

T.C.
İSTANBUL KÜLTÜR ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

ASTROFİZİKTE MAKİNE ÖĞRENMESİ ALGORİTMALARI KULLANILARAK
GALAKSİLERİN ÖZELLİKLERİNİN TESPİT EDİLMESİ

YÜKSEK LİSANS TEZİ

Harun Tayfun SÖYLEMEZ
1800005654

Anabilim Dalı: Fizik

Program: Fizik

Tez Danışmanı: Dr. Öğr. Üyesi Ayşegül F. YELKENCİ

Nisan 2021

T.C.
İSTANBUL KÜLTÜR ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

ASTROFİZİKTE MAKİNE ÖĞRENMESİ ALGORİTMALARI KULLANILARAK
GALAKSİLERİN ÖZELLİKLERİNİN TESPİT EDİLMESİ

YÜKSEK LİSANS TEZİ

Harun Tayfun SÖYLEMEZ
1800005654

Anabilim Dalı: Fizik

Program: Fizik

Tez Danışmanı: Dr. Öğr. Üyesi Ayşegül F. YELKENCİ

Jüri Üyeleri: Dr. Öğr. Üyesi Sinan Aliş (İstanbul Üniversitesi)

Prof.Dr. Gülce Öğrüş İldız

Nisan 2021

ÖNSÖZ

Bu tez çalışmasında SDSS DR16 kataloğundaki galaksilerde karanlık madde halo kütlesi, MICECATv2 mock kataloğundaki galaksiler kullanılarak eğitilmiş bir makine öğrenimi programı ile tahmin edilmiştir.

Tez konusunu seçerken isteklerimi göz önünde bulundurup bana yardımcı olan tez danışmanım Dr. Öğr. Üyesi Ayşegül F. Yelkenci'ye teşekkürlerimi sunarım.

İstanbul 2021

Harun Tayfun Söylemez

İçindekiler Tablosu

KISALTMALAR	iii
TABLO LİSTESİ	iv
ŞEKİL LİSTESİ	v
SİMGE LİSTESİ	vi
KISA ÖZET	vii
ABSTRACT	viii
1. GİRİŞ	1
1.1. Karanlık madde	2
1.2 Karanlık Maddenin Kanıtı ve Galaksilerin Kütle Hesapları	6
1.2.1 Viryal Teoremi Yöntemi	6
1.2.2 X-Işınımı Yöntemi	8
1.2.3 Kütleçekimsel Mercek Etkisi Yöntemi	9
1.3 Makine Öğrenmesi ile Galaksi Halo Kütlelerinin Tayini	13
2. VERİ VE YÖNTEM	13
2.1 SDSS Galaksi Kataloğu	14
2.1.1 Veri Ulaşımı	16
2.2 MICECAT Galaksi Kataloğu	20
2.3 MAKİNE ÖĞRENİMİ VE KARANLIK MADDE KÜTLESİ TAHMİNİ	23
2.3.1 Makine Öğrenimi Uygulaması	24
2.3.2 Algoritma- SVR (Destek Vektör Regresyonu)	27
2.3.3 Veri Analizi	28
2.3.4 SDSS Kataloğundaki Galaksilerin Karanlık Madde Kütlelerinin Tahmini	31
3. SONUÇLAR	33

3.1 Bulgular	34
3.2 Tartışma.....	35
KAYNAKÇA.....	36



KISALTMALAR

SDSS	: Sloan Digital Sky Survey
KM	: Karanlık Madde
Mpc	: Megaparsek
SVR	: Support Vector Regression
SVM	: Support Vector Machine
ML	: Machine Learning
DR16	: Data Release 16
BOSS	: Baryon Oscillation Spectroscopic Survey
APOGEE	: APO Galactic Evolution Experiment
MaNGA	: Mapping Nearby Galaxies at APO
MYSQL	: My Structured Query Language
MiceCat	: Marenstrum Institut de Ciències de l'Espai Simulations
MICE-GC	: MICE-Grand Challenge
CSV	: Comma Separated Values
FITS	: Flexible Image Transport System
ASDF	: Advanced Scientific Data Format
X-Ray	: Energetic High-Frequency Electromagnetic Radiation

TABLO LİSTESİ

Tablo-1 Güçlü ve zayıf merceklenme örnekleri [6]	11
Tablo-2 SQL dizisinde “type” parametresinin anlamı	17
Tablo-3 SQL dizisinde “mode” parametresinin anlamı	18
Tablo-4 SDSS DR16 veri tabanından kullanılan galaksilerin özellikleri	18
Tablo-5 MICECAT v2 veri tabanından çekilen galaksi özellikleri	21
Tablo-6 MICECAT v2 veri tabanından ayıklanmış 20000 farklı galaksiye ait verilerden örnek.....	24
Tablo-7 Sıcaklık haritasına göre, aralarındaki ilişkinin en fazla olduğu özellikler [Ek-1 Out(8)]	27
Tablo-8 (ML) SVR algoritmasının galaksilerin karanlık madde halo kütlesi tahminleri karşılaştırması (Ek-1 Out (24))	29
Tablo-9 SDSS DR16 kataloğundan alınan 20000 galaksi içinden alınan ilk beş galaksinin değerleri (Ek-1 Out(29))	31
Tablo-10 SDSS DR16 kataloğundaki galaksiler için karanlık madde halo kütlesi tahmini (Ek-1 Out(35)).....	32

ŞEKİL LİSTESİ

Şekil-1 Plank'tan sonra evrendeki karanlık madde miktarı [1]	1
Şekil-2. a. Galaksi kümesindeki galaksilerin galaksi merkezinden uzaklığı ile açısal hızlarının değişimi, b. galaksilerin gerçek dönme hızları eğrisi	3
Şekil-3 a. Paneli Vigro galaksi kümesindeki galaksilerin konumları ve hızları; b paneli x-ışınımı görüntüsü [6].....	8
Şekil-4 X-ray ışınımı Hydra ve Coma galaksi kümelerinin x-ışını görüntüleri [6].....	9
Şekil-5 Galaksi kümesindeki kütleçekimsel mercek etkileri [8].....	11
Şekil-6 B 1938+666 kütleçekimsel mercek sistemi [9]	12
Şekil-7 $\theta E, Dd, Ds, DdS, M$ parametrelerinin açıklaması [10].....	12
Şekil-8 SDSS Sloan teleskobu, New Mexico [22]	14
Şekil-9 Ekvator koordinatlarında DR16 eBOSS spektroskopik kapsama alanı [22].	15
Şekil-10 DR16 APOGEE alanlarının dağılımı [22].	16
Şekil-11 DR15 / DR16 MaNGA Ekvatordaki spektroskopik kapsama alanı [22].....	16
Şekil-12 DR16 veri tabanından galaksi verilerinin indirilmesi ¹	18
Şekil-13 MICECAT v2 galaksi kataloğundan seçilen galaksi özelliklerinin birbirleriyle olan ilişkilerini gösteren sıcaklık haritası [Ek-1 Out(5)].....	26
Şekil-14 SVR algoritmasındaki hiper düzlem [33]	27
Şekil-15 Test verisindeki 20000 galaksinin karanlık madde halo logaritmik kütlesine ilişkin SVR tahmini ve hedef değerlerin dağılımı ($\text{dex}M_{\text{Güneş}}$) (Ek-1 Out(25)).....	30
Şekil-16 Test veri setindeki 20000 galaksinin kütlesi ile galaksilerin (ML) SVR algoritması ile tahmin edilen karanlık madde halo kütlesi ($\text{dex}M_{\text{Güneş}}$) (Ek-1 Out(27))	31
Şekil-17 SDSS DR16 veri setinden alınmış 20000 galaksi kütlesi değerleri ile galaksilerin (ML) SVR algoritması ile tahmin edilen karanlık madde halo kütlesi dağılımı ($\text{dex}M_{\text{Güneş}}$) (Ek-1 Out(36)).....	33

SİMGE LİSTESİ

E_k : Kinetik Enerji

E_p : Potansiyel Enerji

G : Evrensel çekim sabiti

R_{cl} : Galaksinin çapı

σ : Galaksi (veya galaksilerin) hız dispersiyonu

M_{cl} : Galaksinin kütlesi

M_{\odot} : Güneş kütlesi

L_{cl} : Galaksinin parlaklığı

L_{\odot} : Güneşin parlaklığı

c : Işık hızı

ϵ : Hangi kanunların geçerli olduğunu gösteren boyutsuz bir sayı

θ_E : Açısal yarıçap

D_s : Kaynak ile gözlemci arasındaki uzaklık

D_L : Kütle çekim merceğinin gözlemciden olan uzaklığı

D_{LS} : Kütle çekim merceğinin kaynağa olan uzaklığı

z : Kırmızıya kayma

Üniversite : İstanbul Kültür Üniversitesi
Enstitüsü : Lisansüstü Eğitimi Enstitüsü
Dalı : Fizik
Programı : Fizik Tezli Yüksek Lisans
Tez Danışmanı : Dr. Öğr. Üyesi Ayşegül F. Yelkenci
Tez Türü ve Tarihi : Yüksek Lisans – Nisan 2021

KISA ÖZET

ASTROFİZİKTE MAKİNE ÖĞRENMESİ ALGORİTMALARI KULLANARAK GALAKSİLERİN ÖZELLİKLERİNİN TESPİT EDİLMESİ Harun Tayfun SÖYLEMEZ

Bu çalışmada galaksilerin halosundaki karanlık maddenin makine öğrenimi algoritması kullanılarak belirlenen kütle miktarı tahminin, matematiksel yöntemlerle elde edilmiş kütle sonuçlarına ne ölçüde yaklaştığı araştırılmaktadır.

Araştırmada galaksilerin halo kısımlarındaki karanlık madde miktarının tahmini ile bunun için hızlı ve daha az işlem zamanı gerektiren bir yöntem geliştirilmesi amaçlanmıştır. MICECAT v2 sentetik galaksi kataloğundan çekilen 20000 galaksiyi içeren veri setinin yüzde 80'ni makine öğrenme algoritmasında modelin eğitilmesi için kullanılmıştır. Geriye kalan veriler modelin testi için kullanılmış ve tahmin edilen sonuçlarla test verileri karşılaştırılmış. Sonuç olarak geliştirilen makine öğrenme SVR algoritması modelinin eğitiminde gerçek değerle tahmin edilen değer arasında hesaplanan standart sapmanın 0.00082 olduğu hesaplanmıştır. Modelin güvenilirliği test edildikten sonra gözlemsel galaksi verileri için geliştirilen model kullanılarak SDSS DR16 kataloğunda yer alan 20000 farklı galaksinin halosundaki karanlık madde kütlesi belirlenmiştir. 10^7 ile 10^{10} $M_{\text{Güneş}}$ arasında kütleyle sahip galaksiler için karanlık madde halo kütlelerinin galaksi görünür kütlelerinden yaklaşık 100 kat fazla olduğu görülmüştür. Galaksilerin Halo kütlesi arttıkça görünür kütle değerinin de arttığı tespit edilmiştir. Çalışmamız makine öğrenimi algoritmaları kullanılarak galaksilerdeki karanlık madde miktarını hızlı bir şekilde tahmin edilebildiğini göstermiştir. Bu çalışmada $z=0$ değerine sahip galaksiler incelenmiştir.

Anahtar Sözcükler: Karanlık madde, halo kütlesi, makine öğrenimi, SDSS galaksi kataloğu, SVD algoritması, galaksiler

Bilim Dalı Sayısal Kodu:

University : Istanbul Kültür University
Institute : Institute of Graduate Education
Department : Physics
Programme : Physics Master's Degree (Thesis)
Tez Danışmanı : Dr. Ayşegül F. Yelkenci
Tez Türü ve Tarihi : MS – April 2021

ABSTRACT

DETERMINING THE PROPERTIES OF GALAXIES USING MACHINE LEARNING ALGORITHMS IN ASTROPHYSICS

Harun Tayfun SÖYLEMEZ

In this study, the relationship between the amount of dark matter in galaxies determined using the machine learning algorithms and the mass results obtained by mathematical methods regarding the dark matter in the halos of galaxies is investigated.

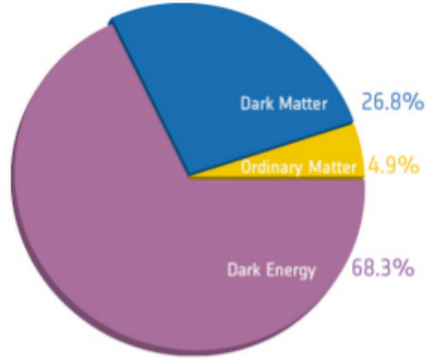
The aim of this research is to estimate the amount of dark matter in the halo parts of galaxies and to develop a method that requires faster and less processing time for this analysis. 80 percent of the dataset including 20000 galaxies from the MICECAT v2 synthetic galaxy catalog was used to train the model in the machine learning SVR algorithm. The remaining data were used for testing the model, and the predicted results were compared with the test data. The standard deviation of prediction data and real data in training model is calculated as 0.00082. After testing the reliability of the model, dark matter halo masses of the 20000 different observed galaxies from SDSS DR16 catalog were determined with our trained model. As a result dark matter halos are found to be approximately 100 times more massive than the stellar masses for galaxies with 10^7 to 10^{10} M_{solar} masses. Dark matter halo mass is observed to be increasing with the raising galaxy stellar masses. Our study also showed that it was possible to quickly estimate the amount of dark matter in galaxies using a machine learning algorithm. All galaxies z value is zero in this study.

Keywords: Dark matter, halo mass, machine learning, SDSS galaxy catalog, SVD algorithm, galaxies.

Science Code :

1. GİRİŞ

Karanlık maddenin (KM) temel doğasını ortaya çıkarma arayışı, Astrofizik, Parçacık Fiziği ve Kozmolojinin kesişme noktasında devam etmektedir. Astrofiziksel veya parçacık çarpıştırıcı aramalarından bugüne kadar KM'nin varlığına dair doğrudan bir kanıt elde edilmemiştir, sadece astrofiziksel gözlemlerden çıkarılan kütle çekim etkileri, bu görünmez parçacık hakkında ipuçları sunmuştur. Yapılan çalışmalar sonucu, Şekil-1'de görüldüğü gibi evrende karanlık madde miktarı evrenin yaklaşık yüzde 26.8'ni ve normal madde miktarı da evrenin yaklaşık yüzde 4.9 nu oluşturduğu hesaplanmıştır.



Şekil-1 Plank'tan sonra evrendeki karanlık madde miktarı [1]

Evrenimizde meydana gelen, bilinen görünür madde yardımıyla açıklanamayan olaylar vardır. Bunlardan biri, evrenin hızlanarak genişlemesi, diğeri ise karanlık maddedir. Evrenimizdeki galaksiler yüksek hızlarda dönmektedirler ve gözlemlenebilir maddelerinin oluşturduğu kütle çekimi galaksileri bir arada tutamayacak kadar zayıf kalmaktadır. Hesaplamalarda sadece görünür maddelerin kütle çekiminin kullanılması halinde galaksilerin dağılması gerektiği hesaplanmıştır ancak gözlemler bunun böyle olmadığını göstermektedir. Bu sebepten dolayı, galaksileri bir arada tutan ve kütle çekimi oluşturan bir maddenin varlığının olması gerektiği düşüncesi ortaya çıkmıştır.

Ünlü gökbilimci Zwicky, arařtırmalarını Coma Takımı Yıldızındaki 90 Megaparsek uzaklıktaki gökada kümesinde ve Bařak takımı Yıldızındaki kümesinde, viryal yöntemiyle galaksi kümelerinin kütlelerini hesaplamıřtır. Daha sonra aynı galaksi kümelerinin kütle-parlaklık iliřkisine bakarak her bir kümenin kütlelerini hesaplamıř ve bu iki kütle arasında büyük bir çeliřki bulmuřtur ve daha sonra da karanlık maddenin varlıđını öngörmüřtür.

