

Data Platform  
Information Technologies

Sayın AisanKheiri

2024-10-02 11:19:04 saatinde SVM 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.

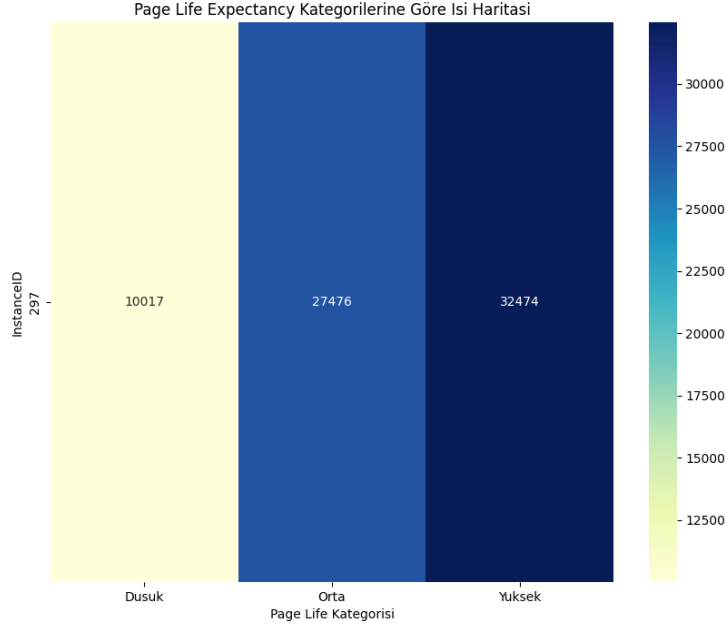
Seçtiğiniz SVM Modeli parametreleri arasında Kernel:sigmoid , C:2 , sutun1:2 , sutun2:7 , test\_size=0.2 ,random\_state=42 ve Hedef Etiket  $y = df['page\ life\ expectancy'] > 300$  değerleri bulunmaktadır, bu parametrelerin sonuçları da raporda detaylı bir şekilde açıklanmıştır.

