ASTRONOMİK GÖZLEM KALİTESİ TAHMİNİNDE MAKİNE ÖĞRENMESİ ALGORİTMALARININ KULLANIMI[+]

***

USE OF MACHINE LEARNING ALGORITHMS IN ASTRONOMICAL OBSERVATION QUALITY ESTIMATION

Ömer Çağrı YAVUZ[*]
Ersin KARAMAN[**]

DOI: 10.33461/ uybisbbd.860955

Öz

Bu çalışmada gözlem kalitesi tahmininde makine öğrenmesi algoritmaların uygulanabilirliğini ortaya koymak amacıyla elde edilen veri seti farklı şekillerde kümelenmiştir. Sonrasında alanyazında en çok kullanılan sınıflandırma algoritmalarından Naive Bayes ve K En Yakın Komşu algoritmaları kullanılarak tahminler yapılmıştır. Çalışma sonucunda, verilerin ait olduğu kümelerin gözlem türünü ifade ettiği varsayılarak, kullanılan makine öğrenmesi algoritmalarının performansı ölçülmüştür. Yapılan tahminlerde Naive Bayes uygulamasının daha iyi performans gösterdiği görülmüştür. Sonuç olarak, astronomik gözlem verilerinin makine öğrenmesi algoritmaları ile tahmin edilebileceği ve bu alanda buna yönelik bir veri seti oluşturulması ihtiyacının hâsıl olduğu söylenebilir.

***Anahtar Kelimeler**: astronomik gözlem kalitesi, makine öğrenmesi, naive bayes, k en yakın komşu*

Abstract

In this study, the data set, which is obtained in order to reveal the applicability of machine learning algorithms in observation quality estimation, were clustered in different number of clusters. Afterwards, estimations were made using Naive Bayes and K Nearest Neighbor algorithms, which are the mostly used classification algorithms in the literature. As a result of the study, the performance of the machine learning algorithms used was measured, assuming that the clusters to which the data belonged express the type of observation. It was found that the Naive Bayes application performed better estimation performance. In conclusion, it can be said that there is a need to create astronomical observation data set so that can be used for comparing machine learning algorithms.

***Keywords:** astronomical observation quality, machine learning, naive bayes, k-nearest neighbor e-government, local e-government applications, disabled information system.*

[+] Bu çalışma, 10 Aralık 2020 tarihinde 7. Uluslararası Yönetim Bilişim Sistemleri Konferansı'nda sözlü bildiri olarak sunulmuştur.

[*] Arş. Gör. Karadeniz Teknik Üniversitesi, İİBF, Yönetim Bilişim Sistemleri Bölümü, omercagriyavuz@ktu.edu.tr, Trabzon, Türkiye, ORCID: 0000-0002-6655-3754
[**] Doç. Dr. Ankara Hacı Bayram Veli Üniversitesi, İİBF, Yönetim Bilişim Sistemleri Bölümü, ersinkaraman@atauni.edu.tr, Ankara, Türkiye, ORCID: 0000-0001-5459-4172

Geliş Tarihi / Received: 14.01.2021
Kabul Tarihi / Accepted: 25.03.2021

1. GİRİŞ

M.Ö. 600 yıllarında gök cisimlerini gözlemleyen ve gezegen konumlarını kaydeden Babil astronomları tarafından çeşitli gözlemler yapıldığı belirtilmektedir. İlerleyen yıllarda çoğunlukla olgusallığa dayanan Babil astronomisinin yanında çeşitli uygarlıklar tarafından farklı tekniklerle astronomik gözlemler gerçekleştirilmektedir. Bu gözlemlerde kullanılan yöntem ve tekniklerin kullanımı büyük gelişim göstererek günümüzde yapılan gözlemlerde de devam etmektedir (Unat, 2001).