Karanlık maddenin varlıđının kabul edilmesinin ardından simülasyonlar karanlık madde varsayımı üzerine geniř ölçekteki evren, yapı oluřumu ve galaksi evrimini inceliyor ve mock dediđimiz simülasyon verilerine dayanan galaksi katalogları oluřturuluyor. Bu projede kullanacađımız MICECATv2 mock katalođu “Marenostrum Institut de Ciències de l'Espace Simulations” tarafından oluřturulmuř bir katalogdur [2]. Simülasyon hesapları yüksek iřlem gücü ve zaman gerektiren bir çalıřmadır. Bu çalıřmada, karanlık madde keřif çalıřmalarını günümüz teknolojisi olan yapay zekâ programlaması ile birleřtirerek makine öđrenme algoritmaları kullanarak daha hızlı sonuçlar alınabileceđi düşünölmüř ve var olan katalog verileri ile öđrenen bir program geliřtirilmiřtir. Makine örenme algoritması kullanılarak henüz hesaplanmayan galaksilerdeki karanlık madde miktarı bu program tarafından tahmin edilebilmiřtir.

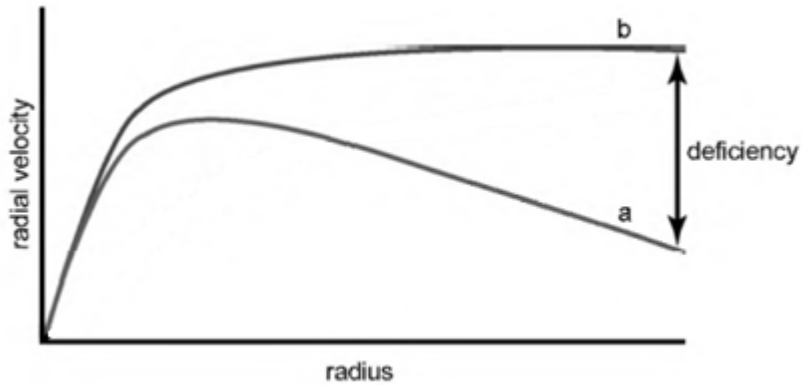
1.1. Karanlık madde

1933 yılında İsviçreli gökbilimci Fritz Zwicky'nin yaptıđı bir gözlemlerle karanlık madde fikri ortaya çıkmıřtır. Galaksiler, kütle çekim etkileřimleriyle galaksi kümeleri oluřturur ve bunlar da bir araya gelerek daha büyük galaksi kümelerini oluřturur. Ünlü astronom Zwicky, Coma galaksi kümesindeki galaksilerin hareketini incelerken, galaksilerin hızlı hareket ettiklerini gözlemlemiř ve incelediđi galaksi kümesinde galaksilerin hareketlerinin rastgele hareket etmediđini, belli bir yörünge izlediđini görmüřtür [3].

Galaksilerdeki parlaklık galaksinin içindeki yıldızlar ve toz bulutlarından gelir ve böylece galaksideki yıldız sayısı hesaplanabilir. Zwicky, gözlemlediđi galaksi kümesinde her bir galaksinin kütlelerini parlaklıklarına bakarak ve Viryal yöntemini kullanarak galaksi kümesinin kütlelerini hesaplamıřtır. Zwicky'nin yaptıđı bu iki farklı hesaplama sonucunda galaksi kümesinin kütlelerini birbirinden nerdeyse 400 kat daha fazla olduđunu görmüřtür ve galaksinin yıldız sayısının gözlemlenenenden daha fazla

olması koşulu altında bunun açıklanabilir olduğunu düşünmüştür. Bu duruma bir açıklama getirmek gerekmiştir. Burada görünmeyen bir maddenin varlığı fikri ortaya çıkmıştır ve buna karanlık madde denilmiştir. Karanlık bir maddenin ışığa yapmadığı düşünülmüştür. Benzer bir gözlem sonucunu da 1936 yılında Sinclair Smith, Virgo kümesinde üzerinde yaparak benzer bir sonuca ulaşmıştır [3]. Kümedeki elemanların hız dağılımları incelenmiştir ve orada ışığa yapan maddeden çok daha fazla miktarda madde olması gerektiği öngörülmüştür. Fakat bu madde miktarı, gezegenler, yıldızlar, karadelikler gibi bildiğimiz gök cisimleri ile açıklanamamıştır çünkü gerekli madde miktarının hesaba katılan cisimlerden çok fazla olması gerektiği hesaplanmıştır. İki farklı küme üzerinde aynı sonuçların çıkmış olması, orada ışığa yapmayan (karanlık) bir madde fikrini desteklemeye başlamıştır.

Sarmal galaksilerin dönüş eğrisinin incelenmesi, galaksideki karanlık maddenin varlığının kanıtlarını gösterir. Sarmal bir galaksinin dönüş eğrisi incelendiğinde, galaksideki bir yıldızın veya gazın dönme hızı $v(r)$, galaktik merkezden uzaklığının bir fonksiyonu olarak ölçülür. Bu hızlar, r yarıçaplı kürenin içinde kalan kütleyle bağlı olacaktır. Sarmal bir galaksi yoğun bir merkez bölgesine sahiptir ve merkez bölgeden uzaklaştıkça görünür kütlenin yoğunluğu azaldığından, yoğun merkez bölgesinden uzaklaştıkça dönüş eğrisinde Kepler'in 3. Yasasına göre şekil-2-a'daki gibi bir düşüş beklenir. Ancak bunun yerine, gözlemsel ölçümler $v(r)$ için Kepler yasasının öngördüğü düşüşü değil, şekil-2-b'de görüldüğü gibi merkezden uzaklaştıkça neredeyse sabit kalan bir hız göstermektedir. Bu yalnızca, galakside muazzam bir görünmeyen kütle veya daha doğrusu görünmeyen karanlık maddeyi barındıran bir halonun varlığı ile mümkün olabileceği düşünülmüştür [4].



Şekil-2. a. Galaksi kümesindeki galaksilerin galaksi merkezinden uzaklığı ile açısal hızlarının değişimi, b. galaksilerin gerçek dönme hızları eğrisi

Gökbilimcilerin galaksi veya galaksi kümelerinin kütlelerini belirlemek için kullandıkları yöntemler vardır. Bunlar Viryal yöntemi, X-Işınımı ve Kütleçekimsel mercek etkisi gibi yöntemlerdir. Bu yöntemlerle ilgili detaylı açıklama bir sonraki bölümde yapılacaktır. Kısacası galaksilerin hareketlerini inceleyerek bu galaksileri veya yıldızları birbirine bağlı tutmak için gereken kütle hesaplanmaya çalışılmıştır. Nesnelerin hareketinin kuvvetlerin toplamıyla açıklanabileceğini ilk gösteren Sir Isaac Newton'dur. Farklı kuvvetler etki ettiğinde, ortaya çıkan hareket, her bir farklı kuvvetin etkisinin toplamıdır. Newton yasaları makro evrende her yerde eşit olarak geçerlidir. Hareket ve kütle çekimi arasındaki denge genellikle açıktır, belki de en iyi şekilde Newton'un bir kuvvetler dengesi terimleriyle düşünülerek görselleştirilir. Basitçe iki cisim arasındaki kütle çekimi nedeniyle Dünya'nın etrafında matematiksel hassasiyetle kusursuzca dönen Ay gibi bazı örnekler verebiliriz. Diğer bir örnek, Satürn'ün sayısız parçacıktan oluşan halkalarıdır. Halkalar bir asimetri sergiler, parçacıkların bir sokak lambasının etrafındaki güve halesi gibi gezegenin etrafında dönmezler yani uzaya doğru uçmazlar ya da gezegene düşmezler. Aslında, uzun zaman önce, Satürn'ü çevreleyen parçacıklar gezegenin bir parçasıydı. Bugün görünen halka, geriye kalan hareket ve kütle çekimi kuvvetleri arasında hassas bir denge içinde hareket eden kaya ve buz parçalarıdır. Bu denge, asteroitlerin asteroit kuşağında toplanmasında veya Samanyolu içindeki yıldızların ve gazın sarmal oluşumunda vardır.

Newton'un kütle çekimi hakkındaki fikirleri gezegen veya yıldız kütleleri gibi çeşitli ölçeklerde kullanılabilir kadar iyidir. Newton'un kanunlarının bilinen gezegenlerin çoğunun hareketlerini açıklamadaki başarısına rağmen, 1830'ların başında bir süreliğine Newton'un çekim yasalarını takip etmeyen yedinci gezegen Uranüs ile başarısız oluyor gibi görünmüştür.

İngiliz matematikçi, John Couch Adams ve Fransız Urbain Jean Joseph Leverrier 1830'ların başlarında, Uranüs'ün yörünge problemi üzerinde çalışmaya başladıktan sonra, Uranüs'ün yörüngesindeki bozuntuya Güneş'ten daha uzak başka bir gezegenin varlığının sebep olup olmadığını merak etmeye başlamışlardır. Ancak bir gezegenin Uranüs'ü tahmin edilen konumdan çekerek, kütleçekimsel bir kuvvet uygulayabileceğini düşünmüşlerdir. Adams, görünmeyen gezegenin kütlelerini, Güneş'ten uzaklığını ve en önemlisi gökyüzündeki konumunu doğru bir şekilde tahmin eden ilk kişi olmuştur. Ekim 1843'te Adams, yeni gezegeni gökyüzünde nerede

bulabileceğine dair makul bir tahminde bulunmuş, ancak Kraliyet Astronomu George Biddell Airy'nin ilgisini çekememiştir. İki yıl sonra, İngiliz Kanalı'nın diğer tarafında Leverrier, önceki sonuçlardan habersiz, Adams'inkine benzer hesaplamalar yapmıştır. Leverrier, çalışmalarını 18 Eylül 1846'da tamamladıktan ve sonuçları Alman gökbilimci Johann Gottfried Galle'ye iletmiştir. Galle, son zamanlarda şans eseri, yeni gezegenin tahmin edilen konumunu içeren gökyüzü alanını kapsayan yeni yıldız haritaları seti satın almış ve yeni gezegeni aramaya başlamıştır ve onu sadece birkaç gün sonra 23 Eylül 1846'da bulmuştur [4].

Karanlık maddenin varlığını kütle çekim yasaları veya diğer yöntemlerle hesaplamak mümkündür ancak karanlık maddenin varlığı ortaya çıkmış olsa da onun ne olduğunu anlamak için çalışmalar devam etmektedir. İlk olarak karanlık maddenin uzaydaki dağılımını yani galaksiler, galaksi kümelerindeki dağılımı anlaşılmalıdır. Karanlık maddenin yoğunluğu her galakside değişir mi yoksa eşit bir şekilde mi dağılmıştır? Araştırmalar karanlık maddenin yoğunluğunun farklı yerlerde farklı olduğunu göstermiştir. Örneğin Güneş sistemimizde karanlık madde yoğunluğu, galaktik merkez gibi kütle çekimin fazla olduğu bölgeden daha fazla olabilir. Karanlık maddenin dağılımında sadece yoğunluğu değil yoğunluk profilini de göz önünde bulundurmak gereklidir [3]. Yapılan araştırmalar ve çalışmalar sonucunda, karanlık maddenin özelliklerini aşağıdaki gibi özetlenebilir:

- Karanlık madde ışıksız bir nesnedir. Fotonlarla etkileşimi yoktur ve herhangi bir elektromanyetik radyasyon yayamaz.
- Karanlık madde, herhangi bir elektromanyetik etkileşime girmediğinden, yüksüz nötr parçacıklardan oluşmalıdır.
- Karanlık madde evrende her yeri kaplamaktadır ve kütle çekimi galaksi kümeleri gibi büyük ölçekli yapıların oluşmasına yardımcı olur.
- Karanlık madde parçacığı karardır; aksi takdirde, bilinen temel parçacıklara parçalanabilir ve laboratuvar deneylerinde tespit edilirdi.
- Karanlık maddenin diğer standart model parçacıklarla etkileşimi çok zayıf olmalıdır.

Evrendeki gezegenlerin, yıldızların veya galaksilerin keşfi için elektromanyetik spektrum ışınımının yanında kütle çekim yasaları veya modern fizikte kullanılan Genel Görecelik teorisini ya da farklı yöntemler kullanmak gerekmektedir. Karanlık madde elektromanyetik spektrumda bir ışınım yaymadığı için etkileşime girmez, tam anlamıyla görünmez bir maddedir ama bu onun var olmadığı anlamına gelmez çünkü

hesaplarımız karanlık maddenin varlığını ortaya çıkarmaktadır. Bu çalışmada galaksiler incelenmiştir.

1.2 Karanlık Maddenin Kanıtı ve Galaksilerin Kütle Hesapları

Bir galaksinin kütlesi birçok yöntem kullanılarak hesaplanabilir. Galaksinin yayımladığı parlaklık, elektromanyetik ışınım, Viryal yönetimi ya da kütle çekim yöntemleri gibi yöntemler kullanılır. Bir galaksinin kütlesi bilindiğinde galaksinin oluşturduğu çekim etkisini hesaplamak mümkündür, yani kütlelerinden, görünür madde ve karanlık madde oranları hesaplanabilir. Galaksilerin kütlelerini hesaplamak için kullanılan Viryal yöntemi, denge varsayımı altında galaksi hız dağılımını gösterir, X-ışınları, hidrostatik denge varsayımı altında sıcak gaz dispersiyonunu ve kütle çekim kırılmasında da arka plan galaksilerin görüntülerindeki kütle çekim bozulmalarını gösterir [5].

1.2.1 Viryal Teoremi Yöntemi

Bu teoremle hız dağılımını kullanarak bir galaksinin kütlelerini hesaplamak amaçlanır. Viryal teoreminin kullanılması için galaksinin kararlı halde olması gerekmektedir yani galaksi çökme ya da genişleme durumunda olmamalıdır. Galaksilerin hızları Doppler yöntemi kullanarak ölçülebilir ve bu yöntemle kümenin toplam kütlesi hesaplanabilir. Galaksiler oluşurken kararsız durumdadırlar, içerisindeki çarpışmalar ve etkileşimler (bu etkileşimler karanlık madde ve gaz bulutları arasındaki etkileşimlerdir) sistem kararlı hale gelene kadar devam eder. İçindeki potansiyel ve kinetik enerjinin dağılımı sürekli değişir. Süper simetri görünümünde olan galaksi kararlıdır ve bu yöntemin kullanılması uygundur. Viryal teoremi, bir sistemde toplam kütle çekim potansiyel enerjisinin, sistemin toplam kinetik enerjisinin iki katına eşit olduğunu ifade eder.

$$E_k = \frac{E_p}{2} \quad (1)$$

$$\frac{m_g \sigma^2}{2} = \frac{G m_g M_{cl}}{2 R_{cl}} \quad (2)$$

$$M_{cl} = \frac{\sigma^2 R_{cl}}{G} \quad (3)$$

Burada R_{cl} galaksinin çapı, σ galaksi hızlarının dağılımı, G evrensel çekim sabiti ve M_{cl} galaksinin kütlesidir. Bu yöntem galaksi kümelerine de uygulanabilir. Genel olarak bir galaksi kümesindeki yıldızların hız dağılımı $\sigma \sim 500-1500$ km/s ve galaksi kümesinin çapı $R_{cl} \sim 3-5$ Mpc olduğu hesaplanmıştır. Bu parametrelerden genel bir galaksi kümesinin kütlesi $M_{cl} \sim 10^{14}-10^{15} M_{\odot}$ güneş kütlesi kadardır [6].

