

Data Platform
Information Technologies

Sayın AisanKheiri

2024-10-03 10:25:37 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.

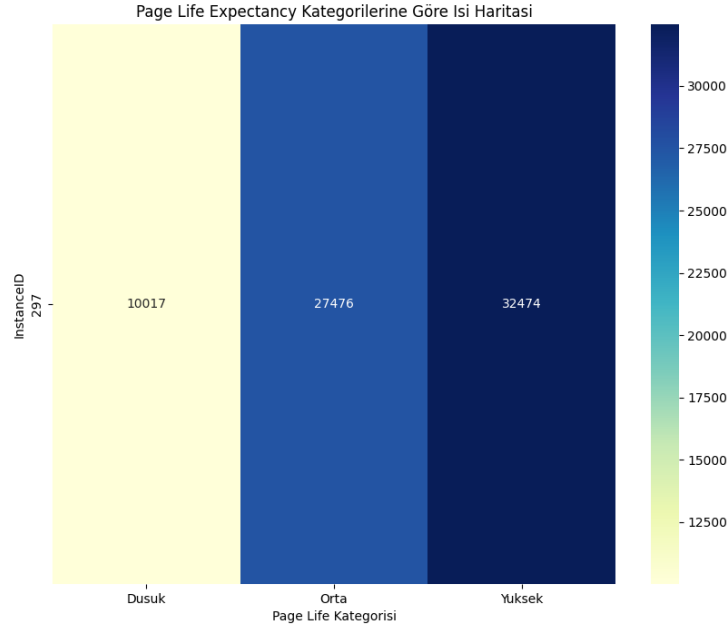
Seçtiğiniz SVM Modeli parametreleri arasında Kernel:sigmoid , C:2 , sutun1:2 , sutun2:7 , test_size=0.2 ,random_state=42 ve Hedef Etiketi $y = df['page\ life\ expectancy'] > 300$ değerleri bulunmaktadır, bu parametrelerin sonuçları da raporda detaylı bir şekilde açıklanmıştır.

