

T.C.  
BEYKENT ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ  
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİMDALI  
BİLGİSAYAR MÜHENDİSLİĞİ BİLİM DALI

**VERİ MADENCİLİĞİ TEKNİKLERİ İLE KONUT  
FİYATI BELİRLEME**

Yüksek Lisans Tezi

Tezi Hazırlayan:  
**Tuğrul UĞURLU**

İstanbul, 2015

T.C.  
BEYKENT ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ  
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİMDALI  
BİLGİSAYAR MÜHENDİSLİĞİ BİLİM DALI

**VERİ MADENCİLİĞİ TEKNİKLERİ İLE KONUT  
FİYATI BELİRLEME**

Yüksek Lisans Tezi

Tezi Hazırlayan:

**Tuğrul UĞURLU**

Öğrenci No:

120820003

Danışman:

Yrd. Doç. Dr. Turhan KARAGÜLER

İstanbul, 2015

## YEMİN METNİ

Yüksek lisans tezi olarak sunduğum “Veri Madenciliği Teknikleri İle Konut Fiyatı Belirleme” başlıklı bu çalışmanın, bilimsel ahlak ve geleneklere uygun şekilde tarafımdan yazıldığını, yararlandığım eserlerin tamamının kaynaklarda gösterildiğini ve çalışmamın içinde kullanıldıkları her yerde bunlara atıf yapıldığını belirtir ve bunu onurumla doğrularım. 25.05.2015

**Tuğrul UĞURLU**

T.C.  
BEYKENT ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZ/PROJE SAVUNMA SINAVI SONUÇ TUTANAĞI

Beykent Üniversitesi  
Fen Bilimleri Enstitüsü Müdürlüğü'ne,

Aşağıda tez/proje adı belirtilen yüksek lisans öğrencisi 120820.003 no'lu ...TUFERUL UZURLU... in 15/06/15 tarihinde yapılan tez/proje savunma sınavı<sup>1</sup> sonucunda 45... dakika süreyle sunduğu ve savunduğu tezi/projesi hakkında<sup>2</sup> oybirliğiyle, ..KABUL.. kararı verilmiştir.

Bilgilerinize saygılarımızla arz ederiz.

---

Anabilim Dalı : Bilgisayar Mühendisliği  
Programı : Bilgisayar Mühendisliği  
Tez/Proje-Başlığı<sup>3</sup> : Veri Madenciliği Tehnikleri ile Konut Fiyatları Belirleme

---

	Tez/Proje Sınav Jürisi	Öğretim Üyesi	İmza
Danışman	: <u>Yrd. Doç. Dr. Turhan KARAGÜNER</u>		<u>[İmza]</u>
Üye	: <u>Doç. Dr. Gökhan SİLİHTARÖRÜ</u>		<u>[İmza]</u>
Üye	: <u>Yrd. Doç. Dr. Ediz SAYKOL</u>		<u>[İmza]</u>

<sup>1</sup> Jüri üyeleri söz konusu tezin kendilerine teslim edildiği tarihten itibaren en geç bir ay içinde toplanarak öğrenciyi tez savunma sınavına alır. Belirlenen günde yapılamayan jüri toplantısı, katılanların hazırladığı bir tutanakla enstitü yönetimine bildirilir. Bu durumda jüri en geç onbeş gün içinde toplanarak adayı tez savunma sınavına alır. Tez savunma sınav süresi en az 45 dakikadır. Yüksek lisans tez savunma sınavı, tez çalışmasının sunulması ve bunu izleyen soru-yanıt bölümlerinden oluşur ve dinleyiciye açıktır. (Beykent Lisansüstü eğitim ve Öğretim Yönetmeliği-Madde30-3)

<sup>2</sup> Tez sınavının tamamlanmasından sonra jüri, tez hakkında "kabul", "düzeltme" veya "red" kararı verir. Jüri başkanı, jüri üyelerince imzalanmış sınav tutanağını, tez sınavını izleyen üç gün içinde ilgili enstitü yönetimine teslim eder. Tezi başarısız bulunan öğrencinin Enstitü ile ilişkisi kesilir. Tezi hakkında düzeltme kararı verilen öğrenci en geç üç ay içinde gerekli düzeltmeleri yaparak ve yönetmelikte belirtilen usullere uygun olarak tezini aynı jüri önünde yeniden savunur. Bu savunma sınavında da tezi kabul edilmeyen öğrencinin enstitü ile ilişkisi kesilir. (Beykent Lisansüstü eğitim ve Öğretim Yönetmeliği-Madde30-4)

<sup>3</sup> İleride doğabilecek aksaklıkların engellenmesi için tezin başlığını yazılması gerekmektedir.

# VERİ MADENCİLİĞİ TEKNİKLERİ İLE OPTİMUM KONUT FİYATI BELİRLEME

Tezi Hazırlayan: Tuğrul UĞURLU

## ÖZET

Çalışma, emlak sitesinden alınan bir veri tabanını kullanmaktadır. Burada amaç konut satmak isteyen mal sahibinin, emlakçının ve inşaat firmasının konut fiyatını belirlemesinde yardım önermektir. Tez içerisinde, veri tabanı tanıtılıp KNIME ve WEKA ile yapılan analizlerin sonuçları değerlendirilerek, karar ağacı çıkartılıp fiyat ile ilgili sonuçlar ve öneriler verilecektir.

Tezde veri madenciliği teknikleri kullanarak ev satışlarında konut fiyatı aralıkları belirlenecektir. Bu işlem yapılırken çeşitli grafik analizlerden ve karar ağacından faydalanılmıştır. K-Means(K-Ortalama) ve Fuzzy C-Means(Bulanık C-Ortalama) algoritmalarını kullanarak kümeleme işlemi yapılmıştır. Daha sonra bu algoritmalarından elde edilen sonuçlar karar ağaçlarına sokularak emlak fiyatlarının belirlenmesinde etkili olan özellikler tespit edilmiştir. Sonrasında çıkan sonuçlar değerlendirilerek uygulama geliştirilmiştir. Bu sayede yeni girilen bir konutun alabileceği fiyat aralığı tespit edilebilmiştir.

**Anahtar Kelimeler:** Karar ağacı, Veri madenciliği teknikleri, Veri madenciliği algoritmaları

# HOUSE PRICING WITH DATA MINING TECHNIQUES

**Presented by: Tuğrul UĞURLU**

## ABSTRACT

In this study a database from the real estate website has been used. The aim is to detect several important factors which will directly affect the price of housing. The organization of this work follows. First the database will be introduced, then the analyse results which are constructed by using KNIME and WEKA will be discussed and finally the prices predicted by the decision-trees will be analyzed.

This work predicts price range of a housing by using the following steps. First various analyses are performed by using decision trees. Then clustering is performed by using K-means and Fuzzy C-means algorithms. Afterwards several important factors which affect housing prices are predicted and by using the results, the application is improved. Thus the application is now able to define the price range of any house.

**Key Words:** Decision tree learner, Data mining technique, Data mining algorithms

## İÇİNDEKİLER

### ÖZET

### ABSTRACT

<b>ŞEKİLLER</b> .....	iv
<b>TABLO</b> .....	vi
<b>DENKLEMLER</b> .....	vii
<b>KISALTMALAR</b> .....	viii
<b>1. GİRİŞ</b> .....	1
<b>2. VERİ MADENCİLİĞİ TEKNİĞİ</b> .....	3
2.1 Kümeleme .....	3
2.1.1 Kümeleme Analizinde Kullanılan Benzerlik ve Uzaklık Ölçütleri .....	4
2.2 Sınıflandırma.....	6
<b>3. VERİLERİN TOPLANMASI</b> .....	7
3.1 WEKA.....	7
3.2 KNIME.....	7
<b>4. VERİ TABANININ TANITIMI</b> .....	8
4.1 Değişken Bazlı Grafik Analiz Çalışması .....	12
<b>5. VERİLERİN VERİ MADENCİLİĞİ TEKNİKLERİYLE KÜMELENMESİ</b> .....	21
5.1 Fuzzy C-Means İle Kümeleme ve DTL .....	22
5.2 K-Means İle Kümeleme ve DTL.....	28
5.3 Weka ile Class'a Etki Eden Özelliklerin Bulunması .....	36
5.4 AttributeSelection İşlemi Sonrası Fuzzy C-Means İşleminin DTL Sokulması	38
5.5 AttributeSelection İşlemi Sonrası K-Means İşleminin DTL Sokulması.....	43
5.6 Kümeleme Algoritmaları .....	46
5.6.1 K-Means Algoritması .....	47
5.6.2 Fuzzy C-Means Algoritması.....	48
5.7 Sınıflandırma Algoritması.....	49
<b>6. UYGULAMA</b> .....	51
<b>SONUÇ</b> .....	52
<b>KAYNAKLAR</b> .....	53

## ŞEKİLLER

Şekil 1. Küme Yapısı .....	4
Şekil 2. Gözetimli Öğrenme Akış Şeması Örneği .....	6
Şekil 4. BanyoSayisi Değişkeninin Histogram Değeri .....	13
Şekil 5. BinaYasi Değişkeninin Histogram Değeri .....	13
Şekil 6. KatSayisi Değişkeninin Histogram Değeri.....	14
Şekil 7. BulunduguKat Değişkeninin Histogram Değeri.....	14
Şekil 8. Aidat Değişkeninin Histogram Değeri .....	15
Şekil 9. Isitma Değişkeninin Histogram Değeri .....	15
Şekil 10. Esyali Değişkeninin Histogram Değeri .....	16
Şekil 11. KullanimDurumu Değişkeninin Histogram Değeri.....	16
Şekil 12. SiteIcerisinde Değişkeninin Histogram Değeri .....	17
Şekil 13. KrediyeyeUygun Değişkeninin Histogram Değeri.....	17
Şekil 14. Kimden Değişkeninin Histogram Değeri .....	18
Şekil 15. Takas Değişkeninin Histogram Değeri.....	18
Şekil 16. Latitude Değişkeninin Histogram Değeri .....	19
Şekil 17. Longitude Değişkeninin Histogram Değeri.....	19
Şekil 18. Fiyat Değişkeninin Histogram Değeri .....	20
Şekil 19. Veriler Hakkında Tanımlayıcı İstatistik Bilgileri .....	20
Şekil 20. Data.csv Dosyasından Okunan Verilerin Bir Bölümü.....	21
Şekil 21. Fiyat Değişkeninin Histogram Üzerinde İncelenmesi .....	22
Şekil 22. Fuzzy C-Means Algoritmasından Çıkan Grupların Bir Kısmı .....	22
Şekil 23. Fuzzy C-Means DTL Akış Şeması .....	23
Şekil 24. DTL Elde Edilen Sonucun Bir Kısmı .....	23
Şekil 25. Scorer İçin Akış Şeması.....	27
Şekil 26. DTL Doğrulaması .....	28
Şekil 27. K-Means Kümeleme Sonucu Oluşan Küme Yapısının Bir Parçası .....	28
Şekil 28. K-Means DTL Akış Şeması.....	29
Şekil 29. K-Means DTL Sonucunun Bir Parçası .....	29
Şekil 30. Scorer Akış Şeması.....	35



Şekil 31. DTL Doğrulaması .....	35
Şekil 32. Weka Fuzzy C-Means AttributeSelection İşlemi Sonucu .....	37
Şekil 33. Weka K-Means AttributeSelection İşlemi Sonucu .....	37
Şekil 34. Weka Fuzzy C-Means Scorer Sonucu .....	42
Şekil 35. Weka K-Means Scorer Sonucu .....	46
Şekil 37. Örnek Karar Ağacı .....	50
Şekil 38. Örnek Uygulama Arayüzü .....	51

## TABLO

Tablo 1. OdaSayisi Deęerlerinin Gruplandırılmıř Durumu .....	8
Tablo 2. BanyoSayisi Deęiřkeninin Gruplandırılmıř Durumu .....	9
Tablo 3. BinaYasi Deęiřkeninin Gruplandırılmıř Durumu .....	9
Tablo 4. KatSayisi Deęiřkeninin Gruplandırılmıř Durumu.....	9
Tablo 5. BulunduguKat Deęiřkeninin Gruplandırılmıř Durumu.....	10
Tablo 6. Isitma Deęiřkeninin Gruplandırılmıř Durumu .....	10
Tablo 7. KullanimDurumu Deęiřkeninin Gruplandırılmıř Durumu.....	11
Tablo 8. Kimden Deęiřkeninin Gruplandırılmıř Durumu .....	11

## DENKLEMLER

Denklem 1. Minkowski Uzaklığı .....	5
Denklem 2. Manhattan City-Block Uzaklığı .....	5
Denklem 3. Öklid Uzaklığı .....	5
Denklem 4. Mahalanobis Uzaklığı.....	5
Denklem 5. Öklid Uzaklığı Formülü .....	47
Denklem 6. K-Means k Adet Rasgele Nesne Seçimi.....	47
Denklem 7. K-Means Karesel Hata Formülü.....	47
Denklem 8. K-Means Kare-Hata Formülü.....	48
Denklem 9. Fuzzy C-Means En Küçük Kareler Yöntemi .....	48
Denklem 10. Fuzzy C-Means U Üyelik Matrisi .....	49
Denklem 11. Fuzzy C-Means U Matrisi Karşılaştırma.....	49
Denklem 12. CART Algoritması .....	50

## KISALTMALAR

- DTL** : Decission Tree Learner / Karar Ağacı
- DTP** : Decission Tree Predictor / Karar Ağacı Öğrenici
- EM** : Expectation–maximization / Beklenti Maksimizasyon

## 1. GİRİŞ

Veri madenciliği, depolanan verilerden, bir başka deyişle veri tabanından daha önceden bilinmeyen, gizli kalmış, kullanılabilir, ilginç bilgilerin çekilmesi işlemidir ve bankacılık, pazarlama, sigortacılık, sağlık gibi değişik alanlarda uygulanmaktadır.

Doğru fiyatlandırma kararları bir ürün veya hizmetin piyasada başarılı olabilmesi için en önemli kararlardan birisidir. Özellikle teknolojinin büyük bir hızla değiştiği, tüketicilerin internet üzerinden dünyanın dört bir köşesindeki rakiplerin ürünleriyle karşılaştırma yapabildikleri ve her geçen gün yeni rakiplerin piyasaya girdikleri bir dönemde fiyatlandırma konusunda yapılan yanlışlıklar ürün veya hizmet doğmadan ölmesine neden olur [1].

Fiyatlandırma, pazarlayıcı için önemli olduğu kadar alıcı içinde önemlidir. Fiyatlandırma kararları bütünsel bakışı, stratejik yaklaşımı ve müşteriler için yaratılan değeri göz önüne almayı gerektiren önemli kararlardır.

Türkiye’de gayrimenkul değerlemesi üzerine yapılan ilk çalışmalar 40 yıl öncesine kadar gider [2]. Bu çalışmada gayrimenkul değer saptanmasında karşılaştırma, gelir ve maliyet parametrelerini esas alan yöntemlerin varlığı vurgulanmıştır.

Bu tezin amacı, konut fiyatları belirlenirken veri madenciliği tekniklerini kullanarak konut fiyatının olması gereken (optimum) fiyat aralıklarını bulmaktır. Tezde, emlak sitesinden alınan gerçek verileri kullanarak satışa sunulmuş konutların belirtilmiş özelliklerine bakarak yeni girilen bir konutun satışında fiyatının belirlenmesi için satıcıya konutun fiyatlandırılması konusunda yardımcı olunacaktır. Çalışmada pilot bölge seçilmiş ve bu bölge üzerinde denemeler planlanmıştır. Pilot bölge olarak İstanbul ili Maltepe ilçesi Yalı Mah. seçilmiştir. Ev fiyatlarındaki değişkenlik aralığı göz önüne alındığında bu işlem için seçilen bölgenin uygun olduğu düşünülmüştür. Çalışmanın bu bölümünde, KNIME ve WEKA araçlarından yararlanılıp çeşitli grafik analizler ve karar ağaçları kullanılmıştır. Veri boyutu büyük olduğundan hepsi görüntülenemeyeceği için verilerin bir kısmı görüntülenmiştir.

Kısım 2’de veri madenciliđi tekniđinden bahsedilmiřtir. Kısım 3’de verilerin nasıl toplandıđı anlatılmıřtır. Kısım 4’de veri tabanı tanıtılmıřtır. Kısım 5’de veri madenciliđi teknikleri uygulanıp kümele iřlemi ve sınıflandırma iřlemi yapılmıřtır. Kısım 6’da örnek bir uygulama geliřtirilmiřtir. Kısım 7’de sonuçlar deđerlendirilmiřtir.

## 2. VERİ MADENCİLİĞİ TEKİNİĞİ

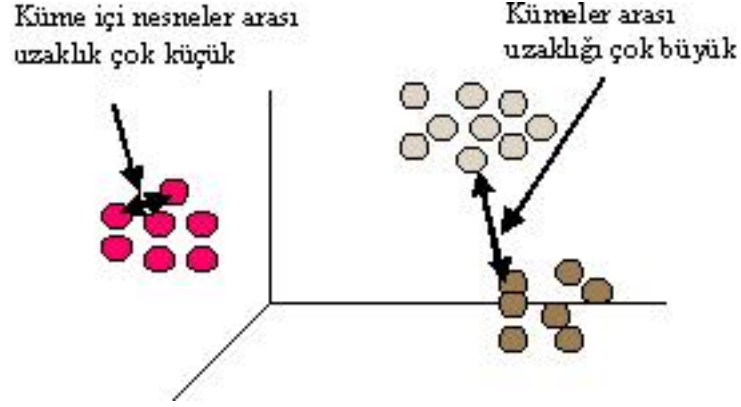
Veri madenciliđi, daha 6nceden bilinmeyen, geęerli ve uygulanabilir bilgilerin geniř veri tabanlarından elde edilmesi ve bu bilgilerin iřletme kararları verirken kullanılmasıdır [3]. Veri madenciliđi, veri tabanları, istatistik ve yapay 6đrenme konularının kavramlarına dayanır ve onların tekniklerini kullanır.

Veri madenciliđi 6zerine matematikçiler 1950’li yıllarda mantık ve bilgisayar bilimleri alanlarında alıřmıř 1960’lı yıllarda ise istatistikçiler yeni bir algoritma keřfetmiřlerdir. 6rneđin regresyon analizi, en b6y6k olabilirlik, sinir ađları vb. metotlar veri madenciliđinin ilk adımlarını oluřturmuřtur. 1970-1990’lı yıllarda yeni programlama dilleri ve yeni bilgisayar tekniklerinin geliřtirilmesi genetik algoritmalar, EM algoritması, K-Means algoritması ve DTL gibi algoritmaları da iermiřtir.

Burada altı izilmesi gereken konu elde edilecek bilginin 6nceden bilinmeyen olmasıdır. 6nceden tahmin edilebilen veya bařka y6ntemlerle ıkarılmıř sonular veri madenciliđinin uygulama alanlarına girmemektedir.