Astronomik gözlemlerden elde edilen verilerin paylaşımı yaygınlaşmakta ve gözlem verilerine erişim kolaylaşmaktadır. Örnek olarak teleskoplardan elde edilen verilerin gözlemcilerle paylaşıldığı ve sanal gözlem yapma imkânı sunan Uluslararası Sanal Gözlemevi Birliği kurulmuştur (Esenoğlu ve Gülseçen, 2019). Ancak elde edilen büyük miktardaki verilerin seçilmesi, normalleştirilmesi, işlenmesi, kaydedilmesi ve bu verilerden çıkarımların yapılmasında çeşitli zorluklarla karşılaşılmaktadır. Yapay Zekâ uygulamalarında kullanılan makine öğrenmesi algoritmalarıyla verilerin işlenme hızı artış göstermektedir. Bu artış ile birlikte Yapay Zekâ disiplini de ilerleme kaydetmektedir (Boffin, Jerabkova, Mérand ve Stoehr, 2019). Bu tekniklerle artan verilerin işlenerek karar almaya yardımcı uygulamaların geliştirilmesi, uygulanması, süreçlerin kolaylaştırılması ve otomatikleştirilmesi amaçlanmaktadır.

Makine öğrenmesi algoritmaları verilerin seçilmesi, dönüştürülmesi, tahmini, işlenmesi gibi amaçlarla çeşitli alanlarda kullanılmaktadır. Benzer şekilde astronomide de bu algoritmaların kullanımına rastlanmaktadır. Örnek olarak Johnston (1987), geliştirdiği prototip ile o yıllarda sorun olarak görülen veri analiz işleminde uzman sistem yaklaşımının uygulanabilirliğini ortaya koymayı amaçlamaktadır. (Albrecht, 1987).

Başka bir çalışmada düşük kırmızıya kaymalı galaksilerin renklerinden belirli bir yıldız oluşum oranının tahmin edilmesinde K En Yakın Komşu algoritmasının başarı oranının test edilmesi amaçlanmıştır. Çalışma sonucunda K En Yakın Komşu algoritmasının son teknoloji astronomik modelden daha iyi performans gösterdiği vurgulanmıştır (Stensbo-Smidt, Igel, Zirm ve Pedersen, 2013). Benzer bir çalışmada da çoklu dalga boyu verilerinden özellikler seçmek için Relief isimli bir filtre yaklaşımı uygulanmıştır. Sonrasında nesneleri alt kümeleriyle sınıflandırmak ve öznitelik ağırlıklarını karşılaştırmak amacıyla Naive Bayes algoritması kullanılmıştır. Çalışma sonucunda Naive Bayes algoritmasının aktif galaktif çekirdek adaylarını seçmek için etkili olduğu belirtilmiştir (Zhang, Luo ve Zhao, 2004).

Bu çalışmada gözlem kalitesi tahmininde makine öğrenmesi algoritmaların uygulanabilirliğini ortaya koymak amaçlanmıştır. Makine öğrenmesi algoritmaları gözetimli öğrenme (supervised learning), gözetimsiz öğrenme (unsupervised learning) ve pekiştirmeli öğrenme olmak üzere üç farklı grupta ele alınmaktadır (Chug, 2018). Gözetimli öğrenme algoritmalarında da sınıflandırma ve regresyon algoritmaları ele alınmaktadır. Gözetimli öğrenme algoritmalarında genellikle sınıflandırılmış örüntülerin bulunduğu veri setleri üzerinden tahminler yapılır. Gözetimsiz öğrenme algoritmalarında da verilerin gruplara ayrıldığı uygulamalar yapılmaktadır.

2. YÖNTEM

Bu çalışmada astronomik gözlem kalitesi tahminine yönelik makine öğrenmesi algoritmaların uygulanabilirliğini ortaya koymak amacıyla sırasıyla verilerin hazırlanması, kümelenmesi ve elde edilen kümeler üzerinde sınıflandırma algoritmalarının uygulanması adımları izlenmiştir.

