



AISECLAB

Artfical Intelligence Security & Defense Lab

Data Preprocessing ve Machine Learning Oluřturma

Hazırlayan:Emre BUYRUK

Makine Öğrenmesi Nedir?

Tanımlama

Makine öğrenmesi, algoritmaların, verileri kullanarak görevleri kendilerinin belirlediği koşullar altında gerçekleştirebilmesini sağlar.

Kullanım Alanları

Makine öğrenmesi uygulamaları arasında oyuncaklar, akıllı cihazlar, resim tanıma sistemleri, spam filtreleri ve hatta daha karmaşık sistemler yer alır.

Temel İlkeleri

Karar vericiler, eğitim ve sınıflandırma gibi konular makine öğrenmesi temel ilkeleri arasındadır.

Makine Öğrenmesi İçin Görsel Veri

Neden Görsel Veri Kullanırız?

Görsel veri, makine öğrenmesinde veri sonuçlarını optimize etmek ve anlamak için öncelikli olarak kullanılır.

Görsel Veri Kullanma Örnekleri

Yüz tanıma, nesne tanıma, resim işleme, video tanıma ve doğal dil işleme gibi birçok alanda görsel verinin kullanımı yaygınlaşmaktadır.

Görsel Veri Toplama Yöntemleri

Görsel veri elde etmek için, web taraması, veri örnekleme ve veri izleme gibi teknikleri kullanabiliriz.

Model Eğitimi

1

Algoritma Seçimi

Doğru algoritmanın seçimi, modelin başarısını etkiler.

Bu nedenle, uygun bir algoritma seçilmelidir.

2

Model Eğitimi

Eğitim verileri üzerinde model oluşturulmalıdır. Veriler, belirli bir oranda eğitim ve test setlerine bölünerek kullanılmalıdır.

3

Hiperparametrelerin Ayarlanması

Hiperparametreler, farklı sonuçlar ve performanslar verir. Doğru hiperparametrelerin kullanımı, modelin doğruluğunu ve başarısını artırır.

Model Testi

Doğruluk Oranı

Modelin doğruluğu, ölçülebilir bir metrik olmalıdır. Doğru sonuçların oranı, modelin başarısını gösterir.

Kesinlik ve Hassasiyet

Doğru sonuç, olumlu olarak etiketlenen tüm sonuçlara karşı olumlu olarak etiketlenen gerçek sonuçların oranıdır. Hassasiyet ise olumlu olarak etiketlenen tüm sonuçların ne kadarının gerçek olduğunu temsil eder.

F1 Skoru

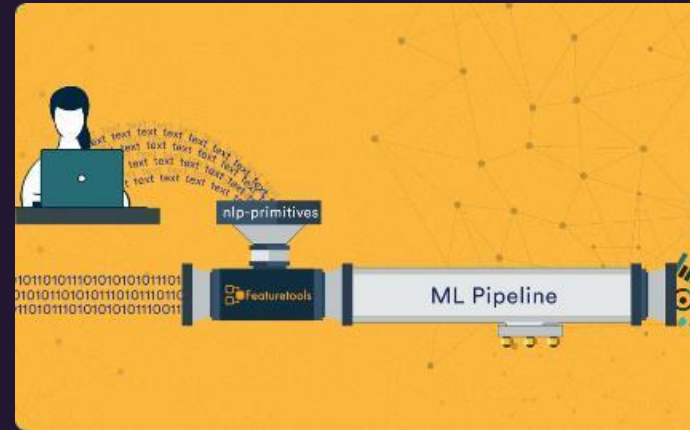
F1 skoru, ölçümler arasında bir denge kurmak için kullanılır ve kesinlik ile hassasiyet arasındaki dengeyi çözer.

Veri Toplama ve Formülasyonu



Veri Toplama

Veri toplama, doğru sonuçlar elde etmek için temel bir unsurdur. İlgili veriler, URL yapısı, web sayfası içerikleri, HTTP trafikleri vb. özellikleri de içermelidir.



Özellik Mühendisliği

Doğru özelliklerin seçimi,
makine öğrenmesi modelleri
için son derece önemlidir.
Özellik mühendisliği, veri
toplama işlemi sonrasında,
verilerin modelleme işleminde
kullanılabilecek hale
getirilmesi işlemidir.



Veri Temizleme

Verilerin doğru, tam ve tutarlı olması gerekir. Veri temizleme, verilerin sağlıklı ve homojen olmasını sağlamak için kullanılan bir yöntemdir.

Bağımlı Bağımsız Değişken Nedir?

Bir bağımsız değişken, bir bağımlı değişkeni etkileyebilen veya ona neden olan bir değişkendir. Bağımlı değişken, bağımsız değişkenlere bağlı olarak değişir. Bu kavramlar, makine öğrenmesi ve istatistiksel veri analizi için temel öneme sahiptir.

Independent vs. Dependent Variable

Sort

INDEPENDENT

determines

DEPENDENT

Years spent smoking

Likelihood of getting cancer

Temperature

Ice cream cones sold

Age of car (in years)

Value of car (in \$)

The number of apples picked

The number of pies that can be made

Time of the year

Minutes of daily sunlight

Number of miles ran

Calories burnt

Tickets presold to a play

Money for decorations

Hours spent studying

Grade earned on test

depends on

Sonuç ve Özet

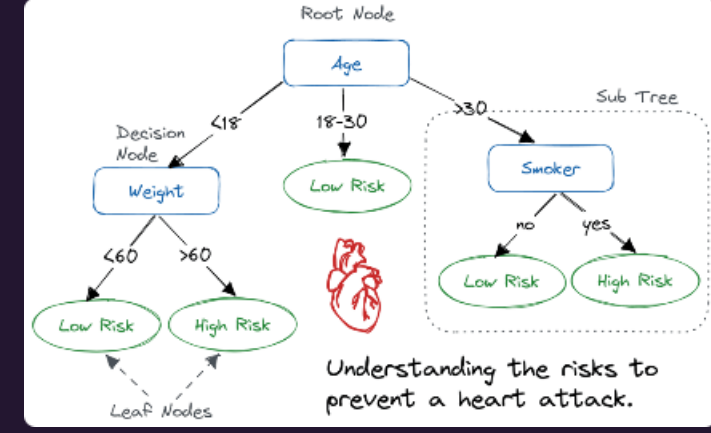
Bağımlı ve bağımsız değişkenler, veri analizi ve makine öğrenmesinde temel öneme sahiptir.

