



**YILDIZ TEKNİK ÜNİVERSİTESİ  
KİMYA-METALÜRJİ FAKÜLTESİ  
MATEMATİK MÜHENDİSLİĞİ BÖLÜMÜ**

**BİTİRME ÇALIŞMASI**

**FIFA 2018 VERİSİNİN VERİ MADENCİLİĞİ  
TEKNİKLERİ İLE İNCELENMESİ**

Proje Yöneticisi: Yrd. Doç. Dr. Nilgün GÜLER BAYAZIT

13053001 Selçuk HACIAHMET

İstanbul, 2018



## İÇİNDEKİLER

	Sayfa
ŞEKİL LİSTESİ .....	iv
ÖNSÖZ .....	v
ÖZET .....	vi
1. GİRİŞ .....	1
2. VIDEO OYUN ANALİZİ VE UYGULAMALARI.....	1
3. FIFA 2018 VERİLERİ ÜZERİNDE KULLANILAN VERİ MADENCİLİĞİ TEKNİKLERİ .....	2
3.1 LİNEER REGRESYON .....	2
3.2 POLİNOM REGRESYON .....	3
3.3 LASSO .....	4
3.4 SUPPORT VECTOR REGRESSION (SVR) .....	5
4. FIFA 2018 VERİSİNİN VERİ MADENCİLİĞİ TEKNİKLERİ İLE ANALİZ EDİLMESİ UYGULAMASI .....	6
4.1 PYTHON NEDİR? .....	6
4.2 PYTHON PROGRAMLAMA DİLİNİN VERİ MADENCİLİĞİNDEKİ ÖNEMİ .....	8
4.2.1 VERİ MADENCİLİĞİ UYGULAMALARINDA KULLANILAN TEMEL PYTHON KÜTÜPHANELERİ .....	8
4.2.1.1 NUMPY.....	8
4.2.1.2 PANDAS .....	8
4.2.1.3 MATPLOTLİB .....	9
4.2.1.4 SCİPY .....	9
4.2.1.5 SCİKİT-LEARN .....	9
4.3 FIFA 2018 VERİSİNİN TEMİZLENME .....	10
4.4 VERİLERİN EĞİTİLMİŞ VE TEST KÜMELERİNE AYIRMA İŞLEMİ.....	12
4.5 VERİLERE LINEER REGRESYON METOTUNUN UYGULANMASI .....	13
4.6 VERİLERE POLİNOM REGRESYON METOTUNUN UYGULANMASI .....	14
4.7 VERİLERE LASSO METOTUNUN UYGULANMASI .....	15
4.8 VERİLERE SVR UYGULANMASI .....	16
5.SONUÇ .....	18
EKLER .....	19
KAYNAKÇA .....	25
ÖZGEÇMİŞ .....	26

## ŞEKİL LİSTESİ

**Şekil 1 :** FIFA 2018 Veri Kümesi Üzerinde Oyuncu Ortalama Puanı ile Oyuncu Değerini Karşılaştırılan Lineer Regresyon Modeli

**Şekil 2 :** FIFA 2018 Veri Kümesi Üzerinde Oyuncu Ortalama Puanı ile Oyuncu Değerini Karşılaştırılan Polinom Regresyon Modeli

**Şekil 3 :** FIFA 2018 Veri Kümesi Üzerinde Oyuncu Ortalama Puanı ile Oyuncu Değerini Karşılaştırılan Lasso Regresyom Modeli

**Şekil 4 :** FIFA 2018 Veri Kümesi Üzerinde Oyuncu Ortalama Puanı ile Oyuncu Değerini Karşılaştırılan Destek Vektörü Regresyom Modeli

**Şekil 5 :** Python Logosu

**Şekil 6 :** 2017 Yılıının En Gözde Programlama Dilleri

**Şekil 7 :** FIFA 18 Veri Kümesinin ('Complete Dataset') Belirli Kısmı

**Şekil 8 :** Nümerik olmayan verilerin boş bir değerle değiştirilmesi

**Şekil 9 :** Value ve Wage değerlerindeki 'M' ve 'K' harflerinin veriden temizlemesi sırasında kullanılan eşitlikler

**Şekil 10 :** Value ve Wage değerlerindeki 'M' ve 'K' harflerinin boşlukla değiştirilmesi ve önlerindeki sayılarının belirlenen sayılarla çarpılması işlemi

**Şekil 11 :** Temizlenen verilerin yeni dosyaya kaydedilmesi

**Şekil 12 :** Verilerin Eğitilmiş ve Test Kümelerine Ayrılması

**Şekil 13 :** Lineer Regresyonun Uygulanması

**Şekil 14 :** Tahminleme İşlemi

**Şekil 15 :** 94 'Overall' puanına sahip bir oyuncunun transfer değerinin tahmini çıktı olarak veren kod satırı

**Şekil 16 :** Tahmin Sonucu

**Şekil 17:** Lineer Regresyonun Figür Olarak Gösterimi

**Şekil 18:** Lineer Regresyon Figürü

**Şekil 19:** Polinom Regresyonun Uygulanması

**Şekil 20:** Polinom Regresyon Figürü

**Şekil 21:** 94 'Overall' puanına sahip futbolcunun tahmin edilen transfer değeri

**Şekil 22:** LASSO Metotunun Figürü

**Şekil 23:** LASSO Metotunun 94 puana sahip futbolcunun transfer değerinin tahmini

**Şekil 24:** SVM Figürü

**Şekil 25:** SVR Tahmin

## **ÖNSÖZ**

Bu çalışmada özellikle günümüz dünyasının değişilmez bir parçası haline gelen veri kümelerinden veri madenciliği teknikleri ile anlamlı bilgi çıkarma prosesi işlenmiştir. Veri madenciliği içinde bulunan bir çok teknikleri kullanarak FIFA 2018 adlı video oyununun içinde bulunan futbolcu ve futbolcu özelliklerinin verileri kullanılarak anlamlı bilgileri açığa çıkarılmaya çalışılmıştır.

Proje konusunu seçerken isteğimi göz önünde bulundurup bana yardımcı olan ve bu çalışmayı yürüttüğüm süre boyunca benden yardımlarını esirgemeyen proje ve akıl danışmanım sayın Yrd. Doç. Dr. Nilgün GÜLER BAYAZIT'a teşekkürü bir borç bilirim.

Çalışmalarım süresince maddi ve manevi destekleri için sevgili aileme ve Öge Tuana KALAYCIOĞLU'na sonsuz teşekkürlerimi sunarım.

**İstanbul, 2018**

**Selçuk HACIAHMET**

## ÖZET

Bu tezde veri madenciliğinde çokça kullanılan veri madenciliği teknikleri Python 3.6 programlama dili kullanılarak uygulanmıştır. Tezde video oyunların analiz edilmesi, bu analizlerin hangi tekniklerle daha verimli ve anlamlı bir bilgi ortaya çıkması ve günümüz dünyasında yavaş yavaş hayatımıza giren “Video Oyunları Analizi” kavramı incelenmiştir.

Python 3.6 programlama dilinin veri madenciliği tekniklerini içerdiği kütüphaneler yardımıyla önceden elde edilmiş FIFA 2018 video oyununun veri kümesi üzerinde oyun yapımcıları tarafından belirlenen ortalama puanları ile oyun içerisindeki oyunun değeri arasında ilişki farklı veri madenciliği teknikleri ve bu tekniklerin arasındaki farkı incelenmiştir.

## ABSTRACT

In this thesis, data mining techniques, which are widely used in data mining, have been implemented using the Python 3.6 programming language. The concept of "Video Games Analysis", which analyzes the video games in thesis, which techniques produce more efficient and meaningful information and which gradually enters our life in today's world, has been examined.

Different data mining techniques and the differences between these techniques were examined in relation to the average score of the FIFA 2018 video game data set previously obtained with the help of the libraries that the Python 3.6 programming language included data mining techniques and the game scores in the game.

## 1. GİRİŞ

Veri madenciliği büyük ölçekli veriler arasından anlamlı bilgiye ulaşılması veya büyük veri yığınları içerisinde gelecekle ilgili tahminde bulunulmasını sağlayabilecek bağıntıların aranmasıdır. Bu büyük verilerin işlenmesinde bilgisayar programlarından veya programlama dillerinden yararlanılır. Veri madenciliğinde ham bilginin kullanılabilmesi için bazı adımlardan geçmesi gerekmektedir. Bu adımlar veri temizleme, veri bütünleştirme, veri dönüşümü, veri seçme, örüntü değerlendirme şeklinde olabilir.<sup>[1]</sup> Bu çalışmada da bu veri temizleme adımlarından bazıları ve veri işleme kısmında ise Python 3.6 programlama dili kullanılmıştır. Kullanılan veri madenciliği teknikleri ise lineer regresyon, polinomik regresyon, LASSO, Support Vector Regression (SVR) şeklindedir.

Bu çalışmada bir veri kümesi üzerinde gerçekleştirilen teknikler ve bu tekniklerin farklılıkları, FIFA 2018 oyununun veri kümesi üzerinden futbol oyuncularının ortalama puanları ve transfer değerleri üzerinden anlamlı bilgilerin çıkarılması, bu çıkarımlardan yola çıkarak herhangi bir ortalama puana sahip oyuncunun farklı veri madenciliği teknikleri ile transfer değerinin tahmin edilmesi incelenmektedir. Ayrıca video oyunları üzerinde veri madenciliğinin dünyada uygulamaları, Python 3.6 programlama dili ve kullanılan veri madenciliği teknikleri hakkında bilgiler verilmiştir.

## 2. VIDEO OYUN ANALİZİ VE UYGULAMALARI

Video oyun analizi, analizi yapılacak oyunun veri kümesi üzerinden anlamlı bilgi veya bilgiler topluluğu çıkarılmasıdır. Bu analiz video oyununu oynayan kişiden video oyununu oluşturan şirketin çalışanlarına kadar herkes tarafından yapılabilir. Video oyun analizi genellikle veri madenciliği teknikleri kullanılarak yapılmaktadır. Literatürde genellikle “Game Data Mining” ya da “Data Mining for Games” olarak geçmektedir.