Hız dağılımı hesaplanırken galaksi kümesinin önündeki veya arkasındaki galaksilerde göz önünde bulundurulup hesaplamalardan çıkartılmalıdır. Örnek olarak şekil-3'deki Virgo galaksi kümesinin kütlesi hesaplandığında, Virgo galaksi kümesindeki galaksi hızlarının dağılımı Dopler yöntemi kullanılarak (optik ve gözlem) yaklaşık olarak $\sigma = 550$ km/s olduğu hesaplanmıştır ve kümenin çapı 3 Mpc yaklaşık 9.257×10^{19} km dir. Kümenin kütlesini güneşin kütlesi cinsinden hesaplanacak olursa, ($G = 6.67408 \times 10^{-20}$ km³ kg⁻¹ s⁻²) aşağıdaki oran bulunur:

$$\frac{M_{cl}}{M_{\odot}} = \frac{550^2 \times 9.257 \times 10^{19}}{6.67408 \times 10^{-20} \times 1.989 \times 10^{30}} = 2.11 \times 10^{14} \quad (4)$$

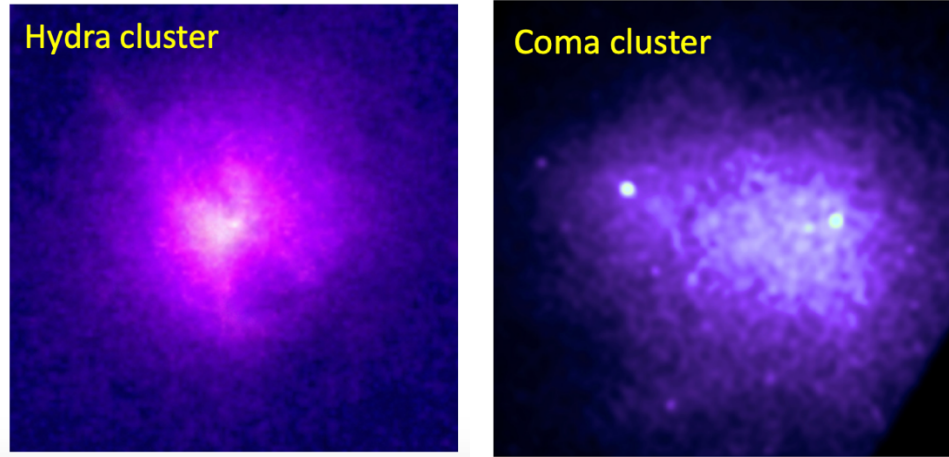
Optik ışık kullanılarak yapılan gözlem sonucu Virgo galaksi kümesi ile ilgili aşağıdaki eşitlik çıkarılmıştır. Kümenin parlaklığı güneşin parlaklığının 10^{12} katıdır,

$$\frac{L_{cl}}{L_{\odot}} \approx 10^{12} \frac{M_{cl}}{M_{\odot}} \quad (5)$$

Eşitlik 4 ve 5'e bakacak olursak, Viryal yöntemi kullanarak hesaplanan galaksi kümesinin kütlesi parlaklıklara bakarak yapılan kütle hesabından yaklaşık 211 kat daha fazladır. Bu da galaksi kümesinde çok fazla karanlık maddenin varlığını gösterir.

Şekil-3-a panelinde Virgo galaksi kümesinin haritası, şekil-3-b de galaksi kümesindeki X-ışınımı görüntüsü görülmektedir. Uzaklığı yaklaşık dünyadan 16 Mpc olan galaksi kümesinin çapı 3 Mpc dir ve yaklaşık 2000 galaksiden oluşmuştur [7].

çok ısındığını düşünelim, öyle bir sıcaklık olsun ki tüm atmosferdeki gazlar genişleyerek hızlı bir şekilde uzaya dağılır ve Güneş sistemini çok kolay bir şekilde terk eder çünkü Güneşin kütle çekimi dağılan gazları tutmak için yeterli değildir. Virgo galaksi kümesinde veya diğer bütün galaksi kümelerinde, sıcak gaz genişlemiyor gibi görünür. Aslında, kümenin merkezine yakın bir yerde küresel bir yığın halinde toplanmış gibi görünür. Şekil-4'de Hydra ve Coma galaksi kümelerinin X-ışını yoğunlukları görülmektedir. Gözlemlenen büyüklükte bir yığın halinde gazı bir arada tutmak için gereken kuvvet hesaplanırsa, yine gerekli kütle çekimi için kütlelerin çok büyük olması gerektiği görülür. Galaksi kümesinin kütlesi için kümedeki gazının radyasyonunun kütlesini hesaplandığında $M_{(gaz)} \approx 3 \times 10^{14} M_{\odot}$ bulunur. X-ışını yöntemi kullanmanın birçok avantajı vardır. Hangi galaksilerin gerçekten kümeye ait olduğunu belirlememize gerek yoktur çünkü hesaplamayı etkilemez ancak halen birleşme veya oluşma sürecinde olan bir kümeyi incelerken hidrostatik dengenin yine doğru bir varsayım olmadığı varsayılırken, kümelerden gerekli olan verileri X-ışını görüntüleme yoluyla elde etmek zahmetlidir.



Şekil-4 X-ray ışınımı Hydra ve Coma galaksi kümelerinin x-ışını görüntüleri [6]

1.2.3 Kütleçekimsel Mercek Etkisi Yöntemi

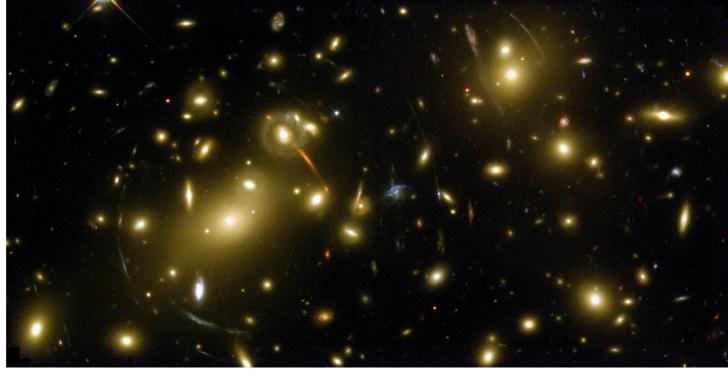
Einstein'ın Genel Görelilik teorisi, kütle çekimi ile uzay-zamanın eğriliğini ve uzay-zamanın geometrisini açıklar ve Einstein'ın teorisi kozmolojik hesaplamalar için daha uygundur. Nötron yıldızları, kara delikler gibi büyük nesnelere varlığında Genel Görelilik teorisi, uzay-zaman eğriliğinin oldukça büyük oranda büküldüğünü göstermektedir. Bununla beraber Newton'un kütle çekim kanunları, Genel Görelilik

etkisinin ihmal edilebilir olduđu sıradan bir yıldız sistemi için iyi sonuç vermektedir. Genel bir yaklaşımla Einstein'ın Genel Görecelik teorisi veya Newton'un çekim kanununun hangisinin etkili bir sonuç vereceğini uygulanan sistem için hesaplanan boyutsuz ε değerine bakmamız gerekmektedir.

$$\varepsilon = \frac{GM}{c^2 R} \quad (6)$$

Burada G evrensel çekim sabiti, M sistemin kütlesi, c ışık hızı ve R cismin çapıdır. Bu formül Güneş için uygulanırsa $\varepsilon \approx 10^{-6}$ çıkar ve Newton kütle çekim kanunları uygulanabilir. Ama $\varepsilon \approx 1$ sonucuna ulaşılan sistemlerde Einstein'ın Genel Görecelik teorisi geçerlidir [3].

Çekimsel mercecek etkisi, küme kütle dağılımını doğrudan ölçmek için çok güçlü bir yöntemdir. Işık bir kümenin yakınından geçtiğinde, kümenin kütle çekimi, ışığın bükülmesine neden olur ve kümenin arkasında ve görüş alanımızın dışında kalan galaksilerin şekli hafifçe görünür. Küme etrafındaki farklı alanlarda arka plandaki galaksilerin ortalama şekil değişimini ölçerek, yüzey kütle profilini tahmin edebiliriz. İki farklı tür kütle çekimi merceği etkisi vardır; zayıf ve güçlü mercekleme. Güçlü mercekleme, çoğunlukla kümenin merkezi alanı etrafındaki gözlemler için kullanılır. Daha spesifik olarak, Einstein Yarıçapı* içinde bulunan kümenin kütlesini hesaplamak için kullanılan bir yöntemi ifade eder. Işık bozulması, kümenin çekirdeği etrafında gözlenen Şekil-5'deki gibi güçlü yaylar şeklinde görünür. Zayıf mercekleme ise çoğunlukla kümenin dış bölgelerinde gözlenen kütleyi tahmin etmek için kullanılır. Teknik olarak, aynı zamanda güçlü mercekleme ile aynı varsayımlara dayanmaktadır. Zayıf ve güçlü merceklemenin en önemli yönü, görünür (veya yakın kızılötesi) banttaki verilere dayanabilmeleridir [8].

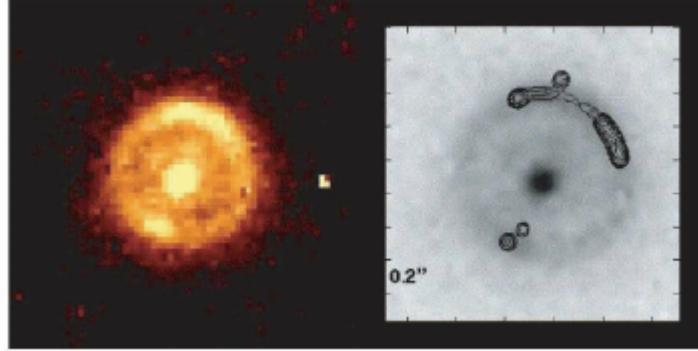


Şekil-5 Galaksi kümesindeki kütleçekimsel mercek etkileri [8]

Kütleçekimsel mercek etkisi yöntemiyle hesaplanan galaksi kümeleri kütlelerine Tablo-1'deki örnekler verilebilir [6]. M/L oranı karanlık madde miktarının yoğunluğunu göstermektedir. Viryal kütle tahminleriyle oldukça büyük bir uyum içinde, kümelerde çok sayıda karanlık madde olduğu görülmektedir. Tablo-1: Kütleçekimsel mercek etkisi yöntemiyle hesaplanan galaksi kümeleri kütleleri.

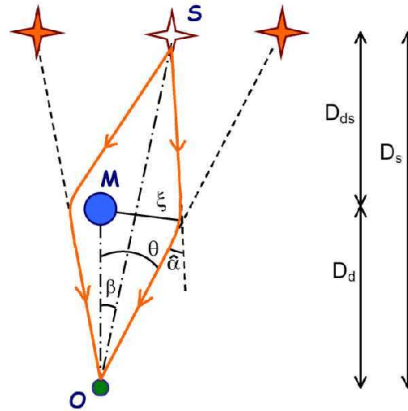
Tablo-1 Güçlü ve zayıf merceklemeye örnekleri [6]

Güçlü merceklemeye örnekleri		
Galaksinin Adı	Kütle (Güneş Kütleleri cinsinden)	M/L Güneş=0.7 (Güneş [9])
A370 M	$5 \times 10^{13} h^{-1}$	270h
A2390 M	$8 \times 10^{13} h^{-1}$	240h
MS2137 M	$3 \times 10^{13} h^{-1}$	500h
A2218 M	$1.4 \times 10^{14} h^{-1}$	360h
Zayıf merceklemeye örnekleri		
MS1224		800h
A1689		400h
CL1455		520h
A2218		310h
CL0016		180h
A851		200h
A2163		300h
Güneş = 0.7		



Şekil-6 B 1938+666 kütleçekimsel mercek sistemi [9]

Tek bir galaksiden kaynaklanan kütleçekimsel mercek etkisi sonucunda Şekil-6'da görüldüğü gibi, ışık kaynağı merceği saran bir halka şeklini alır. Bu halkaya yukarıda da bahsettiğimiz gibi Einstein halkası (Einstein ring) denir. Şekil-6'da halkanın merkezine yakın bir yerde bulunan mercek gökadasıyla birlikte aktif gökadanın haritalandığı tam bir Einstein halkası açıkça görülmektedir. Şekil-6'de gösterilen B 1938+666 gibi bu halkaların birçok örneği keşfedilmiştir. Şekildeki kütle dağılımına Şekil-7'deki gibi bir nokta kütle olarak yaklaşım yapılırsa, uzaktaki cismin görüntüsü θ_E açısal yarıçapına sahip tam bir halka olarak görünür ve bu açı aşağıdaki formül ile hesaplanabilir.



Şekil-7 $\theta_E, D_d, D_s, D_{ds}, M$ parametrelerinin açıklaması [10]

$$\theta_E = \sqrt{\frac{4GM D_{ds}}{c^2 D_d D_s}} \quad (7)$$

Formülde; D_s Kaynak ile gözlemci arasındaki uzaklık, D_L kütle çekim merceğinin gözlemciden olan uzaklığı ve D_{LS} kütle çekim merceğinin kaynağına olan uzaklıktır. Denklemden mercek etkisi gösteren galaksi kümesinin kütlesi M hesaplanabilir. Bu formül tam bir halka oluşturmayan benzer ve az simetrik durumlarda da uygulanabilir [10].

1.3 Makine Öğrenmesi ile Galaksi Halo Kütlelerinin Tayini

Makine öğrenimi (ML) tekniklerini astronomiye uygulayan çalışmaların sayısında son yıllarda önemli bir artış olmuştur. ML'nin astronomideki en önemli uygulamalarından biri, çeşitli nesnelerin sınıflandırılmasıdır, örneğin geçiş olayları ve galaksi morfolojisi bu uygulamalardan birkaç tanesidir [11,12]. Diğer uygulamalar arasında, geniş bant filtresi ile galaksilerin fotometrik kırmızıya kaymalarının belirlenmesi, N-cisim simülasyonlarından sentetik kataloglar oluşturulması için karanlık madde halolarının atanması ve Samanyolu'nun yapısının incelenmesi gibi çalışmalar vardır [13,14,15,16,17]. Ayrıca ML, galaksi kümelerinin dönüş hızlarının bilgilerini kullanarak galaksi kümelerinin dinamik kütle ölçümlerini iyileştirmek için de kullanılmıştır [18]. Daha yakın zamanlarda, dinamik ve X-ışını verilerinin bir kombinasyonunu kullanarak galaksi kümelerinin kütle ölçümü için ML algoritmaları kullanılmıştır [19]. Kozmolojik simülasyonlardan alınan sentetik X-ışını görüntüleri kullanılarak galaksi kümelerinin kütlelerini tahmin etmek için daha karmaşık algoritmalar uygulanmıştır [20]. Galaksi halolarının tahmin edilmesi konusunda yapılan önemli bir çalışma da, Calderon ve Berlind'in 2019'daki makine öğrenmesi yaklaşımı ile SDSS DR7 galaksi kataloğundaki galaksilerin halo kütleleri için tahminlerde bulunduğu çalışmadır [21].

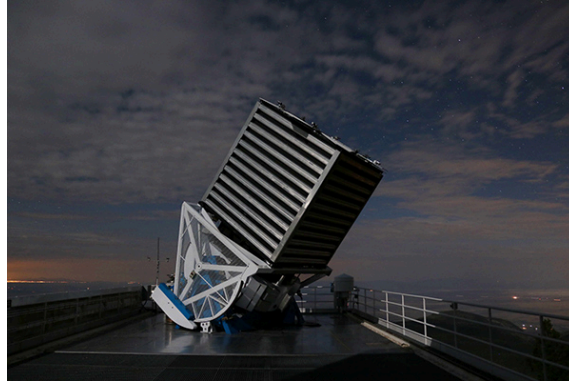
2. VERİ VE YÖNTEM

Bu çalışmada, geniş bir kütle yelpazesindeki galaksilerin halo kütlelerini tahmin etmek için makine öğrenimi tekniğini kullanılmıştır ve SVR ML algoritması incelenmiştir. Hem galaksilerin hem de gruplarının gözlemlenen özelliklerini

kullanarak ML algoritmaları sentetik veriler üzerinde eğitilmiştir. Sonuçlar, bir galaksinin karanlık madde halosunun gerçek kütesini çıkarmak için makine öğrenimi algoritmalarını kullanmanın güçlü bir yöntem olduğunu göstermektedir. Sentetik kataloglarla yaptıkları testler, SDSS DR16 galaksileri için ML ile tahmin edilen halo kütlelerinin hesaplanan değerlere yakın olabileceğini göstermektedir [21].