Bu değişkenler ve algoritmalar, lineer regresyon ve karar ağacı gibi yöntemler, veri analizi ve görsel veri elde etmek için URL kullanımının ve son olarak veri analizi sürecinde çıkabilecek problemlere karar verme konusunda yardımcı olabilir.

Veri analizi, büyük veri kümelerinde farklı desenler ve trendleri anlamamızı kolaylaştırır.

Veri analizinin önemi, veri görselleştirme, veri örnekleme, doğru/bütünsel verinin toplanması, analiz edilmesi ve karar vermeye bağlıdır.

Karar Ağaçları



Ne Anlama Geliyor?

Karar ağacı, bir hedefe ulaşmak için kararların hiyerarşik olarak yapıldığı bir ağaç yapısındaki bir algoritmadır.

Nasıl Çalışır?

Kararlar sırayla her bir düğümde alınır, ağaç dalları ise veri kümesinin değişkenleri tarafından temsil edilir.

Örnek Kullanım Alanları

Karar ağacı algoritmaları, veri madenciliği, makine öğrenmesi ve yapay zeka gibi birçok alanda kullanılır.

Lineer Regresyon

Ne İçin Kullanılır?

Çoklu lineer regresyon analizi, fenomenleri göstermek, tahmin yapmak ve neden-sonuç ilişkilerini anlamak için kullanılır.

1

Ne Anlama Geliyor?

Lineer regresyon, bağımlı bir değişkenin, bir veya daha fazla bağımsız değişken tarafından açıklanabildiği bir istatistiksel modellemedir.

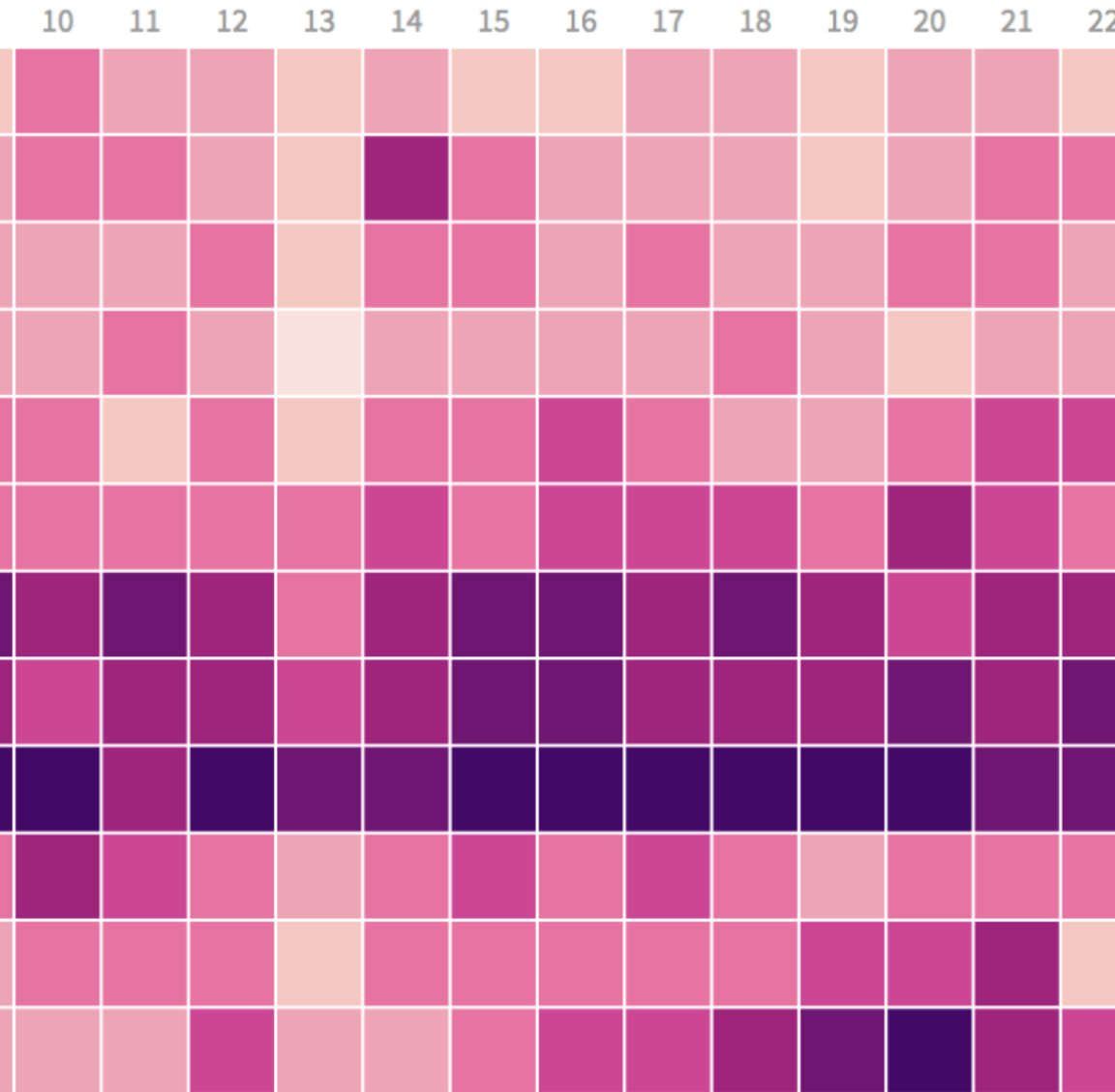
2

3

Örnek Kullanım Alanları

Ekonomi, finans, makine öğrenmesi, meteoroloji ve sağlık alanlarında lineer regresyon yaygın bir şekilde kullanılmaktadır.

nth and day.



