

**ELEKTRONİK POSTALARIN AYRIŞTIRILMASINDA  
NAİVE BAYESIAN VE BULANIK MANTIK YÖNTEMLERİNİN  
KARŞILAŞTIRILMASI**

**Burhan YUMAK**

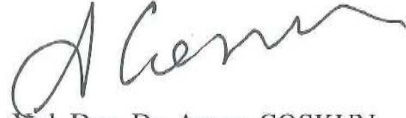
**YÜKSEK LİSANS TEZİ  
BİLGİSAYAR EĞİTİMİ**

**GAZİ ÜNİVERSİTESİ  
BİLİŞİM ENSTİTÜSÜ**

**HAZİRAN 2011**

**ANKARA**

Burhan YUMAK tarafından hazırlanan ELEKTRONİK POSTALARIN AYRIŞTIRILMASINDA NAİVE BAYESIAN VE BULANIK MANTIK YÖNTEMLERİNİN KARŞILAŞTIRILMASI adlı bu tezin Yüksek Lisans tezi olarak uygun olduğunu onaylarım.



Yrd. Doç. Dr. Aysun COŞKUN

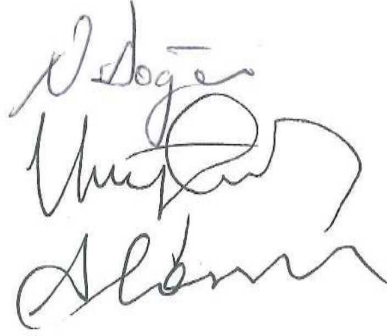
Tez Yöneticisi

Bu çalışma, jürimiz tarafından oy birliği / oy çokluğu ile Bilgisayar Eğitimi Anabilim Dalında Yüksek lisans tezi olarak kabul edilmiştir.

Başkan: : Yrd. Doç. Dr. Nurettin DOĞAN

Üye : Yrd. Doç. Dr. Hüseyin ÇAKIR

Üye : Yrd. Doç. Dr. Aysun COŞKUN



Tarih : 13/06/2011

Bu tez, Gazi Üniversitesi Bilişim Enstitüsü tez yazım kurallarına uygundur.

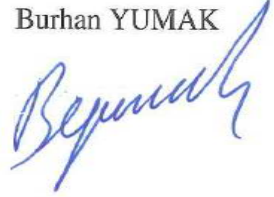


Prof. Dr. Halim SONCUL  
Enstitü Müdürü

## TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada orijinal olmayan her türlü kaynağa eksiksiz atıf yapıldığını bildiririm.

Burhan YUMAK



**ELEKTRONİK POSTALARIN AYRIŞTIRILMASINDA  
NAIVE BAYESIAN VE BULANIK MANTIK YÖNTEMLERİNİN  
KARŞILAŞTIRILMASI  
(Yüksek Lisans Tezi)**

**Burhan YUMAK**

**GAZİ ÜNİVERSİTESİ  
BİLİŞİM ENSTİTÜSÜ  
Haziran 2011**

**ÖZET**

İnternet teknolojilerinin sağlamış olduğu olanaklardan en büyüklerinden biri olan elektronik posta ile haberleşmedir. Ancak kullanıcı sayılarının artması, artan elektronik posta sayıları, elektronik posta üzerinden yeni kazançlar elde edilmesi, elektronik reklamlar gibi nedenlerden dolayı SPAM posta kavramı ortaya çıkmıştır. SPAM postaların artış göstermesi kullanıcılara için önemli oranda zarar vermeye başlamıştır. Kullanıcılara zaman kaybettirmesi, rahatsız edici içeriğe sahip olması, internet üzerinden dolandırıcılığa zemin oluşturması, servis sağlayıcılar için ise aşırı yük taşıyor olması SPAM postaların verdiği zararlardandır. Yapılan araştırmalarda SPAM postaların istenmeyen bir durum olduğunu göstermiştir. Bu nedenle SPAM postaların elektronik postalardan ayrıştırılması büyük bir ihtiyaç haline gelmiştir. Gelen SPAM posta sayısı çok fazla olduğundan, çözüm için yapay zekâ uygulamaları denenmiştir. Kullanılan yöntemlerin performansları, yetenekleri ve ayrıştırma gücü önemli olduklarından dolayı bu konuda da araştırmalara ihtiyaç vardır.

Bu çalışmada, elektronik postaların SPAM postalardan ayrıştırılması üzerine araştırmalar yapılmıştır. SPAM postaların ayrıştırılmasında sık kullanılan iki farklı yöntem olan Naive Bayesian ve Bulanık Mantık yöntemleri incelenmiştir. Bir simülatör uygulama geliştirilerek sanal bir posta sunucu yapısı

oluřturulmuř, 6nceden eklenmiř sanal kullanıcılar tarafından posta g6nderim iřlemleri yapılmıřtır. Seilen bir kullanıcıya gelen elektronik postalar 6zerinden analizler yapılmıřtır. Kullanıcıya gelen postalar ile 6rnek veri sınıflandırma tablosu ıkarılmıřtır. Bu veri tablosu 6zerinden kullanıcıya ait bir SPAM olasılık tablosu ıkarılmıřtır. Bu y6ntemlerin 6nemli 6zellięi de SPAM olasılıklarının kullanıcıya daha 6nceden gelen postalar 6zerinden ıkıyor olmasıdır. Her iki y6ntem iin de kullanıcıya gelen postalar 6zerinden ayrı ayrı SPAM durumları ıkarılmıřtır. Uygulama sonularından elde edilen verilere g6re SPAM postaların ayrıřtırılmasında Naive Bayesian y6nteminin daha etkili olduęu g6r6lm6řt6r.

**Bilim Kodu** : 702.3.006  
**Anahtar Kelime** : spam, naive bayes, istenmeyen posta, bulanık mantık  
**Sayfa Adedi** : 80  
**Tez Y6neticisi** : Yrd.Do.Dr. Aysun Cořkun

**ANALYSIS OF ELECTRONIC MAIL,  
COMPARISON OF METHODS  
FUZZY LOGIC AND NAIVE BAYESIAN  
(M.Sc. Thesis)**

**Burhan YUMAK**

**GAZİ UNIVERSITY  
INFORMATICS INSTITUTE**

**June 2011**

**ABSTRACT**

**One of the main facilities that internet technologies provide is accepted as communication via e-mail. But within the rise in users, increase in the number of e-mails, new methods of earnings in e-mails and electronic advertisements, a new definition “SPAM” came into the picture. The irresistible rise in SPAM e-mails started disturbing the users in a serious manner. Some of the main troubles can be listed as waste of time for users, disturbing content, letting fakery via internet and for causing capacity problems for service providers. Common researches show that SPAM e-mails are definitely unwanted by the users. Therefore, separating the SPAM e-mails from the other ones became an absolute necessity in the internet society. Because the number of the SPAM e-mails are considerably high, some artificial intelligence methods are tested before. There is a need to keep on analyzing these methods since the performance, capabilities and resolution of the methods are extremely important.**

**In this thesis, a research of separating SPAM e-mails from the standart e-mails is completed. The two methods for separating e-mails form SPAM e-mails, Naive Bayesian and Fuzzy Logic are investigated. First, a virtual e-mail server provider is generated by a simulator application and sending e-mails by the**

virtual users are carried out. After that, the analysis are made among a chosen user. The e-mails that are recieved by that user are used to generate a sample data classification chart. By using this chart, a SPAM probability table is derived. The important characteristic of this method is that the SPAM probabilities are derived among the e-mails that are recieved by the user before. Two different SPAM states are reviewed fort he two different methods. By evaluating the datas obtained by the application results, Naive Bayesian method can be seen as more efficient in seperating the SPAM e-mails.

**Science Code** : 702.3.006  
**Key Words** : spam, naive bayesian, email, fuzzy logic  
**Page Number** : 80  
**Adviser** : Assist. Prof. Dr. Aysun Coşkun

**TEŐEKKÜR**

Çalıőmalarım boyunca deęerli yardım ve katkılarıyla ben yönlendiren Hocam Yrd. Doç. Dr. Aysun Coőkun' a, benim bu aőamaya kadar gelmemde hiçbir emeęini esirgemeyen aileme ve hep desteęiyle yanımda olan eőime teőekkürü bir borç bilirim.



## İÇİNDEKİLER

	<b>Sayfa</b>
ÖZET.....	iv
ABSTRACT.....	vi
TEŞEKKÜR.....	viii
İÇİNDEKİLER .....	ix
ÇİZELGELERİN LİSTESİ.....	xii
RESİMLERİN LİSTESİ .....	xiii
ŞEKİLLERİN LİSTESİ .....	xiv
SİMGELER VE KISALTMALAR.....	xvi
1. GİRİŞ .....	1
2. ELEKTRONİK POSTA VE SPAM .....	3
2.1. Elektronik Posta Sistemi .....	3
2.2. E-Posta Programlarında Görülen Terimler .....	3
2.2.1. Gönderen ya da kimden (from).....	3
2.2.2. Alıcı ya da kime (to) .....	4
2.2.3. Konu (subject).....	4
2.2.4. E-posta bedeni (body) .....	4
2.2.5. Karbon kopya "kk" (cc, carbon copy).....	4
2.2.6. Gizli karbon kopya "gkk" (bcc, blind carbon copy) .....	4
2.2.7. Cevaplama (reply).....	5
2.2.8. Herkesi cevapla (reply all) .....	5
2.2.9. Yönlendirme (forward, fwd).....	5
2.2.10. Ek (attachment) .....	6
2.3. Spam Posta.....	6
2.3.1. SPAM postaların sınıflandırma değerleri.....	8
2.3.2. SPAM postaların içerikleri.....	8
2.3.3. SPAM postaların zararları.....	9
2.4. Türkiye’ de SPAM Postalar Üzerine Yapılan Çalışmalar.....	10
2.4.1. İstenmeyen iletilerin paralelleştirilmiş knn algoritması ile tespiti ..11	11

**Sayfa**

2.4.2. Türkçe morfolojik çözümleme, yapay sinir ağları ve bayes filtreleme tabanlı uyarlamalı spam-önler filtrelemesi .....	11
<b>3. NAIVE BAYESIAN TEOREMİ VE SPAM POSTALARIN AYRIŞTIRILMASI .....</b>	<b>12</b>
3.1. Naive Bayesian Teoremi .....	12
3.2. Naive Bayesian Yönteminin Ana Fikri .....	14
3.3. Naive Bayesian Kuralının Çalışma Yapısı.....	15
3.4. Bayes Teoreminin Değişik Şekilleri .....	17
3.4.1. Bahis oranı ve olabirliklik orantısı şeklinde bayes teoremi .....	18
3.4.2. Olasılık yoğunluk fonksiyonları ile bayes teoremi .....	19
3.4.3. Soyut bayes teoremi .....	20
3.5. Bayes Teoreminin Kapsamının Genişletilmesi.....	20
3.6. Naive Bayes Yöntemi İle İlgili Örnek Uygulamalar.....	21
<b>4. BULANIK MANTIK TEOREMİ VE İSTENMEYEN POSTALARIN AYRIŞTIRILMASI .....</b>	<b>28</b>
4.1. Bulanık Mantık ( Fuzzy Logic ).....	28
4.2. Bulanık Sistem .....	31
4.3. Üyelik Fonksiyonları.....	35
4.3.1. Üyelik fonksiyonunun kısımları.....	39
4.3.2. Bulanıklaştırma .....	43
4.3.3. Durulaştırma.....	44
4.3.4. Durulaştırma işlemleri.....	45
4.4. Bulanık Mantık Yöntemi ile Örnek Uygulama.....	53
<b>5. SPAM POSTALARIN NAIVE BAYESIAN TEOREMİ VE BULANIK MANTIK YÖNTEMLERİYLE AYRIŞTIRILMASININ KARŞILAŞTIRILMASI.....</b>	<b>58</b>
5.1. Uygulama Üzerinden Naive Bayes ve Bulanık Mantık Yöntemlerinin Kullanılması .....	59
5.1.1. SPAM postaların ayrıştırılmasında sınıflandırma değerleri.....	62
5.1.2. Gelen postalara ait oluşturulan veri kümeleri .....	65

5.1.3. Gelen postalara ait oluşturulan olasılıklar.....	67
5.1.4. Naive Bayesian yöntemine göre elektronik postanın ayrıştırılması.....	70
5.1.5. Bulanık mantık yöntemine göre elektronik postanın ayrıştırılması.....	71
5.2. Naive Bayesian ve Bulanık Mantık Yöntemine Göre Elektronik Posta Ayrıştırılmasının Karşılaştırılması .....	70
6. SONUÇLAR VE ÖNERİLER .....	77
KAYNAKLAR .....	79
ÖZGEÇMİŞ .....	81

## ÇİZELGELERİN LİSTESİ

<b>Çizelge</b>	<b>Sayfa</b>
Çizelge 2.1. SPAM e-postaların 2003 – 2004 yılları arasındaki içerik değişimi...8	
Çizelge 2.2. 2003 yılında 1272 e-posta kullanıcısı üzerinde yapılan bir araştırmanın sonuçları .....9	
Çizelge 3.1. Daha önceden yapılan şutların veri kümesi .....21	
Çizelge 3.2. Daha önceden yapılan şut şekillerine göre gol durum frekansları...22	
Çizelge 3.3. Gol durum frekanslarına göre gol olasılıkları .....23	
Çizelge 3.4. Adayların kabul durumları veri kümesi .....24	
Çizelge 3.5. Aday bilgilerine göre kabul durum frekansları .....25	
Çizelge 3.6. Aday bilgi frekanslarına göre kabul durum olasılıkları .....26	
Çizelge 4.1. En büyük üyelik derecesi durulaştırması .....40	
Çizelge 5.1. Gelen postaların SPAM sınıflandırma değerleri ve durumları .....64	
Çizelge 5.2. Kara liste durumuna göre veri frekans tablosu .....65	
Çizelge 5.3. Kelime durumuna göre veri frekans tablosu .....65	
Çizelge 5.4. Domain durumuna göre veri frekans tablosu .....66	
Çizelge 5.5. İsim geçme durumuna göre veri frekans tablosu .....66	
Çizelge 5.6. Önceden gelme durumuna göre veri frekans tablosu .....67	
Çizelge 5.7. Kara liste durumuna göre olasılık tablosu .....67	
Çizelge 5.8. Kelime durumuna göre olasılık tablosu .....68	
Çizelge 5.9. Domain durumuna göre olasılık tablosu .....68	
Çizelge 5.10. İsim geçme durumuna göre olasılık tablosu .....68	
Çizelge 5.11. Önceden gelme durumuna göre olasılık tablosu .....69	
Çizelge 5.12. Yeni gelen postanın sınıflandırma değerleri .....69	
Çizelge 5.13. Kullanıcıya gelen ve olması gereken SPAM e-posta sayıları .....75	
Çizelge 5.14. Kullanıcıya gelen e-postaların ayrıştırılmasında Naive Bayesian ve Bulanık Mantık yöntemlerinin SPAM ayrıştırma sonuçları .....75	

**RESİMLERİN LİSTESİ**

<b>Resim</b>	<b>Sayfa</b>
Resim 5.1. Uygulama ana ekran görüntüsü .....	58
Resim 5.2. Uygulama veritabanı genel görüntüsü .....	59
Resim 5.3. Sisteme giriş yapan kullanıcının diğer kullanıcıya e-posta göndermesi .....	60
Resim 5.4. Kullanıcıya gelen e-postanın okunması ve SPAM olarak işaretlenmesi.....	60
Resim 5.5. Kullanıcı gelen e-posta kutusu.....	61
Resim 5.6. Kullanıcı SPAM e-posta kutusu .....	61

## ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 4.1. Bulanık mantık modeli.....	29
Şekil 4.2. Klasik mantık modeli.....	29
Şekil 4.3. Klasik sistem.....	31
Şekil 4.4. Genel bulanık sistem.....	31
Şekil 4.5. TSK bulanık sistem.....	33
Şekil 4.6. Bulanıklaştırma-durulaştırma birimli bulanık sistem .....	34
Şekil 4.7. Bitişik dikdörtgen gösterimi .....	35
Şekil 4.8. Bitişik üçgen gösterimi .....	36
Şekil 4.9. Örtüşmeli üçgen gösterimi.....	37
Şekil 4.10. Bulanık küme .....	38
Şekil 4.11. Yamuk ve çan eğrisi üyelik fonksiyonları .....	39
Şekil 4.12. Bulanık kümeler, (a) normal, (b) normal olmayan .....	40
Şekil 4.13. Bulanık kümeler, (a) dış bükey, (b) dış bükey olmayan .....	34
Şekil 4.14. Dış bükey olmayan kümelerin kesişimi.....	41
Şekil 4.15. Gauss (a) bulanık kümesi (b) dağılım fonksiyonu.....	43
Şekil 4.16. Hassaslık (a) bulanık (b) klasik.....	44
Şekil 4.17. İki bulanık kümenin (a) birleşimi, (b) kesişimi .....	45
Şekil 4.18. Tipik bulanık küme çıktısı, (a) bulanık girdi ilk kısım, (b) bulanık girdi ikinci kısım, (c) ikisinin birleşimi.....	46
Şekil 4.19. En büyük üyelik derecesi durulaştırması .....	47
Şekil 4.20. Sentroid yöntemi ile durulaştırma.....	47
Şekil 4.21. Ağırlıklı ortalama yöntemi durulaştırması.....	49
Şekil 4.22. Ortalama en büyük üyelik durulaştırılması.....	50
Şekil 4.23. Toplamların merkezi durulaştırması .....	51
Şekil 4.24. En büyük alan merkezi ile durulaştırma .....	52
Şekil 4.25. İlk ve son en büyük üyelik dereceleri ile durulaştırma .....	53
Şekil 5.1. Elektronik postaların bulanık ilişkisi .....	71
Şekil 5.2. Yeni gelen elektronik postanın özellikleri .....	72

Şekil 5.3. Kullanıcıya gelen ve örnek alınan postanın Naive Bayesian ve Bulanık Mantık yöntemine göre ayrıştırma oranları.....	74
Şekil 5.4. Beklenen SPAM posta sayısı ile Naive Bayesian ve Bulanık Yöntemi ile ayrıştırılan SPAM postaların karşılaştırılması .....	76
Şekil 5.5. Naive Bayesian ve Bulanık Yöntemlerinin Doğruluk Oranları.....	76

## SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış bazı simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

<b>Kısaltmalar</b>	<b>Açıklama</b>
<b>Naive</b>	Yalın
<b>E-mail</b>	Elektronik Posta
<b>SPAM</b>	İstenmeyen Posta
<b>CC</b>	Carbon Copy (Bilgi Kopyası)
<b>BCC</b>	Blind Carbon Copy (Gizli Bilgi Kopyası)
<b>Fwd</b>	Forward (Yönlendirme)
<b>UCE</b>	Unsolicited Bulk E-mail (İstenmeyen Toplu E-Posta)
<b>MMF</b>	Make Money Fast (Hızlı Para Kazanma)
<b>TSK</b>	Takagi-Sugeno-Kank
<b>MX</b>	Mail İnternet Protokol Numarası
<b>DNS</b>	Domain Name Service (Alan Adı Servisi)
<b>POP</b>	Post Office Protocol (Postane Protokolü)
<b>POP3</b>	Post Office Protocol 3 (Postane Protokolü 3)
<b>TCP/IP</b>	Transmission Control Protocol/ İnternet Protocol (İletim Denetim Protokolü / İnternet Protokolü)
<b>A-priori</b>	Sınıflandırma Problemi Öncül



## 1. GİRİŞ

Tüm dünyada bilgisayar kullanımının yaygınlaşması, internetin de hızla gelişmesine sebep olmuştur. Gelişen ve büyüyen internet teknolojisinin insan hayatına getirdiği en büyük olanaklardan birisi de elektronik posta ile haberleşmedir. Bu haberleşme olanağı sayesinde uzun mesafelere iletilmek istenen bilgiler, kısa ve çok hızlı bir şekilde ulaştırılmaktadır. Kullanımının kolay olması da, zaman içerisinde internet kullanıcılarının bu iletişim şeklini tercih etmesine sebep olmuştur. Hızla büyüyen bu haberleşme şekli sadece kişiler arasında ki iletişimi değil, elektronik ticaret ile ürünlerini müşterilerine sunmak isteyen işletmeler için de yeni bir ulaşım ortamı sağlamıştır.

Bu ucuz, kolay ve hızlı haberleşme aracının yoğun bir şekilde kullanılması, ticari reklamların yaygınlaşması beraberinde bazı problemleri de getirmiştir. Şüphesiz bunlardan en önemlisi ise SPAM postalarıdır. İnternet üzerinde aynı mesajın yüksek sayıdaki kopyasının, bu tip bir mesajı alma talebinde bulunmamış kişilere, zorlayıcı nitelikte gönderilmesi SPAM (İstenmeyen Posta) olarak adlandırılır [1]. SPAM çoğunlukla ticari reklam niteliğinde olup, bu reklamlar sıklıkla güvenilmeyen ürünlerin, çabuk zengin olma kampanyalarının, yarı yasal servislerin duyurulması amacıyla yöneliktir. SPAM gönderici açısından çok küçük bir harcama ile gerçekleştirilebilir olsa da, bu büyük mali yük mesajın alıcılara ulaşmasını sağlayan taşıyıcı ya da servis sağlayıcı kurumlara çıkmaktadır. Kullanıcı tarafından bakıldığında ise elektronik posta gelen kutusunda sürekli artan bir istenmeyen posta çöplüğü durumu söz konusudur. 2010 yılında gönderilen elektronik postaların % 89.1' lik kısmının SPAM olması ise durumun ne kadar önemli bir hal aldığı sonucunu doğrulamaktadır [2].

Kullanıcılar açısından aşırı zaman kaybı, işletmeler için ise verimlilik kayıpları zaman içerisinde bu tür postaların ayrıştırılmasının önemini daha da arttırmıştır. SPAM postalar yapıları itibariyle birbirlerine benzeyen posta yapılarından oluşurlar. Bu doğrultuda gelen postaların SPAM olup olmadığının bilinmesi ve diğer postalardan ayrıştırılması mümkün olabilmektedir.

Bu ayrımı yapabilecek çeşitli yapay zekâ yaklaşımları mevcuttur. Bu yaklaşımlardan diğerlerine göre daha etkili olanları Naive Bayesian ve Bulanık Mantık yöntemleridir. Bu yöntemler ile işletmeler için yüksek maliyet ve verimlilik kayıpları önlenilmekte, kullanıcılar için ise sade ve güvenilir postalara ulaşılması sağlanabilmektedir.

Bu tez çalışması ile örnek bir veri tablosu üzerinden her iki yöntem ayrı ayrı uygulanmış ve çıkan sonuçlara göre SPAM postaların ayrıştırılmasında hangisinin daha etkili olduğu değerlendirilmiştir. Yöntemlere özgü ayrıştırma teknikleri açıklanmış, her iki yöntem için de ayrı ayrı örneklere yer verilmiştir.

## **2. ELEKTRONİK POSTA VE SPAM**

Bu bölümde elektronik postaların genel yapısı ve özelliklerine yer verilmiştir. SPAM postalar hakkında ayrıntılı bilgiler ve karakteristik özellikleri bu bölümde incelenmiştir.

### **2.1. Elektronik Posta Sistemi**

Elektronik posta (e-posta), İnternet üzerinde bilgisayarlar ve insanlar arasında bilgi alışverişini sağlayan ve en yaygın kullanılan internet uygulamasıdır. Elektronik posta ya da kısa adıyla e-mail, bilgisayar ağlarında kullanıcılarının birbirleriyle yazılı olarak haberleşmesini sağlayan bir haberleşme yöntemidir. Bilgisayar ağlarının oluşturulma nedenlerinden biri, kişilerin, bir yerden diğerine (hızlı ve güvenli bir şekilde) elektronik ortamda mektup gönderme ve haberleşme isteğidir. E-posta (electronic mail, e-posta), bu amaçla kullanılan servislere verilen genel addır.

E-posta, başlangıçta sadece düz yazı mesajlar göndermek amacıyla geliştirilmişken, 1995'li yıllardan sonra geliştirilen tekniklerle, e-posta içinde kompozit yapıların (resim, ses, video, html dökümanları, çalışabilir program vb) kullanımı mümkün hale gelmiştir. Artık günümüzde hemen hemen her türlü e-mail programları bu formatları desteklemektedir.

