



# Bilgi Kuramında Shannon Entropisi ve Uygulamaları

Oğuzhan Bahadır<sup>1\*</sup>, Hande Türkmençalıkoğlu<sup>2</sup>

<sup>1</sup>Kahramanmaraş Sütçü İmam Üniversitesi, Fen Edebiyat Fakültesi, Matematik Bölümü, Kahramanmaraş, Türkiye, (ORCID: 0000-0001-5054-8865), [obahadir@ksu.edu.tr](mailto:obahadir@ksu.edu.tr)

<sup>2</sup>Kahramanmaraş Sütçü İmam Üniversitesi, Fen Bilimleri Enstitüsü, Matematik Bölümü, Kahramanmaraş, Türkiye, (ORCID: 0000-0003-3971-8143), [handee.turkmenn@gmail.com](mailto:handee.turkmenn@gmail.com)

(Uluslararası Araştırma-Geliştirme ve Tasarım Konferansı – 15-18 Aralık 2021)

(DOI: 10.31590/ejosat.1039771)

**ATIF/REFERENCE:** Bahadır, O., Türkmençalıkoğlu, H. (2021). Bilgi Kuramında Shannon Entropisi ve Uygulamaları. *Avrupa Bilim ve Teknoloji Dergisi*, (32), 491-497.

## Öz

21. yüzyılda iletişim ve teknoloji çağında fizikten istatistiğe, mühendislikten sosyal bilimlere kadar pek çok alanda uygulamaları olan entropi, güvenli veri aktarımı, kayıpsız veri sıkıştırma, veri madenciliği, makine öğrenmesi ve sınıflandırma gibi alanlarda bilgi teorisinin bir uygulaması olarak karşımıza çıkmaktadır. Shannon tarafından olasılık hesaplamaları kullanılarak oluşturulan entropi formülü, bilgi teorisinde kullanılan önemli bir ölçüm yöntemidir. Bu çalışmada, Shannon'ın entropi formülünün bilgi teorisindeki bazı uygulamaları ile gösterilmektedir. İlk olarak adil ve hileli madeni paralar ile oynanan bir yazı-tura oyununda bilgi ve entropi miktarları hesaplanmış, elde edilen sonuçlar birbiri ile karşılaştırılmıştır. Adil madeni paradan elde edilen entropi değeri, hileli madeni paradan daha yüksek çıkmış ve bu durum entropinin bir sistemdeki belirsizliğin ölçüsü olduğunu kanıtlamıştır. Entropinin bilgi kuramında kullanılmasıyla birlikte etkili iletişim ve minimum bit kullanarak veri depolama gibi işlemlerde büyük gelişme sağlanmıştır. Yapılan ikinci uygulamada bir veri grubundaki karakterler kodlanırken, gereken minimum bit sayısı entropi formülü ile hesaplanmış ve örnek bir kodlama yapılmıştır. Bilgi teorisinde Shannon entropisi, bilgi iletimi ve depolama alanlarının yanı sıra veri madenciliği, makine öğrenmesi ve yapay tahmin mekanizmalarında da kullanılmaktadır. Yapılan son uygulamada yapay öğrenme alanlarından 'Karar Ağacı' oluşturulmasında entropi formülünün, ağacın oluşumundaki 'kök düğüm'ün tespit edilmesinde kullanıldığı gösterilmiştir. Yapılan uygulamalar Shannon entropisinin bilgi teorisindeki işlevselliğini göstermiştir.

**Anahtar Kelimeler:** Entropi, Bilgi kuramı, Bilgi kazancı

## Shannon Entropy and Its Applications in Information Theory

### Abstract

Entropy has applications in many fields from physics to statistics, from engineering to social sciences in the age of communication and technology in the 21st century. It appears as an application of information theory in areas such as secure data transfer, lossless data compression, machine learning and classification. The entropy formula created by Shannon using probability calculations is an important measurement method used in information theory. In this study, some uses of Shannon's entropy formula in information theory are demonstrated with applications. Firstly, in a coin flip game played with fair and fraudulent coins, the amount of information and entropy were calculated, and the results were compared with each other. The entropy value obtained from the fair coin was higher than that of the fraudulent coin, proving that entropy is a measure of uncertainty in a system. With the use of entropy in information theory, great progress has been made in processes such as effective communication and data storage using minimum bits. In the second application, while coding the characters in a data group, the required minimum number of bits was calculated with the entropy formula and an example coding was made. In information theory, Shannon entropy is used in data mining, machine learning and artificial prediction mechanisms as well as information transmission and storage. In the last application, it has been shown that the entropy formula is used to determine the 'root node' in the formation of the tree in the creation of the 'decision tree' from artificial learning areas. The applications have shown the functionality of Shannon entropy in information theory.

