



LOTUS AI

YAPAY ZEKA VE BİLİŞİM ÇÖZÜMLERİ A.Ş.

BİRLİKTELİK ANALİZİ RAPORU

HAZIRLAYANIN;

ADI: SENANUR
SOYADI: BAYRAM

06/12/2024

Giriş

Bu çalışmada, **alışveriş sepeti analizi** üzerinden **birliktelik analizi** yapılacaktır. **Groceries Dataset** adlı veri seti, alışveriş yapan üyelerin satın aldığı ürünlerin listelerinden oluşmaktadır. Veri setinde **Member_number**, **Date** ve **itemDescription** gibi sütunlar bulunmaktadır. **Member_number** her bir müşteri için benzersiz bir kimlik numarası iken, **Date** alışverişin yapıldığı tarihi, **itemDescription** ise her bir alışverişte satın alınan ürünleri belirtmektedir. Bu veri seti, özellikle alışveriş sepeti analizi yapmak amacıyla sıklıkla kullanılan ve ürünler arasındaki ilişkilerin keşfedilmesini sağlayan bir yapıya sahiptir.

Birliktelik Kuralı Analizi

Birliktelik kuralı analizi, özellikle perakende sektöründe yaygın olarak kullanılan bir tekniktir. Bu analiz, birbiriyle ilişkili ürünler arasındaki **gizli bağlantıları** keşfetmeyi amaçlar. Temel olarak, **Apriori algoritması** gibi yöntemlerle, bir ürünün satın alınmasıyla birlikte hangi diğer ürünlerin satın alındığını belirlemeye çalışır. Bu tür analizler, **cross-selling** stratejileri ve müşteri alışveriş davranışlarını anlamak için oldukça faydalıdır.

Birliktelik analizi genellikle üç temel metrikle değerlendirilir:

- **Destek (Support):** Bir ürün kümesinin tüm işlemler içerisinde ne sıklıkla görüldüğünü gösterir. Destek değeri, ürünlerin birbirleriyle birlikte görünme sıklığını temsil eder. Düşük destek değerleri, daha nadir kuralları işaret ederken, yüksek destek değerleri yaygın kuralları ifade eder.
- **Güven (Confidence):** Bir ürünün diğerini tahmin etme olasılığını gösterir. Güven değeri, bir ürün setinin diğerini tahmin etme doğruluğunun bir ölçüsüdür. Yüksek güven, bir ürünle birlikte diğer ürünün yüksek ihtimalle satın alınacağı anlamına gelir.
- **Lift:** İki ürün arasındaki ilişkinin ne kadar güçlü olduğunu gösterir. Lift, iki ürün arasındaki ilişkinin **bağımsızlıktan ne kadar farklı olduğunu** ölçer. Lift değeri **1'den büyükse**, bu iki ürün arasındaki ilişki bağımsızlıktan daha güçlüdür.

Birliktelik analizi, **Python** araçları kullanılarak yapılacaktır. **Python mlxtend** kütüphanesi ile daha esnek ve özelleştirilmiş analizler yapmayı mümkün kılar.

1. PYTHON

Bu çalışma, **birliktelik analizi** yapmak amacıyla **Groceries Dataset** veri setini kullanmaktadır. Bu tür analizler, alışveriş sepetindeki ürünler arasındaki ilişkileri keşfetmek için yaygın olarak kullanılır. Veri setinde her bir müşterinin aldığı ürünler, **itemDescription** sütununda listelenmiştir. Kod, önce veriyi **Member_number** ve **itemDescription** sütunlarıyla gruplandırarak, her müşteri için satın aldığı ürünlerin varlığını bir **ikili matris** formatına dönüştürür. Bu matris, her ürünün alışverişte bulunup bulunmadığını **1 veya 0** ile belirtir.

