

Different Mathematical Models for Entropy in Information Theory

F. Bulut^{1*}

Abstract: Entropy is a metric that measures statically irregularity and randomness in a given certain dataset. The calculated entropy value has a positive real number starting from 0. The entropy value is also directly proportional to the increase of the irregularity in the dataset. In numerous fields including statistics, mathematics, and information theory it has been used widely with an aggregated formula of sigma symbol, logarithm and probability in mathematics. In this study, from different perspectives some new formulas and inferences are proposed that gives the same entropy result. The same entropi results, on the experiments, have been taken over the UCI benchmark datasets.

Keywords: Entropy, Information theory, Mathematical model

Bilgi Kuramındaki Entropi Kavramıyla İlgili Farklı Matematiksel Modeller

Özet: Entropi, belirli bir veri kümesindeki düzensizliği ve dağınıklığı ölçen bir istatistiksel metriktir. Hesaplanan entropi değeri 0'dan başlayan pozitif bir reel değere sahiptir. Veri kümesindeki düzensizliğin artması ile doğru orantılı olarak entropi değeri de artmaktadır. İstatistik, matematik ve bilgi kuramı dâhil birçok alanda yaygın bir şekilde kullanılan entropi formülünde matematikteki toplam sembolü, logaritma ve olasılık hesaplamaları vardır. Bu çalışmada entropi formülüne farklı açılardan bakarak aynı sonucu veren yeni formüller ve çıkarımlar ile farklı matematiksel modeller türetilmiştir. UCI veri setleri üzerinde yapılan deneysel uygulamalarda da aynı entropi değerleri elde edilmiştir.

Anahtar Kelimeler: Entropi, Bilgi kuramı, Matematiksel model

1. Giriş

Entropi termodinamik disiplinde, ısı kaynağının belirli bir sıcaklığında, ısı kaynağı ile yapılan ısı alışverişi olarak tanımlanmaktadır. Bilgi kuramında ise kabaca bir sistemindeki kesinsizliğin sayısal değeridir. Bilgi teorisi açısından bakıldığında, entropi durumu işlevi basitçe sistemin tam mikro durumunu belirtmek için gerekli olan bilgi miktarıdır. Bilgi teorisinde entropi, bilgi miktarının ölçüsüdür (Balian, 2004).

Bilgi teorisi olarak da isimlendirilen bilgi kuramı, verinin ya da bilginin niceliğeleştirilmesi (sayısallaştırılması) ile ilgili uygulamalı bilgisayar mühendisliği dalıdır. Bilgi kuramı ilk olarak 1948'da Claude E. Shannon tarafından güvenli

şekilde verinin bilgisayar ortamında depolanması, sıkıştırılması ve bilgisayarlar arası iletilmesi gibi aşamalarda sinyal işleme işlemlerinin kısıtlamalarını bulmak için geliştirilmiştir (Shannon, 1949). Bilgi kuramı, istatistik (Salkind, 2016), ekonomi (Rosser, 2016), matematik (Dehmer, 2016), bilgisayar bilimi (Anderson, 2008), ekoloji (Özkan, 2016), doğa bilimi (Özkan, 2017), fizik (Toggerson, 2017) ve elektrik-elektronik mühendisliği (Zhou, 2006) gibi bilim dallarında sıklıkla kullanılır. Bilgisayar ortamında verinin işlenmesi ve depolanması gibi işlemlerde entropi değeri önemli bir araçtır. Entropi, bir rastgele değişkenin değerini tahmin ederken belirsizliği sayısallaştırır. Örneğin, bir bozuk para ile oynanan yazı tura oyununda sonuç eşit olasılıklı ya yazı ya da turadır. Fakat zar atma oyununda ise

¹Istanbul Rumeli Üniversitesi, Mühendislik ve Mimarlık Fakültesi, Bilgisayar Mühendisliği, 34570, İstanbul, Türkiye

*Corresponding author (İletişim yazarı): faruk.bulut@rumeli.edu.tr

Citation (Atıf): Bulut, F. (2017). Bilgi Kuramındaki Entropi Kavramıyla İlgili Farklı Matematiksel Modeller. Bilge International Journal of Science and Technology Research, 1 (2): 167-174.

eşit olasılıklı altı sonuç vardır. Bu nedenle yazı tura oyunu zar atmaya göre daha düşük entropiye sahiptir.