Video oyun analizinde geçmişe baktığımızda diğer alanlarda olduğu kadar fazla örneği bulunmamaktadır. Leonardo Cotta, Pedro O.S. Vaz de Melo, Fabrício Benevenuto ve Antonio A.F. Loureiro tarafından hazırlanan “FIFA Futbol Video Oyununun Verileri kullanılarak Futbol Analizi” başlıklı çalışmada 2014 yılındaki Brezilya ve Almanya Milli Takımları arasındaki farklara ve FC Barcelona takımının diğer takımlardan farklı futbol stili incelenmiştir.<sup>[2]</sup>

2015 yılında ise Filip Johansson ve Jesper Wikström'ün ortak çalışmasında ortaya çıkan “DOTA 2 de Tekrar Madenciliğinin Sonuç Tahminleri” adlı tezinde ise DOTA 2 oyununda oynanan maçların tekrar görüntüleri üzerinden veri madenciliği tekniklerini kullanarak sonraki oynanacak olan maçlarda hangi takımın kazanacağını tahmin etme üzerine çalışılmıştır. [3]

2017 yılında Peter Braun, Alfredo Cuzzocrea, Timothy D. Keding, Carson K. Leung, Adam G.M. Padzor ve Dell Sayson'un ortaklaşa yürüttüğü “Oyun Veri Madenciliği: Siber-Fiziksel Dünyalarda Çevrimiçi Oyun Verilerinin Kümelenmesi ve Görselleştirilmesi” adlı çalışmada Overwatch adlı oyunun veri madenciliği tekniklerinden olan kümeleme yardımıyla oyun verilerini görselleştirilmiş ve bazı çıkarımlarda bulunulmuştur. [4]

### **3. FIFA 2018 VERİLERİ ÜZERİNDE KULLANILAN VERİ MADENCİLİĞİ TEKNİKLERİ**

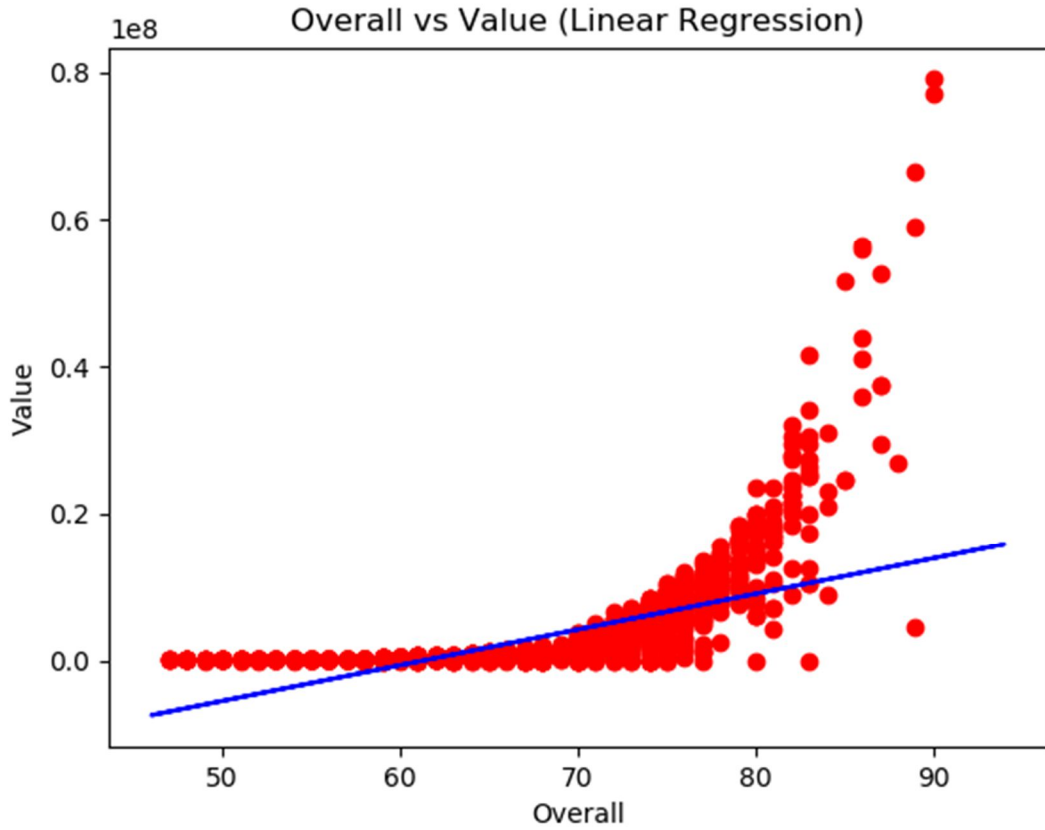
#### **3.1. Lineer Regresyon**

Regresyon analizi bir bağımlı değişken ile bir bağımsız değişken (basit regresyon) veya birden fazla bağımsız değişken (çoklu regresyon) arasındaki ilişkilerin matematiksel eşitlik ile açıklanması sürecidir. Regresyon analizinde değişkenler arasında ilişki doğrusal ise doğrusal regresyon, değil ise doğrusal olmayan regresyon olarak adlandırılır. [5]

Lineer regresyon, veri kümesindeki gözlenen yanıtlar ve doğrusal yaklaşım ile tahmin edilen yanıtlar arasındaki ilişkiyi analiz eder. [6]

Lineer regresyon modeli örneği aşağıda gösterilmiştir.



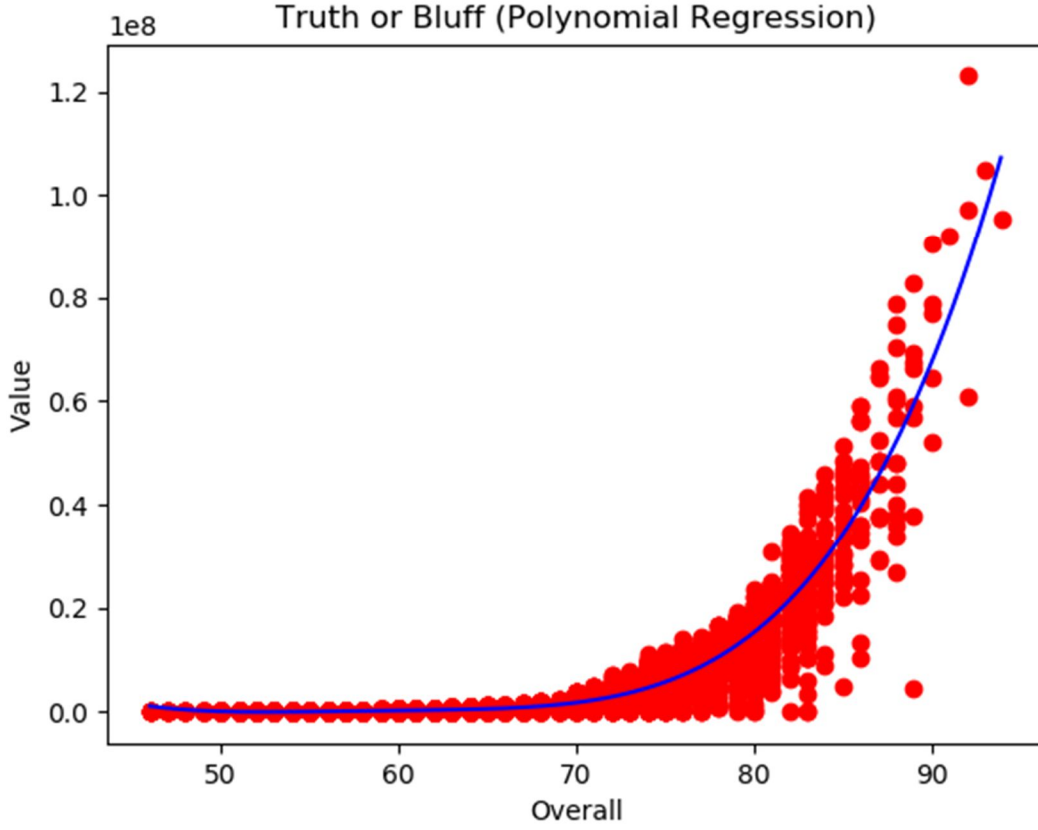


**Şekil 1:** FIFA 2018 Veri Kümesi Üzerinde Oyuncu Ortalama Puanı ile Oyuncu Değerini Karşılaştırılan Lineer Regresyon Modeli

### 3.2. Polinom Regresyonu

Bu regresyon genellikle makine öğrenmesinde kullanılan bir regresyondur. Doğrusal olmayan fonksiyonun sahip olduğu verilerle eğitilmektedir ve polinomik bir eğri oluşturmaktadır. Bu regresyon çok daha geniş bir veri aralığında eğitilmesine ve çalışılmasına izin vermekle birlikte lineer (doğrusal) regresyon kadar da hızlı çalışır.<sup>[7]</sup>

Polinom regresyon modeli örneği aşağıda gösterilmiştir

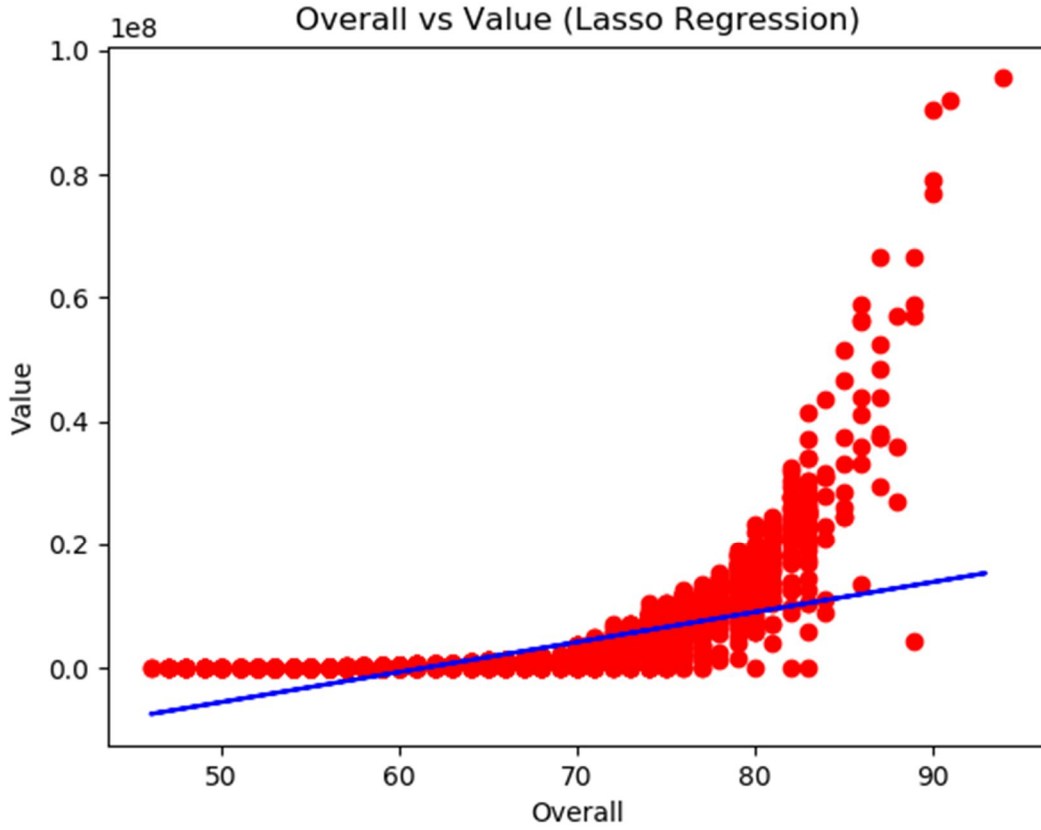


**Şekil 2:** FIFA 2018 Veri Kümesi Üzerinde Oyuncu Ortalama Puanı ile Oyuncu Değerini Karşılaştırılan Polinom Regresyon Modeli