Focus on your business
We'll manage your databases

✉ info@datapplatform.com.tr

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

☎ 0532 320 22 39

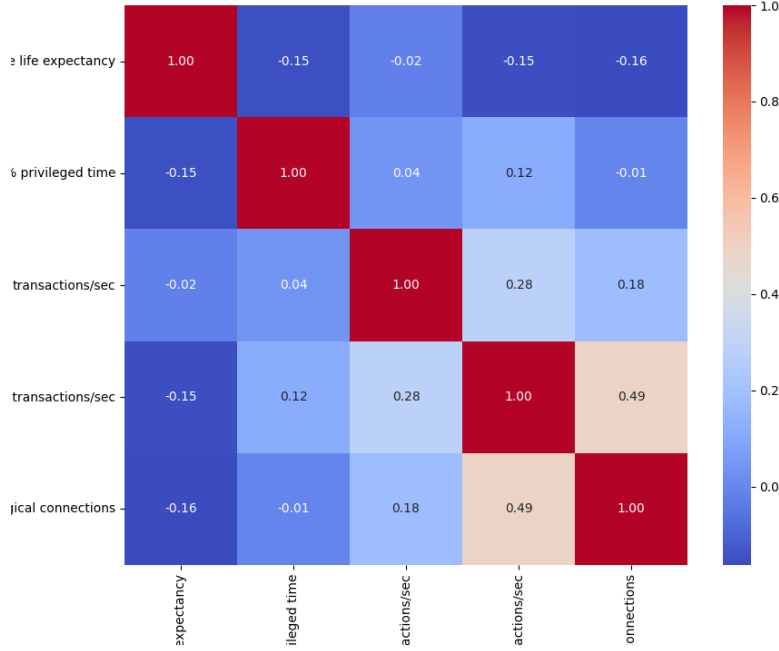


page_life_category Dusuk Orta Yuksek

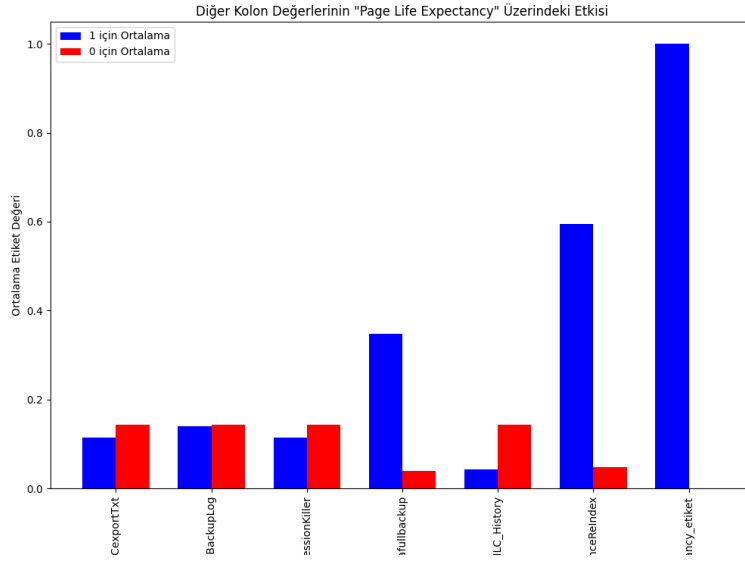
InstanceID

297 10017 27476 32474

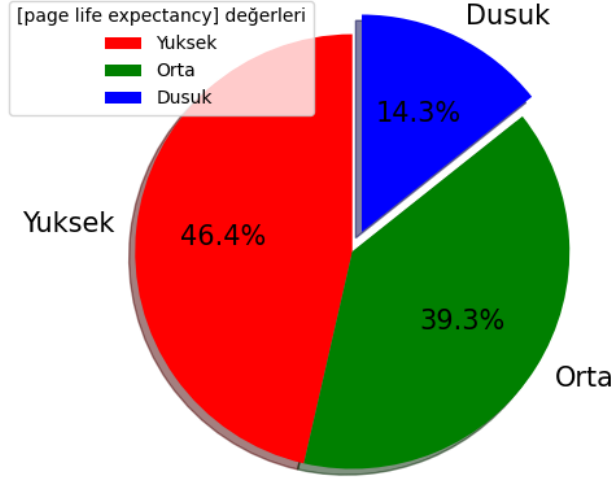
Chat-GPT tarafından üretilen analiz: "Bu veri setinde, page_life_category değişkeni ile InstanceID'ler arasındaki ilişki incelendiğinde, Düşük kategorisine ait InstanceID'lerin 10,017 değerine, Orta kategorisine ait InstanceID'lerin 27,476 değerine ve Yüksek kategorisine ait InstanceID'lerin ise 32,474 değerine sahip olduğu görülmektedir."



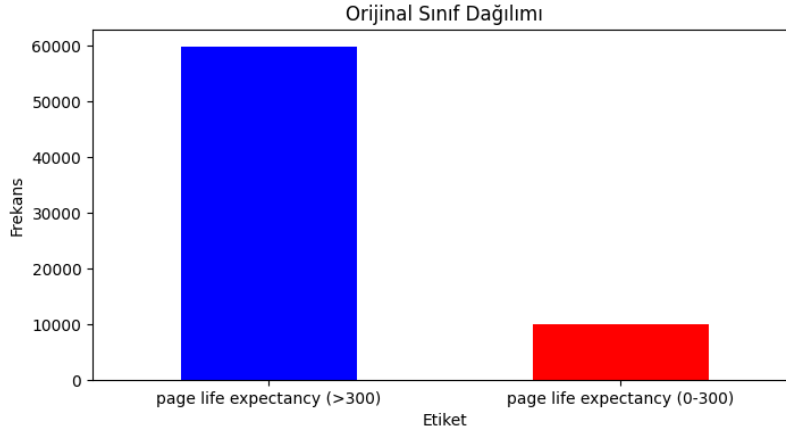
Chat-GPT tarafından üretilen analiz: 'page life expectancy ile logical connections arasında güçlü bir negatif korelasyon mevcutken, % privileged time ile herhangi bir korelasyon bulunamamıştır. Ayrıca, write transactions/sec ile logical connections arasında güçlü bir pozitif korelasyon olduğu gözlemlenmiştir. transactions/sec ve logical connections arasındaki korelasyon ise zayıf bir şekilde pozitif yönlüdür.'