**Keywords:** Entropy, Information theory, Information gain

## 1. Giriş

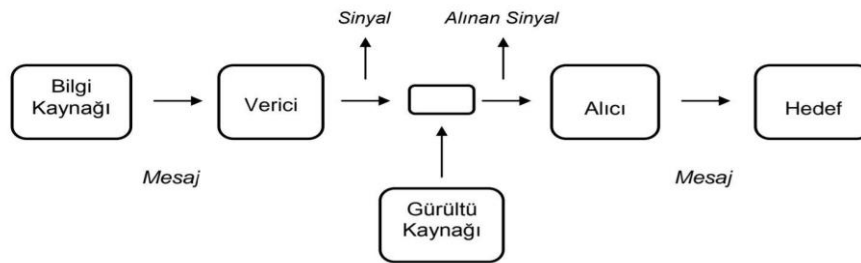
Günümüzde fizikten istatistiğe, mühendislikten sosyal bilimlere pek çok uygulama alanı olan entropi, termodinamikten alınan bir kavramdır. Termodinamikte, entropi bir sistemin mekanik işe çevrilemeyen, evrene dağılan termal enerjisini temsil eder. Örneğin; belli bir potansiyel enerjiye sahip bir top bir yamaçtan aşağı doğru düşerken mevcut potansiyel enerjisinin tamamı kinetik enerjiye dönüşmemekte, bir kısmı evrene ısı enerjisi olarak dağılmaktadır. İşte ısı olarak açığa çıkan ve evrene rastgele dağılan bu enerji entropiyi temsil etmektedir. Termodinamiğin ikinci yasasına göre herhangi bir süreç sırasında sistemin entropi değişimi sıfır veya pozitif olmak zorundadır. Çünkü evrendeki bütün süreçler düzenden düzensizliğe doğru gider ve entropisini artırma eğilimindedir. Genel olarak fiziksel sistemlerde entropi, sistemdeki kesinsizliğin, belirsizliğin sayısal değeri olarak tanımlanabilir.

Evrenin düzenden düzensizliğe doğru giden entropik yönelimi iletişim sürecinde de karşımıza çıkmaktadır. Bir kanal boyunca bir mesajın iletimi sırasında gözlenen bozulma, Claude E. Shannon tarafından entropinin bir türü olarak tanımlanmıştır. Bilgi kuramının babası olarak anılan Shannon, entropiyi bir iletinin içerdiği bilgi miktarının ölçüsü olarak tanımlamıştır (Shannon, 1948). Bu konuda yapılan ilk çalışmalar, iyi ve etkili bir iletişim sürecinin nasıl olacağı ile ilgilidir.

Bilgi iletimi ile ilgili yapılan bu çalışmalar Shannon'un entropi formülünün oluşmasına zemin hazırlamıştır.

Bilgi iletimi ile ilgili ilk çalışma 1924 yılında Harry NYQUIST tarafından yapılmıştır. Harry NYQUIST, 'Certain Factors Affecting Telegraph Speed' adlı makalesinde telgraf üzerinden maksimum hızda ve bozulma olmadan mesaj göndermenin nasıl yapılacağını açıklamıştır (Nyquist, 1924). 1928 yılında ise Ralph HARTLEY 'Transmission of Information' adlı makalesinde şu soru üzerinde durmuştur: 'Bilgi ölçülebilir mi?' Hartley bu konu üzerinde yaptığı çalışmaların sonucunda aşağıdaki formülü geliştirmiştir:

$$H = \log_2 n \quad (1)$$



Şekil 1. Shannon'un Genel İletişim Sistemi

Bu tabloda; bilgi kaynağı, alıcı birime iletilmesi için mesaj ya da mesajlar dizisi üreten sistemi gösterir. Verici, bilgi kaynağı tarafından üretilen mesajı, kanal boyunca iletilebilecek biçime dönüştüren elemandır. Sinyal, mesajın kanalda iletilebilecek biçimini ifade eder. Kanal, sinyalin vericiden alıcıya iletiildiği ortamdır. Alıcı, sinyali yeniden mesaj şekline dönüştürerek vericinin yaptığı işlemin tersini yapan elemandır. Hedef ise mesajın ulaştığı son noktayı göstermektedir (Değirmenci, 2011).