Bilgi teorisinin temel uygulamaları ile ilgili örnek verilecek olursa kayıpsız veri sıkıştırma (RAR dosyaları), kayıplı veri sıkıştırma (JPEG ve MP3 dosyaları) gösterilebilir (Bulut, 2016a). Ayrıca Yapay Öğrenme alanındaki sınıflandırma işlemlerinde kullanılan Karar Ağaçlarının oluşturulmasında entropi kavramı kullanılmaktadır (Mu vd., 2017).

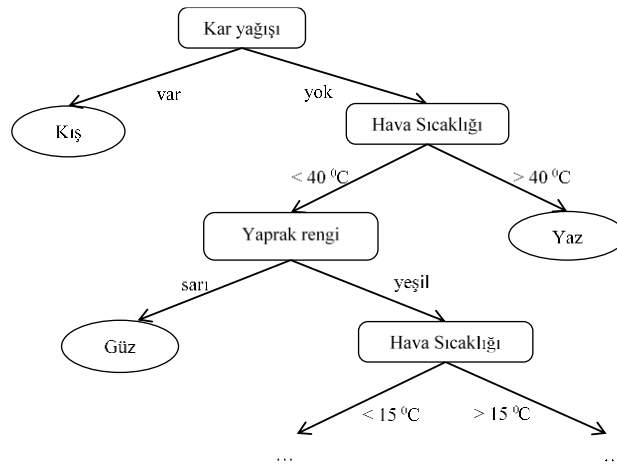
Bilimsel çalışmalarda kullanılan Entropi (belirsizlik), veri kümesi içerisinde bulunan öğelerin düzensizliğini reel sayılarla ifade eden bir ölçü birimidir. Negatif bir sayı ile ifade edilemeyen entropi, sıfır ile başlayan bir değere sahiptir ve sayı değeri arttıkça, sistemdeki düzensizliği o kadar fazla olduğunu gösterir. Eğer bir veri kümesi tam düzenli bir dağılıma sahip ise

entropi değeri de sıfır olabilir. Farklı entropi formülleri bulunsa da bilgi kuramındaki Shannon Entropi yasası rassal değişken olarak genelde Formül 1 ile ifade edilir:

$$E = - \sum_{i=1}^N P_i \log_2 P_i \quad (1)$$

Claude E. Shannon tarafından tanımlan 1 numaralı formülde P_i veri kümesindeki i . veri sınıfının tüm sınıf içerisindeki bulunma olasılığını ifade etmektedir (Gray, 1990).

Entropi, bilgisayar bilim dalında daha çok veri analizi, veri madenciliği, makine öğrenmesi ve sınıflandırma, yapay tahmin mekanizmalarında kullanılmaktadır.



Şekil 1. Örnek Karar Ağacı

Özellikle bir veri setinde veri gruplarının bilgi kazancını (*Information Gain*) hesaplanmasında entropi kullanılmaktadır (Srinivas ve Kumar, 2013).

Bilgi kazancı, karar ağaçlarının oluşturulmasında kullanılan nemli bir kriterdir. Veri seti içerisinde karara en çok etki eden özneliğin sayısal değerini bulmak için kullanılır. Örneğin yapay bir karar destek sistemi, tabiat olaylarına bakarak hangi mevsimde bulunulduğunu tespit etmek istesin. Yağmur, kar, fırtına, ağaçtaki yaprakların rengi, çiçeklerin açıp açmadığı, hava sıcaklığı, nem ve basınç gibi bir dizi öznelik ve bu özneliklerin

farklı mevsimlerde alabilecekleri değerler bir veri kümesinde aktarılınsın. Karar destek sistemi yani karar ağacı da bu veri kümesi ile de eğitilsin. Bilindiği üzere kar sadece kış aylarında yağar, başka mevsimlerde yağışı görülmez. Bu sebepten dolayı eğer havada kar yağışı var ise mevsim mutlaka kıştır denilebilir. Veri setinde bulunan sayısal bilgilere bakarak “kar yağışı” özneliğinin bilgi kazancı mutlaka çok yüksek çıkacak ve karar ağacının inşa edilmesinde ilk ölçüt kar yağışı olarak hesaplanacaktır.