Makina Öğrenmesinde Isı Haritası Nedir ve Ne İşe Yarar?

Makina öğrenmesinde verilerin analizi ve anlaşılması oldukça önemlidir. Bu nedenle ısı haritaları, verilerin görselleştirilmesinde büyük bir rol oynar. Bu sunumda, ısı haritalarının ne olduğu, ne işe yaradığı, nasıl elde edildiği, hangi verileri gösterdiği, uygulamaları ve problemleri çözmek için kullanılan yöntemleri ele alacağız.

Isı Haritası Nedir?



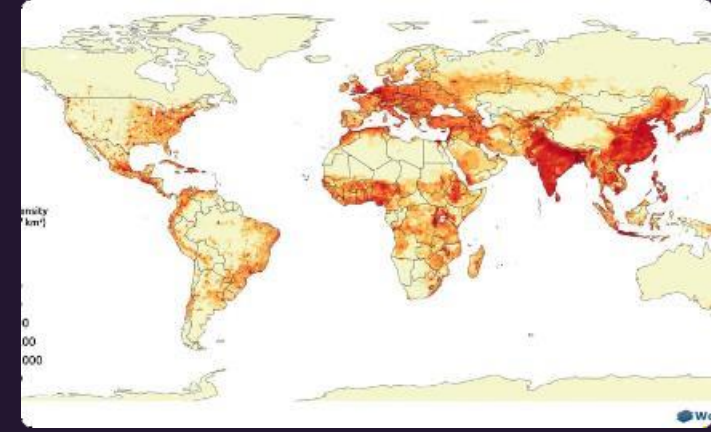
Bulut Haritası

Bulut haritaları, meteorolojik verileri görselleştirmek için kullanılır.



Beyin Haritası

Beyin aktivitesini görselleştirmek için kullanılan bir yöntemdir.



Nüfus Yoğunluk Haritası

Nüfus yoğunluğunu dünya haritası üzerinde gösteren bir yöntemdir.

Isı Haritası Ne İşe Yarar?

Düzenlilikleri Gösterir

Verilerde gözle görülür bir şekilde var olan düzenlilikleri, dağılımları ve kümeleri ortaya koyar.

Analizi Kolaylaştırır

Verileri görselleştirmek, analizi daha kolay hale getirir ve daha hızlı sonuca ulaşmanızı sağlar.

Farklı Boyutları Gösterir

Bir veri kümesinin farklı boyutlarını (zaman, konum, sektör, vs.) görselleştirerek daha kapsamlı bir analiz yapabilirsiniz.

Isı Haritası Nasıl Elde Edilir?

1

Veri Toplanması

Öncelikle, var olan verileriniz veya yeni verileriniz toplanır.

2

Verilerin Analizi

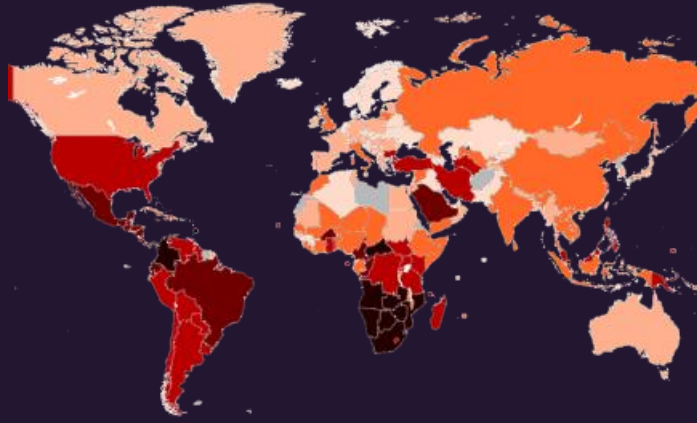
Veriler analiz edilerek, verilerdeki düzenlilikler ortaya çıkarılır.

3

Verilerin Görselleştirilmesi

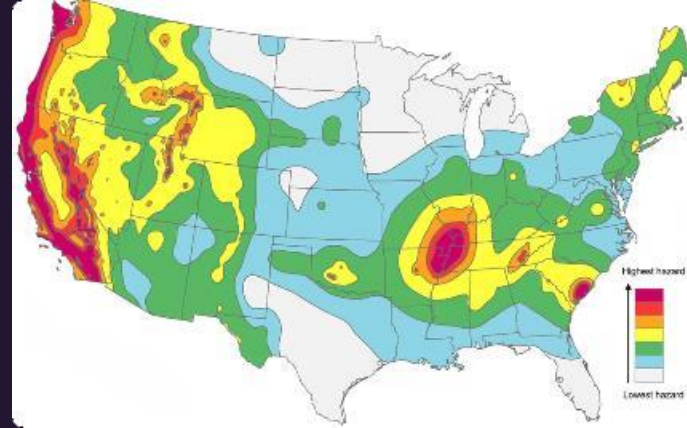
Veriler, istenilen şekilde görselleştirilir. Bu sırada önemli olan, en uygun renk skalasını seçmek ve verileri anlamlı bir şekilde göstermek.

Isı Haritası Hangi Verileri Gösterir?



Gelir Dağılımı Haritası

Gelir dağılımını
görselleştirmek için kullanılır.



Deprem Haritası

Depremleri görselleştirmek
için kullanılır.



Trafik Haritası

Trafik akışını ve yoğunluğunu
görselleştirmek için kullanılır.

Isı Haritası Uygulamaları

Kullanıcı Davranışlarının Analizi

Kullanıcılardan gelen verilerin analizi için kullanılır.

Şirket Performans Analizi

Şirket içi verilerin analizi için kullanılır.

