**DEÜ Fen Fakültesi**

**Bilgisayar Bilimleri Bölümü**

**BİL3013 Veri Madenciliğine Giriş**

**Ödev Konusu: K-Means Kümeleme ile Fiyat Analizi**

**Öğrenci: Yılmaz AYĞIN**

**Öğretim Üyesi: Prof. Dr. Efendi NASİBOĞLU**

**Ödev Kodlarının ve Verileirnin Bukunduğu Github Linki:**

**https://github.com/yilmazaygin/bil3013-veri-madenciligine-giris/tree/yilmazaygin-odev2**

**Tarih: İzmir, 2024**

**İÇİNDEKİLER**

1. Ödevin Tanımı
2. Veri Setinin Temizlenmesi ve Ön İşleme
3. Kullanılan Yöntem ve Teknolojiler
4. Kümeleme Analizi
   * K-Means Algoritması
   * Kümeleme Sonuçları
5. Değerlendirme ve Sonuçlar
6. Kaynakça

**I. ÖDEVİN TANIMI**

Bu ödevde, bir önceki ödevde web kazıma yöntemiyle elde edilen konut kira verilerinin fiyat dağılımını analiz ederek piyasayı sınıflandırmak amaçlanmıştır. İlgili veri setindeki "Kira Bedeli" niteliği kullanılarak, konutların fiyat aralıkları "Ucuz", "Orta" ve "Pahalı" olmak üzere üç gruba ayrılmıştır. Analiz, K-Ortalamalar (K-Means) kümeleme algoritması ile gerçekleştirilmiş ve kümeler arasındaki mesafe ölçümü için Öklit uzaklığı tercih edilmiştir.

Bu çalışmada:

* Kira bedellerine göre piyasadaki konutlar üç kümeye ayrılacak,
* Kümeleme sonucunda her kümedeki nesne sayısı, minimum, maksimum, ortalama ve standart sapma gibi değerler hesaplanacak,
* Her bir kümenin histogramları oluşturularak fiyat dağılımları görselleştirilecektir.

**II. VERİ SETİNİN TEMİZLENMESİ VE ÖN İŞLEME**

İlk adım olarak, ham veri seti temizlenmiş ve analiz için uygun hale getirilmiştir. Temizleme işlemleri aşağıdaki adımlardan oluşmaktadır:

1. Eksik Değerlerin Giderilmesi: Ham veri setindeki eksik veya boş olan gözlemler çıkarılmıştır.
2. Fiyat Anomalilerinin Tespiti: Kira bedeli için mantıksız veya aşırı uç değerler (outliers) analiz edilerek, yanlış veya tutarsız olanlar veri setinden çıkarılmıştır.
3. Veri Dönüşümü ve Standartlaştırma: Kira bedeli verileri K-Means algoritmasına uygun hale getirilmesi için ölçeklendirilmiştir. Bu aşama, K-ortalama algoritmasında kümelerin daha doğru belirlenmesi açısından önemlidir.

Bu süreç sonunda, veri seti analiz edilebilir bir forma getirilmiş ve kümeleme analizine hazır hale getirilmiştir.

**III. KULLANILAN YÖNTEM VE TEKNOLOJİLER**

Bu projede, Python programlama dili ve ilgili kütüphaneler kullanılarak veri temizleme, kümeleme ve görselleştirme işlemleri gerçekleştirilmiştir. Kullanılan başlıca kütüphaneler şunlardır:

* Pandas: Veri temizleme, düzenleme ve manipülasyon işlemlerinde kullanılmıştır. Ham ve temizlenmiş veri setlerinin işlenmesinde önemli rol oynamıştır.
* Scikit-Learn: Makine öğrenimi kütüphanesi olan Scikit-Learn, K-Means algoritmasının uygulanmasında kullanılmıştır.
* Matplotlib ve Seaborn: Verilerin görselleştirilmesi için kullanılmıştır. Özellikle kümeleme sonuçlarının histogram ve dağılım grafiklerinde etkili bir biçimde kullanılmıştır.

**IV. KÜMELEME ANALİZİ**

***1. K-Means Algoritması***

K-Means algoritması, verileri belirli bir K sayısına göre gruplamak için kullanılan bir kümeleme algoritmasıdır. Bu çalışmada K değeri 3 olarak belirlenmiş ve fiyat aralıklarına göre konutlar üç kümeye ayrılmıştır:

* Küme 0 (Ucuz): Kira bedeli düşük olan konutlar
* Küme 1 (Orta): Kira bedeli orta seviyede olan konutlar
* Küme 2 (Pahalı): Kira bedeli yüksek olan konutlar

K-Means algoritması, her veri noktasını en yakın kümeye atamak için Öklit uzaklığını kullanır. Bu analizde, her bir konutun kira bedeli diğer konutlarla karşılaştırılarak ait olduğu küme belirlenmiştir.