Karar ağaçları, yapay karar destek sistemleri içerisinde sıklıkla kullanılmaktadır. Karar ağacı

sınıflandırıcısında, Şekil 2’de görüldüğü üzere ters bir ağaç yapısı kurulur. Gövdeden yani ana düğümden yapraklara doğru giden düğümler üzerinde karşılaştırmalar ve yapraklar üzerinde de hesaplanan sınıf türleri yer almaktadır. Ağacın inşa edilmesi sırasında, her bir özneliğin bilgi kazancı değerine bakılarak bölütleme işlemi yapılır. Özyinelemeli (*recursion*) olarak önemini yitirene kadar bu işlem sürekli olarak tekrar eder (Bulut, 2016b). Özyineleme, öztekrarlı olarak da ifade edilebilir ve bilgisayar biliminde ilgili fonksiyonun ve sistemin kendini gerekli parametrelerle tekrar etmesi anlamına gelmektedir. Şekil 1’de de görüldüğü üzere örnek bir karar ağacı yapısı verilmektedir.

Bilgi kazancının hesaplanması, entropi değerinin çarpmaya göre tersi alınarak yapılır. Ayrıca yapay karar destek sistemlerinde karar ağaçları (*Decision Tree*) algoritması en yaygın kullanım alanına sahiptir (Tosun, 2015). Özyinelemeli diğer bir deyişle öztekrarlı bir şekilde oluşturulan bu ağaç yapısında ağacın düğümlerinde bulunan sorguların seviyesini belirlemede de entropi kullanılmaktadır.

Ayrıca görüntü işleme alanında kullanılan video özetleme (*video summarization*) (Cernekova vd., 2006) ve bilgisayar ağları ile ilgili veri transferi ile topoloji belirleme (El Gamal ve Kim, 2011) gibi alanlarda da entropi ölçütü kullanılmaktadır. David R. Anderson isimli bir araştırmacı, Information Theory and Entropy isimli kitap çalışmasında entropinin yaygın kullanım alanlarını sıralamış ve detaylı bir inceleme yapmıştır (Anderson, 2008).

Açıklandığı üzere Shannon entropisi, bilgi teorisinde ve termodinamiğin uygulamalarını bulan geniş ve genel bir kavramdır. Bilgisayar ortamında saklanan veya iletilen bir mesajdaki verinin miktarını ve karmaşıklığını incelemek üzere tasarlanmıştır.

İletilen veya kaydedilen verilerdeki bu olasılıklar, verideki ortalama bilgi miktarının bir ölçütüdür. Her ne kadar bu olasılıklar farklı da olsa Shannon entropisi verideki “bit” karmaşasını incelemek için kullanılır. Bit, bilgisayar ortamındaki en küçük bilgi birimidir ve sadece 1-0 ile ifade edilebilir. ASCII kodlamasında bir karakter 8 bit ile ifade edilir. Doğal olarak 1 ve 0’lardan oluşan 8 basamaklı kaç farklı dizilim gerçekleştirilir ya da kaç farklı karakter kodlanabilir sorusunun cevabı $2^8=256$ olacaktır (Çakır, 2015).