Coğrafik Haritalama

Bölgelerin coğrafi özelliklerine göre analiz yapmak için kullanılır.

Sağlık Verilerinin Analizi

Sağlık verilerinin analizi için kullanılır.

Isı Haritası ile Problemleri Çözmek İçin Kullanılan Yöntemler



Sonuç ve Değerlendirme

Veri analizi sürecinde, verilerin görselleştirilmesi oldukça önemlidir. ısı haritaları, verilerin görselleştirilmesinde kullanışlı ve etkili bir yöntemdir. Tulpar Analytica olarak, ısı haritalarını verilerin analizi ve anlaşılması için sıklıkla kullanıyoruz ve bu yöntemin sonuçlarına her zaman çok değer veriyoruz.



Makina Öğrenmesinde Label Encode Nedir ve Neden Yapılır?

Makina öğrenmesi algoritmalarının etkili çalışması için verileri sayısal hale getirmek gerekiyor. Label Encode bu işlemi yapmamıza yardımcı olur.



Label Encode Tanımı

Basit bir tanımla:

Label Encode, makina öğrenmesinde kullanılabilecek şekilde, kalitatif verileri sayısal hale getiren bir işlemdir.

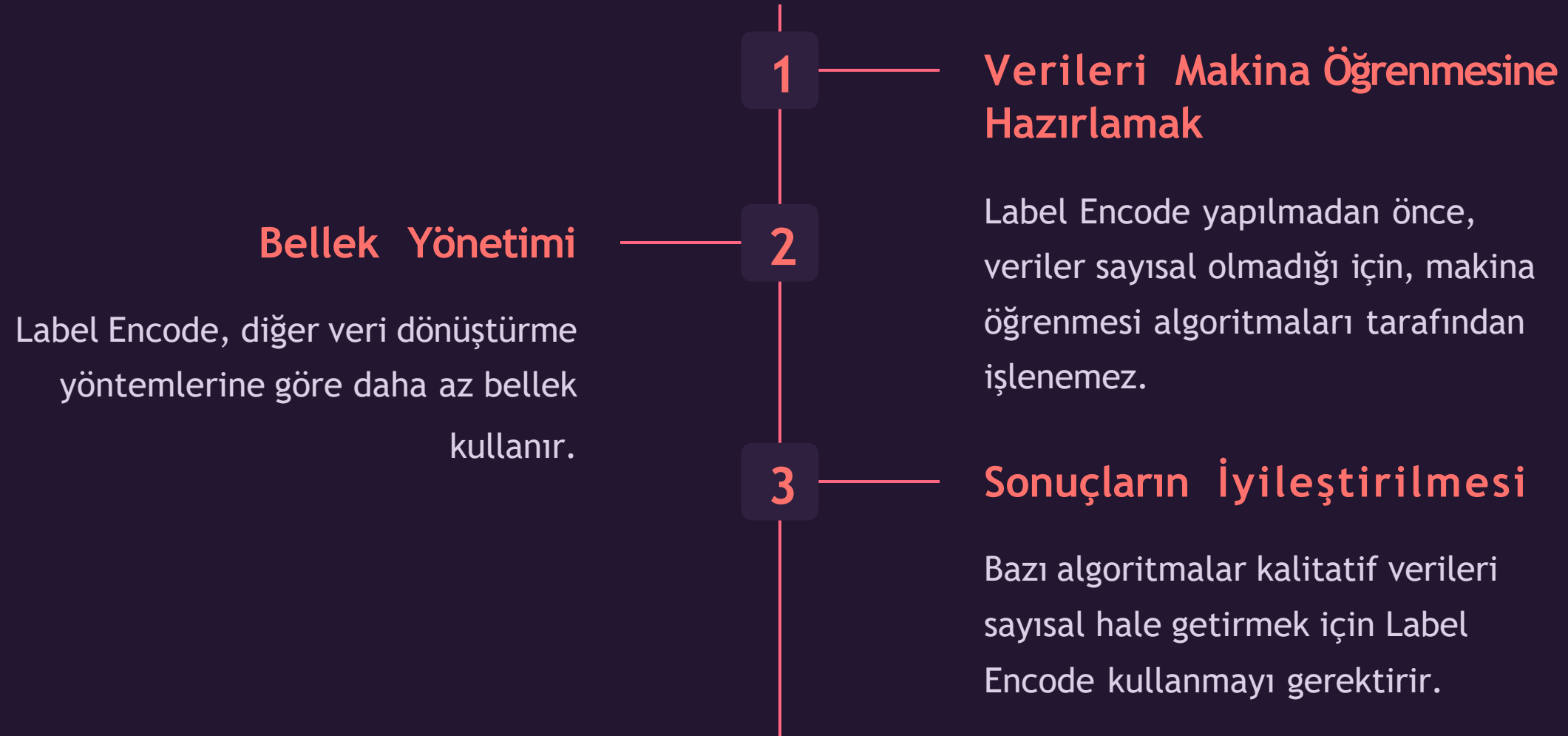
Bir örnek:

Verilerimiz arasından "Kırmızı", "Mavi", ve "Yeşil" gibi renk terimleri varsa, Label Encode bu terimleri sırasıyla 1, 2, ve 3 sayılarına tahsis eder.