### 2.1 K6meleme

ok deđiřkenli analiz tekniklerinden biri olan k6meleme analizinin 6ncelikli amacı, birey ya da nesnelerin temel 6zelliklerini dikkate alarak onları gruplandırmaktır. Diđer bir deyiřle k6meleme analizi, gruplanmamıř verileri benzerliklerine g6re gruplandırarak arařtırmacıya 6zet bilgiler sunmaktır [4]. Analiz sonunda elde edilen k6melerin kendi ilerinde homojen, kendi aralarında ise heterojen bir yapıda olmaları beklenir. K6meleme analizinin ileriye y6nelik kestirim amalı kullanımı s6z konusu deđildir [5].



**Şekil 1.** Küme Yapısı

Şekil 1’de de görüldüğü gibi ortaya çıkacak kümelerin kendi içlerinde homojen, kendi aralarında ise heterojen bir yapıda olmaları beklenir. Belirlenen her bir gruba küme ismi verilir. Sınıflandırma işleminde, sınıflar önceden belirli iken kümeleme işleminde sınıflar önceden belirli değildir. Kümeleme analizi genel olarak dört aşamadan oluşmaktadır.

#### **Analizin aşamaları**

- Veri matrisinin hazırlanması
- Benzerlik veya uzaklık matrislerinin elde edilmesi
- Kümeleme yönteminin belirlenmesi
- Sonuçların yorumlanması

#### **2.1.1 Kümeleme Analizinde Kullanılan Benzerlik ve Uzaklık Ölçütleri**

Kümeleme analizi, gözlemlerin benzerlik veya farklılıklarını göz önüne alarak gözlemleri homojen gruplara ayırmaktadır [6].

Benzerlik, iki nesne veya özellik arasındaki ilişkinin kuvveti olarak açıklanır. Uzaklık ise, iki nesne arasındaki zıtlık veya uyumsuzluğun farklılıklarını ölçer. Benzerlik ve uzaklık ölçümleri gözlemlerin birbirinden ayırt edilmesini sağlar ve bu sayede gözlemler gruplara ayrılır.



Nicel veriler için uzaklık ölçümü çeşitli matematiksek yöntemler ile yapılmaktadır. Sıklıkla kullanılan ölçüler aşağıdaki gibi verilmiştir.

Her biri p tane nicel değişken içeren  $x_i$ , ve  $x_j$  gözlem çifti arasındaki uzaklık,

$d_y = d(x_i, x_j)$  olsun.

**Minkowski Uzaklığı:**

(1)

**Manhattan City-Block Uzaklığı ( $\lambda=1$  durumunda):**

(2)

**Öklid Uzaklığı ( $\lambda =2$  durumunda):**

(3)

**Mahalanobis Uzaklığı:**

(4)

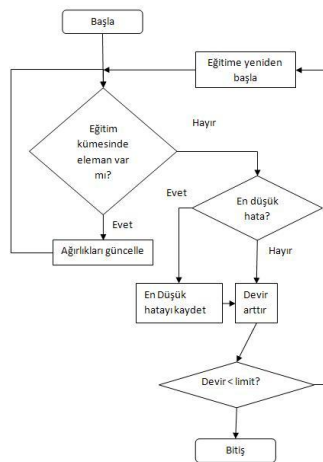
Bağıntılarıyla hesaplanır [7].

## 2.2 Sınıflandırma

Sınıflandırma kavramı, basitçe bir veri kümesi üzerinde tanımlı olan çeşitli sınıflar arasında veriyi dağıtmaktır. Sınıflandırma algoritmaları, verilen eğitim kümesinden bu dağılım şeklini öğrenirler ve daha sonra sınıfının belirli olmadığı test verileri geldiğinde doğru şekilde sınıflandırmaya çalışırlar. Veri kümesi üzerinde verilen bu sınıfları belirten değerlere etiket ismi verilir ve gerek eğitim gerekse test sırasında verinin sınıfının belirlenmesi için kullanılırlar [8].

Sınıflandırma en çok bilinen veri madenciliği tekniklerinden birisidir. Yapay sinir ağları başta olmak üzere bilgisayar bilimlerinin alt dalı olan yapay zekânın çeşitli alanlarında kullanılan öğrenme çeşididir. Sınıflandırma tahminleyici bir modeldir. Sınıflandırma, gizli kalmış kuralları veri setinden çıkartarak daha sonra girilen verinin hangi gruba ait olup olmadığına karar verir.

Gözetimli öğrenme alternatifini olan gözetimsiz öğrenmeye göre daha hızlı çalışmaktadır. Bunun en önemli sebebi ise gözetimli öğrenme, eldeki giriş verilerinin ve istenen sonuçların verilerek bu bilgilerden giriş ve çıkış arasında ilişki oluşturulup makinenin konuyu öğrenmesinin sağlanmasıdır. Gözetimsiz öğrenme ise, gözetimli öğrenmeden farklı olarak verileri sebep-sonuç ya da giriş-çıkış şeklinde etiketlemeden, veri içerisinde var olan ilişkilerin ve yapıların öğrenilmeye çalışılmasıdır. Şekil 2’de basit bir akış diyagramı ile gözetimli öğrenmenin nasıl yapılabileceği gösterilmiştir.



Şekil 2. Gözetimli Öğrenme Akış Şeması Örneği

### 3. VERİLERİN TOPLANMASI

Verilerimizin doğru olması için veri setimizi çok güvendiğimiz ve temiz bir veriye sahip olduğunu düşündüğümüz [sahibinden.com](http://sahibinden.com) adlı internet sitesinin emlak kısmından geliştirilen web site kopyalayıcı uygulaması yardımıyla aldık. Alınan verileri daha sonra html etiketlerinden temizlemek için regex kuralları oluşturarak verileri alıp temiz bir yapıya kavuşturduk.

Uygulamanın modellenmesinin kolay olması için bir pilot bölge seçimine karar verildi. Bu sayede işlemi gözlemlene ve test etme şansı yakalana bilinecektir. Pilot bölge olarak giriş bölümünde de değinildiği gibi İstanbul ili Maltepe ilçesi Yalı Mahallesi seçildi. Projede verileri incelemek için aşağıda kısaca tanıtilen WEKA ve KNIME yazılım araçlarından yararlanılmıştır. KNIME, GUI ara yüzü kullanım kolaylığından dolayı tercih edilmiştir. WEKA, içinde bulunan *AttributeSelection* algoritması yardımıyla, KNIME da karar ağacı yapısı incelenirken yol göstermesinden dolayı tercih edilmiştir.

#### 3.1 WEKA

WEKA, Waikato Üniversitesi'nde geliştirilen ücretsiz bir veri madenciliği ve makine öğrenmesi yazılımıdır. WEKA Java programlama dili ile geliştirilmiştir. Java dili birçok değişik öğrenme algoritmaları için düzenli bir platform sağlamaktadır. WEKA'nın en güçlü özelliği birçok sınıflandırma tekniğini içinde barındırmasıdır. Diğer bir önemli özelliği ise uygulamaların komut girilerek gerçekleştirilmesine olanak tanımasıdır [9].

#### 3.2 KNIME

KNIME, WEKA gibi bir veri madenciliği ve makine öğrenmesi yazılımıdır. KNIME'da Weka gibi Java dili ile geliştirilmiştir. KNIME GNU lisansı ile kullanılabilen açık kaynak kodlu bir yazılımdır. KNIME, 60'dan fazla veri türü üzerinde kullanılmaktadır. Örneğin, örüntü tanıma, basit ve karmaşık analizlerde, insan biyolojisi üzerinde vs. gibi konular yaygın olarak denenilen alanlardır [10].

#### 4. VERİ TABANININ TANITIMI

Veriler, *Microsoft SQL Server 2012* veri tabanı programında saklanmıştır. Veri tabanında bulunan *Model* tablosunun içinde yaralan değişkenler aşağıda gösterilmiştir.

**Fiyat:** Alıcıya sunulan satış fiyatıdır.

**M2:** Alıcıya sunulan konutun m2 cinsinden büyüklüğüdür. 100-1000 m2 arasında konutlar mevcuttur.

**OdaSayisi:** Alıcıya sunulan konutun oda sayısıdır. Oda sayısının veri madenciliği uygulanırken gruplandırılmış değerleri aşağıdaki tabloda gösterilmiştir. Verileri gruplandırarak için *string* olan değerleri veri madenciliğinde işlemek için *numeric* değer karşılıklarıyla gruplandırılmıştır.

Gerçek Değer	Anlamlandırılmış Değer
1	1
1+1	11
10 Üzeri	10
2	2
2+1	12
2+2	13
3	3
3+1	14
3+2	15
4+1	16
4+2	17
5	5
5+1	18
5+2	19
5+3	20
6+1	21
6+2	22
6+3	23
7+2	24
8+1	25

**Tablo 1.** OdaSayisi Değerlerinin Gruplandırılmış Durumu

**BanyoSayisi:** Alıcıya sunulan konutun banyo sayısıdır. Banyo sayısının veri madenciliği uygulanırken gruplandırılmış değerleri aşağıdaki tabloda gösterilmiştir.

Gerçek Değer	Anlamlandırılmış Değer
1	1
2	2
3	3
4	4
5	5
Yok	0

**Tablo 2.** BanyoSayisi Değişkeninin Gruplandırılmış Durumu

**BinaYasi:** Alıcıya sunulan konutun yaşıdır. Bina yaşının veri madenciliği uygulanırken değerleri aşağıdaki tabloda gösterilmiştir.

Gerçek Değer	Anlamlandırılmış Değer
NULL	?
0	0
1	1
11-15 arası	13.5
15	15
16-20 arası	19.5
2	2
21-25 arası	23.5
26-30 arası	28.5
3	3
30	30
31 ve üzeri	31
4	4
5-10 arası	7.5

**Tablo 3.** BinaYasi Değişkeninin Gruplandırılmış Durumu

**KatSayisi:** Alıcıya sunulan konutun kat sayısıdır.

Gerçek Değer	Anlamlandırılmış Değer
1	1
10	10
11	11
15	15
2	2
20	20
21	21
3	3
4	4
5	5
6	6
7	7
8	8
9	9

**Tablo 4.** KatSayisi Değişkeninin Gruplandırılmış Durumu

**BulunduguKat:** Alıcıya sunulan konutun bulunduğu kattır. Bulunduğu kat bilgisinin veri madenciliği uygulanırken değerleri aşağıdaki tabloda gösterilmiştir.

Gerçek Değer	Anlamlandırılmış Değer
NULL	?
1	1
10	10
17	17
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
Bahçe Katı	0
Çatı Katı	0.10
Giriş Katı	0.20
Kot 1	0.30
Müstakil	0.40
Villa Tipi	0.50
Yüksek Giriş	0.60
Zemin Kat	0.70

**Tablo 5.** BulunduguKat Değişkeninin Gruplandırılmış Durumu

**Aidat:** Alıcıya sunulan konutun TL cinsinden aylık aidat ödemesidir.

**Isitma:** Alıcıya sunulan konutun ısıtma sistemi türüdür. Isıtma bilgisinin veri madenciliği uygulanırken değerleri aşağıdaki tabloda gösterilmiştir.

Gerçek Değer	Anlamlandırılmış Değer
NULL	?
Doğalgaz (Kombi)	1
Doğalgaz Sobası	2
Kat Kaloriferi	3
Merkezi Sistem	4
Merkezi Sistem (Isı Pay Ölçer)	5
Soba	6
Yerden Isıtma	7
Yok	0

**Tablo 6.** Isitma Değişkeninin Gruplandırılmış Durumu

**Eşyalı:** Alıcıya sunulan konutun eşyalı olup/olmadığı bilgisidir. Konutlar eşyalı ise 1 değilse 0 rakamı ile gösterilmiştir. Bilinmeyen değerler ? işareti ile gösterilmektedir. Bilinmeyen değerler veri madenciliği teknikleri ile tahmin edilmeye çalışılacaktır.

**KullanımDurumu:** Alıcıya sunulan konutun kullanımda olup/olmadığı bilgisidir. Kullanım durumu bilgisinin veri madenciliği uygulanırken değerleri aşağıdaki tabloda gösterilmiştir. Bilinmeyen değerler ? İşareti ile gösterilmektedir. Bilinmeyen değerler veri madenciliği teknikleri ile tahmin edilmeye çalışılacaktır.

Gerçek Değer	Anlamlandırılmış Değer
NULL	?
Boş	1
Kiracı Oturuyor	2
Kiracılı	3
Mülk Sahibi	4
Mülk Sahibi Oturuyor	5

**Tablo 7.** KullanımDurumu Değişkeninin Gruplandırılmış Durumu

**Siteİçerisinde:** Alıcıya sunulan konutun site içerisinde olup/olmadığı bilgisidir. Konutlar site içerisinde ise 1 değilse 0 rakamı ile gösterilmiştir. Bilinmeyen değerler ? işareti ile gösterilmektedir. Bilinmeyen değerler veri madenciliği teknikleri ile tahmin edilmeye çalışılacaktır.

**KrediyeUygun:** Alıcıya sunulan konutun krediye uygun olup/olmadığı bilgisidir. Konutlar krediye uygunsa ise 1 değilse 0 rakamı ile gösterilmiştir. Bilinmeyen değerler ? işareti ile gösterilmektedir.

**Kimden:** Alıcıya sunulan konutun kim tarafından satıldığını gösteren bilgisidir. Kimden bilgisinin veri madenciliği uygulanırken değerleri aşağıdaki tabloda gösterilmiştir.

Gerçek Değer	Anlamlandırılmış Değer
Emlak Ofisi	1
İnşaat Firmasından	2
Sahibinden	3
Üreticiden	4

**Tablo 8.** Kimden Değişkeninin Gruplandırılmış Durumu

**Takas:** Alıcıya sunulan konutun takas imkanı olup/olmadığını gösteren bilgisidir. Konutlar takasa uygunsa ise 1 değilse 0 rakamı ile gösterilmiştir. Bilinmeyen değerler ? işareti ile gösterilmektedir. Bilinmeyen değerler veri madenciliği teknikleri ile tahmin edilmeye çalışılacaktır.

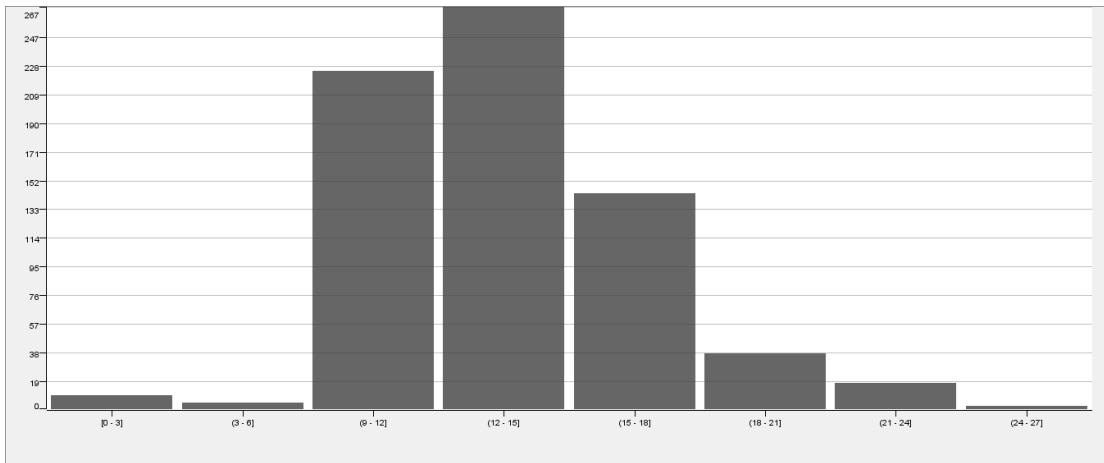
**Latitude:** Alıcıya sunulan konutun enlem bilgisidir. Bilinmeyen enlem bilgileri ? işareti ile gösterilmektedir ve bu değerler veri madenciliği teknikleri ile tahmin edilmeye çalışılacaktır.

**Longitude:** Alıcıya sunulan konutun boylam bilgisidir. Bilinmeyen boylam bilgileri ? işareti ile gösterilmektedir ve bu değerler veri madenciliği teknikleri ile tahmin edilmeye çalışılacaktır.

#### 4.1 Değişken Bazlı Grafik Analiz Çalışması

Analizde kullanılacak değişkenler veri madenciliğinin grafiksel araçları ile incelenmiş ve aşağıdaki sonuçlar elde edilmiştir. Şekil 19’da veriler üzerindeki istatistiksel bilgiler paylaşılmıştır.

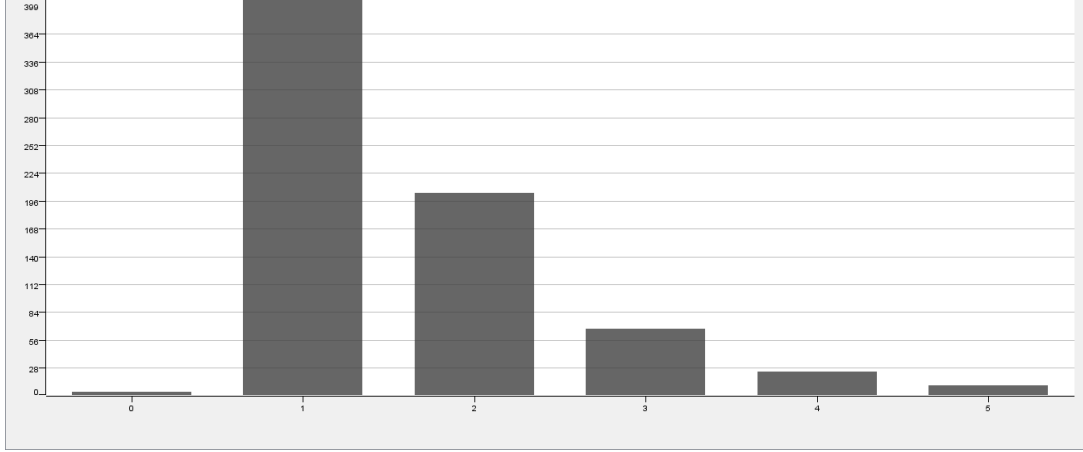
**OdaSayisi Değişkeni:** Aşağıdaki Histogram da görüldüğü gibi OdaSayisi değişkeninin büyük bir bölümü 12-15 kategorik grubu bulunmaktadır. Bunu 9-12 arası grubu kategorik takip etmektedir.



