



YAZAR BİLGİLERİ

ADI: MÜRVET

SOYADI: ARSLAN

OKULU: ERZURUM TEKNİK ÜNİVERSİTESİ

BÖLÜMÜ: BİLGİSAYAR MÜHENDİSLİĞİ

OKUL NUMARASI: 190303034

GİTHUB HESABI: <https://github.com/mrvetarslan>

RAPOR BİLGİLERİ

KONUSU: DOĞAL DİL İŞLEME TEKNOLOJİSİ

BAŞLIĞI: DOĞAL DİL İŞLEME İLE GEÇMİŞTEN GELECEĞE

TARİHİ: 20.12.2023

ÖĞRETİM ÜYESİ: CEM İBRAHİM ARI

Doğal Dil İşlemenin Temel Prensipleri

Doğal Dil İşleme Nedir?

Wikipedia ya göre NLP(Natural Language Processing)'nin Türkçe karşılığı Yapay zeka ve Dil biliminin alt kategorisi ve Dünyada var olan bir çok doğal dilleri kullanarak araştırmalar yapılan, bu araştırmalarla insanlık için fayda sağlanması çalışan bir bilim dalıdır. Dünyada sadece insanlar değil diğer canlılar da birbirleriyle iletişim halindedir. İki tarafında birbirlerini anlaması, iletişim kurabilmesi ve sorunları bu denli iletişim halinde olarak çözmesi çok önemlidir. Teknoloji ilerledikçe artık sadece insanlar ve ya canlılar birbirlerini anlamakla kalmıyor, cansız varlıklar olan makine türevleri olan cansız türler de artık biz canlılarla iletişim halinde olabiliyor. Tam bu noktada Natural Language Processing cümlesinin baş harflerinden oluşan NLP yani "Doğal Dil İşleme" bilimi ortaya çıkıyor.



NLP, Fritz Perls, Virginia Satir, ve Milton Erickson gibi öncülerin çalışmalarının bir sentezi olarak geliştirilmiştir. Richard Bandler ve John Grinder, bu büyük düşünürlerin başarılarını modelleyerek insan davranışını anlama ve iyileştirme yöntemlerini oluşturmuşlardır. NLP'nin temelleri, antropolog Gregory Bateson'ın görüşleri üzerine inşa edilmiştir. 1970'lerde John Grinder'in dilbilim, Richard Bandler'in ise matematik, bilgi teknolojisi ve psikoloji alanındaki uzmanlıklar ile birleşerek geliştirilmiştir. İlk etapta, Dr. Milton Erickson'un hipnoterapi becerileri ve Virginia Satir'in ilişki terapisi konusundaki başarıları, NLP'nin temelini oluşturan önemli bileşenler olmuştur. Richard Bandler, Virginia Satir ve Milton H. Erickson'un çalışmalarından ilham alarak NLP sürecini oluşturmuş ve fobi tedavisi üzerine odaklanmıştır. Bu başarıdan sonra, NLP'nin uygulama alanını genişleterek, birçok farklı alanda etkili bir yöntem olarak benimsenmesine katkıda bulunmuştur. 1975 ile 1977 arasında yayımlanan "The Structure of Magic" ve "Patterns" adlı kitaplar, bu dönemde elde ettikleri başarıları detaylandırılmıştır. NLP, zaman içinde geniş bir takipçi kitlesi bulmuş ve günümüzde

birçok alanda popülerliğini korumaktadır.

Doğal Dil İşleme, özünde bilgisayarların insan dilini anlamak ve kullanmak amacıyla geliştirilmiş bir teknoloji dalıdır. Yani, bilgisayarlar aracılığıyla metinleri, konuşmaları ve dilin diğer formlarını anlamak ve işlemek söz konusudur. Ancak, NLP'nin çekici yönü, sadece kelimeleri tanımakla sınırlı kalmayıp, dilin derinliklerine inerek insanlar arasındaki anlam karmaşıklığını da anlamasıdır. NLP, sadece metinleri kelime ve cümle olarak algılamakla kalmaz, aynı zamanda içerdikleri anlam ve bağlamı anlama yeteneğine sahiptir. Ayrıca, konuşma seslerini tanıma becerisi sayesinde, konuşma dilini de başarılı bir şekilde işleyebilir. Bu yetenekleri sayesinde duygusal ifadeleri anlama kapasitesine de sahiptir, böylece metinlerin ve konuşmaların ardından duygusal tonları ve anamları çözümleyebilir.



Temel Kavramlar Ve İşlemler Nelerdir?

Her teknolojide olduğu gibi NLP'nin de temel ilkeleri mevcuttur. Bu ilkeler doğal dilin karmaşıklığını anlamaya odaklanırken alınan metinleri kelime seviyesinden ziyade anlam derinliği, cümle yapıları ve dillerin semantik zenginliği bağlamında da anlamayı amaçlamaktadır. Bu ilkeler,

1) Dil Modellemesi: Dil Modellemesi, Doğal Dil İşleme'nin (NLP) önemli bir alanını oluşturan bir kavramdır. Bu modellemeler, metinlerin yapısını ve anlamını anlama yeteneği ile bilgisayarların dil üzerinde daha karmaşık görevleri yerine getirmesine olanak tanır.

- Kelimeler Arasındaki İlişkileri Anlama: Dil modelleri, bir kelimenin diğer kelimelerle olan ilişkilerini anlamak için eğitilebilir. Bu, kelimenin bağlam içindeki anlamını daha iyi kavramlarına olanak tanır.
- Ardışık Kelimelerin Tahmin Edilmesi: Dil modellemesi, ardışık kelimelerin birbirine nasıl bağlandığını ve cümleleri nasıl oluşturduğunu tahmin etme

yeteneğini içerir. Bu, bir metindeki cümle yapısını anlama ve benzer cümlelerin oluşturulması sürecinde kullanılabilir.

- Probabilistik Tahminler: Dil modelleri, kelimelerin ardışık olarak nasıl geleceğini tahmin etme işlemini genellikle bir olasılık dağılımı üzerinden gerçekleştirir. Yani, bir kelimenin bir diğerinden sonra gelme olasılığını belirler.
- Eğitim ve Öğrenme: Bu modeller, büyük miktarda metin verisi üzerinde eğitilir. Eğitim süreci, modelin dil yapısını ve ilişkileri öğrenmesine olanak tanır. Öğrenilen bu bilgiler, modelin daha sonra yeni metinleri anlamasına ve üretmesine yardımcı olur.
- Anlam Karşıklığını Çözme: Dil modellemesi, kelimenin bağlamındaki anlam karışıklığını çözebilir. Bu, bir kelimenin birden fazla anlam taşıması veya bir cümlenin farklı yorumlamalara açık olması durumlarında önemlidir.

2) Anlamsal Analiz: Anlamsal Analiz, Doğal Dil İşleme'nin (NLP) önemli bir yönüdür ve metinlerdeki anlamı çözümlemek için kullanılır. Bu analiz, kelimenin veya ifadenin içeriği anlamı ve bağlamı anlamayı hedefler.

- Anlam Çıkarma: Anlamsal Analiz, metinlerde geçen kelimelerin veya ifadelerin anlamlarını çıkarmayı amaçlar. Bu, kelimenin tek başına taşıdığı anlamın yanı sıra, bağlam içindeki anlamını da anlamayı içerir.
- Bağlam Analizi: Anlamsal Analiz, bir kelimenin veya ifadenin geçtiği cümle veya metindeki bağlamı inceleyerek anlamını belirlemeye çalışır. Bir kelimenin çeşitli bağamlarda farklı anlamlar taşıma olasılığı göz önünde bulundurulur.
- Kelime İlişkileri: Analiz, metindeki kelimeler arasındaki ilişkileri anlamayı içerir. Bu, eşanlamlılık, karşılık veya diğer anlam ilişkilerini değerlendirerek metindeki bağlamı daha iyi anlamayı sağlar.
- Çözümleme ve Parçalama: Anlamsal Analiz, bir cümleyi veya metni parçalayarak farklı öğeleri ayırmayı içerir. Bu, cümlenin dilbilgisel yapısını ve içeriği öğelerin rollerini belirlemeye yardımcı olur.
- Duygu Analizi: Analiz, metindeki ifadelerin taşıdığı duygusal tonları anlamayı içerir. Bu, metnin olumlu, olumsuz veya nötr bir duyguya içerip içermediğini belirlemeye yönelik bir özelliktir.

3) Duygu Analizi: Doğal Dil İşleme'nin (NLP) önemli bir uygulama alanıdır ve metinlerdeki duygusal ifadeleri belirleme ve analiz etme sürecini içerir. Bu, metinlerin taşıdığı duygusal tonları ve ifadeleri anlamaya yeteneği ile ilgilidir.