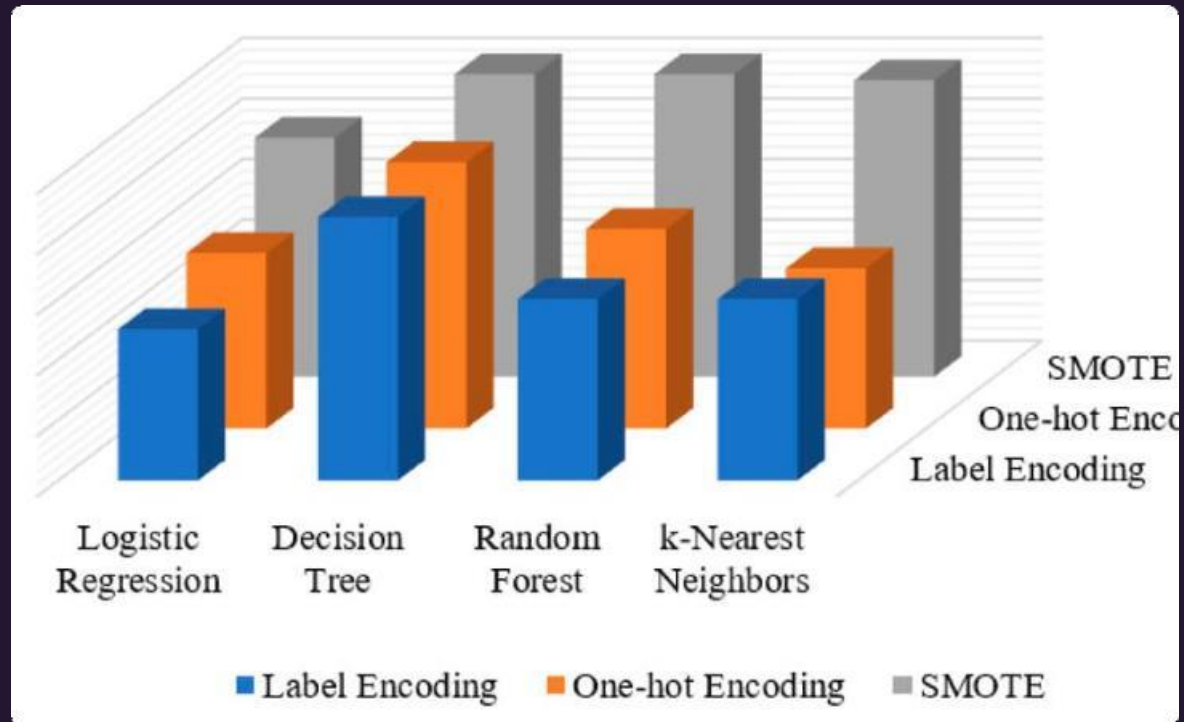
Alternatifleri:

Benzer bir işlem olan One-Hot Encoding, Label Encode'a alternatif bir çözümdür.

Neden Label Encode Yapılır?



Label Encode Örneği



Örnek Bir

Label Encode Nasıl Yapılır?

Scikit-Learn Kullanarak:

Python dilinde Scikit-Learn kütüphanesi, verileri Label Encode etmek için kullanılabilir.

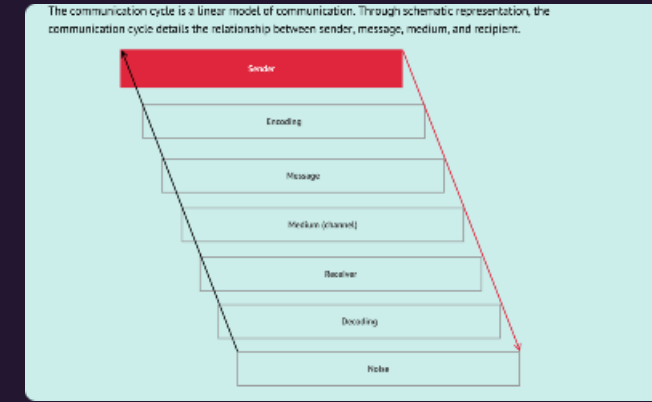
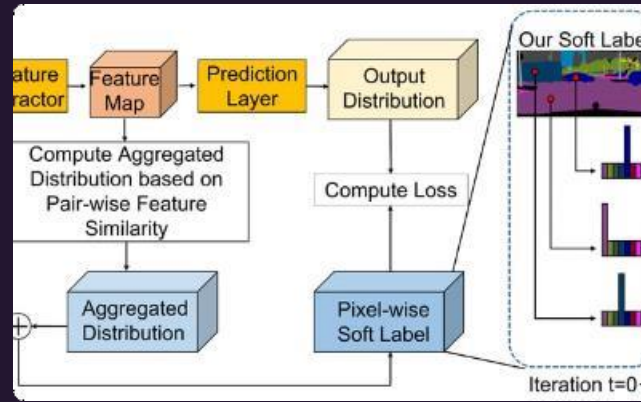
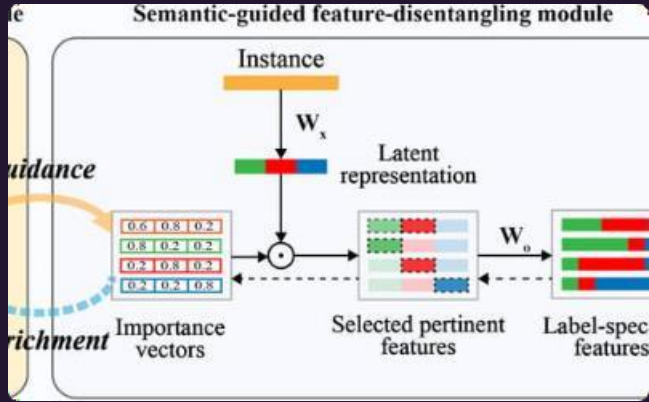
Manuel Yöntem:

Manuel olarak, Pandas kütüphanesi kullanılarak, LabelEncode fonksiyonuyla kolayca yapılabilir.

Label Encode'ın Sonuçları ve Etkileri



Label Encode'ın Ciddi Durumda Yapılması Gerektiği Durumlar



Önemli Faktörler:

- Yapılan sektöre göre, kullanılan algoritmalarla göre, Label Encoding önemli olabilir.
- Bazı verilerin doğru şekilde kalitatif veri olarak tanımlanması gerektiği durumlar vardır.

Penalties:

Bazı makine öğrenmesi projeleri, ciddi hatalar ve yanlış kaynaklardan dolayı Label Encoding çözümleri kullanmakta ve bu durum projenin başarısını etkilemektedir.