Shannon Entropi hesaplamasına bir örnek olarak şöyle bir senaryo verebiliriz. Bilgisayarlar arası mesajlaşma işlemi iki kişiden biri diğerine bir mesaj göndersin. Gönderilen mesaj içerisinde farklı bulunma olasılıklarında sadece {a,b,c,d,e} harfleri bulunsun. Bu durumda mesajların 0 ve 1’leri ile en az bit kullanarak kodlanabilmesi için entropi hesaplamasına ihtiyaç vardır. Harflerin eşit bulunma olasılığında olduğu varsayıldığında,

$$\text{Bit sayısı} = - \sum_{i=1}^N P_i \log_2 P_i$$

$$\text{Bit sayısı} = - \sum_{i=1}^5 \frac{1}{5} \log_2 \frac{1}{5}$$

$$\text{Bit sayısı} \cong 2.323$$

P_i her bir karakterin metin içerisindeki olasılığıdır. Çıkan 2.323 değeri bir üst tam sayıya yuvarlandığında 3 değeri elde edilir. Demek ki sadece 3 bit ile elde bulunan harfler kodlanabilecektir. Zaten 3 bit ile ($2^3 = 8$ olacağı için) en fazla 8 karakter kodlanabilmektedir. Harflerin metin içerisinde bulunma olasılıkları eşit olmayıp farklı olması durumunda ise daha farklı sonuçlar elde edilebilir. Örneğin harflerin bulunma olasılıkları {0.05, 0.05, 0.05, 0.05, 0.8} şeklinde olduğunda entropi sonucu yaklaşık olarak 1.12 çıkar. Bu durumda kullanılacak en az bit sayısı 2 olarak hesaplanır fakat bu işlemlerde en kötü senaryo ele alındığı için kodlama için gerekli olan bit sayısı eşit dağılımlı durumlarda hesaplanabilmektedir.

Harflerin bitler ile kodlanması ile ilgili örnek diziliş Çizelge 1’de verilmiştir. Bir metin içerisinde bulunan N adet farklı karakter için en az ne kadar bit’e ihtiyaç duyulduğu bilgi kuramındaki entropi formülü ile hesaplanır. Görüldüğü üzere 8 bit ASCII ile kodlanması gereken her bir karakter 3’er bit ile kodlanarak büyük bir tasarruf sağlanmıştır.

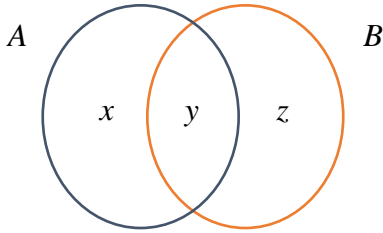
Çizelge 1. Bit kodlama

Karakter	Bit Kodlama
a	000
b	001
c	010
d	011
e	100

Makale toplam beş ana bölümden oluşmaktadır. Takip eden ikinci bölümünde entropinin detaylı açıklamasına ve önerilen yönteme, üçüncü bölümde elde edilen çıkarımlara ve son bölümdeki değerlendirmelerden önce veri setleri üzerindeki deneysel uygulamalara yer verilmiştir.

2. Yöntem

Çalışmada kullanılan logaritma, toplam (sigma) sembolü ve olasılıkla ilgili bazı konuların kısa tanımlarını açıklamak uygun olacaktır.



Şekil 2. İki veri kümesi ve dağılım sayıları

Şekil 2’de kesişimli iki farklı küme verilmiştir. A kümesi içerisinde olan elemanlar x ve y ile; B içerisinde olan elemanlar y ve z ile gösterilsin. x , y ve z burada birer eleman olmayıp küme elemanlarının sayısını göstermektedir. y ise $A \cap B$ içerisinde bulunan elemanları temsil etsin. Bu durumda B kümesi içerisinde A ’nın olma olasılığını araştırmaktadır. P yani olasılık (Probability) Formül 2 ile hesaplanır.

$$P(A|B) = \frac{y}{y+z} \quad (2)$$

$a_r + a_{r+1} + a_{r+2} + \dots + a_n$ toplamını bulan Toplam Sembolü ise Formül 3 ile hesaplanır.

$$\sum_{k=r}^n a_k = a_r + a_{r+1} + a_{r+2} + \dots + a_n \quad (3)$$

Bu gösterimde kullanılan \sum harfine toplam sembolü denir. a ’nın r ’den başlayıp n ’e kadar olan toplamını toplam sembolü sayesinde kısa yolla hesaplanabilir.

Logaritma ise $x=a^y$ şeklinde yazılan bir ifadeyi $\log_a^x=y$ şeklinde yazma amacıyla kullanılır.