### 3.3. LASSO

Lasso, seyrek katsayıları tahmin eden doğrusal bir modeldir. Bazı bağlamlarda, daha az parametre değerine sahip çözümleri tercih etme eğilimi nedeniyle verilen çözümün bağımlı olduğu değişken sayısını etkili bir şekilde azaltılarak daha kullanışlı hale getirilebilmektedir. Bu nedenle Lasso ve türevleri, sıkıştırılmış algılama alanı için çok önemli bir yere sahiptir.<sup>[8]</sup>

Lasso'nun görsel modeli aşağıda gösterilmiştir.



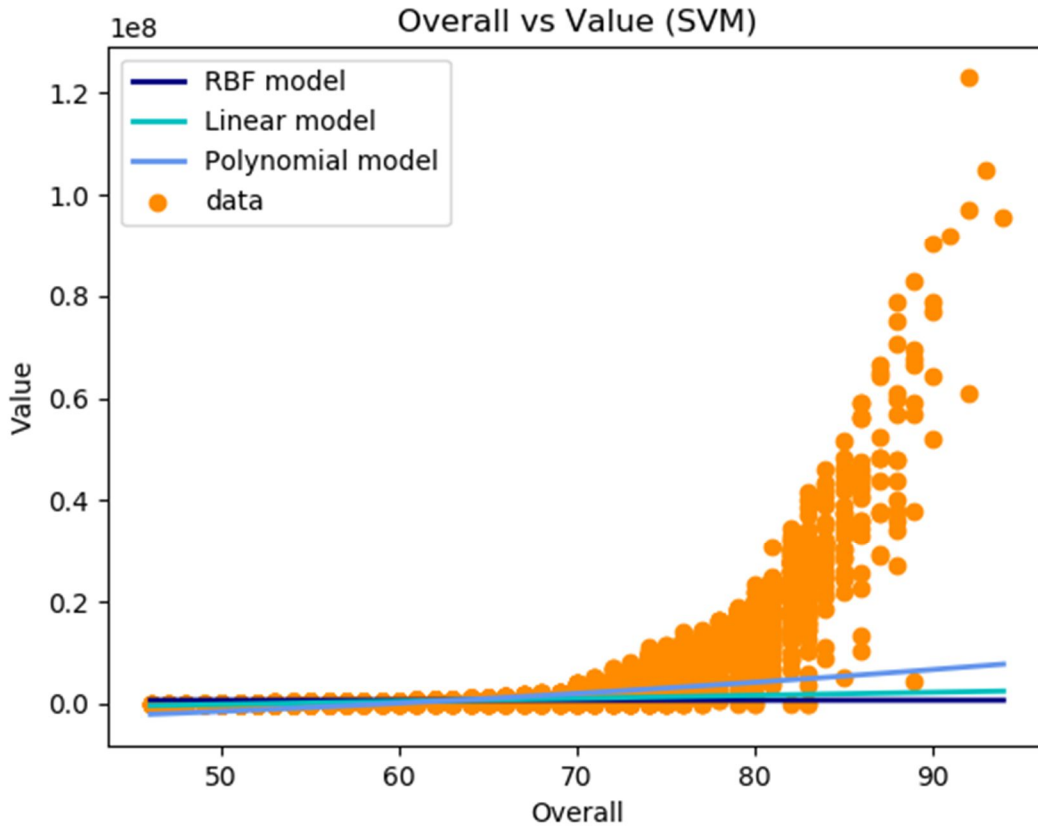
**Şekil 3:** FIFA 2018 Veri Kümesi Üzerinde Oyuncu Ortalama Puanı ile Oyuncu Değerini Karşılaştırılan Lasso Regresyon Modeli

### 3.4. Support Vektör Regression (SVR) – Destek Vektörü Regresyonu

Destek Vektör Sınıflandırma (Support Vector Classification) yöntemi, regresyon problemlerini çözmek için genişletilmiştir. Bu genişleme yöntemine Destek Vektör Regresyonu (Support Vector Regression) denir.

Destek vektör sınıflandırılması ile üretilen model, sadece eğitim verilerinin bir alt kümesine bağlıdır. Çünkü modelin oluşturulmasında rol oynayan maliyet fonksiyonu doğrusal doğrunun altında kalan eğitim noktaları ile ilgilenmemektedir. Benzer şekilde, Destek Vektörü Regresyonu tarafından üretilen model sadece eğitim kümesinin verilerine bağlıdır. Çünkü modelin oluşturulmasında kullanılan maliyet fonksiyonu model tahminine yakın herhangi bir eğitim verisini göz ardı etmektedir.<sup>[9]</sup>

Destek Vektörü Regresyon model örneği aşağıda gösterilmiştir.



**Şekil 4:** FIFA 2018 Veri Kümesi Üzerinde Oyuncu Ortalama Puanı ile Oyuncu Değerini Karşılaştırılan Destek Vektörü Regresyon Modeli

#### 4. FIFA 2018 VERİSİNİN VERİ MADENCİLİĞİ TEKNİKLERİYLE ANALİZ EDİLMESİ UYGULAMASI

##### 4.1. Python Nedir?

Python, Hollandalı programcı Guido Van Rossum tarafından geliştirilmiş bir programlama dilidir. Geliştirilmesi 1990 yılında başlanan Python; C ve C++ gibi programlama dilleri ile kıyaslanırsa aşağıdaki gibi sonuçlar elde edilebilir:

- Python'un öğrenilmesi daha kolaydır.
- Program geliştirme sürecini kısaltır. Başka bir deyişle diğer programlama dillerine göre daha hızlı yazılır.

- Yine diğer programlama dillerine karşın herhangi bir derleyici programlara ihtiyaç duymaz.
- Diğer programlama dillerine kıyasla hem daha okunaklı, hem de daha temiz kodsal bir söz dizime sahiptir.

Yukarıdaki maddeler gibi daha bir çok maddeden dolayı Python programlama dili son yıllarda kullanıcısını oldukça artıran bir programlama dilidir. Ayrıca Python Windows, Linux/Unix ve Mac-OS üzerinde çalışmasının yanı sıra Java ve .NET sanal makinelerine de entegre edilmiş durumdadır. Python'un diğer önemli bir özelliği ise açık kaynak kod lisanslı özgür ve ücretsiz bir yazılımdır. Python ile nesne yönelimli programlama, fonksiyonel ya da yapısal programlama gibi birden fazla programlama paradigmasını destekleyen bir programa dilidir. Bununla birlikte çalıştırılabilir dosyalar şeklinde paketlenen programlardan, betikler (script) yazmaya kadar farklı amaçlarda da kullanılmaktadır.<sup>[10]</sup>



Şekil 5: Python Logosu



Şekil 6 : 2017 Yılı'nın En Gözde Programlama Dilleri<sup>[11]</sup>

## 4.2. Python Programlama Dilinin Veri Madenciliğindeki Önemi

Python programlama dili, veri analizi, interaktif bilgiyi işleme ve veri görselleştirmede, R, MATLAB, SAS, Stata gibi analiz dilleri ve araçları arasında kendisine yer bulmuştur. Bundaki en büyük etken ise son yıllarda veri analizi yapılmasını kolaylaştıran kütüphanelerine(pandas ve scikit-learn gibi) sağlanan destektir. Bu kütüphaneler ile Python'un yazılımsal söz diziliminin kolaylığı birleştiğinde, veri uygulamaları oluşturmak için birincil dil olarak mükemmel bir seçenektir.<sup>[12]</sup>

### 4.2.1. Veri Madenciliği Uygulamalarında Kullanılan Temel Python Kütüphaneleri

#### 4.2.1.1. NumPy

Numerical Python (Sayısal Python) kelimelerinin kısaltması olan NumPy, uzun zamandır Python'da sayısal hesaplamanın temel taşı oluşturmuştur. Python'da sayısal verileri içeren çoğu bilimsel uygulama için gereken veri yapıları, algoritmalar ve kütüphaneleri içermektedir. NumPy, yukarıdakilerin yanı sıra aşağıdaki maddeleri içermektedir;

- Çok boyutlu dizi nesnesi olan *ndarray* hızlı ve çok verimlidir.
- Dizinlerin içindeki elemanlarla hesaplamalar yapılmaktadır. Ayrıca diziler arası matematiksel işlemler de yapılabilir.
- Dizi tabanlı veri kümelerini okuyabilme ve diske bu verileri yazma aracına sahiptir.
- Lineer cebir işlemleri, Fouier dönüşümleri ve rastgele sayı üretimleri yapılabilir.<sup>[12]</sup>

#### 4.2.1.2. Pandas

Pandas kütüphanesi, yapılandırılmış veya tablo verileriyle hızlı, kolay ve etkileyici bir şekilde çalışmayı sağlamak için tasarlanmış yüksek seviyeli veri yapıları ve fonksiyonları sağlamaktadır. 2010 yılında ortaya çıkmasından bu yana, Python'un güçlü ve üretken bir veri analizi ortamı olmasını sağlamıştır.