- Duygu Tanıma: Duygu Analizi, metinde geçen ifadeleri analiz ederek içerdikleri duygusal tonları belirlemeyi amaçlar. Bu tonlar genellikle olumlu, olumsuz veya nötr olarak sınıflandırılabilir.
- Duygu Yoğunluğu: Analiz, metindeki duygusal ifadelerin yoğunluğunu değerlendirerek ne kadar güçlü bir duygusal etki taşıdıklarını anlamayı hedefler. Yoğunluk, duygusal içeriğin belirli bir konuda ne kadar güçlü olduğunu yansıtabilir.
- Duygusal Bağlamı Anlama: Duygu Analizi, ifadeleri bağlam içinde değerlendirir. Aynı kelimenin veya ifadenin farklı bağlamlarda farklı duygusal anlamlar taşıma olasılığını göz önünde bulundurarak anlam karmaşıklığını çözmeyi amaçlar.
- Duygu Değişimlerini İzleme: Analiz, bir metindeki duygusal ifadelerdeki değişimleri izleyerek zaman içinde duygusal durumu anlamayı hedefler. Bu, bir hikayenin veya metnin duygusal gelişimini anlamak için önemlidir.
- Duygu Analizi Modelleri: Duygu Analizi genellikle özel modeller ve ya algoritmalar kullanılarak gerçekleştirilir. Bu modeller, metindeki kelime seçimleri, cümle yapıları ve diğer özelliklerini analiz ederek duygusal içeriği çıkarmayı amaçlar.

4) Dil Çevirisi: Dil Çevirisi, Doğal Dil İşleme'nin (NLP) önemli bir uygulama alanıdır ve bu alandaki temel ilke, metinlerin anlamını koruyarak farklı diller arasında çeviri yapma yeteneğini içerir.

- Çift Yönlü Çeviri: NLP tabanlı dil çeviri modelleri, metinleri bir kaynak dilinden hedef dile çevirebilir ve aynı zamanda hedef dilden kaynak dile çevirebilir. Bu, çift yönlü dil çevirisinin gerçekleştirilebilmesini sağlar.
- Anlamı Koruma: Dil çevirisinde temel amaç, metnin anlamını koruyarak diğer dile çevirmektir. Bu, çevrilen metnin aynı mesajı doğru bir şekilde iletebilmesi anlamına gelir.
- Sözcük Oyunları ve İfadelerin Yorumlanması: NLP tabanlı çeviri modelleri, sözcük oyunları, deyimler ve ifadeler gibi dilin karmaşıklıklarını anlamak ve doğru bir şekilde çevirmek için geliştirilir.
- Dilbilgisel Kurallar ve Yapı Analizi: Dil çevirisini, dilbilgisel kurallar ve yapı analizini içerir. Cümle yapısını anlamak ve doğru dilbilgisel kurallara uygun bir şekilde çeviri yapmak önemlidir.

- Öğrenen Modeller ve Derin Öğrenme: Gelişmiş NLP modelleri, derin öğrenme yöntemleri kullanarak büyük metin verileri üzerinde eğitilir. Bu modeller, dil çevirisi sırasında daha karmaşık yapıları anlamak ve daha doğru çeviriler üretmek için öğrenir.
- Sektöre Özgü Çeviri Modelleri: Belirli sektörlerde yönelik özel çeviri modelleri geliştirilebilir. Örneğin, tıp, hukuk veya teknik terimleri içeren metinler için özel çeviri modelleri oluşturulabilir.

5) Özetleme: Bu ilke, uzun metinleri analiz ederek ana fikirleri ve önemli detayları çıkarma yeteneğini içerir.

- Ana Fikirleri Belirleme: Özetleme, metindeki ana fikirleri belirleme yeteneğini içerir. Uzun bir metindeki temel mesajı veya odak noktasını çıkarma sürecini kapsar.
- Önemli Detayları Seçme: NLP modelleri, metindeki önemli detayları seçme kabiliyetine sahiptir. Bu detaylar, metnin temel anlamını destekleyen veya güçlendiren unsurları içerebilir.
- Kısaltma ve Öz: Özetleme, metni kısaltma ve özleştirme yeteneğiyle ilgilidir. Uzun bir metni daha kısa ve öz bir şekilde getirme sürecini içerir.
- İlerlemeli Anlam İnşası: NLP modelleri, metni anlam düzeyinde inşa ederken ilerlemeli bir yaklaşım benimser. Başlangıçta metni anlama ve ardından ana fikirleri ve önemli detayları belirleme aşamalarını içerir.
- Çeşitli Metin Türlerine Uyarlanabilme: Özetleme, haber makaleleri, akademik yazılar, blog gönderileri gibi çeşitli metin türleri üzerinde uygulanabilir. NLP modelleri, farklı dil ve konseptlere uyum sağlayabilir.
- Toplu Özetleme ve Metin Gruplama: Birden fazla metni bir araya getirerek toplu özetleme ve metin gruplama yeteneklerini içerebilir. Bu, büyük metin koleksiyonları üzerinde geniş kapsamlı özetleme sağlamak için kullanılabilir.
- Duygu ve Tonu Yansıtma: Özetleme süreci, metnin taşıdığı duygu ve tonu da yansıtabilir. Bu, özellikle duygu analizi ve metnin genel tonunu anlamak açısından önemlidir.

- Bilgi Kaybını Minimuma İndirme: NLP modelleri, özetleme sırasında önemli bilgilerin kaybını minimuma indirme amacını taşır. Ana fikirleri ve önemli detayları korurken, gereksiz bilgileri elemeyi hedefler.

Doğal Dil İşleme içindeki dil modellemesi, metin verilerinin örüntülerini anlama ve dilin yapısını kavrama konusunda etkili bir araç olarak öne çıkıyor. Bu modeller, özellikle metin üretimi, çeviri ve duygusal analizi gibi NLP uygulamalarında başarıyla kullanılmaktadır. Anlamsal analiz, metinlerin derinlemesine anlaşılmasına odaklanarak, bilgi çıkarma, duygusal analizi, otomatik cevaplandırma gibi çeşitli NLP görevlerinde etkili bir biçimde kullanılabilir. Duygu analizi, sosyal medya analizi, müşteri geri bildirimini değerlendirmesi, medya takibi ve duygusal pazarlama gibi birçok alanda önemli bir rol oynar. Bu sayede büyük metin verilerinden duygusal örüntüler çıkarmak ve daha geniş bir bağlamda anlamak mümkün olur. Dil çevirisini ise küresel iletişim kolaylaştırır ve farklı diller arasındaki bariyerleri aşarak bilgiye erişimi artırır. Özellikle çeşitli dil çiftleri arasında etkili ve doğru çeviri yapabilen NLP modelleri, birçok sektörde yaygın bir şekilde kullanılmaktadır. Bu özellikler, NLP'nin özetleme yeteneğini güçlendirir ve metin verilerini daha kapsamlı ve anlam dolu bir biçimde işleme yeteneğini temsil eder.

Doğal Dil İşleme, metin verilerini anlama ve işleme yeteneğiyle öne çıkan bir teknoloji alanıdır. Temel işleyişi, öncelikle metin verilerinin toplanması ve temizlenmesi ile başlar. Ardından, tokenizasyon adımıyla metinler kelimelere veya cümlelere ayrılır. Dilbilgisel analiz kullanılarak kelimelerin dilbilgisel yapısı anlaşılır, kökleri çıkartılır veya lemmatization yapılır. Dil modellemesi, metin verilerinin örüntülerini anlamak ve dilin yapısını kavramak adına önemli bir adımdır. Bu aşamada, metin üzerinde öğrenilen dilbilgisel kurallar ve yapılar kullanılarak bir dil modeli oluşturulur. Anlamsal analiz, metinlerdeki anlamı çözümlemek için kullanılır ve metin içindeki kelimelerin veya ifadelerin bağlamını anlamayı amaçlar.