2.1. Verilerin Toplanması ve Hazırlanması

01.01.2012-01.01.2019 günleri arasında saat 12:00'da günlük olarak elde edilmiş değerlerden oluşan gözlem verileri, Atatürk Üniversitesi Astrofizik Araştırma ve Uygulama Merkezi'nden alınmıştır. Toplamda 16 nitelikten oluşan veriler, etki düzeyleri ve altı niteliğin sonuç üzerinde etkisi bulunmadığı dikkate alınarak 12 niteliğe indirgenmiştir. Sonrasında 2557 satırdan oluşan veri setinde yer alan 428 boş hücreye, yüksek boyutlu veri setlerinde daha tutarlı sonuçların alındığı belirtilen Ortalama Atama (Mean Substitution) yöntemiyle ilgili sütunun ortalaması alınarak atanmıştır (Köse ve Öztemur, 2014).

2.2. K-Ortalama Kümeleme

Alanyazında en çok kullanılan kümeleme algoritmalarından biri olan K-Ortalama Kümeleme yöntemi, bir veri kümesini k gruba bölmek için kullanılan kümeleme yöntemlerinden biridir (MacQueen, 1967). Yapılan kümeleme işlemiyle elde edilen veri kümelerinin, küme içi benzerliklerinin azami seviyede, kümeler arası benzerliklerin de asgari seviyede olması amaçlanır (Sarıman, 2011). Kısaca K-Ortalama Kümeleme algoritmasında veriler öncelikle k kadar kümeye ayrılır, sonrasında her bir kümenin orta noktası belirlenerek her nesne en yakın orta noktaya sahip kümeye atanır ve orta noktası yeniden hesaplanarak işleme devam edilir (Ismayilov, 2019).

2.3. Naive Bayes

En çok kullanılan sınıflandırma tekniklerinden biri olan Bayes teoremi, olasılıklara dayanan bir teoremdir (Ouda ve Hart, 1973). Naive Bayes algoritmasında, eğitim veri seti üzerinden öğrenme gerçekleşir. Öğrenme sonrasında olasılığı en yüksek olan değer ilgili sınıfa dâhil edilir. Olasılık hesaplaması yapılırken aşağıda verilen formül kullanılır (Chandra ve Gupta, 2011'den aktaran Sağbaş ve Balli, 2016).

$$P(A/B) = \frac{P(B/A) * P(A)}{P(B)}$$

Yukarıdaki formülde:

- P (A/B), B durumunda A'nın gerçekleşme olasılığını,
- P (B/A), A durumunda B'nin gerçekleşme olasılığını,
- P (A), A'nın gerçekleşme olasılığını,
- P (B), B'nin gerçekleşme olasılığını ifade eder (Zobu, 2019).

2.4. K En Yakın Komşu

En yaygın kullanılan gözetimli öğrenme algoritmalarından biri olan ve 1967 yılında Cover ve Hart tarafından geliştirilen K En Yakın Komşu algoritmasında "K", en yakın komşuların sayısını ifade eder. Sınıflandırılacak verinin eğitim setinde yer alan verilere olan uzaklıkları hesaplanarak ilgili sınıfa atama yapılır (Cover ve Hart, 1967).

Çalışmada geçmişe dönük verilerin kullanılarak geleceğe yönelik tahminlerin yapılmasında Naive Bayes algoritması ve K En Yakın Komşu algoritmasının başarımı denenmiştir. Bu bağlamda çıkış değerlerine yönelik tahminler yapılmıştır. Burada önemli olan husus, her bir uygulamanın daha önce yapılan veri temizleme ve kümeleme sonuçları üzerinden yapılmasıdır. Diğer bir ifade ile veriler uzman yardımı ile etiketlenmemiş, K-Ortalama Kümeleme algoritması kullanılarak sırası ile 3,4 ve 5 kümeye ayrılmış veriler, etiketlenmiş veri olarak değerlendirilmiştir. Uygulamalar Spyder Tümleşik Geliştirme Ortamı'nda Python dili ile scikit-learn kütüphanesi kullanılarak yapılmıştır.