Elektronik posta, ya da e-mail internetin belki de en çok kullanılan özelliğidir. İnternete, ya da internete bağlı olan herhangi bir networke bağlı olan herkese mesaj gönderilebilir. Milyonlarca insan hergün e-mail alıp göndermektedir. E-mail, uzaktaki akraba, arkadaş, işletmenizin farklı bölümlerinde çalışan iş arkadaşları ve kendi meslektaşlarınızla haberleşmenin en kolay ve hızlı şeklidir.

### **2.2. E-Posta Programlarında Görülen Terimler**

Bütün bu terimler, e-postanın baş kısmı da (header) yer alır ve gönderilen e-posta ile ilgili bir takım "gönderme bilgileri"ni içerir.

### **2.2.1. Gönderen ya da kimden (from)**

İletiyi gönderenin adresi ve adı. Burada otomatik olarak kişinin adı olarak gözükmeye başlarsa da başka kişinin ismi ile de gönderim yapmak mümkündür ya da aynı hesaba bağlı başka hesaplar varsa ve tanımlıysalar buradan başka bir hesap da seçilebilir.

### **2.2.2. Alıcı ya da kime (to)**

İletiyi alanın adresi ve adı. İletinin kime yollandığını gösterir. E-posta adreslerinin arasına ", " virgül koyarak birden fazla kişiye yollamak mümkündür.

### **2.2.3. Konu (subject)**

İletiyi özetleyen başlık, genelde daha sonradan hatırlamakta ve e-postaları ayırmakta yardımcı olur. Etik olarak yazılması tavsiye edilir. zaten gönderirken kullandığınız istemci size konunun boş bırakıldığını anımsatacaktır. Mümkün olduğunca kısa, gönderilen postanın içeriğini açıklayıcı net bir tümce olmalıdır.

### **2.2.4. E-posta bedeni (body)**

Burada iletinin kendisi ve genel olarak kullanıcı imzası yer alır. Etik olarak gövdede birşeylerin yazılı olması gerekmektedir. Aksi takdirde kullanılan istemci bunu hatırlatacaktır.

İnsanlar niye boş bir mail gönderirler. Genelde ekli bir dosya gönderdiklerinde içerik kısmı boş olabilir. Fakat postanın ilerki bir tarihte tekrar okunmak istenmesi halinde içeriğin olmaması sorun yaratabilmektedir.

### **2.2.5. Karbon kopya "kk" (cc, carbon copy)**

Aslen gönderilen kişiye ek olarak iletinin başka alıcılara gönderilmesini sağlar. İleti buraya yazılan kişilere doğrudan hitap etmemektedir, ileti bu kişilere bilgilendirme amacı ile yollanmıştır.

### **2.2.6. Gizli karbon kopya "gkk" (bcc, blind carbon copy)**

Kapalı Karbon Kopya olarak da bilinir. Buraya yazılan adresler iletide gözükmez. Etik olarak birçok kişiye gönderilen gayri resmi iletiler için bu alanın kullanılması uygun düşer. Bu tür adres gizleme yığın iletiye karşı da koruma sağlar. Yani özellikle yönlendirme (forward) yaparken buna dikkat etmek gerekir. Böylece hem gönderenin adresi hem de postanın gönderildiği diğer alıcıların adresleri korumuş olur.

### **2.2.7. Cevaplama (reply)**

Belli bir adresten gelen e-postayı cevaplamak için kullanılır. Alıcı kısmına otomatik olarak e-postayı yollayanın adresine gönderilen yazı alıntı olarak alınır. En çok kullanılan fonksiyonlardan biridir. Yalnız göndermeden önce alıntıdan gereksiz yazıların silinmesi gerekebilir. Böylece alıcı neye cevap yazıldığını bilecektir.

### **2.2.8. Herkesi cevaplama (reply all)**

Cevaplanan iletiyi, gelen e-postanın To: ve CC: kısmında yazılmış olan tüm adreslere yollamak için kullanılır. En çok suistimal edilen fonksiyonlardan biridir. Genelde tüm adresler açık olarak ve gereksiz adreslere gider. Bilgisayara giren virüslerin kullandığı mekanizmalardandır.

### **2.2.9. Yönlendirme (forward, fwd)**

Gelen bir e-postanın bir başka adrese yönlendirilmesini sağlar.

### **2.2.10. Ek (attachment)**

Yazıya ek olarak yollanan dosyalardır.. Ses, resim, video, yazı başta olmak üzere birçok değişik formatta dosya eklenebilir. Virüslerin başlıca yayılma yollarından biridir. Bilinmeyen dosya eklerinin açılması önerilmemektedir.

### **2.3. SPAM Posta**

İnternet üzerinde aynı mesajın yüksek sayıdaki kopyasının, bu tip bir mesajı alma talebinde bulunmamış kişilere, zorlayıcı nitelikte gönderilmesi SPAM olarak adlandırılır [3].

SPAM çoğunlukla ticari reklam niteliğinde olup, bu reklamlar sıklıkla güvenilmeyen ürünlerin, çabuk zengin olma kampanyalarının, yarı yasal servislerin duyurulması amacıyla yöneliktir. SPAM gönderici açısından çok küçük bir harcama ile gerçekleştirilebilirken mali yük büyük ölçüde mesajın alıcıları veya taşıyıcı, servis sağlayıcı kurumlar tarafından karşılanmak zorunda kalınır.

İnternet kullanıcıları üzerindeki etkileri incelendiğinde birkaç tip SPAM vardır. Elektronik posta aracılığıyla gönderilen spam doğrudan gönderilen mesajlarla, bireysel kullanıcıları hedef alır. E-Posta SPAM listeleri genellikle Usenet gönderilerinin taranması, tartışma gruplarının üye listelerinin çalınması veya web üzerinden adres aramalarıyla oluşturulur. Email tipindeki spam gönderileri tipik olarak alan kullanıcının masraf yapmasına sebep olur. E-Posta erişimi için süreye bağlı telefon parası ödeyen her kullanıcı için bir bedel ortaya çıkması kaçınılmazdır. Bununla birlikte, istenmeyen postaların taşınmasının servis sağlayıcılar ve diğer on-line servisler üzerinde oluşturduğu mali yük de doğrudan abonelere yansımaktadır [4].

E-Posta yolu ile gönderilen SPAM türlerinden ticari içerikli olan UCE (Unsolicited Commercial e-mail- Talep Edilmemiş Ticari e-posta) adından da anlaşılacağı gibi

istenmediđi halde kullanıcıya gönderilen bir ürünü ya da hizmeti tanıtıcı elektronik posta iletileridir [5].

İçeriğinin mutlaka ticari olması gerekmeyen UBE (Unsolicited Bulk e-mail Talep Edilmemiş Kitlesel E-Posta), aynı anda yüzbinlerce E-Posta hesabına gönderilen E-Posta iletileridir. Bu iletiler ticari içerikli olabileceđi gibi politik bir görüşün propagandasını yapmak ya da bir konu hakkında kamuoyu oluşturmak amacı ile gönderilen E-Posta iletileri de olabilir. SPAM hakkında önemli bir nokta, bir iletinin SPAM olarak nitelendirmek için kullanılacak ölçütün iletinin içeriđi ile hiç alakalı olmamasıdır. Herkesin üzerinde hemfikir olduđu, önemli bir toplumsal duyarlılıđa sahip bir konu hakkında görüş bildirmek için kitlesele olarak gönderilen bir iletide aslında SPAM olarak nitelendirilebilir.

Bir diđer sık rastlanılan E-Posta SPAM tipi ise MMF (Make Money Fast – Kolay Para Kazanın) iletileri; zincir iletiler ya da piramit benzeri pazarlama yapıları ile ilgili gelen iletilerdir. Piramitin en üstündeki isme para gönderip listenin altına kendisi eklediğinde para kazanmaya başlanılacağına ilişkin iletiler bu tip SPAM iletilerine örnek olarak verilebilir.

E-Posta türündeki SPAM' in rahatsız edici bir tipi ise, iletinin tartışma listelerine gönderilmesi durumudur. Birçok tartışma listesinde, kimi işlemler sadece liste üyeleri tarafından gerçekleştirilebildiğinden, SPAM göndericileri, mümkün olduđu kadar çok listeye üye olmaya çalışarak, liste üyelerinin adreslerini temin ederler.

Diđer bir SPAM tipi ise, iptal edilebilir (cancellable) Usenet mesajları aracılıđı ile yapılan spamdir. 20 veya daha fazla haber öbeğine aynı anda gönderilen bir ileti spam kapsamında incelenir. Usenet kullanıcıları açısından bu kadar çok sayıda haber öbeğine gönderilen bir iletinin genellikle öbeklerin çođu, hatta hepsi açısından konu dışı kaldığı tesbit edilmiştir. Bu tür spam, sıklıkla haber öbeklerini okuyan ancak çok ender veya hiç gönderi yapmadıklarından email adresleri elde edilemeyen kullanıcı grubunu hedefler. Usenet spamleri haber öbeklerini reklamlar veya ilgisiz iletilerle doldurarak kullanıcı açısından faydasız ve kullanılması zor hale getirir.

### 2.3.1. SPAM postaların sınıflandırma değerleri

SPAM Postalar yapısal durumları itibariyle bazı ortak özellikleri taşırlar;

- Birden fazla alıcıya aynı içerik ile gönderilirler.
- Çoğunlukla alıcıya hiçbir şey ifade etmezler.
- Çirkin ya da yasadışı içerikle gelirler ya da onlara yönlendirirler.
- İçerikleri yalan ya da yanıltıcı olur.
- Mesajın başlık bilgileri tahrip edilmiş olur.
- Dolayısıyla geriye dönük izleme hayli zor olur.
- Alıcıların bu dağıtımdan ileti almak istemediklerini belirtebilecekleri geçerli/fonksiyonel bir adres sunmazlar.
- Elde edilmesi ve kullanılması kişilik haklarına tecavüz niteliği taşıyan içeriklere sahip olurlar ya da bu yolla toplanan bilgiyi, kitleyi kategorize etmek için kullanırlar [3].

### 2.3.2. SPAM postaların içerikleri

SPAM Postalar çok çeşitli içeriklerle kullanıcının karşısına çıkabilmektedir. SPAM hareketinin doğası itibarı ile kanun dışı ve normal koşullarda pazarlanması yasak olan ürünler ve pornografi, diğer alternatiflerinin önüne geçmektedir.

Çizelge 2.1., SPAM E-Postaların 2003 – 2004 yılları arasındaki içerik değişimini göstermektedir [3].



Çizelge 2.1. SPAM e-postaların 2003 – 2004 yılları arasındaki içerik değişimi [3]

Ürün	2003 Yılı	2004 Yılı	Değişim	Açıklama
Cinsel (Grafiksiz)	%17	%34	+%17	Cinsel içerikli sitelere linkler, cinsel yazılar
Sigorta Hizmetleri	%1	%4	+%3	Ev, otomobil, sağlık ile ilgili sigorta hizmetleri
Bitki / İlaç	%8	10	+%2	Ucuz ilaçlar, uyuşturucular
Finansal	%12	%13	+%1	Çabuk para kazanma yöntemleri
Seyahat / Kumar	%2	%3	+%1	Uçak biletleri, rezervasyonlar, İnternet kumarhane reklamları
Saadet Zincirleri	%8	%7	-%1	İnsan getirdikçe kazanacaksınız konsepti
Haberler	%9	%6	-%3	Kullanıcıya hiçbir şey ifade etmeyen haberler
Diğerleri	%13	%8	-%5	Geri kalan ve SPAM gibi görünen her şey
Cinsel (Grafikli)	%13	%7	-%5	Cinsel resimler içeren her şey
Şüpheli Ürünler	%20	%10	-%10	Kırılmış yazılımlar, düzmece diplomalar vs.

### 2.3.3. SPAM postaların zararları

SPAM E-Posta bilgisayar ağlarının yoğunluğundan kullanıcıların verimli çalışmasına kadar birçok alanda bilgisayar dünyasına ve onun bir bileşeni olan insana zarar vermektedir. İnternet kullanıcıları SPAM' den, içerdiği çirkin ve rahatsız edici içerikten, durduramadıkları bir hareket olduğundan ve istemedikleri halde sürekli aldıklarından dolayı rahatsız olmaktadır. Çizelge 2.2.'de 2003 yılında 1272 e-posta kullanıcısı üzerinde yapılan bir araştırmanın sonuçlarından son kullanıcı

tarafından duyulan rahatsızlıklar hakkında bir öngörü sahibi olmak için bilgi vermektedir:

Çizelge 2.2. 2003 yılında 1272 e-posta kullanıcısı üzerinde yapılan bir araştırmanın sonuçları [3]

<b>Rahatsızlık Nedeni</b>	<b>Kullanıcı Yüzdesi</b>
SPAM'in istenmeyen bir şey olması	% 84
Rahatsız edici çirkin içerik	% 80
Bilgisayara zarar verme potansiyeli	% 79
SPAM E-postaların yüksek boyutu	% 77
Müstehcen içerik	% 76
Gizlilik haklarına tecavüz	% 76
Durdurulamıyor oluşu	% 75
Zaman kaybettirmesi	% 69

Görüldüğü gibi kullanıcıların çok büyük bir kısmı SPAM' i istenmeyen bir durum olarak nitelendirmekte ve bu durumdan rahatsız olmaktadır [3].

#### **2.4. Türkiye' de SPAM Postalar Üzerine Yapılan Çalışmalar**

Ülkemizde istenmeyen postalar konusunda çok fazla çalışmaya rastlanamasa da bu alanda yapılan çalışmalar da mevcuttur. Çeşitli üniversitelerde yapılan SPAM Posta Filtreleme konusunun yanı sıra, **Türkiye Anti-Spam Organizasyonunu** adı altında çalışmalarını sürdüren bir anti spam grubu da ülkemizde bu konu hakkındaki çalışmalarını sürdürmektedir [4].

Bu alanda akademik olarak yapılan çalışmalar yine SPAM Postaların filtrelenmesi üzerine kurulmuştur. Bunlar:

#### **2.4.1. İstenmeyen iletilerin paralelleştirilmiş knn algoritması ile tespiti**

Elektronik posta kullanımının arttığı günümüzde, istenmeyen iletilerin sayısı da artmıştır. Çalışma, elektronik iletilerin belirlenmesi için kişisel bir filtreleme modeli geliştirilmesine yöneliktir. Bunun için geliştirilen sınıflandırıcı sayesinde, gelen iletinin istenen ya da istenmeyen ileti olup olmadığına karar verilir. Bu kararın verilme süresinin kısa olması gerektiği düşünülürse, sınıflandırıcının sonucu bildirme süresinin kısa olması gerekmektedir. Çalışmada, sürenin kısalması için, geliştirilen sınıflandırıcının paralelleştirilmesi sağlanmıştır [6].

#### **2.4.2. Türkçe morfolojik çözümleme, yapay sinir ağları ve bayesian filtreleme tabanlı uyarlamalı spam-önler filtrelemesi**

Yapay Sinir Ağları uygulamalarını temel olarak Türkçe spam mesajlarını filtreleyen bir algoritma ortaya koyulmuştur. Sonuçta, Microsoft Outlook ile bağlantılı çalışabilen bir spam-önler filtresi ortaya çıktığı için, ürün son kullanıcıya yöneliktir ve bundan dolayı kullanıcıya özeldir; program kullanıcının spam ve normal mesajlarını öğrenerek kendini her kullanıcı için uyumlu hale getirir. Algoritmanın iki temel kısmı var: Birinci kısım Türkçe kelimelerin morfolojisini incelerken ikinci kısım Morfolojik incelemeden gelen kelime köklerini kullanarak spam mesajları filtreler. Öğrenme algoritmalarının girdi vektörleri iki şekilde belirlenir: ikili model ve olasılık modeli. Yapay Sinir Ağlarının tek katmanlı ve çok katmanlı algılayıcı birimleri kullanılırken; Bayesian Filtresi de ikili modeli, olasılık modelini ve ileri olasılık modelini kullanmaktadır. Bu çalışma için yaklaşık 750 (410 spam, 340 normal) mesaj kullanılmıştır. Filtrelemede 90%' dan yüksek başarı oranı elde edilmiştir [7].

### 3. NAİVE BAYESIAN YÖNTEMİ VE SPAM POSTALARIN AYRIŞTIRILMASI

Naive Bayesian elektronik postaların SPAM postalardan ayrıştırılması konusunda yapısı itibariyle uygun özellikler taşıyan bir yapıya sahiptir. Bu nedenle elektronik postaların ayrıştırılmasında etkisi büyüktür.

#### 3.1. Naive Bayesian Yöntemi

Bayesian teoremi olasılık kuramı içinde incelenen önemli bir konudur. Bu teorem bir rastlantısal değişken için olasılık dağılımı içinde koşullu olasılıklar ile marjinal olasılıklar arasındaki ilişkiyi gösterir. Bu şekli ile Bayesian teoremi bütün istatistikçiler için kabul edilir bir ilişkiyi açıklar. Bu kavram için Bayesian kuralı veya Bayesian kanunu adları da kullanılır. Ancak bazı istatistikçiler için Bayesian teoremi özel olarak değişik bir önem de taşır. Felsefi temelde olasılık değerlerinin nesnel bir özellik değil, gözlemcinin meydana çıkardığı subjektif bir değer olarak kabul eden subjektivist olasılık düşünürlerine göre Bayesian teoremi, yeni kanıtlar ışığında olasılık değeri hakkındaki subjektif inanışların güncelleştirilip değiştirilmesini sağlayan temel bir gereçtir; yani sonsal bir yaklaşımın temelidir.

Naive Bayesian, hedef değişkenle bağımsız değişkenler arasındaki ilişkiyi analiz eden tahminci ve tanımlayıcı bir sınıflama algoritmasıdır [8].

Olasılık teorisi içinde incelenen bir olay olarak B olayına koşullu bir A olayı (yani B olayının bilindiği halde A olayı) için olasılık değeri, A olayına koşullu olarak B olayı (yani A olayı bilindiği haldeki B olayı) için olasılık değerinden farklıdır. Ancak bu iki birbirine ters koşulluluk arasında çok belirli bir ilişki vardır ve bu ilişkiye (ilk açıklayan istatistikçi İngiliz Thomas Bayes (1702–1761) adına atfen) Bayes Teoremi denilmektedir [9].

Son yıllarda, istatistiksel analizlerde kullanılan temel yaklaşımlar dışında, Bayes yaklaşımının kullanım alanları artmaktadır. 1763 yılında Thomas Bayes tarafından

ortaya konulan Bayes teoremi, bu yaklaşımın temelini oluşturmaktadır. Jeffreys, de Finette, Savage ve Lindley gibi birçok araştırmacı Bayesci analizin gelişimine katkıda bulunmuşlardır [10].

Formel bir teorem olarak Bayes teoremi, olasılık kavramını inceleyen her türlü değişik felsefi temel fikre bağlı olan her türlü istatistikçi tarafından kabul edilir. Ancak olasılığı objektif bir değer olarak gören ve relatif çokluluk olarak tayin eden çoklulukçu (en:frequentist) ekolüne bağlı olan istatistikçiler ile sübjektivist (veya Bayes tipi) ekoline bağlı olan istatistikçiler arasında bu teoremin pratikte nasıl kullanılabileceği hakkında büyük bir fikir ayrılığı bulunmaktadır. Çoklulukcu ekolüne dâhil olanlar olasılık değerlerini rastgele olaylarda meydana çıkma çokluluğuna göre veya anakütlenin altsetlerinin tam anakütleyle orantısı olarak saptanması gerektiğini kabul etmektedirler. Bunlara göre yeni kanıtlar karşısında olasılık değerinin değişme imkânı yoktur. Bu nedenle çoklulukcu ekolü için Bayes teoremi sadece koşulluluklar arasında ilişkiyi gösterir ve bunun pratikte kullanılma gücü küçüktür. Hâlbuki sübjektivist ekolüne göre olasılık gözlemcinin sübjektif belirsizlik ifadesidir. Bu nedenle olasılık değeri sübjektif olup yeni kanıtlar geldikçe değiştirilebileceğine inanmakta ve böylece Bayes teoremini istatistik bir incelemenin temel taşı saymaktadırlar [11].

Bayes karar kuralı, çeşitli kararlar arasındaki farkları, bu kararlara eşlik eden olasılıkları ve maliyetleri kullanarak ölçen istatistiksel bir yaklaşımdır [12].

Öznitelikler arasındaki bazı ilişkiler ve bağımlılıklar gösterilememiş olsa da Bayes karar verme kuralının birçok sınıflandırma probleminde oldukça başarılı sonuçlar verdiği görülmüştür [13].

Naive Bayes, modelin öğrenilmesi esnasında, her çıktının öğrenme kümesinde kaç kere meydana geldiğini hesaplar. Bulunan bu değer, öncelikli olasılık olarak adlandırılmıştır. Örneğin; bir banka kredi kartı başvurularını “iyi” ve “kötü” risk sınıflarında gruplandırmak istemektedir. İyi risk çıktısı toplam 5 vaka içinde 2 kere meydana geldiyse iyi risk için öncelikli olasılık 0,4’tür. Bu durum, “Kredi kartı için

başvuran biri hakkında hiçbir şey bilinmiyorsa, bu kişi 0,4 olasılıkla iyi risk grubundadır” olarak yorumlanır Naive Bayes aynı zamanda her bağımsız değişken/bağımlı değişken kombinasyonunun meydana gelme sıklığını bulur. Bu sıklıklar öncelikli olasılıklarla birleştirilmek suretiyle tahminde kullanılır [14].

### **3.2. Naive Bayes Yönteminin Ana Fikri**

Belirsizlik insan hayatında büyük bir alan kaplamıştır. Bu belirsizlik bilgi eksikliğine, unutulmuş geçmişe veya henüz gerçekleşmemiş durumlara bağlı olabilir. Günlük yaşantıda sürekli olarak kullanılan düşünceler, sezgiler bilimsel amaçlar için kullanılabilir. Bu tür oluşumların bilimsel olarak kullanılmasında Bayesci yaklaşım alternatif olarak ortaya çıkmaktadır [15].

Bir olasılıksal sınıflandırma olan Naive Bayes sınıflandırmanın ana fikri, bir belgenin sınıfının olasılığını tahmin etmek için verilen bir kelimenin sınıfının koşullu olasılıklarını kullanmaktır. Belge sınıflandırma gibi bazı öğrenme problemlerinde yaygın olarak kullanılan en pratik yaklaşımdır. Bu yaklaşımın “Naive” kısmı içindeki kelime bağımsızlığı varsayımından kaynaklanmaktadır. Çünkü kelime kombinasyonlarının olasılıklarını tahminci olarak kullanmaz. Bu yüzden Karar Ağacı (Decision Tree) gibi algoritmaların üstel karmaşıklığından öte verimli bir yaklaşımdır ve performansı Yapay Sinir Ağları (Neural Networks) ve Karar Ağacı ile karşılaştırılabilir [16].

Naive Bayes’de Artımlı (Incremental) olarak tabir edilen online bir öğrenme durumu vardır; her bir talim örneği artımlı olarak bir hipotezin doğru olma olasılığını artırır veya azaltır. Öncül bilgi gözlemlenen verilerle birleştirilebilir.

### 3.3. Naive Bayes Kuralının Çalışma Yapısı

Klasik yaklaşım, gözlemlenmiş veriler yardımı ile hakkında bilgi sahibi olunması istenen kitleye ait sonuçlar çıkarılmasıdır. Hipotez testleri gibi istatistiksel analizler, klasik yaklaşım kullanılarak yapılan çalışmalardır. Samuel Butler, “Hayat, yetersiz önsellerden yeterli sonuçlar çıkarma sanatıdır.” sözüyle kitleye ilişkin parametrelere ait çıkarsamalara yeni bir boyut getirmiştir. Thomas Bayes, bu mantıkla örtüşen biçimde, nedenlerden sonuçlara doğru olan mantık izleyişini sonuçlardan nedenlere doğru değiştirmiştir [15].

Naive Bayes yaklaşımı, genellikle sonrasal olasılıkları hesaplamakta kullanılan ve iki rastgele olayın koşullu ve marjinal olasılıklarını ilişkilendiren bir teoremdir. Maksimum Olabilirlik ilkesi üzerine kurulu bir teoremdir. Bu durumda Bayes Teoremi, konulan olasılıkların doğruluk oranını hesaplamak için kullanılabilir.

A ve B rastgele olaylar olsun;

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)} \quad (3.1)$$

$P(A)$  : A olayının bağımsız olasılığı

$P(B)$  : B olayının bağımsız olasılığı

$P(B | A)$  : A olayının olduğu bilindiğinde B olayının olasılığı  
(Likelihood, Şartlı Olasılık)

$P(A | B)$  : B olayının olduğu bilindiğinde A olayının olasılığı  
(Posterior, Artçıl Olasılık)

Naive Bayes kuralına dayanarak  $P(A | B)$  yi maksimum yapan durumlar hesaplanabilir.

