

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
ELEKTRİK-ELEKTRONİK FAKÜLTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

Uzman Sistemlere Giriş
PROJE

Projenin Konusu

Çeşitli ML Algoritmaları ile Ev Kiralarını Tahmin Eden Uzman Sistem

Dersi Veren Öğretim Üyesi

Doç. Dr. Mehmet Fatih AMASYALI

Projeyi Yapan Öğrenciler

Ç16014033	Ferhat TAŞ
Ç16014023	Eray Zeki SAH
Ç17014901	Emre KURU

Proje Teslim Tarihi

06.01.2022

İÇİNDEKİLER

İÇİNDEKİLER	1
KONU TANITIMI	2
GELİŞTİRME SÜRECİ	3
1. Veri Seti Hazırlanması	3
2. Makine Öğrenmesi Algoritmaları ile Eğitim İşlemleri	3
3. Sunucu Oluşturulması	4
4. Mobil Uygulama Geliştirilmesi	4
ÖRNEK ÇALIŞTIRMALAR	5
SİSTEM BAŞARISI VE YORUMLAR	8
KAYNAKLAR	9

Proje Konusu: Çeşitli ML Algoritmalarını Deneyerek Ev Kiralarını En Başarılı Şekilde Tahmin Eden Uzman Sistem Geliştirmek

Proje Özeti: Projenin temel amacı İstanbul'un 10 ilçesinde (Esenyurt, Beşiktaş, Pendik, Küçükçekmece, Bahçelievler, Üsküdar, Kadıköy, Sarıyer, Bakırköy, Şişli) bulunan ev kiralarını başarılı bir şekilde tahmin eden bir uzman sistem geliştirmektir. Bu kapsamda projede çeşitli Makine Öğrenmesi algoritmalarıyla modeller oluşturulup ev kiralarını tahmin eden modellerin oluşturulması hedeflenmektedir. Daha sonra bir mobil uygulama geliştirilerek ve makine öğrenmesinden elde edilen modeller kullanılarak kullanıcının girdi verip çıktı olarak ev kira tahminini alması sağlanacaktır. Böylelikle hem ev kiralarını tahmin eden bir uzman sistem (bir nevi cebimizdeki emlakçı) geliştirilecek, hem de hangi Makine Öğrenmesi algoritmasının bu konuda daha başarılı olduğu tespit edilecektir.

Projeyi gerçekleştirmek için izlenen yol raporun ikinci bölümünde, yani Geliştirme Süreci'nde detaylı olarak anlatılmıştır.

Projeyi tanıtmak için bir video çekilip YouTube'a yüklenmiştir. Videoya bu linkten ulaşılabilir: https://youtu.be/d-pe_WwLu5U

Projede yazılan kodların tümü boyuttan dolayı zip dosyasına koyulmamıştır, tüm kodlara bu drive linkinden ulaşılabilir:

<https://drive.google.com/drive/folders/1EA2wzcXAoGCYpFeb75fDSUvfZfbahdZ1?usp=sharing>

1. Veri Seti Hazırlanması

Projede ilk olarak bir veri seti elde etmek için Python ile yazılan bir crawler ile <https://www.emlakjet.com> sitesinden İstanbul'un belirlenen 10 ilçesinde bulunan kiralık evlerin verisi çekilmiştir. Daha sonra bu çekilen verilerle Makine Öğrenmesi algoritmalarında kullanılmak üzere bir veri seti oluşturulmuştur.

Her bir ilçe için ayrı ayrı veri setleri oluşturulmuştur. Aşağıda Beşiktaş ilçesinde bulunan verilerden birkaç örnek gösterilmiştir.