Şekil 3. OdaSayisi Değişkeninin Histogram Değeri

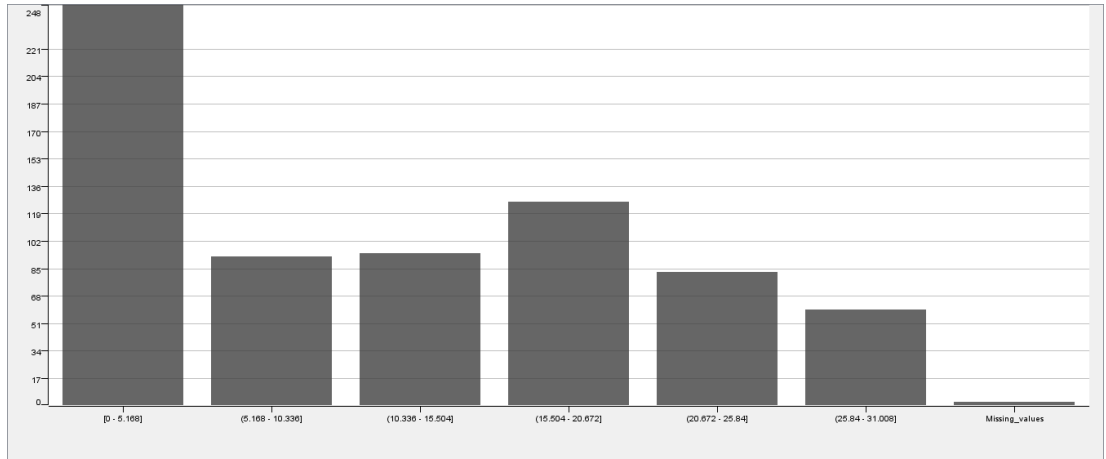


**BanyoSayisi Değişkeni:** Histogramda da görüldüğü gibi BanyoSayisi değişkeninin en büyük bölümü 1 kategorik grubunda bulunmaktadır. Bunu 2 kategorik grubu takip etmektedir.



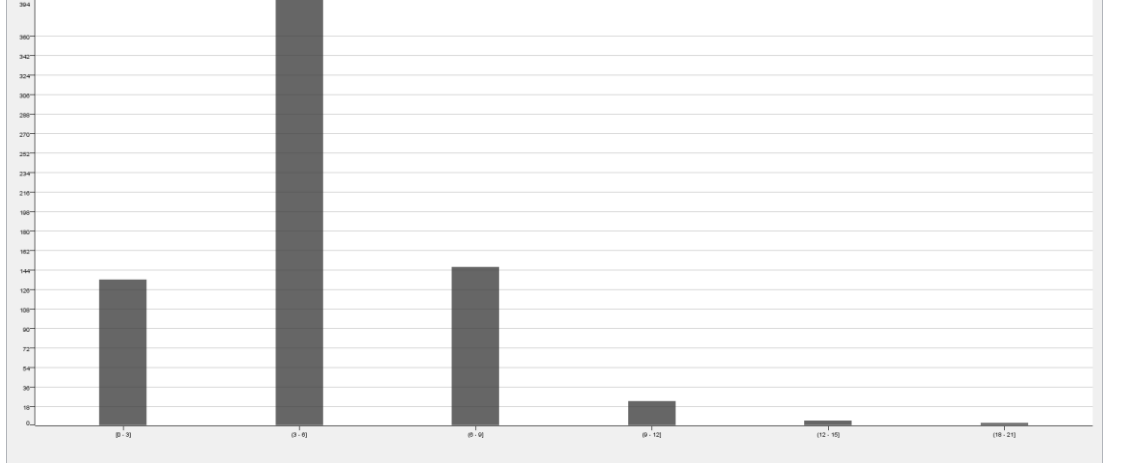
Şekil 4. BanyoSayisi Değişkeninin Histogram Değeri

**BinaYasi Değişkeni:** Histogramda da görüldüğü gibi BinaYasi değişkeninin en büyük bölümü 0-5.168 kategorik grubunda bulunmaktadır.



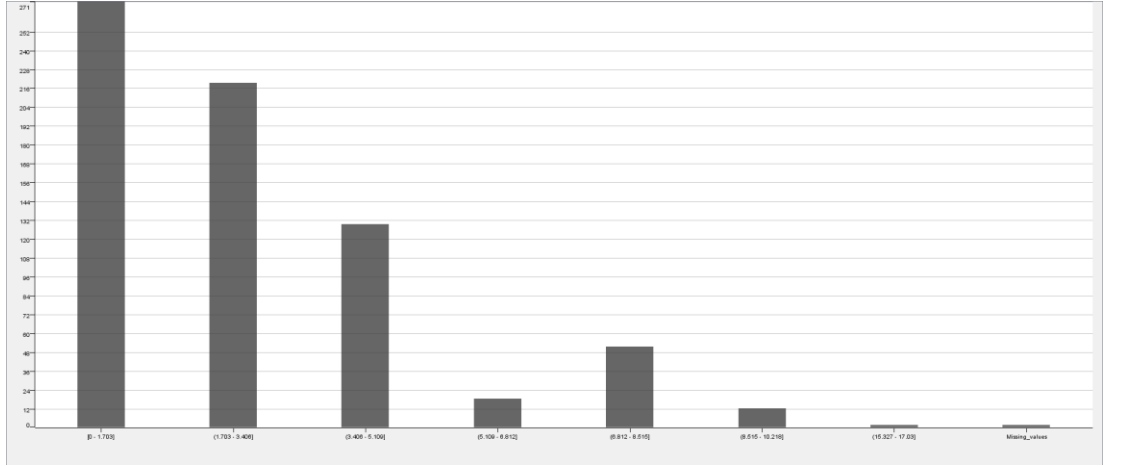
Şekil 5. BinaYasi Değişkeninin Histogram Değeri

**KatSayisi Değişkeni:** Histogramda da görüldüğü gibi KatSayisi değişkeninin en büyük bölümü 3-6 kategorik grubunda bulunmaktadır.



**Şekil 6.** KatSayisi Değişkeninin Histogram Değeri

**BulunduguKat Değişkeni:** Histogramda da görüldüğü gibi BulunduguKat değişkeninin en büyük bölümü 0-1.703 kategorik grubunda bulunmaktadır.



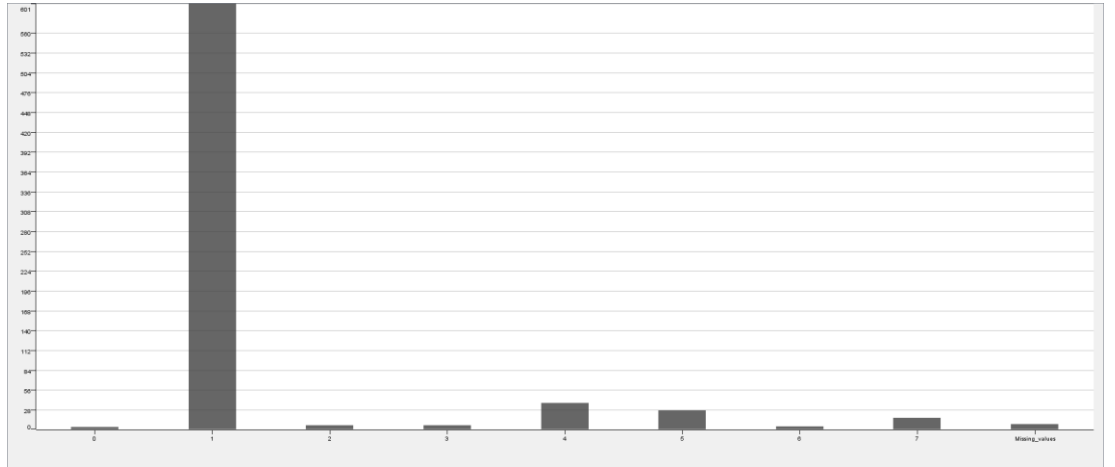
**Şekil 7.** BulunduguKat Değişkeninin Histogram Değeri

**Aidat Değişkeni:** Histogramda da görüldüğü gibi Aidat değişkeninin en büyük bölümünü kayıp değerler oluşturmaktadır. Bunu 2. Sırada 0-80 TL arası grup takip etmektedir.



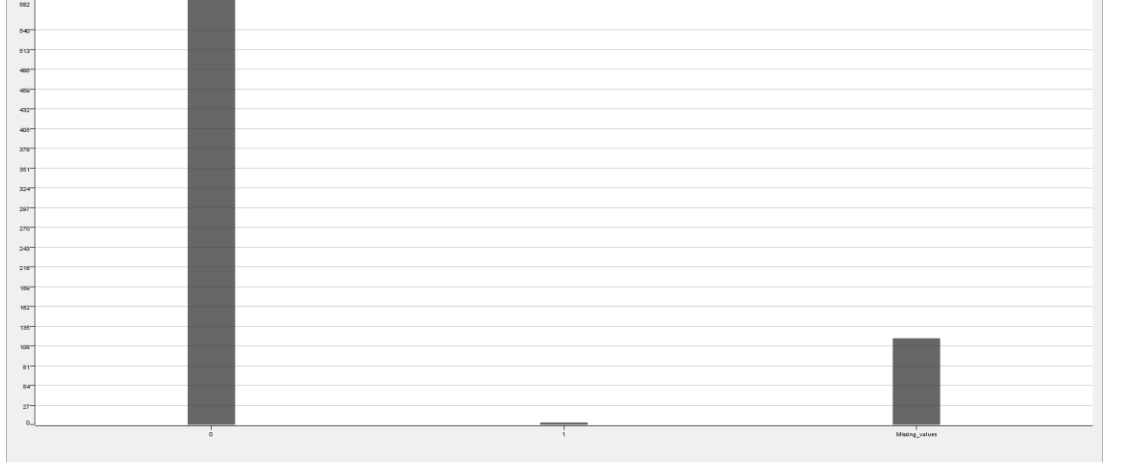
**Şekil 8.** Aidat Değişkeninin Histogram Değeri

**Isitma Değişkeni:** Histogramda da görüldüğü gibi Isitma değişkeninin en büyük bölümünü 1 kategorik grubu oluşturmaktadır.



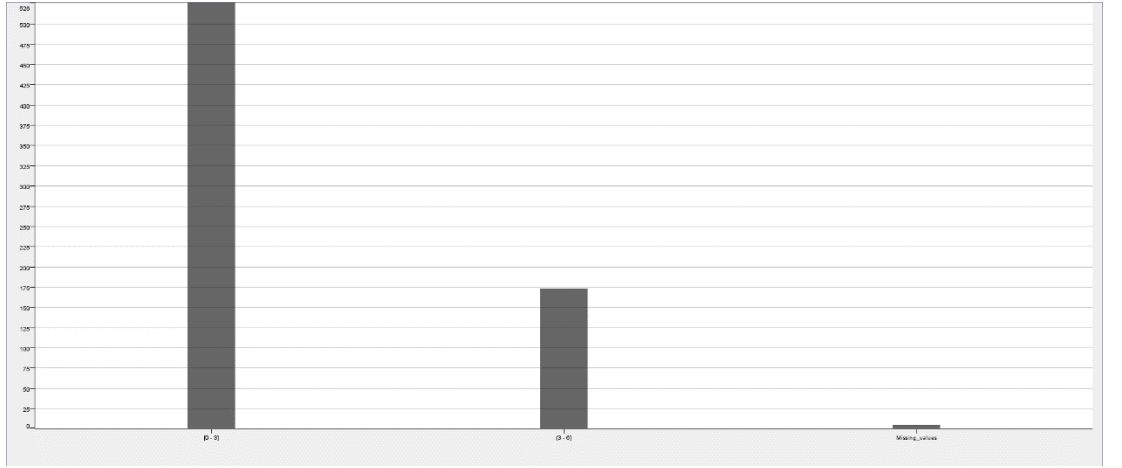
**Şekil 9.** Isitma Değişkeninin Histogram Değeri

**Esyali Değişkeni:** Histogramda da görüldüğü gibi Esyali değişkeninin en büyük bölümünü 0 kategorik grubu oluşturmaktadır.



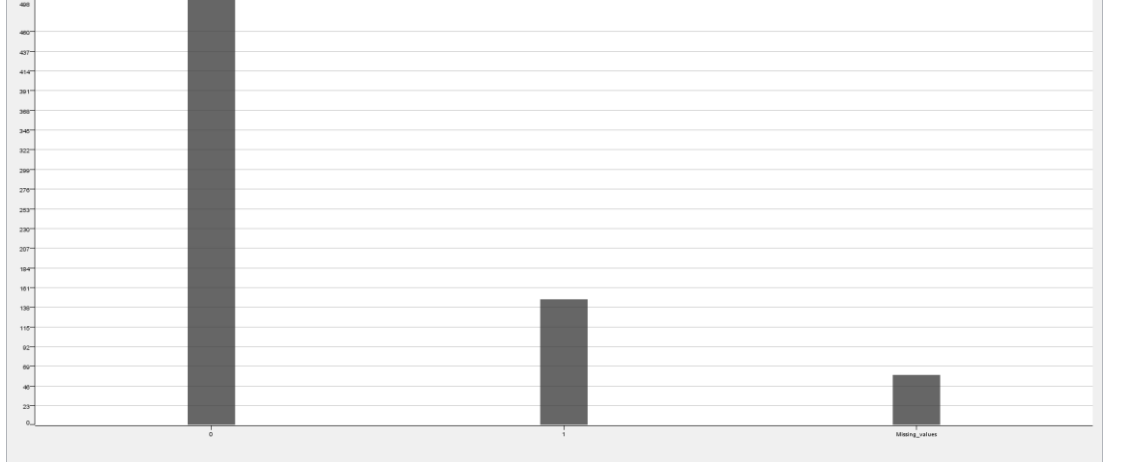
**Şekil 10.** ESYALI Değişkeninin Histogram Değeri

**KullanımDurumu Değişkeni:** Histogramda da görüldüğü gibi KullanımDurumu değişkeninin en büyük bölümünü 0-3 kategorik grubu oluşturmaktadır.



**Şekil 11.** KullanımDurumu Değişkeninin Histogram Değeri

**SiteIcerisinde Değişkeni:** Histogramda da görüldüğü gibi SiteIcerisinde değişkeninin en büyük bölümünü 0 kategorik grubu oluşturmaktadır.



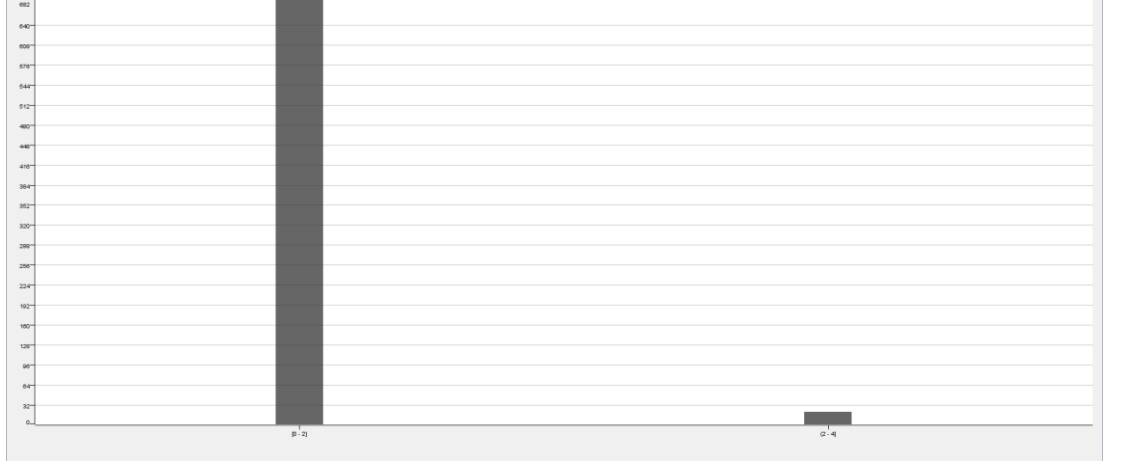
**Şekil 12.** SiteIcerisinde Değişkeninin Histogram Değeri

**KrediyeUygun Değişkeni:** Histogramda da görüldüğü gibi KrediyeUygun değişkeninin en büyük bölümünü 1 kategorik grubu oluşturmaktadır.



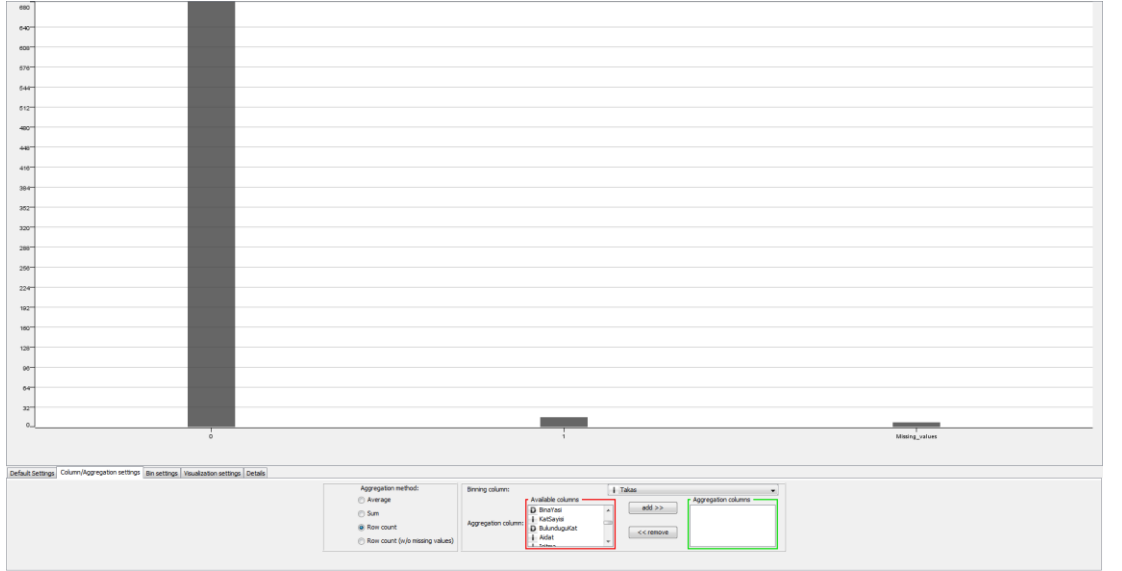
**Şekil 13.** KrediyeUygun Değişkeninin Histogram Değeri

**Kimden Değişkeni:** Histogramda da görüldüğü gibi Kimden değişkeninin en büyük bölümünü 0-2 kategorik grubu oluşturmaktadır.



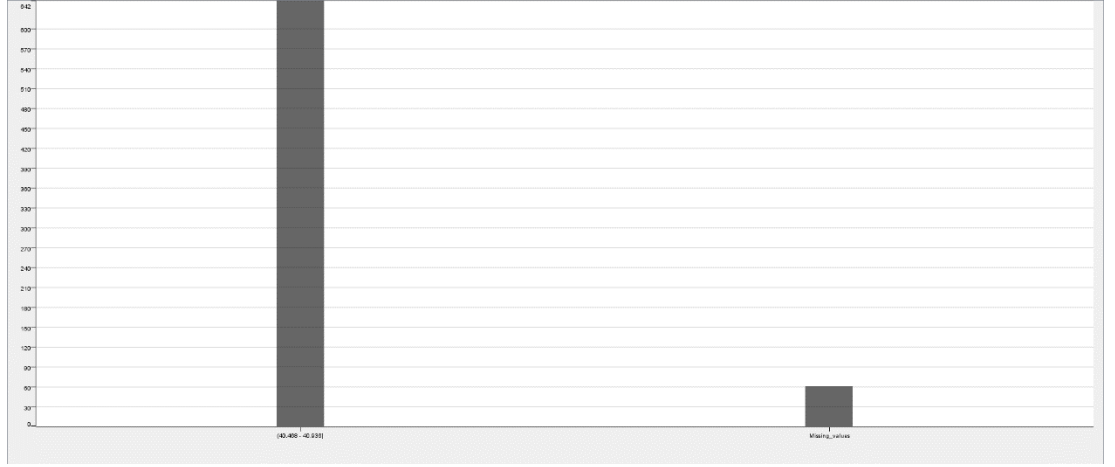
**Şekil 14.** Kimden Değişkeninin Histogram Değeri

**Takas Değişkeni:** Histogramda da görüldüğü gibi Takas değişkeninin en büyük bölümünü 0 kategorik grubu oluşturmaktadır.