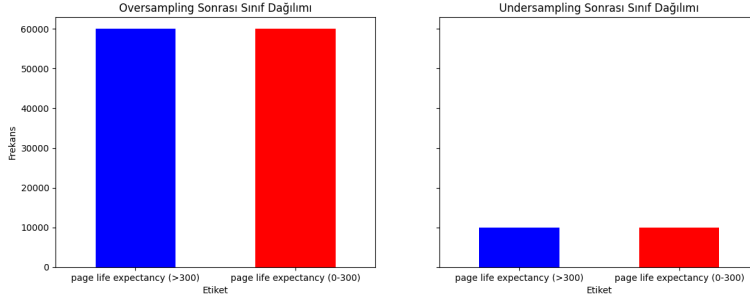
Chat-GPT tarafından üretilen analiz: 'page life expectancy' sütunu üzerinde inceleme yapılırken, farklı sütunların bu değeri etkileyip etkilemediği değerlendirilmiştir. Sonuçlara bakıldığında, 'dbaHCDOCexportTxt' ve 'dbaSessionKiller' sütunları 'page life expectancy' üzerinde benzer bir etkiye sahipken, 'dbaBackupLog' ve 'OTS_Instrument_OHLC_History' sütunlarının etkileri birbirine yakın görünmektedir. Öte yandan, 'dbafullbackup' ve 'dbaMaintenanceReIndex' sütunlarının 'page life expectancy' üzerindeki etkileri diğerlerine göre belirgin bir şekilde farklılık göstermektedir. 'dbaMaintenanceReIndex' sütunu kullanıldığında, 'page life expectancy' değeri oldukça yüksek bir ortalama gösterirken, 'dbafullbackup' sütununun kullanılması 'page life expectancy' üzerinde düşük bir etkiye işaret etmektedir. Son olarak, 'page_life_expectancy_etiket' sütunu incelendiğinde, bu sütunun 'page life expectancy' değerini en fazla etkileyen sütun olduğu görülmektedir.



Chat-GPT tarafından üretilen analiz: 'Yüksek page life expectancy değerine sahip olan gözlem sayısı 32474 iken, orta seviyede bulunan gözlem sayısı 27476 ve düşük seviyede olan gözlem sayısı 10017'dir. Bu verilere göre, veri setindeki page life expectancy değerlerinin çoğunlukla yüksek seviyede olduğu görülmektedir.'



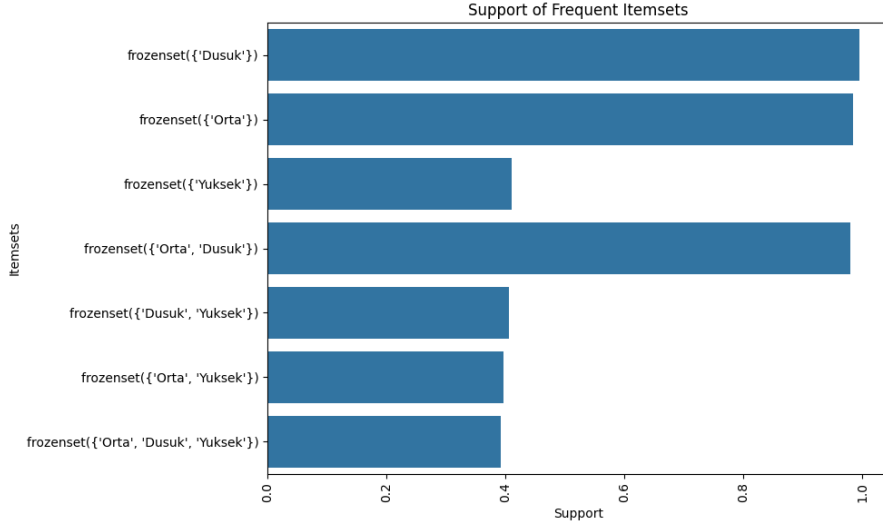
Chat-GPT tarafından üretilen analiz: "Page life expectancy değerlerine göre yapılan etiketleme, 59950 değerinin etiket 0 ve 10017 değerinin etiket 1 olduğunu gösteriyor. Bu sonuçlar, veri setindeki sayısal değerlerin iki farklı kategoriye ayrıldığını göstermektedir. Page life expectancy değerlerinin bu şekilde etiketlenmesi, veri setindeki farklı performans seviyelerini yansıtabilir. Düşük değerlere sahip olanlar bir kategorideyken yüksek değerlere sahip olanlar diğer kategoride yer almaktadır. Bu sayede veri setinin daha anlamlı bir şekilde analiz edilmesi ve kararlar alınması sağlanabilir."