" E " A olayının bütün durumlarının kümesi;

$$A_{MAX} = \arg \max_{A \in E} P(A | B) \quad (3.2)$$

$$= \arg \max_{A \in E} \frac{P(B | A)P(A)}{P(B)} \quad (3.3)$$

$$= \arg \max_{A \in E} P(B | A)P(A) \quad (3.4)$$

Burada  $P(B)$  sabit olarak göz ardı edilebilir. Sınıflandırıcının görevi yeni bir örnek için doğru sınıfı tahmin etmek olacaktır.

Sınıflandırma problemi öncül (a-priori) olasılıklar kullanılarak şu şekilde formüle edilebilir:

$P(v | X)$  = örnek durum için olasılık

$X = \langle a1, \dots, ak \rangle$  v sınıfının örnekleri

$P(\text{sınıf} = \text{tahmin edilen sonuç} | \text{parametre1, parametre2, parametre3, ...})$

X örneğine  $P(v | X)$  olasılığını maksimum yapan sınıf etiketi atanır.

Bayes Teoremi:

$$P(v | X) = P(X | v)P(v)/P(X) \quad (3.4)$$

$P(X)$  bütün sınıflar için sabittir

$P(v)$  = sınıfın bütün örneklere karşı görel frekansı

$P(v | X)$  maksimum =  $P(X | v)P(v)$  maksimum

$P(a1, \dots, ak | v) = P(a1 | v) \dots P(ak | v)$



Eğer  $i$ . Özellik kesikli değerde ise  $P(ai | v)$ ,  $v$  sınıfında  $i$ . Özellik olarak  $ai$  değerine sahip örneklerin göreceli frekansı olarak tahmin edilir.

Eğer  $i$ .ci özellik sürekli ise;

1. Kesikli hale getirilebilir
2. veya  $P(ai | v)$  Gauss (Normal) dağılımına sahip olduğu varsayılarak tahmin edili, sadece ortalama varyans gereklidir.

İki durumda da hesaplanabilir.

### 3.4. Bayes Teoreminin Değişik Şekilleri

Bayes teoremi çok kere daha ek kavramlar eklenerek, sanki daha da genişletilerek, ifade de edilir. Bunun için önce şu ifade kullanılır:

(3.5)

$$P(B) = P(A \cap B) + P(A^c \cap B) = P(B|A)P(A) + P(B|A^c)P(A^c)$$

Burada  $A^c$  (çok kere  $A$  olmayan olarak ifade edilen)  $A$  olayının tamamlayıcısı olur. Bu Bayes teoremi formülüne konulunca Bayes teoremi için yeni alternatif bir formül elde edilir:

$$P(A | B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)} \quad (3.6)$$

Daha genel olarak ,  $\{A_i\}$  olay uzayının bir bölüntüsünü oluşturduğu göz önüne alınınca, bu bölüntü içinde bulunan herhangi bir  $A_i$  için şu ifade elde edilir:

$$P(A_i | B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)} \quad (3.7)$$

### 3.4.1. Bahis oranı ve olabilirlik orantısı şeklinde bayes teoremi

Bayes teoremi bir olabilirlik orantısı olan  $\lambda$  ile bahis oranı olan  $O$  terimleri ile şöyle ifade edilir:

$$O(A | B) = O(A) \cdot \Lambda(A|B) \quad (3.8)$$

Burada

$$O(A | B) = \frac{P(A|B)}{P(A^c|B)} \quad (3.9)$$

B verilmiş ise A olayının bahis oranı;

$$O(A) = \frac{P(A)}{P(A^c)} \quad (3.10)$$

A kendi bahis oranı ve

$$\Lambda(A|B) = \frac{L(A|B)}{L(A^c|B)} = \frac{P(B|A)}{P(B|A^c)} \quad (3.11)$$

olabilirlik orantısı olur.

### 3.4.2. Olasılık yoğunluk fonksiyonları ile bayes teoremi

Bayes teoreminin sürekli olasılık dağılımlarına uygun olan bir şekli de vardır. Olasılık yoğunluk fonksiyonları tıpatıp olasılık olmadıkları için bu şeklin ispatı biraz daha karmaşıktır. Bu şekilde Bayes teoremi bir limit işlemin geliştirilmesi sonucu ile ortaya çıkar.

$$f(x|y) = \frac{f(x, y)}{f(y)} = \frac{f(y|x)f(x)}{f(y)} \quad (3.12)$$

Buna benzer olan bir diğer ifade de toplam olasılık yasası için şöyle ortaya çıkartılabilir:

$$f(x|y) = \frac{f(y|x)f(x)}{\int_{-\infty}^{\infty} f(y|x)f(x)dx} \quad (3.13)$$

Aynı genel aralıklı hâl gibi bu formülde bulunan parçalara da özel isimler verilmiştir:

- $f(x, y)$  X ve Y için bileşik dağılımdır;
- $f(x|y)$  Y=y verilmiş iken X in sonsal dağılımıdır;
- $f(y|x) = L(x|y)$  (x in bir fonksiyonu olarak) Y=y verilmiş ise X in olabilirlik fonksiyonudur;
- $f(x)$  X in marjinal dağılımı ve X in önsel dağılımı olur;
- $f(y)$  Y nin marjinal dağılımı olur;

Burada her bir terim için f notasyonu kullanılmıştır ancak bunların hepsi değişik birer fonksiyondur. Burada verilen hali ile fonksiyonların birbirinden değişik oldukları ancak içlerinde bulunan terimlerin farklı olmaları ile anlaşılabilir. Buradaki terimlerin farklı olmaları ile anlaşılabilir.

### 3.4.3. Soyut bayes teoremi

Olasılık uzayında verilmiş olan iki mutlak sürekli olasılık çözümleri  $P \sim Q(\Omega, F)$  ve bir sigma-cebiri  $(G, \subset F)$  olsun. Bu halde  $F$  ölçülmeli rassal değişken  $X$  için soyut Bayes teorem şöyle ifade edilir:

$$E_P(X|G) = \frac{E_Q\left[\frac{dP}{dQ} X|G\right]}{E_Q\left[\frac{dP}{dQ}|G\right]} \quad (3.14)$$

Bu formulasyon şekli Kalman filtreleme tekniğinde denklemleri bulmak için kullanılır. Bu şekil ayrıca finansman matematiği içinde numeraire değişmesi tekniklerinde uygulanır.

### 3.5. Bayes Teoreminin Kapsamının Genişletilmesi

İkiden daha fazla değişken kapsayan problemler için de Bayes teoremine benzer teoremler oluşturulabilir, Örneğin;

$$P(A|B, C) = \frac{P(A)P(B|A)P(C|A, B)}{P(B)P(C|B)} \quad (3.15)$$

Bu Bayes teoreminin ve koşullu olasılık tanımlamasının üzerine birkaç işlem yaparak genişletilmesi sağlanabilir.

$$\begin{aligned}
P(A|B,C) &= \frac{P(A,B,C)}{P(B,C)} = \frac{P(A,B,C)}{P(B)P(C|B)} = \\
&= \frac{P(C|A,B)P(A,B)}{P(B)P(C|B)} = \frac{P(A)P(B|A)P(C|A,B)}{P(B)P(C|B)}
\end{aligned} \tag{3.16}$$

Bu çalışmalar için uygulanacak genel strateji ortak olasılık için parçalama ile çalışmaya başlayıp ilgili değişkenleri entegrasyon ile ayırmaktır. Uygulanan parçalama şekline göre, bazı entegrallerin 1' e eşit olup parçalama ifadelerinden düşmesi sağlanabilir, bu şekilde hesaplamalarda önemli şekilde azalmalar görülür.

### 3.6. Naive Bayes Yöntemi İle İlgili Örnek Uygulamalar

#### Örnek 1;

Kaleye sürekli şut çeken bir futbolcunun atacağı şutun şekline göre, topun gol olma olasılığı hesaplanabilir. Bunun için atış şekillerinden faydalanılmaktadır (Topu Attığı Yön, Topa Vuruş Sertliği, Topun Yüksekliği gibi).

Çizelge 3.1.'de görüldüğü gibi şut şekillerine göre daha önce yaptığı vuruşlar ve gol durumları listelenmiştir.

Çizelge 3.1. Daha önceden yapılan şutların veri kümesi

Şut	Topu Attığı Yön	Topa Vuruş Sertliği	Topun Yüksekliği	Gol Durumu
1.Şut	Sağ	Sert	Yerden	Evet
2.Şut	Sol	Sert	Havadan	Evet
3.Şut	Sağ	Yumuşak	Havadan	Hayır
4.Şut	Orta	Sert	Yerden	Hayır
5.Şut	Sağ	Yumuşak	Havadan	Evet
6.Şut	Orta	Yumuşak	Havadan	Evet

Şut	Topu Attığı Yön	Topa Vuruş Sertliği	Topun Yüksekliği	Gol Durumu
7.Şut	Sol	Sert	Yerden	Hayır
8.Şut	Sol	Yumuşak	Havadan	?

Tabloda bulunan değerler, şut şekillerine göre gol olma frekanslarını gösterir. Bu frekanslar daha önceden verilen şut veri kümesine göre oluşturulmuştur. Buna göre futbolcu, topu sağ tarafa attığı 3 şuttan 2' sini, ortaya attığı 2 şuttan 1' ini, sola attığı 2 şuttan 1' ini, sert yaptığı 4 vuruştan 2' sini, yumuşak yaptığı 3 vuruştan 2' sini, yerden atılan 3 toptan 1' ini, havadan atılan 4 şuttan da 3' ünü gol ile sonuçlanmıştır. Ayrıca atılan 7 şuttan 4' ü gol ile sonuçlanmıştır.

Çizelge 3.2. Daha önceden yapılan şut şekillerine göre gol durum frekansları

Topu Attığı Yön	Topa Vuruş Sertliği		Topun Yüksekliği		Gol Durumu					
	EVET	HAYIR	EVET	HAYIR	EVET	HAYIR				
Sağ	2	1	Sert	2	2	Yerden	1	2	Evet	Hayır
Orta	1	1	Yumuşak	2	1	Havadan	3	1	4	3
Sol	1	1								

Çizelge 3.3.'deki tablo toplanan bu frekanslara göre oluşturulmuş gol olasılıklarını içermektedir. Bu olasılıklar, Topu Attığı Yönüne göre çıkan dağılıma göre sağ taraftan atılanlar için 2/4, ortadan atılanların için 1/4 ve sol taraftan atılanların için 1/4 şeklindedir. Topa Vuruş Sertliğine göre çıkan dağılıma bakıldığında, sert yapılan vuruşlarda 2/4, yumuşak yapılan vuruşlarda ise 2/4 şeklindedir. Topun Yüksekliğine göre bakıldığında ise bu olasılık oranları, yerden yapılan vuruşlarda 1/4, havadan yapılan vuruşlarda ise 3/4 değerlerini vermektedir.

Çizelge 3.3. Gol durum frekanslarına göre gol olasılıkları

Topu Attığı Yön	Topa Vuruş Sertliği		Topun Yüksekliği		Gol Durumu					
	EVET	HAYIR	EVET	HAYIR	EVET	HAYIR				
Sağ	2/4	1/3	Sert	2/4	2/3	Yerden	1/4	2/3	Evet	Hayır
Orta	1/4	1/3	Yumuşak	2/4	1/3	Havadan	3/4	1/3	4/7	3/7
Sol	1/4	1/3								

Bu frekans ve olasılıklar eşliğinde atılacak yeni şutun tipine göre atışın gol olma olasılığı hesaplanabilir. Futbolcu topu sağdan, yumuşak ve havadan olacak şekilde attığında gol olasılığını hesaplanır.

Topu Attığı Yön	Topa Vuruş Sertliği	Topun Yüksekliği	Gol
Sol	Yumuşak	Havadan	?

Talim kümesindeki olasılık dağılımına dayanarak (Çizelge 3.3) her bir sınıf için olasılık hesaplanır. Önce her bir özelliğin olasılığı hesaba katılır, bütün özelliklere eşdeğer önemde ele alınır ve olasılıklar çarpılır [6].

$$P(\text{EVET}) = 1/4 \cdot 2/4 \cdot 1/4 = 0,0313$$

$$P(\text{HAYIR}) = 1/3 \cdot 2/3 \cdot 1/3 = 0,0370$$

Yukarıda çıkan sonuçlar, atılacak olan şutun durumlarına göre gol olma olasılıklarının çarpılmasıyla hesaplanmıştır. Burada  $P(\text{EVET})$  oluşturulurken alınan çarpım değerleri, yapılacak olan atışın durumuna göre gol olan olasılık değerlerinden oluşurken,  $P(\text{HAYIR})$  oluşturulurken alınan çarpım değerleri ise yapılacak atışın durumuna göre gol olmayan olasılık değerlerinden meydana gelmektedir.

Bayes teoremine göre sadece her bir özelliğin çarpılmasıyla oluşan olasılık kendi başına yeterli olmamaktadır. Bu olasılıklar verilen bir sınıfın toplam olasılığı da hesaba katılarak çarpılır;

$$P(\text{EVET}) = 0.0313 \cdot 4/7 = 0,0179$$

$$P(\text{HAYIR}) = 0.0370 \cdot 3/7 = 0,0159$$

Çıkan değerler dikkate alındığında bir sonuç çıkarılacak olursa, bu olasılığı maksimum yapan sınıfı seçmek yeterli olmuştur, buna göre  $P(\text{EVET}) > P(\text{HAYIR})$  olduğu için yeni durum “EVET” olarak sınıflandırılmış, bu da olağan verilere bağlı olarak atılan bir şutun gol olacağı sonucunu vermiştir.

### Örnek 2;

Bir iş için yapılan başvurular arasından seçmeler yapılmış ve bir eğitim veri kümesi oluşmuştur. Bunun için adayların Eğitim, Yaş ve Cinsiyet bilgileri dikkate alınmıştır.

Aşağıda ki örnek veri kümesine göre adayların Eğitim, Yaş ve Cinsiyet bilgilerine göre kabul durumları verilmiştir.

Çizelge 3.4. Adayların kabul durumları veri kümesi

Aday	Eğitim	Yaş	Cinsiyet	Kabul Durumu
1. Aday	Ortaöğretim	Yaşlı	Erkek	Evet
2. Aday	İlköğretim	Genç	Erkek	Hayır
3. Aday	Yüksek Öğretim	Orta	Kadın	Hayır
4. Aday	Ortaöğretim	Orta	Erkek	Evet
5. Aday	İlköğretim	Orta	Erkek	Evet
6. Aday	Yüksek Öğretim	Yaşlı	Kadın	Evet
7. Aday	İlköğretim	Genç	Kadın	Hayır



Aday	Eğitim	Yaş	Cinsiyet	Kabul Durumu
8. Aday	Ortaöğretim	Orta	Kadın	Evet
9. Aday	Yüksek Öğretim	Orta	Kadın	?

Tabloda bulunan değerler, aday bilgilerine göre kabul edilme durum frekanslarını gösterir. Bu frekanslar daha önceden verilen aday kabul durum veri kümesine göre oluşturulmuştur. Buna göre adaylardan, eğitim durumu İlköğretim olanların 1'i, Ortaöğretim olanların 3'ü ve Yüksek Öğretim olanların 1'i kabul edilmiştir. Yaş durumlarının bakıldığında, yaşı Orta olanların 3'ü, Yaşlı olanların 2'si kabul edilirken, Genç adaylardan kabul işlemi olmamıştır. Cinsiyet durumlarına göre kabul durumlarına bakıldığında ise, Erkek adayların 3'ü kabul edilirken, Bayan adayların 2' si kabul edilmiştir. Ayrıca yapılan 8 başvurudan 5' işe kabul edilmiştir.

Çizelge 3.5. Aday bilgilerine göre kabul durum frekansları

Eğitim	Kabul Durumu		Yaş	Kabul Durumu		Cinsiyet	Kabul Durumu		Kabul Durumu	
	EVET	HAYIR		EVET	HAYIR		EVET	HAYIR	EVET	HAYIR
İlköğretim	1	2	Genç	0	2	Erkek	3	1	Evet	Hayır
Ortaöğretim	3	0	Orta	3	1	Kadın	2	2	5	3
Yüksek Öğretim	1	1	Yaşlı	2	0					

Çizelge 3.6.'daki tablo toplanan bu frekanslara göre oluşturulmuş kabul durum olasılıklarını içermektedir. Bu olasılıklar, Eğitim Durumundan çıkan dağılıma göre İlköğretim için 1/5, Ortaöğretim için 3/5 ve Yüksek Öğretim için 1/5 şeklindedir. Yaşa göre çıkan dağılıma bakıldığında, Genç adaylar için 0/5, Orta yaşta ki adaylar için 3/5 ve Yaşlı adaylar için ise 2/5 şeklindedir. Cinsiyete göre bakıldığında ise bu olasılık oranları, Erkek adaylar için 3/5, Kadın adaylar için ise 2/5 değerlerini vermektedir.

Çizelge 3.6. Aday bilgi frekanslarına göre kabul durum olasılıkları

Eğitim	EVET	HAYIR	Yaş	EVET	HAYIR	Cinsiyet	EVET	HAYIR	Kabul Durumu	
									Evete	Hayıra
İlköğretim	1/5	2/3	Genç	0/5	2/3	Erkek	3/5	1/3	5/8	3/8
Ortaöğretim	3/5	0/3	Orta	3/5	1/3	Kadın	2/5	2/3	5/8	3/8
Yüksek Öğretim	1/5	1/3	Yaşlı	2/5	0/3					

Bu frekans ve olasılıklar eşliğinde yeni adayın aday bilgilerine göre kabul edilip edilmeyeceği olasılığı hesaplanabilir. Yeni adayın eğitim durumu “Yüksek Öğretim”, yaşı “Orta” ve cinsiyeti “Kadın” şeklindedir. Bu bilgilere göre yeni adayın kabul durumu hesaplanmıştır.

Eğitim	Yaş	Cinsiyet	Kabul Durumu
Yüksek Öğretim	Orta	Kadın	?

Talim kümesindeki olasılık dağılımına dayanarak (Çizelge 3.6) her bir sınıf için olasılık hesaplanır. Önce her bir özelliğin olasılığı hesaba katılır, bütün özelliklere eşdeğer önemde ele alınır ve olasılıklar çarpılır [6].

$$P(\text{EVET}) = 1/5 \cdot 3/5 \cdot 2/5 = 0,048$$

$$P(\text{HAYIR}) = 1/3 \cdot 1/3 \cdot 2/3 = 0,074$$

Yukarıda çıkan sonuçlar, yeni adayın bilgilerine göre kabul edilme olasılıklarının çarpılmasıyla hesaplanmıştır. Burada  $P(\text{EVET})$  oluşturulurken alınan çarpım değerleri, adayın bilgilerine göre kabul edilme değerlerinden oluşurken,  $P(\text{HAYIR})$  oluşturulurken alınan çarpım değerleri ise aday bilgilerine kabul edilmeme olasılık değerlerinden meydana gelmektedir.

Bayes teoremine göre sadece her bir özelliğın çarpılmasıyla oluşan olasılık kendi başına yeterli olmamaktadır. Bu olasılıklar verilen bir sınıfın toplam olasılığı da hesaba katılarak çarpılır;

$$P(\text{EVET}) = 0,048 \cdot 5/8 = 0,03$$

$$P(\text{HAYIR}) = 0,074 \cdot 3/8 = 0,028$$

Çıkan değerler dikkate alındığında bir sonuç çıkarılacak olursa, bu olasılığı maksimum yapan sınıfı seçmek yeterli olmuştur, buna göre  $P(\text{EVET}) > P(\text{HAYIR})$  olduğu için yeni durum “EVET” olarak sınıflandırılmış, bu da olağan verilere bağlı olarak başvuru yapan adayın işe kabul edileceği sonucunu vermiştir.

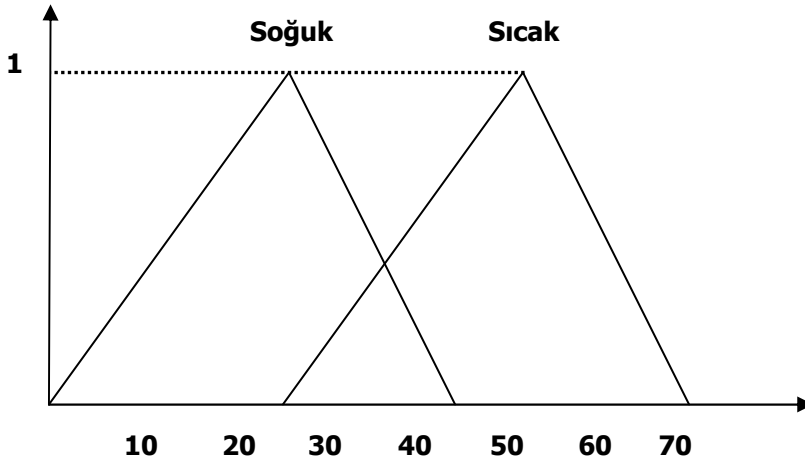
## 4. BULANIK MANTIK YÖNTEMİ VE İSTENMEYEN POSTALARIN AYRIŞTIRILMASI

Bulanık mantık elektronik postaların SPAM postalardan ayrıştırılması konusunda yapısı itibariyle uygun özellikler taşıyan bir yapıya sahiptir. Bu nedenle elektronik postaların ayrıştırılmasında bu yöntemin de kullanılması mümkündür.

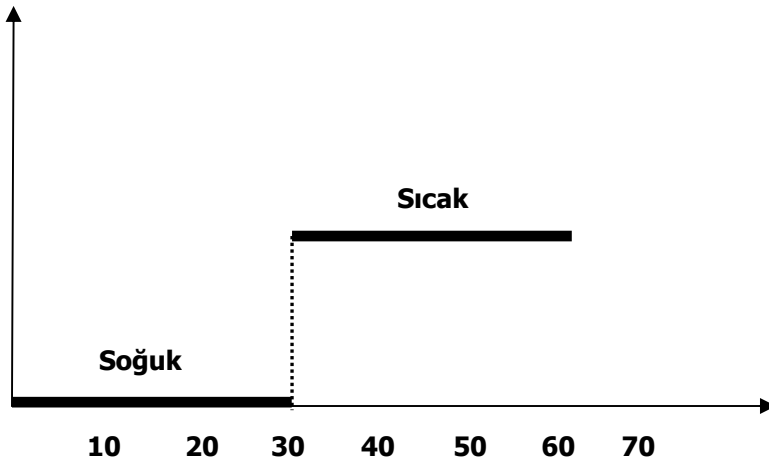
### 4.1. Bulanık Mantık ( Fuzzy Logic )

Bulanık mantık, bir şey hakkında yargı ortaya atarken, aynı anda, bu yargıyı oluştururken dayandığı matematiksel sınıflandırmaların ne kadar içinde, ne kadar dışında olduğundan bahseder. Verinin ne kadar o yargı kümesine ait, ne kadar ait olmadığı bilgisine dayanarak o veriye yeni bir tanım getirir. Örnek olarak, farz edin ki işyerinizde klimanızın otomatik devreye girme sistemi bozuldu ve bu nedenle içerideki sıcaklığı dengede tutulamıyor. Çalışmak için içerisi “çok sıcak” Sekreterinizi çağırıp, ondan sürekli olarak termometreyi gözlemlemesini ve içerideki sıcaklığın 30 C° yi geçmemesini sağlamasını isteniyor. Bu durumda muhtemelen sekreteriniz, bu sorunu klimanın kumandasıyla, klimayı açıp kapatarak çözebileceğini düşünecektir. Sekreteriniz klasik bir yaklaşım gösterir ise, termometre 30 C° ye yakın bir değeri gösterdiğinde klimayı çalıştırması gerektiğini düşünecek, içerisinin yeterince serin olduğu kanısına varınca da klimayı kapatacaktır. Eğer ki sekreteriniz akılcı bir yaklaşım gösterir ise, oda sıcaklığı için uygun sabit bir değer belirleyecektir. İşte Bulanık Mantık, Aristoteles’in “Sadece doğrular ve yanlışlar vardır” mantığına alternatif olarak kendini ifade eder. Modern Teknolojinin kullandığı kodlama biçimi olan 0,1 mantığına karşın Bulanık Mantık, 0 ile 1 arasındaki değerlerin varlığından bahseder... Klasik Mantık 30 °C yi “sıcak” kümesinin sınırı olarak kabul ediyorsa, 29,9 °C yi sıcak olarak kabul etme hakkını kaybeder... Oysaki aradaki bu küçük fiziksel fark, Klasik Mantık için hayati anlam ifade etmektedir... Çünkü bu değerın üyelik kümesi değişmiştir artık. 30 °C, 29,9 °C olmakla, “sıcak” olmayı reddetmiş ve bunun sonunda “sıcak” olma kümesinden dışlanmıştır... Bu sosyal manada oldukça klasik bir tutum olsa gerek... Hangi birimiz

fiziksel dünyada 30 °C yi sıcak kabul ederken, 29,9 °C nin sıcak olmadığını iddia ederiz ki... Oysaki Bulanık Mantık bu tür keskin sınırları kaldırarak, 29,9 °C ‘yi “hemen hemen” tamamen (1’e yakın bir değerle) sıcak olarak kabul eder. Fiziksel dünya ile bilimsel dünyanın paralel bakış açısıyla çalışan bir sistemi ortaya atar Bulanık Mantık... Klasik mantık için “soğuk” ya da “sıcak” olma vardır... Oysaki Bulanık Mantık “soğuk-sıcak” gibi kavramların yanında, “az soğuk”, “çok sıcak”, biraz sıcak” gibi söylemleri de kabullenir ve bunları matematiksel olarak tanımlamaya çalışır. Böylece fiziksel hayat birdenbire matematiksel bir söyleme dönüşür ya da matematik hayatın ta kendisi olur [17]. Diğer bir ifadeyle, kesin ilişkiler ile bulanık ilişkiler arasındaki fark, kesin ilişkilerde sınırların kesin bir şekilde belirlenmiş olması, bulanık ilişkilerde ise sınırların kesin olmayan sınırlarla belirlenmiş olmasıdır [18].