Entropi’yi E ile gösterirsek N elemanlı bir küme

içerisindeki belirsizlik 4 numaralı formül ile hesaplanır.

$$E = - \sum_{i=1}^N P_i \log_2 P_i \quad (4)$$

$$= -P_1 \log_2 P_1 - P_2 \log_2 P_2 - \dots - P_N \log_2 P_N$$

P_i burada A kümesi içerisinde bulunan i . elemanın daha doğrusu çeşidin bulunma ihtimalidir.

Örneğin bir torbada bulunan 4 bilye özdeş ve mavi renkli ise bu kümenin entropisi yani düzensizliği hesaplanmak istenirse:

Eleman sayısı $N = 4$ ve $P(\text{mavi}|A) = \frac{4}{4} = 1$ olsun. Burada $P(\text{mavi}|A) = P(\text{Mavi})$ olarak da yazılabilir. İki gösterim şekli de aynı anlama gelmektedir.

$$E = - \sum_{i=1}^N P_i \log_2 P_i$$

$$E = - \left(\frac{4}{4} \log_2 \frac{4}{4} \right)$$

$$E = 0$$

Entropi’nin 0 çıkması düzensizliğin olmadığını göstermektedir. Diğer bir taraftan bir torbada bulunan mavi ve sarı renklerdeki ikişer bilyenin oluşturduğu bu kümenin $S(A) = N = 2$ ve $P(\text{mavi}) = P(\text{sarı}) = \frac{1}{2}$ şeklinde olsun.

Düzensizlik hesaplaması:

$$E = - \sum_{i=1}^N P_i \log_2 P_i$$

$$E = - \left(\frac{1}{2} \log_2 \frac{1}{2} \right) - \left(\frac{1}{2} \log_2 \frac{1}{2} \right)$$

$$E = 1$$

E ’nin 1 çıkması yani 0’dan uzaklaşması entropinin düzensiz olduğunu gösterir. Bu örnek 10 mavi ve 10 sarı ile de yapılsa aynı sonuç bulunur. Buna göre çeşit ve oranların eşit olduğu durumlarda entropiler de eşit çıkmaktadır.

Farklı bir senaryoda bir torbada bulunan 4 bilye $A = \{\text{sarı, sarı, mavi, kırmızı}\}$ ve $S(A)=N = 4$, $P(\text{sarı}) = \frac{2}{4}$, $P(\text{mavi}) = P(\text{kırmızı}) = \frac{1}{4}$ şeklinde olsun. Bu durumda hesaplama şu şekilde olacaktır:

$$E = - \sum_{i=1}^N P_i \log_2 P_i$$

$$= -\frac{2}{4} \log_2 \frac{2}{4} - \frac{1}{4} \log_2 \frac{1}{4}$$

$$- \frac{1}{4} \log_2 \frac{1}{4} = -\left(-\frac{1}{2} - \frac{2}{4} - \frac{2}{4}\right)$$

$$= 1,5$$

Örnekte de görüldüğü gibi çeşit sayısının artmasıyla doğru orantılı olarak entropi 0'dan uzaklaşmıştır diğer bir deyişle dağınıklık artmıştır.

3. Entropi Formülünden Elde Edilen Çıkarımlar

Entropi formülünü kullanarak dağınıklık hesaplamalarına farklı perspektiflerden bakarak farklı çıkarımlar yapılmak istensin.

3.1. Birinci çıkarım

$$E = - \sum_{i=1}^N P_i \log_2 P_i$$

$$E = -P_1 \log_2 P_1 - P_2 \log_2 P_2 - \dots - P_N \log_2 P_N$$

$$E = -\log_2 P_1^{P_1} - \log_2 P_2^{P_2} - \dots - \log_2 P_N^{P_N}$$

$$E = -(\log_2 P_1^{P_1} + \log_2 P_2^{P_2} + \dots + \log_2 P_N^{P_N})$$

$$E = -(\log_2 P_1^{P_1} \cdot P_2^{P_2} \dots P_N^{P_N})$$

$$E = -\left(\log_2 \prod_{i=1}^N P_i^{P_i}\right)$$

$$E_1 = - \sum_{i=1}^N P_i \log_2 P_i$$

$$= -\left(\log_2 \prod_{i=1}^N P_i^{P_i}\right) \quad (5)$$

Toplam sembolünü içeren entropi formülü Formül 5'de görüldüğü üzere \prod sembolüyle de hesaplanabilmektedir. Logaritmanın açılımından

faydalanarak toplam sembollü bir formül çarpım sembollü bir formüle dönüştürülmüş oldu.