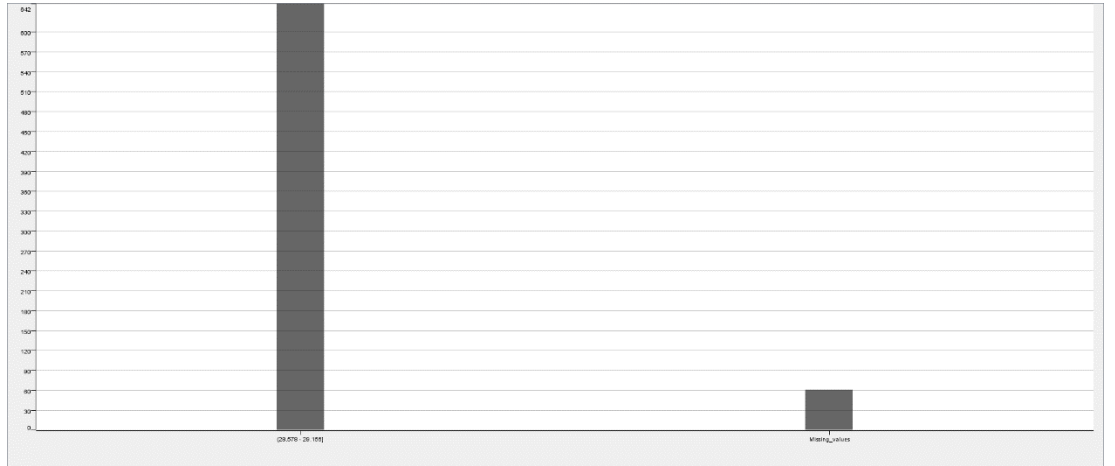
**Şekil 15.** Takas Değişkeninin Histogram Değeri

**Latitude Değişkeni:** Histogramda da görüldüğü gibi Latitude değişkeninin en büyük bölümünü 40.468-40.936 arası enlem bilgisi oluşturmaktadır.



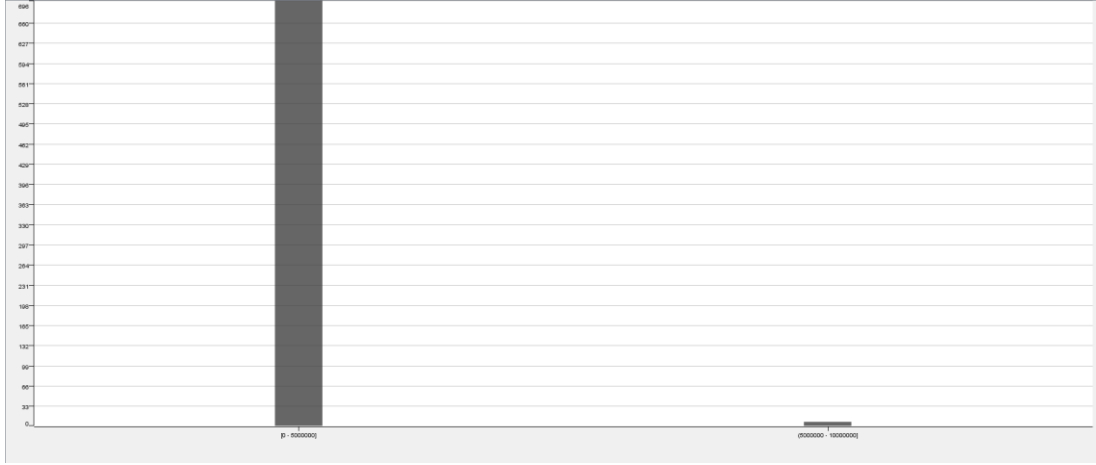
**Şekil 16.** Latitude Değişkeninin Histogram Değeri

**Longitude Değişkeni:** Histogramda da görüldüğü gibi Longitude değişkeninin en büyük bölümünü 28.578-29.156 arası boylam bilgisi oluşturmaktadır.



**Şekil 17.** Longitude Değişkeninin Histogram Değeri

**Fiyat Değişkeni:** Histogramda da görüldüğü gibi Longitude değişkeninin en büyük bölümünü 0-5000000 fiyat aralığı oluşturmaktadır.



**Şekil 18.** Fiyat Değişkeninin Histogram Değeri

Numeric columns		Nominal columns															
Row ID	D M2	D OdaSayisi	D BanyoS...	D BinaYasi	D KatSayisi	D Bulund...	D Aidat	D Isitma	D Esvall	D Kullani...	D SiteEcer...	D Krediy...	D Kmnden	D Takas	D Latitude	D Longitude	D Fiyat
Minimum	11	1	0	0	1	0	0	0	0	1	0	0	1	0	40.908	29.116	26.000
Maximum	520.000	25	5	31	21	17	760	7	1	5	1	1	4	1	40.933	29.153	10.000.000
Mean	616.856	14.047	1.824	11.757	5.151	2.631	89.767	1.496	0.003	1.933	0.227	0.899	1.06	0.022	40.918	29.133	763.653.148
Std. deviation	12.132.735	2.947	0.883	10.007	2.331	2.322	128.763	1.358	0.058	1.32	0.419	0.301	0.345	0.145	0.094	0.005	981.494.663
Variance	147.203.26...	8.687	0.78	100.13	5.432	5.394	16.579.945	1.844	0.003	1.742	0.176	0.091	0.119	0.021	0	0	963.331.65...
Overall sum	428.715	9.861	1.140	8.242	3.616	1.844.1	7.720	1.040	2	1.349	146	624	744	15	26.269.434	18.703.493	536.084.510
No. missings	7	0	0	1	0	1	616	7	118	4	58	8	7	0	60	60	0
Median	9	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
Row count	702	702	702	702	702	702	702	702	702	702	702	702	702	702	702	702	702

**Şekil 19.** Veriler Hakkında Tanımlayıcı İstatistik Bilgileri



## 5. VERİLERİN VERİ MADENCİLİĞİ TEKNİKLERİYLE KÜMELENMESİ

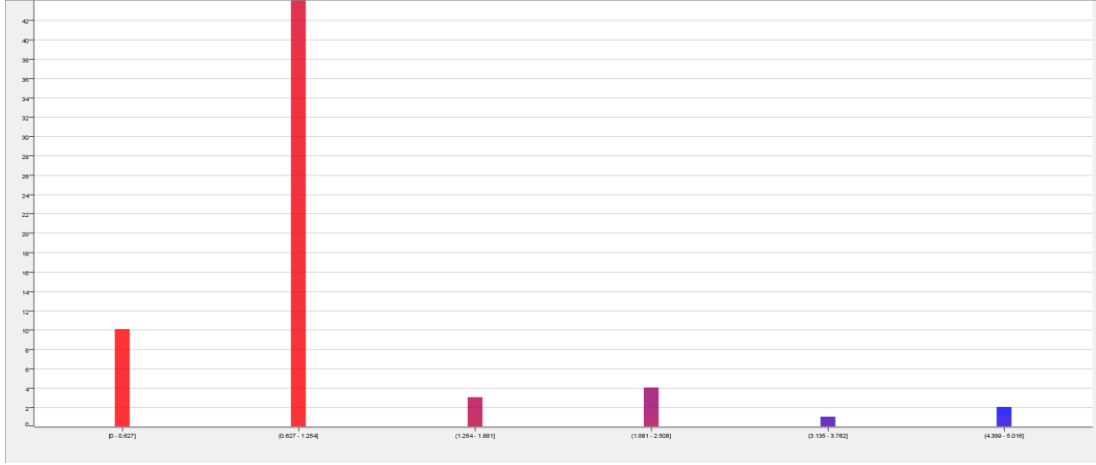
Kümeleme işlemi için iki algoritma seçilerek elde edilen sonuçlar karşılaştırılacak ve en iyi çıktıyı veren algoritma kullanılacaktır. Yüksek başarımları bilinmesinden dolayı KNIME aracı üzerinde K-Means ve Fuzzy C-Means algoritmaları kullanılarak veriler kümelere ayrılacaktır. Verilerin kümeleme işlemine girmesi için önce verilerin kaç gruba ayrılacağına bildirilmesi gerekmektedir. Bu sebeple *Fiyat* değişkenini histogram yardımıyla incelenip verilerin kaç gruba ayrılması gerektiği konusunda fikir elde edilecektir.

Verileri KNIME aracının *CSV Reader* özelliğini kullanarak *data.csv* dosyasından okunmuştur. Şekil 20’de okunan verinin bir bölümü gösterilmektedir.

Row ID	M2	OdaSayisi	Banyo...	D BinaFasi	KatSayisi	D Bulund...	Aldat	Isitma	Esyali	Kullari...	Siteicer...	Krediy...	Kimden	Takas	D Labtude	D Lor	
Row0	259	19	2	23.5	4	4	?	1	0	4	0	1	1	0	40.922	29.128	
Row1	90	12	1	31	4	0.2	?	1	0	1	0	1	1	0	?	?	
Row2	155	14	3	13.5	3	2	?	1	0	1	0	1	1	0	40.919	29.13	
Row3	90	12	1	23.5	4	4	?	1	0	4	0	1	1	0	40.92	29.133	
Row4	130	14	1	0	4	1	?	5	0	1	0	1	1	0	40.92	29.133	
Row5	193	16	2	2	4	1	?	1	0	1	1	1	1	0	40.918	29.13	
Row6	130	14	2	13.5	9	8	60	1	0	1	1	1	1	0	40.911	29.152	
Row7	165	14	2	13.5	4	2	?	1	0	1	0	0	1	0	40.918	29.131	
Row8	90	12	1	23.5	4	4	?	1	0	1	0	1	1	0	40.919	29.135	
Row9	155	14	1	28.5	3	3	150	1	?	1	0	1	1	0	40.922	29.129	
Row10	110	14	1	0	5	0	?	1	0	1	0	1	1	0	40.918	29.131	
Row11	60	11	1	0	4	0	?	1	0	1	0	1	1	0	40.926	29.124	
Row12	70	11	1	23.5	3	0.2	?	1	0	1	0	1	1	0	?	?	
Row13	160	13	2	2	5	5	?	1	?	1	0	1	1	0	40.918	29.134	
Row14	300	22	3	0	5	5	?	1	0	1	?	1	1	0	40.924	29.124	
Row15	155	14	3	13.5	3	2	0	1	0	1	0	1	1	0	40.92	29.131	
Row16	150	14	2	0	9	7	?	1	0	1	0	1	1	0	40.915	29.135	
Row17	100	12	1	1	4	2	50	1	?	4	1	1	1	0	?	?	
Row18	125	14	1	7.5	4	0.6	40	1	0	1	0	1	1	3	0	40.922	29.13
Row19	70	11	1	19.5	4	0.6	?	1	0	1	0	1	1	0	40.918	29.135	
Row20	135	14	1	19.5	4	3	?	1	0	4	0	1	1	0	40.927	29.124	
Row21	90	12	1	28.5	4	0.2	?	1	0	1	0	1	3	0	40.922	29.131	
Row22	190	15	1	3	3	3	?	1	0	4	0	1	1	0	?	?	
Row23	180	14	3	3	5	4	?	1	?	1	?	1	1	0	?	?	
Row24	200	15	1	31	4	0.4	?	2	0	4	0	1	3	0	40.922	29.129	

Şekil 20. Data.csv Dosyasından Okunan Verilerin Bir Bölümü

Veriler içinde kayıp verileri diğer değişkenlerin yardımı ile bulabilmek için KNIME içinde bulunan *Missing Value* (Kayıp Değer) özelliği kullanılmıştır. Verilerdeki gürültüyü azaltmak için *Normalizer* özelliği kullanılmıştır. Bu işlemlerden sonra verileri K-Means ve Fuzzy C-Means algoritmalarına işleme alıp elde edilen sonucu DTL incelenecektir. Histogramda boş bin grupları arttıkça modelin kolay ve incelenebilir olması açısından karar ağacını 8 gruba ayırmanın doğru olduğuna karar verilmiştir. Şekil 21’de histogram incelemesi gösterilmiştir.



Şekil 21. Fiyat Değişkeninin Histogram Üzerinde İncelenmesi

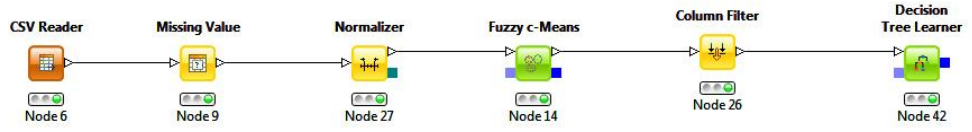
### 5.1 Fuzzy C-Means İle Kümeleme ve DTL

Fuzzy C-Means algoritmasını kullanarak veriler 8 gruba ayrılmıştır. Şekil 22’de Fuzzy algoritmasından çıkan grupların bir kısmı gösterilmektedir.

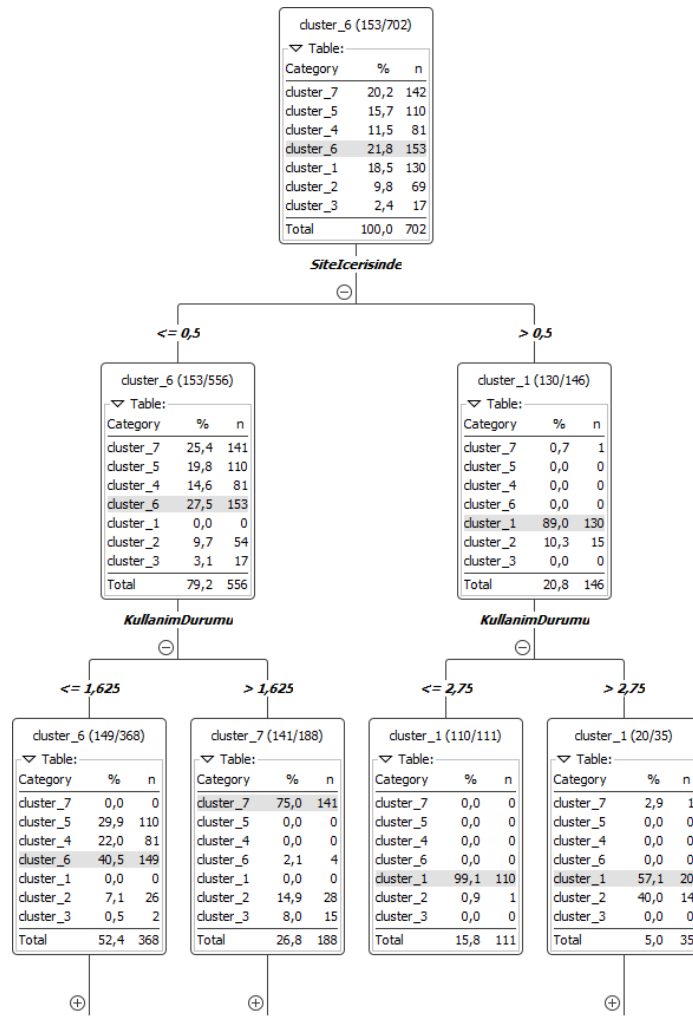
Row ID	D_M2	D_Ödeme	D_Bayesc	D_Sınas	D_KatSayı	D_Buand	D_Abit	D_Bitno	D_Fiyat	D_Rubun	D_Siteler	D_Kr...	D_Kinon	D_Telas	D_Latt	D_Longitude	D_Fiyat	D_ikuster_1	D_ikuster_2	D_ikuster_3	D_ikuster_4	D_ikuster_5	D_ikuster_6	D_ikuster_7	D_ikuster_8	Wöhner Cluste
Row0	1.901	2.975	2.3	0.911	1.175	1.039	1.033	1.140	0.5	0.879	0.5	5	0.5	0.5	1.008	1.947	0.517									cluster_7
Row1	1.901	2.962	1.4	0	1.175	1.033	1.033	1.140	0.5	0.5	0.5	5	0.5	0.5	1.004	1.96	0.59									cluster_5
Row2	1.902	2.938	3.2	2.46	0.95	1.039	1.033	1.140	0.5	0.5	0.5	5	0.5	0.5	1.021	2.202	0.739									cluster_4
Row3	1.901	2.962	1.4	0.911	1.175	1.039	1.033	1.140	0.5	0.879	0.5	5	0.5	0.5	1.008	1.947	0.517									cluster_7
Row4	1.902	2.938	1.4	0.5	1.175	1.039	1.033	1.140	0.5	0.5	0.5	5	0.5	0.5	1.004	1.96	0.59									cluster_5
Row5	1.902	2.938	1.4	0.5	1.175	1.039	1.033	1.140	0.5	0.5	0.5	5	0.5	0.5	1.004	1.96	0.59									cluster_5
Row6	1.902	2.938	2.3	0.79	1.175	1.039	1.033	1.140	0.5	0.5	0.5	5	0.5	0.5	1.004	1.96	0.59									cluster_1
Row7	1.902	2.938	2.3	2.46	1.175	1.039	1.033	1.140	0.5	0.5	0.5	5	0.5	0.5	1.021	2.202	0.739									cluster_5
Row8	1.901	2.962	1.4	0.911	1.175	1.039	1.033	1.140	0.5	0.5	0.5	5	0.5	0.5	1.004	1.96	0.59									cluster_5
Row9	1.902	2.938	1.4	0.507	0.95	1.094	1.089	1.140	0.5	0.5	0.5	5	0.5	0.5	1.015	2.02	0.724									cluster_5
Row10	1.901	2.938	1.4	0.5	1.175	1.039	1.033	1.140	0.5	0.5	0.5	5	0.5	0.5	1.004	1.96	0.59									cluster_5
Row11	1.901	2.975	1.4	0.5	1.175	1.039	1.033	1.140	0.5	0.5	0.5	5	0.5	0.5	1.004	1.96	0.59									cluster_5
Row12	1.901	2.975	1.4	0.911	1.175	1.039	1.033	1.140	0.5	0.5	0.5	5	0.5	0.5	1.004	1.96	0.59									cluster_5
Row13	1.902	2.938	2.3	0.79	1.175	1.039	1.033	1.140	0.5	0.5	0.5	5	0.5	0.5	1.004	1.96	0.59									cluster_5
Row14	1.901	2.962	1.4	0.5	1.175	1.039	1.033	1.140	0.5	0.5	0.5	5	0.5	0.5	1.004	1.96	0.59									cluster_5
Row15	1.902	2.938	3.2	2.46	0.95	1.039	1.033	1.140	0.5	0.5	0.5	5	0.5	0.5	1.021	2.202	0.739									cluster_4
Row16	1.902	2.938	2.3	0.5	1.175	1.039	1.033	1.140	0.5	0.5	0.5	5	0.5	0.5	1.004	1.96	0.59									cluster_4
Row17	1.901	2.962	1.4	0.911	1.175	1.039	1.033	1.140	0.5	0.5	0.5	5	0.5	0.5	1.004	1.96	0.59									cluster_4
Row18	1.902	2.938	1.4	1.989	1.175	1.039	1.033	1.140	0.5	0.5	0.5	5	0.5	0.5	1.004	1.96	0.59									cluster_5
Row19	1.901	2.975	1.4	0.911	1.175	1.039	1.033	1.140	0.5	0.5	0.5	5	0.5	0.5	1.004	1.96	0.59									cluster_5
Row20	1.902	2.938	1.4	0.331	1.175	1.094	1.033	1.140	0.5	0.879	0.5	5	0.5	0.5	1.043	1.472	0.639									cluster_7
Row21	1.901	2.962	1.4	0.911	1.175	1.039	1.033	1.140	0.5	0.5	0.5	5	0.5	0.5	1.004	1.96	0.59									cluster_5
Row22	1.902	2.938	1.4	0.911	1.175	1.039	1.033	1.140	0.5	0.879	0.5	5	0.5	0.5	1.004	1.96	0.59									cluster_7
Row23	1.902	2.938	3.2	0.95	1.175	1.039	1.033	1.140	0.5	0.5	0.5	5	0.5	0.5	1.004	1.96	0.59									cluster_4

Şekil 22. Fuzzy C-Means Algoritmasından Çıkan Grupların Bir Kısmı

Fuzzy C-Means algoritmasından çıkan sonuç DTL yardımıyla işleme alınmıştır. Elde edilmek istenen Fiyat değişkeninin hangi grupta olduğunu bulmak olduğundan, veriler DTL işlemine girmeden önce *Fiyat* değişkenini *Column Filter* yardımıyla veri setinden çıkartılmıştır. Şekil 23’de DTL akış şeması gösterilmiştir.