Pandas kütüphanesi, NumPy'ın yüksek performanslı dizi-hesaplamalarını elektronik tabloların ve ilişkisel veritabanlarının (SQL gibi) esnek veri işleme yetenekleriyle harmanlar. Yeniden şekillendirmeyi, dilimlemeyi, küp küp ayırmayı, kümeleri gerçekleştirmeyi ve verilerin alt kümeleri seçmeyi kolaylaştıran gelişmiş bir dizin oluşturma işlevi sağlamaktadır. Veri manipülasyonu, hazırlanması ve temizlenmesi veri analizinde çok önemli olduğundan, Pandas kütüphanesi veri analistlerinin ana odaklarından biridir.<sup>[12]</sup>

#### 4.2.1.3. Matplotlib

Matplotlib, parselleri ve diğer iki boyutlu veri görselleştirmelerini üretmek için en popüler Python kütüphanesidir. Bu kütüphane başlangıçta John D. Hunter tarafından yaratılmıştır ve şu sıralarda büyük bir geliştirici ekip tarafından geliştirilmesi sürdürülmektedir. Görselleştirme için uygun alan oluşturmak için tasarlanmıştır. Python'da başka görselleştirme kütüphaleleri olmasına karşın, Matplotlib bunlar arasında en yaygın olanıdır. Bunun sebebi diğer kütüphanelerle genel olarak iyi bir bütünleşme içerisine girebilmektedir.<sup>[12]</sup>

#### 4.2.1.4. SciPy

SciPy, bilimsel hesaplamalara hitap eden paketler topluluğudur. Bu kütüphane ile NumPy kütüphanesi birçok geleneksel bilimsel hesaplama uygulamaları için makul ve verimli bir ortam oluşturmaktadır.<sup>[12]</sup>

#### 4.2.1.5. Scikit-Learn

Bu kütüphane projesinin 2010'daki başlangıcından bu yana, Scikit-Learn, Python programcıları için önde gelen genel amaçlı makine öğrenme aracı haline gelmiştir. 7 yıl içerisinde 1500 geliştirici tarafından hazırlanan alt modüller ve algoritmalar aşağıda listelenmiştir;

- Sınıflandırma: SVM(Support Vector Machine – Destek Vektörü Makinesi), En Yakın Komşular (Nearest Neighbors), Rastgele Orman (Random Forest), Lojistik Regresyon (Logistic Regression) gibi...
- Regresyon: Lasso, Ridge Regresyon vb.
- Kümeleme: K-Means, Spektral Kümeleme (Spectral Clustering) vb.

- Boyut Azaltma: PCA, Özellik Seçimi (Feature Selection), Matris Faktörizasyonu (Matrix Factorization) vb.
- Model Seçimi: Izgara Arama (Grid Search), Çapraz Doğrulama (Cross-validation), Metrikler (Metrics)
- Ön İşleme: Özellik Çıkarma (Feature Extraction), Normalleştirme (Normalization)

Pandas, Statsmodels ve IPython kütüphaneleri ile birlikte Scikit-Learn kütüphanesi Python'un verimli bir veri bilimi programlama dili olması için kritik öneme sahiptir.<sup>[12]</sup>

### 4.3. FIFA 2018 Verisinin Temizlenmesi

FIFA 2018 oyunun oyuncu verileri kaggle.com'dan AkshatUppal adlı profilin "FIFA 18 Complete Player Dataset" adlı çalışmasından alınmıştır. Yaklaşık veri kümesinde 17 000 adet veri bulunmaktadır. Bu veri kümesinde;

- FIFA 2018 oyununun içinde yer alan bütün oyuncuların özellikleri,
- Bu her oyuncunun 70'den fazla nitelikleri,
- Oyuncuların ve ülkelerin bayrak resimleri,
- Oyuncunun oynadığı pozisyon verileri,
- Oyuncuların bütün oyun stil nitelikleri (Hızlanma Kalecilik Becerisi gibi)
- Oyuncuların kişisel verileri (Uyruk, Fotoğrafı, Oynadığı Klüp, Yaşı, Transfer Değeri gibi)

özelliklerini içermektedir.

Bu çalışmada overall denilen oyuncunun oyun tarafından belirlenen ortalama puanlar ile value denilen oyuncunun transfer değerleri ile çıkarımlarda bulunulmuştur.

Oyuncunun overall puanı sayısal bir değere sahip iken value niteliği veri kümesi içinde hem sayısal hem de özel karakterle içerdiği için bu veri ile işlem yapılabilmesi için sayısal değere dönüştürülmesi gerekmektedir.

Aşağıdaki görselde oyuncuların niteliklerini gösteren verilerden kısıtlı bir kesit bulunmaktadır:



A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U						
1	Player No	Name	Age	Photo	Nationality	Flag	Overall	Potential	Club	Club Logo	Value	Wage	Special	Acceleration	Aggression	Agility	Balance	Ball control	Composure	Crossing	Curve	Dribbling	Finishing	Free kick accuracy	Height	
0	Cristiano Ronaldo	32	https://cdn.sofifa.org/48/18/players/20801.png	Portugal	https://cdn.sofifa.org/flags/38.png	94	94	Real Madrid CF	https://cdn.sofifa.org/24/18/teams/243.png	~95.5M	~565K	2228	89	63	89										1,85	
1	L. Messi	30	https://cdn.sofifa.org/48/18/players/158023.png	Argentina	https://cdn.sofifa.org/flags/52.png	93	93	FC Barcelona	https://cdn.sofifa.org/24/18/teams/241.png	~105M	~565K	2154	92	48	90	95	95	96	7							1,70
2	Neymar	25	https://cdn.sofifa.org/48/18/players/190871.png	Brazil	https://cdn.sofifa.org/flags/54.png	92	94	Paris Saint-Germain	https://cdn.sofifa.org/24/18/teams/73.png	~123M	~280K	2100	94	56	96	82	95	92								1,75
3	L. Suárez	30	https://cdn.sofifa.org/48/18/players/176580.png	Uruguay	https://cdn.sofifa.org/flags/60.png	92	92	FC Barcelona	https://cdn.sofifa.org/24/18/teams/241.png	~97M	~510K	2291	88	78	86	60	91	83	75							1,73
4	M. Neuer	31	https://cdn.sofifa.org/48/18/players/167495.png	Germany	https://cdn.sofifa.org/flags/21.png	92	92	FC Bayern Munich	https://cdn.sofifa.org/24/18/teams/21.png	~91M	~230K	1493	58	92	55	42	58	47								1,88
5	R. Lewandowski	28	https://cdn.sofifa.org/48/18/players/188545.png	Poland	https://cdn.sofifa.org/flags/37.png	91	91	FC Bayern Munich	https://cdn.sofifa.org/24/18/teams/21.png	~92M	~355K	2143	70	80	78	80	78	80								1,80
6	De Gea	26	https://cdn.sofifa.org/48/18/players/193080.png	Spain	https://cdn.sofifa.org/flags/45.png	90	92	Manchester United	https://cdn.sofifa.org/24/18/teams/11.png	~64.5M	~125K	1458	57	60	43	40	42	64								1,88
7	E. Hazard	26	https://cdn.sofifa.org/48/18/players/183277.png	Belgium	https://cdn.sofifa.org/flags/7.png	90	91	Chelsea	https://cdn.sofifa.org/24/18/teams/5.png	~90.5M	~295K	2096	93	54	93	91	92	87	80	82	93					1,83
8	T. Kroos	27	https://cdn.sofifa.org/48/18/players/182521.png	Germany	https://cdn.sofifa.org/flags/21.png	90	90	Real Madrid CF	https://cdn.sofifa.org/24/18/teams/243.png	~91M	~240K	2165	60	70	69	69	89	85	80							1,83
9	G. Higuaín	29	https://cdn.sofifa.org/48/18/players/167664.png	Argentina	https://cdn.sofifa.org/flags/52.png	90	90	Juventus	https://cdn.sofifa.org/24/18/teams/45.png	~77M	~275K	1961	78	50	75	69	85	86	68							1,88
10	Sergio Ramos	31	https://cdn.sofifa.org/48/18/players/155862.png	Spain	https://cdn.sofifa.org/flags/45.png	90	90	Real Madrid CF	https://cdn.sofifa.org/24/18/teams/243.png	~92M	~310K	2153	75	84	79	60	84	87								1,83
11	K. De Bruyne	26	https://cdn.sofifa.org/48/18/players/192985.png	Belgium	https://cdn.sofifa.org/flags/7.png	89	92	Manchester City	https://cdn.sofifa.org/24/18/teams/10.png	~83M	~285K	2162	68	68	80	70	87	84								1,81
12	J. Courtois	25	https://cdn.sofifa.org/48/18/players/192119.png	Belgium	https://cdn.sofifa.org/flags/7.png	89	89	Chelsea	https://cdn.sofifa.org/24/18/teams/5.png	~59M	~190K	1282	46	23	61	45	25	14	19	13						1,93
13	A. Sánchez	28	https://cdn.sofifa.org/48/18/players/184941.png	Chile	https://cdn.sofifa.org/flags/55.png	89	89	Arsenal	https://cdn.sofifa.org/24/18/teams/1.png	~67.5M	~265K	2181	88	68	87	87	86	78	90							1,80
14	L. Modrić	31	https://cdn.sofifa.org/48/18/players/177003.png	Croatia	https://cdn.sofifa.org/flags/10.png	89	89	Real Madrid CF	https://cdn.sofifa.org/24/18/teams/243.png	~57M	~340K	2228	75	62	93	92	84	84								1,80
15	E. Bale	27	https://cdn.sofifa.org/48/18/players/173731.png	Wales	https://cdn.sofifa.org/flags/50.png	89	89	Real Madrid CF	https://cdn.sofifa.org/24/18/teams/243.png	~69.5M	~370K	2263	93	75	77	69	97	87	87							1,88
16	A. Agüero	29	https://cdn.sofifa.org/48/18/players/153079.png	Argentina	https://cdn.sofifa.org/flags/52.png	89	89	Manchester City	https://cdn.sofifa.org/24/18/teams/10.png	~66.5M	~325K	2074	90	63	86	91										1,80
17	B. Chiellini	32	https://cdn.sofifa.org/48/18/players/138956.png	Italy	https://cdn.sofifa.org/flags/27.png	89	89	Juventus	https://cdn.sofifa.org/24/18/teams/45.png	~38M	~120K	1367	68	48	59	74	52	58	68	58						1,80
18	B. Buffon	35	https://cdn.sofifa.org/48/18/players/1179.png	Italy	https://cdn.sofifa.org/flags/27.png	89	89	Juventus	https://cdn.sofifa.org/24/18/teams/45.png	~4.5M	~111K	1835	49	38	55	49	70	70	13	20	15					1,90
19	P. Dybala	23	https://cdn.sofifa.org/48/18/players/211119.png	Argentina	https://cdn.sofifa.org/flags/52.png	88	89	Juventus	https://cdn.sofifa.org/24/18/teams/45.png	~79M	~215K	2063	64	81	95	93	84	80	88							1,80
20	J. Oblak	24	https://cdn.sofifa.org/48/18/players/200389.png	Slovenia	https://cdn.sofifa.org/flags/44.png	88	89	Atlético Madrid	https://cdn.sofifa.org/24/18/teams/240.png	~57M	~82K	1290	43	67	49	45	16	15	15							1,85
21	A. Griezmann	26	https://cdn.sofifa.org/48/18/players/194765.png	France	https://cdn.sofifa.org/flags/18.png	88	91	Atlético Madrid	https://cdn.sofifa.org/24/18/teams/240.png	~75M	~150K	2104	87	69	80	90	80									1,80
CompleteDataset1																										