Çalışmada kullanılan veri setinde giriş değerleri Tablo 1'de verilmiştir:

Tablo 1: Giriş Değerleri

blh	Sınır tabakası yüksekliği
d2m	Çiğlenme noktası sıcaklığı
fal	Albedo tahmini
sp	Yüzey basıncı
t2m	2 metre sıcaklık
tcc	Toplam bulut örtüsü
tco3	Toplam sütun ozon verisi
tcwv	Toplam sütun su buharı
u10	10 metre U rüzgar bileşeni
v10	10 metre V rüzgar bileşeni
WS	Rüzgar hızı
WD	Rüzgar yönü

Çıkış değerleri ise;

- Üç küme için C1,C2 ve C3,
- Dört küme için C1,C2,C3 ve C4,
- Beş küme için C1,C2,C3,C4 ve C5 olarak isimlendirilmiştir.

3. BULGULAR

Elde edilen veri seti, öncelikle k-ortalama kümeleme yöntemiyle sırasıyla üç, dört ve beş küme olacak şekilde kümelenmiştir. Sonrasında Naive Bayes ve K En Yakın Komşu algoritmasıyla elde edilen üç farklı veri setine yönelik tahminler yapılmıştır.

Uygulamalarda değişik yıllara ait gözlem verileri benzerlik gösterdiğinden gözlem türüne ait değerlerin ortaya koyulması amacıyla 1 Aralık 2012 – 30.04.2013 ve 01.05.2013 – 30.11.2013 tarih aralıkları dikkate alınmıştır. Her uygulamada küme isimleri rastgele atandığından farklı uygulamalarda küme isimleri değişiklik gösterebilmektedir. Örneğin ilk uygulamada 01.12.2012 ve 30.04.2013 tarihleri arası en yüksek değeri alan 130 gün ile C1 iken, ikinci uygulamada en yüksek değeri alan 105 gün ile C3 değeridir.

Uygulama 1, Uygulama 2 ve Uygulama 3'te 3,4 ve 5 kümeye ayrılmış veri setleri üzerinde Naive Bayes algoritmasıyla tahminler yapılmıştır. Uygulama 4 ve Uygulama 5'te de 3 ve 4 kümeye ayrılmış veriler üzerinde tahmin yapılırken K-En Yakın Komşu algoritması kullanılmıştır.

3.1. Uygulama 1

Naive Bayes algoritmasının kullanıldığı ilk uygulamada 1 Aralık 2012 - 30.04.2013 tarihleri arasında 151 günlük gözlem verileri incelendiğinde üç küme için 130 gün C2, 11 gün C1 ce 10 gün için C3 değeri elde edilmiştir. Ek olarak 01.05.2013 – 30.11.2013 tarihleri arasında 214 günlük gözlem verileri için 132 gün C3, 73 gün C1, 9 C2 değeri elde edilmiştir. Ayrıca ilk uygulamada toplam 747 C1, 953 C2 ve 857 C3 değeri elde edilmiştir.

Tablo 2: Uygulama 1 (Kümeler)

K= 3 için (01.12.2012-30.04.2013 / 151 gün)		
C1	C2	C3
130 gün	10 gün	11 gün
K=3 için (01.05.2013-30.11.2013 / 214 gün)		
C1	C2	C3
73 gün	9 gün	132 gün

3.2. Uygulama 2

İkinci uygulamada 1 Aralık 2012 - 30.04.2013 tarihleri arasında 151 günlük gözlem verileri incelendiğinde dört küme için 105 gün C3, 32 gün C1, 9 gün C4 ve 5 gün için C2 değeri elde edilmiştir. Ek olarak 01.05.2013 – 30.11.2013 tarihleri arasında 214 günlük gözlem verileri için 108 gün C2, 71 gün C4, 31 gün C1 ve 4 gün C3 değeri elde edilmiştir.