2.1 SDSS Galaksi Kataloğu

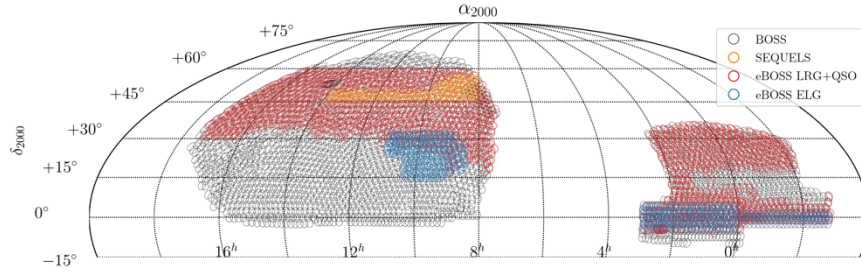
Sloan Dijital Gökyüzü Taraması (SDSS) evrenin yapısını ve nasıl oluştuğunu anlamak için sürekli veri toplayan bir projedir. Gökyüzünü tarayan özel bir teleskop sayesinde görüntüler kaydedilip arşivlenmektedir. SDSS, gökyüzünün renkli derin uzay görüntüleri ve üç milyondan fazla astronomik nesnenin spektrumu ile evrenin şimdiye kadar yapılmış en ayrıntılı üç boyutlu haritalarını oluşturmuştur. Bu tez çalışmasında "16. veri sürümü" (DR16) verileri kullanılmıştır. DR16, Sloan Dijital Gökyüzü Taramasının (SDSS-IV) dördüncü aşamasının dördüncü veri sürümü olup ve Ağustos 2018'e kadar olan SDSS gözlemlerini içermektedir [22].



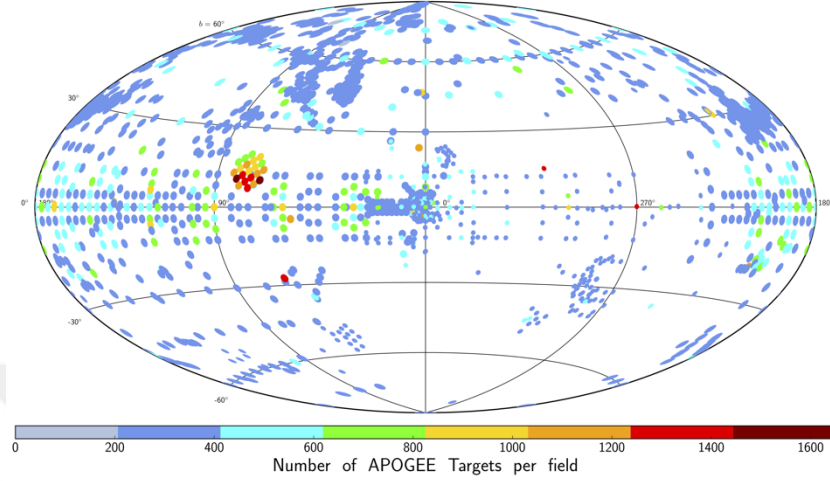
Şekil-8 SDSS Sloan teleskobu, New Mexico [22]

Sloan Dijital Gökyüzü Taraması SDSS adını projeye büyük ölçüde bağış yapan Alfred P. Sloan Kuruluşu'ndan almıştır. New Mexico ABD'de bulunan Apache Point Gözlemevi'ndeki 2.5 metre ve 1 metre çaplı optik teleskoplar ile Şili'de bulunan Las Campanas Gözlemevi'ndeki The Irénée du Pont teleskobu kullanılmaktadır. 2000 yılında veri toplama süreci başlamıştır. Ocak 2011'de yayınlanan 8.veri sürümü (DR8) 14.555 derece karelik bir alanı (tam gökyüzünün %35'ini biraz aşkın) kapsayan

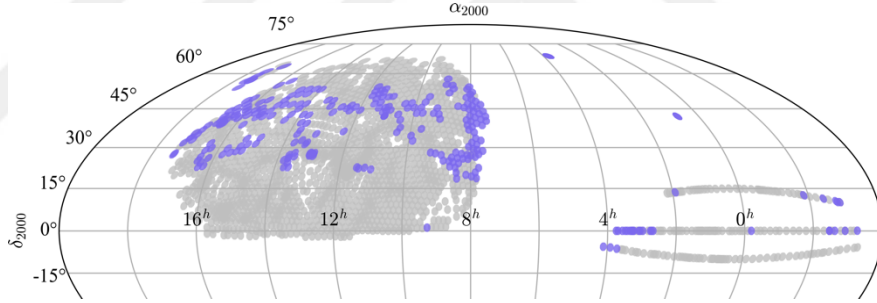
gözlemleri içermektedir. 31 Temmuz 2012 tarihinde yayınlanan 9.veri sürümü Baryon Salınım Spektroskopik Taraması'ndan (BOSS) gelen 800.000'den fazla yeni spektrum içeren ilk sonuçları kapsamaktadır. Bunların içinde 500.000'den fazla spektrum evrende 7 milyar yıl önce var olan nesnelere aittir. 31 Temmuz 2013 tarihinde halka açıklanan 10.veri sürümü (DR10) önceki sürümlerde yayınlanan tüm verileri ve Samanyolu yıldızlarının Apache Point Gözlemevi Galaktik Evrim Deneyi (APOGEE) spektrografı ile alınan 57.000'den fazla, yüksek çözünürlüklü (kızılötesi) spektrumunu içeren verileri kapsamaktadır. Ayrıca DR10 verileri uzak evrendeki galaksilerin ve kuasarların 670.000'den fazla yeni BOSS spektrumunu içermektedir [23]. DR16, Sloan Digital Sky Survey IV'ün dördüncü sürümüdür. DR16, Şili'deki Las Campanas Gözlemevi tarafından gözlemlenen ilk kızılötesi spektrumlar dahil olmak üzere yeni optik ve kızılötesi spektrumları içerir. Daha önce piyasaya sürülen haritalar, yıldız kütüphanesi spektrumları ve görüntüler de DR16'ya dahildir. DR16 Sloan Digital Sky Survey, Ağustos 2018'e kadar alınan SDSS verilerini içerir ve tüm göksel kürenin üçte birinden fazlasını kapsar. Şekil-9, Şekil-10 ve Şekil-11 eBOSS, APOGEE ve MaNGA araştırmaları için DR16'daki gökyüzü kapsama alanları gösterilmiştir [22].



Şekil-9 Ekvator koordinatlarında DR16 eBOSS spektroskopik kapsama alanı [22]



Şekil-10 DR16 APOGEE alanlarının dağılımı [22].



Şekil-11 DR15 / DR16 MaNGA Ekvatordaki spektroskopik kapsama alanı [22]

2.1.1 Veri Ulaşımı

SDSS tüm veri sürümleri halka açık olup bütün verilere internet üzerinden ulaşım mümkündür. SDSS verilerine Sky Server sunucusu üzerinden ulaşılmaktadır¹. SkyServer, SQL (Structured Query Language) dili kullanan bir arayüze sahiptir. Sadece koordinat parametresi verilerek gökyüzünün bir bölgesinin SDSS ile alınmış renkli görüntülerinin bu sunucudan elde edilebilmesi mümkündür. SkyServer SQL sorgu aracı ile spektrumlar ve görüntüler ulaşılabilir durumdadır ve arayüzler kolay kullanılabilir haldedir. Veriler halka açık olup yazılı izin olmaksızın sadece ticari olmayan amaçlar için kullanılabilir durumdadır. Ayrıca Google Gökyüzü²

¹ <http://skyserver.sdss.org/>

² <https://Güneş.google.com/sky/>

uygulamasında SkyServer sunucusuna doğrudan erişim sağlayan SDSS fotometri ve spektroskopi katmanları için eklentiler de vardır. Hayden Planetariumu Dijital Evren uygulamasında¹ da evrenin 3 boyutlu atlası oluşturularak SDSS verileri görüntülenebilmektedir.

Bu tez çalışmasında SDSS CasJobs arayüzü kullanılmıştır ve DR16 kataloğunda bulunan galaksiler ile ilgili tüm parametreler SQL dizisi ile veri alanından çekilmiş ve indirilmiştir. Örneğin bu çalışmada, DR16 galaksi tablosundan veri toplanmıştır. Şekil-12’de görüldüğü gibi yazılan SQL dizilimi sayesinde “PhotoPrimary” tablosundan “type=3” (Tablo-2’de “type” parametresinin anlamları verilmiştir) ve “PhotoPrimary” tablosundan “PhotoObjAll” “mode=1” (Tablo-3’de “mode” parametresinin anlamları verilmiştir) tercihleri yapılarak SQL dizisi yazılmış ve çalışmada kullanılacak veri setinin oluşturulması için sunucuya bilgiler gönderilmiştir.

Tablo-2 SQL dizisinde “type” parametresinin anlamı

#type	Değeri	Açıklama
Unknown	0	Nesnenin ne olduğu bilinmiyor
Cosmic_Ray	1	Kozmik ışın izi (kullanılmıyor).
Defect	2	Nesne, teleskop veya işleme hattındaki bir kusurdan kaynaklanıyor.
Galaxy	3	Galaksi: Birçok yıldızdan ve diğer maddelerden oluşan genişletilmiş bir nesne.
Ghost	4	Hayalet: Yansıyan veya kırılan ışıkla oluşturulan nesne.
Knownobj	5	Bilinen Nesne: Nesne SDSS kataloğu dışından bir katalogdan geliyor
Star	6	Yıldız: Kendinden ışık yayan bir gök cismi.
Trail	7	İz: Bir uydu veya yıldız veya meteor izi. (Henüz kullanılmadı)
Sky	8	Gökyüzü: Boş gökyüzü spektrumu
Notype	9	Hiçbir tipte değil

¹ <https://Güneş.amnh.org/research/hayden-planetarium/digital-universe>

Tablo-3 SQL dizisinde “mode” parametresinin anlamı

#mode	Değeri	Açıklama
Primary	1	Birincil Nesne
Secondary	2	İkincil Nesne
Family	3	Ailesel Nesne
Outside	4	Yığın sınırının dışındaki nesne
Hole	5	Delik içindeki nesne (Field.quality = 5)

SciServer sunucu sayfasından¹ SQL dizisi aşağıdaki gibi yazılmış ve seçilen DR16 kataloğundan SQL dizisi çalıştırılarak galaksi veri tabanına ulaşılmıştır.



Şekil-12 DR16 veri tabanından galaksi verilerinin indirilmesi¹

DR16 veri tabanından indirilen galaksilerin özellikleri aşağıdaki gibidir:

Tablo-4 SDSS DR16 veri tabanından kullanılan galaksilerin özellikleri²

specObjID	ID
ra	Sağ açıklık (derece)
dec	Dik açıklık (derece)
z	Kırmızıya kayma

¹ <http://skyserver.sdss.org/CasJobs/SubmitJob.aspx>

² <http://skyserver.sdss.org/dr16/en/help/browser/browser.aspx#&&history=description+stellarMassFSPS+GranWideNoDust+U>

logMass	Galaksi veya kümnin logaritmik kütlesi ($\text{dex } M_{\text{Güneş}}$)
metallicity	Metallilik
dust1	Genç yıldızların etrafındaki toz miktarının azalması
dust2	Yaşlı yıldızların etrafındaki toz miktarının azalması
tau	Yıldız oluşum tarihi (Gyr)
sfr	Yıldız oluşum hızı ($\log \text{Gyr}^{-1}$)

Sağ açıklık (ra) ve dik açıklık (dec) gökyüzünde bir cismin ekvatoryal koordinat sisteminde (eşlek konum düzeneği) gök küresi üzerindeki konumunu veren parametrelerdir [24].

Z Kırmızıya kayma: Bir galaksinin kırmızıya kayması (veya maviye kayması), hem Doppler kayması ile ilgili bir bileşeni hem de evrenin genişlemesinden ötürü oluşan kozmolojik kırmızıya kaymayı içerir ve galaksinin spektrumundaki çizgilerin daha uzun (veya daha kısa) dalga boylarına kayması olarak gözlenir. Yakınıımızdaki evren dışında, kırmızıya kaymalar veya görünen radyal hızlar kozmolojik genişleme tarafından belirlenir [25].

logMass: Galaksi veya galaksi kümesinin hesaplanan logaritmik kütlesi. Bir önceki konuda galaksi kümeleri veya galaksilerin kütlelerinin nasıl hesaplandığından bahsedilmiştir. Galaksilerin gözlenen spektrofotometrik verileri (ugriz parlaklıları) spektroskopik kırmızıya kayma ve spektroskopik enerji dağılımlarına en iyi uyan yıldız popülasyonu modelleri kullanılarak hesaplanan (görünür) galaksi kütlesi değeridir.¹

metallicity: Metallilik, bir gökcisminin ihtiva ettiği hidrojen veya helyumdan daha ağır olan tüm elementlerin bolluğudur. Evrendeki normal fiziksel maddenin çoğu hidrojen ve helyumdur ve "metaller" kelimesini "hidrojen ve helyum dışındaki tüm elementler" için uygun kısa bir terim olarak kullanılır. Bu kullanım, katı metalin olağan fiziksel tanımından farklıdır. Nispeten yüksek miktarda karbon, nitrojen, oksijen ve neon içeren yıldızlar ve bulutsular, kimyada bu elementler metal olmamasına rağmen astrofiziksel terimlerle "metal açısından zengin" olarak adlandırılır. Güneşimizin metal bolluğu veya metallilik özelliği $Z_{\text{Güneş}}=0.019$ dur. [26]

¹ https://Güneş.sdss.org/dr16/spectro/galaxy_portsmouth/

SDSS DR16 galaksi veri tabanında metallilik parametresi Güneş metalliliği (solar metalicity) biriminde verilmiştir. Ancak MICECAT kataloğunda galaksilerin ya da galaksi kümelerinin metallilik birimi $\text{Log} \left(\frac{O}{H} \right) + 12$ biriminde verilmektedir. Dolayısı ile oluşturulan veri setinde SDSS kataloğundaki galaksi verilerindeki metallilik birimleri aşağıdaki formül kullanılarak çevrilmiştir.

$$\text{Log} \left(\frac{O}{H} \right) + 12 = \text{Log} \left(\text{Metallik} * \frac{0.0127}{0.74*16} \right) + 12 \quad (8)$$

dust1: Galaksilerdeki toz zayıflamasını ölçmek için temel teknikler vardır. Bu parametre genç yıldızların etrafındaki toz zayıflamasının değerini gösterir¹

dust2: Aynı şekilde bu parametrede yaşlı yıldızların etrafındaki toz zayıflamasının değeridir.

Tau: Yıldız oluşum tarihi (star formation history), galaksilerin evrimini tanımlamak ve anlamak için kullanılan temel bileşenlerden biridir. Yıldız oluşum tarihi, yıldızların kısa veya uzun dönemlerde zaman ve uzayda nasıl oluştuğudur. Tau exponansiyel olarak azalan bir yıldız oluşum tarihini göstermektedir [27].

Sfr: Yıldız oluşum oranı (star formation rate) Yıldız oluşum oranı, Güneş kütlesi olarak bir yılda oluşan toplam yıldız kütlesidir.

2.2 MICECAT Galaksi Kataloğu

MICECAT kataloğu Marenstrum Institut de Ciències de l'Espai (MICE) simülasyon verileri kullanılarak oluşturulan bir sentetik galaksi kataloğudur [28]. MICE projesi Barselona'daki Marenstrum süper bilgisayarını kullanarak çok büyük bir nümerik kozmolojik simülasyon geliştirme projesidir [29]. 1-500 Giga parsek küplük hacimler içinde 109-1011 tane karanlık madde parçacığını hesaba katarak N-cisim simülasyonları yapılır. Sonucunda tüm gökyüzü merceklemeye haritaları ve galaksi halo katalogları oluşturulur.