Şekil 7: FIFA 18 Veri Kümesinin (‘Complete Dataset’) Belirli Kısmı

‘Complete Dataset’ dosyasında bulunan “Value” ve “Wage” niteliklerinin verileri bazı oyuncularda nümerik bir sonuç olmadığından ötürü aşağıdaki kod satırlarıyla bu veriler boş bir değerle değiştirilmiştir.

```
dataset["Value"] = dataset.Value.str.replace('[^\x00-\x7F]', '')
dataset["Wage"] = dataset.Wage.str.replace('[^\x00-\x7F]', '')
```

Şekil 8 : Nümerik olmayan verilerin boş bir değerle değiştirilmesi

```
dataset.loc[dataset.Value.str.contains('M') == True, "Value"]
dataset.loc[dataset.Value.str.contains('K') == True, "Value"]
dataset.loc[dataset.Wage.str.contains('K') == True, "Wage"] =
```

Şekil 9 : Value ve Wage değerlerindeki ‘M’ ve ‘K’ harflerinin veriden temizlemesi sırasında kullanılan eşitlikler

```
= pd.to_numeric(dataset.loc[dataset.Value.str.contains('M') == True, "Value"].str.replace('M', ''), errors='ignore')*1000000
= pd.to_numeric(dataset.loc[dataset.Value.str.contains('K') == True, "Value"].str.replace('K', ''), errors='ignore')*1000
pd.to_numeric(dataset.loc[dataset.Wage.str.contains('K') == True, "Wage"].str.replace('K', ''), errors='ignore')*1000
```

Şekil 10 : Value ve Wage değerlerindeki ‘M’ ve ‘K’ harflerinin boşlukla değiştirilmesi ve önlerindeki sayılarının belirlenen sayılarla çarpılması işlemi

Şekil 9 ve şekil 10 ile gösterilen satırlar sırasıyla birbiriyle eşitlenmiştir. Veri kümesinde bulunan ‘Value’ ve ‘Wage’ sütunlarındaki değerler nümerik değildir. Belirli veri madenciliği

tekniklerinde veriler uygulanabilmesi için nümerik olması gerekmektedir. ‘Value’ ve ‘Wage’ sütunlarındaki verilerin sonunda oyununcunun değerine göre değerler bulunmaktadır. ‘M’ harfi önünde bulunan sayı ile birlikte ‘Milyon’u, ‘K’ harfi ise önünde bulunan sayı ile birlikte ‘Bin’i temsil etmektedir. Bu verilerin nümerik verilere dönüşmesi için yukarıdaki metod kullanılmıştır.

Veriler nümerik hale dönüştükten sonra ise yeni bir dosya oluşturularak kaydedilmektedir. Bu kayıt yöntemi aşağıdaki gibidir.

```
dataset.to_csv('CompleteDataset_updated.csv')
```

**Şekil 11:** Temizlenen verilerin yeni dosyaya kaydedilmesi

#### 4.4. Verilerin Eğitilmiş ve Test Kümelerine Ayırma İşlemi

Çalışmanın bu bölümünde veri setindeki tüm veriler eğitilmiş ve test kümelerine ayrılmıştır. Bu ayırma işlemi makine öğrenmesi nedeniyle yapılmaktadır. İleriki bölümlerde de atılacağı üzerine bir oyununcunun herhangi bir overall puanı üzerinden transfer değerinin ne kadar olduğunu tahmin etmek için kullanılır. Bu çalışmada Python programlama dilinde aşağıdaki gibi veriler eğitilmiş ve test kümelerine ayrılmıştır.

```
X = dataset.iloc[:, [6]].values
y = pd.to_numeric(dataset.loc[:, 'Value'].values)

# Splitting the dataset into the Training set and Test set
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.4, random_state = 0)
```

**Şekil 12:** Verilerin Eğitilmiş ve Test Kümelerine Ayrılması

Yukarıdaki kod satırlarında öncelikle x ve y parametlerinin ne olacağına karar verilmiştir. X parametresinde 7. sütunda bulunan bütün oyuncuların ‘Overall’ değeri, y parametresinde ise verisi temizlenen ‘Value’ sütunundaki veriler entegre edilmiştir.

Bu x ve y parametrelerindeki bütün verileri , Scikit- Learn kütüphanesinde bulunan train\_test\_split metoduyla X\_train, X\_test, y\_train, y\_test şeklinde kümelere ayrılmıştır. Bu ayrılma oranı bütün verilerin 10’da 6’sının eğitilmiş, 10’da 4’ünün test kümesinde olacak şekilde veriler kümelendirilmiştir.

#### 4.5. Verilere Lineer Regresyon Metotunun Uygulanması

Veriler her iki kümeye de ayrıldıktan sonra eğitilmiş olan x ve y parametrelerini kullanarak ‘Lineer Regresyon’ metodu aşağıdaki gibi uygulanmıştır.

```
from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(X_train, y_train)
```

**Şekil 13:** Lineer Regresyonun Uygulanması

Lineer regresyon uygulandıktan sonra bu regresyon sonu tahminleme mekanizması ise aşağıdaki gibidir.

```
y_pred = regressor.predict(X_test)
```

**Şekil 14:** Tahminleme İşlemi

Bu tahminleme işleminde y parametresi yani ‘Overall’ puanı bilinen bir futbolcunun transfer değerinin ne kadar olması gerektiğini tahmin etmektedir. Tahminlemede x parametresinin test kümesi verileri kullanılmıştır.

```
print(regressor.predict(94))
```

**Şekil 15:** 94 ‘Overall’ puanına sahip bir oyuncunun transfer değerinin tahmini çıktı olarak veren kod satırı

Bu tahminleme kod satırlarından sonra derlenen kod satırları aşağıdaki gibi bir tahminleme yapmaktadır.

```
[15802619.50682703]
```

**Şekil 16:** Tahmin Sonucu

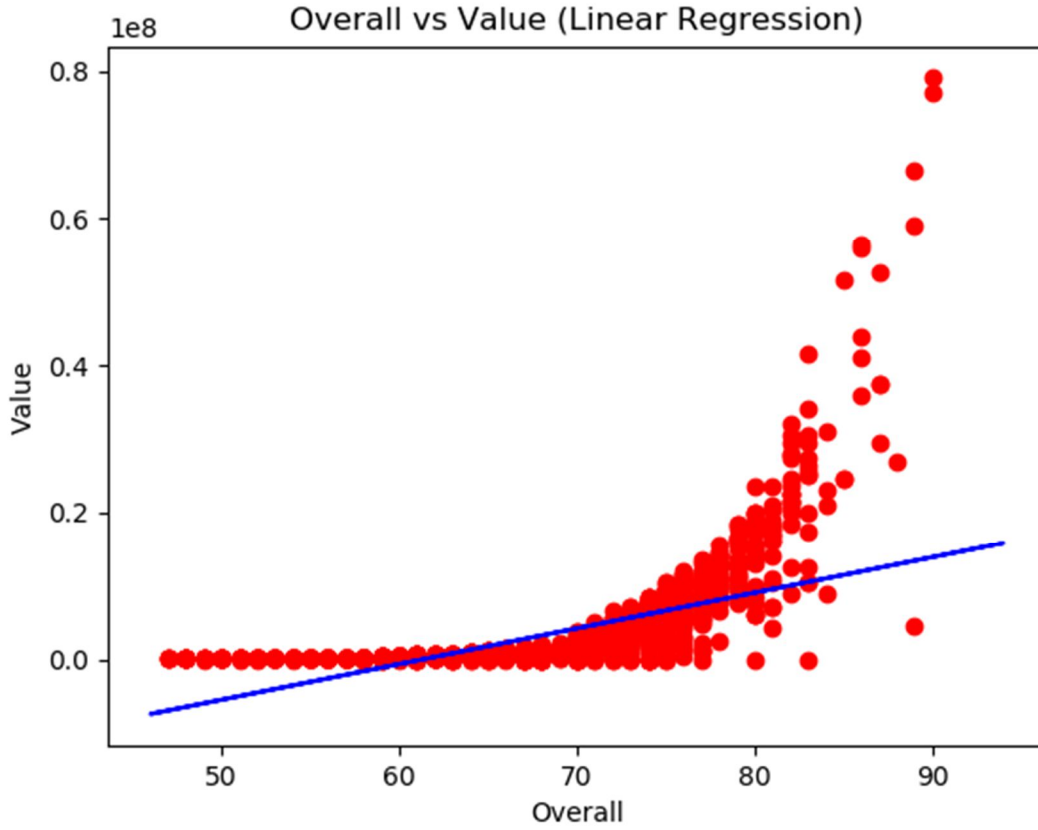
Lineer regresyona göre 94 ‘Overall’ puanına sahip olan bir futbolcu yaklaşık 16 milyon Avro değerinde bir transfer değerine sahiptir.

Ayrılan veri kümelerinin ve lineer regresyonun koordinat ekseninde gösterimi ise aşağıdaki gibidir.

```
plt.scatter(X_test, y_test, color = 'red')
plt.plot(X_train, regressor.predict(X_train), color = 'blue')
plt.title('Overall vs Value (Linear Regression)')
plt.xlabel('Overall')
plt.ylabel('Value')
plt.show()
```

**Şekil 17:** Lineer Regresyonun Figür Olarak Gösterimi

Yukarıdaki kod satırlarının çıktısı ise aşağıdaki gibidir.