Tablo 3: Uygulama 2 (Kümeler)

K= 4 için (01.12.2012-30.04.2013 / 151 gün)			
C1	C2	C3	C4
32 gün	5 gün	105 gün	9 gün
K=4 için (01.05.2013-30.11.2013 / 214 gün)			
C1	C2	C3	C4
31 gün	108 gün	4 gün	71 gün

3.3. Uygulama 3

Beş küme için yapılan bu uygulamada 1 Aralık 2012 - 30.04.2013 tarihleri arasında 95 gün C4, 32 gün C2, 14 gün C3, 7 gün C5 ve 1 gün C1 değeri elde edilmiştir. Ek olarak 01.05.2013 – 30.11.2013 tarihleri arasında 214 günlük gözlem verileri için 90 gün C5, 55 gün C1, 49 gün C3, 17 gün C2 ve 3 gün C4 değeri elde edilmiştir.

Tablo 4: Uygulama 3 (Kümeler)

K= 5 için (01.12.2012-30.04.2013 / 151 gün)				
C1	C2	C3	C4	C5
1 gün	32 gün	14 gün	95 gün	7 gün
K=5 için (01.05.2013-30.11.2013 / 214 gün)				
C1	C2	C3	C4	C5
55 gün	17 gün	49 gün	3 gün	90 gün

3.4. Uygulama 4

K En Yakın Komşu algoritmasıyla 1 Aralık 2012 - 30.04.2013 tarihleri arasında 151 günlük gözlem verileri incelendiğinde üç küme için 133 gün C3, 12 gün C2 ve 6 gün için C3 değeri elde edilmiştir. Ek olarak 01.05.2013 – 30.11.2013 tarihleri arasında 214 günlük gözlem verileri için 136 gün C3, 73 gün C1, 5 C2 değeri elde edilmiştir.

Tablo 5: Uygulama 4 (Kümeler)

K= 3 için (01.12.2012-30.04.2013 / 151 gün)		
C1	C2	C3
133 gün	12 gün	6 gün
K=3 için (01.05.2013-30.11.2013 / 214 gün)		
C1	C2	C3
73 gün	5 gün	136 gün

3.5. Uygulama 5

Son uygulamada da 1 Aralık 2012 - 30.04.2013 tarihleri arasında 151 günlük gözlem verileri incelendiğinde dört küme için 102 gün C3, 37 gün C1, 8 gün C4 ve 4 gün C2 elde edilmiştir. Ek olarak 01.05.2013 – 30.11.2013 tarihleri arasında 214 günlük gözlem verileri için 109 gün C2, 73 gün C4, 30 C1 ve 2 gün C3 değeri elde edilmiştir.

Tablo 6: Uygulama 5 (Kümeler)

K= 4 için (01.12.2012-30.04.2013 / 151 gün)			
C1	C2	C3	C4
37 gün	4 gün	102 gün	8 gün
K=4 için (01.05.2013-30.11.2013 / 214 gün)			
C1	C2	C3	C4
30 gün	109 gün	2 gün	73 gün

K-Ortalama Kümeleme algoritması kullanılarak veriler 3, 4 ve 5 kümeye ayrılmıştır. Ancak 5 kümeye ayrılan verilerin dağılımına bakıldığında 01.12.2012-30.04.2013 tarihleri arasında kümelerin birinde sadece bir veri bulunduğu görülmüştür. Her ne kadar k sayısının artması kümeleme hassasiyetini artıracak olsa da bu çalışma kapsamında kümeleme algoritması ön işleme amacıyla veri etiketleme için kullanıldığından analizler 3 ve 4 küme üzerinden yapılmıştır.