Uygulama alanına özgü işlemler, genellikle belirli bir görevi yerine getirmek için özel olarak tasarlanan NLP uygulamalarını içerir. Elde edilen sonuçlar incelenir, değerlendirilir ve gerektiğinde model geliştirme ve yeniden eğitme adımlarıyla sürekli olarak iyileştirilir. NLP'nin temel işleyişi, büyük metin verilerini etkili bir şekilde analiz ederek dilin karmaşıklığını anlamak ve çeşitli uygulamalarda kullanmak üzerine odaklanır. Bu işleyişi daha iyi anlamak için bir de kodsal şekilde inceleyelim. Bunun için ilk adımıız,

1) Veri Toplama ve Temizleme: NLP projelerine başlarken, işlem yapacağınız dilde veya konuda uygun metin verileri toplanması gereklidir; bu adım, projenin temelini oluşturur. Örnek olarak şöyle bir textimiz olsun:

```
text="Doğal Dil İşleme, bilgisayarların insan dilini anlama, yorumlama ve üretme yeteneği üzerine odaklanan bir teknoloji alanını ifade eder. NLP, metin verilerini analiz ederek dilin örüntülerini çözümleme, anlamsal anlayış geliştirme ve dilin karmaşıklığını anlama amacını taşıır. Bu teknoloji, metin tabanlı verileri anlamlandırmak, dildeki duygusal anlayışını çözümlemek, çeviri yapmak gibi bir dizi uygulama alanında kullanılır. Dil modellemesi, dilbilgisel analiz, anlamsal analiz gibi alt dalları içeren NLP, makine öğrenimi ve yapay zeka teknikleriyle birleştirilerek metin verilerinin daha derin bir düzeyde anlaşılmasını ve işlenmesini sağlar. NLP, çağımızın dijitalleşen dünyasında büyük veri setlerinin etkili bir şekilde yönetilmesi, içgörü elde edilmesi ve çeşitli uygulamalarda kullanılması açısından önemli bir rol oynar."
```

Ancak, toplanan veriler genellikle temizlenmelidir, çünkü gereksiz karakterler, özel etiketler ve anlamsız veriler işleme sürecini zorlaştırbılır ve sonuçları etkileyebilir. İhtiyaç durumu ve proje yapısına bağlı olarak birçok data collection yöntemi bulunmaktadır.

```
import re
import string
text = text.lower()
text = re.sub(f"[{re.escape(string.punctuation)}]", "", text)
text
```

Çıktı:

```
text='doğal dil işleme bilgisayarların insan dilini anlama yorumlama ve üretme yeteneği üzerine odaklanan bir teknoloji alanını ifade eder nlp metin verilerini analiz ederek dilin örüntülerini çözümleme anlamsal anlayış geliştirme ve dilin karmaşıklığını anlama amacını taşıır bu teknoloji metin tabanlı verileri anlamlandırmak dildeki duygusal anlayışını çözümlemek çeviri yapmak gibi bir dizi uygulama alanında kullanılır dil modellemesi dilbilgisel analiz anlamsal analiz gibi alt dalları içeren nlp makine öğrenimi ve yapay zeka teknikleriyle birleştirilerek metin verilerinin daha derin bir düzeyde anlaşılmasını ve işlenmesini sağlar nlp çağımızın dijitalleşen dünyasında büyük veri setlerinin etkili bir şekilde yönetilmesi içgörü elde edilmesi ve çeşitli uygulamalarda kullanılması açısından önemli bir rol oynar'
```

Cleaning adımı sonrasında, data çıktısında büyük-küçük harf karmaşası, noktalama işaretleri gibi normalizasyon gerektiren etkenlerden kurtulmuş olduğumuzu görüyoruz. Cleaning adımı, sonraki süreçler için büyük önem taşır. Burada minik bir örnek üzerinden cleaning işlemine baktık, ancak projenin ihtiyacına göre cleaning adımını özelleştirmek ve birçok işlem yapmak mümkündür. Bu esneklik, projenin gereksinimlerine uygun ve optimize edilmiş bir veri seti elde etme sürecinde kritik bir rol oynar.

2) Metin Ön İşleme: Metin verilerini düzenledikten sonra, onları NLP algoritmaları tarafından işlenebilecek bir biçimde getirmek oldukça kritiktir. Bu aşama, metin verilerini daha anlamlı, yapısal ve işlenebilir hale getirerek NLP modellerinin etkili bir şekilde çalışmasını sağlar. Tokenization gibi tekniklerle metinleri kelimeye veya cümlelere ayırmak, dilbilgisel analiz kullanarak kelimelerin köklerini çıkarmak veya lemmatization yapmak, anlamsal analizle metinlerdeki anlamsal çözümlemek bu

işlemlere örnektir. Bu adımlar, metin verilerini daha önce belirlenmiş dil modellerine uygun bir formata dönüştürerek, NLP algoritmalarının daha etkili bir şekilde çalışabilmesine olanak tanır. Bu sayede, dilin karmaşıklığı ve anlam derinliği üzerinde daha gelişmiş analizler gerçekleştirilebilir, duygusal analizi, çeviri, özetleme gibi uygulamalar daha hassas sonuçlar verebilir.

- Tokenization İşlemi: Bir metni daha küçük parçalara, yani "tokenlara ayırma işlemidir. Bu işlem, metin verilerini daha küçük ve işlenebilir parçalara böler, bu parçalar genellikle kelimeler veya cümleler olabilir.

```
from nltk.tokenize import word_tokenize  
tokens=word_tokenize(text)  
tokens
```

Çıktı:

```
[1:'doğal',  
 'dil',  
 'işleme',  
 'bilgisayarların',  
 'insan',  
 'dilini',  
 'anlama',  
 'yorumlama',  
 've',  
 'üretme',  
 'yeteneği',  
 'üzerine',  
 'odaklanan',  
 'bir',  
 'teknoji',  
 'alanını',  
 'ifade',  
 'eder',  
 'nlp',  
 'metin',  
 'verilerini',  
 'analiz',  
 'ederek',  
 'dilin',  
 'örüntülerini',  
 'çözümleme',  
 'anlamsal',  
 ]
```

Tokenizasyonun temel amacı, metin verilerini dilbilgisel analiz ve işlemler için uygun bir hale getirmektir. Örneğin, bir cümleyi kelimelere ayırmak, bir paragrafi cümlelere bölmek gibi. Tokenizasyon, metin madenciliği, duygusal analizi, dil modellemesi ve benzeri NLP görevlerinde genellikle ilk adım olarak uygulanır.

- Stop words removal: Doğal dil işleme (NLP) uygulamalarında sıkça kullanılan bir ön işleme tekniğidir. Stop words, genellikle dilbilgisel analizlerde pek fazla anlam taşımayan, sık kullanılan kelimelerdir (örneğin, "ve", "ama", "veya"). Bu

kelimeler, genellikle analiz veya modelleme süreçlerinde gereksiz gürültüye neden olabilir. Metin verilerinden bu tür sık kullanılan kelimelerin çıkarılması işlemidir. Bu, metin verilerinin daha temiz, odaklı ve anlam dolu hale gelmesini sağlar. Bu işlem, metin madenciliği, duyu analizi, sınıflandırma ve benzeri NLP görevlerinde kullanılabilir.

```
Orijinal Metin |
['doğal','dil','işleme','bilgisayarların','insan','dilini','anlama','yorumlama','ve','üretme','yeteneği',
'überine','odaklanan','bir','teknoloji','alanını','ifade','eder','nlp','metin','verilerini','analiz',
'ederek','dilin','örüntülerini','çözümleme','anlamsal','anlayış','geliştirme','ve','dilin','karmaşıklığını',
'anlama','amacını','taşır','bu','teknoloji','metin','tabanlı','verileri','anlamlandırmak','dildeki',
'duygu','ve','anlamı','çözümlemek','çeviri','yapmak','gibi','bir','dizi','uygulama','alanında','kullanılır',
'dil','modellemesi','dilbilgisel','analiz','anlamsal','analiz','gibi','alt','dalları','içeren','nlp',
'makine','öğrenimi','ve','yapay','zeka','teknikleriyle','birleştirilerek','metin','verilerinin','daha',
'derin','bir','düzeyde','anlaşılmasını','ve','işlenmesini','sağlar','nlp','çağımızın','digitalleşen',
'dünyasında','büyük','veri','setlerinin','etkili','bir','şekilde','yönetilmesi','içgörü','elde','edilmesi',
've','çeşitli','uygulamalarda','kullanılması','açısından','önemli','bir','rol','oynar']
```

```
from nltk.corpus import stopwords
stop_words = set(stopwords.words('turkish'))

from nltk.corpus import stopwords
turkish_stop_words = set(["ve", "veya", "ama", "bir", "daha", "gibi"])

filtered_sentence = [w for w in tokens if not w.lower() in turkish_stop_words]
filtered_sentence = []

print(filtered_sentence)
```

Çıktı:

```
Stop Words Removal Sonrası
['doğal','dil','işleme','bilgisayarların','insan','dilini','anlama','yorumlama','üretme','yeteneği',
'überine','odaklanan','teknoloji','alanını','ifade','eder','nlp','metin','verilerini','analiz',
'ederek','dilin','örüntülerini','çözümleme','anlamsal','anlayış','geliştirme','dilin','karmaşıklığını',
'anlama','amacını','taşır','teknoloji','metin','tabanlı','verileri','anlamlandırmak','dildeki',
'duygu','anlamı','çözümlemek','çeviri','yapmak','dizi','uygulama','alanında','kullanılır',
'dil','modellemesi','dilbilgisel','analiz','anlamsal','analiz','dalları','içeren','nlp',
'makine','öğrenimi','yapay','zeka','teknikleriyle','birleştirilerek','metin','verilerinin',
'derin','düzeyde','anlaşılmasını','işlenmesini','sağlar','nlp','çağımızın','digitalleşen',
'dünyasında','büyük','veri','setlerinin','etkili','şekilde','yönetilmesi','içgörü','elde','edilmesi',
'çeşitli','uygulamalarda','kullanılması','açısından','önemli','rol','oynar']
```

NLP projelerinde bu işlem sayesinde metin verileri daha etkili bir şekilde işlenmekte ve analiz edilmektedir.