Chat-GPT tarafından üretilen analiz: 'Senin analizin'

Dengesiz sınıf dağılımı sorununu ele almak için gerçekleştirdiğim örnekleme tekniklerinden biri Oversampling, diğeri ise Undersampling yöntemleridir. İlk olarak Oversampling yapıldığında, etiket 0 ve 1 için 59950'şer adet veriye ulaşılmıştır. Bu işlemle azınlık sınıfın veri sayısı artırılarak sınıflar arasındaki dengesizlik giderilmeye çalışılmıştır. Diğer yandan Undersampling işlemi sonucunda ise her iki sınıftan da 10017'şer adet veri bulunmaktadır. Azınlık sınıfın veri sayısını azaltarak sınıflar arasında denge sağlanmaya çalışılmıştır.

Özetle, Oversampling yöntemiyle sınıflar arasındaki veri sayısı farkı giderilirken, Undersampling yöntemiyle her iki sınıfın veri sayıları dengelenmeye çalışılmıştır. Bu tekniklerin kullanılmasıyla elde edilen veri dağılımlarının model performansı üzerinde olumlu etkilere sahip olması beklenmektedir.

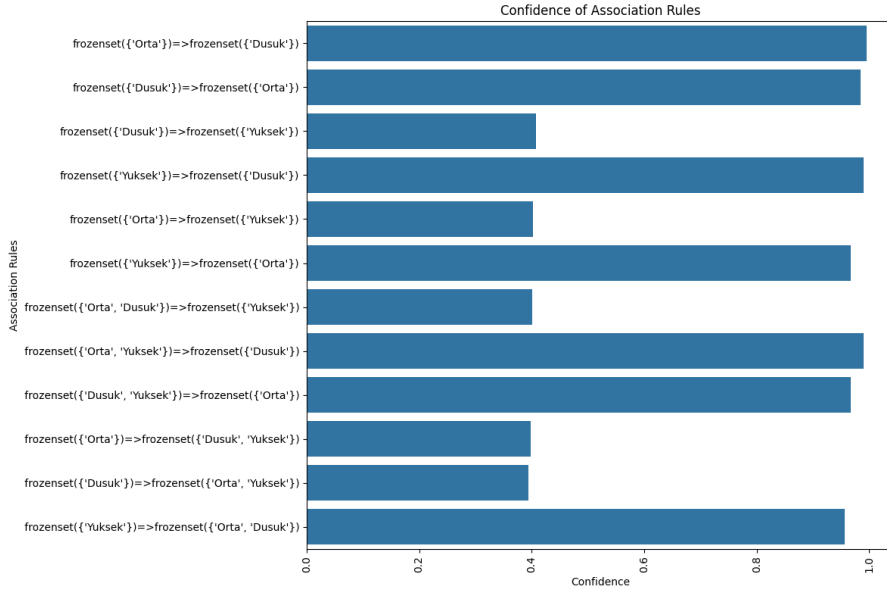


Chat-GPT tarafından üretilen analiz: 'Bu veri setinde sıklıkla görülen kombinasyonlar incelendiğinde, en yüksek destek değerine sahip olan tekli öge "Düşük" olarak belirlenmiştir (%99,55). Bu durumu takip eden diğer tekli ögeler sırasıyla "Orta" (%98,48) ve "Yüksek" (%41,03) olarak tespit edilmiştir.

İkili öge kümelerine baktığımızda, en yüksek destek değerine sahip kombinasyon "Orta" ve "Düşük" (%98,03) olarak belirlenmiştir. Ardından ise "Düşük" ve "Yüksek" (%40,62) kombinasyonu gelmektedir. "Orta" ve "Yüksek" ikilisi ise %39,67 destek değeri ile üçüncü sırada yer almaktadır.

Üçlü öge kümesine bakıldığında ise en yüksek destek değerine sahip kombinasyon "Orta", "Düşük" ve "Yüksek" (%39,26) şeklinde hesaplanmıştır.

FP-Growth algoritması yardımıyla elde edilen bu sonuçlar, veri setindeki ögeler arasındaki ilişkilerin güçlü bir şekilde incelenebileceğini göstermektedir.'



Chat-GPT tarafından üretilen analiz: 'Senin analizin'

FP-Growth algoritması kullanılarak veri seti üzerinde yapılan analiz sonucunda, Orta seviyesinde olan eğitim düzeyine sahip kişilerin Düşük seviyesinde olan eğitim düzeyine sahip kişilerle ilişkili olduğu gözlemlenmiştir. Bu ilişkide, Orta seviyedeki eğitim düzeyine sahip kişilerin Düşük seviyedeki eğitim düzeyine sahip olma olasılığı, 0.984836 olan conviction değeriyle ifade edilmiştir.