Sinyalin kanalda iletimi sırasında, ortamın özelliklerinden dolayı bazı değişimler oluşur. İstenmeyen bu

Burada  $H$ ; bilgiyi,  $n$ ; sembol sayısını,  $s$ ; her seçimde kullanılabilecek farklı sembol sayısını göstermektedir. Bu denklemden, ileti uzayının logaritmasının bize bilgi miktarını verdiği sonucunu çıkarabiliriz (Çetinkaya, 2011).

Hartley'in bu formülünü bir örnek ile anlamlandırmaya çalışalım. 0 ile 50 arasında bir sayı tutup bu sayıyı tahmin ettirmeye çalışalım. Tahmin edilmesi için kaç tane 'evet' 'hayır' sorusu sormamız gerektiğini hesaplayalım. 'Bu sayı 1 midir?', 'Bu sayı 2 midir?', 'Bu sayı 3 müdür?' şeklinde soru sorabiliriz fakat bu çok mantıklı olmayacaktır. Çünkü en iyi ihtimal ile ilk soruda sayıyı bilmemize rağmen, en kötü ihtimalle 50. soruda cevabı bulabiliriz. Onun yerine soruyu 'Bu sayı 25'den küçük müdür?' şeklinde sormak ihtimallerin yarısını elememizi sağlayacaktır ve bir sonraki soruda kalan soruları yarılıyarak devam etmek daha az soru ile cevabı bulmamızı sağlayacaktır. Sorulması gereken ortalama soru sayısı Hartley'nin yöntemine göre:  $\log_2 50 =$  soru sayısı biçiminde hesaplanır. Burada logaritmanın tabanında kullanılan 2 sayısı 'evet' ya da 'hayır' seçeneklerini, 50 ise; toplam soru sayısını temsil etmektedir. Bu durumda sonuç:  $\log_2 50 = 5,64$  olduğundan ortalama soru sayısı 5,64 bulunur. Hartley'e göre burada sormamız gereken ortalama soru sayısı 'bit' olarak adlandırılmıştır. Yani Hartley, 'Bir soru bir bite eşittir' der. Buradan bu veri 5,64 bit ile depolanabilir sonucunu çıkarabiliriz.

Hartley'nin geliştirdiği bu formül bilgi kuramının oluşmasına zemin hazırlamıştır. Bugünkü bilgi kuramının doğuşu ise, Claude E. Shannon'un 1948 yılında yayınlanan "A Mathematical Theory of Communication" adlı çalışmasına dayanmaktadır. Shannon, etkili ve iyi bir iletişimin nasıl olabileceğini açıkladığı bu çalışmasında iletişimin matematiksel bir modellemesini yaparak, ortaya koyduğu yaklaşım ile bilgi kuramının temelini oluşturmuştur. Daha sonra Shannon tarafından güvenli şekilde veri sıkıştırma, depolama ve iletme gibi sinyal işleme işlemlerinin kısıtlamalarını bulmak için geliştirilmiştir (Shannon, 1948). Shannon'un tanımladığı iletişim sistemi aşağıda verilmiştir:

değişimler gürültü olarak adlandırılır. Gürültü var ise, kaynaktan çıkan mesaj ile hedefin aldığı mesaj birbirinden farklı olur. Vericiden alıcıya gönderilen mesajlarda, bozulma en sık rastlanan durumdur. Buna karşın, mesajların hiç değişime uğramadan alıcıya ulaşması çok küçük olasılığa sahiptir. Shannon, iletilen mesajlarda zamanla artan bu bozulmayı entropinin bir türü olarak belirtmiştir ve çok küçük bir hata ile gürültülü bir kanal boyunca bilgi iletiminin mümkün olduğunu kanıtlamıştır. Shannon tarafından ortaya konan bilgi kuramı iletişim teknolojilerinde çok önemli bir soruna, bilginin en kısa yoldan doğru biçimde iletilmesine çözüm getirir (Ruelle, 2014).

Bu çalışmada entropinin bilgi iletimi ve veri depolama gibi uygulama alanlarında ne denli işlevsel olduğu, bilgi-iletişim teknolojilerinin gelişimine sağladığı katkılar ve yapay öğrenme alanında karar ağaçlarının oluşturulmasındaki kullanımının gösterilmesi amaçlanmıştır. İlerleyen bölümde Shannon entropisi tanıtılmış ve uygulama alanlarından birkaçı örneklendirilmiştir.