¹<http://skyserver.sdss.org/dr16/en/help/browser/browser.aspx#&&history=description+stellarMassFSPSGranWideNoDust+U>

Bu şekilde oluşturulan kataloglara mock katalogları denir ve gözlemler yerine modellere dayanan sentetik veriler içerir. Bu projede sunulan güncel veri tabanı MICECAT v2 yaklaşık 200 milyon galaksi ve kırmızıya kayma $z = 1,4$ 'e kadar kapsamlı galaksi verisi içerir. Bu son sentetik galaksi veri tabanı, ana MICE-Grand Challenge (MICE-GC) ışık konisi N-cisim simülasyonundan ve ilgili halo kataloğundan oluşturulmuştur. Kütle çözünürlük etkisi 5-10% $k > 1$ h/Mpc olarak tanımlanmaktadır.

Bu tez çalışmasında 20000 adet galaksi için sentetik veri kullanılmıştır. ML makine öğrenimi programında kullanmak için MICECAT v2 veri tablolarından seçilen parametreler aşağıdaki gibidir. MICECAT v2 kataloğunda 2'nin anlamı 2. versiyon olmasıdır.

Tablo-5 MICECAT v2 veri tabanından çekilen galaksi özellikleri¹

ra_gal	Sağ açıklık (derece)
dec_gal	Dik açıklık (derece)
z_cgal	Galaksinin kırmızıya kayması
cgal	Galaksinin koordinat mesafesi (Mpc) (comoving distance)
z_cgal_v	Gözlenen kırmızıya kayma (özgün hızı ile birlikte)
cgal_v	z_cgal_v karşı gelen koordinat mesafesi (Mpc)
flag_central	0= Merkezil, 1=uydu Galaksi
lmhalo	Karanlık madde halosunun (logaritmik) kütlesi ($\text{dex}M_{\text{Güneş}}$)
v_vir	Karanlık madde halosunun viryal hızı (km/viryal)
nsats	Uydu galaksi sayısı
xhalo	Karanlık madde halosunun uzaklığı x bileşeni (Mpc)
yhalo	Karanlık madde halosunun uzaklığı Güneş bileşeni (Mpc)
zhalo	Karanlık madde halosunun uzaklığı z bileşeni (Mpc)
vxhalo	Karanlık madde halosunun koordinat mesafesi x bileşeni (Mpc)
vyhalo	Karanlık madde halosunun koordinat mesafesi Güneş bileşeni (Mpc)

¹ <https://cosmohub.pic.es/catalogs/1>

vzhalo	Karanlık madde halosunun koordinat mesafesi z bileşeni (Mpc)
unique_gal_id	Galaksi ID
lsfr	Yıldız oluşma oranı (logaritmik) $[(M_{\text{Güneş}}/\text{Güneş})/\text{yıl}]$
lmstellar	Galaksinin yıldız kütlesi (logaritmik) $(\text{dex}M_{\text{Güneş}})$
metallicity	Metallilik özelliği birimi $12+\log O/\text{GÜNEŞ}$
(Burada bahsi geçen parametrelerin İngilizce tanımları kullanılmış olup EK-1'de verilmiştir.)	

MICECAT v2 verilerine ulaşmak için COSMOHUB sunucusu¹ kullanılmıştır. Bu sunucu üzerinden MICECAT v2 kataloğundan veri almak için SQL dizini kullanılarak karanlık madde ile ilgili parametreler seçilmiştir. Seçilen parametreler Tablo-5'de tanımlanmıştır. COSMOHUB kullanıcılara galaksilerle ilgili çok önemli analiz sonuçları verebilmektedir. COSMOHUB sunucusunun en önemli özelliği, kullanıcıların herhangi bir SQL bilgisi olmadan, CSV, FITS veya ASDF gibi çeşitli formatlarda indirebileceği ve analiz edebileceği kendilerine özel kataloglarını oluşturabilmesidir. Kullanıcılar ayrıca önceden oluşturulmuş katalogları analiz etmek için gerekli olabilecek diğer dosyaları da indirebilirler. Kataloglar, bunların yanı sıra 1 boyutlu histogramlar ve 2 boyutlu ısı haritası çizimlerini içeren entegre bir görselleştirme aracı ile kullanılmıştır. Görselleştirme aracı, Flagship simülasyon galaksi kataloğunu keşfetmek, doğrulamak ve dağıtmak için anahtar olmuştur. Son olarak, kullanıcılar daha da hızlı sonuçlar elde etmek için toplam kataloğun bir alt kümesi üzerinde çalışmak üzere örneklemeler de kullanılabilir. Bu tez çalışmasında veriler CSV formatında indirilip python programlama dilinde kullanılan "pandas", "numpy" ve öğrenme kütüphanesi olan scikit-learn kütüphaneleri kullanılmıştır.

Python, nesne yönelimli, modüler ve etkileşimli yüksek seviyeli bir programlama dilidir. Bu proje Python 3.8 versiyonu ile yazılmıştır ve Anaconda, Jupyter Notebook arayüzü ile geliştirilmiştir [30]. Python geliştiricilere kolaylık sağlamak için Anaconda isimli bir Enterprise piyasaya sürmüştür. Bu sürümde 100 hazır kütüphane

¹ <https://cosmohub.pic.es/>

bulundururken 700 kütüphaneyi de conda isimli paket yöneticisini kullanarak kurma olanağı sağlamıştır. Bu sayede veri bilimi ve diğer alanlarla ilgili kütüphaneler projelerde kullanılabilmiştir. Anaconda ile birlikte Jupyter Notebook kurulumu da yapılmıştır. Jupyter Notebook, çeşitli programlama dilleri için etkileşimli bir ortam sağlayan açık kaynak kodlu bir programdır.

Projede kullanılan pandas kütüphanesi, python programlama dilinde yazılmış veri işleme ve analizi için kullanılan bir yazılım kütüphanesidir. Pandas kütüphanesi 2008 yılında geliştirilmeye başlanmış ve 2009'un sonunda açık kaynaklı hale getirmiştir [31]. Bu kütüphane temel olarak sayısal tabloları işlemek için bir veri yapısı oluşturur. Diğer bir kütüphane numpy kütüphanesidir, NumPy, Python ile sayısal hesaplamayı etkinleştirmeyi amaçlayan açık kaynaklı bir kütüphanedir. 2005 yılında Numerical ve Numarray kütüphanelerinin ilk çalışmaları üzerine inşa edilmiş açık kaynaklı bir kütüphanedir [32]. Makine öğrenmesi için kullanılan kütüphane ise scikit-learn kütüphanesidir. 2007 yılında David Cournapeau tarafından bir Google Summer of Code projesi olarak başlatılmıştır. 2010 yılında, INRIA şirketi tarafından açık kaynaklı kod olarak sunulmuştur. Scikit-learn yapay öğrenme alanında en yaygın olarak kullanılan kütüphanelerden biridir. Doğrusal regresyon, lojistik regresyon, karar ağaçları, rastgele orman gibi birçok temel yöntemi içeren bu kütüphane, Anaconda ile beraber gelmektedir.

2.3 MAKİNE ÖĞRENİMİ VE KARANLIK MADDE KÜTLESİ TAHMİNİ

Makine öğrenimi, sadece astrofizik ve astronomide değil değil farklı disiplinlerde de kullanılmaktadır. Makine öğrenimi algoritmalarının en önemli özelliği analitik bir çözüm yapma zorunluluğu olmadan, bazı girdiler ile beklenen bir çıktı arasındaki parametrik olmayan ilişkileri öğrenebilmesidir ve daha sonra da öğrenme yeteneğini kullanarak farklı girdiler için sonucu çok küçük hatalarla tahmin edebilir olmasıdır. Bir makine öğreniminin türü olan "supervised learning" denetimli öğrenme algoritması için, bir eğitim veri kümesi sağlanır ve makine öğrenimi algoritmaları, özellikler kümesi arasındaki eşleşmeyi öğrenmeye çalışır. Algoritma eğitildikten sonra, ne kadar iyi çalıştığını ölçmek için farklı bir "test" veri kümesinde test edilir. Nihayetinde amaç, algoritmayı sonucu bilinmeyen bir veri kümesine uygulamaktır.

Bu çalışmada denetimli öğrenme algoritması olan SVR (Support Vector Regression) algoritması kullanılmıştır.

Bu çalışmada kullanılan programın amacı galaksi veya galaksi kümelerindeki karanlık madde halosunun kütlelerini tahmin etmektir. Bu amaçla, makine öğrenimi programı için MICECAT v2 kataloğundaki galaksiler veya galaksi kümelerinin verileri kullanılmıştır. MICECAT v2 kataloğunda galaksilerin karanlık madde halo kütleleri Bölüm 2’de anlatıldığı gibi simülasyonlar yani matematik hesaplamalar sonucunda oluşturulmuş verileri içerir. Bu veriler kullanılarak programın eğitilmesi sağlanmıştır.

2.3.1 Makine Öğrenimi Uygulaması

Bu çalışmaya adapte edilen makine öğrenimi (ML) algoritmasında, Python programlama dilinin bir kütüphanesi olan scikit-learn kütüphanesi kullanılmıştır. Bu projede kullanılan galaksilerdeki karanlık madde miktarı bir sınıflandırma problemi olduğundan dolayı denetimli öğrenme algoritması olan SVR algoritması kullanılmıştır. ML algoritmasını eğitmek için öncelikle hangi özelliklerin kullanılacağına karar verilmiştir. Özellikler (feature) seti, galaksinin özelliklerini ve galaksinin ait olduğu grubun özelliklerini içerir. Bu özellikler MICECAT v2 kataloğundan seçilmiş özelliklerdir.

MICECAT v2 kataloğu CSV formatında Python programlama dilinde kullanılan “pandas” kütüphanesi ile ayıklanmıştır. Bu işlem sonucunda 20000 farklı galaksinin koordinatları, kırmızıya kayması, karanlık madde halosunun viryal hızı, karanlık madde halosunun uzaklığı, yıldız oluşma oranı, yıldız kütleleri ve metalliliği gibi Tablo-5’de bahsi geçen özelliklerini içeren bir veri seti elde edilmiştir. Tablo-6’da bu elde edilen veri setinden bir örnek gösterilmiştir. MICECAT v2 kataloğundan alınan galaksi verileri 6 ile 12 $\text{dex}M_{\text{Güneş}}$ aralığındaki galaksi kütleleri ile sınırlıdır. Bu da veri setimizin sınırlılığını göstermektedir.

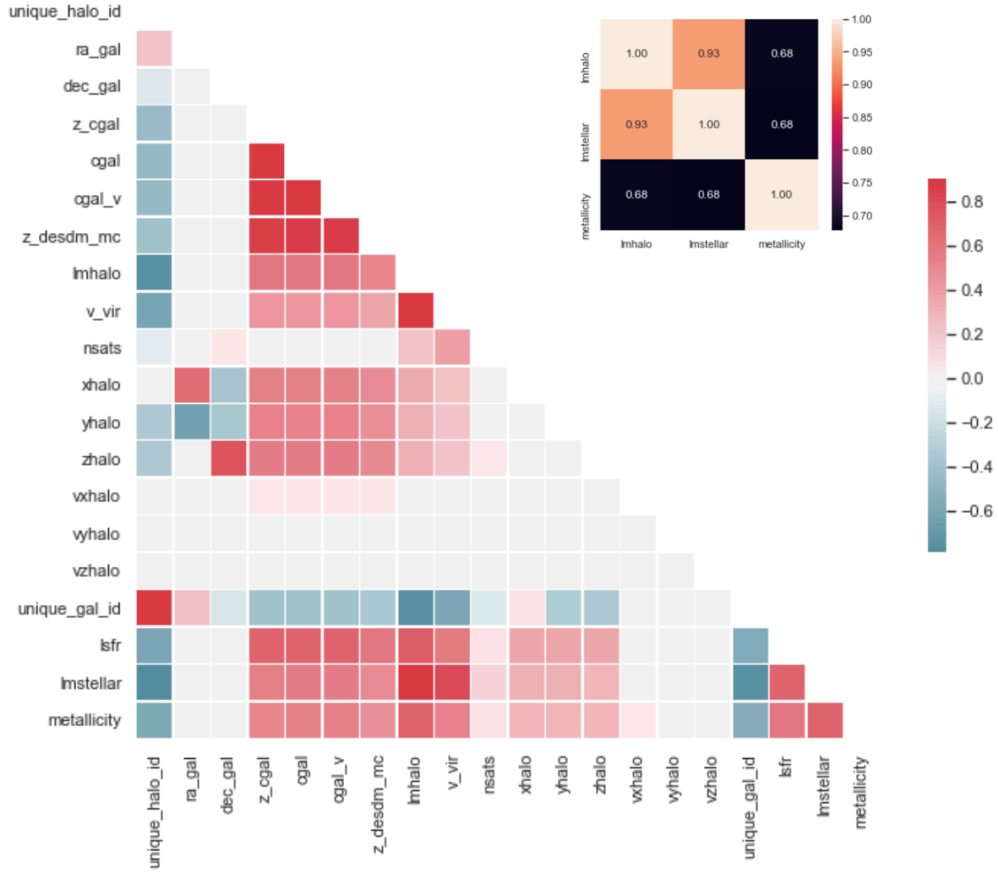
Tablo-6 MICECAT v2 veri tabanından ayıklanmış 20000 farklı galaksiye ait verilerden örnek

#	0	1	2	3	4
unique_halo_id	4999492	6742522	4753997	1010335	3538498
ra_gal (derece)	27.8703	2.35478	9.30961	24.4254	18.7483
dec_gal (derece)	70.2169	36.41	67.8885	73.7125	22.7034

z_cgal	0.62369	0.30398	0.56225	0.32461	0.19342
cgal (Mpc/Güneş)	1630	855.85	1491.17	909.71	557.83
cgal_v (Mpc/Güneş)	1631.35	859.74	1491.8	908.63	553.94
z_desdm_mc	0.51689	0.46083	0.49681	0.49235	0.53175
lmhalo (dexM _{Güneş})	12.1339	14.0009	12.4957	12.1191	11.921
v_vir (km/Güneş)	175.4	735.2	231.6	173.4	149
nsats	7	577	13	9	2
xhalo (Mpc/Güneş)	257.9	28.3	90.8	105.5	165.4
yhalo (Mpc/Güneş)	487.7	688.2	553.9	232.3	487.3
zhalo (Mpc/Güneş)	1533.8	508	1381.5	873.2	215.3
vxhalo (km/Güneş)	-135	-413	242	-278	-516
vyhalo (km/Güneş)	134	294	51	47	-40
vzhalo (km/Güneş)	99	198	20	-77	-428
lsfr [(M _{Güneş} /Güneş)/yıl]	0.791	1.792	0.461	0.154	0.639
lmstellar (dexM _{Güneş})	10.083	10.602	10.514	10.436	9.977
metallicity	8.972	8.963	9.128	8.996	8.898

Tablo-6'da bir örneği gösterilen ve 20000 farklı galaksi için elde edilen veriler bir sonraki aşamada korelasyon matrisine sokulmuştur. Korelasyon matrisi veriler arasındaki ilişkileri hesaplayan bir fonksiyondur ve yine Python'un "pandas" kütüphanesinde kullanılır. Korelasyon matrisinin sonucunu görsel olarak anlatmak ve özellikler arasındaki ilişkiyi görmek için yine Python kütüphanelerinden biri olan

“numpy” kütüphanesi kullanılarak özelliklerin sıcaklık haritası çıkartılmıştır.



