METİN MADENCİLİĞİYLE KANAL İSTANBUL HAKKINDA DUYGU ANALİZİ ELDE EDİLMESİ

Hamdi Uraz Alkış-161307051

Bilişim Sistemleri Mühendisliği Bölümü

Kocaeli Üniversitesi

alkisuraz@gmail.com

Özet

*Bu projede,metin madenciği gerçekleştirmek adına,son zamanlarda çok fazla polemik yaratmış kanal İstanbul projesi hakkında,duygu analizi gerçekleştirilmiştir..*

# Giriş

Duygu analiziyle birlikte,insanların kanal İstanbul hakkındaki görüşlerini,olumlu olumsuz nötr olarak açıklayacabileceğiz.Projede,geleneksel makine öğrenimi olarak naive bayes,derin öğrenme yöntemleri olarak cnn kullanılmıştı***r.***

# Kullanılan Algoritmalar

**2.1.Naive Bayes**

Naive Bayes Sınıflayıcı ile Bayes Teoremi hesaplarında dikkat edilmesi gereken en önemli fark sınıflayıcıların olasılık değerinden ziyade hedef sınıfı bulmaya odaklanmasıdır. Bu yüzden paydada bulunan değer, tüm hedef sınıflara ait olasılık hesaplarında ortak olduğundan ihmal edilebilir.

Bu yüzden Bayes Teoreminde bulunan sınıf olasılıkları toplamı 1 olmak zorunda iken Naive Bayes Sınıflayıcı ile bulunan değerlerin toplamı,ihmal edilen paydadan dolayı 1 olamaz. Buna göre Bayes Teoremi ile bir durumun olası sonuçlarının olasılıkları bulunurken Naive Bayes ile normalizasyonsuz olasılıklar üzerinden sınıflandırma yapılabilir

***Gaussian Naive Bayes***: Eğer özelliklerimiz sürekli değer(continuous value) ise bu değerlerin bir gauss dağılımıveya diğer bir değişle normal dağılımdan örneklendiğini varsayıyoruz.

***Multinominal Naive Bayes*:**Örneğin bir e-ticaret siteniz varsa pek çok kategoriniz olur. Çok sınıflı kategorileri -spor, kozmetik, giyim, petshop- sınıflandırmak için kullanılır.

**Bernoulli Naive Bayes:**Multinominal Naive Bayes’e benzer şekilde sınıflandırma yapar. Ancak tahminler sadece boolean(ikili) şekildedir. Evet/Hayır, Spam/Spam Değil, 1/0, Var/Yok gibi.

**2.2.Decision Tree**

Karar ağacı algoritması, veri madenciliği sınıflandırma algoritmalarından biridir.Önceden tanımlanmış bir hedef değişkene sahiplerdir. Yapıları itibariyle en tepeden en aşağı inen bir strateji sunmaktadırlar.

Bir karar ağacı, çok sayıda kayıt içeren bir veri kümesini, bir dizi karar kuralları uygulayarak daha küçük kümelere bölmek için kullanılan bir yapıdır. Yani basit karar verme adımları uygulanarak, büyük miktarlardaki kayıtları, çok küçük kayıt gruplarına bölerek kullanılan bir yapıdır.

**2.3.** **Convolutional neural network**

Convolutional neural network, yapay sinir ağlarının özel bir türüdür.Özellikle resim tanıma içinkullanılan çok etkili bir mekanizmadır. Klasik neural networks yapısına benzer şekilde,nöronlardan oluşur ve öğrenmek için ağırlıklar ve bias değerine sahiptir. Her nöron girişleri alır, birleştirir ve genellikle non-lineer bir fonksiyon ile çıkış üretir. CNN uygulamaları girişleri görüntü olarak varsayar ve özellikleri mimariye kodlamamızı sağlar.

1-Convolutional Layer

Bu katman CNN’nin ana yapı taşıdır. Resmin özelliklerini algılamaktan sorumludur. Bu katman, görüntüdeki düşük ve yüksek seviyeli özellikleri çıkarmak için resme bazı fitreler uygular. Örneğin, bu filtre kenarları algılayacak bir filtre olabilir. Bu filtreler genellikle çok boyutludur ve piksel değerleri içerirler.(5x5x3) 5 matrisin yükseklik ve genişliğini, 3 matrisin derinliğini temsil eder.

2.Pooling Layer

Bu katman, CNN’deki ardışık convolutional katmanları arasına sıklıkla eklenen bir katmandır. Bu katmanın görevi, gösterimin kayma boyutunu ve ağ içindeki parametreleri ve hesaplama sayısını azaltmak içindir. Bu sayede ağdaki uyumsuzluk kontrol edilmiş olur. Birçok Pooling işlemleri vardır, fakat en popüleri max pooling’dir. Yine aynı prensipte çalışan average pooling, ve L2-norm pooling algoritmalarıda vardır.

3-Flattening Layer

Bu katmanın görevi basitçe, son ve en önemli katman olan Fully Connected Layer’ın girişindeki verileri hazırlamaktır. Genel olarak, sinir ağları, giriş verilerini tek boyutlu bir diziden alır. Bu sinir ağındaki veriler ise Convolutional ve Pooling katmanından gelen matrixlerin tek boyutlu diziye çevrilmiş halidir.

4.Full-Connected Layer

Bu katman Convolutional Network’un son ve en önemli katmanıdır. Verileri Flattening işleminden alır ve Sinirağı yoluyla öğrenme işlemini geçekleştirir.

**2.4 Multi Layer Perceptron**

Perceptron Model bir Yapay Sinir Ağları modelidir ve bugünkü Yapay Sinir Ağları için önemli bir temel oluşturmaktadır. Supervised (denetimli) bir training (öğrenme) algoritmasıdır. Yani ağa hem giriş hem de çıkış kümesi verili ve öğrenme beklenir. Perceptron Modeli’nde en önemli faktör eşik değeridir. Bu değer kullanılarak güzel bir sınıflandırma yapılabilmektedir. Saptanacak olan eşik değeri probleme göre belirlenebilir. Bu modelde iterasyon sayısı artırılarak öğrenme derecesi artırılabilir. Tek Katmanlı Algılayıcı’lar aşağıdaki şekilde modellenmiştir.

# Deney Kurulumu

Veri kümesi,twitter üzerinden #kanalistanbul,#yakanalyaistanbul,#kanalistanbulaevet adında üç farklı hashtag ile yaklaşık dokuzbin adet veri toplanmıştır.Verilerin yaklaşık 4000bin tanesi nötr,2000bin tanesi pozitif,geri kalan bölüm ise negatiftir.Tarih aralıkları, genel olarak 20 aralık 2020-1 mart 2020 arasındadır.Python programlama diliyle,selenium kullanılarak,googlechrome api üzerinden veriler çekilmiştir.Çekilen tweetleri temizlemek için,re kütüphanesinden yararlandım.Metinleri,olumlu-olumsuz olarak ayırmayı ise texblob kütüphanesi ile gerçekleştirdim.Elde ettiğimiz metinleri,ayrı olarak numerik hale getirmemiz gerektiğinden,bunun için sklearnden TfidfVectorize methodunu kullandım.Verileri de %80test %20 train olarak böldüm.Sonra verileri one hot encoding yapıp reshape ile tekrar biçimlendirdim.Yalnızca a bölümünde accuracy değeri olarak naive bayes ve decision tree ile 0.4286 değeri aldım.Diğer sunumda yaptığım bütün alogritmalarda hata aldım ve çözemedim.

.