Şekil 4.1. Bulanık mantık modeli



Şekil 4.2. Klasik mantık modeli

Bulanık Mantığın tarihi çok eskilere dayanmaktadır. Aristoteles'in "Var ya da Yok" yasalarına karşın Heraclitus, bir şeyin hem doğru hem yanlış olabileceği fikrini ortaya sürmüştür. Plato ise bu durumu ileriye götürerek "doğru" ve "yanlış" olmanın dışında, doğru ve yanlışın iç içe olduğu üçüncü bir durumdan bahseder. Ancak ilk kez Lukasiewicz 1900'lerin başında "olası" kavramını ortaya atmıştır. Bu kavram Bulanık Mantığın temelini oluşturur. Lukasiewicz, Doğru ile Yanlış arasında sonsuz farklı değer olduğundan bahsetmiş ve ancak bu mantık uygulamalarda çok başarı elde edememiştir. Nihayet 1965 yılında Lotfi A. Zadeh, bu değerleri [0.0, 1.0] aralığında ki sayılarla ifade ettiği teorisinin adına "Bulanık Mantık (Fuzzy Logic)" ismini vermiştir [17].

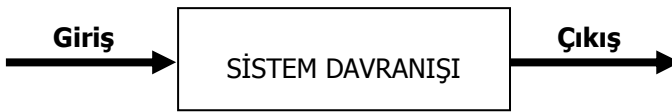
Bulanık Mantık çözümlene yaparken, bilimin kendine has, insanların öznel tecrübelerini küçümseyen tarzından farklı olarak, gözlemlere yer verir. Odanın sıcak oluşu öznel bir gözlemdir. Bu konuda ortaya bir yargı atabilmek için bu gözleme tabi ki ihtiyaç vardır. Ancak sınırları da aynı öznel tecrübe belirlemelidir. Odanın sıcaklığının ne olması gerektiğine sekreteriniz karar verir. Bu durumda Klasik Mantık işlemdeki "sıcak" oluş, öznel bir gözlem olsa da sınırlar ve ara geçişleri matematik tarafından (Aristoteles tarafından...) belirlenir. Bulanık Mantık işlemlerinde ise yargılar öznel (bilirkişi) gözlemlere dayanır ve bu yargılar matematiksel olarak analize hazır bir şekilde modellenir. Matematik kural olmaktan çıkar, hayat için birkaç rakam oluverir.

Bulanık Mantığın tıpkı matematik gibi uygulamasının olmadığı bir alandan bahsetmek çok zor... Endüstriyel Sistem modellemelerinden, yazılım geliştirmeye; otomatik kontrol sistemlerinden, veri analizine; yöneylem araştırma tekniklerinden, sosyolojik değişim kurallarını izleme gibi birçok alanda Bulanık Mantık uygulamalarını başarılı bir şekilde görmek mümkün. Özellikle Modern Kontrol Sistemleri, Bulanık Mantık bilimini üstlenmiş durumda... Bu beraberlikten en çok yarar gören başarılı uygulamalarıyla Otomatik Kontrol Sistemleri bilimi gibi görünüyor. Bunun yanında Bulanık mantık önüne çıkan daha karmaşık problemlerle kendini ispatlama fırsatını yakalıyor... Örnek olarak, günümüzde Robotik Hareket

Sistemlerinin karmaşık kontrol problemleriyle çoğunlukla Bulanık Mantık ilgileniyor.

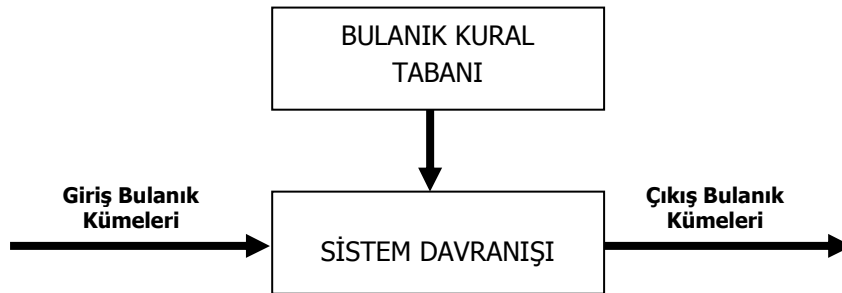
## 4.2. Bulanık Sistem

Klasik sistemlerin hemen hepsi Şekil 4.3 'de verilen üç ayrı birimden ibarettir.



Şekil 4.3. Klasik sistem

Bu kısımlar giriş, sistem davranışı ve çıkış kısımlarıdır. Bulanık sistemlerin bu klasik tasarımdan farkı, Şekil 4.4'te gösterildiği gibi sistem davranışı kısmının ikiye ayrılarak aralarında bağlantılı dört birimin olmasıdır.



Şekil 4.4. Genel bulanık sistem

Burada bulunan birimlerin her birinin farklı, fakat birbiri ile ilişkili olabilen aşağıdaki görevleri vardır.

- Genel Bilgi Tabanı Birimi: incelenecek olayın maruz kaldığı girdi değişkenlerini ve bunlar hakkındaki tüm bilgileri içerir. Buna veri tabanı veya

kısaca giriş adı da verilir. Genel veri tabanı denmesinin sebebi buradaki bilgilerin sayısal ve/veya sözel olabilmesidir.

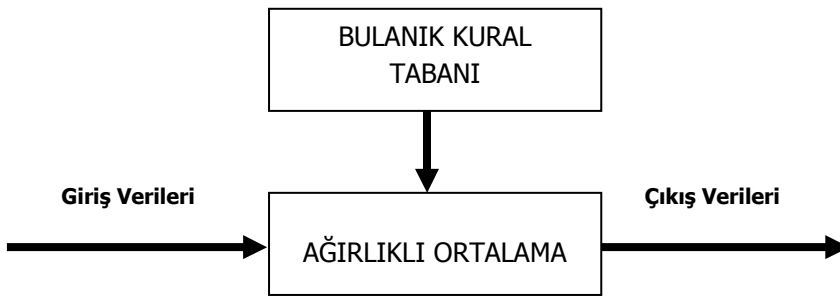
- **Bulanık Kural Tabanı Birimi:** Veri tabanındaki girişleri çıkış değişkenlerine bağlayan mantıksal EGER-İSE türünde yazılabilen bütün kuralların tümünü içerir. Bu kuralların yazılmasında sadece girdi verileri ile çıktılar arasında olabilecek tüm aralık (bulanık küme) bağlantıları düşünülür. Böylece, her bir kural girdi uzayının bir parçasını Çıktı uzayına mantıksal olarak bağlar. İşte bu bağlamların tümü kural tabanını oluşturur.
- **Bulanık Çıkarım Motoru Birimi:** Bulanık kural tabanında giriş ve çıkış bulanık kümeleri arasında kurulmuş olan ilişkilerin hepsini bir araya toplayarak sistemin bir çıkışlı davranmasını temin eden işlemler topluluğunu içeren bir mekanizmadır. Bu motor, her bir kuralın çıkarımlarını bir araya toplayarak tüm sistemin girdiler altında nasıl bir çıktı vereceğinin belirlenmesine yarar .
- **Çıktı Birimi:** Bilgi ve bulanık kural tabanlarının, bulanık çıkarım motoru vasıtası ile etkileşimi sonunda elde edilen çıktı değerlerinin topluluğunu belirtir [19].

Şekil 4.4 genel bir bulanık sistemi temsil eder. Burada dikkat edilmesi gereken bir nokta genel olarak girdi yani veri tabanındaki bilgilerin ve çıktıların bulanık değerler olmasıdır. Yani Şekil 4.4'teki sistemde, her birim tamamen bulanık kümelerden oluşmaktadır. Temel bulanık sistemin en önemli mahzuru, sayısal olan veri tabanının böyle bir genel bulanık sisteme girememesi ve çıktıların sayısal olmaması, dolayısı ile mühendislik tasarımlarında doğrudan kullanılamamasıdır.

Genel bulanık sistemin mahzurlarını bir dereceye kadar ortadan kaldırmak için *Takagi ve Sugello(1985)* ve *Sugello ve Kallk (1988)* tarafından teklif edilen ve Takagi-Sugeno-Kank (TSK) bulanık sistemi denilen sistem kullanılır. Burada veri tabanındaki girdiler birer sayı, bulanık kural ve çıkarım motorunun çalışması sonunda elde edilen çıktılar ise girdilerin bir fonksiyonu şeklindedir. Yani kural



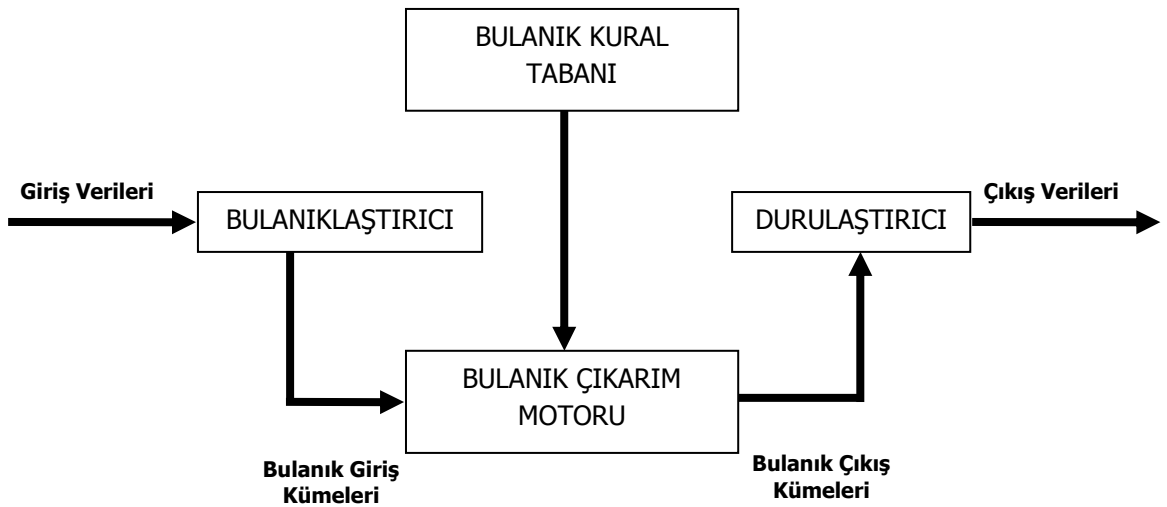
tabanındaki öncül kısımların değişkenleri olduğu gibi, İSE kelimesinden sonraki kural soncul kısmına bu değişkenlerin birer doğrusal fonksiyonu olarak yansıtıldığı düşünülmüştür. Mesela, 3 tane öncül değişkenin ( $X_1$ ,  $X_2$  ve  $X_3$ ) bulunması halinde soncul değişken olan  $y$  genel olarak bulanık sistemin kurallarından birinde EĞER  $X_1$  çok ve  $X_2$  yüksek ve  $X_3$  dar İSE  $y = 3X_1 + X_2 + 2X_3$  şeklinde ifade edilebilir. Bütün kuralların soncul kısımları sanki çok terimli bir doğrusal denklemden ibarettir. Böyle bir yapıya sahip olan bulanık sistemde sonuçlar bulanık küme şeklinde olmadıklarından Şekil 3.3'teki Bulanık Çıkarım Motoru birimi yerine, her bir kuralın öncül kısmından hesaplanan üyelik dereceleri ağırlık olmak üzere ağırlıklı çıkarım hesaplaması birimi gelir (Şekil 4.5.)



Şekil 4.5. TSK bulanık sistem

Aslında böyle bir bulanık sistemde çıktı uzayı girdilerin fonksiyonu olarak, her bir alt uzayda geçerli bir kural olmak üzere temsil edilmiştir. TSK yaklaşımı ile çıktı yüzeyinin doğrusal olmaması halinde bile, bu yüzeyin alt uzaylar üzerinde girdi değişkenleri cinsinden düzlem parçaları şeklinde modellendiği anlaşılır. Ancak, TSK bulanık sisteminin mahzurları arasında İSE kısmından sonra matematik bir ilişki bulunduğundan, kuralların soncul kısımlarının insan tarafından verilecek sözel bilgileri modelleyememesi ve giriş-çıkış değişkenleri arasında yazılması mümkün olan tüm kuralların soncul kısımlarının bulanık olmaması dolayısı ile yazılamamasıdır. İşte bu mahzurları ortadan kaldırabilmek için Şekil 4.5'te verilen ve girdi ve çıktı birimlerinde sırası ile bulanıklaştırma ve durulaştırma işlemleri yapıldığından bu birimlerin de kutu şeklinde gösterildiği bir bulanık sistem karşımıza çıkar.

Burada genel bir bulanık sistemdeki bulanık kural tabanı ve çıkarım motoru aynen kalmaktadır. Girişlerin sayısal olanları duruma tabi tutularak bulanıklaştırılmasına yarayan bulanıklaştırıcı birim ile yine bulanık olan çıktıların sayısallaştırılmasına yarayan durulaştırıcı birim ilave edilmiştir. Bulanıklaştırma ve durulaştırma sırası ile giriş sayılarını bulanıklaştırması ve bulanık sayıların sayısallaştırılması anlamına gelir. Bu sisteme, bulanık sözel bilgiler ile bulanıklaştırılmış sayısal bilgiler bir arada toplanarak, sanki Şekil 4.3'te gösterilen genel bulanık sistemin girdisine indirgenmiş bir durum ortaya çıkar. Bulanık sistem çıkışlarının mühendislik tasarımlarında kullanılması amacı ile sayısallaştırılması için durulaştırma birimi ilave edilmiştir. Bu bulanıklaştırıcı durulaştırıcı sistem, genel ve TSK bulanık sistemlerinde bulunan tüm mahzurları ortadan kaldırır [19].



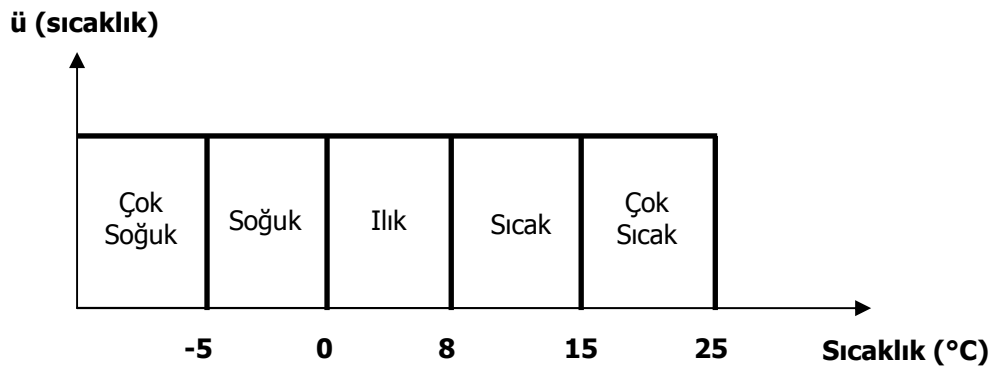
Şekil 4.6. Bulanıklaştırma-durulaştırma birimli bulanık sistem

Bulanık sistemlerin başlıca özellikleri arasında en önemli, konu olarak, çoklu girdileri, kural tabanı ve çıkarım motoru ile işleyerek tek çıktı haline dönüştürmesi gelir. Bazı özel durumlarda, çıktılar birden fazla olabilir. Ancak, hemen her mühendislik çalışmasında en az bir tane çıktı bulunur. Bulanık sistem doğrusal olmayan bir şekilde girdileri oluşturan değişkenleri, Çıktı değişkenine dönüştürerek, sistemin davranışını tespit eder. Böylece bilgi tabanının doğrusal olmayan dönüşümlere maruz bırakılması ile istenen sonuçlara ulaşmak için incelenen sistemin

kontrol altına alınması mümkün olmaktadır. Bulanık sistemler sayesinde mühendislikte görüntü işleme, zaman serileri esaslı tahmin yapmak, kontrol sorunlarını çözmek ve haberleşme yani iletişim konularında uygulamalar yapmak mümkün olmaktadır. Bunun dışında bulanık sistemler mühendislik, tıp, sosyoloji, psikoloji, işletme, uzman sistemler, yapay zekâ, sinyal işlenmesi, ulaştırma, kavşak sinyalizasyon onu gibi birçok alanda rahatlıkla kullanılabilir [19].

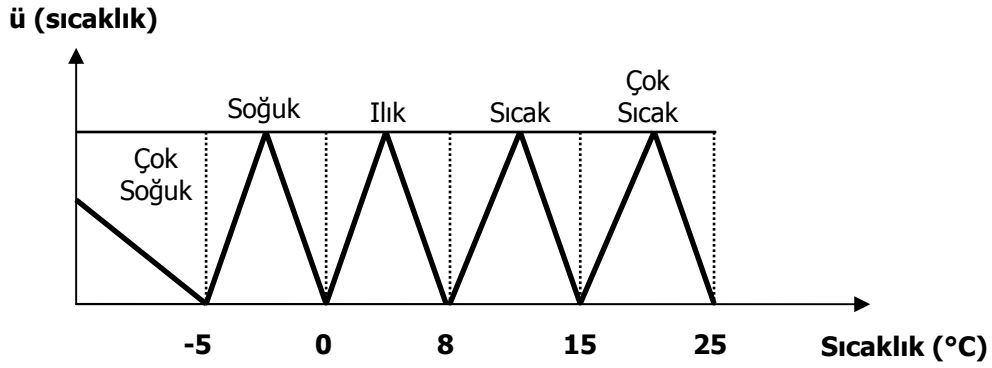
### 4.3. Üyelik Fonksiyonları

Araştırılmak istenilen ya da üzerinde çalışılan verilerin temsil ettiği sayısal aralık, konu hakkında bilgi sahibi olan uzman kişiler tarafından belirlenebilir. Örneğin, Ankara'da sıcaklık derecesinin değişim aralığının  $-5\text{ }^{\circ}\text{C}$ 'den  $+35\text{ }^{\circ}\text{C}$  olduğu bilinsin. İşte bu aralık sıcaklık kümesinin Ankara için öğelerinin bulunabileceği aralığı belirtir. Böylece tüm sıcaklık uzayı belirlenmiştir. Ancak, günlük konuşmalarda bu sıcaklık uzayının da bir takım alt aralıklardan oluştuğu düşünülür. Mesela, 'çok soğuk', 'soğuk', 'ılık', 'sıcak', 'aşırı sıcak' gibi... Burada, önce her bir alt terimin aralığının ne olduğuna karar veriniz gibi bir emirle karşılaşırsa, belki mühendis olanlar bu alt kümelerin her birinin üst üste örtüşmeyen, ancak birbirinin sınırında devamlarıymış gibi olduklarını söyleyebilir. Mesela çok soğukun  $-5\text{ }^{\circ}\text{C}$  ile  $0\text{ }^{\circ}\text{C}$ , soğukun  $0\text{ }^{\circ}\text{C}$  ile  $+8\text{ }^{\circ}\text{C}$ , ılığın  $+8\text{ }^{\circ}\text{C}$  ile  $+15\text{ }^{\circ}\text{C}$ , sıcaklığın  $+15\text{ }^{\circ}\text{C}$  ile  $+25\text{ }^{\circ}\text{C}$ , çok sıcaklığın ise  $+25\text{ }^{\circ}\text{C}$  den başladığı söylenebilir. Burada dikkat edilirse aralık tahminlerinde bulunulmuş ve her bir alt aralıktan biri bitince diğeri başlamıştır. (Şekil 4.7)



Şekil 4.7. Bitişik dikdörtgen gösterimi

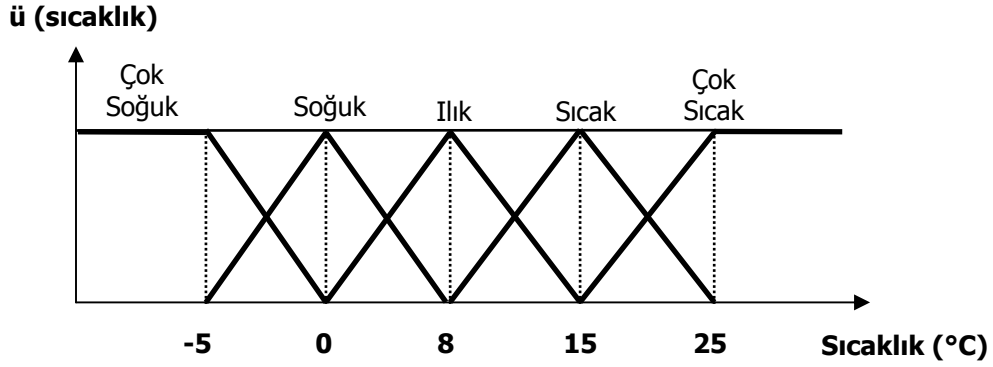
Bu aralıkların sınırlarında yine Aristo mantığına göre katı kararlar alınmalıdır. Örneğin, 7,9 °C' nin soğuk, 8.1 °C' nin ise ılık olduğuna karar verilir. Bu şekilde gösterim bakımından önemli bir nokta, her alt aralığa düşen sıcaklık değerinin üyelik derecesinin, sadece o aralıkta 1' e, diğer aralıklarda ise 0' a eşit olduğudur. Bu nedenle, her sıcaklık alt kelimesinin üyelik fonksiyonu yüksekliği 1' e eşit olan bir dikdörtgen şeklindedir.



Şekil 4.8. Bitişik üçgen gösterimi

Şekil 4.8'de yukarıdaki tartışmanın bir doğal sonucu olarak en basit üçgen üyelik fonksiyonları bitişik olarak alınmıştır. Bu üçgenlerin de sıcaklık alt kümelerini tam yansıtmadığı açıktır. Çünkü burada da sınırlardaki sıcaklık değerlerinin üyelik dereceleri sıfır olarak düşünülmüştür. Ayrıca, bu sınır değerleri, ne alttaki ne de üstteki sıcaklık alt kümelerine dâhildir. Böylece, sınır değerler için tam anlamı ile bir belirsizlik vardır. Diğer taraftan, bu şekildeki alt aralıklar halen Aristo mantığına göre işlem görür. Çünkü bir alt aralığa düşen sıcaklık değeri, sadece o alt aralığa aittir. Fakat Şekil 4.7'den farklı olarak üyelik derecesi 1'e eşit değildir. Biraz daha makul düşünen birisi, bu aralıkların arasındaki geçiş kısımlarının böyle birbirinin devamı olmayacağını ve bir örtüşmenin söz konusu olabileceğini söylerse, daha mantıklı, günlük hayatta geçerli ve uzlaştırıcı çözümlere gitmiş olur. Çünkü herkesin ılık sınırlarının +5 ile +15 °C'de sıfır üyelik derecelerine sahip olacağını kabul etmesini savunmak mümkün değildir. Halbuki, günlük hayatta sınıra yakın olan değerlerin hangi aralığa düşeceği oldukça müphem ve şüpheli, yani bulanıktır.

Böylece, sıcaklık alt aralıklarının birbiri ile örtüşmeli geçişlere sahip olmasının gerekliliği ile sonuçta Şekil 4.9'da verilen üyelik fonksiyonları karşımıza çıkar.