Şekil-13 MICECAT v2 galaksi kataloğundan seçilen galaksi özelliklerinin birbirleriyle olan ilişkilerini gösteren sıcaklık haritası [Ek-1 Out(5)]

Şekil-13’deki korelasyon matrisi pembe renge kayan kareler yani sayısal değer olarak 0.2’ye yakın olan özelliklerin birbirleri arasında bir ilişki olabileceği ve tam tersine, sayısal değeri -0.6 değerine yani yeşilimsi karelere sahip özelliklerin birbirleri arasında bir ilişkinin olmadığı anlaşılmaktadır. Bu sıcaklık haritası sayısal değerlerine bakılarak incelendiğinde galaksinin karanlık madde halosunun kütlesi (lmhalo), galaksi kütlesi (lmstellar) ve metallilik (metallicity) özellikleri arasında yüksek bir ilişki olduğu görülmüştür. Veri setinden her bir galaksinin bu özellikleri Ek-1 Out(7) olarak ayıklanmıştır. Bu parametreler arasındaki korelasyon ayrıntılı incelendiğinde Tablo-7’deki sayısal ilişkiler görülmüştür.

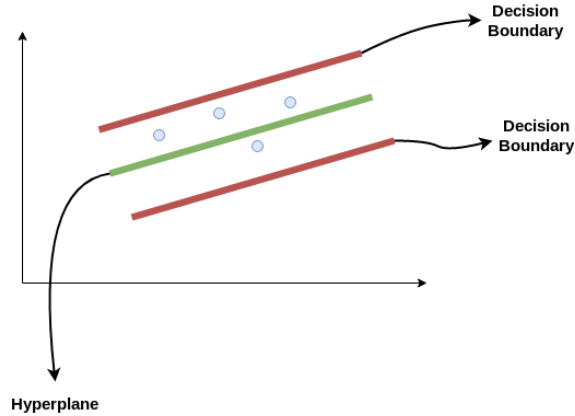
Tablo-7 Sıcaklık haritasına göre, aralarındaki ilişkinin en fazla olduğu özellikler [Ek-1 Out(8)]

	Halo Kütlesi	Galaksi Kütlesi	Metallilik
Halo Kütlesi	1	0.934626	0.682461
Galaksi Kütlesi	0.934626	1	0.676657
Metallilik	0.682461	0.676657	1

Bu ilişkili parametreler belirlenerek ML algoritmasının eğitimi için kullanılacak galaksi özellikleri seçilmiştir. Bu çalışmada ML algoritmasının eğitimi için SVR makine öğrenmesi algoritması kullanılmıştır.

2.3.2 Algoritma- SVR (Destek Vektör Regresyonu)

SVM (Destek Vektör Makineleri) ve SVR (Destek Vektör Regresyonu), makine öğreniminde sınıflandırma problemleri için popüler ve yaygın olarak kullanılmaktadır.



Şekil-14 SVR algoritmasındaki hiper düzlem [33]

Destek Vektör Regresyonu (SVR), SVM ile aynı prensibi kullanır, ancak regresyon problemleri içindir. SVR'de amaç, bir eğitim için kullanılan veri örneğinin girdilerinden sayısal değerlere eşleşmeyi yaklaşık olarak yapan bir fonksiyon bulmaktır. Şekil 14'de görülen iki kırmızı çizgi karar sınırı ve yeşil çizgi de hiper düzlem olarak düşünülür. SVR ile ilerlerken amaç temelde karar sınırı çizgisi içinde kalan noktaları dikkate

almaktır. En iyi uyum çizgisi, maksimum sayıda noktaya sahip olan hiper düzlemdir. Öncelikle, karar sınırlarının ne olduğunu belirlemek gerekir ve bunlar şekil 14'deki kırmızı çizgilerdir. Bunlar, hiper düzlemden Güneş+ a' ve Güneş-a' uzaklıkta çizilen çizgilerdir. "a", temelde epsilon olarak anılır. Buradaki amaç hiper düzleme en yakın veri noktaları veya destek vektörleri bu sınır çizgisi içinde olacak şekilde orijinal hiper düzlemden 'a' uzaklıkta bir karar sınırı belirlemektir. Hiper düzlemin denklemi +a ve - a arasında olmalıdır [33].

Kullanılan SVR algoritmasının parametreleri aşağıdaki gibi seçilmiştir. SVR(kernel='poly', C=1, gamma='auto', degree=3, epsilon=.1, coef0=1);

- **Kernel:** {'doğrusal', 'poli', 'rbf', 'sigmoid', 'önceden hesaplanmış'}, varsayılan = 'rbf' Algoritmada kullanılacak çekirdek türünü belirtir. "Doğrusal", "poli", "rbf", "sigmoid", "önceden hesaplanmış" veya çağrılabilir özelliklerden biri olmalıdır. Hiçbiri belirtilmezse, 'rbf' kullanılacaktır.
- **C:** varsayılan = 1.0 Düzenlilik parametresi. Düzenlemenin gücü C ile ters orantılıdır. Kesinlikle olumlu olmalıdır.
- **Gamma:** {'scale', 'auto'} veya float, default = 'scale' 'Rbf', 'poli' ve 'sigmoid' için çekirdek katsayısı.
- **Degree:** varsayılan = 3 Polinom çekirdek fonksiyonunun derecesi ('poli'). Diğer tüm çekirdekler tarafından yok sayılır.
- **Epsilon:** varsayılan = 0.1 Epsilon-SVR modelinde Epsilon. Gerçek değerden epsilon mesafesi içinde tahmin edilen noktalarla eğitim kaybı fonksiyonunda hiçbir hatanın ilişkilendirilmediği belirtir.
- **coef0:** varsayılan = 0.0 Çekirdek işlevinde bağımsız terim. Yalnızca 'poli' ve 'sigmoid'de anlamlıdır

Dağınıklık matrisi (confusion matrix) bize kaç tahminin doğru veya yanlış sınıflandırıldığını söyler. Bunu tahmin sonucuna ve gerçek değere uygulamadan önce bir sınır değeri seçmemiz gereklidir. Seçtiğimiz ayrılma değeri (cut-off) 12 (logaritmik galaksi kütlesi) olduğunda dağınıklık matrisi [[3471, 65], [122, 342]] şeklinde elde edilir.

2.3.3 Veri Analizi

Bu çalışmada ML algoritmasının eğitimi için SVR makine öğrenmesi algoritması tercih edilmiştir. Python'da SVR algoritmasının çözümü için "scikit-learn" kütüphanesi

kullanılmıştır [34]. ML Algoritmalarının gerçeğe yakın sonuçlar alabilmesi için veri setinin çok iyi analiz edilmesi gerekmektedir. Matris korelasyonu bu yöntemlerden biridir. Matris korelasyonun ardından hangi değerlerin özellik (feature), hangilerinin etiket (label) olduğunun tespiti yapılır. Daha sonra veri seti eğitim (training) ve test olarak ikiye ayrılır. Bu çalışmada MICECATv2 kataloğundan elde edilen 20000 galaksiye ait veri setinin %80'i eğitim ve %20'si test için kullanılmıştır.

Özellik (feature) olarak galaksi kütlesi ve metallilik özelliği, etiket (label) olarak galaksinin karanlık madde halo kütlesi seçilmiştir. Sklearn kütüphanesi özellikleri (feature) X, etiketleri (label) GÜNEŞ olarak kabul edip veri setini iki parçaya aşağıdaki Python komutu ile bölmektedir (Ek-1 Out(12)):

```
X_train, X_test, Güneş_train, Güneş_test = train_test_split(X, Güneş, test_size=0.2, random_state=0)
```

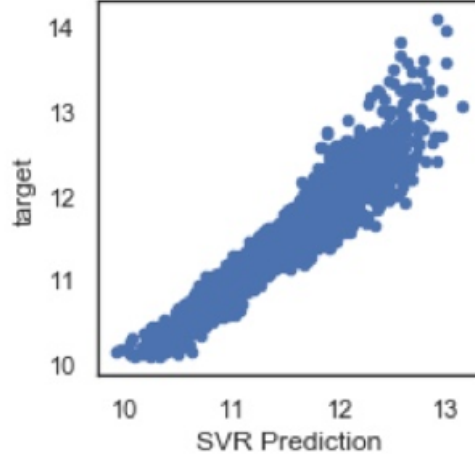
Burada test_size parametresinin 0.2 olması test için kullanılacak verinin tüm verinin yüzde 20'si olacağını göstermektedir. Veri seti eğitiminin ardından standart sapma değeri bir çıktı olarak yer almaktadır, standart sapma geçek değerle tahmin edilen değer arasında hesaplanır. Bu çalışmada programın öğrenmesinin ardından standart sapma değeri 0.0008208852901361433 (Ek-1 Out(23)) ve bunun yanında, ortalama değeri -0.029 ve sonuç değeri de -0.028 olarak bulunmuştur. Programın testi için test verisinin içinden rastgele galaksiler seçilmiş ve programın galaksinin karanlık madde halo (logaritmik) kütlelerine ait tahminleri ile MICECAT v2 kataloğundan elde edilen test verisindeki hedef değerler (Imhalo) karşılaştırılmıştır. Bu karşılaştırma Tablo-8'de gösterilmiştir.

Tablo-8 (ML) SVR algoritmasının galaksilerin karanlık madde halo kütlesi tahminleri karşılaştırması (Ek-1 Out (24))

	Galaksinin karanlık madde halo logaritmik kütlesi ($\text{dexM}_{\text{Güneş}}$)		
#Galaksi	Test verisindeki hedef değer (Imhalo) ($\text{dexM}_{\text{Güneş}}$)	ML (SVR) algoritmasının tahmini ($\text{dexM}_{\text{Güneş}}$)	Hata (%)

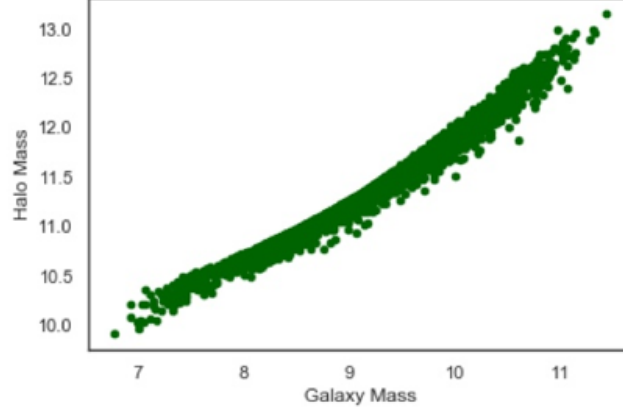
19134	11.4789	11.421379	0.501102022
4981	12.0249	11.951199	0.612903226
16643	11.4397	11.392184	0.415360543
19117	11.4458	11.917531	4.121433189
5306	11.243	11.241322	0.014924842
230	10.7342	10.667523	0.62116413
3148	11.0259	11.13101	0.953300864
11525	11.5512	11.478229	0.631717917
13672	11.5092	11.552885	0.379565912
1624	11.2098	11.111447	0.877384075

Benzer şekilde test verisindeki galaksinin karanlık madde halo logaritmik kütlesine ilişkin tüm hedef değerler ile (ML) SVR algoritmasının tahminleri karşılaştırıldığında Şekil-15'deki grafik elde edilmiştir.



Şekil-15 Test verisindeki 20000 galaksinin karanlık madde halo logaritmik kütlesine ilişkin SVR tahmini ve hedef değerlerin dağılımı ($\text{dex}M_{\text{Güneş}}$) (Ek-1 Out(25))

MICECAT v2 kataloğundan elde edilen test verisindeki galaksilerin Imstellar parametresi ile verilen galaksi kütlesi yıldız kütlesi (logaritmik) ($\text{dex}M_{\text{Güneş}}$) ile (ML) SVR algoritmasının tahmin ettiği galaksinin karanlık madde halo kütlesi değerleri arasındaki ilişkiyi görmek için Şekil-16'daki grafik çizdirilmiştir.



Şekil-16 Test veri setindeki 20000 galaksinin kütlesi ile galaksilerin (ML) SVR algoritması ile tahmin edilen karanlık madde halo kütlesi ($dexM_{Güneş}$) (Ek-1 Out(27))

Bu iki grafiğe göre oluşturulan ML algoritmasının güvenilirliği ve hassasiyeti görülmektedir. Graifiklerden çıkan skor değeri -0.029083030638803853 olup algoritmamızın tahminlerinin oldukça doğru olduğunu söylemektedir.

2.3.4 SDSS Kataloğundaki Galaksilerin Karanlık Madde Kütlesinin Tahmini

Eğitimi ve testi tamamlanan (ML) SVR algoritması SDSS DR16 kataloğundaki farklı galaksilerin karanlık madde halo kütlelerini tespit etmek için kullanılmıştır. SDSS DR16 Bölüm-2.1.1'de anlatıldığı gibi Tablo-4'de verilen galaksi parametrelerine göre CSV formatında 20000 farklı galaksinin verileri toplanmıştır. Bu veriler Python ile ayıklanmıştır ve ilk beş galaksi verisi örnek olması açısından Tablo-9'da sunulmuştur.

Tablo-9 SDSS DR16 kataloğundan alınan 20000 galaksi içinden alınan ilk beş galaksinin değerleri (Ek-1 Out(29))

#	0	1	2	3	4
ID	2994896769 75171584	2994896774 44933632	2994899518 53078528	2994899523 22840576	2994905016 08892416
ra	146.714	146.714	146.919	146.744	146.86
dec	-1.0413	-1.0413	-0.9905	-0.6522	-0.8089
z	0.02128	0.02122	0.21393	0.20378	0.12655

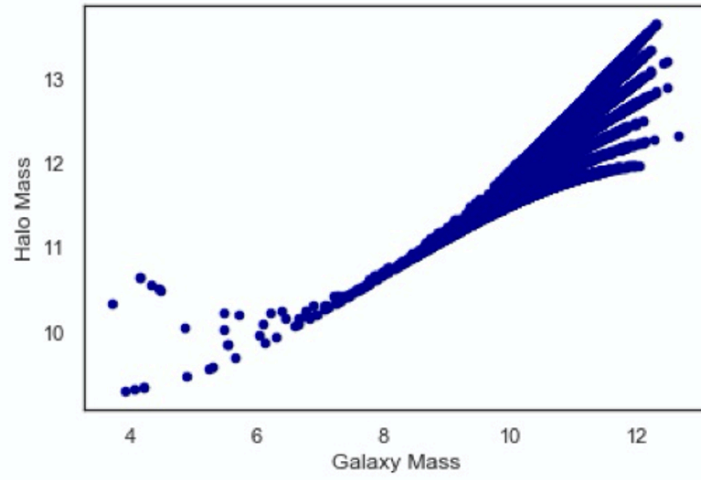
metallicity	0.015	0.015	0.019	0.019	0.03
dust1	0.15015	0.15015	0.15015	0.15015	0.15015
dust2	0.05	0.05	0.05	0.05	0.05
tau	2.56669	2.56669	0.51372	0.51372	0.01547
age	10.0012	10.0012	2.65309	5.30648	11.7898
logMass	10.4842	10.482	11.1885	11.4241	11.328
t_age	12.4703	12.4703	3.16006	5.82013	11.8053
ssfr	-2.2861	-2.2861	-2.1465	-4.3786	-33
ssfr_mean	-2.3408	-2.3408	-7.9944	-13.202	-29.701

(ML) SVR algoritmamız SDSS DR16 kataloğundaki $z=0$ ile $z=1$ arasındaki galaksilere uygulanmış ve bu galaksilerin karanlık madde halo kütlesi tahmin edilmiştir. SDSS DR16 kataloğundaki bazı galaksiler için (ML) SVR algoritmamızın tahmin ettiği karanlık madde halo kütlesi ile SDSS DR16 kataloğundaki galaksi kütlesi ve metalilik değerleri Tablo-10'da verilmiştir. (ML) SVR algoritmasında metalilik özelliği etiket olarak kullanılmıştır ve eğitilen program SDSS DR16 da bulunan 20000 galaksiye uygulanmış ve ilk beş sonuç örnekleme açısından Tablo-10'da gösterilmiştir.