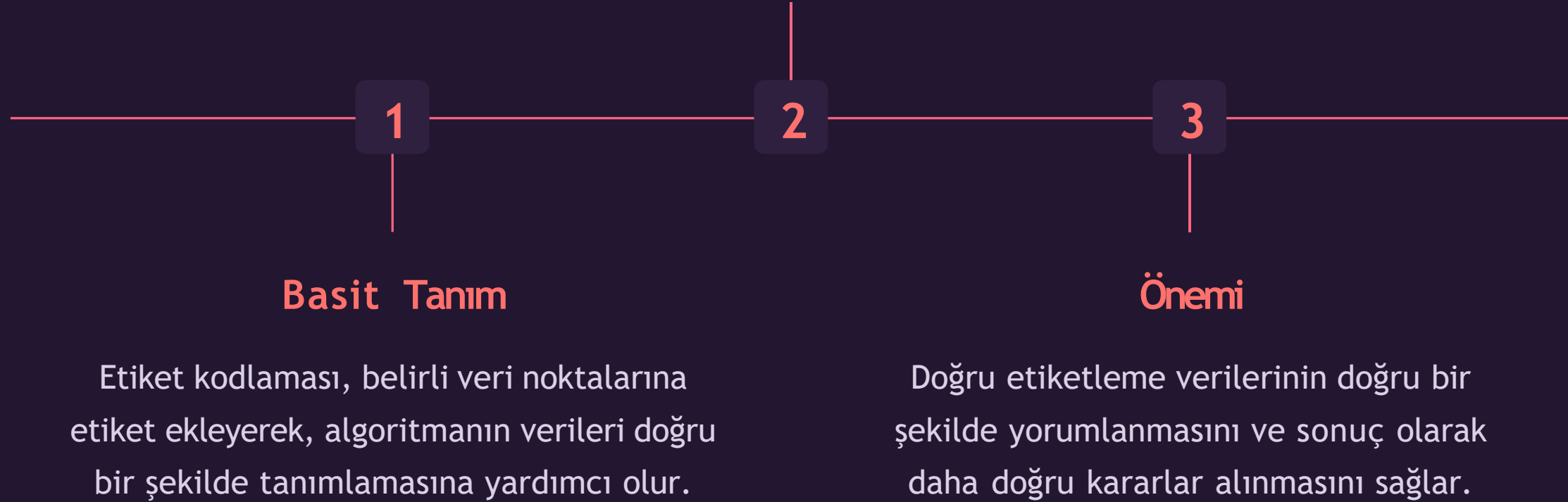
Verinin Önemi:

Doğru veri toplama, verilerin doğru ve kapsamlı olarak tasnif edilmesi önemlidir, bu nedenle Label Encoding işlemi, doğru şekilde uygulanmazsa, verilerin gürültüye eklenmesine neden olabilir.

Etiket Kodlama Nedir?

Farklı Veri Türleri

Metin, ses, görüntü ve veri tabanı gibi çeşitli veri türlerini etiketleyebilirsiniz.



Etiket Kodlama Yöntemleri

Tanımlayıcı Etiketleme

Etiketler öğeleri açıklar ve etiketleme işlemi daha belirgin hale getirir.

Sezgisel Etiketleme

İnsan beyninin kalıpları kullanarak, verileri tanımlaması şeklinde yapılır.

Genel Etiketleme

Bir veri setinin tüm noktalarının etiketlenmesini gerektirmez.



Makine Öğrenmesi Etiket Kodlama Mantığı

Herhangi bir nesneyi ya da olayı tanımlamanın en etkili yollarından biri, onu etiketlemektir. Bu sunumda, makine öğrenmesi ile nasıl etiket kodlaması yapabileceğinizi öğreneceksiniz.

Makine öğrenmesi, bir bilgisayarın veri örneklerindeki kalıpları tanımasını ve benzer şekilde etiketlenmemiş verileri sınıflandırmasını sağlayan bir yapay zeka dalıdır. Etiket kodlaması, makine öğrenmesi modellerinin eğitilmesinde kullanılan bir yöntemdir.

Bu sunumda, etiket kodlaması nedir, nasıl yapılır, farklı yöntemleri nelerdir, zorlukları nelerdir ve etiket kodlama ile ilgili örnekler nelerdir gibi konuları ele alacağız. Sunumun sonunda, makine öğrenmesi etiket kodlama mantığı hakkında daha iyi bir anlayışa sahip olacaksınız.

Etiket Kodlama Nasıl Yapılır?

```
25
26 def check_db():
27     if not os.path.isfile(FILE_URI):
28         db.create_all()
29
30 @app.route("/")
31 def home():
32     check_db()
33     all_books = db.session.query(Book).all()
34     return render_template("index.html", books=all_books)
35
36 @app.route("/edit", methods=['GET', 'POST'])
37 def edit():
38
39     if request.method == 'POST':
40         book_id = request.form["id"]
41         book_to_update = Book.query.get(book_id)
42         book_to_update.rating = request.form["rating"]
43         db.session.commit()
44         return redirect(url_for("home"))
```

Manuel Kodlama

Veri setindeki her noktanın elle etiketlenmesi şeklinde yapılır.



Otomatik Kodlama

Makine öğrenmesi teknikleri kullanarak, algoritma doğrudan verileri etiketleyebilir.



Yarı Otomatik Kodlama

İnsanlar ve algoritmaların birlikte çalışması ile gerçekleştirilir.

Etiket Kodlamada Karşılaşılan Zorluklar

- Kapsamlı veri setleri
- Sınıflandırma ihtiyacı
- Manuel işlem olduğu zaman yavaş kalma
- Etiketleme önyargısı



Etiket Kodlama ile İlgili Örnekler

Yüz Tanıma

Yüz tanıma sistemleri etiket kodlaması yolunu kullanır.

Otomatik Dil Çevirisi

Çevirinin doğru sonuç üretmesi için etiket kodlama sistemi uygulanır.