	A	B	C	D	E	F	G	H	I
1	Kira	Alan	Oda Sayısı	Banyo Sayısı	Bina Yaşı	Isıtma Tipi	Balkon Durumu	Eşya Durumu	Site İçerisinde
2	30000	220	4+1	2	4	Merkezi (Pay Ölçer)	Var	Eşyalı	Evet
3	25000	120	3+1	1	21 Ve Üzeri	Merkezi Doğalgaz	Var	Eşyalı	Evet
4	26000	190	3+1	1	21 Ve Üzeri	Merkezi Doğalgaz	Yok	Eşyalı	Hayır
5	8000	58	2+1	1	21 Ve Üzeri	Kombi Doğalgaz	Yok	Eşyalı	Hayır
6	7650	100	2+1	2	21 Ve Üzeri	Merkezi Doğalgaz	Yok	Boş	Hayır
7	7000	100	3+1	1	21 Ve Üzeri	Merkezi Doğalgaz	Yok	Boş	Hayır
8	6000	55	1+1	1	21 Ve Üzeri	Kombi Doğalgaz	Yok	Eşyalı	Hayır
9	5000	40	1+1	1	2	Kombi Doğalgaz	Yok	Eşyalı	Hayır
10	9500	100	3+1	1	21 Ve Üzeri	Kombi Doğalgaz	Yok	Eşyalı	Hayır
11	7000	70	1+1	1	21 Ve Üzeri	Kombi Doğalgaz	Yok	Eşyalı	Hayır
12	7250	110	3+1	2	21 Ve Üzeri	Merkezi Doğalgaz	Var	Boş	Hayır
13	7500	55	1+1	2	21 Ve Üzeri	Kombi Doğalgaz	Yok	Eşyalı	Hayır
14	6250	90	2+1	1	16-20	Kombi Doğalgaz	Yok	Boş	Hayır
15	22000	160	3+1	2	21 Ve Üzeri	Merkezi Doğalgaz	Yok	Eşyalı	Hayır
16	7000	75	2+1	1	11-15	Yerden Isıtma	Yok	Eşyalı	Hayır
17	30000	150	3+1	2	21 Ve Üzeri	Merkezi Doğalgaz	Yok	Boş	Hayır
18	28000	95	2+1	1	21 Ve Üzeri	Kombi Doğalgaz	Yok	Boş	Hayır

2. Makine Öğrenmesi Algoritmaları ile Eğitim İşlemleri

Projenin bu kısmında oluşturulan veri seti üzerinde Lineer Regresyon, Random Forest, KNN ve SVR algoritmaları ile eğitim işlemleri yapılmış ve modeller elde edilmiştir. Bu eğitim işlemleri yapılırken Python'da bulunan Scikit-Learn kütüphanesi kullanılmış ve bu işlemler için bir notebook dosyası üzerinde çalışılmıştır.

Eğitim işlemi için veri setinin %90'ı, test işlemi içinse %10'luk kısmı ayrılmıştır. Daha sonrasında her bir ilçe için 4 adet makine öğrenmesi modeli elde edilerek kaydedilmiştir.

3. Sunucu Oluřturulması

Projenin bu kısmında Python’da bulunan Flask kütüphanesi kullanılarak bir sunucu oluşturulmuřtur. Bu sunucuda bir POST metodu yazılmış ve bu metod sayesinde mobil uygulamadan gelecek girdiler alınıp karşılığında önceden eğitilen modeller kullanarak çıktılar (ev kira tahminleri) üretilmektedir. Kullanıcıdan alınan her bir girdi için dört adet algoritma (Lineer Regresyon, Random Forest, KNN ve SVR) ile toplamda dört adet çıktı üretilmekte ve mobil uygulamaya gönderilmektedir.

Sistemin adım adım çalışması řu şekildedir:

- Kullanıcı girdisi üzerinde düzenleme yapılarak, girdi makine öğrenmesi modeli için uygun formata getirilir.
- İstenilen ilçeye göre 4 adet makine öğrenmesi modeli yüklenir.
- Modeller ile kullanıcı girdisi kullanılarak çıktı üretilir.
- Üretilen çıktı sözlük formatında mobil uygulamaya gönderilir.

4. Mobil Uygulama Geliřtirilmesi

Projede kullanıcıyla etkileřim için Flutter kullanılarak bir mobil uygulama (bir nevi elimizin altındaki emlakçı) geliřtirilmiřtir. Bu mobil uygulama kullanıcıdan istediğı ev özelliklerini alıp çıktı olarak evin tahmini aylık kirasını vermektedir. Mobil uygulamada kullanıcıdan girdi olarak ařağıdaki özellikler alınmaktadır:


- İlçe bilgisi (Örnek: Beřiktař)
- Evin alan bilgisi (Örnek: 100 m²)
- Evin oda sayısı (Örnek: 3+1)
- Bina yaşı (Örnek: 21 ve Üzeri)
- Banyo sayısı (Örnek: 1)
- Isıtma tipi (Örnek: Merkezi Doğalgaz)
- Balkon durumu (Var-Yok)
- Eřya durumu (Eřyalı-Boř)
- Site içerisinde (Evet-Hayır)

Kullanıcı mobil uygulamada yukarıdaki girdileri verdikten sonra bu bilgiler sunucuya gönderilip karşılığında dört algoritmadan gelen ev kira tahminlerini almaktadır.