Tablo-10 SDSS DR16 kataloğundaki galaksiler için karanlık madde halo kütlesi tahmini (Ek-1 Out(35))

ID	Logaritmik tahmin edilen karanlık madde kütlesi (dexM _{Güneş})	Logaritmik Galaksi kütlesi (dexM _{Güneş})	Metallilik ($\text{Log} \left(\frac{O}{H} \right) + 12$)
299489676975171584	11.986	10.484	0.015
299489677444933632	11.985	10.482	0.015
299489951853078528	12.502	11.188	0.019
299489952322840576	12.639	11.424	0.019
299490501608892416	12.895	11.328	0.030

SDSS DR16 Veri setinden alınmış 20000 galaksinin kütlesi ve metallilik özelliği ML SVR ile eğitilmiş programa girdi olarak verildiğinde her galaksiye ait karanlık madde halo kütlesi tespit edilmiştir ve tahmin sonuçları Şekil-17’de gösterilmiştir. Şekil-17 ve Şekil-16 karşılaştırıldığında özellikle galaksi kütlesi için 7 ile 11 ($\text{dex}M_{\text{Güneş}}$) arasına bakıldığında her iki grafikte benzer bir uyum olduğu görülmüştür.



Şekil-17 SDSS DR16 veri setinden alınmış 20000 galaksi kütlesi değerleri ile galaksilerin (ML) SVR algoritması ile tahmin edilen karanlık madde halo kütlesi dağılımı ($\text{dex}M_{\text{Güneş}}$) (Ek-1 Out(36))

3. SONUÇLAR

Bu çalışmanın sonucunda SDSS DR16 kataloğunda bulunan 20000 galaksi için karanlık madde halo kütleleri MICECATv2 mock kataloğundaki 20000 sentetik galaksi verisi ile eğitilen ve test edilen ML algoritması tarafından tahmin edilmiştir. Galaksilerdeki karanlık madde gözlem sonucu elde edilemediğinden, uzun işlem zamanı gerektiren programlar kullanılır ancak ML algoritması kullanılarak hesaplamalar saniyeler içinde yapılmaktadır. Araştırmanın sonunda karanlık madde halosunun kütlesi 0.00082 standart sapma ile tahmin edilmiştir. Eğitilecek verinin

artırılması hata değerini küçültecektir. Hata değerini küçültmenin başka bir yolu da başka ML algoritmaları kullanmak olabilir. Bu çalışmada galaksinin metallilik özelliği etiket olarak seçilip SVR algoritması kullanılarak tahmin edilen sonuçlardaki hatanın çok küçük olması sebebiyle bu algoritmanın kullanılmasına karar verilmiştir.

3.1 Bulgular

- MICECAT V2 sentetik kataloğundaki 20000 farklı galaksi için elde edilen veriler Şekil-13'deki korelasyon matrisi oluşturulmuştur. Pembe renge kayan kareler yani sayısal değer olarak 0.2'ye yakın olan özelliklerin birbirleri arasında bir ilişki olabileceği ve tam tersine, sayısal değeri -0.6 değerine yani yeşilimsi karelere sahip özelliklerin birbirleri arasında bir ilişki olmadığı görülmüştür. Bu sıcaklık dağılımı haritası sayısal değerlerine bakılarak incelendiğinde galaksinin karanlık madde halosunun kütlesi (Mhalo), galaksi kütlesi (Mstaller) ve metallilik (metallicity) özellikleri arasında bir ilişki bulunmuştur.
- Bu parametreler arasındaki korelasyon ayrıntılı incelendiğinde galaksinin Metallilik özelliğinin korelasyon matrisinde en düşük değere sahip olduğu ortaya çıkmıştır. Bu sebepten geliştirilen ML programını eğitmek için Metallilik özelliği etiket olarak seçilmiştir [Ek1 Out(8) ve Ek-1 Out(5)].
- ML algoritmasıyla programın eğitiminde test tamamlandıktan sonra gerçek değer ile tahmin edilen değerler arasında hesaplanan standart sapma değeri 0.0008208852901361433 olarak bulunmuştur. [Ek-1 Out(23)]
- Algoritmamızda hassasiyeti gösteren skor değeri -029083030638803853 olarak hesaplanmıştır. [Ek-1 Out(26)]
- MICECATv2 mock kataloğundan alınan 20000 galaksi karanlık madde halo kütlesi değeri ile geliştirilen ML algoritmasının tahmin ettiği halo kütlesi değerleri karşılaştırıldığında (Şekil-15) tahmin değerlerinin hedef değerlerle çok iyi bir uyum içinde olduğu ve hatanın küçük olduğu görülmüştür [Ek-1 Out(24)].
- MICECATv2 galaksi veri setindeki galaksilerin kütleleri ve galaksilerin (ML) SVR algoritması ile tahmin edilen karanlık madde halo kütleleri

karşılaştırıldığında (Şekil-16) Galaksilerdeki karanlık madde halo kütlesi arttıkça görünür kütle değerinin de arttığı tespit edilmiştir.

- SDSS DR16 veri setindeki 20000 farklı galaksi için geliştirilmiş ML SVR algoritması kullanılarak galaksilerin karanlık madde halo kütleleri belirlenmiştir (Şekil-17). Galaksi kütlelerinin 7-11 (dexMGüneş) değerleri arasındaki veriler için Şekil-16 ve Şekil-17'de benzer bir eğilim olduğu görülmüştür. Bu eğilim bu değerler arasında galaksilerin tahmin edilen halo kütlelerinin doğruluğunu göstermektedir [Ek-1 Out(36)]. 10^7 ile 10^{10} $M_{Güneş}$ arasında kütleyle sahip galaksiler için karanlık madde halo kütlelerinin galaksi görünür kütlelerinden yaklaşık 100 kat fazla olduğu görülmüştür. 10^{11} $M_{Güneş}$ 'den büyük galaksi kütleleri için ML algoritmasının karanlık madde halo kütlesi tahminleri saçılma göstermektedir. 20000 galaksinin halo kütlelerini ve galaksi kütlelerini içeren CSV formatındaki çıktı dosyasına aşağıdaki linkten ulaşılabilir ¹.

3.2 Tartışma

Galaksilerdeki karanlık madde doğrudan gözlem sonucu elde edilemediğinden ve bunun için uzun işlem zamanı gerektiren programlar kullanılması gerekirken, ML algoritması kullanılarak sonuçlara saniyeler içinde ulaşabilmektedir. ML hızlı olmasının yanısıra literatürdeki mevcut kullanılan bir çok yöntemden de daha iyi sonuç vermektedir.

Bu çalışma için önerilebilecek iyileştirme yöntemlerinden biri ML algoritması eğitilirken kullanılan veri setinin artırılmasıdır çünkü bu tahminlerdeki hatayı değerini küçültecektir. Hata değerini küçültmenin başka bir yolu da başka ML algoritmaları kullanmak olabilir. Bu çalışmada galaksinin halo kütlesi etiket, metallilik ve galaksi kütlesi özellik olarak seçilmiştir. Kullanılan ML SVR algoritması uygulandığında hatanın çok küçük çıkması sebebiyle bu algoritmanın kullanılmasına karar verilmiştir. Bu çalışma ileride farklı ML algoritmaları (Random Forest, XGBoost, Neural Network vb.) ve farklı etiketler kullanılarak tekrarlanabilir. Hangi algoritmaların hangi özellikler için daha iyi sonuçlar vereceğini görmek için karşılaştırmalar yapılabilir.

¹ <https://dropbox.com/s/4hhqccnxw9u7ycm/output.csv?dl=0>

KAYNAKÇA

- [1] Esa, "Planck", 10.11.2020 <https://sci.esa.int/web/planck/-/51557-planck-new-cosmic-recipe>
- [2] P.Fosalba, M.Crocce, E.Gaztanaga, and F.J.Castander, "The MICE Grand Challenge Lightcone Simulation I: Dark matter clustering", Mon. Not. R. Astron. Soc., 2015
- [3] Debasish Majumdar, Dark Matter An Introduction, US: CRC Press, 2014.
- [4] Ken Freeman ve Geof McNamara, In Search of Dark Matter, Germany: Praxis Publishing, 2006.
- [5] Ş.Ç.Türksoy, "Glaksi Grupları ve Kümeleri", Ankara Üniversitesi Açık Ders Malzemeleri, 09.11.2020.
<https://acikders.ankara.edu.tr/course/view.php?id=4461#section-5>
- [6] George Djorgovski , California Institute Of Technology, Caltech Astronomy, Ay 127, Winter 2017, Lecture04, The Contents of the Universe. 09.11.2020.
https://sites.astro.caltech.edu/~george/ay127/Ay127_SGD_Lec04.pdf
- [7] M.Girardi, A. Biviano, G.Giuricin, F.Mardirossian and M.Mezzetti, "Velocity Dispersions in Galaxy Clusters", The Astrophysical journal, 404:38-50,1992 August5
- [8] Alexandros Papastergiou, "The M/L Ratio Of Clusters And Ω_m Parameter", Aristotle University Of Thessaloniki, Faculty of Physics Department of Science , February 2015.
- [9] Riess, Macri, "A Redetermination Of The Hubble Constant With The Hubble Space Telescope From A Differential Distance Ladder", et al., ApJ, 699, 539 (2009)

- [10] G. Bisnovatyi-Kogan, "The gravitational Lens as a Radiospectrometer", Mem. S.A.It. Vol. 81, 144, (2010)
- [11] Mahabal A., et al., 2008, *Astronomische Nachrichten*, 329, 288
- [12] Banerji M., et al., 2010, *Monthly Notices of the Royal Astronomical Society*, 406, 342
- [13] Ball N. M., Brunner R. J., Myers A. D., Strand N. E., Alberts S. L., Tchong D., Llorca X., 2007, *The Astrophysical Journal*, 663, 774
- [14] Gerdes D. GÜNEŞ., Sypniewski A. J., McKay T. A., Hao J., Weis M. R., Wechsler R. GÜNEŞ., Busha M. T., 2010, *The Astrophysical Journal*, 715, 823
- [15] Xu X., Ho S., Trac GÜNEŞ., Schneider J., Poczós B., Ntampaka M., 2013, *Astrophysical Journal*, 772, 147
- [16] Kamdar GÜNEŞ. M., Turk M. J., Brunner R. J., 2016a, *Monthly Notices of the Royal Astronomical Society*, 455, 642
- [17] Riccio G., et al., 2016, in Bassis S., Esposito A., Morabito F. C., Pasero E., eds, *Advances in Neural Networks*. Springer International Publishing, Cham, pp 27–36, doi:10.1007/978-3-319-33747-0_3, http://link.springer.com/10.1007/978-3-319-33747-0_3
- [18] Ntampaka M., Trac GÜNEŞ., Sutherland D. J., Battaglia N., Póczós B., Schneider J., 2015, *The Astrophysical Journal*, 803, 50
- [19] Armitage T. J., Kay S. T., Barnes D. J., 2019, *Monthly Notices of the Royal Astronomical Society*, 484, 1526

[20] Ntampaka M., “A Deep Learning Approach to Galaxy Cluster X-ray Masses”, et al., 2018, arXiv e-prints

[21] Victor F. Calderon, Andreas A. Berlind, “Prediction of Galaxy Halo Masses in SDSS DR7 via a Machine Learning Approach”, 08.02.2019, MNRAS 000, 1–13

[22] The Sloan Digital Sky Survey: Data Release 16, 21.11.2020
<https://Güneş.sdss.org/dr16/>

[23] Wikipedia, Sloan Dijital Gökyüzü Araştırması, 09.11.2020
https://tr.wikipedia.org/wiki/Sloan_Dijital_G%C3%B6ky%C3%BCz%C3%BC_Ara%C5%9Ft%C4%B1rmas%C4%B1

[24] Prf. Dr Ethem Derman, 09.11.2020,
<http://Güneş.ethemderman.com/media/kitap/7.html>

[25] John Huchra, Harvard-Smithsonian Center for Astrophysics, Extragalactic Redshifts, 21.11.2020, <https://ned.ipac.caltech.edu/help/zdef.html>

[26] M. S. Bothwell, R. Maiolino, R. Kennicutt, Jr, G. Cresci, F. Mannucci, A. Marconi and C. Cicone, “A fundamental relation between the metallicity, gas content and stellar mass of local galaxies”, Advance Access publication, MNRAS 433, 1425–1435, 2013 June 17

[27] M. Boquien, V. Buat, and V. Perret, “Impact of star formation history on the measurement of star formation rates”, Astronomy & Astrophysics manuscript no. Article, (2018)

[28] Marenstrum Institut de Ciències de l'Espai Simulations, 21.11.2020,
<http://maia.ice.cat/mice/>

[29] F.Castender, J.Carretero, M. Croce, p.Fosalba and E. Gaztanaga, Building Galaxy Mock Catalogues with MICE, 2012, August 05.

[30] Anaconda, Inc. (formerly Continuum Analytics, Inc.), 10.12.2020
<https://docs.anaconda.com/ae-notebooks/4.2.2/user-guide/basic-tasks/apps/jupyter/>

[31] NumFOCUS, 10.12.2020 <https://pandas.pydata.org/docs/>

[32] NumFOCUS, 10.12.2020 <https://numpy.org/>

[33] Alakh Sethi, "Support Vector Regression Tutorial for Machine Learning",
09.11.2020, <https://Güneş.analyticsvidhya.com/blog/2020/03/support-vector-regression-tutorial-for-machine-learning/>

[34] INRIA, 09.11.2020, <https://scikit-learn.org/stable/>

EKLER

ML programı Python Jupyter Notebook aracı kullanılarak geliştirilmiştir. Aşağıda programın kaynak kodu verilmiştir.

```
In [1]: # Linear Regression Using SGD - Supervised Continuous ...
# Machine Learning SVR algorithm
# predict Halo Mass from galaxy properties
# MICECAT v2 catalogue
# Prediction has been approved by using MICECAT version 2 dataset
#unique_halo_id      unique halo id
#ra_gal              galaxy right ascension (degrees)
#dec_gal             galaxy declination (degrees)
#z_cgal              galaxy true redshift
#cgal                galaxy comoving distance (Mpc/h)
#z_cgal_v            galaxy observed redshift (including peculiar velocity)
#cgal_v              galaxy comoving distance corresponding to z_cgal_v (Mpc/h)
#z_desdm_mc          desdm photometric redshift monte carlo assigned to follow the benchm
ark N(z)
#flag_central        flag 0==central 1==satellite
#lmhalo              log base 10 of FoF halo mass
#v_vir               halo virial velocity (km/h)
#nsats               number of satellites
#xhalo               halo comoving distance x-axis (Mpc/h)
#yhalo               halo comoving distance y-axis (Mpc/h)
#zhalo               halo comoving distance z-axis (Mpc/h)
#vxhalo              halo physical (peculiar) velocity in x-component (km/h)
#vyhalo              halo physical (peculiar) velocity in y-component (km/h)
#vzhalo              halo physical (peculiar) velocity in z-component (km/h)
#unique_gal_id       unique galaxy id
#lsfr                logarithm of star formation rate in (Msolar/h)/year
#lmstellar            logarithm of the stellar mass (Msolar/h)
#metallicity         oxygen metallicity in units 12+logO/H

# SDSS DR16 stellarMassFSPSGranEarlyDust
#specObjID           Unique ID
#ra                  deg Right ascension of fiber, J2000
#dec                 deg Declination of fiber, J2000
#z                   Redshift used (corresponds to z_nogso in specObjAll)
#logMass              Best-fit stellar mass of galaxy
#t_age               Best-fit look-back formation time
#metallicity          Best-fit metallicity, where Z_sun=0.019
#dust1               Best-fit value for dust attenuation around young stars
#dust2               Best-fit value for dust attenuation around old stars
#tau                 Best-fit star formation history e-folding time (tau)
#age                 Best-fit mass-weighted average age of the stellar population
#ssfr                Best-fit specific star formation rate
#ssfr_mean           Best-fit specific mean of star formation rate

# predict the galaxies halo mass in SDSS DR16
```