Ayrıca, Düşük seviyesinde olan eğitim düzeyine sahip kişilerin Yüksek seviyede eğitim düzeyine sahip olma olasılığı da incelenmiştir. Bu durumda, Düşük seviyedeki eğitim düzeyine sahip kişilerin Yüksek seviyedeki eğitim düzeyine sahip olma olasılığı 0.996164 conviction değeriyle ifade edilmiştir.

Son olarak, Orta ve Düşük seviyedeki eğitim düzeylerine sahip olan kişilerin birlikte Yüksek seviyede

eğitim düzeyine sahip olma olasılığı da araştırılmıştır. Bu durumda, Orta ve Düşük seviyedeki eğitim düzeylerine sahip kişilerin birlikte Yüksek seviyede eğitim düzeyine sahip olma olasılığı 0.983709 conviction değeriyle ifade edilmiştir.

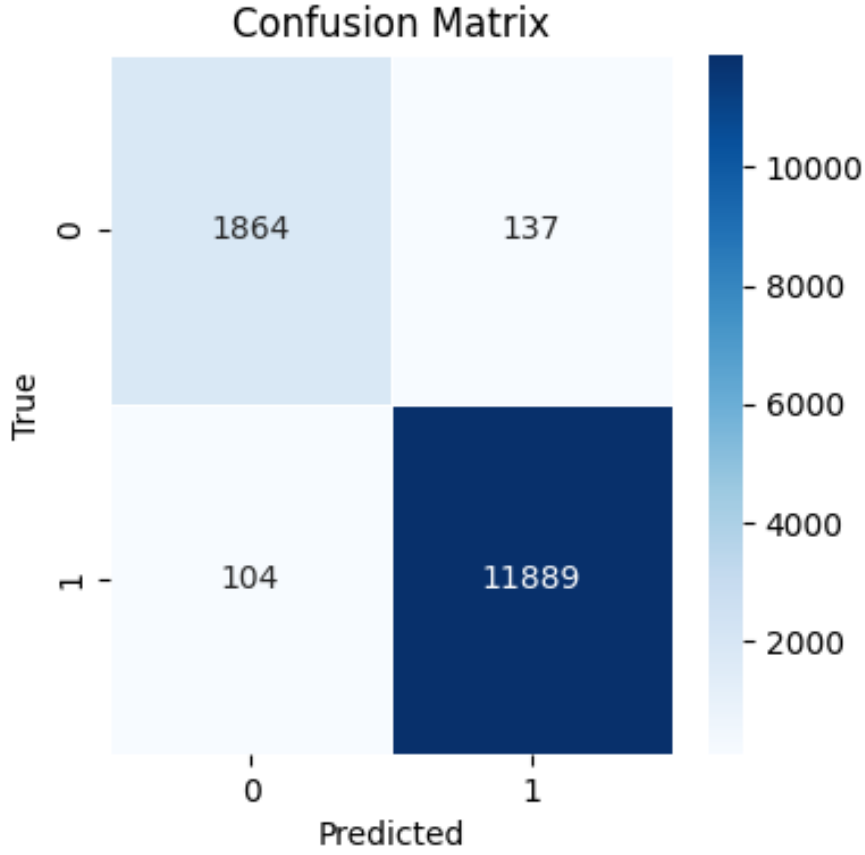
Bu analiz sonuçlarına göre, eğitim düzeyi kategorileri arasında belirli ilişkilerin olduğu ve bu ilişkilerin belirli olasılık değerleri ile ölçülebileceği görülmektedir.

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



Çalışan algoritma: SVM

Chat-GPT tarafından üretilen analiz: Bu sonuçlara göre SVM makine algoritması oldukça başarılı bir performans sergilemiş. Confusion Matrix'e baktığımızda, sınıflandırma işleminde 1864 doğru pozitif, 137 yanlış negatif, 104 yanlış pozitif ve 11889 doğru negatif sonuç elde edilmiştir. Recall değeri 0.991 ile oldukça yüksek bir başarı oranına işaret ederken, Accuracy 0.982 ile oldukça yüksek bir doğruluk seviyesini temsil etmektedir. F1 Score ise 0.989 değeriyle tatmin edici bir denge sağlamıştır. Bu sonuçlara dayanarak, SVM algoritmasının veri setini iyi bir şekilde sınıflandırdığı ve genel olarak yüksek doğrulukla çalıştığı söylenebilir.