- Kök Çıkarma (Stemming) ve Kökten Türetme (Lemmatization): Doğal dil işleme (NLP) uygulamalarında metin verilerini işleme ve analiz etme

süreçlerinde kullanılan iki farklı dilbilgisel ön işleme tekniğidir. Her ikisi de kelime formunu daha temel bir hale getirme amacı taşır, ancak farklı yaklaşımlar kullanırlar.

Kök çıkarma, bir kelimenin kökünü elde etmeye çalışan bir süreçtir. Bu süreçte, bir kelimenin sonundaki ekleri kaldırarak kelimenin temel kısmını bulmaya çalışır. Bu yöntem, benzer anlamlı kelimeleri aynı köke indirmeye çalışır, ancak bu işlem sırasında ortaya çıkan kökler genellikle gerçek bir kelimenin kökü olmayabilir. Örneğin, "running" kelimesinin kökü "run" olarak çıkarılabilir.

Kökten türetme, bir kelimenin gerçek dilbilgisel kökünü (lemmasını) bulma işlemidir. Bu yöntem, dilbilgisel kuralları ve sözlük bilgisini kullanarak kelimenin gerçek kökünü belirlemeye çalışır. Bu sayede elde edilen kök, genellikle bir sözlükte bulunan bir kelimenin temsilcisidir. Örneğin, "running" kelimesinin kökü "run" olarak belirlenebilir.

```
from snowballstemmer import TurkishStemmer
from nltk.stem import WordNetLemmatizer
from nltk.tokenize import word_tokenize

Kök çıkarma (Stemming)
turkish_stemmer = TurkishStemmer()
stemmed_text = [turkish_stemmer.stemWord(word) for word in original_text]

Kökten türetme (Lemmatization)
lemmatizer = WordNetLemmatizer()
lemmatized_text = [lemmatizer.lemmatize(word) for word in original_text]

Sonuçları yazdırma
print("Original Text:", original_text)
print("Stemmed Text:", stemmed_text)
print("Lemmatized Text:", lemmatized_text)
```

Çıktı:

```
Original Text: ['doğal', 'dil', 'işleme', 'bilgisayerların', 'insan', 'dilini', 'anlama', 'yorumlama',
'üretme', 'yeteneği', 'üzerine', 'odaklanan', 'teknoloji', 'alanını', 'ifade', 'eder', 'nlp', 'metin',
'verilerini', 'analiz', 'ederek', 'dilin', 'örüntülerini', 'çözümleme', 'anlamsal', 'anlayış',
'geliştirme', 'dilin', 'karşılıklığını', 'anlama', 'amacını', 'taşır', 'teknoloji', 'metin', 'tabanlı',
'verileri', 'anlamlandırmak', 'dildeki', 'duyu', 'anlamı', 'çözümlemek', 'geviri', 'yapmak', 'dizi',
'uygulama', 'alanında', 'kullanılır', 'dil', 'modellemesi', 'dilbilgisel', 'analiz', 'anlamsal', 'analiz',
'dalları', 'içeren', 'nlp', 'makine', 'öğrenimi', 'yapay', 'zeka', 'teknikleriyle', 'birleştirilerek',
'metin', 'verilerinin', 'derin', 'düzyede', 'anlaşılmasını', 'işlenmesini', 'sağlar', 'nlp', 'çağımızın',
'dijitalleşen', 'dünyasında', 'büyük', 'veri', 'setlerinin', 'etkili', 'şekilde', 'önetilmesi', 'icgörü',
'elde', 'edilmesi', 'çeşitli', 'uygulamalarla', 'kullanılması', 'açısından', 'önemli', 'rol', 'oynar']

Stemmed Text: ['doğal', 'dil', 'işleme', 'bilgisayar', 'insan', 'dil', 'anla', 'yorumla', 'üret',
'yetenek', 'üzer', 'odakla', 'teknoloji', 'alan', 'ifade', 'et', 'nlp', 'metin', 'veri', 'analiz',
'eder', 'dil', 'örüntü', 'çözümle', 'anlamsal', 'anlays', 'geliş', 'dil', 'karşık', 'anla', 'amac',
'taşır', 'teknoloji', 'metin', 'taban', 'veri', 'anlamla', 'dil', 'duyu', 'anlam', 'çözümle', 'çevir',
'yap', 'dizi', 'uygula', 'alan', 'kullan', 'dil', 'modelleme', 'dilbilgisel', 'analiz', 'anlamsal',
'analiz', 'dal', 'iger', 'nlp', 'makine', 'ögren', 'yapay', 'zeka', 'teknik', 'birleştir', 'metin',
'veri', 'derin', 'düzey', 'anla', 'işle', 'sağla', 'nlp', 'çağ', 'dijitalleş', 'dünya', 'büyük', 'veri',
'set', 'etkili', 'şekil', 'yönet', 'icgörü', 'elde', 'et', 'çeşitli', 'uygula', 'kullan', 'açı', 'önemli',

Lemmatized Text: ['doğal', 'dil', 'işleme', 'bilgisayerların', 'insan', 'dilini', 'anlama', 'yorumlama',
'üretme', 'yeteneği', 'üzerine', 'odaklanan', 'teknoloji', 'alanını', 'ifade', 'eder', 'nlp', 'metin',
'verisi', 'analiz', 'ederek', 'dilin', 'örüntü', 'çözümleme', 'anlamsal', 'anlays', 'geliştirme',
'dilin', 'karşılıklığını', 'anlama', 'amacını']
```

3) Analiz ve İşleme(Analysis and Processing): Doğal Dil İşleme, metin verilerini analiz etmek için kullanılan güçlü bir alandır. Metin sınıflandırma, veri düzenleme ve filtreleme süreçlerinde etkili bir rol oynar. Örneğin spam e-postalarını sınıflandırmak veya makaleleri konularına göre etiketlemek gibi görevlerde kullanılır. Duygu analizi, metinlerin içерdiği duygusal tonları belirleme konusunda önemli bir araçtır ve kullanıcı görüşlerini anlama veya marka itibarını değerlendirme gibi uygulamalarda değerlidir. Konu modellemesi, büyük metin koleksiyonlarının karmaşıklığını yönetmede yardımcı olur ve metinlerdeki ana temaları tanımlayarak içgörü sağlar. Özellikle büyük veri analitiği projeleri ve içerik yönetimi sistemlerinde kritik bir rol oynar. Entite tanıma, metinlerdeki önemli varlıklar tanımlar ve bu da bilgi çıkarmada ve metin madenciliği projelerinde oldukça değerlidir. NLP'nin temel görevleri, metin verilerini anlamlı ve yapılandırılmış bir şekilde kullanmaya olanak tanır, bu da çeşitli sektörlerde bilgi çıkarma, veri sınıflandırma ve karar destek sistemlerinin geliştirilmesinde önemli bir araç olarak hizmet verir.

Günlük Hayattaki Uygulamalar

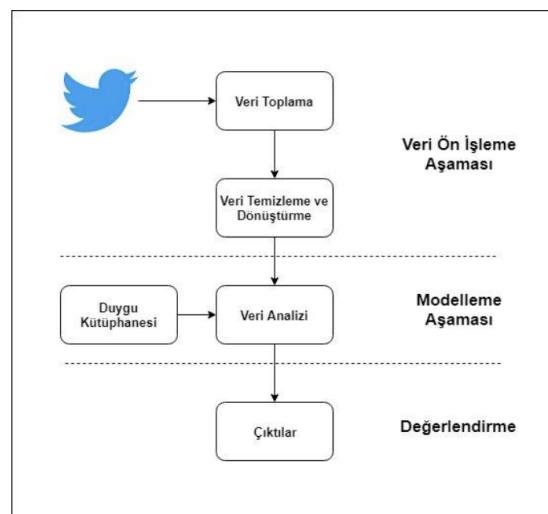
NLP'nin Sosyal Medya Analizi Ve Duygu Analizi Üzerindeki Etkisi

NLP, sosyal medya analizi ve duygu analizi alanlarında büyük bir etki yaratmaktadır. Sosyal medya, birçok insanın günlük yaşamını, düşüncelerini ve duygularını paylaştığı bir platformdur. NLP, bu büyük veri setlerini analiz ederek önemli içgörüler elde etme yeteneği ile sosyal medya analizinde kritik bir rol oynar. Sosyal medya üzerindeki duygusal içerikleri anlama yeteneği, duygu analizi adı verilen bir NLP görevi aracılığıyla gerçekleşir. Bu analiz, metin verilerindeki duygusal tonları belirlemek için kullanılır. Örneğin, marka itibarı yönetimi, ürün geri bildirimleri analizi veya genel kamu duyarlılığı hakkında bilgi edinme amacıyla duygu analizi sıkça kullanılır. NLP'nin sosyal medya analizi ve duygu analizi üzerindeki etkisi, şirketlerin müşterileri memnuniyetini değerlendirmelerine, pazar trendlerini anlamalarına ve hatta acil durumları önceden tahmin etmelerine olanak tanır. Bu sayede, hızlı ve doğru kararlar almak, müşteri ihtiyaçlarına daha iyi cevap vermek ve genel olarak iş stratejilerini geliştirmek mümkün olur.