1. Machine Learning algorithm has been train and test using MICECAT dataset

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

```
In [2]: fn_s = "/Users/typhoons/Dropbox/Pyhsics_Master/Pyhton/Galaxy_Halo_ML/HaloData/MicecatCatalogue/7800.csv"
```



```
In [7]: hdf_mcc.head()
```

Out[7]:

	lmhalo	lmstellar	metallicity
0	12.1339	10.083	8.972
1	14.0009	10.602	8.963
2	12.4957	10.514	9.128
3	12.1191	10.436	8.996
4	11.9210	9.977	8.898

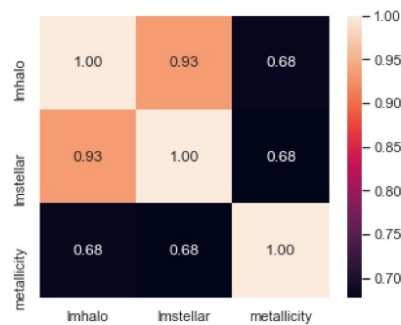
```
In [8]: corr_matrix1 = hdf_mcc.corr()
corr_matrix1
```

Out[8]:

	lmhalo	lmstellar	metallicity
lmhalo	1.000000	0.934626	0.682461
lmstellar	0.934626	1.000000	0.676657
metallicity	0.682461	0.676657	1.000000

```
In [9]: plt.figure(figsize=(5, 4))
sns.heatmap(corr_matrix1,
            xticklabels=corr_matrix1.columns,
            yticklabels=corr_matrix1.columns, annot=True, fmt=".2f")
```

Out[9]: <matplotlib.axes._subplots.AxesSubplot at 0x7f884b0456d8>



sklearn SVR

```
In [10]: from sklearn.preprocessing import Imputer, MinMaxScaler, StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.base import BaseEstimator, TransformerMixin
from sklearn.pipeline import Pipeline, FeatureUnion
```

```
In [11]: #splitting feature and label
dataset_h=hdf_mcc.copy()
X=dataset_h.drop(["lmhalo"], axis=1) #stellar mass has been deleted form datset (drop)
y=dataset_h["lmhalo"]
```

```
In [12]: #splitting test and train set
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)
```

pipelines minmax_pipe and st_pipe

```
In [13]: #define a custom transformer class for feature scaling
class FeatureSelector(TransformerMixin, BaseEstimator):
    def __init__(self, feature_list):
        self.feature_list = feature_list
    def fit(self, X, y=None):
        return self
    def transform(self, X):
        return X[self.feature_list].values
```

```
In [14]: minmax_feature=["metallicity"]
st_feature= ["lmstellar"]
```

```
In [15]: from sklearn.impute import SimpleImputer

#minmax scaling pipeline
#imputer... If "mean", then replace missing values using the mean along the axis.
minmax_steps= [("mm_selector", FeatureSelector(minmax_feature)),
               ("imputer", SimpleImputer(strategy= 'mean')),
               ("mm_scaler", MinMaxScaler())
              ]
```

```
In [16]: minmax_pipe= Pipeline(minmax_steps)
```

```
In [17]: #standard scaling pipeline
st_steps = [("st_selector", FeatureSelector(st_feature)),
            ("imputer", SimpleImputer(strategy= "mean")),
            ("st_scaler", StandardScaler())
           ]
```

```
In [18]: st_pipe=Pipeline(st_steps)
```

```
In [19]: #combined pipeline to concateante output of the above two pipeline
prep_pipe=FeatureUnion([("minmax", minmax_pipe), ("st", st_pipe)])
```

Regression problem using SVR

```
In [20]: from sklearn.ensemble import RandomForestRegressor
from sklearn.svm import SVR
from sklearn.linear_model import SGDRegressor
from sklearn.model_selection import cross_val_score

import warnings
warnings.filterwarnings('ignore')
```

```
In [21]: def display_scores(scores):
print("Scores:", scores)
print("Mean:", scores.mean())
print("Standard Deviation:", scores.std())
```

```
In [22]: #SVR Regression
SV_steps=[("prep", prep_pipe), ("SVR", SVR(kernel='poly', C=1, gamma='auto', degree=3, epsilon=.1, coef0=1))]
SV_regressor=Pipeline(SV_steps)
```

```
In [23]: CV_score=cross_val_score(SV_regressor, X_train, y_train, cv=5, scoring="neg_mean_squared_error")
display_scores(CV_score)
```

```
Scores: [-0.02825239 -0.02876696 -0.03058813 -0.02855039 -0.02925729]
Mean: -0.029083030638803853
Standard Deviation: 0.0008208852901361433
```

```
In [24]: SV_regressor.fit(X_train, y_train)
report = pd.DataFrame({"target": y_test, "SVR Prediction": SV_regressor.predict(X_test)})
report.head(10)
```

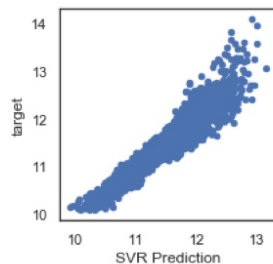
Out[24]:

	target	SVR Prediction
19134	11.4789	11.421379
4981	12.0249	11.951199
16643	11.4397	11.392184
19117	11.4458	11.917531
5306	11.2430	11.241322
230	10.7342	10.667523
3148	11.0259	11.131010
11525	11.5512	11.478229
13672	11.5092	11.552885
1624	11.2098	11.111447

```
In [25]: report.plot.scatter(x="SVR Prediction", y = "target", figsize=[3,3])
```

'c' argument looks like a single numeric RGB or RGBA sequence, which should be avoided as value-mapping will have precedence in case its length matches with 'x' & 'y'. Please use a 2-D array with a single row if you really want to specify the same RGB or RGBA value for all points.

Out[25]: <matplotlib.axes._subplots.AxesSubplot at 0x7f884b0745c0>

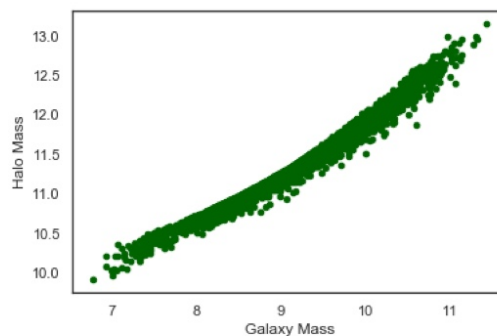


```
In [26]: score = SV_regressor.score(X_test, y_test)
print(score)
```

0.9034719072262455

```
In [27]: galaxyM_haloM=hdf_mcc.copy()
galaxyM_haloM["PrdHaloMass"] = report[["SVR Prediction"]]
galaxyM_haloM.plot.scatter(x='lmstellar', y='PrdHaloMass', c='DarkGreen')
plt.xlabel("Galaxy Mass")
plt.ylabel("Halo Mass")
```

Out[27]: Text(0, 0.5, 'Halo Mass')



In []:

2. predict the Halo mass of the SDSS DR16 galasiex

```
In [28]: fn_sdss = "/Users/typhoons/Dropbox/Pyhsics_Master/Pyhton/Galaxy_Halo_ML/HaloData/SDSSDR16/galaxy mass2SDSS16.csv"
```

```
In [29]: hdf_sdss = pd.read_csv(fn_sdss)
hdf_sdss.head()
```

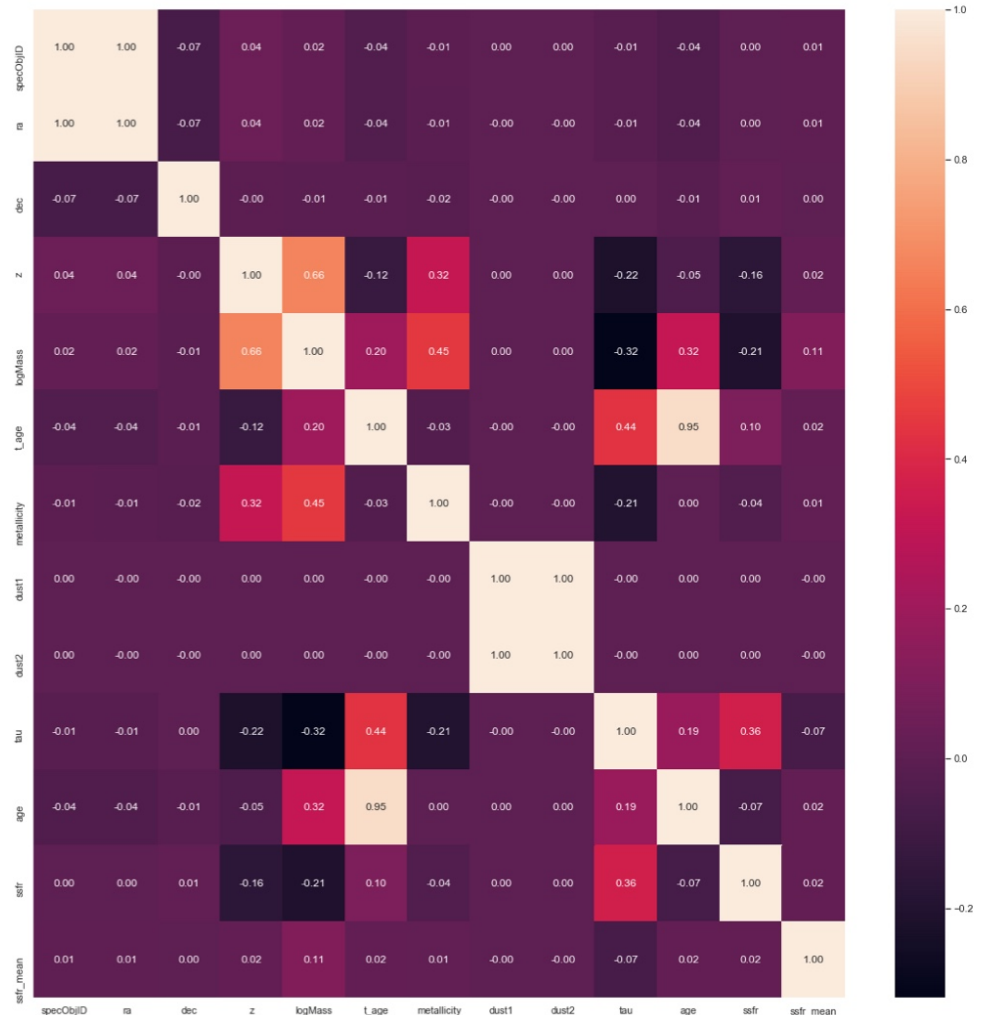
Out[29]:

	specObjID	ra	dec	z	logMass	t_age	metallicity	dust1	dust2	tau	
0	299489676975171584	146.71421	-1.041304	0.021275	10.484221	12.470287	0.015	0.15015	0.05	2.566691	10.00
1	299489677444933632	146.71421	-1.041304	0.021222	10.482012	12.470287	0.015	0.15015	0.05	2.566691	10.00
2	299489951853078528	146.91945	-0.990492	0.213925	11.188459	3.160064	0.019	0.15015	0.05	0.513718	2.60
3	299489952322840576	146.74413	-0.652191	0.203783	11.424051	5.820127	0.019	0.15015	0.05	0.513718	5.30
4	299490501608892416	146.85983	-0.808902	0.126554	11.328028	11.805271	0.030	0.15015	0.05	0.015473	11.70

```
In [30]: corr_matrixSDSS = hdf_sdss.corr()
```

```
In [31]: # plot the heatmap
import seaborn as sns
import matplotlib.pyplot as plt
plt.figure(figsize=(20, 20))
sns.heatmap(corr_matrixSDSS,
            xticklabels=corr_matrixSDSS.columns,
            yticklabels=corr_matrixSDSS.columns, annot=True, fmt=".2f")
```

Out[31]: <matplotlib.axes._subplots.AxesSubplot at 0x7f8835796668>



```
In [32]: hdf_sdsss = hdf_sdsss.copy()
hdf_sdsssS = hdf_sdsss.drop(["specObjID", "ra", "dec", "z", "t_age", "dust1", "dust2", "tau", "age", "ssfr", "ssfr_mean"], axis = 1)
hdf_sdsssS.head()
```

Out[32]:

	logMass	metallicity
0	10.484221	0.015
1	10.482012	0.015
2	11.188459	0.019
3	11.424051	0.019
4	11.328028	0.030

```
In [33]: hdf_sdsssS['metallicity'] = 12 + (np.log10((0.35*hdf_sdsssS['metallicity'])/(16.0*0.74)))
hdf_sdsssS.columns = ["lmstellar", "metallicity"]
hdf_sdsssS.head()
```

Out[33]:

	lmstellar	metallicity
0	10.484221	8.646808
1	10.482012	8.646808
2	11.188459	8.749470
3	11.424051	8.749470
4	11.328028	8.947838

```
In [34]: prd_halo_mass_SDSS = pd.DataFrame({ "SVR Prediction Halo mass of Galaxy":SV_regressor.predict(hdf_sdsssS)})
prd_halo_mass_SDSS.head()
```

Out[34]:

	SVR Prediction Halo mass of Galaxy
0	11.986316
1	11.985166
2	12.502143
3	12.639224
4	12.894924

```
In [35]: prd_halo_mass_SDSS["specObjID"] = hdf_sdsss["specObjID"]
prd_halo_mass_SDSS["logMass"] = hdf_sdsss["logMass"]
prd_halo_mass_SDSS["metallicity"] = hdf_sdsssS["metallicity"]

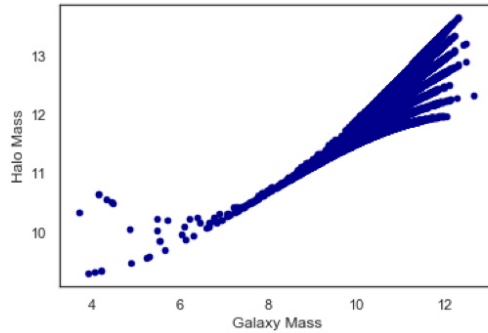
pd.options.display.float_format = '{:,.3f}'.format
prd_halo_mass_SDSS.applymap(lambda x: str(int(x)) if abs(x - int(x)) < 1e-6 else str(round(x,3)))
prd_halo_mass_SDSS.head()
```

Out[35]:

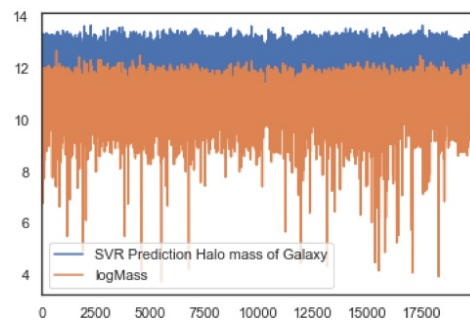
	SVR Prediction Halo mass of Galaxy	specObjID	logMass	metallicity
0	11.986	299489676975171584	10.484	8.647
1	11.985	299489677444933632	10.482	8.647
2	12.502	299489951853078528	11.188	8.749
3	12.639	299489952322840576	11.424	8.749
4	12.895	299490501608892416	11.328	8.948

```
In [36]: hdf_plot1 = prd_halo_mass_SDSS.drop(["specObjID", "metallicity"], axis = 1)
hdf_plot = hdf_plot1.iloc[:20000, :]
hdf_plot.plot.scatter(x='logMass', y='SVR Prediction Halo mass of Galaxy', c='DarkBlue')
plt.xlabel("Galaxy Mass")
plt.ylabel("Halo Mass")
```

Out[36]: Text(0, 0.5, 'Halo Mass')



```
In [37]: lines = hdf_plot.plot.line()
```



```
In [38]: axes = hdf_plot.plot.line(subplots=True)
type(axes)
```

Out[38]: numpy.ndarray