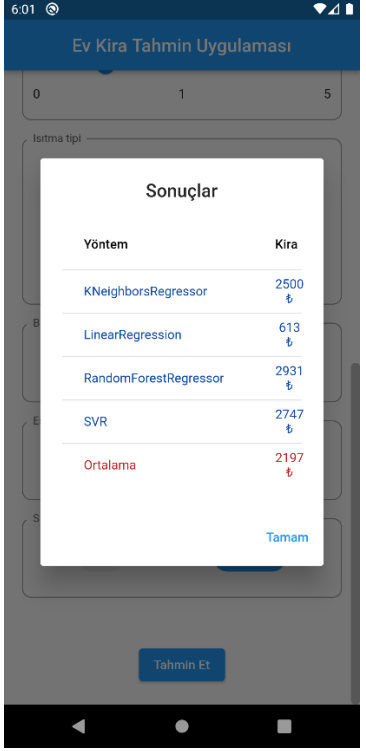
ÖRNEK ÇALIŞTIRMALAR

Örnek Çalıştırma 1:

Kullanıcı Girdileri



Kira Tahmini

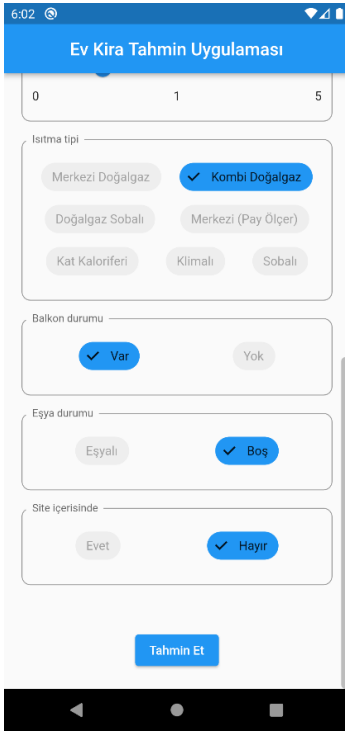
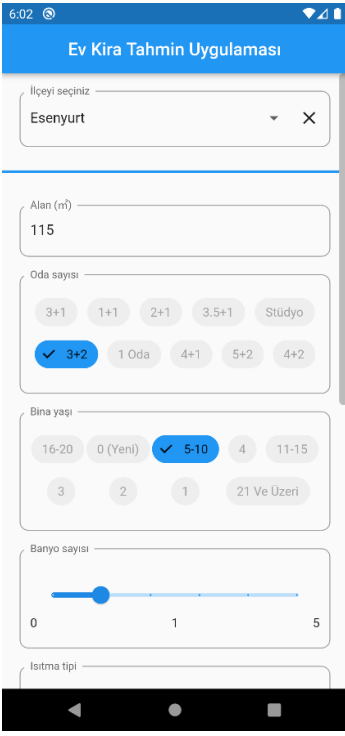


```
127.0.0.1 - - [01/Jan/2022 20:56:57] "POST /getpredictions HTTP/1.1" 200 -
  Alan Oda Sayısı ... Isıtma Tipi_Sobalı Isıtma Tipi_Yerden Isıtma
0 105 5 ... 0 0

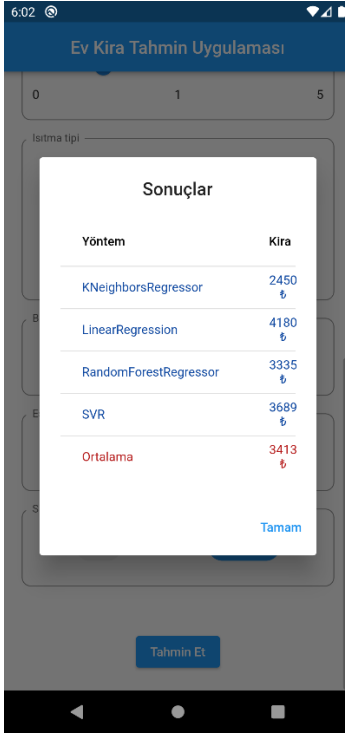
[1 rows x 14 columns]
{'LinearRegression': 613, 'RandomForestRegressor': 2931, 'KNeighborsRegressor': 2500, 'SVR': 2747}
```

Örnek Çalıştırma 2:

Kullanıcı Girdileri



Kira Tahmini



Yöntem	Kira
KNeighborsRegressor	2450 ₺
LinearRegression	4180 ₺
RandomForestRegressor	3335 ₺
SVR	3689 ₺
Ortalama	3413 ₺


Tamam

```
127.0.0.1 - - [01/Jan/2022 21:02:04] "POST /getpredictions HTTP/1.1" 200 -
  Alan Oda Sayısı ... Isıtma Tipi_Merkezi Doğalgaz Isıtma Tipi_Sobalı
0  115      7      ...                      1                      0

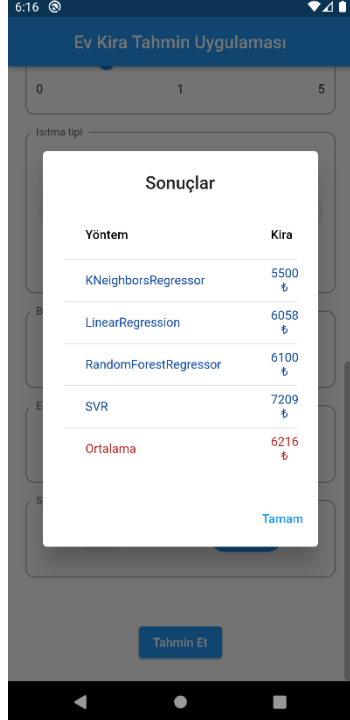
[1 rows x 14 columns]
{'LinearRegression': 4180, 'RandomForestRegressor': 3335, 'KNeighborsRegressor': 2450, 'SVR': 3689}
```

Örnek Çalıştırma 3:

Kullanıcı Girdileri



Kira Tahmini



Yöntem	Kira
KNeighborsRegressor	5500 ₺
LinearRegression	6058 ₺
RandomForestRegressor	6100 ₺
SVR	7209 ₺
Ortalama	6216 ₺

Tamam

Tahmin Et

```
127.0.0.1 - - [01/Jan/2022 21:15:57] "POST /getpredictions HTTP/1.1" 200 -
Alan Oda Sayısı ... Isıtma Tipi_Merkezi Doğalgaz Isıtma Tipi_Yerden Isıtma
0 120 5 ... 1 0

[1 rows x 13 columns]
{'LinearRegression': 6058, 'RandomForestRegressor': 6100, 'KNeighborsRegressor': 5500, 'SVR': 7209}
```


SİSTEM BAŞARISI VE YORUMLAR

Sistemde kullanılan modellerin farklı özelliklerinden dolayı farklı girdilerde daha başarılı olmaları beklenmektedir. Örneğin K Neighbour Regressor, eğer dataset üzerinde kullanıcı girdisine yakın bir veri varsa çok yakın bir sonuç verirken, dataset verisinden belirli oranlarda farklı girdilere sonuç dönme başarısında geride kalabilir.

Sistemde başarı ölçütü olarak r2 score kullanılmıştır. Aşağıdaki tabloda her bir ilçe için farklı algoritmaların r2 score değerleri verilmiştir.

İlçe	Linear Regression	Random Forest Regressor	KNeighborsRegressor	SVR
esenyurt	0.169	0.837	0.962	0.084
besiktas	0.547	0.909	0.937	0.295
pendik	0.813	0.85	0.995	0.3
kucukcekmece	0.643	0.95	0.997	0.296
bahcelievler	0.508	0.905	0.964	0.307
uskudar	0.745	0.922	0.954	0.277
kadikoy	0.735	0.929	0.997	0.605
sariyer	0.597	0.943	0.985	0.481
bakirkoy	0.708	0.897	0.822	0.502
sisli	0.595	0.927	0.994	0.433

Sonuçlardan yola çıkarak en başarılı modellerin Random Forest Regressor ve K Neighbours Regressor olduğu söylenebilir. Veri seti boyutunun ilçeden ilçeye değişkenlik göstermesi de bu başarıları etkilemektedir.

- [1] I. Bershadskiy, «Predicting House Price Using Regression Algorithm for Machine Learning,» [Çevrimiçi]. Available: <https://yalantis.com/blog/predictive-algorithm-for-house-price/>.
- [2] danvickmiller.dev, «Flutter FormBuilder,» [Çevrimiçi]. Available: https://pub.dev/packages/flutter_form_builder.
- [3] P. Misra, «House Price Predictions Using Keras,» [Çevrimiçi]. Available: <https://towardsai.net/p/deep-learning/house-price-predictions-using-keras>.