NLP'nin sosyal medya analizi ve duygu analizi üzerindeki etkisi konusunda birçok deney ve araştırma yapılmıştır. Bu alandaki çalışmalar, doğal dil işlemenin büyük veri setlerini anlamlandırma, duygusal içerikleri analiz etme ve sosyal medya üzerindeki eğilimleri belirleme konularında ne kadar etkili olabileceğini göstermektedir. Araştırmacılar, sosyal medya platformlarından elde edilen metin verilerini kullanarak NLP tekniklerini test ederler. Bu deneyler genellikle duygu analizi, konu modellemesi, kullanıcı davranışlarının tahminlenmesi gibi konulara odaklanır. Akademik literatürde

ve endüstriyel uygulamalarda bu konuda birçok yayın bulunmaktadır. Bu yayınlar, NLP'nin sosyal medya analizi ve duyu analizi konularında nasıl kullanılabileceğine dair çeşitli örnekleri ve başarı hikayelerini içermektedir. Avrupa Bilim ve Teknoloji Dergisi'nin 16. sayısında Semra Yıldırım, Yücel Batu Salman ve Serkan Ayvaz tarafından yazılan "Türkçe Duygu Kütüphanesi Geliştirme: Sosyal Medya Verileriyle Duygu Analizi Çalışması" raporunda temel amaç, Türkçe için kapsamlı bir duyu kütüphanesi oluşturmaktır. Yeni kütüphane, Türkçe sosyal medya paylaşımlarında etkili duyu analizi çalışmalarına katkı sağlamayı amaçlamaktadır. Mevcut diğer kütüphaneler incelenerek genişletilmiş ve temelde emoji karakterleri ve puanlama altyapısı eklenmiştir.

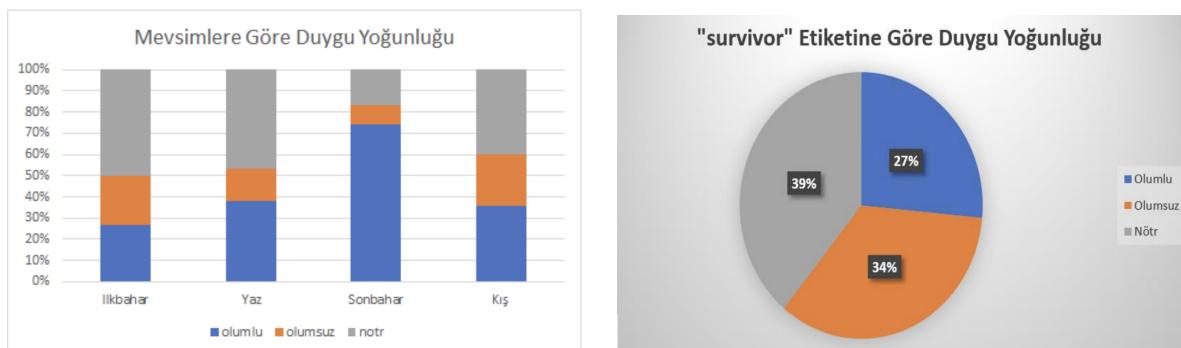
Türkçe Duygu Kütüphanesi'nin etkinliğini değerlendirmek için Twitter'da belirli etiketlerle oluşturulan paylaşımlar toplanmış ve bu veri üzerinden duyu analizi çalışmaları yapılmıştır. İki farklı konu başlığı altında gerçekleştirilen analiz çalışmalarından birincisi, hava değişikliklerinin kullanıcılar üzerindeki etkisini ölçmeyi amaçlamıştır. Bu analizde yaz ve kış aylarında toplanan veri setleri incelenerek, Türkçe paylaşım yapan kullanıcıların kış aylarında daha olumsuz paylaşımlar yaptığı, bahar ve yaz aylarında ise daha olumlu paylaşımlar yaptığı gözlemlenmiştir. Diğer analiz, Türk televizyonlarında belirli bir süre boyunca popüler olan Survivor adlı yarışma programı izleyicilerinin Türkçe paylaşımlarını ele almıştır. Bu analizde, kullanıcıların genellikle bu programla ilgili olumsuz paylaşımlar yaptığı ve aynı zamanda paylaşımlarda bozuk Türkçe kullanım alışkanlıklarını tespit edildiği görülmüştür. Araştırmada, Türkçe Twitter paylaşımları üzerinde gerçekleştirilen çeşitli deney ve araştırmalar, dil yapısı ve kullanıcı alışkanlıklarını olmak üzere iki ana engel grubuya karşılaşımıştır.



Türkçe'nin esnek ve diğer dillere göre farklı bir yapıya sahip olması, kullanıcıların İngilizce klavye kullanma alışkanlıkları gibi faktörler, Türkçe duyu analizi çalışmalarını zaman alıcı kılmaktadır. Bu zorluklara rağmen, Türkçe kelimeler

üzerinden duyu analizi konusunda birçok çalışma yapılmıştır. Firmaların bu alana odaklanmak için özel departmanlar kurduğu, internet üzerindeki Türkçe verinin analiz edilmesinin yeni hizmetlerin ve iş alanlarının oluşmasına katkı sağladığı gözlemlenmiştir. Yapılan çalışmalarda çeşitli graplama teknikleri, analiz seviyeleri ve teknolojiler kullanılmıştır. Örnek çalışmalardan biri, Erkut Evirgen'in Türkçe tweetlerde duyu analizi üzerine yaptığı tez çalışmasıdır. Bu çalışmada, R programlama dili kullanılarak düzensiz formatlı kelimelerin işlenmesi ve yaygın makine öğrenme yöntemlerinin karşılaştırılması gibi adımlar izlenmiştir. Diğer örnek çalışmalar arasında, Dilek Kayahan, Aslı Sergen ve Banu Diri'nin Türkiye'deki televizyon programlarının reytinglerini tahminleme çalışması, B. İbrahim Sevindi'nin Türkçe film yorumları üzerindeki duyu analizi, ve Sinem Demirci'nin duyu analizi için "finer-grained emotion analysis" yöntemini kullanması yer almaktadır. Bu çalışmaların ortak noktası, Türkçe duyu analizi konusunda çeşitli metodolojilerin ve tekniklerin başarıyla uygulanmış olmasıdır.

Sonuç olarak ise, iki deneyin gerçekleştirildiği çalışmanın birincisinde, kullanıcıların hava durumuna göre değişen duyguları ölçülmüş ve kış aylarında olumsuz, yaz aylarında ise pozitif paylaşım oranlarının arttığı gözlemlenmiştir. Mevsim değişikliklerinin kullanıcı duygularını etkilediği tespit edilmiştir. İkinci deneyde ise Türkçe paylaşım yapan Twitter kullanıcılarının Survivor etiketi ile yaptıkları paylaşımlar incelemiş ve genel olarak olumsuz içerikli paylaşımların olduğu gözlemlenmiştir.



Bu çalışma ile, Twitter üzerinden belirli etiketlerle toplanan Türkçe veriler üzerinde duyu analizi gerçekleştirmiştir. Analizlerde varlık-görüş seviyesi ve cümle seviyesi duyu analizi yöntemleri kullanılmış, ayrıca geliştirilen duyu kütüphanesine duyu yönelik gruplamaları ve duyu yoğunluğu parametreleri eklenmiştir. Bu sayede cümle seviyesinde yapılan analizlerde anlam kayıpları en aza indirilmiştir. Ek olarak, duyu kütüphanesine eklenen emojiler de puanlandırılarak duyu analizini desteklemiştir. Türkçe'deki kelime grupları, ikilemeler ve dil sorunlarına çözüm bulmak amacıyla kütüphane zenginleştirilmiş ve Twitter kullanıcılarının Türkçe

metinleri İngilizce alfabesi kullanarak yazmaları sonrasında oluşan paylaşımları da analiz edebilme kabiliyeti kazandırılmıştır.