Şekil 18: Linear Regresyon Figürü

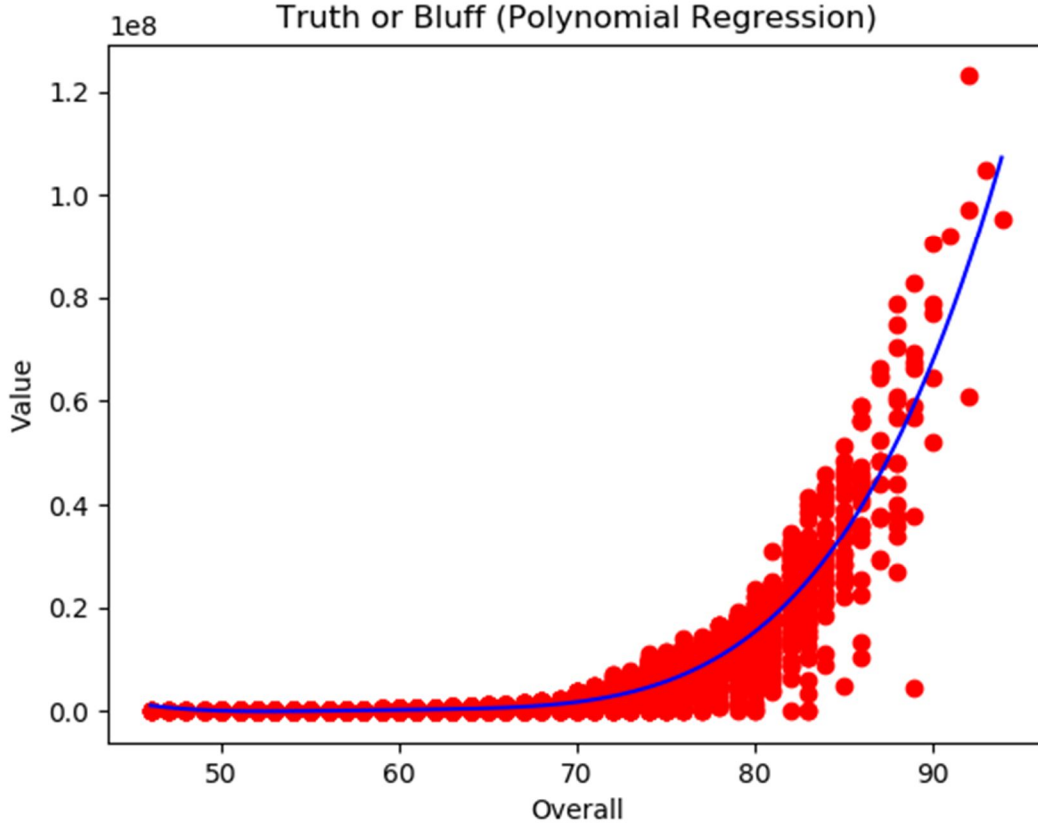
#### 4.6. Verilere Polinom Regresyon Metotunun Uygulanması

Eğitilmiş ve test kümelerine ayrılan verileri kullanılarak gerçekleştirilen bir regresyondur. Yine Scikit-Learn kütüphanesinden PolynomialFeatures metoduyla bu regresyon gerçekleştirilmektedir.

```
from sklearn.preprocessing import PolynomialFeatures
poly_reg = PolynomialFeatures(degree = 4)
X_poly = poly_reg.fit_transform(X)
poly_reg.fit(X_poly, y)
lin_reg_2 = LinearRegression()
lin_reg_2.fit(X_poly, y)
```

Şekil 19: Polinom Regresyonun Uygulanması

Lineer regresyonda kullanılan görselleştirme yöntemine benzer bir yöntemle polinom regresyonun figürü aşağıdaki gibidir.



**Şekil 20:** Polinom Regresyon Figürü

Bu figürde dikkat edileceği üzere regresyon çizgisi verilerin oluşturduğu eğri şeklindedir. Bu nedenle polinom regresyonda tahminleme başarımı oldukça iyi olmakla birlikte gerçeğe yakın sonuçlar da vermektedir. FIFA 2018 veri setinin içinde bulunan verilerde en yüksek ‘Overall’ puanına sahip olan oyuncu ‘Cristiano Ronaldo’dur ve 94 puana sahiptir. Transfer değeri ise 95.5 milyon Avro’dur. Polinom regresyonda yapılan tahminlemede ise bu sonuç aşağıdaki gibidir.

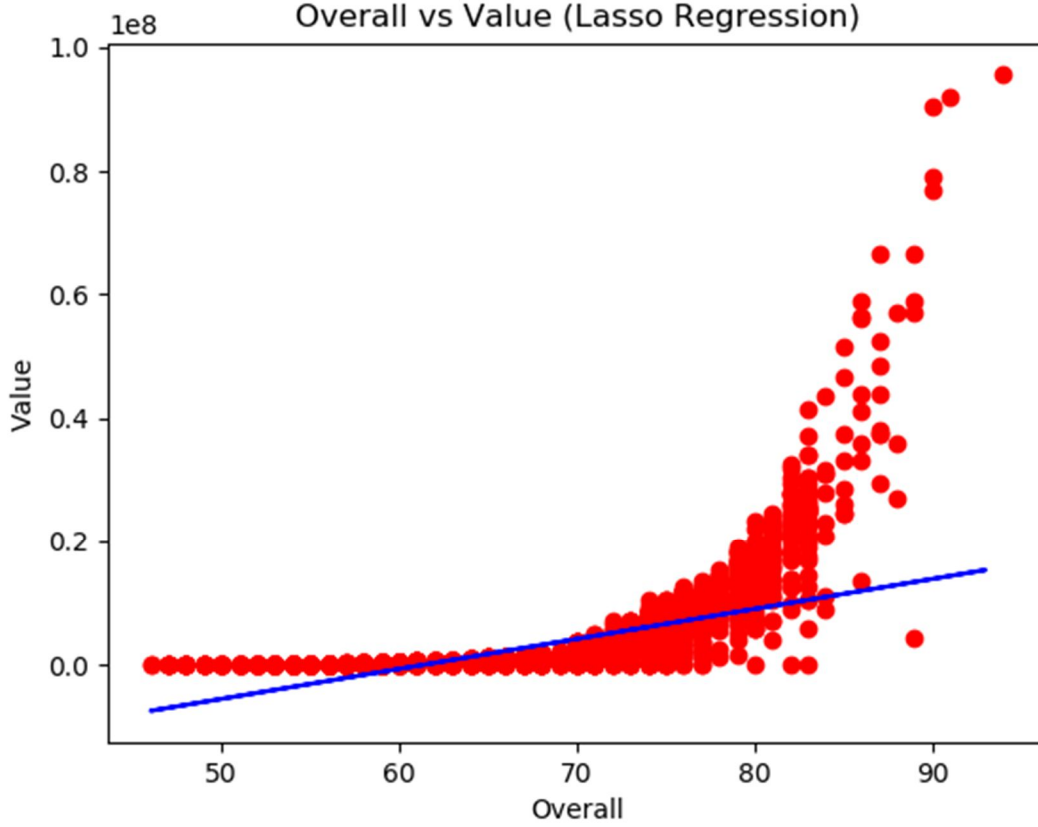
[1.08421296e+08]

**Şekil 21:** 94 ‘Overall’ puanına sahip futbolcunun tahmin edilen transfer değeri (108.42 milyon Avro)

#### 4.7. Verilere LASSO Metotunun Uygulanması

Yapılan uygulamada yukarıdaki iki regresyondaki gibi görselleştirme de ve kodlama bariz bir fark bulunmamaktadır. Bu metotta da x ve y parametrelerinin eğitilmiş ve test verileri

kullanılmıştır. Aralarında ki fark ise eğitilen verilerin değerleri ve tahminlemede ortaya çıkan sonuçtur. LASSO metotunun figürü aşağıdaki gibidir.



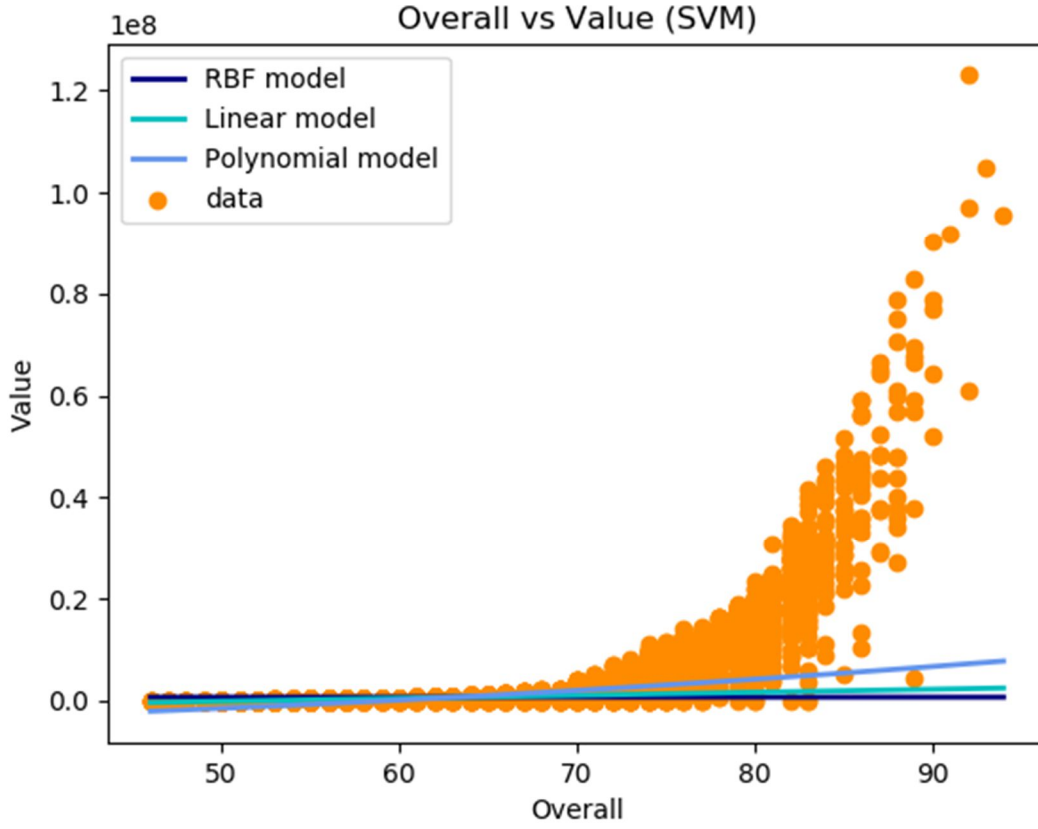
**Şekil 22:** LASSO Metotunun Figürü

[15802619.45017958]

**Şekil 23:** LASSO Metotunun 94 puana sahip futbolcunun transfer değerinin tahmini

#### 4.8. Verilere SVR Uygulanması

SVR metotunda daha çok eğitilen veriler ve bu verilerin alt kümeleri üzerinden çalışmalar yapar. Bu sebepten dolayı yukarıdaki regresyonların tahminlemelerinden farklı olarak bir tahminde bulunmaktadır ve görselleştirilmiş hali diğerlerinden farklıdır. Aşağıda bulunan figürde SVM üzerinde veri madenciliğinde kullanılan ‘Ridge Regression’ doğrusu, lineer doğru ve polinomik eğri bulunmaktadır.