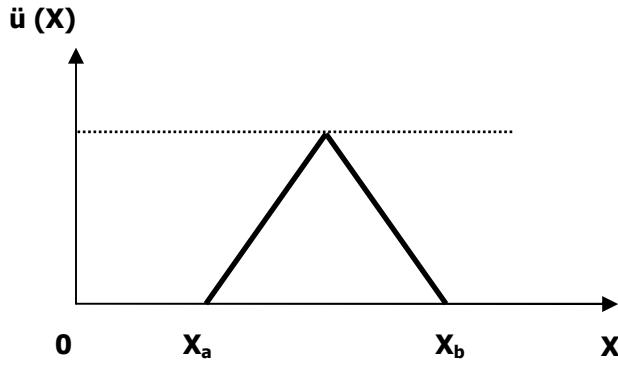


Şekil 4.9. Örtüşmeli üçgen gösterimi

Yukarıda söylenenlerden sonra ilk ve son alt aralıktaki sıcaklık durumlarının 'çok çok soğuğa' veya 'çok çok sığağa' doğru giderken başka alt aralıklar olmadığından, üyelik derecelerinin 1'e eşit kalmasının makul olacağı anlaşılır. Bunun doğal bir sonucu olarak da, ilk ve son üyelik fonksiyonlarının üçgen değil de yamuk şeklinde olacağı sonucuna varılır. Böylece, her alt aralığa girişimli olarak bir üyelik fonksiyonu şekli tayin edilmiştir.

Diğer taraftan, sorun her alt aralığa, örneğin 'ılık' aralığına düşen sıcaklık derecelerinin hepsinin aynı önemde olup olmayacağıdır. Tabii olarak, ılık aralığının alt ve üst uçlarına yaklaştıkça onun komşusu olan altta soğuk, üstte ise sıcak alt kümelerine doğru geçişler beklendiği için, o geçiş bölgelerine rastlayan kısımların tam anlamı ile ılık vasfına sahip olacağı söylenemez. Böylece, her bir alt aralığa düşen sıcaklık derecelerinin, o alt aralığın uçlarına yakın kısımlarında önemlerini ortaya kıyasla göreceli olarak kaybedeceği sonucuna, buradan da eğer bir alt aralıktaki önem derecesi diye bir değer düşünülecek olursa bunun en büyük değerlerinin o alt aralığın ortalarında, en düşük değerlerinin ise uçlarda olacağını söyleyebiliriz. Bu düşünceler bizi Şekil 4.10'da gösterilen bir geometrik gösterime sürükler. Genel olarak, her alt aralığın ayrıık üyelik fonksiyonu bu şekilde gösterildiği gibi olur. Bu

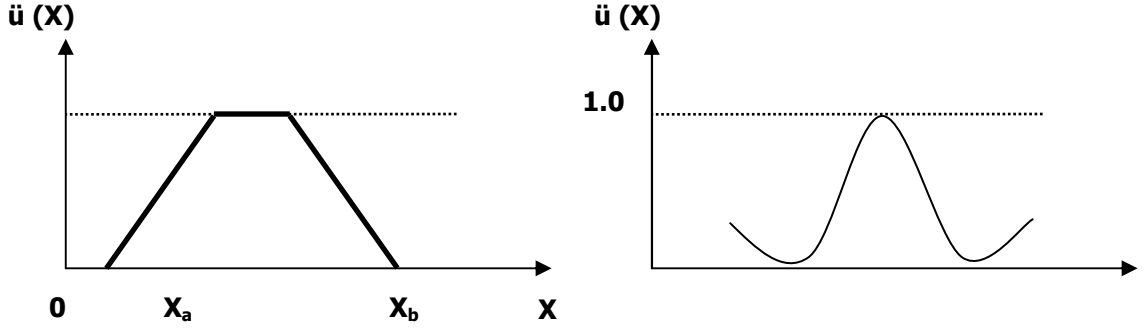
fonksiyonların simetrik olması gerekmez. Böylece  $X_a$  ve  $X_b$  gibi alt ve üst sınırlara sahip  $X$  değişkeninin bu aralıktaki her değerine ayrı bir üyelik derecesi,  $\bar{u}(x)$ , tayin edilmiş olur. Bu aralıkta  $k$  i tüm  $X$  değerleri, o  $X$  değişkeninin bir alt kümesini teşkil eder [19].



Şekil 4.10. Örtüşmeli üçgen gösterimi

Genel olarak, küme üyelerinin değerleri ile değişiklik gösteren böyle bir eğriye üyelik fonksiyonu (önem eğrisi) adı verilir. Bunun en önemli özellikleri, alt küme sınırlarındaki değerlerinin orta ögelerinkine göre daha düşük olmasıdır. Ancak klasik kümelere bir benzerlik teşkil etmesi açısından en büyük önem derecesine sahip olan ortaya yakın ögelere 1 değeri verilirse, diğerlerinin 0 ile 1 arasında ondalıklı ve sürekli değiştiği sonucuna varılır. İşte bu şekilde, 0 ile 1 arasındaki değişimin, her bir öge için değerine, üyelik derecesi, bunun bir alt küme içindeki değişimine ise, üyelik fonksiyonu adı verilir. Böylece, üyelik fonksiyonu şemsiyesi altında toplanan ögeler önem derecelerine göre birer üyelik derecesine sahiptir [19].

Bu kısma son vermeden, matematik kurallarına uygun olarak düzgün şekilli üyelik fonksiyonlarının Şekil 4.11'de gösterilen üçgenden başka, yamuk veya çan eğrisi şeklinde olacağı aklımıza gelebilir (Şekil 4.11). Pratik uygulamalarda bunlardan en fazla üçgen, ondan sonra da yamuk olanı kullanılır [19].



Şekil 4.11. Yamuk ve çan eğrisi üyelik fonksiyonları

#### 4.3.1. Üyelik fonksiyonunun kısımları

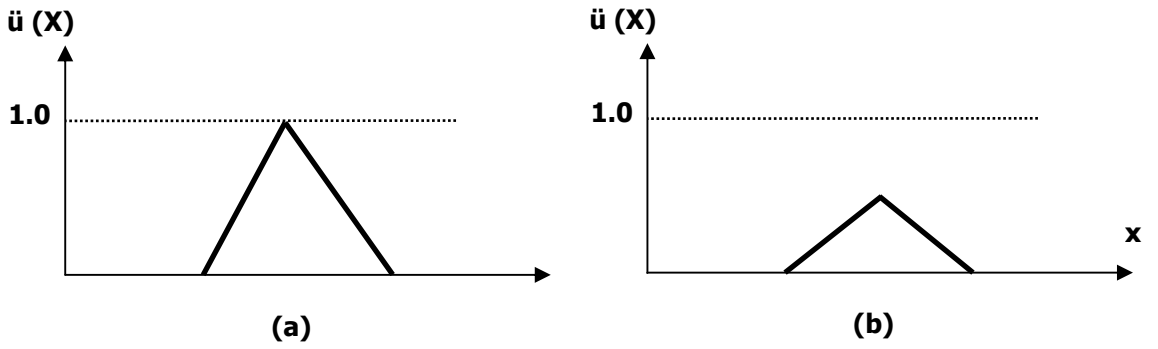
Yukarıda yapılan açıklamalardan, genel olarak bir üyelik fonksiyonunda bulunması gereken kısımlar hakkında fikirler oluşmuştur. Bunların, daha bilimsel terminoloji olarak açıklaması aşağıda yapılacaktır. En genel hali ile, yamuk şeklindeki bir üyelik fonksiyonu, Şekil 4.7'da gösterildiği gibi, değişik kısımlara ayrılabilir [19].

Bu durumda,  $i$  üyelik dereceli öğelerin tam anlamı ile, hiçbir şüpheye düşmeksizin, sadece o alt kümeyle ait olduğu sonucuna varılır. Böyle üyelik derecesine sahip olan öğeler alt kümenin orta kısmında toplanmıştır. İşte üyelik dereceleri 1'e eşit olan öğelerin toplandığı alt küme kısmına, o alt kümenin özü (core) denir. Burada  $\mu(x) = 1$ 'dir. Üçgen şeklindeki üyelik fonksiyonunda bir tane öğenin üyelik derecesi 1'e eşit olduğundan, üçgen üyelik fonksiyonlarının özü bir nokta olarak karşımıza çıkar [19].

Bunun aksine bir alt kümenin tüm öğelerini içeren aralığa o alt kümenin dayanağı (*support*) adı verilir. Dayanakta bulunan her öğenin az veya çok değerinde (0 ile 1 arasında) üyelik dereceleri vardır. Bunun matematik gösterilişi  $\mu(x) > 0$  şeklindedir. Aslında bu öğeler topluluğu önceki kısımda belirtilen aralığa karşı gelir. Üyelik dereceleri 1'e veya 0'a eşit olmayan öğelerin oluşturduğu kısımlara üyelik fonksiyonunun sınırları (*boundary*) veya geçiş bölgeleri denir. Bunun matematik tanımı  $0 < \mu(x) < 1$  şeklindedir. Bunlar alt kümenin kısmi öğeleridir. Aslında bir alt kümeyle bulanıklık özelliğinin takılması bu geçiş yerlerinin bulunması sonucundadır.

Genel olarak, tüm üyelik fonksiyonlarında biri sağda diğeri de solda olmak üzere iki tane geçiş bölgesi vardır. Şekil 4.11' de en sol ve en sağdaki bulanık kümelerde birer tane geçiş bölgesi vardır [19].

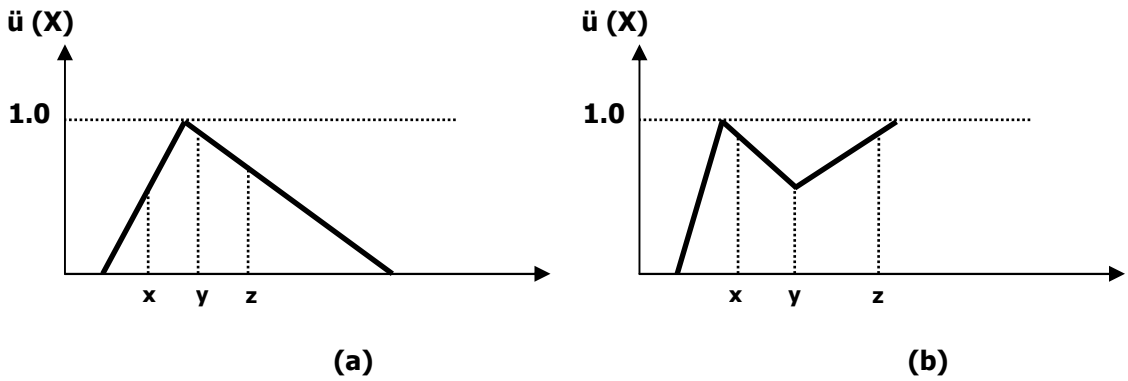
Yukarıda şekil olarak açıklanan bu üç özelliğe ilave olarak üyelik fonksiyonunun sahip olması gerekli olan iki tane daha özellik bulunmaktadır. Bunlardan birincisi, bulanık kümenin normal olduğunu tespit etmemize yarayan bir kavramdır. Buna göre normal bulanık kümede, en azından bir tane üyelik derecesi 1'e eşit olan öğe bulunmalıdır. Şekil 4.12 normal ve normal olmayan bulanık kümeleri göstermektedir.



Şekil 4.12. Bulanık kümeler, (a) normal, (b) normal olmayan

İkinci özellik ise bulanık kümenin dış bükey (konveks) olmasıdır. Dış bükey olan bulanık kümelerde üyelik fonksiyonu kümenin dayanağı üzerinde, ya sürekli artar veya sürekli azalır veya üçgen üyelik fonksiyonunda olduğu gibi önce sürekli olarak üyelik derecesi bir öğede 1'e eşit oluncaya kadar artar ondan sonraki dayanağa düşen öğeler için sürekli azalır. Bunun aksi durumlar da söz konusudur. Ancak, onlar bulanık kümelere üyelik fonksiyonu olamaz. Şekil 4.13'de dış bükey olan ve olmayan bulanık alt kümelere bazı misaller gösterilmektedir.



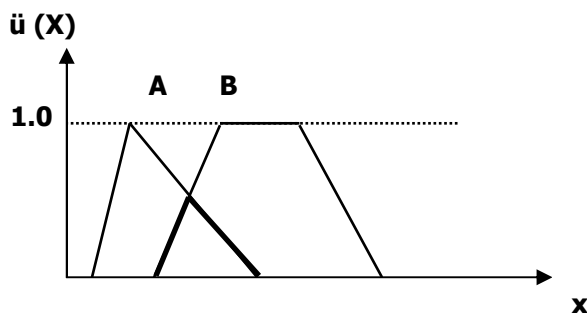


Şekil 4.13. Bulanık kümeler, (a) dış bükey, (b) dış bükey olmayan

Dış bükeyliğin matematik olarak tanımlanmasında, aynı bulanık alt kümeye düşen  $x$ ,  $y$  ve  $z$  gibi üç tane öge düşünülürse ve bunlar arasında değerce büyüklük olarak  $x < y < z$  gibi bir sıra bulunuyor ise, bunlardan ortadakinin üyelik fonksiyonu önceki ve sonrakine göre,

$$\ddot{u}(y) \geq EK [\ddot{u}(x), \ddot{u}(z)] \quad (4.1)$$

şeklindedir ve bu bağıntı daima geçerli olmalıdır. Burada EK en küçükleme (minimizasyonu) işlemi demektir. Yani  $y$ 'nin üyelik derecesi,  $x$  ve  $z$ 'nin üyelik derecelerinin en küçüğünden daha büyüktür. İşte bu durumda o kümeye dış bükey bulanık küme adı verilir [19].



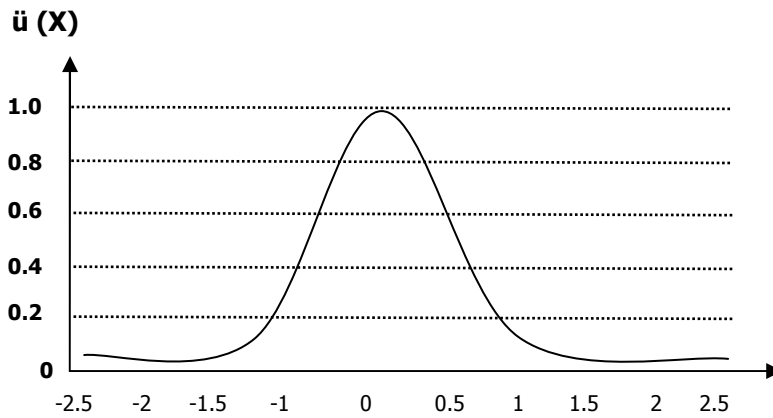
Şekil 4.14. Dış bükey olmayan kümelerin kesişimi

Bulanık kümelerin üyelik fonksiyonlarında üyelik derecesinin 0,5'e eşit olması durumundaki noktaya geçiş noktası (*cross-over*) adı verilir. Yani geçiş noktasında,  $\mu(x) = 0,5$ 'dir. Ayrıca, bulanık kümenin yüksekliği denilen bir büyüklük, üyelik derecesinin en büyük olduğu öğelere karşı gelir. Yukarıda söylenenlerden sonra, normal bulanık kümelerde yüksekliğin 1'e eşit olması gerekliliği anlaşılır. Diğer bir ifade ile yüksekliği 1'e eşit olmayan bulanık kümeler, normal olmadıklarından herhangi bir bulanık küme, mantık ve sistem çalışmasında kullanılamaz. Normal olmayan bulanık kümeleri normal hale dönüştürmek için, kümenin her bir üyelik derecesinin, en büyük üyelik derecesine bölünmesi gereklidir. Böylece normal olmayan bulanık kümelerin dış bükey olmaları şartı ile nasıl normal bulanık kümeler haline dönüştürüleceği anlaşılmış olur.

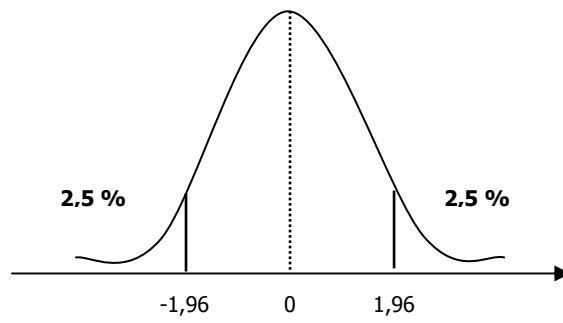
Temel bulanık kümeler normal ve dış bükey olmasına karşılık, küme işleminin yapılması sonucunda elde edilen kümeler, bulanık normal küme çıkamayabilir. Örneğin, iki normal ve dışbükey bulanık alt kümenin birleşimi normal ve dış bükey olmayan bulanık küme çıkabilir (Şekil 4.14).

Üyelik fonksiyonları simetrik olabilir de olmayabilir de. Genel olarak, bir boyutlu uzayda tanımlanan bulanık kümelerin iki veya daha fazla boyutta, az da olsa tanımlanması mümkündür. Bir boyutlu uzayda çizgi şeklinde olan üyelik fonksiyonları iki boyut bu uzayda yüzey şeklinde görülür.

İstatistikteki dağılım fonksiyonları hakkında bilgisi olanlar, üyelik fonksiyonunun sanki dağılım fonksiyonlarına benzediği sonucunu çıkarabilir. Dağılım fonksiyonlarında tepe noktasının 1'e eşit olması söz konusu değildir. Ancak histogram olarak dağılım fonksiyonunun altındaki alanın 1'e eşit olması gereklidir. Şekil 4.15'te Gauss eğrisi şeklinde dağılım ve üyelik fonksiyonları ayrı ayrı gösterilmiştir. Bunlardan bulanık kümeyi temsil eden Gauss eğrisinde üyelik fonksiyonunun tepe noktasının 1'e eşit olduğuna dikkat ediniz. Bu şekilde  $f(x)$ ,  $x$  rasgele değişkeninin ihtimal yoğunluk fonksiyonunu gösterir.



(a)

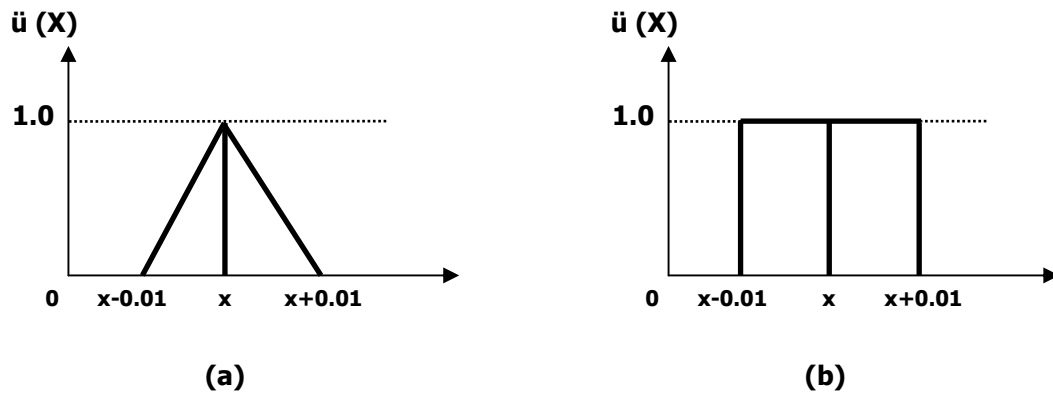


(b)

Şekil 4.15. Gauss (a) bulanık kümesi (b) dağılım fonksiyonu

### 4.3.2. Bulanıklaştırma

Genel olarak, klasik küme şeklinde beliren değişim aralıklarının bulanıklaştırılması, bulanık küme, mantık ve sistem işlemleri için gereklidir. Bunun için, bir aralıkta bulunabilecek öğelerin hepsinin, 1'e eşit üyelik derecesine sahip olacak yerde, 0 ile 1 arasında değişik değerlere sahip olması düşünülür. Bazı öğelerin belirsizlik içerdikleri kabul edilir. Bu belirsizliğin sayısal olmayan durumlardan kaynaklanması halinde bulanıklıktan söz edilir. Özellikle, bazı cihazların prezisyonu diye tabir edilen durumlarda mesela  $\pm 1\%$ 'lik hassaslık (prezisyon), ölçülen  $x$  büyüklüğünün  $x + 0.01$  ve  $x - 0.01$  arasında değişeceği beklentisini ifade eder. Bunun klasik ve bulanık kümelerde gösterilişi Şekil 4.16'da ki gibidir.



Şekil 4.16. Hassaslık (a) bulanık (b) klasik

Buradan, bulanık presizyonun pratikte mantıki olarak daha sağlıklı bir tanım olduğu ortaya çıkar. Buna göre, prezisyon kelimesinden ve değerinden bulanık üyelik fonksiyonunun üçgen şeklinde olması akla ilk gelen durumdur.

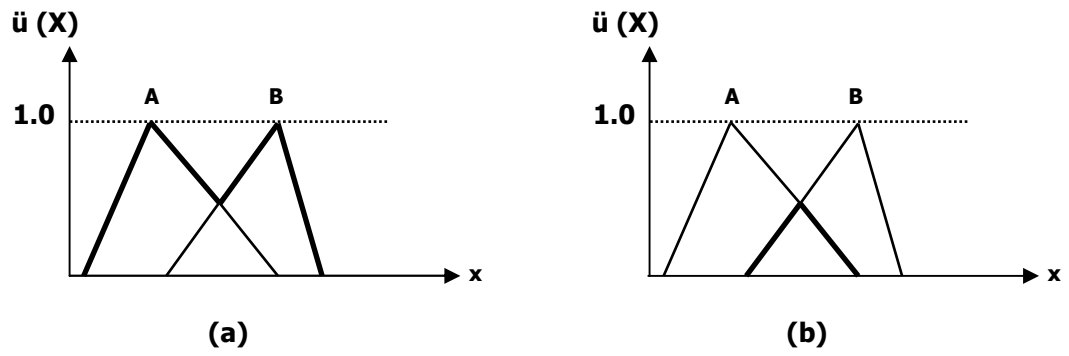
### 4.3.3. Durulaştırma

Bulanık kümelerin üyelik fonksiyonlarının nasıl bulunacağı hakkında gerekli bilgi ve yöntemlerden bahsedildi. Pratik uygulamalarda, özellikle cihaz ve mühendislik plan, proje ve tasarımlarında boyutlandırmalar için kesin sayısal değerlere gerek duyulmaktadır. İşte bu durumlara bulanık olarak elde edilmiş veya verilmiş bilgilerden yararlanarak gerekli cevapların verilmesi için bulanık olan bilgilerin durulaştırılması gerekmektedir. İnsanlar için yapay zekâ çalışmalarında bulanık değişken, küme, mantık ve sistemler öneme sahip olmasına mukabil, bunların bulanık olabilecek çıkarımlarının kesin sayılar haline dönüştürülmesi gerekir. İşte bulanık olan bilgilerin kesin sonuçlar haline dönüştürülmesi için yapılan işlemlerin tümüne birden durulaştırma (defuzzification) işlemleri adı verilir.

### 4.3.4. Durulaştırma işlemleri

Daha önce de belirtildiği gibi, bir bulanık küme işlemi sonucundaki bulanık kümenin tek sayı haline dönüştürülmesi gerekebilir. Bu, bulanıklaştırma işleminin aksi olan

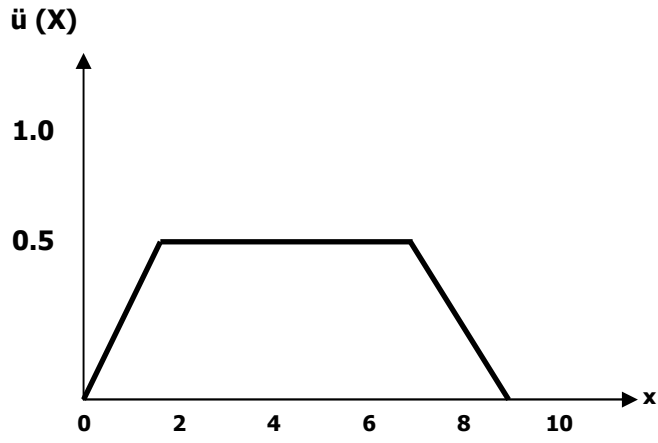
durulaştırma işlemi ile yapılır. Yapılan işlemler sonrasında bulanık sonuçlardan bir tanesi Şekil 4.17 a'daki gibi yamuk, diğerinin ise Şekil 4.17 b'deki gibi üçgen şeklinde olduğunu düşünelim. Bunların ikisinin birleşimi ile yapılan son işlem sonrası bulanık çıkarım kümeleri elde edilir. Şimdi bu son dış bükey olmayan bulanık kümeden tek sayılı bir tasarım büyüklüğünün çıkartılması düşünölsün. İşte bunun için durulaştırma işleminin yapılması gerekecektir.