Sesli Asistanlar (Alexa, Siri) Ve Konuşma Tanıma Teknolojileri

Sesli asistanlar, günümüzde yaygın olarak kullanılan teknolojik araçlardır ve kullanıcıların sesli komutlarını algılayarak çeşitli görevleri yerine getirmelerine olanak tanır. Öne çıkan sesli asistanlardan bazıları Amazon'un Alexa, Apple'in Siri ve Google Asistanıdır. Bu sesli asistanlar, konuşma tanıma teknolojileri üzerine kurulu olan yapay zeka sistemlerini kullanarak kullanıcılarla etkileşimde bulunur. Sesli asistanlar genellikle doğal dil işleme (NLP) ve konuşma tanıma teknolojilerini içerir. Doğal dil işleme, insan dilini anlamak ve yorumlamak için kullanılan bir teknoloji dalıdır. Konuşma tanıma ise sesli komutları metne dönüştürerek anlamlandırma sürecidir. Bu teknolojiler sayesinde sesli asistanlar, kullanıcının talimatlarını anlayabilir ve çeşitli görevleri yerine getirebilir. Alexa, Siri ve Google Asistan gibi sesli asistanlar geniş bir dizi görevi yerine getirebilir. Kullanıcırlara hava durumu bilgisi sağlama, müzik çalma, takvim etkinliklerini hatırlatma, soruları yanıtlama, ev otomasyon sistemlerini kontrol etme gibi işlevlere sahiptirler. Bu teknolojiler, kullanıcı deneyimini zenginleştirmek ve günlük yaşamı kolaylaştmak adına geliştirilmekte ve yaygın olarak benimsenmektedirler.



Peki bu asistanlar nasıl eğitilmektedir? Asistanlar genellikle bir dizi karmaşık eğitim süreçlerinden geçerler. Bu süreçler, yapay zeka (AI) ve doğal dil işleme (NLP) teknolojilerinin geliştirilmesini içermektedir.

1) Veri Toplama ve Etiketleme: Sesli asistanları eğitmek için büyük miktarda veriye ihtiyaç vardır. Bu veriler genellikle sesli komutlar, metin transkripsiyonları ve kullanıcı etkileşimleri içerir. Toplanan veriler, etiketleme sürecinde insan uzmanlar tarafından incelenir ve kategorilere ayrılır.

2) Doğal Dil İşleme Algoritmalarının Geliştirilmesi: Sesli asistanların doğru bir şekilde çalışabilmesi için etkili doğal dil işleme (NLP) algoritmalarına ihtiyaç vardır. Bu algoritmalar, kullanıcıların sesli komutlarını anlama, dilin yapısını analiz etme ve anlam çıkarma yeteneklerini içerir.

3) Konuşma Tanıma Sistemlerinin Eğitimi: Sesli komutların doğru bir şekilde anlaşılabilmesi için konuşma tanıma sistemleri eğitilir. Bu süreçte, çeşitli konuşmacılardan gelen ses dalgaları örneklenir ve algoritmalar, sesli komutları metin formatına dönüştürmeye öğrenir.

4) Yapay Zeka ve Makine Öğrenimi: Sesli asistanlar, kullanıcı davranışları ve talepleri üzerinde öğrenme yetenekleri geliştirebilmek için yapay zeka ve makine öğrenimi tekniklerinden faydalananır. Bu, asistanların zaman içinde kullanıcıya daha iyi hizmet verebilmesini sağlar.

5) Gerçek Kullanıcı Geri Bildirimleri: Sesli asistanlar, kullanıcı etkileşimlerinden elde edilen gerçek zamanlı geri bildirimlere dayanarak sürekli olarak güncellenir ve iyileştirilir. Kullanıcıların sesli asistanı kullanma biçimleri ve verdikleri tepkiler, sistemleri daha akıllı ve duyarlı hale getirmek için kullanılır.

6) Büyük Veri ve Bulut Bilişim Kullanımı: Sesli asistanlar, genellikle büyük veri setleri ile çalışır ve bu veriler bulut bilişim altyapıları üzerinde depolanır. Bu, asistanların daha fazla veriye erişim sağlayarak öğrenme yeteneklerini artırmasına yardımcı olur.

Sesli asistanların eğitimi, karmaşık bir süreçtir ve genellikle birçok disiplini içerir. Veri güvenliği ve kullanıcı gizliliği, bu süreçte önemli bir rol oynar, bu nedenle asistanlar genellikle bu konularda sıkı güvenlik protokollerine tabidir.



Ses tanıma, insan sesini bir mikrofon aracılığıyla algılayarak bilgisayar tarafından tanınmasını sağlayan bir sistemdir. Microsoft'un teknolojik bir devrim olarak tanımladığı konuşma tanıma sistemi 1950'li yılların sonlarına doğru kullanılmaya başlanmıştır. Sonrasında ise insan-bilgisayar iletişiminde önemli bir ihtiyaç haline gelmiştir. Konuşma tanıma disiplininin zorlu bir çalışma alanı ve bir dizi alt problemler içermektedir. Konuşma tanıma problemleri arasında konuşmacı belirleme, konuşmacı tanıma, ayrık sözcük tanıma gibi çeşitli konular bulunmaktadır. Konuşma tanıma teknolojisinin ana bileşenleri ve çalışma prensipleri vardır. Bunlar,

1) Ses Dalgalarının Alınması: Konuşma tanıma süreci, ilk olarak kullanıcının sesini algılayarak ses dalgalarını kaydetmekle başlar. Bu, genellikle bir mikrofon veya başka bir ses algılama cihazı kullanılarak gerçekleştirilir.

2) Analog Ses Dalgalarının Dijital Formata Dönüşürtülmesi: Alınan ses dalgaları daha sonra analogdan digitale dönüştürülür. Bu adım, bilgisayarlar tarafından işlenebilen dijital verilere dönüşümü sağlar.

3) Ön İşleme: Ses dalgalarının dijital formata dönüştürülmesinin ardından, sinyal üzerinde ön işleme adımları uygulanır. Bu adımlar arasında gürültü azaltma, tonlama ve diğer ses sinyali iyileştirme teknikleri bulunabilir.

4) Özellik Çıkarımı: Ses dalgalarının dijital temsilinden, özellik vektörleri çıkarılır. Bu özellik vektörleri, sesin temel özelliklerini ve karakteristiklerini temsil eder. Mel frekans cepstrum katsayıları (MFCC) gibi özellikler sıkılıkla kullanılır.

5) Model Eğitimi: Özellik vektörleri, bir konuşma tanıma modelini eğitmek için kullanılır. Bu model, belirli kelimelerin veya seslerin temsilini öğrenir. Genellikle makine öğrenimi algoritmaları, yapay sinir ağları veya gizli Markov modelleri gibi teknikler kullanılır.

6) Tanıma ve Transkripsiyon: Eğitilen model, gerçek zamanlı olarak ses girdisini analiz eder ve belirli bir dildeki söylenen kelimeleri metin formatına dönüştürür. Bu adıma "tanıma" veya "transkripsiyon" denir.

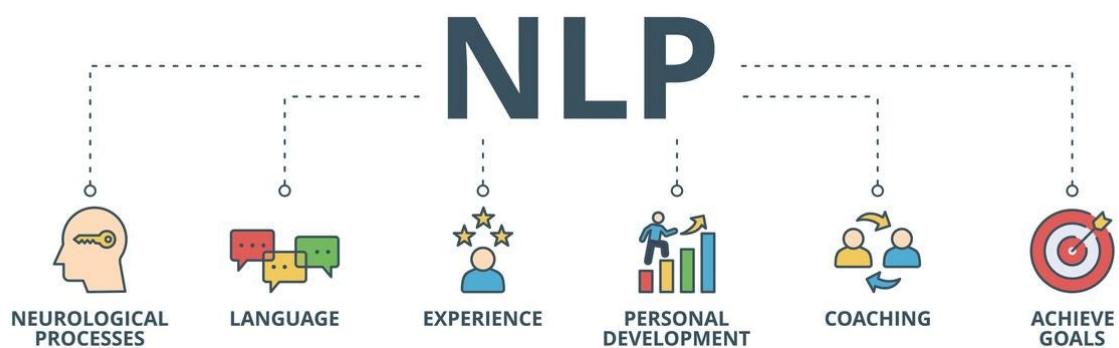
7) Dil Modelleri ve Kontekst: Konuşma tanıma sistemleri, dil modellerini kullanarak kelime sıralamalarını ve bağlamı anlamaya çalışır. Bu, bir kelimenin ardından muhtemel bir diğer kelimenin ne olabileceğini tahmin etmeye yardımcı olur.

8) Gelişmiş Özellikler: Bazı konuşma tanıma sistemleri, konuşan kişinin hissiyatını veya bağlamını daha iyi anlamak için duygusal analizi veya ses tonu analizi gibi gelişmiş özellikler de içerebilir.