Şekil 23. Fuzzy C-Means DTL Akış Şeması



Şekil 24. DTL Elde Edilen Sonucun Bir Kısımı

Elde edilen DTL sonuçlarına göre emlak fiyatları belirleyen en büyük özelliğin SiteIcerisinde olması olduğu tespit edilmiştir. SiteIceriside olanların %21,8

ile cluster\_6 olduđu görülmüştür. Şekil 24'de DTL sonucunun bir kısmı gösterilmiştir.

Elde ettiğimiz başlıca sonuçlar şöyledir:

- **SiteIcerisinde** 0,5 eşit veya altında olup **KullanımDurumu** 1,625 eşit veya altında olanlar %40,5 **cluster\_6** grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında olup **KullanımDurumu** 1,625 üstünde olanlar %75,0 **cluster\_7** grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında olup **KullanımDurumu** 1,625 eşit veya altında olanların **BinaYasi** 2,6774 eşit veya altında olanlar %60,1 **cluster\_6** grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında olup **KullanımDurumu** 1,625 eşit veya altında olanların **BinaYasi** 2,6774 üstünde olanlar %85,8 **cluster\_5** grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında olup **KullanımDurumu** 1,625 eşit veya altında olanların **BinaYasi** 2,6774 eşit veya altında olup **M2** 0,5017 eşit veya altında olanlar %90,2 ile cluster\_6 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında olup **KullanımDurumu** 1,625 eşit veya altında olanların **BinaYasi** 2,6774 eşit veya altında olup **M2** 0,5017 üstünde olanlar %64,7 ile cluster\_4 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında olup **KullanımDurumu** 1,625 eşit veya altında olanların **BinaYasi** 2,6774 eşit veya altında olup **M2** 0,5017 eşit veya altında olanların **BinaYasi** 1,5887 eşit veya altındaysa %94,4 cluster\_6 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında olup **KullanımDurumu** 1,625 eşit veya altında olanların **BinaYasi** 2,6774 eşit veya altında olup **M2** 0,5017 eşit veya altında olanların **BinaYasi** 1,5887 üstündeyse %66,7 cluster\_5 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında olup **KullanımDurumu** 1,625 eşit veya altında olanların **BinaYasi** 2,6774 eşit veya altında olup **M2** 0,5017 eşit veya altında olanların **BinaYasi** 1,5887 eşit veya

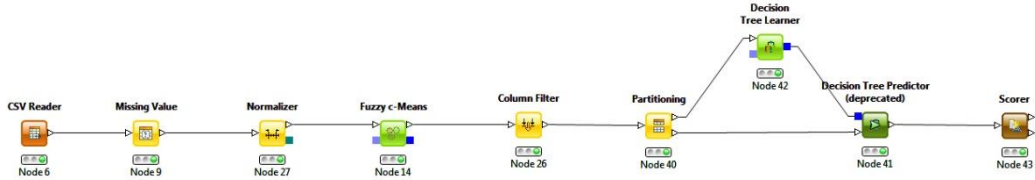
altındaysa **KrediyeUygun** 0,5 altındaysa %66,7 cluster\_2 grubundadır.

- **SiteIcerisinde** 0,5 eşit veya altında olup **KullanimDurumu** 1,625 eşit veya altında olanların **BinaYasi** 2,6774 eşit veya altında olup **M2** 0,5017 eşit veya altında olanların **BinaYasi** 1,5887 üstüneyse **KrediyeUygun** 0,5 üstüneyse %96,7 cluster\_6 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında olup **KullanimDurumu** 1,625 eşit veya altında olanların **BinaYasi** 2,6774 eşit veya altında olup **M2** 0,5017 eşit veya altında olanların **BinaYasi** 1,5887 üstüneyse **KrediyeUygun** 0,5 üstüneyse **BanyoSayisi** 1,4 eşit veya altındaysa %100 cluster\_6 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında olup **KullanimDurumu** 1,625 eşit veya altında olanların **BinaYasi** 2,6774 eşit veya altında olup **M2** 0,5017 eşit veya altında olanların **BinaYasi** 1,5887 üstüneyse **KrediyeUygun** 0,5 üstüneyse **BanyoSayisi** 1,4 üstüneyse %78,9 cluster\_6 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında olup **KullanimDurumu** 1,625 eşit veya altında olanların **BinaYasi** 2,6774 eşit veya altında olup **M2** 0,5017 eşit veya altında olanların **BinaYasi** 1,5887 üstüneyse **KrediyeUygun** 0,5 üstüneyse **BanyoSayisi** 1,4 üstüneyse **BinaYasi** 0,9355 üstüneyse %100 cluster\_4 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında olup **KullanimDurumu** 1,625 eşit veya altında olanların **BinaYasi** 2,6774 eşit veya altında olup **M2** 0,5017 eşit veya altında olanların **BinaYasi** 1,5887 üstüneyse **KrediyeUygun** 0,5 üstüneyse **BanyoSayisi** 1,4 üstüneyse **BinaYasi** 0,9355 eşit veya altındaysa %88,2 cluster\_6 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında olup **KullanimDurumu** 1,625 eşit veya altında olanların **BinaYasi** 2,6774 eşit veya altında olup **M2** 0,5017 eşit veya altında olanların **BinaYasi** 1,5887 üstüneyse **KrediyeUygun** 0,5 üstüneyse **BanyoSayisi** 1,4 üstüneyse **BinaYasi** 0,9355 eşit veya altındaysa **Latitude** 2,2705 eşit veya altındaysa %80,0 cluster\_6 grubundadır.

- **SiteIcerisinde** 0,5 eşit veya altında olup **KullanimDurumu** 1,625 eşit veya altında olanların **BinaYasi** 2,6774 eşit veya altında olup **M2** 0,5017 eşit veya altında olanların **BinaYasi** 1,5887 üstünderse **KrediyeUygun** 0,5 üstünderse **BanyoSayisi** 1,4 üstünderse **BinaYasi** 0,9355 eşit veya altındaysa **Latitude** 2,2705 üstünderse %100 cluster\_6 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında olup **KullanimDurumu** 1,625 eşit veya altında olanların **BinaYasi** 2,6774 eşit veya altında olup **M2** 0,5017 eşit veya altında olanların **BinaYasi** 1,5887 üstünderse **KrediyeUygun** 0,5 üstünderse **BanyoSayisi** 1,4 üstünderse **BinaYasi** 0,9355 eşit veya altındaysa **Latitude** 2,2705 eşit veya altındaysa **Longitude** 2,6807 eşit veya altındaysa %50,0 cluster\_4 ve cluster\_6 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında olup **KullanimDurumu** 1,625 eşit veya altında olanların **BinaYasi** 2,6774 eşit veya altında olup **M2** 0,5017 eşit veya altında olanların **BinaYasi** 1,5887 üstünderse **KrediyeUygun** 0,5 üstünderse **BanyoSayisi** 1,4 üstünderse **BinaYasi** 0,9355 eşit veya altındaysa **Latitude** 2,2705 eşit veya altındaysa **Longitude** 2,6807 üstünderse %100,0 cluster\_6 grubundadır.
- **SiteIcerisinde** 0,5 üstünder olup **KullanimDurumu** 2,75 eşit veya altında olanlar %99,1 cluster\_1 grubundadır.
- **SiteIcerisinde** 0,5 üstünder olup **KullanimDurumu** 2,75 üstünder olanlar %57,1 cluster\_1 grubundadır.
- **SiteIcerisinde** 0,5 üstünder olup **KullanimDurumu** 2,75 üstünder olanlar **BinaYasi** 1,5887 eşit veya altındaysa %100 cluster\_1 grubundadır.
- **SiteIcerisinde** 0,5 üstünder olup **KullanimDurumu** 2,75 üstünder olanlar **BinaYasi** 1,5887 üstünderse %63,6 cluster\_2 grubundadır.
- **SiteIcerisinde** 0,5 üstünder olup **KullanimDurumu** 2,75 üstünder olanlar **BinaYasi** 1,5887 üstünderse **Longitude** 2,6721 eşit veya altındaysa %53,8 cluster\_1 grubundadır.

- **SiteIcerisinde** 0,5 üstünde olup **KullanimDurumu** 2,75 üstünde olanlar **BinaYasi** 1,5887 üstündeysen **Longitude** 2,6721 eşit veya altındaysa Latitude 2,5081 eşit veya altındaysa %87,5 cluster\_1 grubundadır.
- **SiteIcerisinde** 0,5 üstünde olup **KullanimDurumu** 2,75 üstünde olanlar **BinaYasi** 1,5887 üstündeysen **Longitude** 2,6721 eşit veya altındaysa Latitude 2,5081 üstündeysen %80,0 cluster\_2 grubundadır.
- **SiteIcerisinde** 0,5 üstünde olup **KullanimDurumu** 2,75 üstünde olanlar **BinaYasi** 1,5887 üstündeysen **Longitude** 2,6721 üstündeysen %100 cluster\_2 grubundadır.

Karar ağacından çıkan sonuçlarının doğruluk tespitini yapmak için KNIME içerisindeki *Scorer* özelliğinden faydalanılmıştır. Karar ağacını test etmek için test verisine ihtiyaç vardır. Bu sebeple var olan verilerin bir kısmı DTP’de test amaçlı kullanılacaktır. KNIME içerisinde *Partition* özelliğinden yararlanılarak verilerin %80’i karar ağacında, %20’si ise karar ağacını test etmek amaçlı kullanılacaktır. Şekil 25’de *Scorer* için akış şeması verilmiştir.



**Şekil 25.** Scorer İçin Akış Şeması

DTL sonucu doğrulama işlemi sonucunda %90,78 başarı elde edildiği, %9,22 başarısız sonuç elde edildiği görülmüştür. Şekil 26’da DTL için doğrulama sonucu gösterilmiştir.

Winner Cluster \ Prediction (DecTree)	cluster_7	cluster_5	cluster_4	cluster_6	cluster_1	cluster_2	cluster_3
cluster_7	27	0	0	0	0	0	1
cluster_5	0	21	1	0	0	0	0
cluster_4	0	1	11	2	0	1	0
cluster_6	0	1	1	24	0	0	0
cluster_1	0	0	0	0	31	2	0
cluster_2	0	0	0	0	1	12	0
cluster_3	2	0	0	0	0	0	2

Correct classified: 128  
Accuracy: 90,78 %  
Cohen's kappa ( $\kappa$ ) 0,888

Wrong classified: 13  
Error: 9,22 %

Şekil 26. DTL Doğrulaması

## 5.2 K-Means İle Kümeleme ve DTL

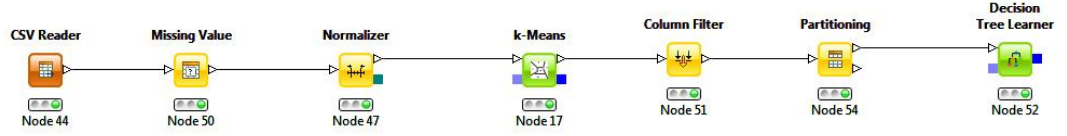
K-Means algoritmasını kullanarak veriler 8 gruba ayrılmıştır. Şekil 27’de K-Means algoritmasından çıkan grupların bir kısmı gösterilmektedir.

Row ID	D M2	D OdaSayisi	D BanyoS...	D BinaYasi	D KatSayisi	D Bülend...	D Aidat	D İstina	D Eyalı	D Kullanı...	D Siteicer...	D Krediy...	D Kİmden	D Takas	D Latitude	D Longitude	D Fiyat	S Cluster
Row0	0.503	3.875	2.3	3.911	1.175	1.559	1.033	1.143	0.5	3.875	0.5	5	0.5	0.5	3.058	1.947	0.917	cluster_3
Row1	0.501	2.562	1.4	5	1.175	0.553	1.033	1.143	0.5	0.5	5	0.5	0.5	0.5	2.404	2.58	0.59	cluster_1
Row2	0.502	2.938	3.2	2.46	0.95	1.029	1.033	1.143	0.5	0.5	5	0.5	0.5	0.5	2.621	2.202	0.739	cluster_2
Row3	0.501	2.562	1.4	3.911	1.175	1.559	1.033	1.143	0.5	3.875	0.5	5	0.5	0.5	2.798	2.555	0.621	cluster_3
Row4	0.502	2.938	1.4	0.5	1.175	0.765	1.033	3.714	0.5	0.5	0.5	5	0.5	0.5	2.798	2.555	0.682	cluster_4
Row5	0.503	3.312	2.3	0.79	1.175	0.765	1.033	1.143	0.5	0.5	5	5	0.5	0.5	2.405	2.223	1.165	cluster_5
Row6	0.502	2.938	2.3	2.46	2.3	2.618	0.855	1.143	0.5	0.5	5	5	0.5	0.5	1.204	4.85	0.646	cluster_5
Row7	0.502	2.938	2.3	2.46	1.175	1.029	1.033	1.143	0.5	0.5	0.5	0.5	0.5	0.5	2.321	2.369	0.745	cluster_7
Row8	0.501	2.562	1.4	3.911	1.175	1.559	1.033	1.143	0.5	0.5	0.5	5	0.5	0.5	2.625	2.865	0.617	cluster_1
Row9	0.502	2.938	1.4	4.637	0.95	1.294	1.388	1.143	0.5	0.5	0.5	5	0.5	0.5	3.181	2.02	0.714	cluster_1
Row10	0.501	2.938	1.4	0.5	1.4	0.5	1.033	1.143	0.5	0.5	0.5	5	0.5	0.5	2.319	2.293	0.669	cluster_4
Row11	0.501	2.375	1.4	0.5	1.175	0.5	1.033	1.143	0.5	0.5	0.5	5	0.5	0.5	3.859	1.513	0.59	cluster_4
Row12	0.501	2.375	1.4	3.911	0.95	0.553	1.033	1.143	0.5	0.5	0.5	5	0.5	0.5	2.404	2.58	0.612	cluster_1
Row13	0.502	2.75	2.3	0.79	1.4	1.824	1.033	1.143	0.5	0.5	0.5	5	0.5	0.5	3.392	2.696	0.712	cluster_4
Row14	0.504	4.438	3.2	0.5	1.4	1.824	1.033	1.143	0.5	0.5	0.5	5	0.5	0.5	3.44	1.513	1.571	cluster_2
Row15	0.502	2.938	3.2	2.46	0.95	1.029	0.5	1.143	0.5	0.5	0.5	5	0.5	0.5	2.741	2.295	0.739	cluster_2
Row16	0.502	2.938	2.3	0.5	2.3	2.353	1.033	1.143	0.5	0.5	0.5	5	0.5	0.5	1.879	2.801	1.21	cluster_2
Row17	0.501	2.562	1.4	0.645	1.175	1.029	0.796	1.143	0.5	3.875	5	5	0.5	0.5	2.404	2.58	0.678	cluster_6
Row18	0.502	2.938	1.4	1.589	1.175	0.659	0.737	1.143	0.5	0.5	0.5	5	3.5	0.5	3.053	2.233	0.646	cluster_4
Row19	0.501	2.375	1.4	3.331	1.175	0.659	1.033	1.143	0.5	0.5	0.5	5	0.5	0.5	2.345	2.759	0.612	cluster_1
Row20	0.502	2.938	1.4	3.331	1.175	1.294	1.033	1.143	0.5	3.875	0.5	5	0.5	0.5	3.943	1.473	0.639	cluster_3
Row21	0.501	2.562	1.4	4.637	1.175	0.553	1.033	1.143	0.5	0.5	0.5	5	3.5	0.5	3.016	2.312	0.59	cluster_1
Row22	0.503	3.125	1.4	0.935	0.95	1.294	1.033	1.143	0.5	3.875	0.5	5	0.5	0.5	2.404	2.58	0.849	cluster_0
Row23	0.502	2.938	3.2	0.935	1.4	1.559	1.033	1.143	0.5	0.5	0.5	5	0.5	0.5	2.404	2.58	0.793	cluster_2
Row24	0.503	3.125	1.4	5	1.175	0.606	1.033	1.786	0.5	3.875	0.5	5	3.5	0.5	3.108	2.069	1.752	cluster_3
Row25	0.502	2.938	2.3	0.5	1.175	0.765	1.033	1.143	0.5	0.5	0.5	5	0.5	0.5	1.262	4.939	0.628	cluster_4
Row26	0.505	3.875	5	0.5	1.175	1.559	1.033	3.071	0.5	0.5	0.5	5	0.5	0.5	2.404	2.58	2.401	cluster_2
Row27	0.501	2.938	1.4	0.5	1.4	1.029	1.033	3.714	0.5	0.5	0.5	5	0.5	0.5	2.477	2.728	0.7	cluster_4
Row28	0.502	2.938	1.4	5	0.95	1.294	1.033	1.143	0.5	0.5	0.5	5	0.5	0.5	3.09	1.894	1.795	cluster_1
Row29	0.501	2.938	1.4	0.5	1.625	1.824	1.033	1.143	0.5	0.5	0.5	5	0.5	0.5	2.368	2.4	0.768	cluster_4
Row30	0.501	2.562	1.4	1.589	1.4	0.659	0.737	1.143	0.5	2.75	0.5	5	3.5	5	2.655	2.611	0.644	cluster_0
Row31	0.502	3.312	2.3	3.331	2.075	1.029	0.796	2.429	0.5	2.75	0.5	5	3.5	0.5	2.083	2.535	0.791	cluster_0
Row32	0.501	2.562	1.4	1.589	1.175	0.553	1.033	1.143	0.5	0.5	0.5	5	0.5	0.5	2.24	2.507	0.594	cluster_4
Row33	0.501	2.562	1.4	3.911	1.175	1.559	1.033	1.143	0.5	3.875	0.5	5	0.5	0.5	2.798	2.555	0.622	cluster_3
Row34	0.502	2.938	2.3	2.46	1.85	1.294	1.033	1.143	0.5	0.5	0.5	5	0.5	0.5	2.357	2.505	0.691	cluster_2