Focus on your business  
We'll manage your databases

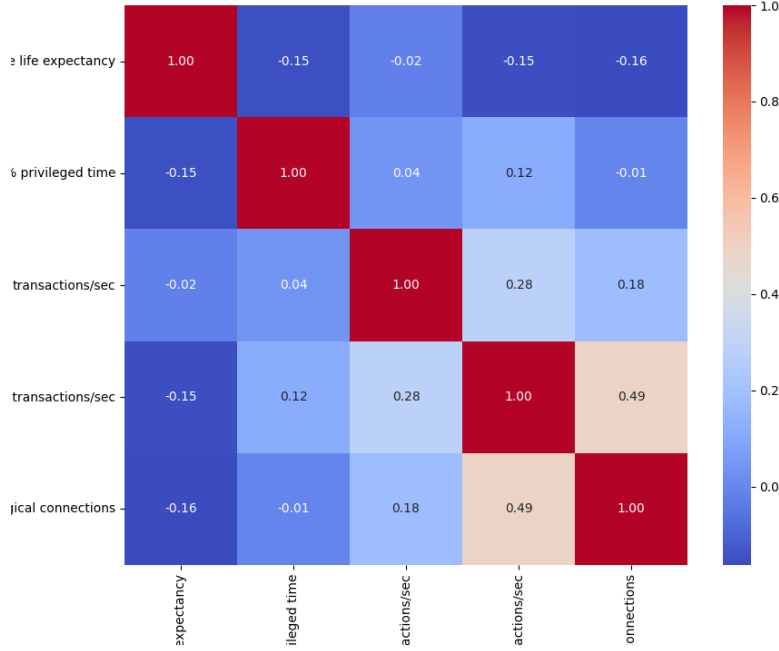
✉ info@dataplatfrom.com.tr

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

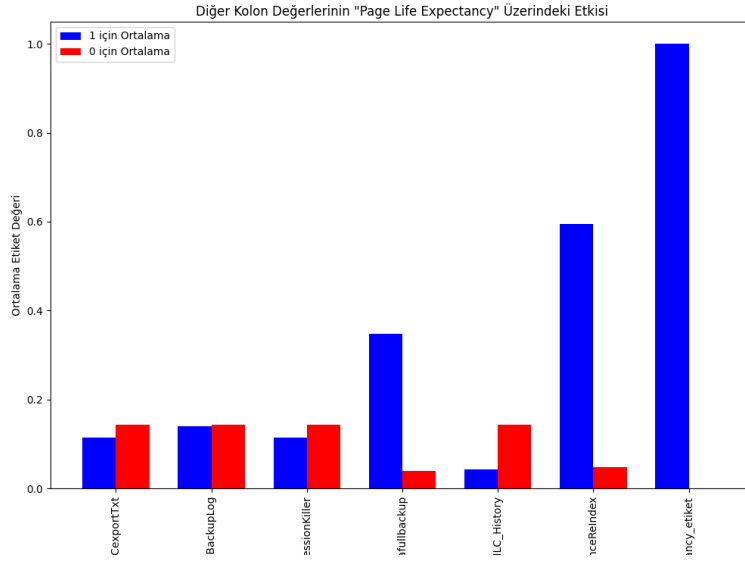
☎ 0532 320 22 39



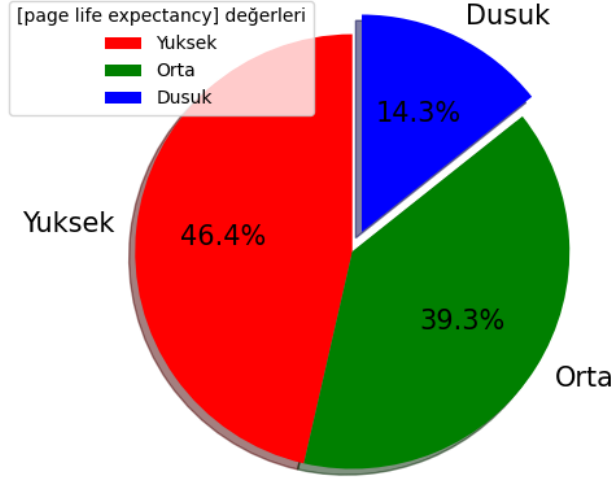
Chat-GPT tarafından üretilen analiz: "Veri setindeki sayfa yaşam kategorileri ile InstanceID 297 arasında şu örüntüyü gözlemliyorum: Düşük kategori için 10017, Orta kategori için 27476 ve Yüksek kategori için 32474 değeri görülmektedir. Bu verilere dayanarak, InstanceID 297'nin sayfa yaşam kategorileri arasında 'Yüksek' kategoride en yüksek değeri aldığını söyleyebiliriz."



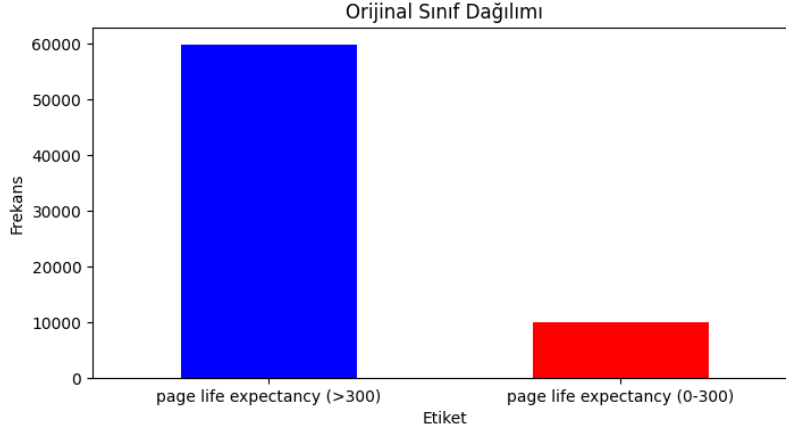
Chat-GPT tarafından üretilen analiz: Bu veri setinde page life expectancy ile diğer değişkenler arasında zayıf düzeyde negatif bir korelasyon bulunmaktadır. Özellikle write transactions/sec ve logical connections arasında güçlü bir pozitif ilişki gözlemlenmektedir. Aynı şekilde transactions/sec ile logical connections arasında da orta düzeyde pozitif bir ilişki bulunmaktadır. Ancak % privileged time ile diğer değişkenler arasında önemsiz bir ilişki tespit edilmiştir. Bu veriler incelendiğinde, sayfa ömrü beklentisi ile yazma işlemleri/saniye ve mantıksal bağlantılar arasında güçlü bir ilişki olduğu görülmektedir.