**Şekil 24: SVM Figürü**

SVM'nin 94 puanlık oyuncu için tahmin ettiği transfer değeri de aşağıdaki gibidir.

**SVR Tahmin(94)= 699518.57680388**

**Şekil 25: SVR Tahmini**

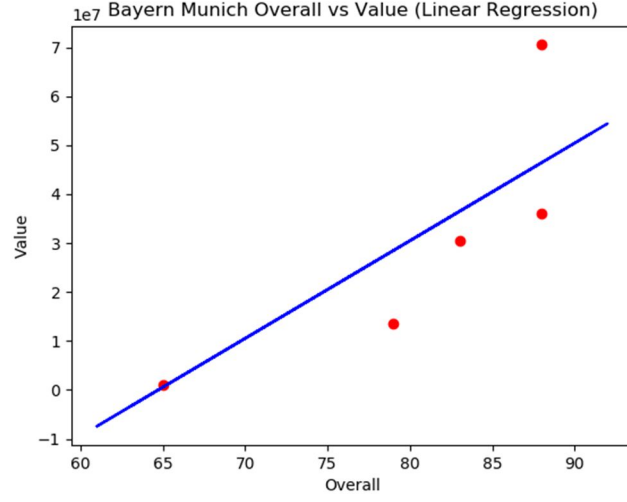
Avrupanın altı büyük takımını ayrı ayrı verilerinin incelenip görselleştirme çalışmaları yapılmıştır. Bu altı takım verilerinde linner ve polinom regresyonları kullanılmış ve bu regresyonlar üzerinden tahminlemeler yapılmıştır. Bu çalışmaların görselleri 'Ekler' kısmında bulunmaktadır. Bu görsel ve tahminler üzerinden bazı çıkarımlar yapmak mümkündür. Görsellerden de anlaşılacağı üzere hangi takımın transfer bütçesi daha fazla, hangi takımlarda puanı yüksek oyuncular var gibi soruların cevabı rahatlıkla bulunabilmektedir.

## 5. SONUÇ

Çalışma kapsamında video oyun analizi hakkında bilgi verilmiş, dünyadaki örnek çalışmalardan bahsedilmiş, veri madenciliği ve Python programlama dili ile ilgili önemli noktalara değinilmiş, Python kütüphanelerinin veri madenciliğinde nasıl kullanıldığı kavranmış ve uygulanmış, bu uygulamalar sonucunda “FIFA 2018 Verisi Üzerinde Veri Madenciliği Uygulaması” adlı somut bir proje ortaya çıkmıştır. Bu projede veri madenciliği tekniklerinin bazıları değinilmiş, bu proje üzerinde kullanılan tekniklerin tanımları ve arasındaki farkları açıklanmış ve projede bu teknikler kullanılarak elde edilen sonuçlar karşılaştırılmıştır. Ayrıca bu projeye ileriki zamanlarda da devam edilecek olup aynı veri kümesi üzerinde farklı veri madenciliği teknikleri kullanılarak farklı çıkarımlar elde edilecektir.



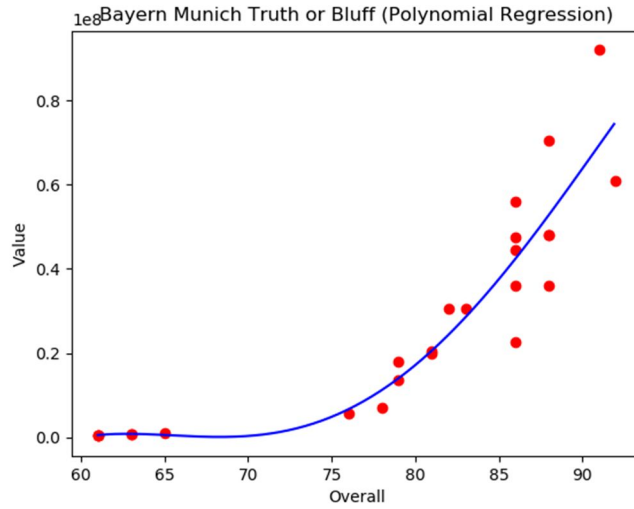
## EKLER



**Ek 1:** Bayern Munich Takımının Linear Regresyon Figürü

[58407093.84460142]

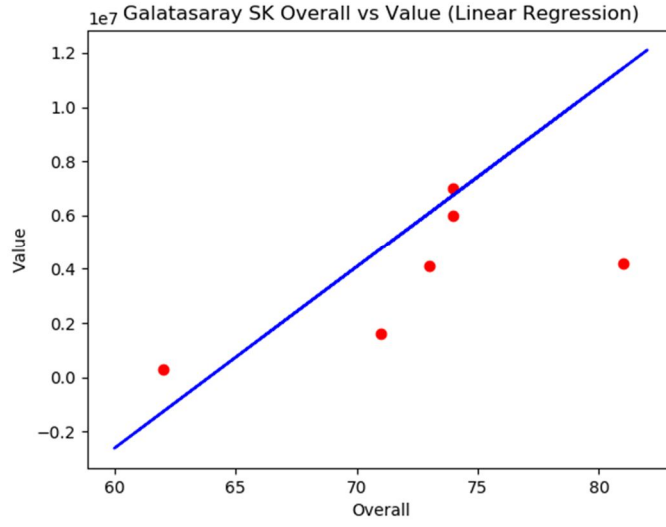
**Ek 2:** Bayern Munich Verisinin Linear Regresyon Tahmini (58.4 milyon Avro)



**Ek 3:** Bayern Munich Takımının Polinom Regresyon Figürü

[85919188.25260162]

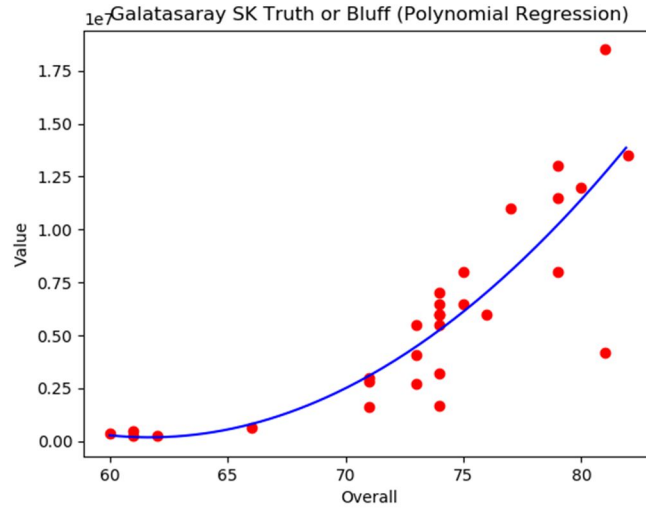
**Ek 4:** Bayern Munich Verisinin Polinom Regresyon Tahmini (85.9 milyon Avro)



**Ek 5:** Galatasaray Takımının Linner Regresyon Figürü

[16121500.8699212]

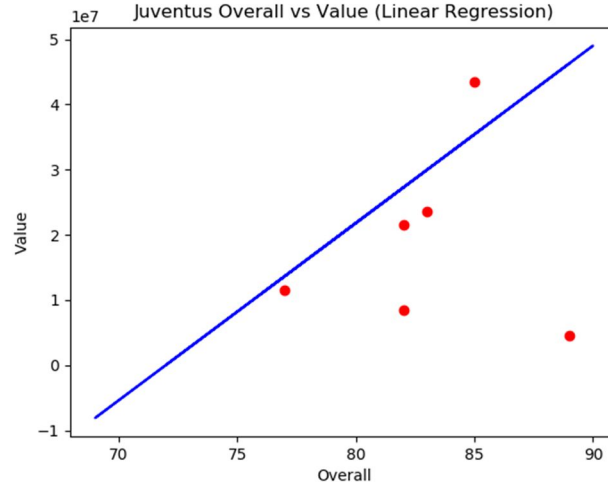
**Ek 6:** Galatasaray Verisinin Lineer Regresyon Tahmini (16.1 milyon Avro)



**Ek 7:** Galatasaray Takımının Polinom Regresyon Figürü

[35129466.23903748]

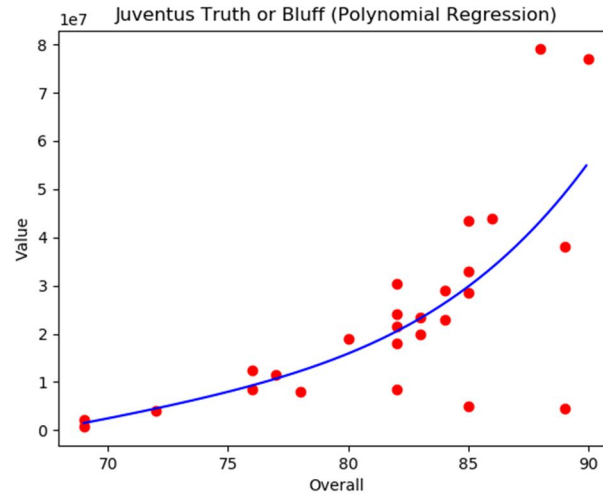
**Ek 8:** Galatasaray Verisinin Polinom Regresyon Tahmini (35.1 milyon Avro)



**Ek 9:** Juventus Takımının Linner Regresyon Figürü

[59882431.39979157]