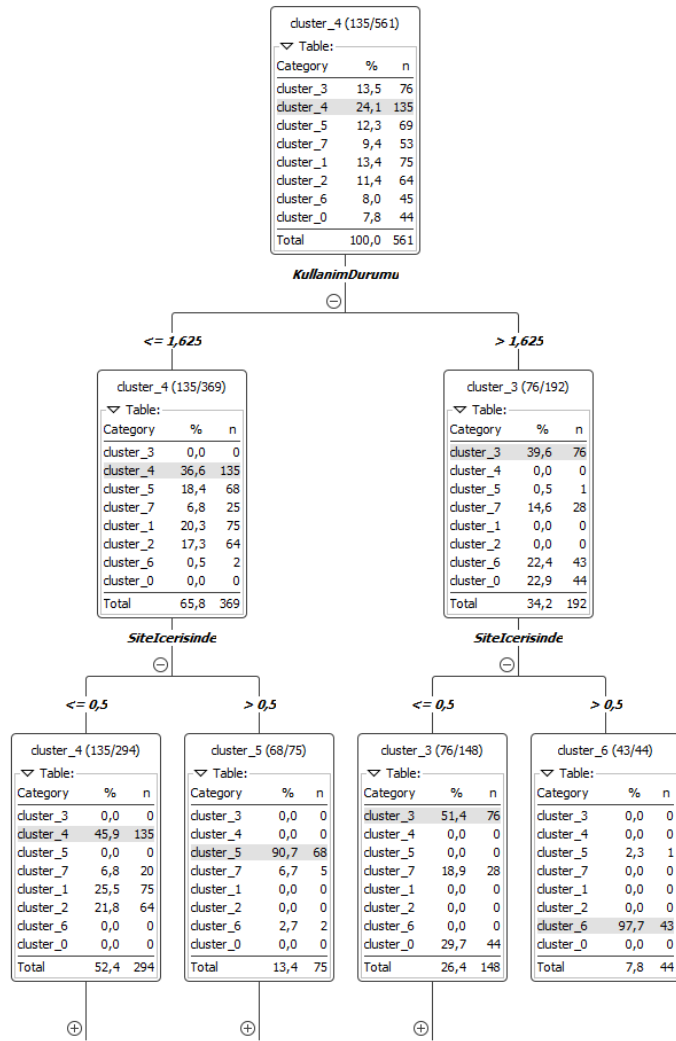
Şekil 27. K-Means Kümeleme Sonucu Oluşan Küme Yapısının Bir Parçası

K-Means algoritmasından çıkan sonucu DTL işleme alınacaktır. Elde edilmek istenen *Fiyat* değişkenin hangi grupta olduğunu bulmak olduğundan veriler DTL girmeden önce *Fiyat* değişkenini *Column Filter* yardımıyla veri setinden çıkartılmıştır. Şekil 28’de DTL akıl şeması gösterilmiştir.





Şekil 28. K-Means DTL Akış Şeması



Şekil 29. K-Means DTL Sonucunun Bir Parçası

Elde edilen DTL sonuçlarına göre emlak fiyatları belirleyen en büyük özelliğin *KullanimDurumu* olduğu tespit edilmiştir. *KullanimDurumu* %24,1 ile

cluster\_4 grubunda olduđu görülmüştür. Şekil 29’da DTL sonucunun bir kısmı gösterilmiştir.

- **KullanımDurumu** 1,625 eşit veya altında olanlar %36,6 cluster\_4 grubundadır.
- **KullanımDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa %45,9 cluster\_4 grubundadır.
- **KullanımDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 üstünderse **KrediyeUygun** 0,5 eşit veya altındaysa %83,3 cluster\_7 grubundadır.
- **KullanımDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 üstünderse **KrediyeUygun** 0,5 üstünderse %97,1 cluster\_5 grubundadır.
- **KullanımDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 üstünderse **KrediyeUygun** 0,5 üstünderse **KullanımDurumu** 0,5 eşit veya altındaysa %100 cluster\_5 grubundadır.
- **KullanımDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 üstünderse **KrediyeUygun** 0,5 üstünderse **KullanımDurumu** 0,5 üstünderse %100 cluster\_6 grubundadır.
- **KullanımDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa **BinaYasi** 2,6774 eşit veya altındaysa %64,0 cluster\_4 grubundadır.
- **KullanımDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa **BinaYasi** 2,6774 üstünderse %85,5 cluster\_1 grubundadır.
- **KullanımDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa **BinaYasi** 2,6774 üstünderse **KrediyeUygun** 0,5 eşit veya altındaysa %100 cluster\_7 grubundadır.
- **KullanımDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa **BinaYasi** 2,6774 üstünderse **KrediyeUygun** 0,5 üstünderse %98,6 cluster\_1 grubundadır.

- **KullanımDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa **BinaYasi** 2,6774 eşit veya altındaysa **BanyoSayisi** 1,4 eşit veya altındaysa %93,9 cluster\_4 grubundadır.
- **KullanımDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa **BinaYasi** 2,6774 eşit veya altındaysa **BanyoSayisi** 1,4 üstündeyse %63,9 cluster\_2 grubundadır.
- **KullanımDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa **BinaYasi** 2,6774 eşit veya altındaysa **BanyoSayisi** 1,4 eşit veya altındaysa **BinaYasi** 1,5887 eşit veya altındaysa %98,1 cluster\_4 grubundadır.
- **KullanımDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa **BinaYasi** 2,6774 eşit veya altındaysa **BanyoSayisi** 1,4 eşit veya altındaysa **BinaYasi** 1,5887 üstündeyse %50,0 cluster\_1 grubundadır.
- **KullanımDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa **BinaYasi** 2,6774 eşit veya altındaysa **BanyoSayisi** 1,4 eşit veya altındaysa **BinaYasi** 1,5887 üstündeyse **Latitude** 1,8326 eşit veya altındaysa %75,0 cluster\_4 grubundadır.
- **KullanımDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa **BinaYasi** 2,6774 eşit veya altındaysa **BanyoSayisi** 1,4 eşit veya altındaysa **BinaYasi** 1,5887 üstündeyse **Latitude** 1,8326 üstündeyse %100 cluster\_1 grubundadır.
- **KullanımDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa **BinaYasi** 2,6774 eşit veya altındaysa **BanyoSayisi** 1,4 üstündeyse **BulunduguKat** 1,2941 eşit veya altındaysa %53,2 cluster\_4 grubundadır.
- **KullanımDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa **BinaYasi** 2,6774 eşit veya altındaysa **BanyoSayisi** 1,4 üstündeyse **BulunduguKat** 1,2941 üstündeyse %88,0 cluster\_2 grubundadır.
- **KullanımDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa **BinaYasi** 2,6774 eşit veya altındaysa

**BanyoSayisi** 1,4 üstündeyse **BulunduguKat** 1,2941 eşit veya altındaysa **BanyoSayisi** 2,3 eşit veya altındaysa %69,4 cluster\_4 grubundadır.

- **KullanimDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa **BinaYasi** 2,6774 eşit veya altındaysa **BanyoSayisi** 1,4 üstündeyse **BulunduguKat** 1,2941 eşit veya altındaysa **BanyoSayisi** 2,3 üstündeyse %90,9 cluster\_2 grubundadır.
- **KullanimDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa **BinaYasi** 2,6774 eşit veya altındaysa **BanyoSayisi** 1,4 üstündeyse **BulunduguKat** 1,2941 üstündeyse **KrediyeUygun** 0,5 eşit veya altındaysa %100 cluster\_7 grubundadır.
- **KullanimDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa **BinaYasi** 2,6774 eşit veya altındaysa **BanyoSayisi** 1,4 üstündeyse **BulunduguKat** 1,2941 üstündeyse **KrediyeUygun** 0,5 üstündeyse %93,6 cluster\_2 grubundadır.
- **KullanimDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa **BinaYasi** 2,6774 eşit veya altındaysa **BanyoSayisi** 1,4 üstündeyse **BulunduguKat** 1,2941 üstündeyse **KrediyeUygun** 0,5 üstündeyse **OdaSayisi** 3,125 eşit veya altındaysa %72,7 cluster\_2 grubundadır.
- **KullanimDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa **BinaYasi** 2,6774 eşit veya altındaysa **BanyoSayisi** 1,4 üstündeyse **BulunduguKat** 1,2941 üstündeyse **KrediyeUygun** 0,5 üstündeyse **OdaSayisi** 3,125 üstündeyse %100 cluster\_2 grubundadır.
- **KullanimDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa **BinaYasi** 2,6774 eşit veya altındaysa **BanyoSayisi** 1,4 üstündeyse **BulunduguKat** 1,2941 üstündeyse **KrediyeUygun** 0,5 üstündeyse **OdaSayisi** 3,125 eşit veya altındaysa **Latitude** 2,2595 eşit veya altındaysa %100 cluster\_2 grubundadır.
- **KullanimDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa **BinaYasi** 2,6774 eşit veya altındaysa

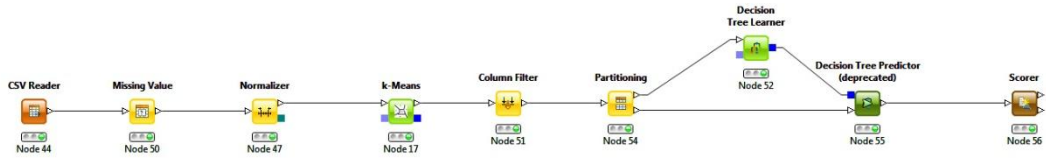
**BanyoSayisi** 1,4 üstüneyse **BulunduguKat** 1,2941 üstüneyse **KrediyeUygun** 0,5 üstüneyse **OdaSayisi** 3,125 eşit veya altındaysa **Latitude** 2,2595 üstüneyse %60 cluster\_4 grubundadır.

- **KullanımDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa **BinaYasi** 2,6774 eşit veya altındaysa **BanyoSayisi** 1,4 üstüneyse **BulunduguKat** 1,2941 üstüneyse **KrediyeUygun** 0,5 üstüneyse **OdaSayisi** 3,125 eşit veya altındaysa **Latitude** 2,2595 üstüneyse **Latitude** 2,4382 eşit veya altındaysa %100 cluster\_4 grubundadır.
- **KullanımDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa **BinaYasi** 2,6774 eşit veya altındaysa **BanyoSayisi** 1,4 üstüneyse **BulunduguKat** 1,2941 üstüneyse **KrediyeUygun** 0,5 üstüneyse **OdaSayisi** 3,125 eşit veya altındaysa **Latitude** 2,2595 üstüneyse **Latitude** 2,4382 üstüneyse %66,7 cluster\_2 grubundadır.
- **KullanımDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa **BinaYasi** 2,6774 eşit veya altındaysa **BanyoSayisi** 1,4 üstüneyse **BulunduguKat** 1,2941 eşit veya altındaysa **BanyoSayisi** 2,3 eşit veya altındaysa **BinaYasi** 1,5887 eşit veya altındaysa %83,3 cluster\_4 grubundadır.
- **KullanımDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa **BinaYasi** 2,6774 eşit veya altındaysa **BanyoSayisi** 1,4 üstüneyse **BulunduguKat** 1,2941 eşit veya altındaysa **BanyoSayisi** 2,3 eşit veya altındaysa **BinaYasi** 1,5887 üstüneyse %50,0 cluster\_7 grubundadır.
- **KullanımDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa **BinaYasi** 2,6774 eşit veya altındaysa **BanyoSayisi** 1,4 üstüneyse **BulunduguKat** 1,2941 eşit veya altındaysa **BanyoSayisi** 2,3 eşit veya altındaysa **BinaYasi** 1,5887 üstüneyse **Latitude** 2,3206 eşit veya altındaysa %100 cluster\_7 grubundadır.

- **KullanımDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa **BinaYasi** 2,6774 eşit veya altındaysa **BanyoSayisi** 1,4 üstündeyse **BulunduguKat** 1,2941 eşit veya altındaysa **BanyoSayisi** 2,3 eşit veya altındaysa **BinaYasi** 1,5887 üstündeyse **Latitude** 2,3206 üstündeyse %100 cluster\_2 grubundadır.
- **KullanımDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa **BinaYasi** 2,6774 eşit veya altındaysa **BanyoSayisi** 1,4 üstündeyse **BulunduguKat** 1,2941 eşit veya altındaysa **BanyoSayisi** 2,3 eşit veya altındaysa **BinaYasi** 1,5887 **Isitma** 1,1429 eşit veya altındaysa %96,2 cluster\_4 grubundadır.
- **KullanımDurumu** 1,625 eşit veya altında olanlar **SiteIcerisinde** 0,5 eşit veya altındaysa **BinaYasi** 2,6774 eşit veya altındaysa **BanyoSayisi** 1,4 üstündeyse **BulunduguKat** 1,2941 eşit veya altındaysa **BanyoSayisi** 2,3 eşit veya altındaysa **BinaYasi** 1,5887 **Isitma** 1,1429 üstündeyse %100 cluster\_2 grubundadır.
- **KullanımDurumu** 1,625 üstünde olanlar %39,6 cluster\_3 grubundadır.
- **KullanımDurumu** 1,625 üstünde **SiteIcerisinde** 0,5 eşit veya altında olanlar %51,4 cluster\_3 grubundadır.
- **KullanımDurumu** 1,625 üstünde **SiteIcerisinde** 0,5 üstünde olanlar %97,7 cluster\_6 grubundadır.
- **KullanımDurumu** 1,625 üstünde **SiteIcerisinde** 0,5 eşit veya altında olanlar **BinaYasi** 2,497 eşit veya altındaysa %78,2 cluster\_0 grubundadır.
- **KullanımDurumu** 1,625 üstünde **SiteIcerisinde** 0,5 eşit veya altında olanlar **BinaYasi** 2,497 üstündeyse %81,7 cluster\_3 grubundadır.
- **KullanımDurumu** 1,625 üstünde **SiteIcerisinde** 0,5 eşit veya altında olanlar **BinaYasi** 2,497 üstündeyse **KrediyeUygun** 0,5 eşit veya altındaysa %100 cluster\_7 grubundadır.
- **KullanımDurumu** 1,625 üstünde **SiteIcerisinde** 0,5 eşit veya altında olanlar **BinaYasi** 2,497 üstündeyse **KrediyeUygun** 0,5 üstündeyse %98,7 cluster\_3 grubundadır.

- **KullanımDurumu** 1,625 üstünde **SiteIcerisinde** 0,5 eşit veya altında olanlar **BinaYasi** 2,497 eşit veya altındaysa **KrediyeUygun** 0,5 eşit veya altındaysa %100 cluster\_7 grubundadır.
- **KullanımDurumu** 1,625 üstünde **SiteIcerisinde** 0,5 eşit veya altında olanlar **BinaYasi** 2,497 eşit veya altındaysa **KrediyeUygun** 0,5 üstündeyse %100 cluster\_0 grubundadır.

K-Means algoritmasının karar ağacı sonuçlarının doğruluk tespitini yapmak için tekrar *Scorer* özelliğinden kullanılacaktır. Fuzzy C-Means algoritmasında olduğu gibi K-Means algoritması içinde karar ağacımızı test etmek amaçlı *Partition* özelliğinden yararlanılarak verilerin %80 karar ağacında %20 ise karar ağacını test etmek için kullanılmıştır. Şekil 30'da *Scorer* için akış şeması verilmiştir.



Şekil 30. Scorer Akış Şeması

DTL sonucu doğrulama işlemi sonucunda %92,908 başarı elde edildiği, %7,092 başarısız sonuç elde edildiği görülmüştür. Şekil 31'de DTL için doğrulama sonucu gösterilmiştir.

Cluster \ P...	cluster_3	cluster_1	cluster_2	cluster_4	cluster_5	cluster_7	cluster_6	cluster_0
cluster_3	23	0	0	0	0	0	0	0
cluster_1	0	32	1	0	0	0	0	0
cluster_2	0	2	11	2	0	1	0	0
cluster_4	0	1	0	17	0	0	0	0
cluster_5	0	0	0	0	19	0	0	0
cluster_7	0	0	0	1	0	8	0	0
cluster_6	0	0	0	0	0	0	6	0
cluster_0	2	0	0	0	0	0	0	15

Correct classified: 131  
Accuracy: 92,908 %  
Cohen's kappa ( $\kappa$ ) 0,916

Wrong classified: 10  
Error: 7,092 %

Şekil 31. DTL Doğrulaması

### 5.3 Weka ile Class'a Etki Eden Özelliklerin Bulunması

Kümeleme işlemi sonucunda elde edilen sınıflara hangi özelliklerin etki ettiğini bulmak için Weka içindeki *AttributeSelection* özelliğinden faydalanılmıştır. Weka'da *AttributeSelection* özelliğine aşağıdaki parametre verilmiştir.

```
AttributeSelection -E "weka.attributeSelection.CfsSubsetEval -P 1 -E 1" -S  
"weka.attributeSelection.BestFirst -D 1 -N 5"
```

Fuzzy C-Means ve K-Means ile elde edilen kümeleme işlemlerine ayrı ayrı Weka'da *AttributeSelection* özelliği uygulanmıştır. Şekil 32 Ve 33'de WEKA sonuçları gösterilmiştir.