Konusma tanımanın ticari potansiyeli ve insanlık için sunduğu avantajlar saymakla bitmez. Gelişen ses tanıma teknolojileri sayesinde, işitme engellilerin duyabilmeleri ve canlı yayınlardaki konuşmaların yanında metne çevrilebilmesi gibi birçok örnek verilebilir. Ayrıca bu teknoloji güvenlik, eğitim ve tıp alanlarında da önemli bir rol almaktadır.

Makine Çevirisi Ve Çok Dilli İletişimdeki Rolü

Doğal Dil İşleme, dilbilgisel analiz ve anlama yetenekleri sağlayan bir yapay zeka dalıdır ve bu alandaki en önemli uygulamalardan biri makine çevirisidir. Makine çevirişi, bir dildeki metinleri başka bir dile otomatik olarak çevirmeyi amaçlayan bir teknoloji olarak NLP'nin önemli bir parçasını oluşturur.



Makine çevirişi, geçmişten günümüze büyük bir evrim geçirmiştir ve dil çiftleri arasında etkili çeviriler yapabilen modellerin geliştirilmesi ile önemli bir noktaya gelmiştir. Bu evrimde NLP modellerinin güçlenmesi, büyük veri setlerinin kullanılabilirliğinin artması ve derin öğrenme tekniklerinin yaygınlaşması etkili olmuştur. Makine çevirişi, temel olarak iki ana yaklaşım üzerine kuruludur,

- 1) İstatistiksel Makine Çevirisi:** Dil çevirisini alanında önemli bir role sahip olan bir yaklaşımı temsil eder. Bu metodoloji, büyük miktardaki paralel metin veri setlerinden yararlanarak dil çiftleri arasındaki istatistiksel ilişkileri modellemeyi amaçlar. İstatistiksel Makine Çevirisi'nin temelinde, bir kaynak dildeki kelimenin veya ifadenin hedef dildeki karşılığını belirlemek için istatistiksel yöntemlerin kullanılması yatar. Bu süreç, dil çiftleri arasındaki sözcük hizalamasını ve çeviri olasılıklarını belirlemek üzere istatistiksel modellerin kullanılmasını içerir. Örneğin, büyük bir paralel veri setindeki çeviri örnekleri üzerinden, bir kelimenin diğer dile çevrilme olasılığı hesaplanabilir. İstatistiksel Makine Çevirisi, veri odaklı bir yaklaşım benimseyerek, çeviri kalitesini artırmak ve doğal dildeki farklı ifadeleri başarılı bir şekilde çevirebilmek için geniş veri setlerinden öğrenilen istatistiksel örüntülerden faydalananır. Bu sayede, dil çevirisini sistemleri, dil çiftleri arasındaki çeşitlilikleri ve bağamları daha etkili bir şekilde kavrayabilir.
- 2) Derin Öğrenme Tabanlı Makine Çevirisi:** Dil çevirisini alanında önemli bir gelişme sağlayan ve daha önceki yöntemlere kıyasla üstün performans gösteren bir yaklaşımı temsil eder. Geleneksel sistemlerin aksine, NMT, büyük ölçüde end-to-end (uçtan uca) öğrenmeye dayanır ve çeviri sürecini bir bütün olarak ele alır. Bu yöntem, dil çevirisindeki karmaşıklıkları anlamak ve çözmek için derin sinir ağlarından faydalananır. NMT'nin en önemli avantajlarından biri, kelime seçimi ve dilbilgisi kurallarını daha iyi kavrayarak daha doğal ve akıcı çeviriler üretilmesidir. Büyük miktarda paralel metin verisi üzerinde eğitilen derin öğrenme modelleri, dil çiftleri arasındaki kompleks ilişkileri öğrenerek daha etkili çeviriler gerçekleştirebilir. NMT, özellikle kelime düzeni, bağlam ve anlam gibi dilbilgisel öğeleri daha başarılı bir şekilde ele alınabilir, bu da çeviri kalitesini önemli ölçüde artırır. Bu nedenle, Derin Öğrenme Tabanlı Makine Çevirisi, dil çevirisindeki standartları yükseltme potansiyeli olan yenilikçi bir yaklaşım olarak öne çıkar.

Google Translate, Microsoft Translator ve OpenAI'nın GPT-3 gibi birçok platform, günümüzde makine çevirisinde önemli başarılar elde etmiş örneklerdir. Bu sistemler, büyük dil veri setleri ile eğitilen karmaşık sinir ağları üzerine kuruludur. GPT-3 gibi dil modeli, genel dil anlayışı ve çok dilli çeviri yetenekleriyle dikkat çeker. Ne kadar çok faydası olsa da zorlukları da vardır. Özellikle dildeki anlam değişiklikleri, kelime oyunları, kültürel nüanslar gibi karmaşık durumlar çeviri sistemlerini zorlayabilir. Bu nedenle, makine çevirisinin gelecekteki gelişimleri, dil anlama yeteneklerini artırarak ve daha fazla dil çiftini kapsayarak gerçekleştirilebilir.

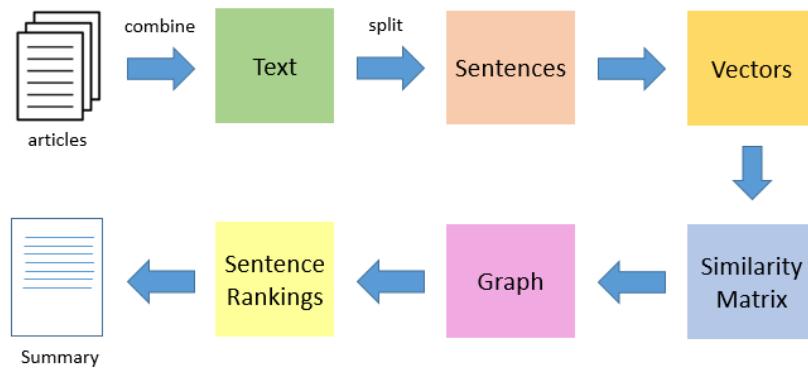
Doğal Dil İşleme, çok dilli metin sınıflandırmasında hayatı bir rol oynar. NLP'nin gelişmiş algoritmalar ve teknikler kullanarak farklı dillerde yazılmış metinlerden anlamlı içgörüler çıkarma yeteneği, metin sınıflandırma algoritmalarının çok dilli

metinlerin konularını, temalarını ve duygusal tonunu daha doğru bir şekilde tanımlamasına olanak tanır. Örneğin, NLP, belirli anahtar kelimeleri tespit ederek veya belirli ifadeleri ayırt ederek metin sınıflandırma algoritmalarına rehberlik edebilir. Bu durum, "otomatik öğrenme" veya "yapay zeka" gibi terimlerin farklı dillerdeki metinlerdeki bağlamını anlamak ve sınıflandırmak için oldukça değerlidir. Ayrıca, NLP'nin entite tanıma yetenekleri, farklı dillerdeki metinlerdeki önemli varlıklarını belirleyerek bilgi çıkarmada kritik bir rol oynar. Bu, çok dilli içeriklerdeki öne çıkan kişileri, kurumları veya yerleri tanımlamak için kullanılır ve metin sınıflandırma algoritmalarının belirli varlıkları içeren metinleri daha etkili bir şekilde sınıflandırmamasına yardımcı olur. Duygu analizi açısından NLP, çok dilli metinlerin duyarlığını anlama ve sınıflandırma konusunda önemli bir araçtır. Farklı dillerde yazılan metinlerde duygusal tonların ve ifadelerin karmaşaklılığı göz önüne alındığında, NLP'nin metin sınıflandırma algoritmalarına duygusal bağlamı anlama yeteneği, çok dilli iletişimde daha etkili ve duyarlı bir anlayış sağlar.

Otomatik Metin Öztleme Ve Öne Çıkarılmış İçerik Oluşturma

Otomatik metin öztleme, doğal dil işleme alanında önemli bir uygulama olarak öne çıkmaktadır. Metin öztleme çalışmaları genellikle ana ve ara uygulamalar olmak üzere iki ana grupta incelenmektedir. Ana uygulamalar arasında bilgisayarla çeviri, otomatik öztleme, bilgi çıkarımı ve bilginin yeniden elde edilmesi gibi kendi başına birer uygulama oluşturan örnekler bulunmaktadır. Ara uygulamalar ise, tümceyi öğelerine ayırma, çözümleme, biçimbilimsel analiz, sözcük anlamını belirginleştirme gibi ana uygulamalar için gerekli işlemleri gerçekleştirmektedir. Internetin metin materyali bolluğu, büyük metinleri manuel olarak öztlemenin zorluğunu ortaya çıkarmaktadır. Ancak, otomatik metin öztleme, bu zorluğu aşmak için geliştirilmiş bir konsepttir. Bu yöntem, çeşitli kaynaklardan elde edilen bilgileri belirli bir amaca yönelik olarak kısaltma işlemini ifade eder.