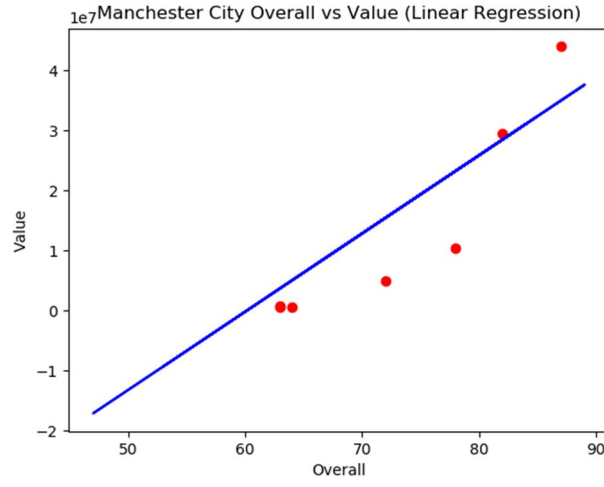
**Ek 10:** Juventus Verisinin Lineer Regresyon Tahmini (59.8 milyon Avro)



**Ek 11:** Juventus Takımının Polinom Regresyon Figürü

[89962317.36191034]

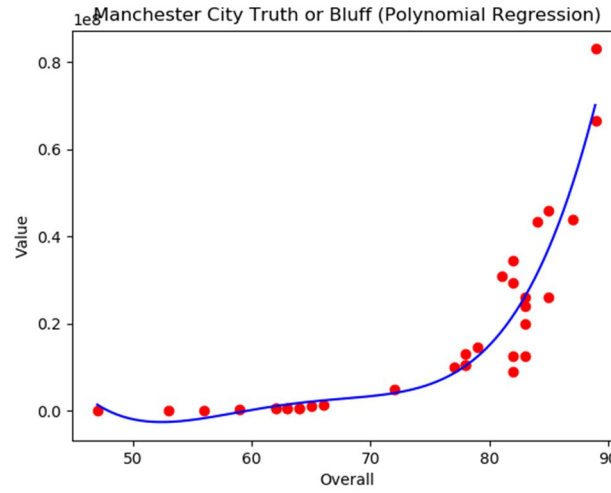
**Ek 12:** Juventus Verisinin Lineer Regresyon Tahmini (89.9 milyon Avro)



**Ek 13:** Manchester City Takımının Linner Regresyon Figürü

[44140073.18019313]

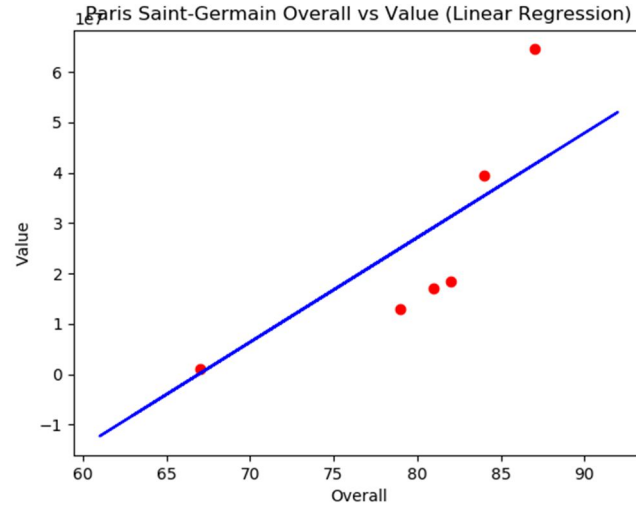
**Ek 14:** Manchester City Verisinin Lineer Regresyon Tahmini (44.1 milyon Avro)



**Ek 15:** Manchester City Takımının Polinom Regresyon Figürü

[1.43511926e+08]

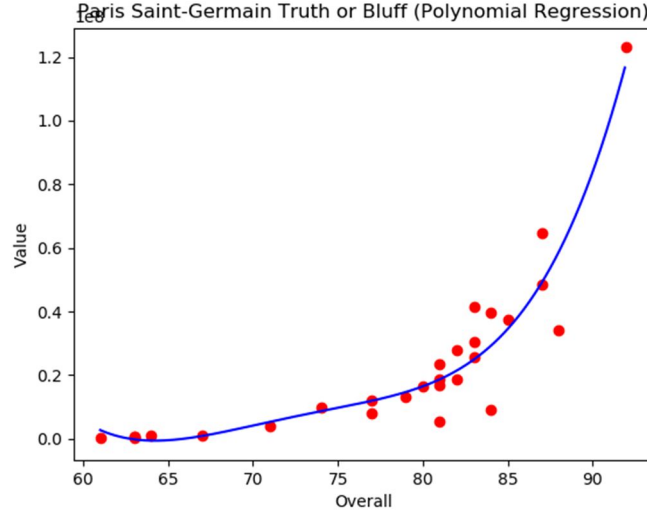
**Ek 16:** Manchester City Verisinin Polinom Regresyon Tahmini (143.5 milyon Avro)



**Ek 17:** Paris Saint-Germain Takımının Linner Regresyon Figürü

[56171139.69198197]

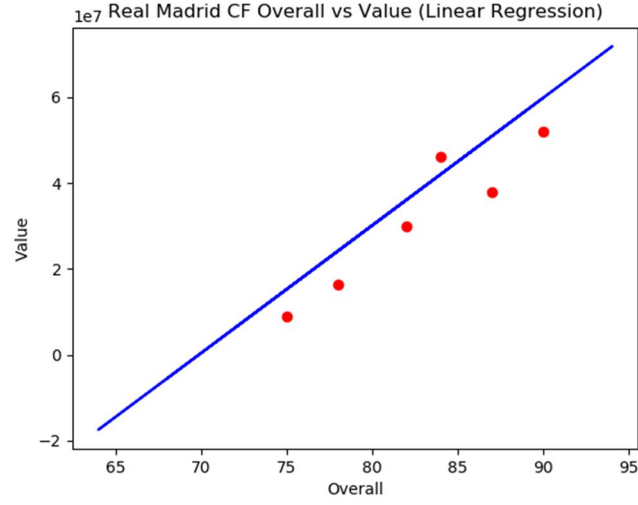
**Ek 18:** Paris Saint-Germain Verisinin Lineer Regresyon Tahmini (56.1 milyon Avro)



**Ek 19:** Paris Saint-Germain Takımının Polinom Regresyon Figürü

[1.65550056e+08]

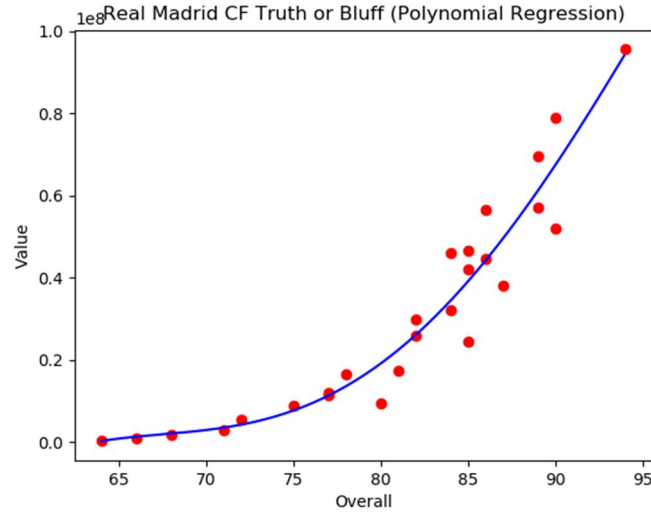
**Ek 20:** Paris Saint-Germain Verisinin Polinom Regresyon Tahmini (165.5 milyon Avro)



**Ek 21:** Real Madrid Takımının Linner Regresyon Figürü

[71757790.34772834]

**Ek 22:** Real Madrid Verisinin Lineer Regresyon Tahmini (71.7 milyon Avro)



**Ek 23:** Real Madrid Takımının Polinom Regresyon Figürü

[94647638.57109165]

**Ek 24:** Real Madrid Verisinin Polinom Regresyon Tahmini (94.6 milyon Avro)

## KAYNAKÇA

- [1] Yıldız, Mümin - Telekomünikasyon Sektöründe Müşteri Sadakatinin Veri Madenciliği Teknikleriyle Modellenmesi – 2017
- [2] Leonardo Cotta, Pedro O.S. Vaz de Melo, Fabrício Benevenuto, Antonio A.F. Loureiro – Using Soccer Video Game Data for Soccer Analytics
- [3] Result Prediction by Mining Replays in Dota 2, Filip Johansson, Jesper Wikström – 2015
- [4] Game Data Mining: Clustering and Visualization of Online Game Data in Cyber-Physical Worlds,, Peter Braun, Alfredo Cuzzocrea, Timothy D. Keding, Carson K. Leung, Adam G.M. Padzor, Dell Sayson (2017)
- [5] Spiers, N., Manktelow, B. Ve Hewitt, M. J.(2009), Practical Statistics Using SPSS, National Institute for Health Research NHS, England.
- [6] [http://scikit-learn.org/stable/modules/linear\\_model.html](http://scikit-learn.org/stable/modules/linear_model.html)
- [7] [http://scikit-learn.org/stable/modules/linear\\_model.html#polynomial-regression-extending-linear-models-with-basis-functions](http://scikit-learn.org/stable/modules/linear_model.html#polynomial-regression-extending-linear-models-with-basis-functions)
- [8] [http://scikit-learn.org/stable/modules/linear\\_model.html#lasso](http://scikit-learn.org/stable/modules/linear_model.html#lasso)
- [9] <http://scikit-learn.org/stable/modules/svm.html#regression>
- [10] python.org
- [11] <https://spectrum.ieee.org/computing/software/the-2017-top-programming-languages>
- [12] Python for Data Analysis, 2nd Edition Data Wrangling with Pandas, NumPy, and IPython, William McKinney, 2017

## ÖZGEÇMİŞ

Ad Soyad : Selçuk HACIAHMET

Doğum Tarihi : 04.08.1995

Doğum Yeri : Bakırköy/İSTANBUL

Lise : 2009 – 2013 Bursa Turhan Tayan Anadolu Lisesi

Staj Yaptığı Yerler : Yapı ve Kredi Bankası A. Ş. Bilişim Teknolojileri Müdürlüğü –  
Gebze – (2016/4 hafta)

Yapı ve Kredi Bankası A. Ş. Bilişim Teknolojileri Müdürlüğü –  
Gebze – (2017/4 hafta)

Gamelab İstanbul Oyun Geliştirme Merkezi – İstanbul – (2017/20  
iş günü)

Encode IT Management and Consultancy – İstanbul – (2018,  
halen devam ediliyor.)