Fuzzy C-Means kümeleme işlemi sonucunda class'a etki eden özellikler aşağıda belirtilmiştir:

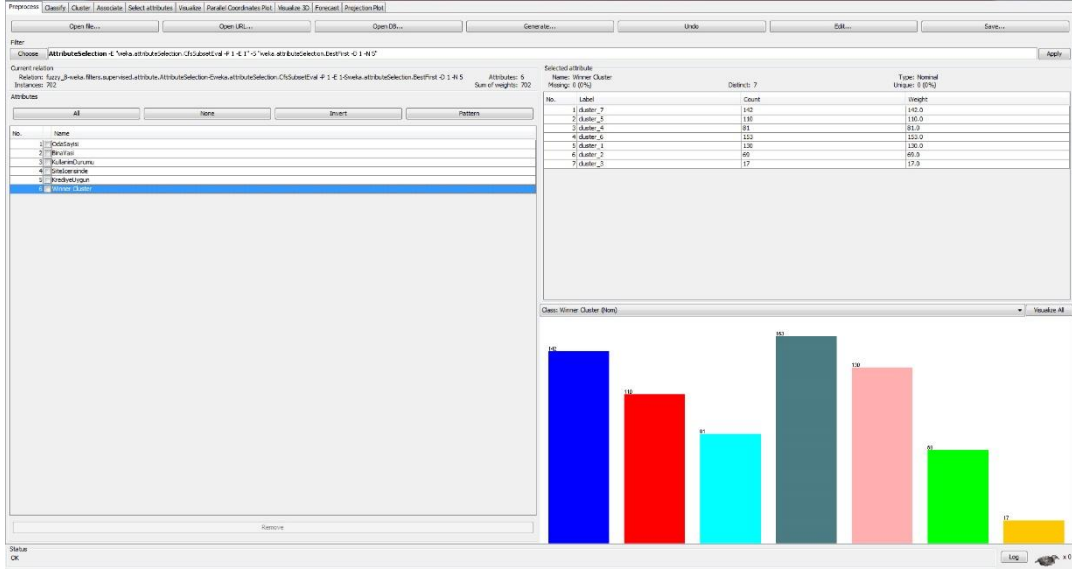
- OdaSayisi
- BinaYasi
- KullanimDurumu
- SiteIcerisinde
- KrediyeUygun
- Winner Cluster

K-Means kümeleme işlemi sonucunda class'a etki eden özellikler aşağıdaki belirtilmiştir:

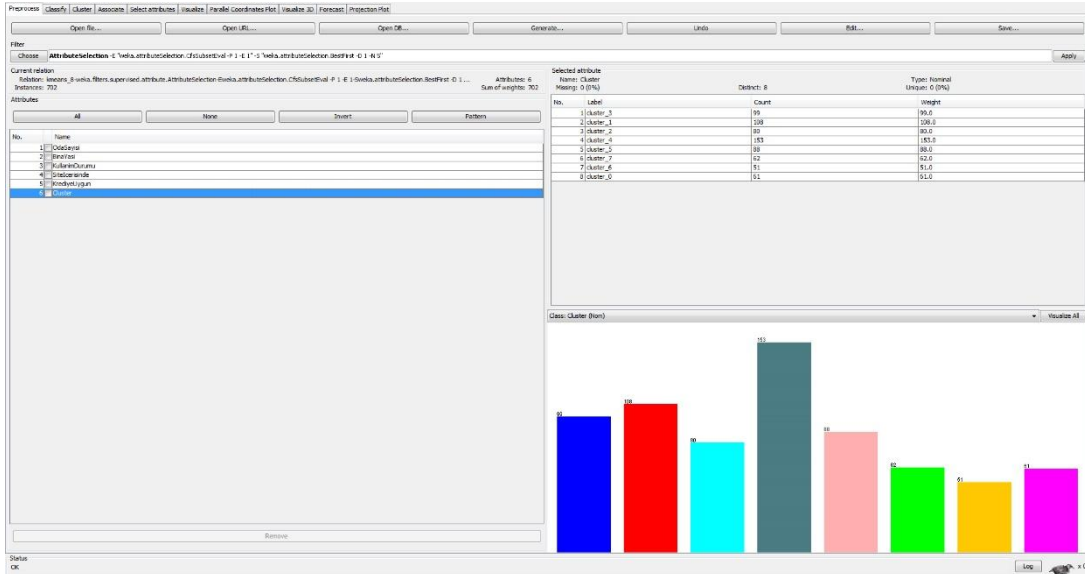
- OdaSayisi
- BinaYasi
- KullanimDurumu
- SiteIcerisinde
- KrediyeUygun
- Cluster



Her iki işlem incelendiğinde class'a etki eden özelliklerin aynı olduğu tespit edilmiştir. Veri tabanından bu değişkenleri alarak DTL yapısına verileri tekrar işleme alınarak sonuçlar tekrar değerlendirilmiştir.



Şekil 32. Weka Fuzzy C-Means AttributeSelection İşlemi Sonucu



Şekil 33. Weka K-Means AttributeSelection İşlemi Sonucu

#### 5.4 AttributeSelection İşlemi Sonrası Fuzzy C-Means İşleminin DTL Sokulması

*Colum Filter* özelliğinin yardımıyla WEKA'dan elde edilen sonuca göre class'a etki eden değişkenler kullanılarak veriler DTL işlemi alınmıştır.

Elde edilen DTL sonuçlarına göre emlak fiyatları belirleyen en büyük özelliğin *SiteIcerisinde* olduğu tespit edilmiştir. *SiteIcerisinde* olanların %21,2 ile cluster\_6 olduğu görülmüştür.

Elde ettiğimiz başlıca sonuçlar şöyledir:

- **SiteIcerisinde** 0,5 eşit veya altında ise %27,0 ile cluster\_6 grubundadır.
- **SiteIcerisinde** 0,5 üstünderse %90,0 ile cluster\_1 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında ise **KullanımDurumu** 1,625 eşit veya altında ise %38,9 ile cluster\_6 grubundadır.
- **SiteIcerisinde** 0,5 üstünderse **KullanımDurumu** 1,625 üstünderse %74,8 ile cluster\_7 grubundadır.
- **SiteIcerisinde** 0,5 üstünderse **BinaYasi** 3,3306 eşit veya altında ise %94,6 ile cluster\_1 grubundadır.
- **SiteIcerisinde** 0,5 üstünderse **BinaYasi** 3,3306 üstünderse %62,5 ile cluster\_2 grubundadır.
- **SiteIcerisinde** 0,5 üstünderse **BinaYasi** 3,3306 üstünderse **KullanımDurumu** 2,75 eşit veya altında ise %66,7 cluster\_1 grubundadır.
- **SiteIcerisinde** 0,5 üstünderse **BinaYasi** 3,3306 üstünderse **KullanımDurumu** 2,75 üstünderse %80,0 cluster\_2 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında ise **KullanımDurumu** 1,625 eşit veya altında ise **BinaYasi** 2,6774 eşit veya altında ise %58,6 cluster\_6 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında ise **KullanımDurumu** 1,625 eşit veya altında ise **BinaYasi** 2,6774 üstünderse %88,0 cluster\_5 grubundadır.

- **SiteIcerisinde** 0,5 eşit veya altında ise **KullanimDurumu** 1,625 üstündeyse **KrediyeUygun** 0,5 eşit veya altında ise %100 cluster\_2 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında ise **KullanimDurumu** 1,625 üstündeyse **KrediyeUygun** 0,5 üstündeyse %89,2 cluster\_7 grubundadır.
- **SiteIcerisinde** 0,5 üstündeyse **BinaYasi** 3,3306 eşit veya altında ise **KrediyeUygun** 0,5 eşit veya altında ise %58,3 cluster\_1 grubundadır.
- **SiteIcerisinde** 0,5 üstündeyse **BinaYasi** 3,3306 eşit veya altında ise **KrediyeUygun** 0,5 üstündeyse %99,0 cluster\_1 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında **KullanimDurumu** 1,625 eşit veya altında **BinaYasi** 2,6774 eşit veya altında **OdaSayisi** 2,9375 eşit veya altında ise %79,4 cluster\_6 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında **KullanimDurumu** 1,625 eşit veya altında **BinaYasi** 2,6774 eşit veya altında **OdaSayisi** 2,9375 üstündeyse %79,0 cluster\_4 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında **KullanimDurumu** 1,625 eşit veya altında **BinaYasi** 2,6774 üstündeyse **KrediyeUygun** 0,5 eşit veya altındaysa %100 cluster\_2 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında **KullanimDurumu** 1,625 eşit veya altında **BinaYasi** 2,6774 üstündeyse **KrediyeUygun** 0,5 üstündeyse %100 cluster\_5 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında **KullanimDurumu** 1,625 üstündeyse **KrediyeUygun** 0,5 üstündeyse **BinaYasi** 1,5887 eşit veya altındaysa %56,0 cluster\_7 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında **KullanimDurumu** 1,625 üstündeyse **KrediyeUygun** 0,5 üstündeyse **BinaYasi** 1,5887 üstündeyse %97,9 cluster\_7 grubundadır.
- **SiteIcerisinde** 0,5 üstünde **BinaYasi** 3,306 eşit veya altında **KrediyeUygun** 0,5 eşit veya altında **KullanimDurumu** 0,5 eşit veya altındaysa %100 cluster\_1 grubundadır.

- **SiteIcerisinde** 0,5 üstünde **BinaYasi** 3,306 eşit veya altında **KrediyeUygun** 0,5 eşit veya altında **KullanimDurumu** 0,5 üstünderse %100 cluster\_2 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında **KullanimDurumu** 1,625 eşit veya altında **BinaYasi** 2,6774 eşit veya altında **OdaSayisi** 2,9375 eşit veya altında **BinaYasi** 1,5887 eşit veya altındaysa %89,3 cluster\_6 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında **KullanimDurumu** 1,625 eşit veya altında **BinaYasi** 2,6774 eşit veya altında **OdaSayisi** 2,9375 eşit veya altında **BinaYasi** 1,5887 üstünderse %46,7 cluster\_4 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında **KullanimDurumu** 1,625 eşit veya altında **BinaYasi** 2,6774 eşit veya altında **OdaSayisi** 2,9375 üstünderse **KrediyeUygun** 0,5 eşit veya altında ise %100 cluster\_2 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında **KullanimDurumu** 1,625 eşit veya altında **BinaYasi** 2,6774 eşit veya altında **OdaSayisi** 2,9375 üstünderse **KrediyeUygun** 0,5 üstünderse %84,5 cluster\_4 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında **KullanimDurumu** 1,625 üstünderse **KrediyeUygun** 0,5 üstünderse **BinaYasi** 1,5887 eşit veya altında **KullanimDurumu** 2,75 eşit veya altında ise %62,5 cluster\_3 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında **KullanimDurumu** 1,625 üstünderse **KrediyeUygun** 0,5 üstünderse **BinaYasi** 1,5887 eşit veya altında **KullanimDurumu** 2,75 üstünde ise %82,4 cluster\_7 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında **KullanimDurumu** 1,625 üstünderse **KrediyeUygun** 0,5 üstünderse **BinaYasi** 1,5887 eşit veya altında **KullanimDurumu** 2,75 eşit veya altında **BinaYasi** 1,0806 eşit veya altında ise %100 cluster\_6 grubundadır.

- **SiteIcerisinde** 0,5 eşit veya altında **KullanımDurumu** 1,625 üstündeyse **KrediyeUygun** 0,5 üstündeyse **BinaYasi** 1,5887 eşit veya altında **KullanımDurumu** 2,75 eşit veya altında **BinaYasi** 1,0806 üstündese %100 cluster\_3 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında **KullanımDurumu** 1,625 üstündeyse **KrediyeUygun** 0,5 üstündeyse **BinaYasi** 1,5887 eşit veya altında **KullanımDurumu** 2,75 üstünde **OdaSayisi** 3,125 eşit veya altında ise %100 cluster\_7 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında **KullanımDurumu** 1,625 üstündeyse **KrediyeUygun** 0,5 üstündeyse **BinaYasi** 1,5887 eşit veya altında **KullanımDurumu** 2,75 üstünde **OdaSayisi** 3,125 üstünde ise %75,0 cluster\_3 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında **KullanımDurumu** 1,625 eşit veya altında **BinaYasi** 2,6774 eşit veya altında **OdaSayisi** 2,9375 eşit veya altında **BinaYasi** 1,5887 eşit veya altında **KrediyeUygun** 0,5 eşit veya altında ise %66,7 cluster\_2 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında **KullanımDurumu** 1,625 eşit veya altında **BinaYasi** 2,6774 eşit veya altında **OdaSayisi** 2,9375 eşit veya altında **BinaYasi** 1,5887 eşit veya altında **KrediyeUygun** 0,5 üstünde ise %91,5 cluster\_6 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında **KullanımDurumu** 1,625 eşit veya altında **BinaYasi** 2,6774 eşit veya altında **OdaSayisi** 2,9375 eşit veya altında **BinaYasi** 1,5887 üstünde **KrediyeUygun** 0,5 eşit veya altında ise %100 cluster\_2 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında **KullanımDurumu** 1,625 eşit veya altında **BinaYasi** 2,6774 eşit veya altında **OdaSayisi** 2,9375 eşit veya altında **BinaYasi** 1,5887 üstünde **KrediyeUygun** 0,5 üstündeyse %53,8 cluster\_4 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında **KullanımDurumu** 1,625 eşit veya altında **BinaYasi** 2,6774 eşit veya altında **OdaSayisi** 2,9375 eşit veya altında **BinaYasi** 1,5887 üstünde **KrediyeUygun** 0,5 üstündeyse **OdaSayisi** 2,5625 üstündeyse %75,0 cluster\_4 grubundadır.

- **SiteIcerisinde** 0,5 eşit veya altında **KullanimDurumu** 1,625 eşit veya altında **BinaYasi** 2,6774 eşit veya altında **OdaSayisi** 2,9375 eşit veya altında **BinaYasi** 1,5887 üstünde **KrediyeUygun** 0,5 eşit veya altında **OdaSayisi** 2,5625 eşit veya altında ise %60 cluster\_5 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında **KullanimDurumu** 1,625 eşit veya altında **BinaYasi** 2,6774 eşit veya altında **OdaSayisi** 2,9375 eşit veya altında **BinaYasi** 1,5887 üstünde **KrediyeUygun** 0,5 eşit veya altında **OdaSayisi** 2,5625 eşit veya altında **OdaSayisi** 1,25 eşit veya altında ise %50 cluster\_4 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında **KullanimDurumu** 1,625 eşit veya altında **BinaYasi** 2,6774 eşit veya altında **OdaSayisi** 2,9375 eşit veya altında **BinaYasi** 1,5887 üstünde **KrediyeUygun** 0,5 eşit veya altında **OdaSayisi** 2,5625 eşit veya altında **OdaSayisi** 1,25 üstünde ise %100 cluster\_5 grubundadır.

DTL sonucu doğrulama işlemi sonucunda %92,908 başarı elde edilmiş, %7,092 başarısız sonuç elde edildiği görülmüştür. Şekil 34’da DTL için doğrulama sonucu gösterilmiştir.

Winner Clu...	cluster_7	cluster_5	cluster_4	cluster_6	cluster_1	cluster_2	cluster_3
cluster_7	34	0	0	0	0	0	0
cluster_5	0	15	1	0	0	0	0
cluster_4	0	0	10	4	0	0	0
cluster_6	0	0	1	33	0	0	0
cluster_1	0	0	0	0	22	0	0
cluster_2	0	0	0	0	0	15	0
cluster_3	3	1	0	0	0	0	2

Correct classified: 131  
Accuracy: 92,908 %  
Cohen's kappa ( $\kappa$ ) 0,913

Wrong classified: 10  
Error: 7,092 %

**Şekil 34.** Weka Fuzzy C-Means Scorer Sonucu

## 5.5 AttributeSelection İşlemi Sonrası K-Means İşleminin DTL Sokulması

WEKA'da elde edilen bilgiler sonucunda veri tabanından veriler okunup *ColumFilter* özelliğini kullanarak class'a etkisi olmayan diğer değişkenler veri setinden çıkartılmıştır. Bu işlemin ardından DTL işlemi uygulanmıştır.

Elde edilen DTL sonuçlarına göre emlak fiyatları belirleyen en büyük özelliğin *SiteIcerisinde* olduğu tespit edilmiştir. *SiteIcerisinde* olanların %20,7 ile cluster\_4 olduğu görülmüştür. Elde ettiğimiz başlıca sonuçlar şöyledir:

- **SiteIcerisinde** 0,5 üstündeyse %60,9 cluster\_5 grubundadır.
- **SiteIcerisinde** 0,5 üstünde **KullanimDurumu** 0,5 eşit veya altında ise %91,7 cluster\_5 grubundadır.
- **SiteIcerisinde** 0,5 üstünde **KullanimDurumu** 0,5 üstünde ise %97,7 cluster\_6 grubundadır.
- **SiteIcerisinde** 0,5 üstünde **KullanimDurumu** 0,5 eşit veya altında **KrediyeUygun** 0,5 eşit veya altında ise %100 cluster\_7 grubundadır.
- **SiteIcerisinde** 0,5 üstünde **KullanimDurumu** 0,5 eşit veya altında **KrediyeUygun** 0,5 üstünde ise %100 cluster\_5 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında ise **BinaYasi** 2,6774 eşit veya altındaysa %47,73 cluster\_4 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında ise **BinaYasi** 2,6774 üstünde ise %43,6 cluster\_1 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında ise **BinaYasi** 2,6774 eşit veya altındaysa **KullanimDurumu** 1,625 eşit veya altında ise %61,7 cluster\_4 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında ise **BinaYasi** 2,6774 eşit veya altındaysa **KullanimDurumu** 1,625 üstündeyse %82,5 cluster\_0 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında ise **BinaYasi** 2,6774 üstünde **KullanimDurumu** 0,5 eşit veya altında ise %82,8 cluster\_1 grubundadır.

- **SiteIcerisinde** 0,5 eşit veya altında ise **BinaYasi** 2,6774 üstünde **KullanimDurumu** 0,5 üstünde ise %84,3 cluster\_3 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında ise **BinaYasi** 2,6774 üstünde **KullanimDurumu** 0,5 üstünde **KrediyeUygun** 0,5 eşit veya altında ise %100 cluster\_7 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında ise **BinaYasi** 2,6774 üstünde **KullanimDurumu** 0,5 üstünde **KrediyeUygun** 0,5 üstünde ise %97,4 cluster\_3 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında ise **BinaYasi** 2,6774 üstünde **KullanimDurumu** 0,5 eşit veya altında **KrediyeUygun** 0,5 eşit veya altında ise %100 cluster\_7 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında ise **BinaYasi** 2,6774 üstünde **KullanimDurumu** 0,5 eşit veya altında **KrediyeUygun** 0,5 üstünde ise %96,5 cluster\_1 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında ise **BinaYasi** 2,6774 üstünde **KullanimDurumu** 0,5 eşit veya altında **KrediyeUygun** 0,5 üstünde **OdaSayisi** 3,5 eşit veya altında ise %100 cluster\_1 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında ise **BinaYasi** 2,6774 üstünde **KullanimDurumu** 0,5 eşit veya altında **KrediyeUygun** 0,5 üstünde **OdaSayisi** 3,5 üstünde ise %100 cluster\_2 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında ise **BinaYasi** 2,6774 eşit veya altındaysa **KullanimDurumu** 1,625 üstünde **KrediyeUygun** 0,5 eşit veya altında ise %100 cluster\_7 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında ise **BinaYasi** 2,6774 eşit veya altındaysa **KullanimDurumu** 1,625 üstünde **KrediyeUygun** 0,5 üstünde ise %100 cluster\_0 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında ise **BinaYasi** 2,6774 eşit veya altındaysa **KullanimDurumu** 1,625 eşit veya altında **OdaSayisi** 3,125 eşit veya altında ise %82,0 cluster\_4 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında ise **BinaYasi** 2,6774 eşit veya altındaysa **KullanimDurumu** 1,625 eşit veya altında **OdaSayisi** 3,125 üstündeyse %78,2 cluster\_2 grubundadır.