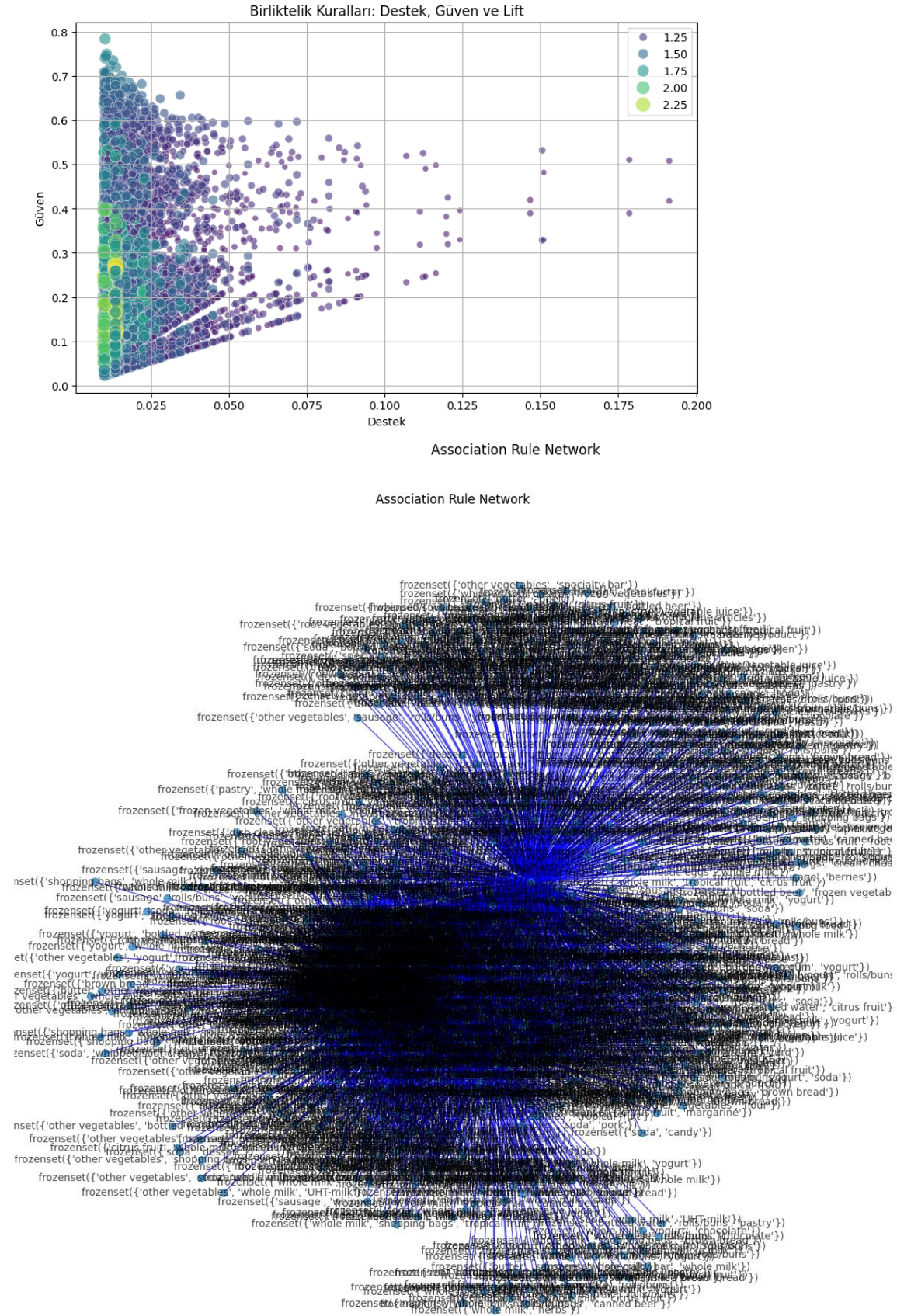
Apriori algoritması kullanılarak, veri setindeki ürünler arasında sıkça görülen kombinasyonlar (itemsets) bulunur. **min_support** parametresi, yalnızca en yaygın ürün kombinasyonlarının seçilmesini sağlar. Daha sonra, bu ürün kombinasyonları arasındaki ilişkiyi belirlemek için **birliktelik kuralları** oluşturulur. **Lift** metriği kullanarak, ürünler arasındaki ilişkinin gücü ölçülür ve **en yüksek lift ve güven** değerlerine sahip kurallar sıralanır.

Sonuçların görselleştirilmesinde, **Seaborn** kütüphanesi ile **scatter plot** (dağılım grafiği) oluşturulur. Bu grafik, ürünler arasındaki ilişkilerin **support** (destek) ve **confidence** (güven) değerlerini görsel olarak gösterir. **Lift** değeri ise, nokta boyutuyla ifade edilir. Bu görselleştirme, güçlü ilişkilerin görsel olarak anlaşılmasını sağlar.