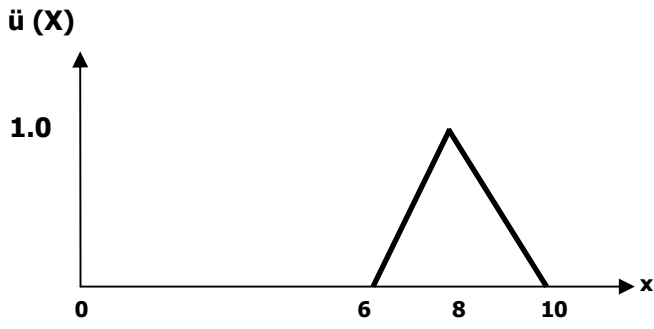
Şekil 4.17. İki bulanık kümenin (a) birleşimi, (b) kesişimi

Tabii olarak Şekil 4.17'da, iki tane bulanık kümenin birleşimi sonucunda elde edilen bulanık çıkarım gösterilmiştir. Hâlbuki değişik şekilleri olan çıkarımların iki veya daha fazla sayıdaki temel bulanık kümelerden çıkması mümkündür.

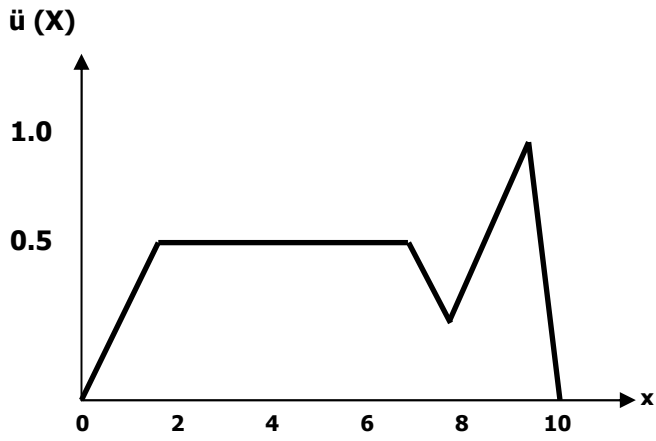
Aşağıda yedi tane durulaştırma işleminin esasları verilecektir. Bunların hangisinin kullanılacağına, elindeki sorunun türüne göre araştırma veya tasarımı yapan mühendisin karar vermesi gereklidir.



(a)



(b)

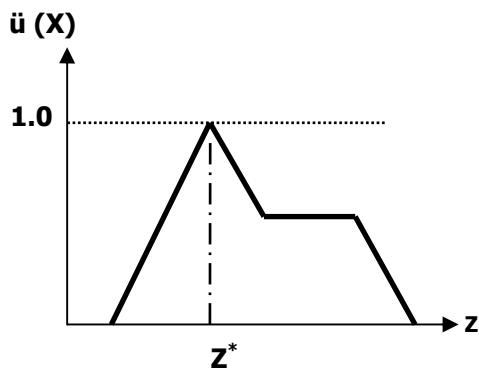


(c)

Şekil 4.18. Tipik bulanık küme çıktısı,  
 (a) bulanık girdi ilk kısım,  
 (b) bulanık girdi ikinci kısım,  
 (c) ikisinin birleşimi

### En büyük üyelik ilkesi

Bunun diğ er bir adı da y ukseklik y ontemidir. Kullanılabilmesi i in tepeleri olan  ıkanın bulanık k umelerine gerek vardır. Őekil 4.19' da g sterilen bu durulařtırma iřleminin aritmetik notasyon Őeklinde g sterimi [19];



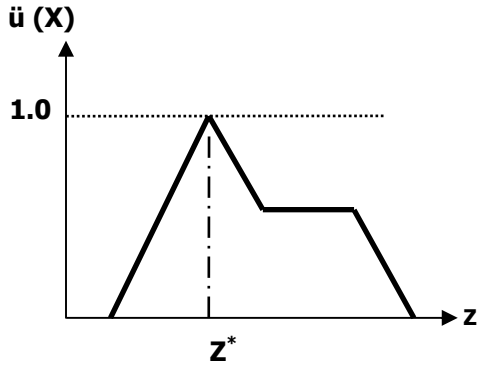
 izelge 4.1. En b y k üyelik derecesi durulařtırması

### Sentroid y ontemi

Bunun diğ er bir adı da ağırlık merkezi y ontemidir. Durulařtırma iřlemlerinde, belki de en yaygın olarak kullanılan iřlem budur. Őekil 4.20' da g sterilmiř olan bu durulařtırmanın matematik iřlemi ařağıdaki denklem vasıtası ile yapılır [18].

$$z^* = \frac{\int \ddot{u}_\zeta(z) \cdot z dz}{\int \ddot{u}_\zeta(z) \cdot dz} \quad (4.2)$$

Burada  $\int$  integral iřaretini g sterir.



Şekil 4.20. Sentroid yöntemi ile durulaştırma

#### Ağırlıklı ortalama yöntemi

Bunun kullanılabilmesi için simetrik üyelik fonksiyonunun bulunması gereklidir. İşlemler matematik olarak

$$z^* = \frac{\sum \ddot{u}_{\xi}(\bar{z}) \cdot \bar{z}}{\sum \ddot{u}_{\xi}(\bar{z})} \quad (4.3)$$

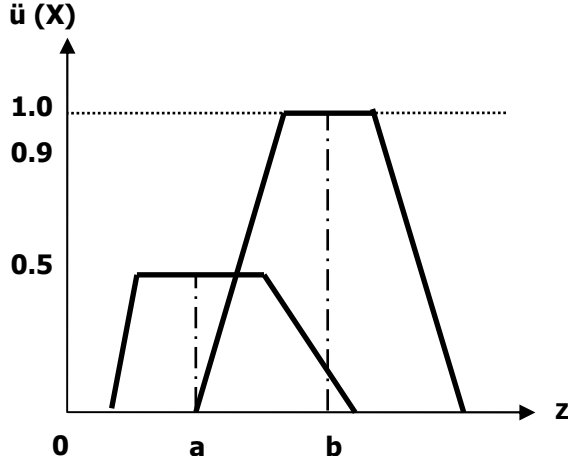
şeklinde yapılır. Burada L işareti cebir anlamında toplamayı gösterir. Bu durulaştırma işlemi Şekil 4.21'de gösterilmiştir. Böylece çıkışı oluşturan bulanık kümelerin üyelik fonksiyonlarının her biri sahip oldukları en büyük üyelik derecesi değeri ile çarpılarak ağırlıklı ortalamaları alınır.

Misal olarak Şekil 4.21'deki iki bulanık kümenin ağırlıklı ortalaması (durulaştırılmış değer).

$$z^* = \frac{a(0.6) + b(0.9)}{0.6 + 0.9} \quad (4.4)$$



olarak bulunur. Bu durulaştırma işlemi sadece simetrik olan üyelik fonksiyonları için geçerli olduğundan, a ve b değerleri temsil ettikleri şekillerin ortalamalarıdır.



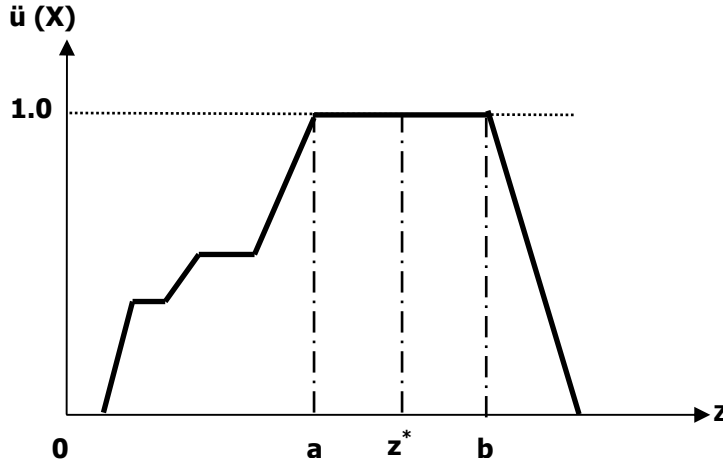
Şekil 4.21. Ağırlıklı ortalama yöntemi durulaştırması

#### Ortalama en büyük üyelik

Bu yöntem aynı zamanda en büyüklerin ortası diye de bilinir. Bu bakımdan birinci durulaştırma ilkesine çok yakındır. Ancak, en büyük üyeliğin konumu tekil olmayabilir. Bunun anlamı üyelik fonksiyonunda en büyük üyelik derecesine sahip olan,  $\ddot{U}A(Z) = 1$ , bir nokta yerine plato gibi düzlük kısmı da bulunabilir. Şekil 4.22'de durulaştırma işlemi gösterilmiş olan bu yöntemle göre durulaştırılmış değer

$$z^* = \frac{a + b}{2} \quad (4.5)$$

olarak bulunur. Buradaki a ve b değerleri şekilde gösterilmiştir [19].



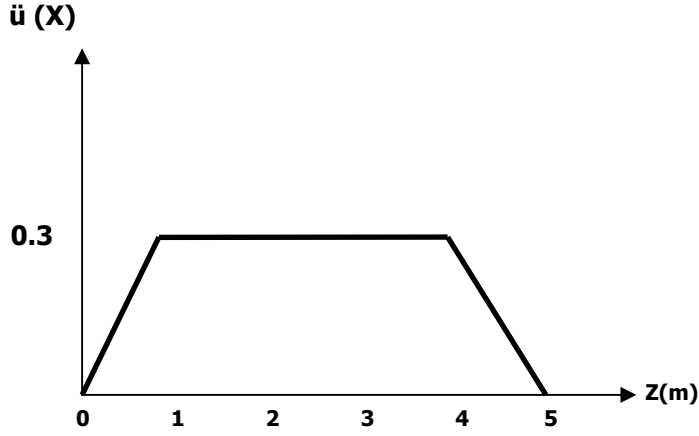
Şekil 4.22. Ortalama en büyük üyelik durulaştırılması

#### Toplamların merkezi

Kullanılan durulaştırma işlemleri arasında en hızlı olanı bu yöntemdir. Bu yöntemde iki bulanık kümenin birleşimi yerine onların cebirsel toplamları kullanılır. Bunun bir mahzuru örtüşen kısımların iki defa toplama girmesidir. Durulaştırılmış değer

$$z^* = \frac{\int_z \sum_{k=1}^n \ddot{u}_{\zeta} (z) dz}{\int_z \sum_{k=1}^n \ddot{u}_{\zeta} (z) dz} \quad (4.6)$$

olarak hesap edilebilir. Bir bakıma, bu hesaplama tarzı, ağırlıklı ortalama durulaştırmasına benzer. Ancak toplamların merkezi yönteminde ağırlıklar ilgili üyelik fonksiyonlarının alanlarıdır. Ortalama ağırlıklar yönteminde ise bu üyelik derecesidir. Toplamların merkezi ile durulaştırma işlemleri Şekil 4.23'de gösterilmiştir [19].



Şekil 4.23. Toplamların merkezi durulaştırması.

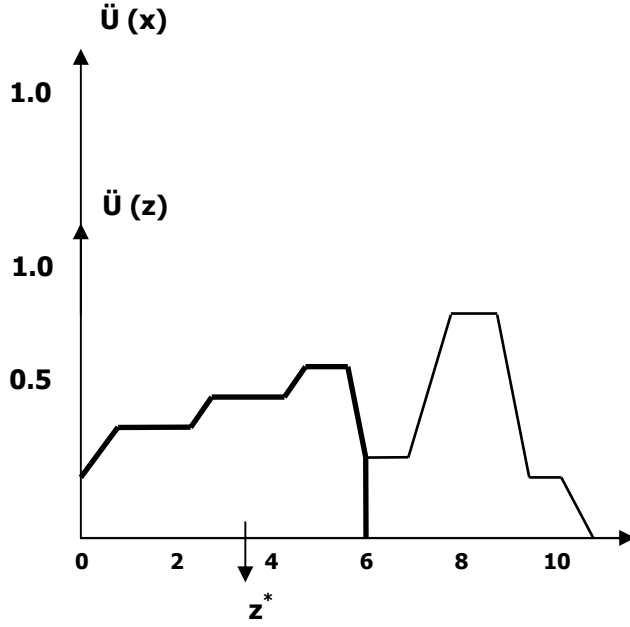
#### En büyük alanın merkezi

Eğer çıkış bulanık kümesi en azından iki tane dış bükey alt bulanık kümeyi içeriyor ise, dış bükey bulanık kümelerin en büyük alanın ağırlık merkezi durulaştırma işleminde kullanılır. Şekil 4.24'de gösterilen durulaştırma işleminin matematik hesaplaması

$$z^* = \frac{\int \ddot{u}_{ab\zeta}(z)zdz}{\int \ddot{u}_{ab\zeta}(z)dz} \quad (4.7)$$

eşitliğine göre yapılır. Burada  $\ddot{U}_{eb\zeta}(Z)$  en büyük alanlı dış bükey bulanık kümenin hakim olduğu alt bölgeyi gösterir.

Bu şart tüm çıkarım bulanık kümesinin dış bükey olmadığı zaman kullanılır, ama tüm çıkarımın dış bükey olması durumunda z. sentroid yöntemi ile elde edilenin aynısıdır [19].



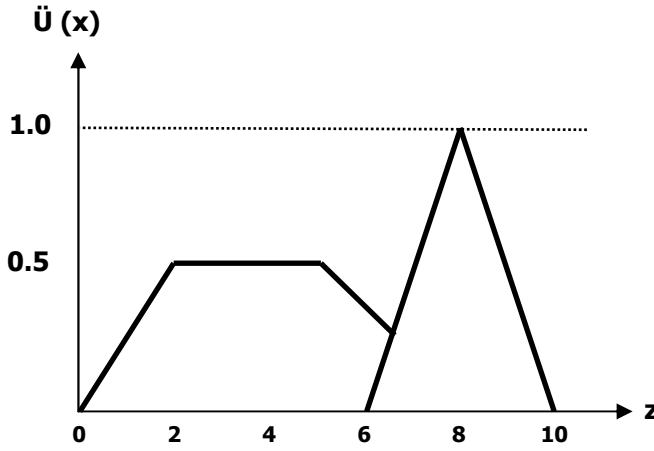
Şekil 4.24. En büyük alan merkezi ile durulaştırma

#### En büyük ilk veya son üyelik derecesi

Bu yöntem de, tüm çıktıların birleşimi olarak ortaya çıkan bulanık küme de en büyük üyelik derecesine sahip olan en küçük (veya en büyük) bulanık küme değerini seçmek esasına dayanır. Hesaplamaların vereceği  $z$ . için aşağıdaki denklemler geçerlidir. Önce bulanık küme çıkarımı,  $B$ , birleşiminde en büyük yükseklik,  $Y_{eb}$  tespit edilir.

$$Y_{eb}(B) = EB[\dot{U}_B(Z)] \quad (4.8)$$

Bundan sonra birinci en büyük değer,  $z$ ., bulunur. Bu yöntemin bir diğer seçeneği ise ilk yerine son en büyük bulanık küme değerinin,  $z$ ., bulunmasıdır. Bu durumlar Şekil 4.25'te gösterilmiştir.



Şekil 4.25. İlk ve son en büyük üyelik dereceleri ile durulaştırma

#### 4.4. Bulanık Mantık Yöntemi ile Örnek Uygulama

Bu örnekte, vadeli krediyle ev almak isteyen üç kişiden hangisine hangi vadeyle kredi verileceğinin bilgisayar tarafından bulanık mantık yaklaşımı kullanılarak belirlenmesi sağlanmıştır.

Örnekte, ev satın almak için vadeli krediye müracaat eden üç kişi bulunmaktadır. Bu kişiler sırasıyla B, C ve D kişileridir. Yalnız, bu kişilerin gelirleri bilgisayarın veritabanında kayıtlı olmadığından kime hangi vadede kredi verileceği bilinmemektedir.

Yalnız, adayların finansal durumları, kullandıkları otomobillerin fiyatlarına bakılarak tahmin edilebilmektedir.

Kredi verilecek adaylar ve sahip oldukları otomobillerin fiyatları aşağıda verilmiştir.

Kişinin İsmi	Otomobilin Fiyatı
B	9.200
C	48.000
D	16.000

Kredi verilecek adayların isimleri ve otomobillerinin fiyatlarının bulunması.

Sezgiler, vadeli ev kredisi verilecek doğru kişinin C kişisi olduğu söylemektedir. Fakat, kredi verilecek en uygun kişinin kim olduğunu bulanık ilişkiler ve bileşim işlemleri yardımıyla bulunabilmektedir.

**Kural 1** : EĞER otomobil “pahalı” İSE gelir “yüksek”

**Kural 2** : EĞER gelir “yüksek” İSE kredi “kısa vadeli” verilebilir.

Bulanık çıkarım için oluşturulan üç adet küme eşitlik I.2’de verilmiştir.

Otomobilin fiyatı = {ucuz, orta, pahalı}

Adayın geliri = {düşük, orta, yüksek}

Kredi ödeme vadesi = {kısa, orta, uzun}

Kural 1, adayın kullandığı “otomobil” ile adayın “geliri” arasındaki ilişkiyi göstermektedir. Bu kural, bulanık ilişki matrisi şeklinde eşitlik I.3’te verilmiştir.

$$\begin{array}{c}
 \text{(gelir)} \\
 \text{düşük} \quad \text{orta} \quad \text{yüksek} \\
 \text{ucuz} \quad \left[ \begin{array}{ccc} 0.6 & 0.3 & 0.3 \end{array} \right] \\
 \text{orta} \quad \left[ \begin{array}{ccc} 0.3 & 0.8 & 0.3 \end{array} \right] \\
 \text{pahalı} \quad \left[ \begin{array}{ccc} 0.1 & 0.3 & 0.7 \end{array} \right]
 \end{array}
 \quad (4.9)$$

Eşitlik 4.9’ da verilen bulanık ilişki matrisinin ilk sütununda, düşük gelire sahip insanların ucuz otomobilleri satın almaya meyilli oldukları görülmektedir. Diğer

sütunlara bakıldığında yine benzer şekilde kişinin geliri ile sahip olduğu otomobilin fiyatı arasında benzer paralel ilişki olduğu gözlenmektedir.

Diğer yandan, adayın geliri ile adaya verilecek kredinin vadesi arasındaki S bulanık ilişkisi eşitlik 4.10'da verilmiştir.

$$\begin{array}{rcc}
 & & \text{(kredi vadesi)} \\
 & & \text{kısa} \quad \text{orta} \quad \text{uzun} \\
 S = (\text{gelir}) & \begin{array}{l} \text{düşük} \\ \text{orta} \\ \text{yüksek} \end{array} & \begin{bmatrix} 0 & 0.3 & 1.0 \\ 0.2 & 0.6 & 0.5 \\ 0.8 & 0.4 & 0.4 \end{bmatrix} & (4.10)
 \end{array}$$

Eşitlik 4.9 ve eşitlik 4.10' da ki bulanık ilişkiler, 4.11 ve eşitlik 4.12' de ki gibi de gösterilebilir.

$$R \subset (\text{otomobilin fiyatı}) \times (\text{gelir}) \quad (4.11)$$

$$S \subset (\text{gelir}) \times (\text{kredi vadesi}) \quad (4.12)$$

Böylece  $R \circ S$  bulanık ilişkisi, (otomobil fiyatı)  $\times$  (kredi vadesi) kartezyen çarpımının sonucunda elde edilmiş olur.

$R \circ S$  birleşimi, max ve min işlemleri kullanılarak eşitlik 4.13'de verildiği gibi elde edilir.

$$\begin{array}{rcc}
 & \text{düşük} & \text{orta} & \text{yüksek} & & \text{kısa} & \text{orta} & \text{uzun} \\
 R \circ S = & \begin{array}{l} \text{ucuz} \\ \text{orta} \\ \text{pahalı} \end{array} & \begin{bmatrix} 0.6 & 0.3 & 0.3 \\ 0.3 & 0.8 & 0.3 \\ 0.1 & 0.3 & 0.7 \end{bmatrix} & \times & \begin{array}{l} \text{düşük} \\ \text{orta} \\ \text{yüksek} \end{array} & \begin{bmatrix} 0 & 0.3 & 1.0 \\ 0.2 & 0.6 & 0.5 \\ 0.8 & 0.4 & 0.4 \end{bmatrix}
 \end{array}$$

$$\begin{array}{l}
 \text{ucuz} \\
 = \text{orta} \\
 \text{pahalı}
 \end{array}
 \begin{array}{l}
 \text{kısa} \quad \text{orta} \quad \text{uzun} \\
 \left[ \begin{array}{ccc}
 0.3 & 0.3 & 0.6 \\
 0.3 & 0.6 & 0.5 \\
 0.7 & 0.4 & 0.4
 \end{array} \right]
 \end{array}$$

Üç adayın otomobillerinin ucuz, orta ve pahalı olmak üzere bulanık matris şeklindeki gösterimi 4.14’de verilmiştir.

$$\begin{array}{l}
 \text{ucuz} \quad \text{orta} \quad \text{pahalı} \\
 \text{B} \\
 \text{C} \\
 \text{D}
 \end{array}
 \begin{array}{l}
 \left[ \begin{array}{ccc}
 0.3 & 0.3 & 0.6 \\
 0.3 & 0.6 & 0.5 \\
 0.7 & 0.4 & 0.4
 \end{array} \right]
 \end{array}
 \tag{4.14}$$

4.14’de verilen bulanık ilişki matrisine göre B adayının otomobili ucuz, D adayının otomobili biraz pahalı, C adayının otomobili ise en pahalı olan otomobildir.

Üç adayın kredi ödeme vadeleri, eşitlik 4.15’de verilen birleşim işlemi sonucunda elde edilir. Örneğin;

B adayının kredi ödeme vadesi eşitlik 4.15’de verilmiştir.

$$\begin{array}{l}
 \text{(otomobil)} \\
 \text{ucuz} \quad \text{orta} \quad \text{pahalı} \\
 [0.9 \quad 0.1 \quad 0]
 \end{array}
 \times
 \begin{array}{l}
 \left[ \begin{array}{ccc}
 0.3 & 0.3 & 0.6 \\
 0.3 & 0.6 & 0.5 \\
 0.7 & 0.4 & 0.4
 \end{array} \right]
 \end{array}
 =
 \begin{array}{l}
 \text{(kredi vadesi)} \\
 \text{kısa} \quad \text{orta} \quad \text{uzun} \\
 [0.3 \quad 0.3 \quad 0.6]
 \end{array}
 \tag{4.15}$$



C adayının kredi ödeme vadesi, eşitlik 4.16'da verilmiştir.

$$\begin{array}{l}
 \text{(otomobil)} \\
 \text{ucuz} \quad \text{orta} \quad \text{pahalı} \\
 [0 \quad 0.2 \quad 0.8]
 \end{array}
 \times
 \begin{bmatrix}
 0.3 & 0.3 & 0.6 \\
 0.3 & 0.6 & 0.5 \\
 0.7 & 0.4 & 0.4
 \end{bmatrix}
 =
 \begin{array}{l}
 \text{(kredi vadesi)} \\
 \text{kısa} \quad \text{orta} \quad \text{uzun} \\
 [0.7 \quad 0.4 \quad 0.4]
 \end{array}$$

D adayının kredi ödeme vadesi, eşitlik 4.17'da verilmiştir.

$$\begin{array}{l}
 \text{(otomobil)} \\
 \text{ucuz} \quad \text{orta} \quad \text{pahalı}
 \end{array}
 \times
 \begin{bmatrix}
 0.3 & 0.3 & 0.6 \\
 0.3 & 0.6 & 0.5 \\
 0.7 & 0.4 & 0.4
 \end{bmatrix}
 =
 \begin{array}{l}
 \text{(kredi vadesi)} \\
 \text{kısa} \quad \text{orta} \quad \text{uzun}
 \end{array}$$

(4.17)

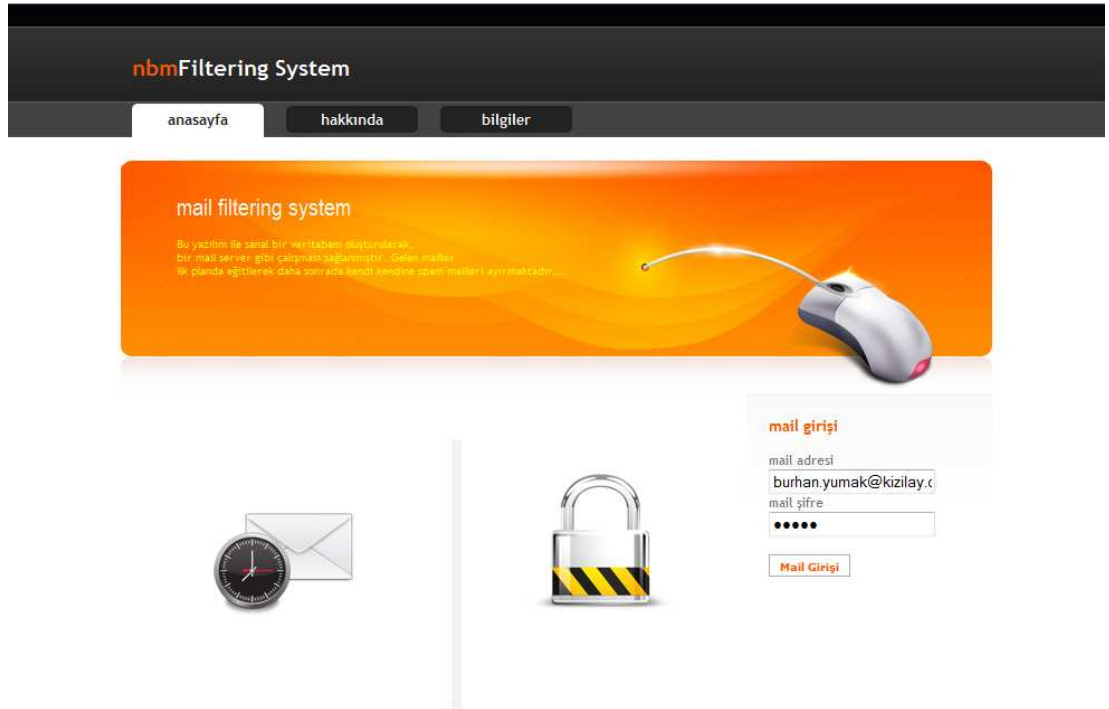
Eğer birleşim işleminin sonuçları karşılaştırılacak olursa, ev almak için B'nin uzun vadeli krediye ihtiyaç duyduğu, D'nin krediyi geri ödeyebilmek için ortalama zamana ihtiyaç duyduğunu ve C'nin krediyi en kısa sürede ödeyebilecek kişi olduğu sonuçlarına ulaşılır.