## 2. Shannon Entropisi ve Uygulamaları

### 2.1. Shannon Entropisi

Shannon bir bilgi kaynağı tarafından üretilen bilginin ne kadar ve hangi oranda olduğunu ölçen bir miktar tanımlamıştır. Varsayalım ki kaynağın olası n tane sonucu,  $A_1, A_2, \dots, A_n$  ( $p_1 = p(A_1), p_2 = p(A_2), \dots, p_n = p(A_n)$  ile ortaya çıkan) kaynak sonuçları kesikli olasılık dağılımı tarafından;

Tablo 1. Olay ve Olasılık

Olay	$A_1, A_2, \dots, A_n$
Olasılık	$p_1, p_2, \dots, p_n$

şeklinde tanımlansın.

Bir  $H(p_1, p_2, \dots, p_n)$  fonksiyonu olduğunu varsayalım. Bu fonksiyon bir olayın seçilmesinde ne kadar 'seçim' olduğunu 'ölçer'. H fonksiyonu aşağıdaki aksiyomları sağlamalıdır. (Shannon aksiyomları)

i) H her  $p_i$ 'de süreklidir.

ii) Eğer  $p_1 = p_2 = \dots = p_n = \frac{1}{n}$  ise o zaman H monoton artan fonksiyondur. Yani daha fazla olay varken eşit derecede olası olaylar için daha fazla belirsizlik vardır.

iii) Bir seçim, birbirini takip eden 2 seçime ayrılırsa orijinal H değeri, her bir H değerinin ağırlıklandırılmış toplamı şeklinde olmalıdır.

Shannon bu üç koşulu birden sağlayan tek fonksiyonunun,

$$H = -k \sum_{i=1}^n (p_i \log_2 p_i) \quad (2)$$

olduğunu göstermiştir. Burada k pozitif bir sabiti ifade eder. Bu Shannon tarafından kesikli sistemler için tanımlanmış bilgi entropisi tanımıdır. Sürekli sistemlerde ise toplam sembolünün yerini integral sembolü almaktadır. Bu ifade istatistiksel mekaniğin belirli formülasyonlarında görülen ifadenin aynı olması açısından dikkat çekicidir (Calin, 2014). Bilginin önemli bir ölçütü, genellikle depolama ve iletişim için gerekli olan parçaların ortalama sayısı olan entropidir.

Shannon aksiyomlarından çıkan sonuca göre entropi, bir rastgele değişkenin değerini tahmin ederken belirsizliği sayısallaştırır. Örneğin, bir yazı tura oyununun sonuç için sağladığı bilgi, bir zar atma oyununun sonuç için sağladığı bilgiden daha azdır. Yazı tura oyununda eşit olasılıklı iki sonuç varken, zar atma oyununda ise eşit olasılıklı altı sonuç vardır. Dolayısıyla yazı tura oyunundaki belirsizlik zar atma oyunundaki belirsizlikten daha azdır. Bu nedenle yazı tura oyunundaki belirsizlik, zar atma oyununa göre daha düşük entropiye sahiptir (Ben Naim, 2008).

Elimizde hilesiz bir madeni para olduğunu varsayalım. Paramızı yazı tura oyunu oynamak için havaya attığımızda aşağıdaki tablodaki bilgileri elde ederiz.

Tablo 2. Adil Para İle Yazı Tura Oyunu

Çıktılar	Yazı	Tura
Olasılık	$\frac{1}{2}$	$\frac{1}{2}$
Bit sayısı	1	
Bilgi : $-\log_2 p$	1	1
Entropi: $p \cdot \log_2 p$	$\frac{1}{2}$	$\frac{1}{2}$
Toplam entropi	1	

Bu tabloda yazı ve tura gelme olasılıkları eşit ve  $\frac{1}{2}$  olduğundan bu olayları ifade etmek için kullanılan bit miktarı 1 dir. Bu olayların gerçekleşme olasılıkları eşit olduğundan dolayı taşıdıkları bilgi miktarları toplamı 2'dir. Toplam entropi ise 1'dir.

Şimdi de elimizdeki madeni paranın hileli olması (yani olasılıkların birbirine eşit olmadığı durumu) inceleyelim.