Sonuç ve Özet

	Avantajları	Dezavantajları
Tanımlama	Doğru tanımlama, verilerin doğru yorumlanmasını sağlar.	Zaman ve işgücü gerektirir.
Sınıflandırma	Veri setini daha verimli hale getirir.	Manuel iş olabilir.
Orientasyon	Doğru veri seti orientasyonu daha iyi sonuçlar üretir.	Doğru sizin istediğiniz oryantasyon olmayabilir.

Makine Öğrenmesinde URL'ler

1 URL Nedir?

URL, internet sayfalarının adresleridir ve verilerin nasıl görüntülendiğini belirlemek için kullanılır.

2 Farklı URL Formatları

Kaynak URL'ler, özel URL'ler, alt URL'ler ve URL parametreleri gibi farklı türdeki URL'ler, makine öğrenmesinde farklı amaçlar için kullanılabilir.

3 URL'lerin Önemi

URL'ler, doğru ve tutarlı veri toplama için kritik öneme sahiptir ve veri analizi için gerekli doğru sonuçlara ulaşılmasını kolaylaştırır.

Zararlı URL

İnternetin karanlık köşelerinde

Kötü niyetli bağlantılar tehdit ediyor

Ancak, bizim güçlü siber güvenlik teknolojimiz sayesinde

Bu tehditleri tespit edebiliyoruz ve engelliyoruz!

Ve en iyi şey?

Bunları yaparken, makine öğrenmesi ile daha da güçleniyoruz!

b

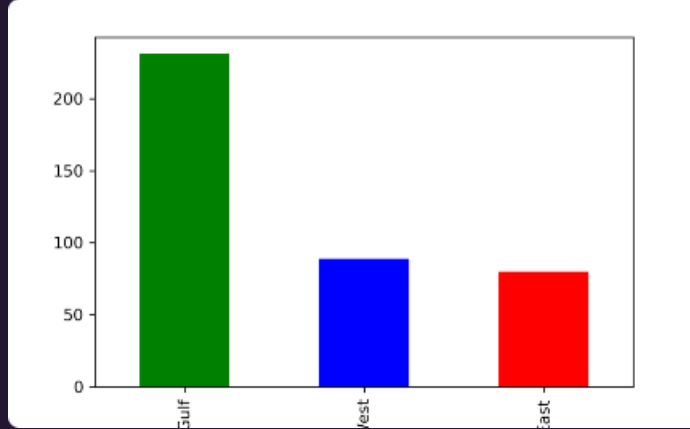


Sonuçlar ve Performans Ölçütleri

Makine öğrenmesi kullanarak zararlı URL'leri tespit etmek, geleneksel yöntemlere göre daha yüksek bir doğruluk oranı sağlar. Kesinlik ve hassasiyet ölçümleri, modelin başarılı bir şekilde performans gösterdiğini göstermektedir.

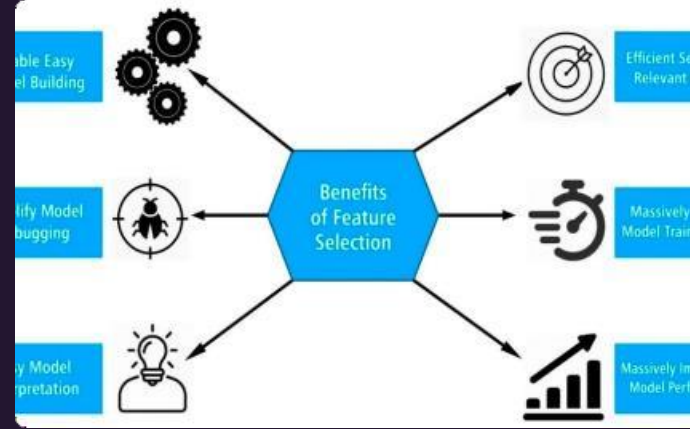
results

Karşılaşılan Zorluklar ve Çözümleri



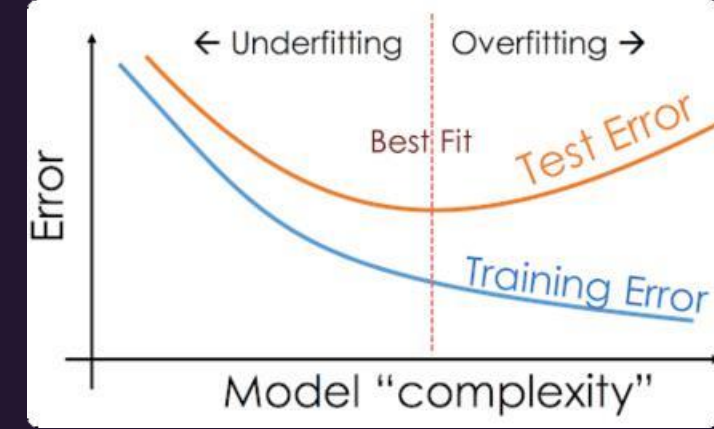
Sınıf Dengesizliği

Verilerin çoğunluğu zararsız URL'lerden oluştuğu için sınıf dengesizliği oluşur. Bu sorun, örnekleme teknikleri veya sınıf ağırlıklarının kullanımı ile çözülebilir.



Özellik Seçimi

Özelliklerin doğru seçimi, modelin sonuçları üzerinde büyük bir etkiye sahiptir. Özellik seçimi, hedef değişkenle korelasyonu olan özneliliklerin seçilmesi ile yapılabilir.



Aşırı Uyum

Aşırı uyum, eğitim verilerine çok fazla uyum sağlaması ve test verileri için iyi bir sonuç vermeyi başaramamasıdır. Bu sorun, düzenleme yöntemleri kullanılarak çözülebilir.