3.2. İkinci çıkarım

$$E_1 = -\left(\log_2 \prod_{i=1}^N P_i^{P_i}\right) \quad (6)$$

2'nin negatif entropi kuvveti görüldüğü üzere tüm olasılıkların kendi kuvvetlerinin birbirleriyle çarpımına eşit çıkmaktadır. Bu durumda 7 numaralı formül elde edilmiş olur.

$$2^{-E_2} = \prod_{i=1}^N P_i^{P_i} \quad (7)$$

Sigma sembolü ile ifade edilen düzensizlik ölçütü tamamıyla pi sembolü ile ifade edilebilmiştir.

3.3. Üçüncü çıkarım

8 numaralı formüle ulaşabilmek öncesinde bir takım işlemler yapılmıştır.

$$E = -\left(\log_2 \left(\frac{x_1}{N}\right)^{\frac{x_1}{N}} \cdot \left(\frac{x_2}{N}\right)^{\frac{x_2}{N}} \dots \dots \left(\frac{x_n}{N}\right)^{\frac{x_n}{N}}\right)$$

$$E = -\log_2 \left(\frac{1}{N} \cdot X_1^{\frac{x_1}{N}} \cdot \frac{1}{N} \cdot X_2^{\frac{x_2}{N}} \dots \frac{1}{N} \cdot X^{\frac{x_n}{N}}\right)$$

$$E = -\log_2 \left(\left(\frac{1}{N}\right)^{\frac{x_1+x_2+\dots+x_n}{N}} \cdot \prod_{i=1}^N x_i^{P_i}\right)$$

Elde edilen çıkarım 8 numaralı formülde görüldüğü gibidir.

$$E_3 = -\log_2 \left(N^{-ort} \cdot \prod_{i=1}^N x_i^{P_i}\right) \quad (8)$$

Üç farklı çıkarım da görüldüğü üzere klasik entropi formülü ile elde edilmiştir. Toplam sembolü içeren klasik formülden çarpım sembolü içeren formüller türetilmiştir.

3.4. Basit örnek senaryo

Bir örneği üç adet entropi formülü ile açıklayarak aynı sonucu elde edip edilemeyeceği kontrol

edilmek istensin. 2 elemanlı kümemizde bir mavi ve sarı bilyenin olduğu düşünölsün. Bu durumda herhangi bir rengin seçilme ihtimalinin entropisi hesaplanmak istensin.

Bu durumda ilk önce klasik entropi formölüyle hesaplama yapmak gerekir.

$$N = 2 \text{ ve } P(\text{Mavi}) = P(\text{Sarı}) = \frac{1}{2}$$

$$E = - \sum_{i=1}^N P_i \log_2 P_i$$

$$E = - \left(\frac{1}{2} \log_2 \frac{1}{2} \right) - \left(\frac{1}{2} \log_2 \frac{1}{2} \right)$$

$$E = 1$$

1. Çıkarıma göre:

$$E_1 = - \left(\log_2 \prod_{i=1}^N P_i^{P_i} \right) = - \left(\log_2 \prod_{i=1}^2 \left(\frac{1}{2} \right)^{\frac{1}{2}} \right) = 1$$