Tablo 3. Adil Olmayan Para İle Yazı Tura Oyunu

Çıktılar	Yazı	Tura
Olasılık	$\frac{1}{4}$	$\frac{3}{4}$
Bit sayısı	1	
Bilgi : $-\log_2 p$	2	0,415
Entropi: $p \cdot \log_2 p$	0,5	0,307
Toplam entropi	0,807	

Bu tabloda ise atılan paranın tura gelme olasılığı yazı gelme olasılığından daha fazla olduğu için bu olayda sonucun tura gelmesi daha muhtemeldir. Taşıdıkları bilgi miktarı incelenirse daha düşük olasılıklı olan yazı gelmesi olayının, daha büyük olasılıklı tura gelmesi olayından daha fazla bilgi içerdiği tablodan kolaylıkla görülebilir. Toplam entropi değerinde ise bir azalma söz konusudur. Bunun sebebi eşit olasılıklı olmayan olaylardaki sürprizliğin daha az olmasıdır. Yani olaydaki sürprizlik, eşit olasılıklı yazı turaya göre daha düşüktür. Buradan; varlığı bilinen bir bilgi ile entropinin birbirine zıt kavramlar olduğu sonucu çıkarılır (Yüksel, 2020).

Shannon'ın tanımladığı entropi, belli birimlerle (bit, napier, desibel) ölçülebilen niceliksel bir büyüklüktür. Böylelikle, bir rastgele sürecin olasılık yapısı belli olduğunda, o sürecin entropisi sayısal olarak hesaplanabilmekte ve söz konusu birimlerle ifade edilebilmektedir (Çetinkaya ve Yiğit, 2019). Bilgisayarlarımızda verileri depolarken minimum bit kullanarak verilerimizi saklamaya ihtiyaç duyarız. Entropi kavramı bu anlamda bizim için oldukça işlevseldir. Hartley'in bilgi formülü kullanılarak aşağıda verilen uygulamada, entropi kullanılarak ne

kadar bilgi kazancı (bit) elde edilebileceği, 2 strateji karşılaştırılarak incelenmiştir. Her iki stratejide de bazı sorular sorulmuş ve alınan cevaplara göre hangi stratejinin soruyu daha hızlı ve daha az bit ile kaydedebileceği hesaplanmıştır.

Üzerlerinde 1'den 50'ye kadar numara yazılı 50 yarışmacının katıldığı bir koşu yarışmasında 8 yarışmacının finale kaldığını ve yarışmacıların final turu için koşu alanına numara sırasına göre yerleştirileceğini varsayalım. Her yarışmacının yan yana dizilmiş sekiz koşu parkurunda aynı anda yarışacağını düşünelim. Amacımız, finale kalan yarışmacılardan

25 numaralı yarışmacının hangi parkurda yarışacağını bulmak olsun. Bildiğimiz tek şey, 25 numaralı yarışmacının final turunda yarışacak yarışmacılardan biri olduğudur ve parkura yarışmacılar numara sırasına göre dizilecektir. Burada ihtiyaç olan bilgi yarışmacının hangi parkurda olduğudur ve bu bilgi keyfi soruların cevapları neticesinde elde edilecektir. Bu sorulara yalnızca "evet" ya da "hayır" cevapları verilecektir. Finale kalan yarışmacıların numaraları ve parkurları aşağıdaki tabloda gösterilmiştir.

Tablo 4. Yarışmacıların Numaraları ve Parkurları

	1. parkur	2. parkur	3. parkur	4. parkur	5. parkur	6. parkur	7. parkur	8. parkur
Yarışmacı numarası	3	6	8	17	25	32	43	50

Burada 2 farklı strateji ele alalım.

#### 1. Strateji:

Soru 1: 25 numaralı yarışmacı 1. parkurda mı?  
Soru 2: 25 numaralı yarışmacı 2. parkurda mı?  
Soru 3: 25 numaralı yarışmacı 3. parkurda mı?  
Soru 4: 25 numaralı yarışmacı 4. parkurda mı?  
Soru 5: ...

#### 2. Strateji:

Soru 1: 25 numaralı yarışmacı 4. parkurda mı? Evet ise amaca ulaşılır. Hayır ise 25 17'den büyük olduğundan 17 ve sol taraftaki parkurlar elenir.

Soru 2: 25 numaralı yarışmacı 6. parkurda mı? Evet ise amaca ulaşılır. Hayır ise 25 32'den küçük olduğundan sağ taraftaki parkurlar elenir.

Soru 3: 25 numaralı yarışmacı 5. parkurda mı? Cevap kesinlikle evet olacaktır.