- **SiteIcerisinde** 0,5 eşit veya altında ise **BinaYasi** 2,6774 eşit veya altındaysa **KullanımDurumu** 1,625 eşit veya altında **OdaSayisi** 3,125 eşit veya altında **BinaYasi** 2,2067 eşit veya altında ise %91,4 cluster\_4 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında ise **BinaYasi** 2,6774 eşit veya altındaysa **KullanımDurumu** 1,625 eşit veya altında **OdaSayisi** 3,125 eşit veya altında **BinaYasi** 2,2067 üstünde ise %35,3 cluster\_1 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında ise **BinaYasi** 2,6774 eşit veya altındaysa **KullanımDurumu** 1,625 eşit veya altında **OdaSayisi** 3,125 eşit veya altında **BinaYasi** 2,2067 üstünde **KrediyeUygun** 0,5 eşit veya altında ise %100 cluster\_7 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında ise **BinaYasi** 2,6774 eşit veya altındaysa **KullanımDurumu** 1,625 eşit veya altında **OdaSayisi** 3,125 eşit veya altında **BinaYasi** 2,2067 üstünde **KrediyeUygun** 0,5 üstünde ise %42,9 cluster\_1 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında ise **BinaYasi** 2,6774 eşit veya altındaysa **KullanımDurumu** 1,625 eşit veya altında **OdaSayisi** 3,125 eşit veya altında **BinaYasi** 2,2067 eşit veya altında ise **OdaSayisi** 2,1875 eşit veya altında ise %60,0 cluster\_2 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında ise **BinaYasi** 2,6774 eşit veya altındaysa **KullanımDurumu** 1,625 eşit veya altında **OdaSayisi** 3,125 eşit veya altında **BinaYasi** 2,2067 eşit veya altında ise **OdaSayisi** 2,1875 üstünde ise %93,7 cluster\_4 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında ise **BinaYasi** 2,6774 eşit veya altındaysa **KullanımDurumu** 1,625 eşit veya altında **OdaSayisi** 3,125 eşit veya altında **BinaYasi** 2,2067 eşit veya altında ise **OdaSayisi** 2,1875 eşit veya altında **OdaSayisi** 0,875 eşit veya altında ise %100 cluster\_4 grubundadır.
- **SiteIcerisinde** 0,5 eşit veya altında ise **BinaYasi** 2,6774 eşit veya altındaysa **KullanımDurumu** 1,625 eşit veya altında **OdaSayisi** 3,125 eşit veya altında **BinaYasi** 2,2067 eşit veya altında ise

**OdaSayisi** 2,1875 eşit veya altında **OdaSayisi** 0,875 üstünde ise %100 cluster\_2 grubundadır.

DTL işleminin *Scorer* özelliği kullanılarak doğrulanması sonucunda %88,652 başarılı sonuç, %11,348 başarısız sonuç elde edildiği görülmüştür. Şekil 35’da DTL işlemini için doğrulama sonucu gösterilmiştir.

Cluster \ P...	cluster_3	cluster_1	cluster_2	cluster_4	cluster_5	cluster_7	cluster_6	cluster_0
cluster_3	24	0	0	0	0	0	0	0
cluster_1	1	18	0	0	0	0	0	0
cluster_2	0	1	12	7	0	0	0	0
cluster_4	0	0	3	34	0	0	0	0
cluster_5	0	0	0	0	9	1	0	0
cluster_7	0	0	0	2	0	9	0	0
cluster_6	0	0	0	0	0	0	8	0
cluster_0	1	0	0	0	0	0	0	11

Correct classified: 125  
Accuracy: 88,652 %  
Cohen's kappa ( $\kappa$ ) 0,864

Wrong classified: 16  
Error: 11,348 %

**Şekil 35.** Weka K-Means Scorer Sonucu

## 5.6 Kümeleme Algoritmaları

Kümeleme analizi, bir veri kümesindeki bilgileri belirli yakınlık ölçütlerine göre gruplara ayırma işlemidir. Kümeleme işleminde küme içindeki elemanların benzerliği fazla, kümeler arası benzerlik ise az olmalıdır. Kümeleme analizi, bireylerin ya da nesnelerin sınıflandırılmasını ayrıntılı bir şekilde açıklamak amacıyla geliştirilmiştir. Bu amaca yönelik olarak, ele alınan örnekte yer alan varlıklar aralarındaki benzerliklere göre gruplara ayrılır, daha sonra bu gruplara giren bireylerin görünüşü ortaya konur. Diğer bir hedef ise, benzer elemanların gruplanmasıyla veri setini küçültmektir [11].

### 5.6.1 K-Means Algoritması

En yaygın kullanılan gözetimsiz öğrenme yöntemlerinden biridir. K-Means'in atama mekanizması her verinin sadece bir kümeye ait olabilmesine izin verir. Bu nedenle, keskin bir kümeleme algoritmasıdır. K-Means algoritmasının genel mantığı n adet veri nesnesinden oluşan bir veri kümesini, giriş parametresi olarak verilen k adet kümeye bölümlenektir. Amaç, gerçekleştirilen bölümlenme işlemi sonunda elde edilen kümelerin, küme içi benzerliklerinin maksimum ve kümeler arası benzerliklerinin minimum olmasını sağlamaktır. Çalışma yönteminde, denklem 5.'de Öklid uzaklığı formülü temel alınarak kümeleme yapılmaktadır.

$$\text{---} = \text{---} \quad (5)$$

K-Means Algoritmasının İşlem Basamakları Aşağıdaki gibidir.

1. k adet nesneyi rastgele seç.

$$\text{---} \quad (6)$$

2. Karesel Hata Formülü formüldeki aşağıdaki gibi hesaplanır.

$$(7)$$

K kümesini içeren bütün kümeler uzayı için kare-hata, küme içindeki değişmelerin toplamıdır. O halde söz konusu kare-hata değeri formül aşağıdaki gibi hesaplanır:

(8)

3. Her bir veriyi kendisine en yakın kümeye ata.
4. Verilerin hepsi en yakın kümelere atandığında tekrar k tane küme için merkezleri hesapla.
5. Küme Merkezlerinde bir değişiklik olmayıncaya kadar 2. ve 3. Adımları tekrarla [12].

### 5.6.2 Fuzzy C-Means Algoritması

Fuzzy C-Means algoritması, bulanık bölünmeli kümeleme tekniklerinden en iyi bilinen ve en yaygın kullanılan yöntemidir. Fuzzy C-Means algoritması 1973 yılında Dunn tarafından ortaya atılmış ve 1981’ de Bezdek tarafından geliştirilmiştir. Fuzzy C-Means algoritması da amaç fonksiyonu temelli bir metottur. Fuzzy C-Means metodu, nesnelere iki veya daha fazla kümeye ait olabilmesine izin verir. Bulanık mantık prensibi gereği her veri, kümelerin her birine  $[0,1]$  arasında değişen birer üyelik değeri ile aittir. Bir verinin tüm sınıflara olan üyelik değerleri toplamı “1” olmalıdır. Nesne hangi küme merkezine yakın ise o kümeye ait olma üyeliği diğer kümelere ait olma üyeliğinden daha büyük olacaktır. Amaç fonksiyonun belirlenen minimum ilerleme değerine yakınsak aşmasıyla kümeleme işlemi tamamlanır.

Algoritma, en küçük kareler yönteminin genellemesi olan aşağıdaki amaç fonksiyonunu öteleyerek minimize etmek için çalışır.

(9)

U üyelik matrisi rasgele atanarak algoritma başlatılır. İkinci adımda ise merkez vektörleri hesaplanır.

(10)

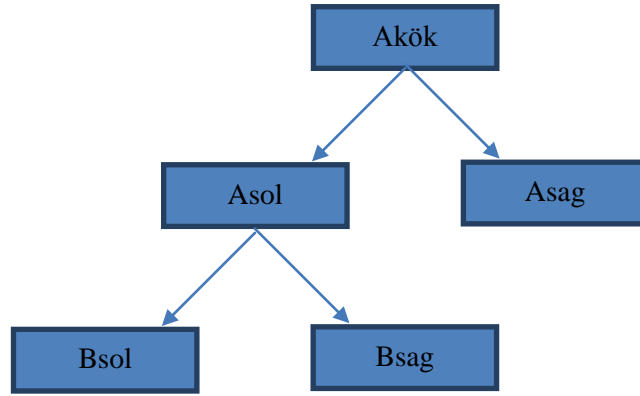
Hesaplanan küme merkezlerine göre, U matrisi aşağıdaki eşitlik kullanılarak yeniden hesaplanır. Eski U matrisi ile yeni U matrisi karşılaştırılır ve fark  $\epsilon$ 'dan küçük olana kadar işlemler devam eder.

(11)

Kümeleme işlemi sonucunda bulanık değerler içeren U üyelik matrisi kümelemenin sonucunu yansıtır. İstenirse, berraklaştırma yapılarak bu değerler yuvarlanıp 0 ve 1'lere dönüştürülebilir [13].

### **5.7 Sınıflandırma Algoritması**

Sınıflandırma en çok bilinen veri madenciliği tekniklerindedir. Sınıflandırma algoritmaları, veri tabanında gizli kalmış kuralları ortaya çıkararak veri tabanında kaydı bulunmayan bir kaydın hangi sınıfın üyesi olup olmayacağına karar verir. Karar ağaçları sınıflandırma işlemlerinde yaygın kullanılan algoritmalarındandır. Karar ağaçlarının yapılandırılması ve anlaşılması daha kolaydır.



**Şekil 36.** Örnek Karar Ağacı

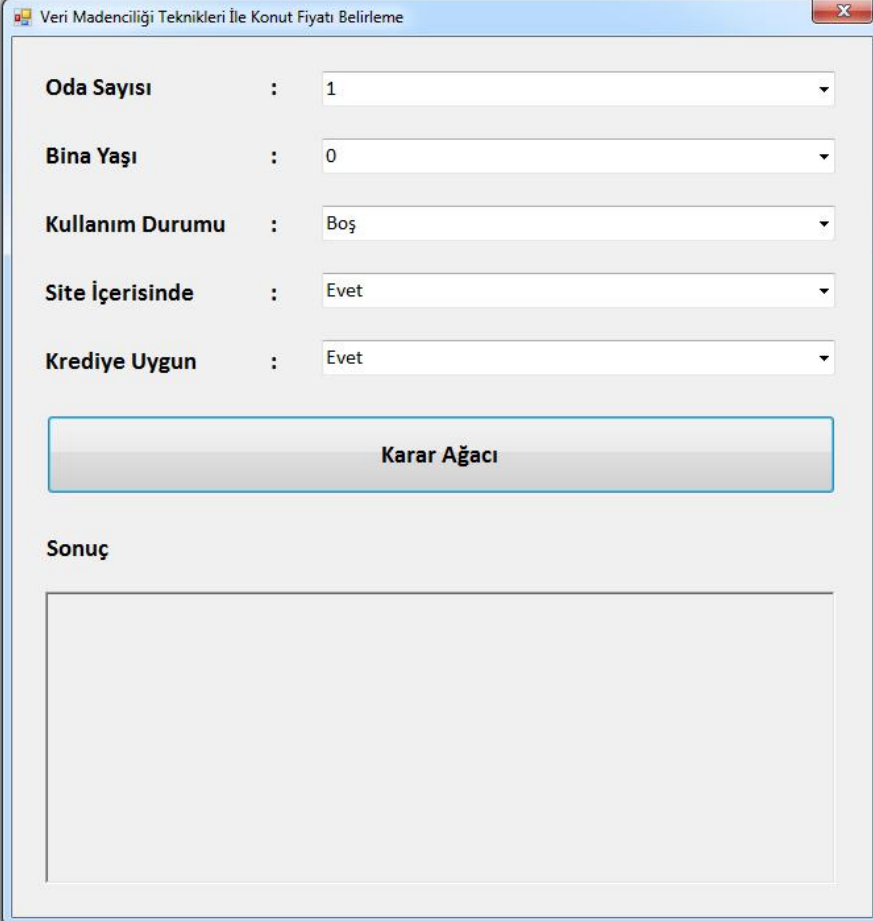
Karar ağaçları oluşturulurken kullanılan algoritmanın ne olduğu önemlidir. Kullanılan algoritmalar ağacın şeklini değiştirebilir. Bu sebeple şekli değişen ağaç farklı sınıflandırma sonuçları verebilecektir.

CART karar ağacı oluşturmada kullanılan yaygın bir algoritmadır. İkili bir ağaç yapısı vardır ve ana düğümden iki yavru düğüm oluşur. CART, sayısal ve nominal değerler üzerinde çalışabilir. CART algoritması dallara ayırma işlemi yaparken kayıp verileri önemsemez, hızlı ve esnek bir yapıya sahip algoritmadır. Denklem 12’de CART algoritması gösterilmiştir.

(12)

## 6. UYGULAMA

Microsoft Visual Studio.NET uygulama geliştirme ortamında C# dili kullanılarak DTL'den elde edilen çıktılara göre bir uygulama geliştirilmiştir. Uygulama *OdaSayisi*, *BinaYasi*, *KullanimDurumu*, *SiteIcerisinde*, *KrediyeUygun* değişkenlerini giriş değeri olarak yeni girilen bir konutun hangi sınıfta olduğunu tespit etmektedir. Şekil 38'de uygulama arayüzü gösterilmiştir.



Veri Madenciliği Teknikleri İle Konut Fiyatı Belirleme

Oda Sayısı : 1

Bina Yaşı : 0

Kullanım Durumu : Boş

Site İçerisinde : Evet

Krediye Uygun : Evet

Karar Ağacı

Sonuç

Şekil 38. Örnek Uygulama Arayüzü

sahibinden.com sitesinden Yalı Mahallesinde satışa sunulan bir konut alınmış test amaçlı kullanılmıştır. Karar ağacına **OdaSayisi**: 2+1, **BinaYasi**: 31 ve üzeri, **KullanimDurumu**: Boş, **SiteIcerisinde**: Hayır, **KrediyeUygun**: Evet bilgileri sonucunda %94,1 ile cluster\_0 grubunda olduğu olduğu tespit edilmiştir.

## SONUÇ

Yapılan çalışmadaki temel amaç özet kısmında da belirtildiği gibi emlak fiyatlarının belirlenmesinde veri madenciliği teknikleri kullanılarak emlak fiyatlarının değer aralığının bulunmasını sağlamaktır. Emlak sektörü gerek dünya çapında gerekse ülkemizde giderek ivme kazanmaktadır. Emlak fiyatının doğru belirlenmesi tüketici ve satıcı açısından önem arz etmektedir.

Yapılan çalışmalar sahibinden.com emlak sitesi üzerinden alınmış veriler ile yapılmıştır. Çalışmalar üzerinde veri madenciliği, kümele algoritmaları, karar ağaçları ve istatistiksel işlemler uygulanmıştır. Projenin veri madenciliği kısmında hesaplamalar doğru bir şekilde yapılmış ve başarılı sonuçlar alınmıştır. Tezde, özet kısmında da belirtildiği üzere İstanbul ili Maltepe ilçesi Yalı Mahallesi pilot bölge olarak seçilmiştir. Burada satışa sunulan emlak fiyatlarının kümelenmesi ve fiyatlarının belirlenmesinde farklı kümeleme algoritmaları kullanılmıştır. Analiz içerisinde son kullanılan teknik karar ağacı olmuştur.

Yapılan kümeleme işlemleri sonrasında K-Means ile kümele işleminin DTL sonucundan %92,908 başarı elde edilmiştir. Fuzzy C-Means kümeleme sonucuna DTL uyguladığımızda %90,78 başarı elde edilmiştir. Weka'da AttributeSelection özelliğini kullanarak class'a etki eden özellikler tespit edilmiştir. K-Means kümeleme işleminin DTL sokulmasından %88,652 başarı elde edildiği görülmüştür. Fuzzy C-Means işleminin sonucunda %92,908 başarı elde edilmiştir. Değişken sayısı azaltılarak konut fiyatlarının belirlenmesinde Fuzzy C-Means ile yüksek başarı elde edilmiştir.

Bu tezin devamı getirilebilir durumdadır. Karar ağacında çıkan sonuçların çeşitli yazılım araçları ile fiyatlandırmaya yansıtılması fiyatlandırmada önemli iyileştirmeler sağlayacaktır. Sadece sahibinden.com verileri üzerinde değil, aynı zamanda emlakçılar ve diğer emlak site verileri üzerinde de genişletilebilmesi söz konusu olabilir. Burada önemli olan karar ağacının işletmenin kullanacağı yazılıma yansıtılacak hale getirmektir. Bu açıdan tüm etkenlerin netleştirilmesi ve bunların ilgili yazılıma doğru şekilde aktarılması gerekmektedir.



## KAYNAKLAR

1. Doğru Fiyat Belirleme Kararları  
<http://www.arguden.net/tr/makaleler/fiyat-belirleme/>, 05 Nisan 2014
2. Alınacak konutun fiyatını belirlemenin 35 kriteri var  
<http://www.milliyet.com.tr/2005/02/21/ekonomi/axeko02.html>, 05 Nisan 2014
3. Akpınar, Haldun, “*Data: Veri Madenciliği Veri Analizi*”, İstanbul, Papatya Yayıncılık, 2014
4. Kalaycı, Şeref, “*SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri*”, Ankara, Asil Yayın Dağıtım, 2010, s.320.
5. Kümeleme Analizi ve Bir Örnek  
<http://80.251.40.59/veterinary.ankara.edu.tr/sgurcan/webkume/kume.html>, 08.05.2015
6. Kümeleme Analizinde Kullanılan Bazı Farklılık ve Benzerlik Ölçülerinin İncelenmesi  
<http://emredunder.blogspot.com.tr/2011/06/kumeleme-analizinde-kullanilan-baz.html>, 08.05.2015
7. Tatlıdil, Hüseyin, “*Uygulamalı Çok Değişkenli İstatistiksel Analiz*”, Ankara, Akademi Matbaası, 2002
8. Silahtaroğlu, Gökhan, “*Kavram ve Algoritmalarıyla Temel Veri Madenciliği*”, İstanbul, Papatya Yayıncılık, 2008
9. Kalua, Botjan, “*Instant Weka How-to*”, Packt Publishing, 2013

- 10. Bakos, Gabor, “KNIME Essentials”, Packt Publishing, 2013**
- 11. Işık, M. ve Çamurcu, A. Y. “K-Means ve Aşırı Küresel C-Means Algoritmaları ile Belge Madenciliği”, Marmara Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 2010**
- 12. Sarıman, Güncel, “Veri Madenciliğinde Kümeleme Teknikleri Üzerine Bir Çalışma: K-Means ve K-Medoids Kümeleme Algoritmalarının Karşılaştırılması”, Süleyman Demirel Üniversitesi Fen Bilimleri Dergisi, 2011**
- 13. Işık, M. ve Çamurcu, A. Y., “K-Means, K-Medoids ve Bulanık C-Means Algoritmalarının Uygulamalı Olarak Performanslarının Tespiti”, İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, 2007**

## **ÖZGEÇMİŐ**

15 Mart 1988 tarihi Ordu ili doğumluyum. Fatsa Lisesi'nden mezun olduktan sonra 2008 yılında IŐık Üniversitesi Grafik Sanatlar ve Grafik Tasarım bölümünü burslu olarak kazandım ve 2012 yılında mezun oldum. Őu anda vakıf üniversitesinde yazılım uzmanı olarak çalışmaktayım. Aynı zamanda Beykent Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliđi yüksek lisans eğitimimi sürdürmekteyim.

**Tuđrul UđURLU**