2. Çıkarıma göre:

$$\prod_{i=1}^N P_i^{P_i} = 2^{-E_2}$$

$$\prod_{i=1}^2 \left(\frac{1}{2} \right)^{\frac{1}{2}} = 2^{-E_2}$$

$$2^{-E_2} = \frac{1}{2} \text{ ve}$$

$$E_2 = 1 \text{ dir.}$$

3. Çıkarıma göre:

$$E_3 = - \log_2 \left(N^{-\text{ort}} \cdot \prod_{i=1}^N x_i^{P_i} \right)$$

$$E_3 = - \log_2 \left(2^{-1} \cdot \prod_{i=1}^2 \left(\frac{1}{2} \right)^{\frac{1}{2}} \right)$$

$$E_3 = 1$$

Göröldüğü üzere elde edilen çıkarımlar aynı senaryo üzerinde kullanılmış ve yine aynı sonuçlar elde edilmiştir.

4. Deneysel Uygulama

California Üniversitesi, veri madenciliği ve makine öğrenmesi algoritmalarının uygulanabileceği ve istatistik alanındaki yeni teorilerin test edilebileceği ortak kullanıma açık UCI veri setleri adında bir veri ambarını kullanıma açmıştır (Lichman, 2018). UCI veri setleri gerçek hayattan alınan birçok türdeki verilerle oluşturulmuştur. Veri setlerindeki sınıf türleri genel olarak dengeli bir dağılıma sahiptir.

Deneysel çalışmalar için yüzlerce veri seti içerisinden birbirinden farklı istatistiksel dağılımlara sahip 10 adet veri seti seçilmiştir. *autos*, *balance-scale*, *coll0*, *diabetes*, *glass*, *primary-tumor*, *segment*, *splice*, *vehicle* ve *vowel* veri setleri içerisinde bulunan toplam örnek sayısı, sınıf tür sayısı ve her bir sınıf türü içerisinde bulunan örnek sayısı Çizelge 2'de verilmektedir. Her bir sınıf türünün örnekleri toplamı, $\sum_{i=1}^N n_i$ şeklinde ifade edilerek toplam örnek sayısı verilmektedir.

Çizelge 2. UCI Veri setlerindeki örnek sayıları ve her bir sınıf türünün eleman sayıları

Veri seti	N Örnek Sayısı	Sınıf Etiketli Sayısı	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9	n_{10}	n_{11}
<i>autos</i>	202	5	22	67	54	32	27						
<i>balance-scale</i>	625	3	288	49	288								
<i>col10</i>	2019	10	603	112	87	140	244	314	294	76	50	99	
<i>diabetes</i>	768	2	500	268									
<i>glass</i>	205	5	70	76	17	13	29						
<i>primary-tumor</i>	302	11	84	20	14	39	14	28	16	24	10	29	24
<i>segment</i>	2310	7	330	330	330	330	330	330	330				
<i>splice</i>	3190	3	767	768	1655								
<i>vehicle</i>	846	4	212	217	218	199							
<i>vowel</i>	990	11	90	90	90	90	90	90	90	90	90	90	90

Çizelge 3. Entropi sonuçları

Veri seti	E, E_1, E_2 ve E_3 formüllerine göre elde edilen entropi değerleri
<i>autos</i>	2.194442
<i>balance-scale</i>	1.318116
<i>col10</i>	2.9289
<i>diabetes</i>	0.933134
<i>glass</i>	2.0094
<i>primary-tumor</i>	3.175773
<i>segment</i>	2.807355
<i>splice</i>	1.48017
<i>vehicle</i>	1.999067
<i>vowel</i>	3.459432

Orijinal entropi formülü E ve türetilen entropi formülleri E_1, E_2 ve E_3 formüllerine göre elde edilen sonuçlar Çizelge 3'de görülmektedir. Her birinde hesaplanan E değerinin, E_1, E_2 ve E_3 için de aynı olduğu tespit edilmiştir.