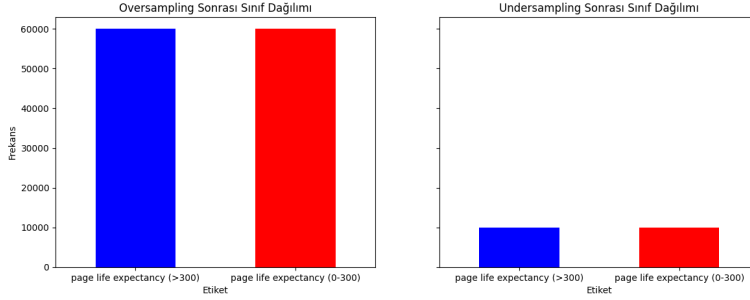
Chat-GPT tarafından üretilen analiz: Veri setindeki belirli sütunların 'page life expectancy' sütunu üzerindeki etkilerini incelemek amacıyla yapılan analizlerde, farklı sütunların bu sürece olan etkileri ortaya konmuştur. Analiz sonuçlarına bakıldığında, örnek olarak 'dbaBackupLog' sütununun 'page life expectancy' üzerindeki etkisinin 0.139045 olduğu gözlemlenmiştir. Diğer bir örnek olarak 'OTS\_Instrument\_OHLC\_History' sütununun ise bu etkiye olan katkısının oldukça düşük olduğu, yani 0.041841 oranında olduğu tespit edilmiştir. Ayrıca, 'dbaMaintenanceReIndex' sütununun 'page life expectancy' üzerindeki etkisinin oldukça yüksek olduğu, 0.595573 olarak hesaplandığı görülmektedir. Son olarak, 'page\_life\_expectancy\_etiket' sütununun bu analizdeki en belirgin etkiye sahip sütun olduğu, oranının da 1.0 olduğu dikkat çekmektedir. Bu analiz sonuçları, farklı sütunların 'page life expectancy' üzerindeki etkileri konusunda önemli ipuçları sunmaktadır.



Chat-GPT tarafından üretilen analiz: "Page life expectancy değerlerine göre kategorilere ayrıldığında yüksek kategoride en fazla gözlem sayısına sahipken, orta ve düşük kategorilerinde ise daha az gözlem sayısı olduğu görülmektedir. Bu da genel olarak page life expectancy değerlerinin yaygın olarak yüksek olduğunu göstermektedir."



Chat-GPT tarafından üretilen analiz: "Veri setindeki sayfa yaşam süresi (page life expectancy) değerlerine göre yapılan etiketlemede elde edilen sonuçlara göre, 59950 sayfa için etiket 0 ve 10017 sayfa için etiket 1 belirlenmiştir. Bu durum, veri setindeki sayfa yaşam süresi değerlerinin geniş bir dağılıma sahip olduğunu göstermektedir. Düşük etikete sahip olan sayfaların daha uzun bir yaşam süresine sahip olduğu, yüksek etikete sahip olan sayfaların ise daha kısa bir yaşam süresine sahip olduğu görülmektedir. Bu bilgiye dayanarak, sayfa yaşam süresi ile ilgili farklı kümelenmeler ve trendler gözlemlenebilir."

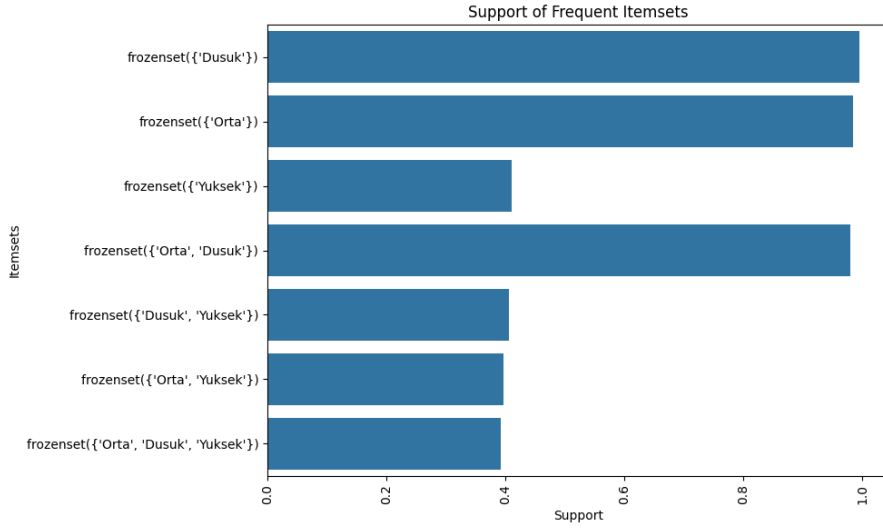


Chat-GPT tarafından üretilen analiz: "Senin analizin"

Oversampling yapıldıktan sonra elde edilen verilere göre, her iki sınıf da eşit sayıda örnek içermektedir. Bu durum, modelin daha dengeli bir şekilde eğitilmesine olanak tanırken, azınlık sınıfının daha iyi temsil edilmesini sağlayabilir.