Astronomi ile ilgili gözlem verilerinin etiketlenmesi ve makine öğrenmesi algoritmaları için hazır hale getirilmesi zahmetli bir süreçtir. Çalışmada ele alınan çıkış değerlerinin (C1,C2,C3.) gözlem türünü ifade ettiği göz önünde bulundurularak gözlemcilerin yönlendirilmesinde bu değerler dikkate alınabilir.

4. TARTIŞMA VE SONUÇ

Astronomik Gözlem Kalitesi, sıcaklık, nem, rüzgâr, bulut, görüş gibi kriterlere göre değişiklik göstermektedir. Astronomik gözlem kalitesiyle birlikte gözlem türü de farklılık göstermektedir. Ayrıca gözlem türü, gözlem yapılacak gök cisimlerine göre değişiklik gösterebilmektedir. Günümüzde yapılan gözlemler uzman yönlendirmesi doğrultusunda yapılmaktadır. Bu çalışmada bu

yönlendirmelerde makine öğrenmesi algoritmalarının uygulanabilirliğini ortaya koymak amaçlanmaktadır.

Uygulamaların sonuçları analiz edildiğinde kümelerin genellikle aralık, nisan ve temmuz aylarında değişiklik gösterdiği görülmüştür. Bütün uygulamalar birlikte ele alındığında aralık ve mayıs ayları arası ilk uygulamada C2, ikinci uygulamada C3 ve son uygulamada C4 değerleri kesişim göstermektedir. Ek olarak mayıs ve aralık aylarında da ilk uygulamada C3, ikinci uygulamada C2 ve son uygulamada C5 değerlerinin kesişim gösterdiği görülmüştür.

Yapılan ilk uygulamada aralık, ocak, şubat ve mart aylarında C2, nisan, mayıs haziran için C3, temmuz ağustos, eylül, ekim ve kasım aylarında C1 değerinin baskın olduğu görülmüştür. Bu durum göz önünde bulundurularak nisan ayından itibaren gözlem türünün değişiklik gösterdiği söylenebilir.

İkinci uygulamada aralık, ocak, şubat, mart ve nisan aylarında C3 değeri, mayıs, haziran, temmuz ve ağustos aylarında C2, eylül ayında C4, ekim ve kasım aylarında da C1 değerinin baskın olduğu görülmüştür. İlk uygulamayla benzer şekilde nisan ayında gözlem türünün değişiklik gösterdiği görülmektedir. K En Yakın Komşu algoritmasıyla ilk iki uygulamaya benzer sonuçların elde edildiği görülmektedir.

Yapılan uygulamalarda Naive Bayes algoritması başarı oranları üç küme için 0.97, dört küme için 0.92, K En Yakın Komşu algoritması başarı oranları üç küme için 0.95, dört küme için 0.91 olarak hesaplanmıştır. Naive Bayes ve K En Yakın Komşu algoritmaları her ne kadar K Ortalama Kümeleme algoritmasından elde edilen kümeleme sonuçlarına göre uygulansa da astronomik gözlem verilerinin makine öğrenmesi algoritmaları ile tahmin edilebileceği ve bu alanda buna yönelik bir veri seti oluşturulması ihtiyacının hâsıl olduğu söylenebilir.

Han, Ding, Zhang ve Zhao (2016) çalışmalarında fotometrik kırmızıya kayma değerlerini KNN ve Destek Vektör Makineleri'nin entegrasyonuyla tahmin etmeyi amaçlamışlardır. Başka bir çalışmada da düşük kırmızıya kaymalı galaksilerin renklerinden belirli bir yıldız oluşum oranının tahmin edilmesi amaçlanarak K En Yakın Komşu algoritması kullanılmıştır (Stensbo-Smidt vd., 2013). Bu çalışmada farklı olarak astronomik gözlem kalitesinin tahmin edilmesinde K En Yakın Komşu ve Naive Bayes algoritmaları ayrı ayrı uygulanarak sonuçlar karşılaştırılmıştır. İlerleyen çalışmalarda astronomik gözlem kalitesinin tahmin edilmesinde çeşitli sınıflandırma algoritmalarının kullanımı ve bu algoritmaların başarımlarının karşılaştırılması gözlemevlerindeki akıllı uygulamaların geliştirilmesi noktasında katkı sağlayacaktır. Ayrıca geleceğe yönelik gözlem kalitesi tahminiyle araştırmacıların yönlendirilmesi ve gözlemevlerinin etkili, verimli kullanılması sağlanacaktır.