5. Tartışma ve Sonuçlar

Rastgeleliği, belirsizliği ve beklenmeyen durumun ortaya çıkma olasılığını gösteren entropi üzerine farklı bir çalışma gerçekleştirilmiştir. Bu makalede belirli bir veri kümesi üzerinde düzensizliği bulma işlemi yani entropi hesaplaması farklı matematiksel modellerle yapılmış ve aynı sonuçlar

elde edilmiştir. Entropi formülüne farklı bir açıdan bakan bu çalışmada türetilen yeni formüller ile aynı senaryo üzerinde ve değişik veri setleri üzerinde yapılan hesaplamalarda aynı sonuçlar elde edilmiştir. Logaritmanın açılımındaki özellikten dolayı toplam sembolü ile hesaplanan entropi, çarpım sembolü de kullanılarak farklı şekillerde hesaplanabileceği gösterilmiştir.

Kaynaklar

- Anderson, D. R. (2008). Information Theory and Entropy (pp. 51-82). Springer, ISBN-13: 978-1441979698, New York.
- Balian, R. (2004). Entropy, a protean concept. In Poincaré Seminar 2003 (pp. 119-144). Birkhäuser Basel.
- Bulut, F. (2016a). Huffman Algoritmasıyla Kayıpsız Hızlı Metin Sıkıştırma. El-Cezeri Journal of Science and Engineering, 3(2), 287-296.
- Bulut, F. (2016b). AdaBoost ile Kalp Krizi Risk Tespiti, Celal Bayar Üniversitesi Fen Bilimleri Dergi, 12(3), 459-472.
- Cernekova, Z., Pitas, I., Nikou, C. (2006). Information theory-based shot cut/fade detection and video summarization. IEEE Transactions on circuits and systems for video technology, 16(1), 82-91.
- Çakır, S. (2015). Bütünleşik Bulanık Shannon Entropi-Bulanık Veri Zarflama Analizi Yöntemiyle Teknoloji Firmalarında

- Etkinlik Ölçümü. Karadeniz Teknik Üniversitesi, Doktora Tezi.
- Dehmer, M., Emmert-Streib, F., Chen, Z., Li, X., Shi, Y. (Eds.). (2016). *Mathematical Foundations and Applications of Graph Entropy* (Vol. 6). John Wiley & Sons.
- El Gamal, A., Kim, Y. H. (2011). *Network information theory*. Cambridge University press.
- Gray R. M., (1990). *Entropy and information*. In *Entropy and Information Theory*, Publisher: Springer, (pp. 21-55) New York, ISBN-13: 978-1441979698.
- Lichman, M. (2018). UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science.
- Mu, Y., Liu, X., Yang, Z., and Liu, X. (2017). A parallel C4.5 decision tree algorithm based on MapReduce. *Concurrency and Computation: Practice and Experience*, 29(8).
- Özkan, K. (2016). Application of Information Theory for an Entropic Gradient of Ecological Sites, *Entropy*, 18(10), 340.
- Özkan, K. (2017). *Doğanın Kuantum Analizi*. Süleyman Demirel Üniversitesi, Orman Fakültesi Yayın No: 102, ISBN: 978-605-9454-08-7, Isparta, 148 s.
- Rosser, J. B. (2016). Entropy and econophysics. *The European Physical Journal Special Topics*, 17(225), 3091-3104.
- Salkind, N. J. (2016). *Statistics for people who (think they) hate statistics*. Sage Publications.
- Shannon, C. E. (1949). Communication theory of secrecy systems. *Bell Labs Technical Journal*, 28(4), 656-715.
- Srinivas, D. S., Kumar, M. A. (2013). Attribute and Information Gain Based Feature Selection Technique For Cluster Ensemble: Hybrid Majority Voting Based Variable Importance Measure. *IJITR*, 1(6), 607-610.
- Toggerson, B., Nguyen, D. (2017). *Physics 131: Forces, Energy and Entropy*, Publisher: University of Massachusetts Amherst Libraries, Amherst, Massachusetts.
- Tosun, S. (2015). Sınıflandırmada yapay sinir ağları ve karar ağaçları karşılaştırması: Öğrenci başarıları üzerine bir uygulama (Doctoral dissertation, Fen Bilimleri Enstitüsü), İTÜ.
- Zhou, F., Maxisch, T., Ceder, G. (2006). Configurational electronic entropy and the phase diagram of mixed-valence oxides: the case of Li_xFePO_4 . *Physical review letters*, 97(15), 155704.