Eğer 1. strateji seçilirse, yarışmacının hangi parkurda olduğu ilk soruda öğrenilebilir ancak bu 1/8 olasılıkla meydana gelebilir ve 7/8 olasılıkla doğru parkur seçilmeyebilir. İlk soruda yarışmacının parkurunun bulunamadığı varsayalım. İkinci soruda bulma olasılığı 1/7'dir ve 6/7 olasılıkla para bu parkurda olmayabilir. Eğer ilk altı sorunun sonunda yarışmacının parkuru bulunamazsa, yedinci sorudan sonra cevaba yani yarışmacının hangi parkurda olduğu bilgisine ulaşılır. Diğer taraftan, eğer 2. strateji seçilirse en kötü senaryoyla gereken bilgiye ilk soruda ulaşamaz, ikinci sorunun sonunda da istenen bilgi elde edilemez ancak üçüncü sorunun sonunda gereken bilgiye kesinlikle ulaşılır. Bu durumda ikinci stratejiyi seçmek çok daha mantıklı olacaktır. Çünkü birinci stratejide ilk sorunun sonunda yarışmacının parkuru bulunabilir fakat bu çok düşük bir olasılıktır. Eğer ilk soruda parkur bulunamadıysa ikinci soruya geçildiğinde yalnızca bir parkur elenmiş olacaktır. Fakat ikinci stratejide ilk soruda doğru cevaba ulaşılmadığı takdirde 17'den küçük olan sol taraftaki 3 parkur da elenmiş olacaktır. İkinci soruda da sağ taraftaki parkurların yarısı elenir ve 3. soruda kesin olarak cevap bulunur (Değirmenci, 2011). O halde verilen örnekteki veriyi kaydederken en kötü ihtimal ile 8 soruda doğru cevaba ulaşılabilirdiğinden, Hartley'in formülü kullanılarak;  $\log_2 8 = 3$  bit bulunur. Aynı zamanda entropi değeri de 3'tür. Yani oyuncunun parkurunu belirlemek için 8 parkur her

defasında ikiye bölünerek doğru parkurun 8 soru yerine 3 soruda bulunması (8 bit yerine 3 bit ile depolanması) sağlanabilir. Dolayısıyla burada bilgi kazancı elde edilebildiği görülmüştür.

Calin, 2014 yılında yaptığı çalışmada; entropinin bilgi teorisinde bir uygulamasının Shannon tarafından bilgi ve iletişimin problemlerini ve yönlerini açıklamak için formüle edildiğini belirtir (Calin, 2014). Bu teoride bilgi kaynağı (alıcıya iletilecek mesaj sayısının bir dizisi) tarafından sisteme verilen mesaj bir miktar bilgi içerir ve bu bilgi sayısal olarak logaritma 2 tabanında ( $\log_2 n$ ) ölçülebilir. Bu durumda ortaya çıkan birimlere ikili rakamlar veya bitler denir. Bu bilgi, tüm olası mesajların arasından bir mesajın seçilebileceği seçim özgürlüğünün bir ölçüsüdür. Bir bit ölçüsü, eşit derecede olası iki seçenek arasındaki seçimi temsil etmektedir. Dolayısıyla Shannon tarafından olasılık hesaplamaları kullanılarak geliştirilen bilgi kuramında entropi kavramı, bilgisayarlarımızda verileri minimum bit kullanarak saklama, hızlı ve etkili iletişim kurabilme gibi alanlarda bizim için oldukça işlevsel bir kavram olmuştur.

Shannon entropisi bilgi analizinde kullanılan önemli bir ölçüm methodudur. Shannon'un formülünün, bir veri grubundaki karakterler kodlanırken gerekli olan minimum bit sayısını bulduğu görülmektedir. Bu uygulamada 'TÜRKİYE' kelimesini bilgisayarımıza kodlarken gerekli minimum bit sayısını, Shannon'un formülü yardımıyla bulmak amaçlanmaktadır.

Aşağıda verilen Shannon'un entropi formülüne göre, her harfin olasılığı ile bu olasılığın logaritma 2 tabanındaki değeri çarpılıp, bulunan değerler toplanır ve en son bulunan değer toplama işlemine göre tersi alınır. Entropi formülü aşağıdaki gibidir.

$$H = -k \sum_{i=1}^n (p_i \log_2 p_i) \quad (3)$$

'TÜRKİYE' kelimesindeki her bir harf veri grubunda bir tane olduğundan bütün harflerin seçilme olasılığı birbirine eşit ve 1/7'dir. Buna göre Shannon entropisi aşağıdaki şekilde hesaplanır:

$$\begin{aligned} H(x) &= -(1/7. \log_2 1/7) + (1/7. \log_2 1/7) + (1/7. \log_2 1/7) + \\ &\quad (1/7. \log_2 1/7) + (1/7. \log_2 1/7) + (1/7. \log_2 1/7) + \\ &\quad (1/7. \log_2 1/7) \\ &= -(1/7). 7 \log_2 (1/7) \end{aligned}$$





Tablo 5. Niteliklere Göre İşe Eleman Alımı

Yaş (<30, 30-40, >30)	Yabancı Dil Yeterliliği (yetersiz, yeterli)	Maaş beklentisi (5000 TL'den az, çok)	İşe alınma durumu (evet, hayır)
<30	Yeterli	Az	Hayır
<30	Yeterli	Çok	Hayır
30-40	Yeterli	Az	Evet
>40	Yeterli	Az	Evet
>40	Yetersiz	Az	Evet
>40	Yetersiz	Çok	Hayır
30-40	Yetersiz	Çok	Evet
<30	Yeterli	Az	Hayır
<30	Yetersiz	Az	Evet
>40	Yetersiz	Az	Evet
<30	Yetersiz	Çok	Evet
30-40	Yeterli	Çok	Evet
30-40	Yetersiz	Az	Evet
>40	Yeterli	Çok	Hayır

Karar ağacı oluşturulurken verilen üç kriterden (yaş, yabancı dil yeterlilik ve maaş beklentisi) hangisinin adayları değerlendirmek için daha önemli bir nitelik olduğu entropi formülü ile hesaplanır. Her bir nitelik için ayrı ayrı entropi hesabı yapılır ve bilgi kazancı bulunur. Elde edilen en büyük kazanç ağacın kökünü belirler. Örneğin 'Yaş' niteliği için örnek bir hesap yapılırsa;

Tablo 6. Yaş Kriteri Bilgi Miktarı

Yaş	Evet sayısı	Hayır sayısı	Info(evet, hayır)
<30	2	3	I(2,3)
30-40	4	0	I(4,0)
>40	3	2	I(3,2)

4 numaralı eşitlik kullanılarak sistemin entropi değeri aşağıdaki şekilde hesaplanır.

$$\begin{aligned} \text{Entropi (İş)} &= I(9,5) \\ &= -\frac{9}{14} \log_2 \frac{9}{14} - \frac{5}{14} \log_2 \frac{5}{14} \\ &= 0,940. \end{aligned}$$

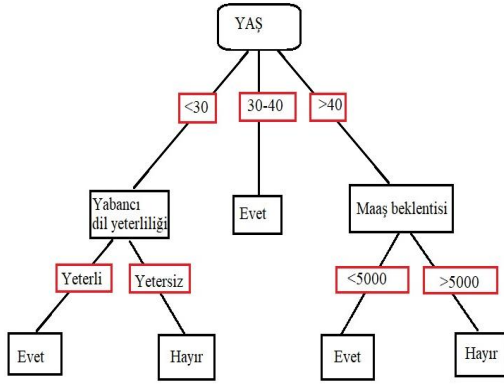
5 numaralı eşitlik kullanılarak yaş kriterinin içerdiği bilgi miktarı aşağıdaki şekilde hesaplanır.

$$\begin{aligned} \text{Info}_{\text{yaş}}(\text{İş}) &= \frac{5}{14} I(2,3) + \frac{4}{14} I(4,0) + \frac{5}{14} I(3,2) \\ &= \frac{5}{14} \left( -\frac{2}{5} \log_2 \frac{2}{5} - \frac{3}{5} \log_2 \frac{3}{5} \right) + \frac{4}{14} \left( -\frac{4}{4} \log_2 \frac{4}{4} - \frac{0}{4} \log_2 \frac{0}{4} \right) + \frac{5}{14} \left( -\frac{3}{5} \log_2 \frac{3}{5} - \frac{2}{5} \log_2 \frac{2}{5} \right) \\ &= 0,694. \end{aligned}$$

6 numaralı eşitlik kullanılarak kazanılan bilgi miktarı aşağıdaki şekilde hesaplanır.

$$\begin{aligned} \text{Gain (Yaş)} &= \text{Entropi (İş)} - \text{Info}_{\text{yaş}}(\text{İş}) \\ &= 0,940 - 0,694 \\ &= 0,246. \end{aligned}$$

Benzer şekilde diğer niteliklerin de entropi ve bilgi kazançları hesaplandığında; Gain (Yabancı dil yeterlilik) = 0,151, Gain (Maaş beklentisi) = 0,048 değerleri elde edilir. Buradan çıkarılan sonuca göre oluşturulmak istenen karar ağacını bilgi kazancı en yüksek olan nitelikten yani yaş kriterinden başlatmak avantajlı olacaktır. Elde edilen sonuçlara göre oluşturulan karar ağacı Şekil 3'te gösterilmiştir.



Şekil 3. Sonuç Karar Ağacı

### 3. Sonuç

Çalışmadan elde edilen sonuca göre bilgi kuramında, bilgi miktarının ölçüsü olarak tanımlanan entropi, bize daha az olasılıklı olayların daha çok bilgi içerdiğini, olasılığı daha fazla olan olayın ise daha az bilgi içerdiğini göstermiştir. İki olayın gerçekleşme olasılıkları birbirine eşit olduğunda hangi sonucun geleceği eşit olasılıklı olmayan durumlara göre daha belirsizdir ve bu yüzden entropi maksimum değeri alır. Uygulamalardan da görüldüğü üzere, Hartley'in ortaya koyduğu ve Shannon'un geliştirdiği entropi tanımı ve formülasyonları, bilgi teknolojisinde, özellikle bilgi iletimindeki sorunları ortaya koyma, bilginin en kısa yoldan doğru biçimde iletilmesini sağlama, daha az bit ile veri depolama ve yapay öğrenme alanlarında karar ağaçlarının oluşumunda oldukça kullanışlıdır. Bu kavramın bilgi teknolojisinin gelişimine oldukça katkı sağladığı görülmüştür.

### Kaynakça

- Ben-Naim, A. (2008). A Farewell to Entropy: Statistical Thermodynamics Based on Information, *World Scientific Publishing*, 9-19. ISBN-13 978-981-270-706-2.
- Bulut, F. (2017). Different Mathematical Models for Entropy in Information Theory. *Bilge International Journal of Science and Technology Research*, 1(2), 167-174 . Retrieved from <https://dergipark.org.tr/tr/pub/bilgesci/issue/32353/335439>.
- Calin, O., Udrişte, C. (2014). Geometric Modeling in Probability and Statistics, *Publisher: Springer*, ISBN 978-3-319-07778-6, 78-106.
- Çetinkaya, O. (2011). Belirsizliğin Ölçülmesi ve Entropi. *İstanbul Üniversitesi İktisat Fakültesi Mecmuası*, (44), (1-4). <https://dergipark.org.tr/tr/pub/iuifm/issue/834/9160>.
- Çetinkaya, C. P., Yiğit, S. D. (2019). Su Kalitesi Gözlem Ağlarının Performansının Değerlendirilmesi için Bir Yöntem Önerisi ve Gediz Havzasında Uygulanması . *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi* , 21 (62) , 483-497 . DOI: 10.21205/deufmd.2019216214.
- Değirmenci, İ. (2011). Entropi Ölçüleri ve Maksimum Entropi İlkesi, *Yüksek lisans tezi, İstatistik Anabilim Dalı, Hacettepe Üniversitesi*, 1-3.
- Kale, E. (2021). Meteorological Drought Forecasting Using Decision Three, *Dissertation, Dissertation Master's*

*Program of Electrical and Computer Engineering, Antalya Bilim University.*

- Mu, Y., Liu, X., Yang, Z. (2017). A Parallel C4.5 Decision Tree Algorithm Based on MapReduce, *Concurrency and Computation: Practice and Experience*, 29(8).
- Nyquist, H. (1924). Certain Factors Affecting Telegraph Speed. *Transactions of the American Institute of Electrical Engineers*, (43), 412-422.
- Ruelle, D. (2014). Rastlantı ve Kaos, *Tübitak Popüler Bilim Kitapları*, Çevirmen: Deniz Yurtören, 102-107.
- Shannon, C.E. (1948). A Mathematical Theory of Communication, *The Bell System Technical Journal*, 27(3), 379-423.
- 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.
- Yüksel, E. (2020). Düzensizlik (Entropi), Çapraz Düzensizlik (Cross Entropi) ve KL-İraksaklığı (KL-Divergence), <https://medium.com/kaveai/tagged/entropi>