Böylelikle, C kişinin ev satın almak için kredi verilecek kişi olması olasılığının en yüksek olduğu sonucuna ulaşılır.

## 5. SPAM POSTALARIN NAIVE BAYES VE BULANIK MANTIK YÖNTEMLERİYLE AYRIŞTIRILMASININ KARŞILAŞTIRILMASI

Bu bölümde her iki yöntemin karşılaştırılması için bir simülasyon yazılım geliştirilmiştir. Yazılım, PHP web programlama dilinde yazılmış, MySQL veri tabanı kullanılmıştır. Yazılım, tıpkı bir e-posta sunucusu gibi çalışabilecek şekilde tasarlanmış, sistem dâhilinde olan e-posta kullanıcıları tarafından birbirlerine e-posta göndermeleri istenmiştir. Bu e-posta alışverişlerinde yazılım arka kısımda her iki yöntem için de olasılıkları hesaplayıp veri tabanına yazmıştır. Kullanıcı istediğinde gelen e-postayı da SPAM olarak işaretleyip sistemin kendi kendine eğitilmesine yardımcı olmaktadır.

İlk aşamada uygulamanın çalışma yapısı ve yöntemlerin uygulama üzerinde nasıl kullanıldığından bahsedilmiş, ikinci aşamada ise uygulama sonuçlarına yer verilmiştir. Yazılımın ana ekran görüntüsü Resim 5.1.'deki gibidir.



Resim 5.1. Uygulama ana ekran görüntüsü

## 5.1. Uygulama Üzerinden Naive Bayes ve Bulanık Mantık Yöntemlerinin Kullanılması

Yazılım, genel bir e-posta sunucusu gibi çalışacak şekilde tasarlanmıştır. Yazılımın veritabanı genel görüntüsü Resim 5.2.' de gösterilmiştir. Kullanıcılara denemeler için sanal e-posta adresleri oluşturulmuştur. Kullanıcılar farklı yerlerden Resim 5.3. de olduğu gibi sisteme giriş yaparak, diğer test kullanıcılarına e-postalar göndermiştir. Kullanıcılara her e-posta geldiğinde sistem otomatik devreye girerek kullanıcı e-postayı görmeden önce çalışmaktadır. Uygulama genelinde de seçilen bir kullanıcıya gelen e-postalar incelemeye alınmış ve sonuçlar değerlendirilmiştir. Naive Bayes ve Bulanık Mantık yöntemlerinin sistemde nasıl uygulandığı incelenen kullanıcıya gelen bir e-posta üzerinden açıklanmıştır.

Sunucu: localhost ▶ Veritabanı: spamfiltering

Tablo:	Eylem	Kayıtlar <sup>1</sup>	Türü	Karşılaştırma	Boyut	Ek Yük
blacklist		0	MyISAM	latin1_swedish_ci	1.0 KİB	-
filter		1	MyISAM	latin1_swedish_ci	3.1 KİB	-
mail		18	MyISAM	latin1_swedish_ci	5.3 KİB	-
spamwords		36	MyISAM	latin1_swedish_ci	3.7 KİB	-
user		5	MyISAM	latin1_swedish_ci	3.3 KİB	-
<b>5 tablo</b>	<b>Toplam</b>	<b>60</b>	<b>InnoDB</b>	<b>latin1_swedish_ci</b>	<b>16.3 KİB</b>	<b>0 B</b>

↑ Tümünü Seç / Hiçbirini Seçme Seçimleri:

Baskı görünümü Veri sözlüğü

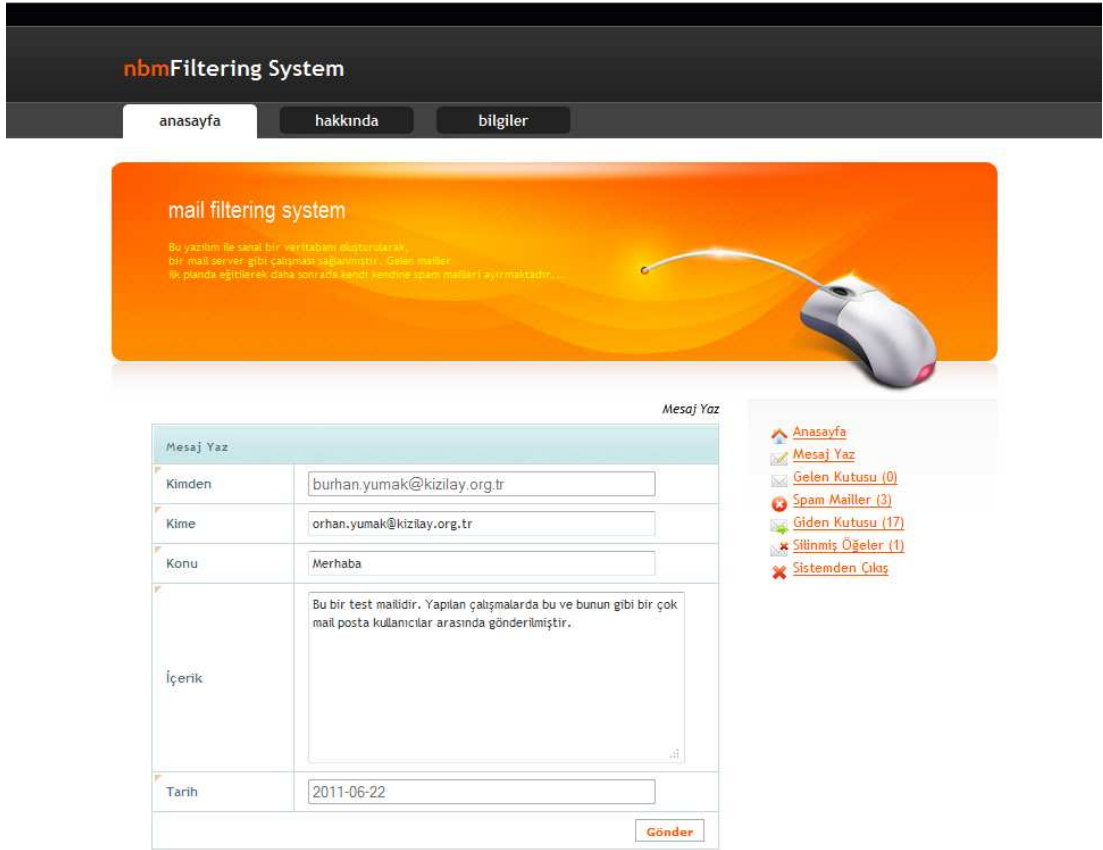
spamfiltering veritabanında yeni tablo oluştur

İsim:  Alan sayısı:

<sup>1</sup> Yaklaşık olabilir. SSS 3.11'e bakın

Yeni phpMyAdmin penceresi aç

Resim 5.2. Uygulama veritabanı genel görüntüsü



Resim 5.3. Sisteme giriş yapan kullanıcının diğer kullanıcıya e-posta göndermesi



Resim 5.4. Kullanıcıya gelen e-postanın okunması ve SPAM olarak işaretlenmesi

**nbmFiltering System**

anasayfa hakkında bilgiler

**mail filtering system**

Bu yazılım ile sanal bir veri tabanı oluşturularak bir mail serveri gibi çalışması sağlanmıştır. Gelen mailler ilk planda eğitilerek daha sonra adı kendi kendine spam mailleri ayırmaktadır....

1 2

Tüm Mailler

Durum	Kimden	Tarih	İşlemler	id
E	burhan.yumak@kizilay.org.tr burhan.yumak@kizilay.org.tr	2011-06-21	📧 📧 ✖ ⚙	16
E	burhan.yumak@kizilay.org.tr burhan.yumak@kizilay.org.tr	2011-06-21	📧 📧 ✖ ⚙	15
E	burhan.yumak@kizilay.org.tr burhan.yumak@kizilay.org.tr	2011-06-21	📧 📧 ✖ ⚙	14
E	burhan.yumak@kizilay.org.tr burhan.yumak@kizilay.org.tr	2011-06-21	📧 📧 ✖ ⚙	13

- Anasayfa
- Mesaj Yaz
- Gelen Kutusu (0)
- Spam Mailler (3)
- Giden Kutusu (17)
- Silinmiş Öğeler (1)
- Sistemden Çılas

Resim 5.5. Kullanıcı gelen e-posta kutusu

**nbmFiltering System**

anasayfa hakkında bilgiler

**mail filtering system**

Bu yazılım ile sanal bir veri tabanı oluşturularak bir mail serveri gibi çalışması sağlanmıştır. Gelen mailler ilk planda eğitilerek daha sonra adı kendi kendine spam mailleri ayırmaktadır....

Spam Mailler

Durum	Kimden	Tarih	İşlemler	id
E	burhan.yumak@kizilay.org.tr burhan.yumak@kizilay.org.tr	2011-06-21	📧 📧 ✖ ⚙	10
E	burhan.yumak@kizilay.org.tr konu1	2011-06-21	📧 📧 ✖ ⚙	2
E	burhan.yumak@kizilay.org.tr burhan.yumak@kizilay.org.tr	2011-06-21	📧 📧 ✖ ⚙	1

Toplam Kayıt Sayısı: 3

- Anasayfa
- Mesaj Yaz
- Gelen Kutusu (0)
- Spam Mailler (3)
- Giden Kutusu (17)
- Silinmiş Öğeler (1)
- Sistemden Çılas

Resim 5.6. Kullanıcı SPAM e-posta kutusu

Her iki yöntemin yazılım üzerinde kullanılabilmesi için SPAM postalar ait bir filtreleme tablosuna ihtiyaç vardır. Bu tabloda her iki yöntem için de bu sınıflandırma değerlerine ait olasılık rakamları tutulmuştur. Sınıflandırma değerleri SPAM postaların ayrıştırılmasında önemli bir öncelik teşkil etmektedir. Bu sınıflandırma değerleri genel SPAM karakteristiklerinden seçilerek sistemde kullanılmıştır.

### **5.1.1. SPAM postaların ayrıştırılmasında sınıflandırma değerleri**

Yazılım SPAM postaların ayrıştırılmasında kullanılan iki yöntemin karşılaştırılmasını incelemiştir. Her iki yöntem ile de bir kullanıcıya gelen postalar üzerinden yola çıkılarak yeni gelmiş olan bir e-postanın kullanıcı için SPAM posta olup olmadığı sonucu çıkarılmıştır. Bu karar verme sürecinde ise adı geçen iki yöntemden de faydalanarak, hangisinin bu durumda karar verme olasılığının daha yüksek olduğu ve doğru sonuçlar verdiği karşılaştırılmıştır.

E-postaların SPAM durumlarının incelenmesinde, Kara liste, kelime durumu, domain durumu, isim geçme durumu, bilinirlik durumu sınıflandırma değerleri kullanılmıştır.

Adı geçen bu sınıflandırma değerleri, postaların SPAM olmasında etkili değerlerdir.

#### Kara liste (black list)

Kara liste, yaygın bir SPAM posta engelleme tekniğidir. Hesaplama yükü yoktur ve gerçekleştirilmesi kolaydır. Bu teknik basitçe bilinen istenmeyen elektronik postacıların IP adreslerinin bir listesini el ile tutan kurumları içerir, böylece bu adreslerden gelen elektronik postalar engellenir. İstenmeyen elektronik postacılar düzenli olarak IP adreslerini değiştirdiğinden ve geniş aralıkta IP adresleri kullandığından, kara listeler kısa zaman aralıklarında, küçük istenmeyen elektronik posta miktarlarını engellemede çok etkindirler. Tek bir kaynağa özgü istenmeyen elektronik postayı engellemede çabuk bir çözüm sağlarlar, fakat genel bir istenmeyen elektronik posta filtreleme çözümü olarak etkin değildirlir.

### Kelime durumu

Belge sınıflandırma (metin sınıflandırma veya kategorize etme olarak da bilinir), içeriğine dayanarak bir belgeyi bir veya daha fazla kategoriye ayırma işlemidir. Gelen metin belgeleri, bir talim kümesinde etiketlenmiş belgelerden çıkarılan bilgiye dayanarak oluşturulmuş, önceden tanımlanmış kategorilere atanır. Her iki yöntemden önce de posta değerlendirmede kullanılacak olan bu sınıflandırma, istenmeyen postaların ayırt edilmesinde kullanılan veri kümesi (kelime grupları) tarafından taranır ve posta bir genel değerlendirme kümesine konulur. Burada sınıflandırma kümeleri Zayıf, Normal ve Güçlü olarak belirlenmiştir. Zayıf posta içerisinde bulunan istenmeyen kelimelerin hiç olmadığını ya da çok az bulunduğunu belirtir.

### Domain kontrol

Bu, sahte bir from veya return adresi kullanan SPAM postacıardan gelen istenmeyen elektronik postaların engellenmesinde etkin bir tekniktir. İstenmeyen elektronik postacılar hileli (fake) adresler kullanırlar ve bu sayede gönderdikleri istenmeyen elektronik posta onları bulmada bir iz olamaz.

From adresinin geçerli olup olmadığını belirlemek için, sistem from adresinde kullanılan domain üzerinde bir arama yapar. Eğer domain geçerli bir MX kaydına sahip değilse, from adresi de geçerli değildir ve elektronik posta istenmeyen elektronik posta olarak etiketlenir.

### İsim kontrol

Bu karakteristik de genelde SPAM postalarda sıkça rastlanan kullanıcının adının direkt olarak yazılarak gönderildiği postaları ayırtırmak için kullanılır. Bu özellik genel bir ayırt edici özellik olsa da genel bir istenmeyen elektronik posta filtreleme çözümü olarak kendi başına etkin değildir.

### Önceden gelme durumu (bilinirlik)

Bir kullanıcıya gelen elektronik postaların SPAM posta olarak ayırt edilmesinde, postayı yollayan kişinin bilinirliği çok önemlidir. Kullanıcı güvendiği ve bildiği bir kullanıcıdan istenmeyen posta geleceğini düşünmez. Genel olarak görüş böyledir. Ancak yapay zekâ düşünüldüğünde daha önceki postalarda kullanıcının bilinebilirliği de çok önemli bir ayırıştırma karakteristiği oluşturur.

Kullanıcıya gelen e-postanın SPAM durumunun tahmin edilmesinde önceden gelen e-postaların sınıflandırma değerlerine bakılarak oluşturulmuş veri kümeleri önemli birer kaynaktır. Her iki yöntem de kullanıcıya gelen önceki postalar üzerinden tahmin olasılıkları çıkarmıştır. Örnek olarak gelen postanın incelenmesi için oluşturulmuş sınıflandırma değerleri tablosu Çizelge 5.1.' de ki gibidir.

Çizelge 5.1. Gelen postaların SPAM sınıflandırma değerleri ve durumları

İstenmeyen Posta (SPAM) Olma Sınıflandırma Değerleri						
Mail	Kara Liste	Kelime Durumu	Domain Var mı	İsim Geçiyor mu	Önceden Geldi mi	SPAM Durumu
	Evet / Hayır	Zayıf / Normal / Güçlü	Evet / Hayır	Evet / Hayır	Evet / Hayır	Evet / Hayır
1	Evet	Normal	Evet	Hayır	Hayır	<b>Evet</b>
2	Hayır	Zayıf	Evet	Hayır	Evet	<b>Hayır</b>
3	Evet	Güçlü	Hayır	Evet	Hayır	<b>Evet</b>
.	.	.	.	.	.	.
.	.	.	.	.	.	.
20	Hayır	Normal	Evet	Evet	Hayır	<b>Hayır</b>
21	Hayır	Zayıf	Hayır	Hayır	Evet	<b>Evet</b>
22 ?	<b>Evet</b>	<b>Zayıf</b>	<b>Evet</b>	<b>Evet</b>	<b>Hayır</b>	<b>?</b>



### 5.1.2. Gelen postalara ait oluşturulan veri kümeleri

Kullanıcıya daha önceden gelen postalar üzerinden Çizelge 5.1.' de olduğu gibi bir tablo çıkarılmıştır. Bu tablo üzerinde önceden gelmiş olan 21 postanın sınıflandırma değerlerine göre sınıflandırılmaları ve SPAM durumları görülmektedir. Bu değerler ışığında bazı frekans tabloları çıkarılmıştır.

Kullanıcıya gelen postalardan Kara Listede olup SPAM olarak değerlendirilen posta sayısı 6, SPAM olarak değerlendirilmeyen posta sayısı 3 iken, Kara Listede olmayıp SPAM olarak değerlendirilen posta sayısı 6, SPAM olarak değerlendirilmeyen posta sayısı da 6 olarak Çizelge 5.2' de çıkarılmıştır.

Çizelge 5.2. Kara liste durumuna göre veri frekans tablosu

<b>Kara Liste</b>	<b>SPAM EVET</b>	<b>SPAM HAYIR</b>
<b>VAR</b>	6	3
<b>YOK</b>	6	6

Gelen postalardan Kelime durumlarına göre “Zayıf” olan ve SPAM olarak değerlendirilen posta sayısı 6, SPAM olarak değerlendirilmeyen posta sayısı 6, Kelime durumlarına göre “Normal” olan ve SPAM olarak değerlendirilen posta sayısı 3, SPAM olarak değerlendirilmeyen posta sayısı 3 iken, Kelime durumlarına göre “Güçlü” olan ve SPAM olarak değerlendirilen posta sayısı 3, SPAM olarak değerlendirilmeyen posta sayısı da 0 olarak Çizelge 5.3' de çıkarılmıştır.

Çizelge 5.3. Kelime durumuna göre veri frekans tablosu

<b>Kelime Durumu</b>	<b>SPAM EVET</b>	<b>SPAM HAYIR</b>
<b>ZAYIF</b>	6	6
<b>NORMAL</b>	3	3
<b>GÜÇLÜ</b>	3	0

Gelen postalardan Domainleri geçerli olan ve SPAM olarak değerlendirilen posta sayısı 6, SPAM olarak değerlendirilmeyen posta sayısı 9 iken, Domainleri geçerli olmayan ve SPAM olarak değerlendirilen posta sayısı 6, SPAM olarak değerlendirilmeyen posta sayısı da 0 olarak Çizelge 5.4' de çıkarılmıştır.

Çizelge 5.4. Domain durumuna göre veri frekans tablosu

<b>Domain Kontrol</b>	<b>SPAM EVET</b>	<b>SPAM HAYIR</b>
<b>VAR</b>	6	9
<b>YOK</b>	6	0

Gelen postalardan içerisinde kullanıcının direkt adı geçen ve SPAM olarak değerlendirilen posta sayısı 6, SPAM olarak değerlendirilmeyen posta sayısı 3 iken, içerisinde direkt adı geçmeyen ve SPAM olarak değerlendirilen posta sayısı 6, SPAM olarak değerlendirilmeyen posta sayısı da 6 olarak Çizelge 5.5' de çıkarılmıştır.

Çizelge 5.5. İsim geçme durumuna göre veri frekans tablosu

<b>İsim Kontrol</b>	<b>SPAM EVET</b>	<b>SPAM HAYIR</b>
<b>EVET</b>	6	3
<b>HAYIR</b>	6	6

Gelen postalardan kullanıcıya daha önce gönderilmiş olan ve SPAM olarak değerlendirilen posta sayısı 6, SPAM olarak değerlendirilmeyen posta sayısı 3 iken, kullanıcıya daha önce gönderilmemiş olup ve SPAM olarak değerlendirilen posta sayısı 6, SPAM olarak değerlendirilmeyen posta sayısı da 6 olarak Çizelge 5.6' da çıkarılmıştır.

Çizelge 5.6. Önceden gelme durumuna göre veri frekans tablosu

<b>Önceden Geldi mi?</b>	<b>SPAM EVET</b>	<b>SPAM HAYIR</b>
<b>EVET</b>	6	3
<b>HAYIR</b>	6	6

### 5.1.3. Gelen postalara ait oluşturulan olasılıklar

Gelen postalardan Kara Listede olanların SPAM olarak değerlendirilme olasılığı 6/12, SPAM olarak değerlendirilmeme olasılığı 3/9 iken, Kara Listede olmayanların SPAM olarak değerlendirilme olasılığı 6/12, SPAM olarak değerlendirilmeme olasılığı da 6/9 olarak Çizelge 5.7’ de çıkarılmıştır.

Çizelge 5.7. Kara liste durumuna göre olasılık tablosu

<b>Kara Liste</b>	<b>SPAM EVET</b>	<b>SPAM HAYIR</b>
<b>VAR</b>	6/12	3/9
<b>YOK</b>	6/12	6/9

Gelen postalardan Kelime Durumu “Zayıf” olup SPAM olarak değerlendirilme olasılığı 6/12, SPAM olarak değerlendirilmeme olasılığı 6/9, Kelime Durumu “Normal” olup SPAM olarak değerlendirilme olasılığı 3/12, SPAM olarak değerlendirilmeme olasılığı 3/12 iken, Kelime Durumu “Güçlü” olup SPAM olarak değerlendirilme olasılığı 3/12, SPAM olarak değerlendirilmeme olasılığı da 0/9 yani 0 olarak Çizelge 5.8’ de ki gibidir.

Çizelge 5.8. Kelime durumuna göre olasılık tablosu

<b>Kelime Durumu</b>	<b>SPAM EVET</b>	<b>SPAM HAYIR</b>
<b>ZAYIF</b>	6/12	6/9
<b>NORMAL</b>	3/12	3/9
<b>GÜÇLÜ</b>	3/12	0/9

Gelen postalardan Domaini geçerli olanların SPAM olarak değerlendirilme olasılığı 6/12, SPAM olarak değerlendirilmeme olasılığı 9/12 iken, Domaini geçerli olmayanların SPAM olarak değerlendirilme olasılığı 6/12, SPAM olarak değerlendirilmeme olasılığı da 0/12 yani 0 olarak Çizelge 5.9' de çıkarılmıştır.

Çizelge 5.9. Domain durumuna göre olasılık tablosu

<b>Domain Kontrol</b>	<b>SPAM EVET</b>	<b>SPAM HAYIR</b>
<b>VAR</b>	6/12	9/12
<b>YOK</b>	6/12	0/12

Gelen postalardan içerisinde kullanıcının adının direkt geçtiği postaların SPAM olarak değerlendirilme olasılığı 6/12, SPAM olarak değerlendirilmeme olasılığı 3/9 iken, içerisinde kullanıcının adının direkt geçmediği postaların SPAM olarak değerlendirilme olasılığı 6/12, SPAM olarak değerlendirilmeme olasılığı da 6/9 olarak Çizelge 5.10' da ki gibidir.

