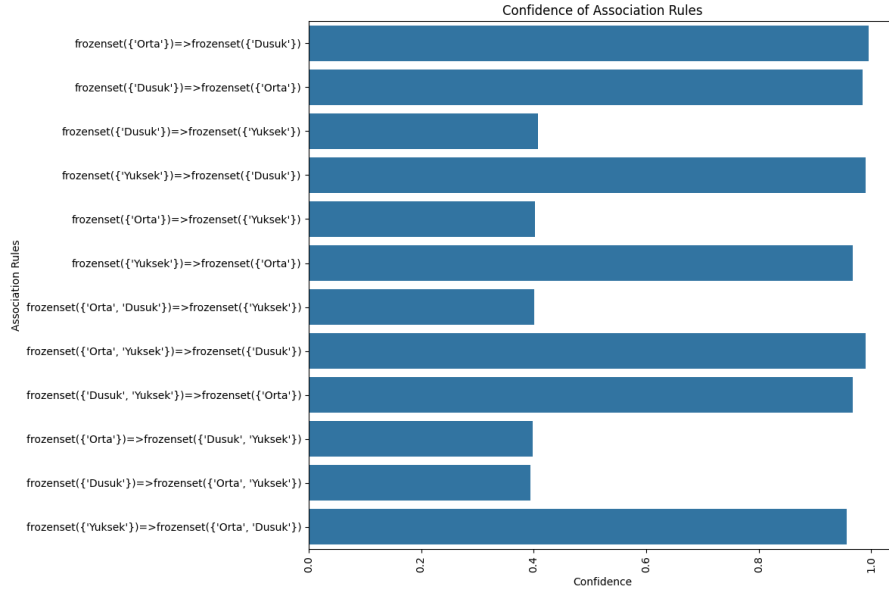
Chat-GPT tarafından üretilen analiz: 'Page life expectancy değerlerine göre yapılan etiketleme sonucunda, 0 etiketine sahip olan değerlerin sayısı 59950 iken, 1 etiketine sahip olan değerlerin sayısı ise 10017 olarak belirlenmiştir. Bu durum, veri setindeki farklı Page life expectancy değerlerinin dağılımını göstermektedir. 0 etiketine sahip olan değerlerin ortalama olarak daha yüksek olduğu düşünülebilirken, 1 etiketine sahip olan değerlerin daha düşük olduğu gözlemlenmektedir. Bu bilgiler ışığında, page life expectancy değerlerinin farklı kategorilere ayrıldığı ve bu ayrımın veri setinin genel analizine katkı sağladığı söylenebilir.'



Chat-GPT tarafından üretilen analiz: 'Bu veri seti dengesiz sınıf dağılımı sorununu ele almak için örnekleme tekniklerinin kullanıldığını gösteriyor. Oversampling yapıldıktan sonra, 0 etiketine sahip örnek sayısı 59950 iken 1 etiketine sahip örnek sayısı da 59950 olarak dengelendi. Benzer şekilde Undersampling yapıldıktan sonra, her iki sınıfa da ait örnek sayıları 10017'ye düşürüldü. Bu işlemler sonucunda sınıf dağılımı dengelendi ve her iki sınıfın da eşit sayıda örneğe sahip olduğu bir veri seti elde edildi.'

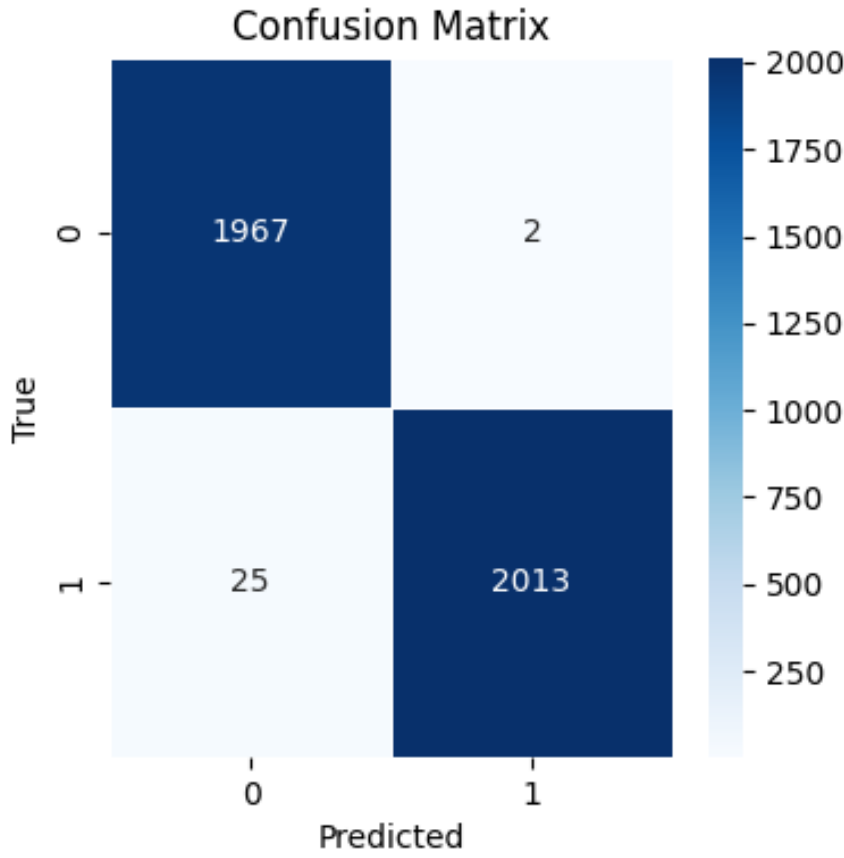


Chat-GPT tarafından üretilen analiz: "Bu veri setinde toplamda dört farklı öge bulunmaktadır: Düşük, Orta ve Yüksek. En sık görülen kombinasyonlar ise şunlardır: Orta ve Düşük birlikte %98,03 sıklıkla görülmektedir. Düşük ve Yüksek kombinasyonu %40,62 ve Orta ile Yüksek kombinasyonu ise %39,67 sıklıkla karşılaşılmaktadır. En sık görülen üçlü kombinasyon ise Orta, Düşük ve Yüksek öğelerini içeren kombinasyon olup bu kombinasyon %39,26 sıklıkla gözlenmektedir."



Chat-GPT tarafından üretilen analiz: 'Senin analizin'

Bu veri seti üzerinde yapılan FP-Growth algoritması analizine göre, orta gelire sahip olanların genellikle düşük gelir grubuna geçiş yaptığı belirlenmiştir. Aynı şekilde, düşük gelire sahip olanların yüksek gelir grubuna geçiş oranı da oldukça yüksektir. Orta gelir grubuna mensup bireylerin genellikle yüksek gelir grubuna geçiş yaptığı durumlar da mevcuttur. Ayrıca, orta ve düşük gelir gruplarına aynı anda sahip olanların genellikle yüksek gelir grubuna geçtiği tespit edilmiştir. Tüm bu verilere dayanarak, gelir seviyeleri arasındaki ilişkilerin ve geçişlerin incelendiği karmaşık bir yapı olduğu gözlemlenmiştir.



Çalışan algoritma: KNN

Chat-GPT tarafından üretilen analiz: Bu KNN algoritması, oldukça yüksek doğruluk oranları elde etmiştir. Confusion Matrix'e baktığımızda, 2004 doğru pozitif tahminleme yapılmışken sadece 25 yanlış pozitif tahmin olduğu görülmektedir. Recall değeri 0.9875 olarak oldukça yüksektir, yani modelin gerçek pozitifleri kaçırmama oranı yüksektir. Benzer şekilde, Accuracy değeri 0.9937 ve F1 Score değeri 0.9937 olarak yüksektir, yani hem doğru sınıflandırma oranı yüksektir hem de modelin precision ve recall dengeli bir şekilde performans gösterdiğini gösterir. Bu sonuçlara göre, KNN modelinin veri setini başarılı bir şekilde sınıflandırdığı söylenebilir.