Ayrıca, **networkx** kütüphanesi ile bir **kural ağı** (association rule network) oluşturulmuş ve görselleştirilmiştir. Bu ağda, her bir **birliktelik kuralı** için bir **kenar** oluşturulur, ve bu kenarların kalınlığı, **lift** değerine bağlı olarak değişir. Bu görselleştirme, ürünler arasındaki ilişkilerin görsel bir temsiliyi sağlar ve hangi ürünlerin birlikte alınmasının daha olası olduğunu gösterir.

Bu analiz, özellikle **pazarlama stratejileri** ve **cross-selling** fırsatları için oldukça faydalıdır, çünkü **güçlü ilişkiler** ve **ürün kombinasyonları** ortaya çıkarmak mümkündür.

Çıktılar ise şu şekildedir:



Association Rule Network:

Bu görsel, ürünler arasındaki birliktelik kurallarını bir ağ yapısıyla görselleştirir. Düğümler, ürün veya ürün gruplarını, bağlantılar ise ürünler arasındaki ilişkileri temsil eder. Ağın yoğunluğu, analize dahil edilen kural ve ürün sayısının çok fazla olduğunu gösterir. Merkezdeki düğümler, en güçlü ilişkili ürün gruplarını işaret eder. Ancak, yoğun bir ağ yapısı olduğu için detayları analiz etmek zorlaşabilir. Bu nedenle, yalnızca merkezi ve belirgin ilişkileri incelemek daha anlamlı sonuçlar sunabilir.

Destek, Güven ve Lift Dağılımı:

Bu görsel, birliktelik kurallarını destek, güven ve lift metrikleri üzerinden analiz eder. Destek, ürünlerin birlikte görülme sıklığını, güven, bir ürün alındığında diğerinin alınma olasılığını, lift ise bu birlikteliğin tesadüfi olmaktan ne kadar uzak olduğunu ifade eder. Görselde düşük destek ve güven değerlerine sahip birçok kural olduğu görülüyor, ancak bazı kurallar yüksek lift değerleriyle öne çıkıyor. Bu yüksek lift değerleri, pazarlama veya stratejik planlamada anlamlı birliktelikleri işaret eder.

En yüksek performansa sahip ürün kombinasyonlarına şu şekilde örnek verebiliriz:

En Yüksek Performansa Sahip Ürün Kombinasyonları (En İyi 10 Kural):

Öncüller: other vegetables, sausage, whole milk -> Sonuç: yogurt, rolls/buns	Destek: 0.0136, Güven: 0.2704, Lift: 2.4287
Öncüller: yogurt, rolls/buns -> Sonuç: other vegetables, sausage, whole milk	Destek: 0.0136, Güven: 0.1221, Lift: 2.4287
Öncüller: other vegetables, yogurt, rolls/buns -> Sonuç: sausage, whole milk	Destek: 0.0136, Güven: 0.2598, Lift: 2.4286
Öncüller: sausage, whole milk -> Sonuç: other vegetables, yogurt, rolls/buns	Destek: 0.0136, Güven: 0.1271, Lift: 2.4286
Öncüller: yogurt, curd -> Sonuç: sausage, whole milk	Destek: 0.0100, Güven: 0.2484, Lift: 2.3220
Öncüller: sausage, whole milk -> Sonuç: yogurt, curd	Destek: 0.0100, Güven: 0.0935, Lift: 2.3220
Öncüller: sausage, whole milk, rolls/buns -> Sonuç: other vegetables, yogurt	Destek: 0.0136, Güven: 0.2789, Lift: 2.3184
Öncüller: other vegetables, yogurt -> Sonuç: sausage, whole milk, rolls/buns	Destek: 0.0136, Güven: 0.1130, Lift: 2.3184
Öncüller: other vegetables, whole milk, yogurt -> Sonuç: sausage, rolls/buns	Destek: 0.0136, Güven: 0.1893, Lift: 2.2986
Öncüller: sausage, rolls/buns -> Sonuç: other vegetables, whole milk, yogurt	Destek: 0.0136, Güven: 0.1651, Lift: 2.2986

Hiperparametre optimizasyonu ile de çıktı şu şekildedir:

