

## Faktör Analizi Çözümlemesine Ortak Entropi Yaklaşımı

Nida Oruç<sup>1</sup>

Doğan Yıldız<sup>2</sup>

### Özet

Faktör analizi, belirli bir olguyla ilgili ve birbirleriyle ilişkili çok sayıda değişkenin birbirleriyle olan ilişkisini inceleyerek, değişkenlerin sayısını azaltmak ve ilişkisiz yeni değişkenler yaratmak için kullanılan çok değişkenli istatistiksel bir analizdir. Bu çalışmada farklı bir yaklaşım olarak, faktör çıkarmada birleşik entropiden yararlanılmıştır. Buradan elde edilen “Faktörleşme” ile “Açıklayıcı Faktör Analizi” sonuçları karşılaştırılarak benzerlik ve farklılıkları tartışılmıştır.

İstatistik ile entropiyi birleştiren bir model kullanılmıştır. Uygulama kısmında ilk olarak istatistikte yaygın olarak kullanılan faktör analizi sonuçları elde edilmiştir. Ardından, kovaryans matrisi yerine birleşik entropi matrisi kullanılan entropi tabanlı yaklaşımla elde edilen matrisle bir karşılaştırma yapılmıştır. Bu bağlamda birleşik entropi matrisi ile elde ettiğimiz sonuçlar ile faktör analizi ile elde ettiğimiz sonuçlar karşılaştırılmıştır. Son durumda ise birleşik entropi matrisi ile elde edilen sonuçlar gösterilmiştir. Ek olarak, bu çalışma ile entropiye teorik destek ile aslında varyans-kovaryans ölçülerinin kullanım alanlarına uygulanabileceği ve desteklenebileceğini göstermek hedeflenmiş ve çözülmesi gereken bazı sorunlara rağmen, alternatif olarak bilinen yöntemlerin yerine önerilmektedir.

### 1. Giriş

Bilim insanları, istatistiksel gözlemler veya deneylerde bilgi arayışından yararlanır. Claude Shannon, bu bağlamda öne çıkan bilim adamlarından biridir. Bu çerçevede bakıldığında, 1948 yılında Claude Shannon’ın “A

1 Araştırma Görevlisi, İstanbul Topkapı Üniversitesi, nidaoruc@topkapi.edu.tr, 0009-0005-0230-2234

2 Dr. Öğr. Üyesi, Yıldız Teknik Üniversitesi, dyildiz@yildiz.edu.tr, 0000-0001-7430-2368

Mathematical Theory of Communication” adlı kitabında istatistikte entropi kavramı ilk kez ortaya çıktı. Bilgi teorisinde bu kavram Shannon entropisi olarak da bilinir. Shannon’nun makalesi yazılmadan önce, bilgi genellikle belirsiz bir sisli akış olarak görülüyordu. Ancak Shannon’ın makalesi ile, bilginin iyi tanımlanmış ve en önemlisi ölçülebilir bir nicelik olduğunu gösterdi (Altaylıgil, 2008).

Boltzmann tarafından ele alınan entropi terimi, 1948 yılına kadar fizik alanında kullanılan bir kavram olmuştur. Bell Laboratuvarlarında bu tarihlerde “Enformasyon Teorisi” üzerinde çalışan Claude Shannon, yüzyılın en ünlü matematikçilerinden biri olan John von Neumann’ın da yardımıyla bu terimi bilgedeki belirsizliğin ölçüsünü göstermek için kullanmıştır (Shannon, 1948).

Entropinin tanımı için ortaya atılan her fikir her alanda düzensizliğin ölçülmesi amacıyla kullanılması için tanımlanmıştır. Bilgi teorisi üzerindeki tanımında ise bir veri kaynağındaki belirsizlik (rastgelelik) miktarını temsil eden bit sayısıdır (Koşan vd., 2019).

İstatistikte entropi kavramını, bilgi teorisinden türetilmiştir. Entropi kavramı başlangıçta termodinamikte düzensizliği veya belirsizliği ölçmek için ortaya çıkmıştır. Bilgi teorisi, ele alınacak şekliyle, matematiksel olasılık ve istatistik teorisinin bir alt dalıdır. Bu bağlamda, soyut matematiksel formülasyonları, herhangi bir olasılıksal veya istatistiksel gözlem sistemi için genel geçerli olan bir nitelik taşımaktadır. Dolayısıyla, bilgi teorisi, olasılık ve istatistik gibi çeşitli disiplinlerde uygulanabilmektedir. Modern iletişim teorisinde önemli bir rol oynar, iletişim sistemini stokastik veya rastgele bir süreç olarak formülize eder. Tuller, iletişim teorisinin genellikle bilgi teorisi olarak adlandırıldığını belirtir. Rothstein, bilgi teorisini ölçülebilir kümelerle ilgilenen, belirli bir doğası belirtilmemiş alternatifler arasındaki seçimlerle ilgilenen soyut matematik şeklinde tanımlamıştır (Kullback, 1960).

Bir sistemin düzensizliğinin ölçülmesinde bilgi teorisindeki entropi kullanılmaktadır. Bilgi teorisinde entropi kavramı  $H$  ile ifade edilmektedir. Entropi, bir bilgi ortamındaki düzensizlik seviyesini belirlemek amacıyla da kullanılmaktadır. Bilginin monoton bir düzen içerisinde akması durumunda düzensizliği azaltabilir. Entropi formülündeki  $n$  genellikle veri miktarını temsil etmektedir (Koşan vd., 2019). Ek olarak  $p_i$  ise ilgili verideki olasılığı temsil etmektedir.

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

Yukarıda (1) ile ifade edilen formül esas alındığında,  $p$  ve  $1-p$  olasılık değerlerine sahip olan iki durum varsayımı altında, bu iki olasılığın eşit

olduğu durumda entropi değerinin maksimum olduğu noktadır (Koşan vd., 2019).

Bu çalışmada asıl ilgilenilen konu bilgi teorisi, olasılık, istatistik ve matematik teorilerinin bir dalıdır. Bu nedenle bilgi teorisine bu alanlarda oldukça sık rastlanmaktadır. Ek olarak modern iletişim teorisinde önemli bir rol oynamakla birlikte iletişim sistemini stokastik veya rastgele bir yöntem olarak formüle eder.

Bir sistemde girilen bilgi miktarındaki artış, sistemdeki belirsizlik seviyesinin azalması sebebiyle entropi değerinin düşmesine sebep olacaktır. Bunun tersi olarak diğer taraftan ise bilgi miktarındaki azalış, belirsizlik seviyesinin artması sebebiyle entropi değerinin yükselmesine neden olacaktır (Koşan vd., 2019).

Bir sistemin bilinmeyen yapısındaki artan bilgi miktarı, sistemin belirsizliğini azaltacaktır. Bu bağlamda istatistiksel bir bakış açısından ele alınan kutulardan top çekme örneğinde ve zar atma deneyinde bilgi miktarları ile entropi düzeyleri arasındaki ilişki, örneklerde detaylı bir şekilde incelenmiştir (Öncül, 2023).

Entropi terimi bilgi teorisinde tanımlanmasında ise belli başlı özellikler altında şekillendirilmiştir (Koşan vd., 2019). Bu bağlamda özelliklerin ilkinde eğer bir sistemin entropisi sıfırsa, istisna bir durum haricinde tüm durumların olasılıklarının sıfır olduğu belirtilebilir. Bu istisna durum için ise tek bir durum kalmaktadır. Bu durum ise, tüm olasılıkların tamamına sahiptir. Diğer bir özellik ise her bir  $p_i$  değeri belirli  $n$  için eşit olduğu zaman entropi en yüksek değere ulaşır ve  $\log(n)$ 'e eşittir. Bu özellik sezgisel açıdan en belirsiz olan durum olarak ifade edilmektedir. Üçüncü özellikte ise ele alınacak bir diğer nokta ise  $x$  ve  $y$  olacak şekilde iki olay ele alındığında ve bu olayların sırasıyla olasılık değerleri  $m$  ve  $n$  olarak belirtildiğinde;

Ortak olasılığı  $p(i, j)$ 'dir. Bu ortak oluşumun entropisi;

$$H(x, y) = -\sum_{i,j} P(i, j) \log_p(i, j) \quad (2)$$

$$H(x) = -\sum_{i,j} P(i, j) \log_p(i, j) \quad (3)$$

$$H(y) = -\sum_{i,j} P(i, j) \log_p(i, j) \quad (4)$$

Bağımsız olaylar söz konusu olduğunda, bireysel entropilerin toplamının ortak olayın belirsizliğine eşit ya da daha büyüktür şeklinde ifade edilebilir.

$$H(x, y) \leq H(x) + H(y) \quad (5)$$

$$p(i, j) = p(i) p(j) \quad (6)$$

Öne çıkan dördüncü özellik olarak  $\{1,0,0,\dots,0\},\{0,1,0,\dots,0\},\dots\{0,0,0,\dots,1\}$  şeklindeki dağılımların entropi değeri  $0$  dir.

Bir başka özelliğe değinmek gerekirse bir olasılık dağılımındaki tüm  $p_i$  değerlerinin birbirine eşit olmasına yönelik bir değişim entropi değerini arttırır.

Shannon entropisi “A Mathematical Theory of Communication” başlıklı araştırma makalesi sayesinde öne sürmüştür. Shannon’nun tanımladığı iletişim sisteminde bilgi kaynağı, alıcı birime iletilmesi amacıyla ileti ya da ileti dizisi üreten sistemdir. Verici ise, bilgi kaynağı ile üretilmiş olan mesajı, kanal boyunca iletim için uygun sinyal oluşturan bir sistemdir. Mesajı kanalda iletmek amacıyla kullanılan biçime sinyal adı verilmektedir. Vericiden alıcıya iletilen sinyalin ortamı kanal olarak adlandırılmaktadır. Alıcı ise bu sinyali yeniden mesaj formatına dönüştürür ve genellikle vericinin yaptığı işlemin tersini yapmaktadır. Alıcı terminal ile iletişim kurmak için bir ileti ya da ileti dizisi üreten sistem, bilgi kaynağı olarak isimlendirilmektedir. Sinyalin vericiden alıcıya iletilmesinde kullanılan ortam kanal olarak adlandırılmaktadır. Bilginin varış yeri yani bir diğer adı hedef, iletinin ulaştığı son noktadır şeklinde tanımlanabilmektedir (Değirmenci, 2011).

Shannon’nun entropisi iletişim kuramı açısından bir ölçü olarak tanımlanmış olmasına rağmen, bu ölçünün geniş çaplı bir kullanım alanı mevcuttur. Bu alanlardan bazıları matematik, mühendislik, fizik, biyoloji, sosyal bilimlerdir. Entropinin bu önemli başarısı, özellikler olasılıksal sistemde belirsizliği ölçebilmesindeki olarak söylenebilmektedir.

Shannon, bir olay hakkındaki bilgi edinme sürecini, olayın belirsizlik içermesi gerektiği durumlarda gerçekleşeceğini belirtmiştir. Bu durumlarda, yüksek olasılıklı olayların ortaya çıkması genellikle çok az bilgi verir, ancak düşük olasılıklı olayların ortaya çıkması genellikle daha fazla bilgi verir. Bu bağlamda, belirli bir alternatfin gerçekleşme olasılığı, ilgili çözümü gösteren işaret, sembol ya da sayının belirsizliğini göstermektedir. Sonuç olarak, edinilen bilgi, giderilen belirsizliğin miktarını dolaylı olarak ölçer (Bahadır vd., 2021).

Shannon, entropi terimini bu yöntemle, bir olayın farklı seçeneklerinin beklenen değeri olarak matematiksel bir ilişkiyle tanımladı. Entropi bit veya desibel gibi birimlerle ölçülebilen bir nicel büyüklüktür. Bu sebeple bir rasgele sürecin olasılıklı yapısının belirlenmesinin ardından bu sürecin entropisi hesaplanır ve belirli birimler ile ifade edilebilmektedir. Shannon geliştirmiş olduğu entropi formülündeki logaritmayı 2 tabanında kullanmıştır. Logaritma kullanılmasının amacı ise bir bağımsız sistem içindeki entropinin toplanabilir

nicelik formatında getirilebilmesi için kullanılmaktadır. Bu kapsamda, logaritma işleminin bilimsel çalışmalarda da sayısal kodlama sistemlerinde de uygun olduğu kanısına varılmıştır ve bu sebeple iki tabanında logaritma kullanılmıştır. İki tabanında logaritmanın tercih edilmesinde en büyük etken ise sayısal kodlama sistemlerine uygun olmasından kaynaklanmaktadır. Ek olarak entropi değerlerinin hesaplanmalarında elde edilen sonuçlarının ikili sayı sisteminde kısaltılarak ifade edilmesine olanak sağladı. Bu şekildeki kısaltma birimine bit denmektedir (Cover, 1999).

Olası durumları  $p_1, p_2, p_3, \dots, p_n$  kümesi olarak tanımlayan Shannon, bu kümedeki hangi olayın gerçekleşeceği bilinmediği için bu oluşan belirsizliği entropi yardımıyla ölçülebileceğini ifade etmiştir. Entropinin genel formülü (7) eşitliğinde gösterildiği gibidir. Denklemden H entropiyi ifade ederken, k sabiti Boltzmann sabiti olarak tanımlanmaktadır (Kullback, 1960).

$$H = -K \sum p_i \log_2 p_i \quad (7)$$

Shannon, k sabitini 1 kabul ederek ortaya Shannon entropisi ya da bir diğer ifadeyle Shannon belirsizliğini ortaya koymuştur. Shannon entropisi (8) denklemi ile ifade edilmektedir.

$$H = -\sum p_i \log_2 p_i \quad (8)$$

Bu bağlamda bu ifadeler ışığında varılan sonuç, rasgele bir değişkenin değerini tahmin ederken ortaya çıkmış olan belirsizliğin sayısallaştırılan entropi değeridir.

Bilgi teorisinin temellerini ortaya atmış olan Claude Shannon, ortak entropi ile Shannon entropisinin arasındaki ilişkiyi açıklamıştır. Bir önceki bölümlerde yazıldığı gibi bir olayın meydana gelme olasılığına göre bulunan bilgi ölçüsü Shannon entropisi ile elde edilmektedir. Bununla birlikte, eğer değişken sayısı iki ya da daha fazla olması halinde rasgele değişkenlerin kendi aralarındaki belirsizliğin ölçümüne ortak entropi denir. Shannon bu kavramlar üzerinde çalışmaları sayesinde gelişen bu alan, günümüzde istatistik, matematik, bilgisayar bilimleri ve birçok başka alanda uygulamalarda kullanılabilmektedir.

Ortak entropinin genel anlamda tanımı  $p(x, y)$  olasılık dağılımına sahip ve  $(X, Y)$  kesikli rasgele değişkeninin entropisi (9) denklemi ile tanımlanır (Cover, 1999).

$$H(X, Y) = -\sum_{x \in X} \sum_{y \in Y} P(x, y) \log P(x, y) \quad (9)$$

Bu çalışmada temel hedefi, belirli bir veri seti ile ilk olarak faktör analizi gerçekleştirilmesidir ve bu analizin hemen ardından ise aynı veri seti kullanılarak faktör analizinin başındaki gözlemlenen değişkenler arasındaki bağımlılığın ölçülmesinde kullanılan korelasyon matrisi yerine entropik tabanlı yöntem kullanılacaktır. Değişkenlerin entropi değerlerinden faydalanılarak birleşik entropi matrisi oluşturulacak ve elde edilen bu matrisi faktör analizindeki varyans-kovaryans matrisi yerine kullanarak bu doğrultuda bu matris yardımı ile boyut indirgeme yönteminin etkinliği incelenecektir. Son aşamada, elde edilen sonuçların karşılaştırılması yapılacak olup bu doğrultuda farklı yaklaşımların avantajları ve dezavantajları incelenecektir.

Bu çalışma kapsamında, kullanılan verinin faktör analizi ve entropi temelli yaklaşımın analizinin sonuçlarının karşılaştırılmasını değerlendirmeyi hedeflenmiştir. İlk aşamada faktör analizi yapılacak ve faktör analizindeki aynı amaç için entropi tabanlı yaklaşım kullanılarak birleşik entropi matrisi elde edilecek ve faktörler bulunacaktır. Bu iki sonuç karşılaştırmalı olarak ölçülecektir. İkinci olarak ise kovaryans matrisi yerine birleşik entropi matrisi kullandığımız entropi tabanlı yaklaşımla elde edilen matris yardımı ile elde edilen sonuçların avantajları ve dezavantajları üzerinde durulacak ve değerlendirilecektir. Bu çalışmasının çıktıları ise entropinin istatistiksel anlamda daha sık kullanılmasına veya yapılacak olan araştırmalarda farklı bir yöntemin de kullanılabilir olmasında yardımcı olması beklenebilir.

## 2. Metodoloji

Çok değişkenli istatistiksel analiz teknikleri, çok sayıda değişkenin bulunduğu durumlarda genellikle tercih edilmektedir. Araştırmacılar bu teknikleri kullanarak, değişkenler arasındaki karmaşık ilişkileri anlamaya çalışmaktadırlar. Analize başlamadan önce ilk yapılması gereken şey değişkenler arasındaki ilişkileri iyi bir şekilde kavrayabilmektir. Sınıflandırılmış verilerin olduğu durumlar için farklı yöntemler geliştirilmiştir. Bu tür veriler üzerinde çalışılırken, doğrusal olmayan yaklaşımların veya sınıflandırılmış değişkenlerin seviyelerinin kukla değişken olarak kullanılması oldukça yaygın kullanılan yöntemlerdendir. Entropiye bağlı hesaplamalarda ise, değişkenler hakkındaki yapılan hesaplamaların kesin sonuçları yerine sınıflandırılmış verilerin olasılıkları dikkate alınarak işlem yapılmaktadır. Bu gibi durumlarda analizin belirsizlik düzeyi veya analizin düzensizliği hesaplanarak, değişkenler arasındaki ilişki çözümlenmeye çalışılmaktadır. Entropiye bağlı hesaplamaları kullanarak değişkenlerin yapısını anlayabilmek ve verileri daha sade bir biçimde yazabilmek için temel bileşenler analizi kullanılmaktadır.

Analiz yaparken çok sayıda değişkenin kullanılması, araştırmacının işini çok daha karmaşık ve güç bir duruma getirebilmektedir. Bu gibi zorlukların önüne geçmek ve analizi daha kolay yönetilebilir hale getirmek, araştırmacının işini çok daha kolay hale getirecektir. Bu sebeple, çok sayıda değişkenin daha az sayıda değişkene indirgenmeye çalışılması oldukça önemlidir. Daha az değişken olmasıyla birlikte, analizin uygulanabilirliği yüksek ölçüde arttırılabilir ve daha net, daha kolay yorumlanabilir sonuçlar elde edilebilmektedir. Değişken sayısının indirgenmesine yardımcı olan yöntemlerden biri temel bileşenler analizidir. Bu yöntemle değişkenler daha az sayıdaki temel bileşenlerine kadar indirgenmekte ve veri seti özet haline getirilmektedir (Büyüköztürk, 2002).

Faktör analizi ile birlikte,  $p$  tane değişkeninin bulunduğu durumlarda birbirleriyle ilişkili olan değişkenlerin birbirleriyle yanyana getirilmesi ve daha az sayıda elde edilecek olan yeni (ortak) ilişkisi olmayan değişkenlerin ortaya çıkması amaçlanmaktadır. Bu sebeple faktör analizi, boyut indirgeme konusunda, araştırmacılar arasında oldukça popüler olan bir seçenektir (Shannon, 1948).

Faktör analizleri yapılar iken, tüm değişkenler bir arada ele alınmaktadır. Sonrasında ortaya çıkabilecek herhangi bir faktör ise, gözlemlenmiş değişkenlerin bir fonksiyonu olarak yani bağımsız bir değişken olarak kabul edilmektedir (Tatlıdil, 1992).

Entropi ve varyans benzer yapıya sahiptir. Benzer özelliklere sahip olan entropi ve varyansı incelediğimizde ilk olarak ancak ve ancak  $p_i$ 'den biri 1 ise, o zaman  $H = 0$ . Bu özellik varyans içinde geçerlidir. Bir diğer ortak özellik ise  $x$  ve  $y$  değişkenleri bağımsız ise  $H(x, y) = H(x) + H(y)$  şeklindedir. Bu durum da varyans içinde geçerlidir.  $x$  ve  $y$  bağımsız ise  $var(xy) = var(x) + var(y)$  şeklinde ifade edilebilir (Wei, 1987).

Belirsizlik ölçüsü için varyans ve entropi arasındaki ilişki ve farklılıklar mevcuttur. Bilgi teorisi çalışmaları ve uygulamaları sırasında, farklı verilere göre varyans ve entropi arasında uygun bir istatistik seçilmelidir. Bu aşamada Tablo 1'den faydalanılabilmektedir (Wei, 1987).

*Tablo 1. Veri tipi ve İstatistik*

Veri Tipi	İstatistik
Değerler (frekans dağılımında dönüştürülmesi güç değerler)	Varyans
Değerler ve olasılıklar	Varyans ya da entropi
Olasılıklar (kukla değişken uygulamak kolay)	Varyans ya da entropi
Olasılıklar (kukla değişken uygulamak zor)	Entropi

Ortak entropi hesaplanmasında kullanılan formül (10) ile gösterilmiştir.

$$H(X, Y) = -\sum_{i,j} p(i, j) \log p(i, j) \quad (10)$$

Buradaki sonuç ise  $X_\alpha$  rassal değişkeninin marjinal entropisine eşittir. Bu durumda ulaşılabacak sonuç ise (11) eşitliğidir. (Çilingir, 2019)

$$H(X, X) = H(X) \quad (11)$$

Birleşik entropi ile her bir olayın ayrı entropileri arasındaki ilişki incelendiğinde, bu ilişki (12) ile ifade edilir.

$$H(X, Y) \leq H(X) + H(Y) \quad (12)$$

Bu eşitsizlik Schwarz-Cauchy eşitsizliğine denktir. Bu eşitsizliğe ulaşabilmemizdeki sebep olayın bağımsız olması sebebiyle  $p(i, j) = p(i)p(j)$  denkleminin geçerli olmasıdır. Sonuç olarak birleşik entropi olayının belirsizliği, X ve Y'nin belirsizliklerinin ayrı ayrı toplamına eşit veya daha küçüktür (Çilingir, 2019).

Bir veri setinde bağımlılık ölçüsünün hesaplanmasında ise (12) eşitliğinden faydalanarak;

$$\beta_{x_\alpha, x_\beta} \geq \frac{H(x_\alpha) + H(x_\beta)}{H(x_\alpha, x_\beta)} \quad (13)$$

denklem (13) ile bağımlılık ölçüsünün hesaplanması mümkündür (Cover, 1999).

Değişkenlerin entropi değerleri ile birleşik entropi değerleri elde edildikten sonra,  $p \times p$  boyutlarında simetrik bir birleşik entropi matrisi oluşturmak mümkündür (Erbaş, 2010).

*Tablo 2. Birleşik entropi matrisinin görselleştirilmiş hali*

	$X_1$	...	$X_p$
$X_1$	$H(X_1)$	...	$H(X_1, X_p)$
...	...	$\ddots$	...
$X_p$	$H(X_1, X_p)$	...	$H(X_p)$



Bu çalışma doğrultusunda ülkelerin sosyo-ekonomik ve sağlık alanındaki faktörleşmesi üzerinde durulması amaçlanmıştır. Literatürde faktör analizi ve entropi alanında çeşitli araştırmalar mevcuttur. Ancak istatistiksel anlamda yapılan entropi tabanlı analizi yöntemlerinin uygulamaları açısından literatür maalesef oldukça zengin değildir. Bu sebeple bu çalışmasıyla birlikte istatistiksel anlamda entropi tabanlı bir uygulama ile hem dikkat çekilmesi hem de literatüre katkıda bulunması amaçlanmaktadır.

Bu çalışmada hedef, veri setini analiz etmek ve veri setindeki temel faktörleri belirlemek için faktör analizi ve entropi tabanlı bir yaklaşım olan birleşik entropi yönteminin kullanılmasıdır. Bu uygulama kapsamında, her iki yöntem de kullanılarak elde edilen faktörlerin belirlenmesi hedeflenmiştir. Bu bağlamda, her iki analiz yöntemiyle elde edilen faktörler karşılaştırılması amaçlanmıştır. Bu karşılaştırma, farklı analiz yöntemlerinin veri setindeki yapıyı anlama ve yorumlama yeteneklerini değerlendirmeyi amaçlamıştır.

Özetle, bu çalışma ile hem faktör analizi hem de entropi tabanlı bir yaklaşım olan birleşik entropi kullanarak veri setindeki yapıyı incelemiş ve bu iki farklı analiz yönteminin sonuçlarını karşılaştırarak araştırmanın genel amacını oluşturmuştur.

Bu çalışma Kaggle üzerinden “Unsupervised Learning on Country Data” başlıklı veri seti alınarak yapılmıştır. Veri setinin ülkeler bazında genel gelişimini belirleyen sosyo-ekonomik ve sağlık faktörlerini kullanarak kategorize edilmesi amaçlanmıştır. Bu bağlamda 167 ülke üzerinden bir değerlendirme söz konusudur. Veri setinde bulunan değişkenler sırasıyla çocuk ölüm oranı, ihracat, sağlık, ithalat, gelir, enflasyon, ortalama yaşam süresi, doğurganlık oranı ve kişi başına gayri safi yurtiçi hasıla (GDPP) şeklinde 9 değişkendir.

Tablo 3. Veri setinde kullanılan değişkenler

Değişkenler	Değişken Adı	Açıklama
Çocuk Ölümü	$X_1$	Her 1000 doğumda, 5 yaş altı ölen çocuk ölümleri.
İhracat	$X_2$	Kişi başına düşen mal ve hizmet ihracatı. Kişi başına düşen GSYİH'nin yüzdesi olarak verilmiştir.
Sağlık	$X_3$	Kişi başına düşen toplam sağlık harcamaları. GSYİH'nin yüzdesi olarak verilmiştir.
İthalat	$X_4$	Kişi başına düşen mal ve hizmet ithalatı. Kişi başına düşen GSYİH'nin yüzdesi olarak verilmiştir.
Gelir	$X_5$	Kişi başına net gelir.
Enflasyon	$X_6$	Toplam GSYİH'nin yıllık büyüme oranı
Beklenen Yaşam Süresi	$X_7$	Mevcut ölüm oranları aynı kalsa, yeni doğan bir çocuğun ortalama yaşayacağı yıl sayısı
Toplam Doğurganlık	$X_8$	Mevcut yaş-doğurganlık oranları aynı kalsa, herkadına doğacak çocuk sayısı
GDPP	$X_9$	Kişi başına düşen GSYİH. Toplam GSYİH'nin toplam nüfusa bölünmesiyle hesaplanır.

## 2.1. Faktör Analizi Uygulaması

Bu başlık altında ilk olarak veri seti SPSS ile faktör analizine sokulup kaç faktöre indirgendiği hesaplanmıştır ve isimlendirilmiştir. İlk aşamada SPSS ile faktör analizi incelemesi yapılmıştır. Bu bağlamda ilk olarak KMO değeri hesaplanmış olup, bu değer 0,68 ile orta seviyede çıkmıştır. Tablo 4'te görüldüğü gibi KMO değerinin faktör analizi için uygun olduğunu göstermektedir.

Tablo 4. Kaiser-Meyer-Olkin ve Bartlett Testi

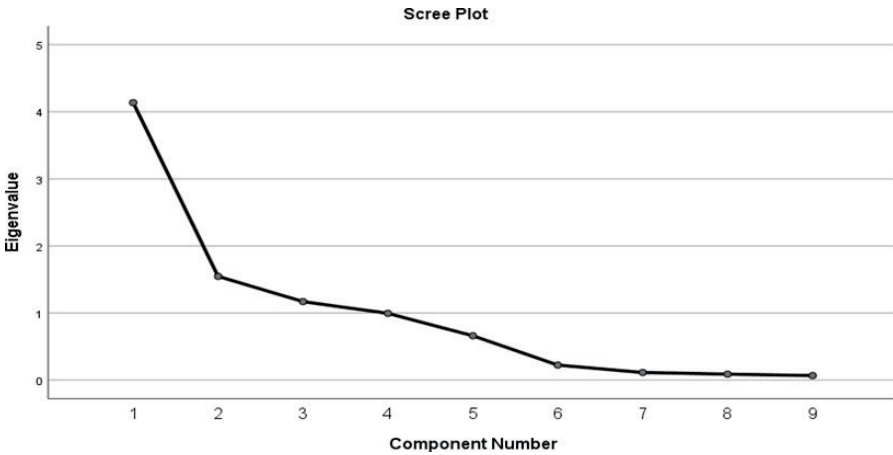
Kaiser-Meyer-Olkin ve Bartlett Testi		
Kaiser-Meyer-Olkin Örnekleme Yeterliliğinin Ölçüsü.	0,678	
Bartlett'in Küresellik Testi	Tahmini Ki-Kare	1169,737
	Serbestlik Derecesi	36
	p-değeri	0,000

Sonrasında ise değişkenler incelenmiş olup bu kapsamda veri setindeki 9 değişkenin de analiz sonucunda kalması kararına varılmıştır. Buna Tablo 5’de görüldüğü gibi 0,5 altında değer olmaması sebebiyle karar verilmiştir.

*Tablo 5. Toplam Varyans*

Değişkenler	Initial	Extraction
cocuk_olumu	1,000	0,786
ihracat	1,000	0,939
saglik	1,000	0,602
ithalat	1,000	0,911
gelir	1,000	0,764
enflasyon	1,000	0,638
yasam_suresi	1,000	0,842
toplam_dogurganlik	1,000	0,712
gdpp	1,000	0,659

Faktör analizine geçiş yapılmasının ardından ise Şekil 6’da gösterilmiş olan Scree Plot ile 3 faktöre indirgenmesine gerektiği gözlemlenmektedir. Scree Plot’a bakıldığında faktör sayısının 3 olarak belirlenmesinin sebebi özdeğerin 1’den büyük olan değerler kadar faktöre indirgeneyeceğidir. Bu kapsamda özdeğerin 1’den büyük olduğunu gösteren 3 nokta mevcuttur. Bu bilgilere ek olarak indirgenen faktörlerin açıklanan toplam varyansı ise Tablo 7 ile gösterilmektedir.



*Şekil 6. Scree plot*

Tablo 7. Açıklanan toplam varyans

Başlangıç Özdeğerleri				Özütleme Karekök Yüklemeleri Toplamı		
Bileşen	Toplam	Varyans Yüzdəsi	Kümülatif %	Toplam	Varyans Yüzdəsi	Kümülatif %
1	4,136	45,952	45,952	3,666	40,734	40,734
2	1,546	17,182	63,133	1,794	19,938	60,672
3	1,170	13,004	76,138	1,392	15,466	76,138
4	,995	11,053	87,191			
5	,661	7,340	94,531			
6	,224	2,484	97,015			
7	,113	1,260	98,276			
8	,088	,981	99,257			
9	,067	,743	10,000			

Şekil 6'da ve Tablo 7'de görüldüğü gibi faktör analizi ile 9 değişkenden 3 faktöre indirgenmiştir. Bu bağlamda ilk faktörün açıklanan toplam varyansı %40,7 iken ikincinin %19,9 olduğu gözükmektedir. Bu bağlamda ilk iki faktörün açıklanabilir toplam varyansının %60,6 olduğu gözlemlenmektedir. Üçüncü faktörün ise %15,5 iken üç faktörün toplam açıklanabilir varyans oranının %76,1 olduğu gözlemlenmektedir. Faktör analizinde Varimax dönüşümü kullanılmıştır.

Tablo 8. SPSS ile faktör analizinden elde edilen faktörler

	Sosyal	Ekonomi	Sağlık
yasam_suresi	0,902	-0,005	-0,169
cocuk_olumu	0,852	-0,037	0,244
gelir	0,834	0,241	0,100
toplam_dogurganlik	0,793	-0,078	0,276
gdpp	0,785	0,180	-0,097
ithalat	-0,022	0,929	-0,219
ihracat	0,375	0,880	0,154
enflasyon	-0,138	-0,216	0,756
saglik	0,158	-0,111	0,752

Mevcut 9 adet değişkenin 3 faktör ile temsil edilebileceğine karar verilir. Bu 3 faktör Tablo 8’de gösterilmiştir. Bu faktörlerin isimlendirilmesi ise değişkenlere bağlı olarak yapılmıştır. Bu bağlamda faktör 1’i temsilen sosyal faktörü olarak isimlendirilmiştir. Sosyal faktöründeki değişkenler ise sırasıyla yaşam süresi, çocuk ölümü, gelir, toplam doğurganlık ve gdp (kişi başına düşen gayri safi yurtiçi hasılat) süresi şeklindedir. Faktör 2’yi temsilen ekonomi faktörü ele alınmıştır. Bahsi geçen faktördeki değişkenler ise ithalat ve ihracat değişkenleridir. Son olarak faktör 3’ü temsilen ise sağlık faktöründe sadece sağlık değişkeni bulunmaktadır.

## 2.2. Ortak Entropi Uygulaması

Bu başlık altında ilk olarak veri seti Python ve Excel yardımları ile kaç faktöre indirgendiği hesaplanmıştır ve isimlendirilmiştir. Bu bağlamda ilk olarak veri setine ait olan 9 değişkenin tek tek entropisi hesaplanmaktadır. Bu değerler sırasıyla Tablo 9’da gösterilmiştir.

*Tablo 9. Değişkenlere ait entropi değerleri*

Değişken	Entropi Değeri
$X_1$	7,0183
$X_2$	7,1351
$X_3$	7,1351
$X_4$	7,1921
$X_5$	7,2520
$X_6$	7,2474
$X_7$	6,8775
$X_8$	6,9944
$X_9$	7,2520

Tablo 10’da ise değişken çiftleri için hesaplanan birleşik entropi değerleri ve değişken hakkında kararlar gösterilmektedir.

Tablo 10. SPSS ile faktör analizinden elde edilen faktörler

	$H(X_\alpha, X_\beta)$	$\frac{H(X_\alpha) + H(X_\beta)}{H(X_\alpha, X_\beta)}$	Değişkenler Hakkında Karar
$X_1, X_2$	7,952450	1,7798	Bağımlı.
$X_1, X_3$	7,998094	1,7696	Bağımlı.
$X_1, X_4$	7,954710	1,7864	Bağımlı.
$X_1, X_5$	8,135143	1,7542	Bağımlı.
$X_1, X_6$	8,029995	1,7766	Bağımlı.
$X_1, X_7$	7,890984	1,7610	Bağımlı
$X_1, X_8$	7,968946	1,7584	Bağımlı
$X_1, X_9$	8,135143	1,7542	Bağımlı
$X_2, X_3$	8,117179	1,7580	Bağımlı.
$X_2, X_4$	7,964127	1,7990	Bağımlı.
$X_2, X_5$	8,193555	1,7559	Bağımlı.
$X_2, X_6$	8,117179	1,7719	Bağımlı.
$X_2, X_7$	7,949890	1,7626	Bağımlı
$X_2, X_8$	8,056506	1,7538	Bağımlı
$X_2, X_9$	8,193555	1,7559	Bağımlı
$X_3, X_4$	8,157627	1,7563	Bağımlı.
$X_3, X_5$	8,193555	1,7559	Bağımlı.
$X_3, X_6$	8,092434	1,7773	Bağımlı.
$X_3, X_7$	8,006342	1,7502	Bağımlı

$X_3, X_8$	7,978662	1,7709	Bağımlı
$X_3, X_9$	8,193555	1,7559	Bağımlı
$X_4, X_5$	8,222028	1,7568	Bağımlı.
$X_4, X_6$	8,169603	1,7675	Bağımlı.
$X_4, X_7$	7,889517	1,7833	Bağımlı
$X_4, X_8$	8,093227	1,7529	Bağımlı
$X_4, X_9$	8,222028	1,7568	Bağımlı
$X_5, X_6$	8,249708	1,7576	Bağımlı.
$X_5, X_7$	8,064754	1,7520	Bağımlı
$X_5, X_8$	8,123167	1,7538	Bağımlı
$X_5, X_9$	8,151639	1,7793	Bağımlı
$X_6, X_7$	8,062494	1,7519	Bağımlı
$X_6, X_8$	7,995534	1,7812	Bağımlı
$X_6, X_9$	8,249708	1,7576	Bağımlı
$X_7, X_8$	7,935953	1,7480	Bağımlı
$X_7, X_9$	8,064754	1,7520	Bağımlı
$X_8, X_9$	8,123167	1,7538	Bağımlı

Tablo 9'da görüldüğü gibi ortak entropi değerleri hesaplanmıştır. Bu bilgiler ışığında bağımlı olup olmadığı da kontrol edilmiştir. Birleşik entropi değerlerinin 1 ile 2 arasında olması beklenmektedir (Çilingir, 2019). Bir sonraki adım ise birleşik entropi matrisinin elde edilmesidir. Birleşik entropi matrisi Tablo 11'de gösterilmektedir.

**Tablo 11. Birleşik Entropi Matrisi**

	1	2	3	4	5	6	7	8	9
1	7,01832	7,95245	7,99809	7,95471	8,13514	8,03000	7,89098	7,96895	8,13514
2	7,95245	7,13514	8,11718	7,96413	8,19356	8,11718	7,94989	8,05651	8,19356
3	7,99809	8,11718	7,13514	8,15763	8,19356	8,09243	8,00634	7,97866	8,19356
4	7,95471	7,96413	8,15763	7,19209	8,22203	8,16960	7,88952	8,09323	8,22203
5	8,13514	8,19356	8,19356	8,22203	7,25197	8,24971	8,06475	8,12317	8,15164
6	8,03000	8,11718	8,09243	8,16960	8,24971	7,24745	8,06249	7,99553	8,24971
7	7,89098	7,94989	8,00634	7,88952	8,06475	8,06249	6,87754	7,93595	8,06475
8	7,96895	8,05651	7,97866	8,09323	8,12317	7,99553	7,93595	6,99437	8,12317
9	8,13514	8,19356	8,19356	8,22203	8,15164	8,24971	8,06475	8,12317	7,25197

Birleşik entropi matrisi yardımıyla entropi-korelasyon matrisine geçiş yapılacaktır. Bu bağlamda entropi-korelasyon matrisi Tablo 12’de gösterilmektedir.

**Tablo 12. Entropi- Korelasyon Matrisi**

	1	2	3	4	5	6	7	8	9
1	1,00000	0,13110	0,03670	0,16343	-0,20953	0,02072	0,12739	0,02144	-0,20953
2	0,13110	1,00000	-0,14976	0,22976	-0,27336	-0,10754	0,09559	-0,09307	-0,27336
3	0,03670	-0,14976	1,00000	-0,20145	-0,24017	-0,01110	-0,02150	0,10495	-0,24017
4	0,16343	0,22976	-0,20145	1,00000	-0,29959	-0,19006	0,24392	-0,13526	-0,29959
5	-0,20953	-0,27336	-0,24017	-0,29959	1,00000	-0,35093	-0,07466	-0,15029	-0,05301
6	0,02072	-0,10754	-0,01110	-0,19006	-0,35093	1,00000	-0,08959	0,11875	-0,35093
7	0,12739	0,09559	-0,02150	0,24392	-0,07466	-0,08959	1,00000	0,03768	-0,07466
8	0,02144	-0,09307	0,10495	-0,13526	-0,15029	0,11875	0,03768	1,00000	-0,15029
9	-0,20953	-0,27336	-0,24017	-0,29959	-0,05301	-0,35093	-0,07466	-0,15029	1,00000

Entropi-korelasyon matrisinin elde edilmesinin ardından ise 9 değişkenden kaç tane faktöre indirgeneceği bulunacaktır. Faktör analizinde de olduğu gibi buradaki işlemlerde de Varimax dönüşümü kullanılmıştır. Mevcut 9 adet değişkenin 3 faktör ile temsil edilebileceğine karar verilir. Bu 3 faktör Tablo 13’de gösterilmiştir. Bu faktörlerin isimlendirilmesi ise değişkenlere bağlı olarak yapılmıştır.

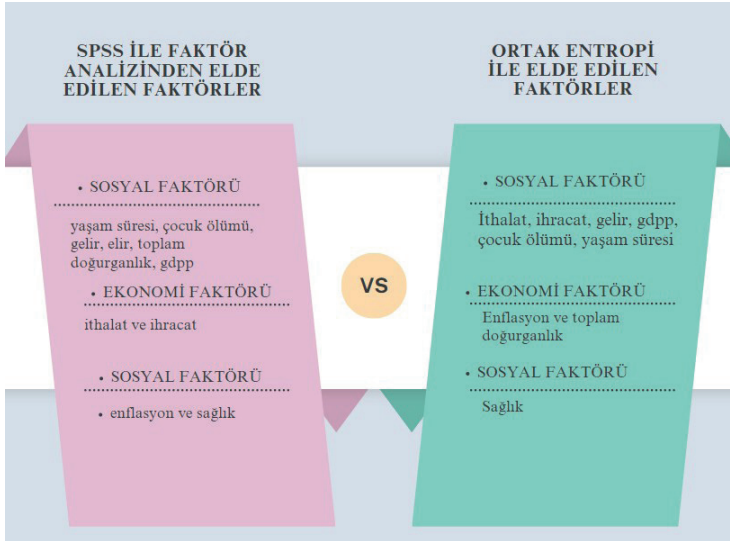


Tablo 13. Ortak entropi ile elde edilen faktörler

	Sosyal	Ekonomi	Sağlık
ithalat	0,765	-0,494	0,070
ihracat	0,613	-0,287	-0,018
gelir	0,603	-0,255	-0,005
gdpp	0,603	-0,255	-0,005
cocuk_olumu	0,539	0,005	0,072
yasam_suresi	0,340	-0,270	0,115
enflasyon	0,332	0,767	-0,566
toplam_dogurganlik	0,034	0,443	-0,018
sağlık	0,090	0,693	0,720

Faktör 1'i temsilen toplum faktörü olarak isimlendirilmiştir. Sosyal faktöründeki değişkenler ise sırasıyla ithalat, ihracat, gelir, ithalat, gdpp (kişi başına düşen gayri safi yurtiçi hasılat), çocuk ölümü ve beklenen yaşam süresi şeklindedir. Faktör 2'yi temsilen ekonomi faktörü ele alınmıştır. Bahsi geçen faktördeki değişkenler ise enflasyon ve toplam doğurganlık değişkenleridir. Son olarak faktör 3'ü temsilen ise sağlık faktöründe sadece sağlık değişkeni bulunmaktadır.

Sonuç olarak yapılan bu iki uygulama sonrasında elde edilmiş olan faktörlerin karşılaştırılması Şekil 14'te gösterilmiştir.



Şekil 14. Faktör Analizi ve Ortak Entropi yöntemlerinden elde edilen faktörlerin karşılaştırılması

### 3. Sonuç ve Tartışma

Çalışma kapsamında öncelikli hedef, istatistiksel anlamda entropi bazı yaklaşımların da uygulamalarda bir alternatif olarak kullanılabilmesinin gösterilmesi hedeflenmiştir. Bu bağlamda ilk olarak entropi kavramı hakkında bilgiler verilmiş ve ortak entropi konusu ele alınmıştır. Bu bağlamda değişkenler arası ilişki entropi ölçüsü ile incelenmiştir ve bu yöntem ile birleşik entropi matrisi kullanılarak boyut indirgemenin sağlanabileceği incelenmiştir. Bu bilgilere ek olarak bibliyometrik analiz üzerinde durulmuş ve bu amaç doğrultusunda yapılmış olan çalışmalara ait yazarlar, anahtar kelimeler ve literatür bilgileri kazanılmıştır.

Uygulama bölümünde ülkeler bazında genel gelişimini belirleyen sosyo-ekonomik ve sağlık faktörlerini kullanarak kategorize edilmiş değişkenlerin bulunduğu veri seti ile farklı analiz yöntemleri kullanılarak karşılaştırma yapılmıştır. İlk olarak veri seti faktör analizinin sonuçlarıyla entropi tabanlı yaklaşımın sonuçlarının karşılaştırması amacıyla faktör analizinin sonuçları elde edilmiştir. Ardında ise değişkenlere ait entropi ölçüleri ile elde edilen birleşik entropi matrisi yardımıyla boyut indirgeme sağlanıp sağlanmadığı teyit edilmiştir. Bu bilgiler doğrultusunda elde edilen bilgiler ile faktör analizinden elde edilen bilgilerin karşılaştırılması yapılmıştır.

Kullanılan veri setiyle entropi yöntemi sayesinde elde edilen birleşik entropi matrisi ile faktör indirgeme yöntemi gerçekleştirilmesi hedeflenmiştir. Bu sayede birleşik entropi yöntemiyle faktör indirgemenin sağlanıp sağlanmadığı araştırılmıştır.

Sonuç olarak bu çalışma, entropinin teorik bir temelini ve varyans-kovaryans ölçümlerini kullanmanın yararlı olduğu alanlara ek olarak katkıda bulunmuştur. Bu çalışmanın, entropinin önemini ve entropiye dayalı araştırma yöntemlerini vurguladığı düşünüldüğünde, istatistik alanında entropi konusuna daha fazla önem verilmesi ve bu alanda daha fazla çalışma yapılması önerilmektedir. Bununla birlikte, çalışmanın ortaya çıkan zorluklar, akademik alanda yeni çalışmaların gerekli olduğunu göstermektedir. Çözülmesi gereken bazı sorunlara rağmen, alternatif olarak bilinen yöntemlerin yerine önerilmektedir.

## Kaynakça

- Altaylıgil, Y. B. (2008). Entropi ölçüsü ve bazı ekonometri uygulamaları.
- Bahadır, O., & Türkmençalıkoğlu, H. (2021). Bilgi Kuramında Shannon Entropisi ve Uygulamaları. *Avrupa Bilim ve Teknoloji Dergisi*, 32, 497-497.
- Büyükoztürk, Ş. (2002). Faktör Analizi: Temel Kavramlar ve Ölçek Geliştirme de Kullanımı. *Kuram ve uygulamada eğitim yönetimi*, 32(32), 470-483.
- Cover, T. M. (1999). *Elements Of Information Theory*. John Wiley & Sons.
- Çilingir, N. B. (2019). Faktör Analizi: Teori ve Uygulama [Master's thesis]. Sosyal Bilimler Enstitüsü.
- Değirmenci, İ. (2011). Entropi Ölçümleri ve Maksimum Entropi İlkesi (Doctoral dissertation, Yüksek Lisans Tezi, Hacettepe Üniversitesi). [Yüksek Lisans Tezi]. Hacettepe Üniversitesi.
- Erbaş, Ü. (2010). Entropi ilkelerinin boyut indirgeme uygulamaları. <https://search.proquest.com/openview/c98c77beacb361f03877297736fc38be/1?pq-origsite=gscholar&cbl=2026366&cdiss=y>
- Koşan, M. A., Coşkun, A., & Karacan, H. (2019). Yapay Zeka Yöntemlerinde Entropi. *Journal of Information Systems and Management Research*, 1(1), 15-22.
- Kullback, S. (1960). *Information Theory and Statistics*.
- Oruç, N. (2024). Faktör analizi çözümlemesinde birleşik entropi yaklaşımı (Yüksek lisans tezi), Yıldız Teknik Üniversitesi
- Öncül, E. (2023). Çok Değişkenli Yöntemlerde Entropi Kullanımı: Mutluluk İndeksi Üzerine Bir Uygulama. Yıldız Teknik Üniversitesi.
- Shannon, C. E. (1948). A Mathematical Theory Of Communication. *The Bell system technical journal*, 27(3), 379-423.
- Tatlıdil, H. (1992). *Uygulamalı Çok Değişkenli İstatistiksel Analiz*. Engin Yayınları.
- Wei, Y. (1987). *Variance, Entropy, and Uncertainty Measure*. Dept. of Statistics, People's Univ. of China.