Gelecekteki Araştırma Yönleri

1 Veri Boyutları

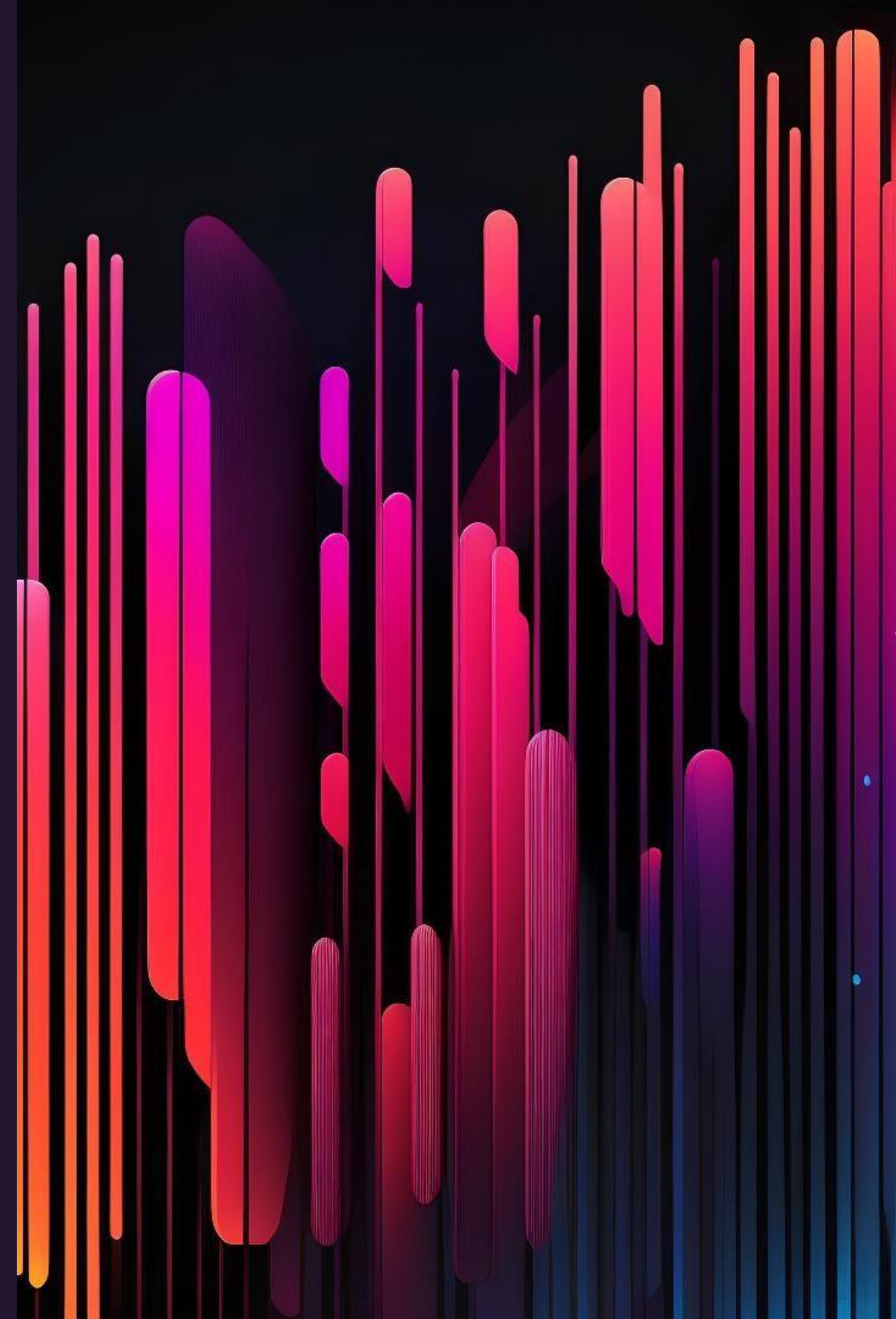
Zararlı URL'leri tespit etmek için daha fazla veri kullanılarak elde edilen sonuçların değerlendirilmesi gerekmektedir.

2 Performans İyileştirmeleri

Daha iyi sonuçlar elde edebilmek için, daha gelişmiş algoritmaların ve yenilikçi özellik seçim tekniklerinin kullanımı ile performansın artırılması gerekmektedir.

3 Çok Dilli Yöntemler

Zararlı URL'lerin tespiti, çoklu diller arasında farklılık gösterir. Bu nedenle, çok dilli yöntemler ile zararlı URL'lerin tespiti sağlanabilir.

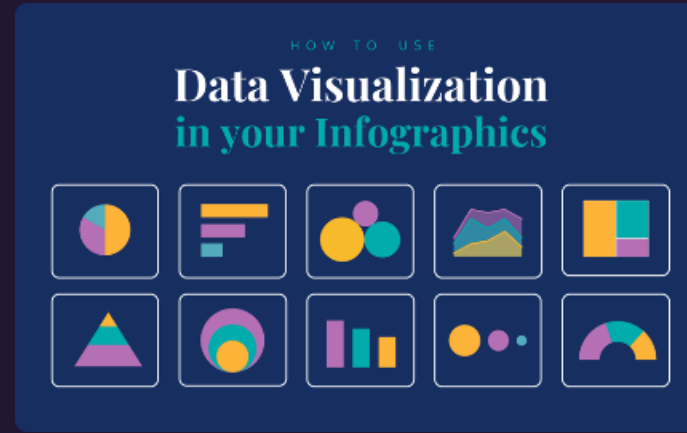


Karar Vermeyi Kolaylařtırmak için Veri Analizi



Veri Analizi

Veri analizi, büyük miktarda veri oluřturma, temizleme, organizasyon ve yorumlama sürecidir.



Veri Görselleřtirme

Veri Görselleřtirme, bir veri kümesini grafik veya resimlerle sunarak, verilerin daha anlaşılır hale gelmesini sağlar.



Karar Verme

Veri analizi, karar verme sürecindeki unsurları doğru anlamamızı sağlar, böylece doğru kararlar alabilir ve potansiyel riskleri en aza indirebiliriz.