Otomatik metin öztleme sistemleri, büyük miktardaki veriyi işlemekte ve kullanıcılara önemli bilgileri hızlı bir şekilde sunmaktadır. Metin öztleme işlemlerinde yaygın olarak kullanılan yöntemlerden biri cümle puanlama yöntemidir. Bu yöntem, belirli kriterler altında metinleri değerlendirir ve bütünden parçaya giderek öztleme gerçekleştirir. Otomatik metin öztleme, özetleyici ve çıkarımsal olmak üzere iki türde gerçekleşir. Çıkarımsal metin öztleme, okunabilirliği yüksek bir yaklaşım sunarken, özetleyici yöntemler geniş kapsamlı kelime bilgisini kullanarak metni kısaltma eğilimindedir. Çıkarımsal metin öztlemenin altında ise kelime frekans değerine göre öztleme ve TextRank algoritmasına göre öztleme gibi alt yöntemler bulunmaktadır. Bu yöntemler, metin içindeki önemli kelimeleri belirleyerek veya belirli algoritmalar kullanarak öztleme sürecini gerçekleştirir. Bu noktada, metin içindeki anahtar kelimelerin kontrolü, cümle konumları, frekans analizi ve anlamı pekiştiren sözcüklerin dikkate alınması gibi kriterler de önemli bir rol oynamaktadır.



Zorluklar ve Etik Sorunlar

NLP Uygulamalarının Karşılaştığı Teknik Zorluklar

Doğal Dil İşleme uygulamalarının karşılaştığı zorluklar geniş bir yelpazeye yayılmaktadır. Özellikle, büyük dil modelleri gibi gelişmiş teknolojilerin çıktılarındaki potansiyel kusurların ve yaniltıcı bilgilerin toplum için olumsuz etkileri vardır. Araştırmacılar, bu modellerin içerik oluşturucuların kötü niyetleri olmaksızın dahi zararlı sonuçlara yol açabileceğini belirtmektedir. Bu zararlar, bir temsilcinin yanlış kanıları öğrenmesi veya model eğitim sürecindeki hatalar nedeniyle kazara yayılabilir. Özellikle, dil araçlarının ayrımcılığı kolayca süredürebilmesi, toplum üzerinde yüksek bir zarar riski taşıdığını işaret eder. Dil modelleri, mevcut sorunlu normlara meydan okumayı zorlaştırbılır ve toplumu değer bağımlılığı yaratacak şekilde etkileyebilir. Ayrıca, dil modelleri, özellikle siyasi amaçlar taşıyan bireyler tarafından şiddet teşvik etme veya toplumsal zarar biçimlerini sağlama potansiyeline sahiptir.

Challenges

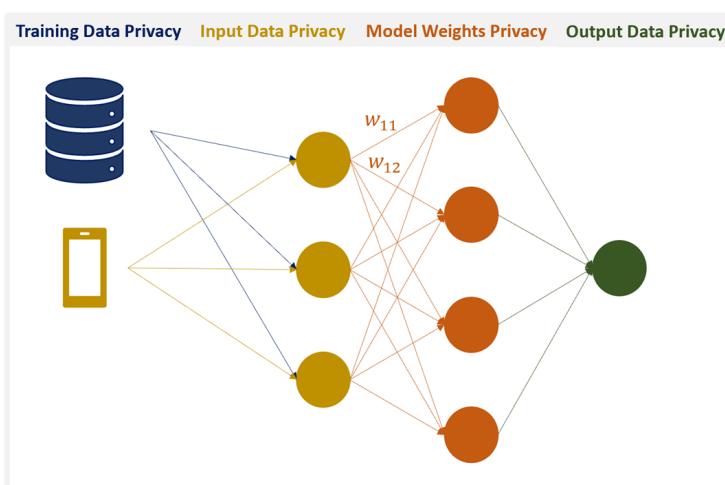


Bu tehlikelerin yanı sıra, NLP uygulamaları genel olarak bazı temel zorluklarla karşılaşır. Dilin karmaşıklığı, çoklu anlamı ve mizahi ifadeler gibi anlam katmanları,

doğru anlamayı zorlaştırbilir. Veri kalitesi eksikliği ve sınıflar arasındaki dengesizlik, etiketli veri toplamanın zorluklarını ortaya koyar. Çok dilli modellerin performansı ve kültürel farklılıklar, genellemeyi zorlaştırbilir. Ayrıca, NLP uygulamalarının eğitim verilerindeki önyargılar ve duygusal analizi gibi hassas konulardaki zorluklar, güvenilirlik sorunlarına yol açabilir. Hesaplama gücü ve altyapı zorlukları, özellikle büyük dil modellerinin eğitiminde enerji tüketimi ve çevresel etkileri vurgular. Türkçe'nin özgün yapısı ve dilin sürekli evrimi, Türkçe odaklı NLP uygulamalarında özel zorluklar yaratır. Bu zorluklara rağmen, NLP alanındaki araştırmalar ve teknolojik gelişmeler devam etmekte olup, gelecekte yapılacak çalışmalarında NLP önyargı analizi gibi konular önemli bir odak noktası olacaktır. Unutulmamalıdır ki NLP modellerinin beslendiği kaynakların dil olması ve bu kaynakların insanlar tarafından oluşturulduğu gerçeği, bu teknolojilerin etik ve güvenilirlik açısından sürekli gözden geçirilmesini gerektirmektedir.

Veri Gizliliği Ve Etik Konuları

Doğal Dil İşlemenin etik konuları, teknolojinin gelişmesiyle birlikte önem arz eden bir dizi meseleyi içermektedir. Bu etik konular, NLP uygulamalarının toplum üzerindeki etkilerini, kullanıcı gizliliğini ve güvenliğini, önyargıyı, şeffaflığı ve daha pek çok unsuru kapsamaktadır. Öncelikle, NLP modellerinin eğitiminde kullanılan veri setlerindeki önyargı, toplumdaki mevcut eşitsizlikleri yansıtarak ve pekiştirerek önemli sorumlara yol açabilir. Dil, cinsiyet, etnik köken gibi faktörlere duyarlılık kazanan modeller, yanlış kararlar alabilir ve toplumsal önyargıları güçlendirebilir. Kullanıcı gizliliği ve veri güvenliği endişeleri de NLP'nin etik konuları arasında öne çıkar. Bu uygulamalar genellikle büyük metin verilerini işlediğinden, bu verilerin içinde yer alan hassas bilgilerin korunması büyük bir sorumluluktur. Veri şifreleme, yetkilendirme ve güvenlik önlemleri, bu konudaki sorumlulukları yerine getirmede kritik bir rol oynar.



NLP'nin şeffaflık ve açıklanabilirlik konularındaki zorlukları da göz önüne alınmalıdır. Gelişmiş modellerin karmaşıklığı, iç işleyişlerini anlamayı zorlaştırbılır, bu da kullanıcılar ve geliştiriciler için güven sorunlarına neden olabilir. Ayrıca, NLP uygulamalarının kullanım alanlarındaki etik sorumluluklar da önemlidir. Örneğin, duygusal analizi uygulamalarının doğru ve adil bir şekilde insanların duygusal durumlarını değerlendirmesi gereklidir. Yanlış anlamalar veya hatalı çıkarımlar, toplum üzerinde olumsuz etkiler yaratabilir. Bu etik konulara yönelik farkındalık ve çözüm odaklı yaklaşımalar, NLP teknolojisinin etik kullanımını teşvik eder ve toplumsal güveni artırabilir. Etik standartlara uygun, şeffaf ve kullanıcı gizliliğine saygılı NLP uygulamaları, teknolojinin daha sürdürülebilir bir evrim geçirmesine olanak sağlar.

KAYNAKLAR

- <https://medium.com/@seval318/do%C4%9Fal-dil-i%C5%9Flemenin-nlp-temel-i%C5%9F%C87lkeleri-ve-i%C5%9Fleyi%C5%9Fi-176c0ff9ddde>
- <https://www.karel.com.tr/blog/dogal-dil-isleme-nlp-natural-language-processing-nedir>
- <https://wozox.com/tr-TR/Blog/nlp/nlp-kuruculari-ve-tarihcesi/>
- <https://dergipark.org.tr/tr/download/article-file/727452>
- <https://tr.shaip.com/blog/how-voice-assistant-understand-what-you-are-saying/>
- <https://uzmanposta.com/blog/nlp/>
- <https://ts2.space/tr/nlp-ve-cok-dillilik-kapsamli-bir-kilavuz/#qsc.tab=0>
- <https://acikerisim.sakarya.edu.tr/xmlui/bitstream/handle/20.500.12619/97078/T09372.pdf?sequence=1&isAllowed=y>
- <https://www.geeksforgeeks.org/ethical-considerations-in-natural-language-processing-bias-fairness-and-privacy/>