KAYNAKÇA

Boffin, H. M., Jerabkova, T., Merand, A., & Stoehr, F. (2019). Report on the ESO Workshop" Artificial Intelligence in Astronomy". Msngr, 178, 61-63.

Chandra, B., ve GUPTA, M. (2011). "Robust approach for estimating probabilities in Naïve–Bayes Classifier for gene expression data". Expert Systems with Applications, 38(3), 1293-1298.

Chugh, J. (2018). Types of Machine Learning and Top 10 Algorithms Everyone Should Know. 01.09.2020 tarihinde https://blogs.oracle.com/ai/types-of-machine-learning-and-top-10-algorithmseveryone-should-know-v2 adresinden alınmıştır.

Cover, T., & Hart, P. (1967). "Nearest neighbor pattern classification". IEEE transactions on information theory, 13(1), 21-27.

Esenoğlu H.H., Gülseçen H. (2019). SRG X Işın Gözlemevi: Büyük Veri. 05.09.2020 tarihinde https://tug.tubitak.gov.tr/sites/images/tug/srg-sunum-hasan-esenoglu.pdf adresinden alınmıştır.

Han B., Ding H.P., Zhang Y.X., & Zhao Y.H. (2016). "Photometric redshift estimation for quasars by integration of KNN and SVM". Research in Astronomy and Astrophysics, 16(5), 074.

Ismailov, E. Büyük veri kavramı, K-means, Optıcs ve Cure algoritmalarının uygulanması ve performans ölçümü. Yüksek Lisans Tezi. Ege Üniversitesi, Fen Bilimleri Enstitüsü, İzmir.

Köse, İ. A., & Öztemur, B. (2014). "Kayıp veri ele alma yöntemlerinin t-testi ve ANOVA parametreleri üzerine etkisinin incelenmesi". Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi.

Macqueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. Proceedings of the Fifth Symposium on Math, Statistics, and Probability, 281–297. Berkeley, CA: University of California Press.

Ouda, R., & Hart, P. (1973). Pattern Classification and Scene AnalysisWiley. New York.

Sağbaş, E. A., & Ballı, S. (2016). İnsan Aktivitelerinin Bilek Hareketleri Kullanılarak Tahmin Edilmesi. Akıllı Teknoloji & Akıllı Yönetim, 52.

Sariman, G. (2011). "Veri madenciliğinde kümeleme teknikleri üzerine bir çalışma: k-means ve k-medoids kümeleme algoritmalarının karşılaştırılması". Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 15(3), 192-202.

Stensbo-Smidt, K., Igel, C., Zirm, A., & Pedersen, K. S. (2013). Nearest neighbour regression outperforms model-based prediction of specific star formation rate. In 2013 IEEE international conference on big data, 141-144.

Unat, Y. (2001). İlk Çağlardan Günümüze Astronomi Tarihi. Ankara. Nobel Yayınları.

Zhang, Y. X., Luo, A. L., & Zhao, Y. H. (2004). An automated classification algorithm for multiwavelength data. In Optimizing Scientific Return for Astronomy through Information Technologies (Vol. 5493, pp. 483-490). International Society for Optics and Photonics.

Zobu, (2019). Naive Bayes: Teorisi ve Python Uygulaması. 02.09.2020 tarihinde https://medium.com/kaveai/naive-bayes-ve-uygulamalar%C4%B1-d7d5a56c689b adresinden alınmıştır.