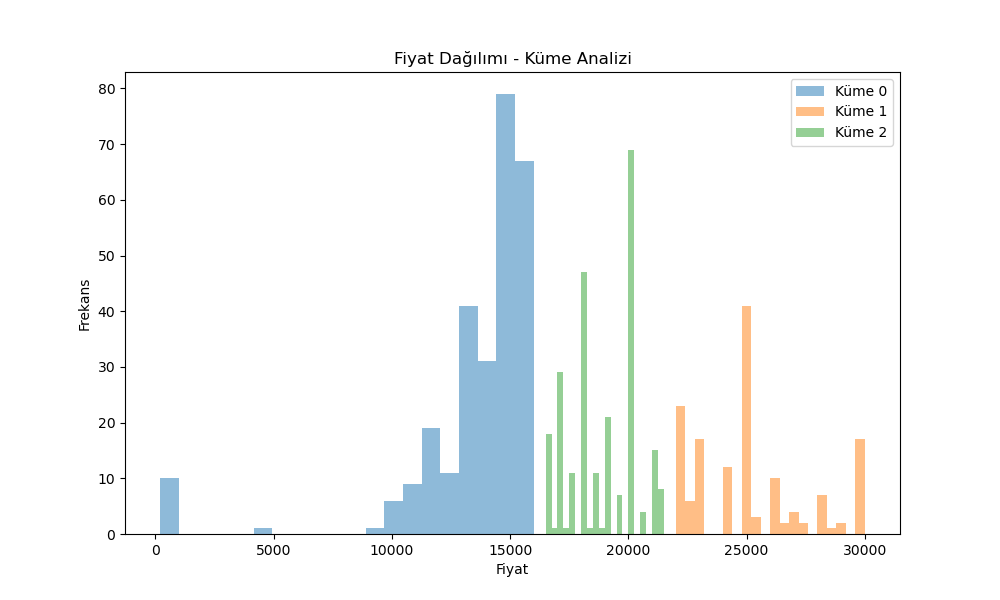
***2. Kümeleme Sonuçları ve Analizi***

Kümeleme analizinin ardından her bir konut, ait olduğu küme ile birlikte yeni bir sütun olarak veri setine eklenmiştir. Kümeleme sonuçları detaylı olarak aşağıda sunulmuştur:

* Küme 0 (Ucuz):
  + Toplam Nesne Sayısı: 275
  + Minimum Kira Bedeli: 200 TL
  + Maksimum Kira Bedeli: 16,000 TL
  + Ortalama Kira Bedeli: 13,650 TL
  + Standart Sapma: 3,070 TL
* Küme 1 (Orta):
  + Toplam Nesne Sayısı: 147
  + Minimum Kira Bedeli: 22,000 TL
  + Maksimum Kira Bedeli: 30,000 TL
  + Ortalama Kira Bedeli: 25,101 TL
  + Standart Sapma: 2,481 TL
* Küme 2 (Pahalı):
  + Toplam Nesne Sayısı: 244
  + Minimum Kira Bedeli: 16,500 TL
  + Maksimum Kira Bedeli: 21,500 TL
  + Ortalama Kira Bedeli: 18,802 TL
  + Standart Sapma: 1,421 TL

Bu analiz, her bir küme için kira bedelinin ortalama ve dağılımını ortaya koyarak fiyat gruplarını net bir şekilde tanımlamaktadır. Ayrıca her bir kümenin fiyat dağılımı, histogram grafikleri ile görselleştirilmiş ve kümeler arasındaki fiyat farklılıkları açıkça gözlemlenmiştir.

Aşağıda, fiyat dağılımı – küme analizi grafiğini görebilirsiniz.



**V. DEĞERLENDİRME VE SONUÇLAR**

Kira bedellerine göre yapılan kümeleme analizi sonucunda İzmir, Buca bölgesindeki konut kiralarının üç ana fiyat grubunda toplandığı belirlenmiştir. Bu analiz, bölgedeki piyasa fiyatları ve kira aralıkları hakkında genel bir bilgi sunmuştur. Aşağıda bu analizin sonuçları ve değerlendirmesi yer almaktadır:

* Ucuz Kira Grubu: Kira bedeli düşük olan bu grup, genellikle daha küçük, eski veya merkezi olmayan bölgelerde yer alan konutları içermektedir.
* Orta Kira Grubu: Bu grupta, piyasanın ortalama fiyatlarına sahip, özellik açısından ortalama koşullara sahip konutlar bulunmaktadır.
* Pahalı Kira Grubu: Bu grupta bulunan konutlar, genellikle daha geniş, yeni veya merkezi bölgelerde yer alarak yüksek kira bedellerine sahiptir.

Bu sınıflandırma, kullanıcıların bütçelerine uygun konutları seçmelerinde kolaylık sağlamakta ve piyasa koşulları hakkında genel bir analiz sunmaktadır. Kira fiyatlarındaki aralığın anlaşılabilir olması, kullanıcıların tercihlerine göre daha bilinçli seçim yapmalarını sağlayacaktır.

**VI. KAYNAKÇA**

1. https://scikit-learn.org
2. https://pandas.pydata.org
3. https://matplotlib.org
4. https://seaborn.pydata.org
5. "Introduction to Data Mining" – Han, Kamber, Pei