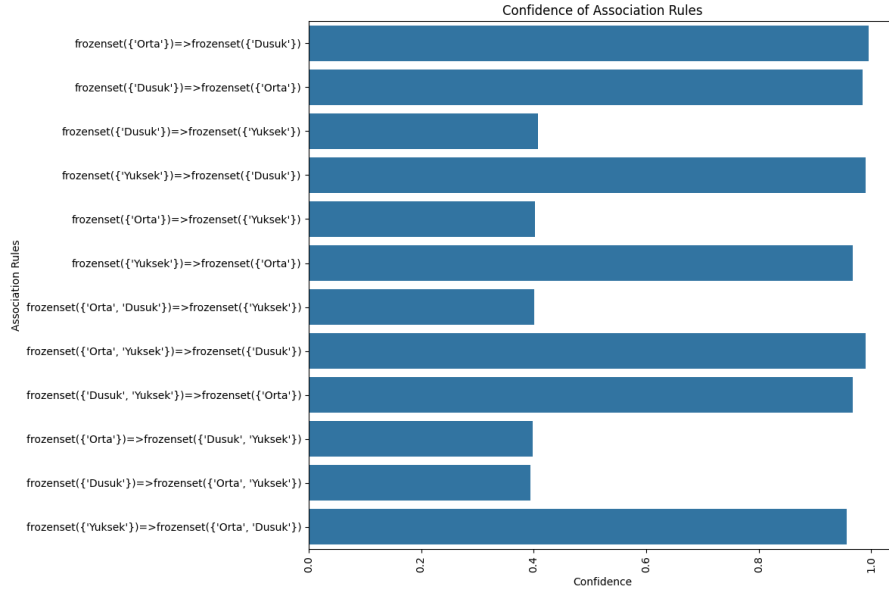
Undersampling yapıldıktan sonra elde edilen verilerde ise her iki sınıfın örnek sayısının azaltıldığı görülmektedir. Bu durum, modelin hızlı eğitimini sağlayabilir ancak azınlık sınıfının temsilini olumsuz yönde etkileyebilir.

Hangi örnekleme tekniğinin tercih edilmesi gerektiği, veri setine ve problem domainine bağlı olarak değişebilir. Bunun için farklı denemeler yaparak hangi yöntemin model performansını arttırdığını belirlemek önemlidir.



Chat-GPT tarafından üretilen analiz: Bu veri setine göre gözlemlenen sık görülen kombinasyonlar incelendiğinde, en yaygın kombinasyonun (Dusuk) seviyesi olduğu görülmektedir. (Dusuk) seviyesi tek başına %99.55 sıklıkla görülmekte iken, (Yuksek) seviyesinin görülme sıklığı %41.03 olarak tespit edilmiştir. En sık görülen kombinasyonlar ise sırasıyla şunlardır: (Orta, Dusuk) kombinasyonu %98.03, (Dusuk, Yuksek) kombinasyonu %40.62 ve (Orta, Yuksek) kombinasyonu ise %39.67 sıklıkla görülmektedir. En nadir kombinasyon ise (Orta, Dusuk, Yuksek) olup %39.26 sıklıkla görülmektedir. Bu verilere göre, genel trend (Dusuk) seviyesinin daha yaygın olduğunu ve (Yuksek) seviyesinin ise diğerlerine nazaran daha az görüldüğünü göstermektedir.





### Chat-GPT tarafından üretilen analiz: "Senin analizin"

- Orta seviye gelire sahip olan müşteriler genellikle düşük seviyede olan fiyatları tercih ediyor. Bu durum, Orta ve Düşük gelir seviyeleri arasında güçlü bir negatif ilişki olduğunu gösteriyor.
- Düşük seviye gelire sahip müşterilerin yüksek seviye fiyatları tercih etme olasılığı oldukça düşük. Düşük ve Yüksek gelir seviyeleri arasında negatif bir ilişki olduğunu gösteriyor.
- Orta seviye gelire sahip müşterilerin yüksek seviye fiyatları tercih etme olasılığı, Düşük seviyedekinden biraz daha yüksek olsa da yine de düşüktür. Orta ve Yüksek gelir seviyeleri arasında negatif bir ilişki olduğunu gösteriyor.
- Düşük seviye gelire sahip müşterilerin ve yüksek seviye gelire sahip müşterilerin, fiyat tercihleri arasında güçlü bir negatif ilişki bulunmaktadır.
- Orta seviye gelire sahip müşterilerin, Düşük ve Yüksek seviyelerdeki fiyatları bir arada tercih etme olasılığı oldukça düşüktür.

- Orta ve Yüksek seviye gelir grupları arasında tercih edilen fiyatlar arasında negatif bir ilişki olduğunu gösteriyor.
- Düşük ve Yüksek seviye gelir grupları, Orta seviye fiyatları bir arada tercih etme olasılığı oldukça düşüktür.

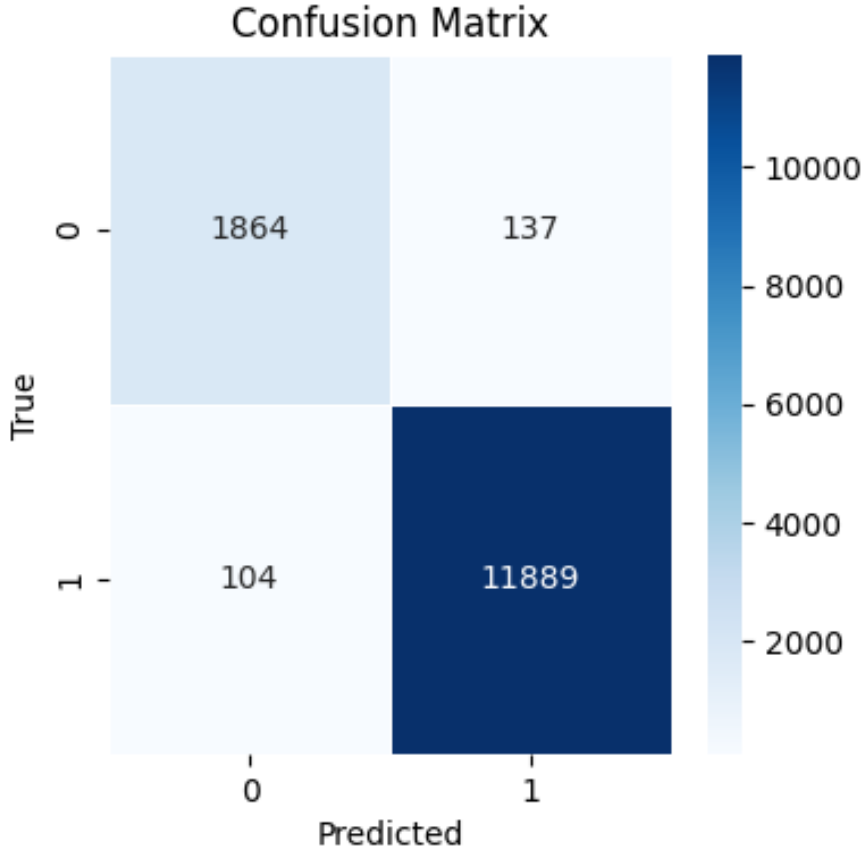
Bu analizler, müşterilerin gelir seviyeleri ile ürün fiyatlarının tercih edilme olasılıkları arasındaki ilişkileri anlamamıza yardımcı olabilir.

Focus on your business  
We'll manage your databases

✉ info@dataplatfrom.com.tr

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

☎ 0532 320 22 39



Çalışan algoritma: SVM

Chat-GPT tarafından üretilen analiz: Bu SVM makine algoritması, toplam 13994 örnekleme üzerinde değerlendirildi. Algoritmanın performansı oldukça yüksek. Confusion Matrix'e göre, modelin 1864 doğru negatif tahmini ve 11889 doğru pozitif tahmini vardır. Recall (duyarlılık) metriği 0.991 olarak yüksektir, yani gerçek pozitiflerin yaklaşık %99'una doğru bir şekilde tahmin yapmaktadır. Accuracy (doğruluk) değeri ise 0.982 olarak oldukça sağlamdır. F1 Score metriği ise 0.99 olarak yüksek bir değer gösterir, bu da modelin hem hassaslık hem de geri çağırma performansının dengeli olduğunu gösterir. Precision (kesinlik) değerlerine baktığımızda, 'False' sınıfı için 0.947 ve 'True' sınıfı için 0.989 olarak oldukça iyi sonuçlar elde edilmiştir. Sonuç olarak, SVM modeli veri setini başarılı bir şekilde sınıflandırmış ve yüksek performans göstermiştir. Bu sonuçlar, modelin güvenilirliğini ve etkinliğini doğrulamaktadır.