Çizelge 5.10. İsim geçme durumuna göre olasılık tablosu

<b>İsim Kontrol</b>	<b>SPAM EVET</b>	<b>SPAM HAYIR</b>
<b>EVET</b>	6/12	3/9
<b>HAYIR</b>	6/12	6/9

Gelen postalardan kullanıcıya daha önce gönderilmiş olanların SPAM olarak değerlendirilme olasılığı 6/12, SPAM olarak değerlendirilmeme olasılığı 3/9 iken, daha önce gönderilmemiş olanların SPAM olarak değerlendirilme olasılığı 6/12, SPAM olarak değerlendirilmeme olasılığı da 6/9 olarak Çizelge 5.11’ de çıkarılmıştır.

Çizelge 5.11. Önceden gelme durumuna göre olasılık tablosu

<b>Önceden Geldi mi?</b>	<b>SPAM EVET</b>	<b>SPAM HAYIR</b>
<b>EVET</b>	6/12	3/9
<b>HAYIR</b>	6/12	6/9

Çizelge 5.12. Yeni gelen postanın sınıflandırma değerleri

<b>İstenmeyen Posta (SPAM) Olma Sınıflandırma Değerleri</b>						
<b>Mail</b>	<b>Kara Liste</b>	<b>Kelime Durumu</b>	<b>Domain Var mı</b>	<b>İsim Geçiyor mu</b>	<b>Önceden Geldi mi</b>	<b>Spam Durumu</b>
	Evet / Hayır	Zayıf / Normal / Güçlü	Evet / Hayır	Evet / Hayır	Evet / Hayır	<b>Evet / Hayır</b>
<b>22</b>	<b>Evet</b>	<b>Zayıf</b>	<b>Evet</b>	<b>Evet</b>	<b>Hayır</b>	<b>?</b>

#### 5.1.4. Naive Bayesian yöntemine göre elektronik postanın ayrıştırılması

Kullanıcıya önceden gelen postalar üzerinden bir sınıflandırma değerleri tablosu çıkarılmıştır. Ayrıca tablo da önceden gelen postaların SPAM durumları da verilmiştir. Naive Bayesian ile elektronik postaların ayrıştırılmasında bu sınıflandırma değerleri tablosu analiz edilerek çıkarılmış olasılık oranları kullanılmıştır. İlk adımda bu olasılık değerlerine göre SPAM durumu ya da SPAM olmama durumuna göre iki ayrı olasılık sonucu çıkarılmıştır.

$$P(\text{EVET}) = 6/12 \cdot 6/12 \cdot 6/12 \cdot 6/12 \cdot 6/12 = 0,0313$$

$$P(\text{HAYIR}) = 6/9 \cdot 3/9 \cdot 9/12 \cdot 6/9 \cdot 6/9 = 0,0741$$

Buna göre gelen postalardan yola çıkılarak yeni gelen postanın SPAM olma olasılığı 0.0313 çıkarken, SPAM olmama olasılığı 0.0741 çıkmıştır. Ancak, Naive Bayes teoremine göre sonuç için sadece bu olasılıklar değil sınıfın toplam olasılık değeri de hesaba katılmalıdır. Buna göre toplam olasılık hesabı da işlem katılmış ve çıkan sonuçlar ile çarpılmıştır;

$$P(\text{EVET}) = 0,0313 \cdot 12/21 = 0,18$$

$$P(\text{HAYIR}) = 0,0741 \cdot 9/21 = 0,32$$

İşlemler sonucunda çıkan değerlere bakılarak maksimum sınıf seçilmiştir.  $P(\text{HAYIR}) > P(\text{EVET})$  olduğu için yeni durum “HAYIR” olarak sınıflandırılmıştır. Bu da olağan verilere bağlı olarak değerlendirilen yeni elektronik postanın % 32 ye, % 18 ihtimalle SPAM posta olmayacağı sonucunu vermektedir.

### 5.1.5. Bulanık mantık yöntemine göre elektronik postanın ayrıştırılması

Elektronik postaların ayrıştırılmasında kullanılan Bulanık Mantık teoreminde, ilk önce veri kümesine göre özellik ve sonuç kümeleri oluşturulmuştur.

Buna göre,

Özellik = {Kara Liste, Kelime Durumu, Domain, İsim, Bilinirlik}

Sonuç = {Evet, Hayır}

Elektronik posta özelliklerinin olasılıkları ve sonuç ilişkileri göz önünde bulundurulduğunda aralarında ki bulanık ilişki aşağıda verilmiştir.

	EVET	HAYIR
Kara Liste	6 / 12	6 / 9
Kelime Durumu	6 / 12	3 / 9
Domain	6 / 12	9 / 12
İsim	6 / 12	6 / 9
Bilinirlik	2 / 4	2 / 3

	EVET	HAYIR
Kara Liste	0.50	0.67
Kelime Durumu	0.50	0.33
Domain	0.50	0.75
İsim	0.50	0.67
Bilinirlik	0.50	0.67

Şekil 5.1. Elektronik postaların bulanık ilişkisi

Elde edilen bulanık ilişkilere bakıldığında, elektronik postanın istenmeyen posta çıkma ihtimali için, “Kara Liste” olasılığının 0.50, “Kelime Durumu” olasılığının 0.50, “Domain” olasılığının 0.50, “İsim” olasılığının 0.50 ve “Bilinirlik” olasılığının

0.50 olduğu görülmüştür. Burada ki sayılara göre kullanıcıya gelen elektronik postaların ne kadarının istenmeyen posta olup olmadığı bilgisi çıkartılmıştır. Burada daha doğru sonuçlar çıkarabilmek için Özellik kümesinin genişletilmesi gerekmiş, bu şekilde daha doğru bir tahmin sağlanmıştır.

Kullanıcıya yeni gelen elektronik postanın özelliklerine dair bilgiler ise Şekil 5.2 de verilmiştir.

Kara Liste	Kelime Durumu	Domain	İsim	Bilinirlik
[9 / 21	12 / 21	15 / 21	9 / 21	12 / 21 ]

Kara Liste	Kelime Durumu	Domain	İsim	Bilinirlik
[0 .43	0 .57	0 .71	0 .43	0 .57 ]

Şekil 5.2. Yeni gelen elektronik postanın özellikleri

Tüm özellikler toplandığında, bulanıklaşan verileri durulaştırma işlemine geçmek için verilerin birleştirilmesi işlemi gerçekleştirilmiştir.

$$[0.43 \ 0.57 \ 0.71 \ 0.43 \ 0.57] \times \begin{bmatrix} 0.50 & 0.67 \\ 0.50 & 0.33 \\ 0.50 & 0.75 \\ 0.50 & 0.67 \\ 0.50 & 0.67 \end{bmatrix}$$



Yeni gelen elektronik postanın SPAM Posta olma ihtimali için,

$$E = \begin{bmatrix} (0.43 \wedge 0.50) \vee (0.57 \wedge 0.50) \vee (0.71 \wedge 0.50) \vee (0.43 \wedge 0.50) \vee (0.57 \wedge 0.50) \\ (0.43 \wedge 0.67) \vee (0.57 \wedge 0.33) \vee (0.71 \wedge 0.75) \vee (0.43 \wedge 0.67) \vee (0.57 \wedge 0.67) \end{bmatrix}$$

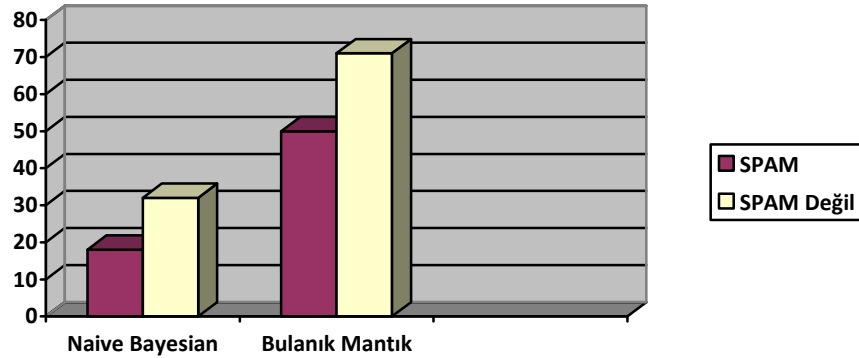
$$E = \begin{bmatrix} (0.43) \vee (0.50) \vee (0.50) \vee (0.43) \vee (0.50) \\ (0.43) \vee (0.33) \vee (0.71) \vee (0.43) \vee (0.57) \end{bmatrix}$$

$$\begin{array}{cc} \text{EVET} & \text{HAYIR} \\ E = [0.50 & 0.71] \end{array}$$

Çıkan sonuçlara göre yeni gelen elektronik postanın %71 e, %50 oranıyla SPAM Posta olma ihtimalinin olmadığı sonucu ortaya çıkmıştır. Buna göre yeni gelen elektronik posta bulanık mantık teoremine göre SPAM posta olarak işaretlenmemiştir.

## 5.2. Naive Bayesian ve Bulanık Mantık Yöntemine Göre Elektronik Posta Ayırıştırılmasının Karşılaştırılması

Kullanıcıya gelmiş olan ve örnek olarak alınan postanın SPAM olmadığı bilgisi bilinmektedir, bu bilgi ışığında postanın ayırıştırılmasında her iki yöntemin de ne kadar etkili olduğu ayrı ayrı denenmiştir. Sonuçlara göre, Naive Bayesian filtreleme metodu % 32 ye, % 18 lik bir tahmin sonucu ile gelen postanın SPAM bir posta olmadığı sonucu çıkarmıştır. Buna göre gelen posta %32 lik ihtimal ile normal bir elektronik postadır ve bu doğru bir sonuçtur. Öte yandan diğer yöntem olan Bulanık Mantık yöntemine göre ise %71 e, %50 lik bir tahmin sonucu ile gelen postanın SPAM bir posta olmadığı sonucu her iki metodunda istenmeyen postaların ayırıştırılmasında doğru bir yol izlendiği sonucunu vermektedir.



Şekil 5.3. Kullanıcıya gelen ve örnek alınan postanın Naive Bayesian ve Bulanık Mantık yöntemine göre ayırıştırma oranları

Şekil 5.3.' de görüldüğü gibi Naive Bayesian yönteminde ki hata payı Bulanık Mantık yönteminde ki hata payına göre daha düşüktür.

Ancak bu rakamlar kullanıcıya gelen herhangi bir posta üzerinden çıkarılmıştır. Genel bir karşılaştırma yapmak için, kullanıcıya gelen toplam 100 e-posta üzerinden de yöntemlerin incelemesi yapılmıştır. Buna göre kullanıcıya gelen e-postaların ne

kadarının SPAM posta olduğu Çizelge 5.13.'de ve karşılaştırılan yöntemlerin çıkardıkları SPAM posta sayıları da Çizelge 5.14.' de verilmiştir.

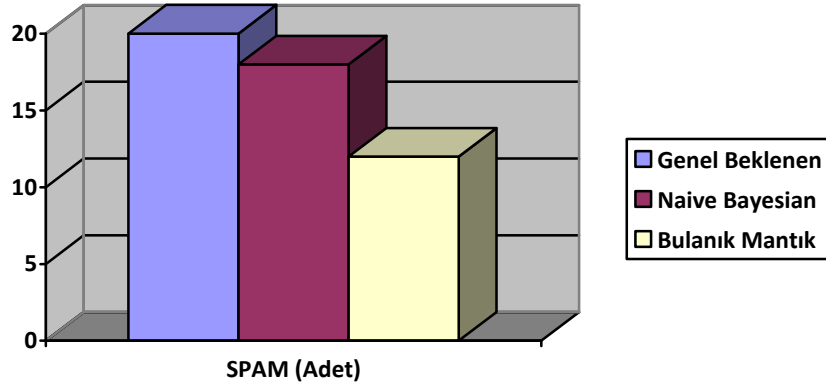
<b>Açıklama</b>	<b>Adet</b>
<b>Kullanıcıya Gelen E-Posta Sayısı</b>	100
<b>Toplam SPAM Olmayan Posta Sayısı</b>	22
<b>Toplam SPAM Posta Sayısı</b>	78

Çizelge 5.13. Kullanıcıya gelen ve olması gereken SPAM e-posta sayıları

Çizelge 5.14.' de ki değerlere bakıldığında, Bulanık Mantık yöntemi elektronik postaların ayrıştırılmasında % 54.6 oranında bir doğruluk sonucu çıkarmıştır. Diğer taraftan Naive Bayesian yöntemi ise % 81.8 oranında bir doğruluk sonucu çıkarmıştır (Şekil 5.5.). Olasılık yüzdelere ve doğruluk sonuçlarına bakıldığında elektronik postaların ayrıştırılmasında Naive Bayesian yönteminin Bulanık Mantık yöntemine göre daha etkili bir sonuç çıkardığı görülmüştür. Bu açıdan bakıldığında Naive Bayesian yöntemi ile SPAM postaların ayrıştırılması daha güvenli hale geleceği sonucuna varılmıştır.

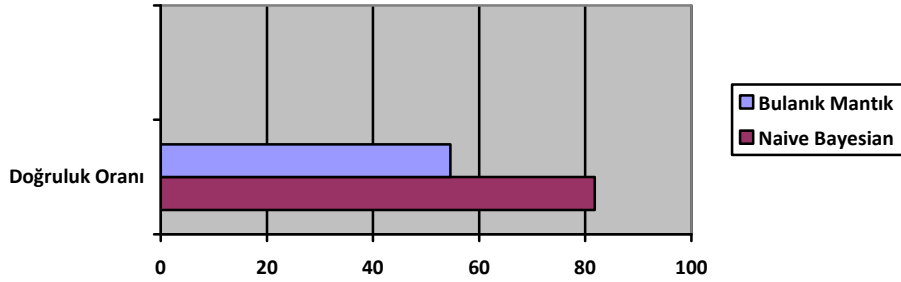
	<b>Toplam Gelen Posta</b>	<b>SPAM Olmayan</b>	<b>SPAM Olan</b>
<b>Genel Beklenen</b>	100	78	22
<b>Naive Bayesian</b>	100	82	18
<b>Bulanık Mantık</b>	100	88	12

Çizelge 5.14. Kullanıcıya gelen e-postaların ayrıştırılmasında Naive Bayesian ve Bulanık Mantık yöntemlerinin SPAM ayrıştırma sonuç tablosu



Şekil 5.4. Beklenen SPAM posta sayısı ile Naive Bayesian ve Bulanık Mantık yöntemi ile ayrıştırılan SPAM postaların karşılaştırılması

Şekil 5.4.' de görüldüğü gibi ayrıştırılması beklenen SPAM sayısına daha yakın sayıda SPAM ayrıştırılan yöntem Naive Bayesian yöntemidir. SPAM olan 22 elektronik posta sayısı, Naive Bayesian yöntemi ile 18, Bulanık Mantık yöntemi ile 12 olarak tahmin edilmiştir.



Şekil 5.5. Naive Bayesian ve Bulanık Yöntemlerinin Doğruluk Oranları

Şekil 5.5.' de Naive Bayesian yönteminin %81.8, Bulanık Mantık yönteminin ise %54.6 oranında bir doğruluk sonucu çıkardığı gösterilmektedir.

## 6. SONUÇLAR VE ÖNERİLER

Tüm dünyada 2010 Haziran itibariyle yapılan araştırmada 1.97 Milyar internet kullanıcısının bulunduğu yayınlanmıştır. Bu oran bir önceki seneye göre % 14 oranında artmış ve bununla birlikte dünyada 1.88 Milyar elektronik posta kullanıcı sayısına ulaşılmıştır. Günde ortalama 294 Milyar elektronik postanın gönderildiği günümüzde son rakamlara göre 2010 yılı içerisinde gönderilen elektronik posta sayısı 107 Trilyonu bulmuştur. Bu kadar önemli rakamlardan bahsedilen internet ortamında, gönderilen elektronik postaların % 89.1 lik kısmının SPAM olması ise durumun ne kadar önemli bir hal aldığı sonucunu doğrular niteliktedir [1].

SPAM, diğer adıyla da istenmeyen postaların maddi kazançlar elde etmede kullanıldığı da düşünüldüğünde, gönderilen SPAM postaların sayısı her geçen gün daha da artacaktır. Kullanıcılara vereceği zararlar da bu oranda artış gösterecektir. Gelen kutusunda sürekli beklenmeyen postaların bulunması hem elektronik posta kullanımını etkileyecek hem de posta güvenliğini giderek tehlike altına atacaktır.

SPAM postaların artış oranlarına bakıldığında, normal olarak gelen postaların ayrıştırılmasının giderek zor olduğu görülmektedir. İnsan gücü ile toplamda gelen 294 Milyar elektronik postanın 261.6 Milyarının ayrıştırılması sayısal olarak da durumun zorluğunu göstermeye yetmiştir. Ancak yapay zekâ uygulamalarının çok sık kullanıldığı günümüzde, doğru yaklaşımlar ile bu durumun çözülebileceği kanıtlanmıştır.

Uygulamadan çıkan sonuçlara göre, Kullanıcıya gelmiş olan ve örnek olarak alınan postanın SPAM olmadığı bilgisi bilinmektedir, bu bilgi ışığında postanın ayrıştırılmasında her iki yöntemin de ne kadar etkili olduğu ayrı ayrı denenmiştir. Sonuçlara göre, Naive Bayesian filtreme metodu % 32 ye, % 18 lik bir tahmin sonucu ile gelen postanın SPAM bir posta olmadığı sonucu çıkarmıştır. Buna göre gelen posta %32 lik ihtimal ile normal bir elektronik postadır ve bu doğru bir sonuçtur. Öte yandan diğer yöntem olan Bulanık Mantık yöntemine göre ise %71 e,

%50 lik bir tahmin sonucu ile gelen postanın SPAM bir posta olmadığı sonucu her iki metodunda istenmeyen postaların ayrıştırılmasında doğru bir yol izlendiği sonucunu vermiştir. Bu duruma göre elektronik postaların ayrıştırılmasında her iki yöntem de kullanılabilir. Çıkan rakamlar değerlendirildiğinde ise Naive Bayesian ile elektronik postaların ayrıştırılmasında doğru sonuca daha da yaklaşıldığı sonucu çıkarılmaktadır. Ancak bu çıkarım kullanıcıya gelen herhangi bir posta üzerinden yapılmıştır. Genel bir karşılaştırma yapmak için, kullanıcıya gelen toplam 100 e-posta üzerinden genel bir sonuç analizi yapılmıştır. Gelen 100 e-postanın içerisinde 22 adet SPAM posta bulunmaktadır. Naive Bayesian yönteminin çıkardığı SPAM sayısı 18 iken, Bulanık Mantık yöntemi ile bulunan SPAM sayısı 12' dir. Buna göre Bulanık Mantık yöntemi elektronik postaların ayrıştırılmasında %54.6 oranında bir doğruluk sonucu verirken, diğer taraftan Naive Bayesian yöntemi ise %81.8 oranında bir doğruluk sonucu vermiştir.

Karşılaştırılan bu yöntemler, e-posta sunucuları için ön ayrıştırma sistemi olarak kullanılabilir. Özellikle ülkemizde ki kurumsal e-postalarda SPAM postaların ayrıştırılması için bu tarz sistemler kullanılmamaktadır. Bu yöntemler e-posta sunucuları için ön ayrıştırma sistemi olarak kullanıldığında herhangi bir manuel müdahaleye ihtiyaç duyulmayacaktır. Bu sayede SPAM postaların yaratmış olduğu olumsuzluklar için bir çözüm oluşturulabilir.

İnsan beyninin özelliklerinden olan öğrenme yolu ile yeni bilgiler türetebilme, yeni bilgiler oluşturabilme, keşfedebilme ve tahmin edebilme gibi yetenekleri herhangi bir yardım almadan geliştirebilen bir sistem olan yapay zekâ yaklaşımları milyonlarca kayıt içerisinden, insan beyninin yetmeyeceği kadar kısa bir sürede sonuca ulaşabilmektedirler. SPAM postaların ayrıştırılmasında kullanılabilen Naive Bayesian ve Bulanık Mantık yöntemleri bunun gibi birçok alanda da farklı durumların tahminlerinde doğru veri analizleriyle kullanılabilirler.

## KAYNAKLAR

1. İnternet: Boğaziçi Üniversitesi “Uyarlamalı Türkçe Spam-Önler Filtrelemesi” <http://www.cmpe.boun.edu.tr/~gungort/papers/Uyarlamali%20Turkce%20Spam-Onler%20Filtrelemesi.doc> (2009).
2. İnternet: Internet World Stats “Internet 2010 İn Numbers” <http://royal.pingdom.com/2011/01/12/internet-2010-in-numbers> (2010).
3. İnternet: Vikipedi Özgür Ansiklopedi “İstenmeyen Posta ya da Spam” <http://tr.wikipedia.org/wiki/Spam> (2011).
4. İnternet: Anti Spam “Spam Nedir?” <http://www.spam.org.tr> (2009).
5. İnternet: Electronic Privacy Information Center “SPAM - Unsolicited Commercial E-Mail” [http://epic.org/privacy/junk\\_mail/spam](http://epic.org/privacy/junk_mail/spam) (2010).
6. Yıldız, T., Yıldırım S. ve Altılar T., “İstenmeyen İletilerin Paralelleştirilmiş KNN Algoritması ile Tespiti”, *Akademik Bilişim 2008*, Çanakkale, (2008).
7. Özgür, L., “Türkçe morfolojik çözümleme, yapay sinir ağları ve bayesian filtreleme tabanlı uyarlamalı spam-önler filtrelemesi”, Yüksek Lisans Tezi, *İstanbul Boğaziçi Üniversitesi Bilgisayar Mühendisliği Bölümü*, İstanbul, 5-10 (2003).
8. Hudairy, H., “Data mining and decision making support in the governmental sector”, Master Thesis, *Louisville University*, Kentucky, 1-5 (2004).
9. Stephen, M. S., “Thomas bayes' bayesian inference”, *Journal of the Royal Statistical Society*, Series A, 250–258 (1982).
10. Yılmaz, R., “Jeffreys' in önseli kullanılarak genelleştirilmiş lineer modellerde bayes analizi ve bir uygulama”, Yüksek Lisans Tezi, *Ondokuz Mayıs Üniversitesi*, Samsun, 61, (2002).
11. Graham, P., “Better bayesian filtering”, *In Proceedings of the Spam Conference*, 5-10, (2003).

12. Duda, R. O., Hart, P. E., “Stark. Pattern Classification (2nd ed.)”, *John Wiley & Sons*, 20-25, (2001).
13. Domingos P., Pazzani M., “On the Optimality of the Simple Bayesian Classifier under Zero-One Loss”, *Machine Learning*, 29, 103-130, (1997).
14. Hudairy, H., “Data mining and decision making support in the governmental sector”, Master Thesis, *Louisville University*, Kentucky, 10-15 (2004).
15. Gürsakal N., “Bayesci İstatistik”, *Uludag Üniversitesi Basımevi*, Bursa, 68, (1992).
16. Zdziarski, Jonathan, A., “Ending Spam: Bayesian Content Filtering and the Art of Statistical Language Classification”, *No Starch Pres Inc.*, 10, (2005).
17. Koyuncu, E., “Yeni matematiksel kod; bulanık mantık”, *Teknolojist, ITU IEEE*, (2004).
18. Tanaka, K.: “An Introduction to Fuzzy Logic for Practical Applications”, *Rassel Inc.*, Japan, (1997).
19. Şen, Z., “Bulanık Mantık ve Modelleme İlkeleri”, *Bilge Kültür Sanat*, İstanbul, (2001).



## ÖZGEÇMİŞ

### Kişisel Bilgiler

Soyadı, adı : YUMAK, Burhan  
Uyruğu : T.C.  
Doğum tarihi ve yeri : 17.09.1984 Giresun  
Medeni hali : Evli  
E-mail : burhanyumak@gmail.com.

### Eğitim

Derece	Eğitim Birimi	Mezuniyet tarihi
Yüksek lisans	Gazi Üniversitesi /Bilgisayar Eğitimi	2011
Lisans	Gazi Üniversitesi/ Bilgisayar Öğretmenliği	2007
Lise	Giresun Anadolu Meslek Lisesi	2001

### İş Deneyimi

Yıl	Yer	Görev
2008-.....	MyBilet Bilgi İletişim Sistemleri	Yazılım Uzmanı
2007-2008	Türk Kızılayı Genel Merkez	Yazılım Uzmanı

### Yabancı Dil

İngilizce

### Hobiler

Spor, Bilgisayar teknolojileri, Yüzme