**EGE ÜNİVERSİTESİ**

**MÜHENDİSLİK FAKÜLTESİ**

**BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ**

**YAPAY ZEKA YÖNTEMLERİ ve UYGULAMALARI**

**2019-2020 BAHAR YARIYILI**

**PROJE 2 – ARA RAPOR**

**YAPAY ZEKA / MAKİNE ÖĞRENMESİ / DERİN ÖĞRENME / ENİYİLEME PROJESİ**

Fatih Cüre

05170000113

1. **PROJE**
   1. **Konusu**

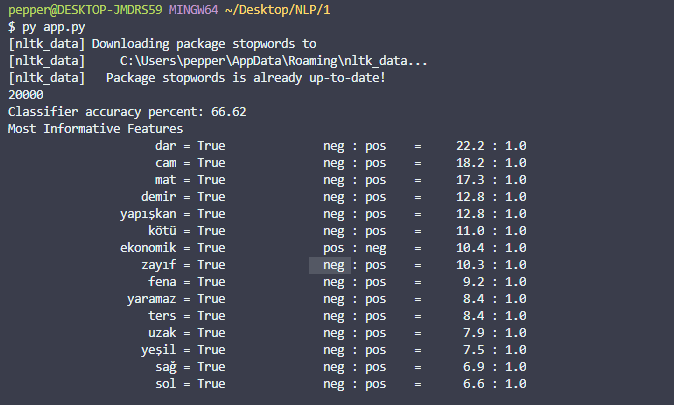
Projede Doğal Dil İşleme ile ürün görüşünün olumluluk durumunun araştırılması konu olarak ele alındı.

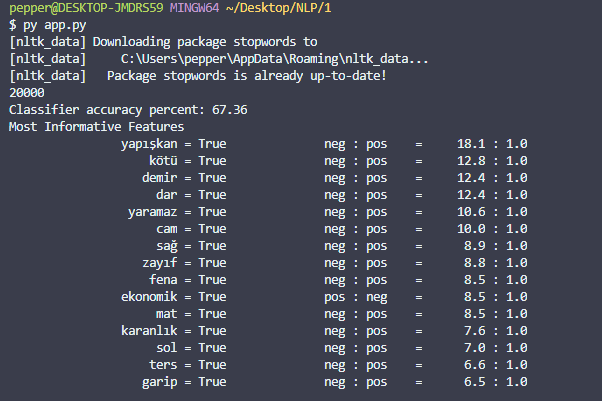
* 1. **Veri seti**

Veri seti olarak internetten ve kendimin elde ettiği ürün yorumları veri seti kullanıldı. İnternetten elde edilen veri setinde de yaklaşım olarak ürün yorumlarına ek olarak verilen puan/yıldız değerleri baz alınarak pozitif/negatif olarak etiketlenmesi ile elde edildiği biliniyor. Kendi ürettiğim veri seti de aynı yaklaşımı kullanıyor.

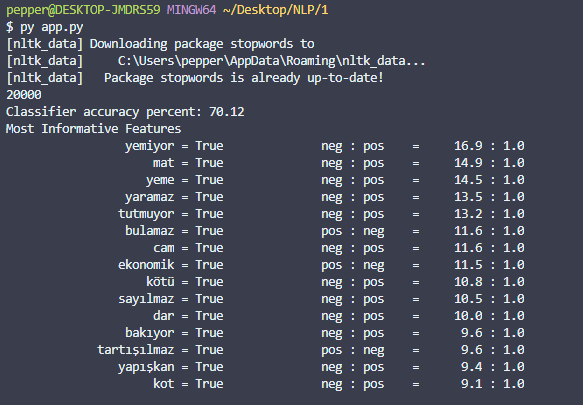
İnternetten elde edilen veri seti: <https://github.com/fthbrmnby/turkish-text-data>

Kendi elde ettiğim veri seti: \Kaynak\_Kod\P2\dataset\kitapyurdu dizini altında bulunuyor.

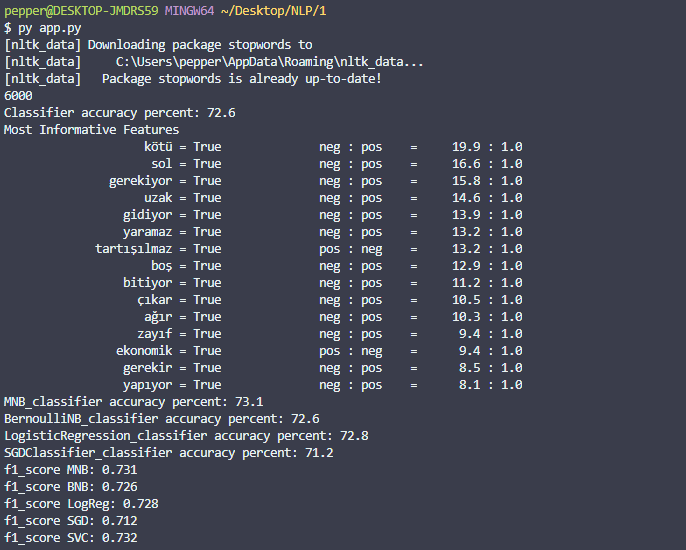
1. **DENEYSEL ÇALIŞMALAR**
   1. Sadece sıfatların ***kelime kökleri*** değerlendirilerek NaiveBayes yöntemi kullanılarak elde edilen doğruluk oranı**: **
   2. “değil” kelimesinin cümlenin gidişatını tersine çevirdiğini hesaba katarak ve ***sadece kelime kökleriyle*** ilgilenen çalışmanın doğrulukoranı:



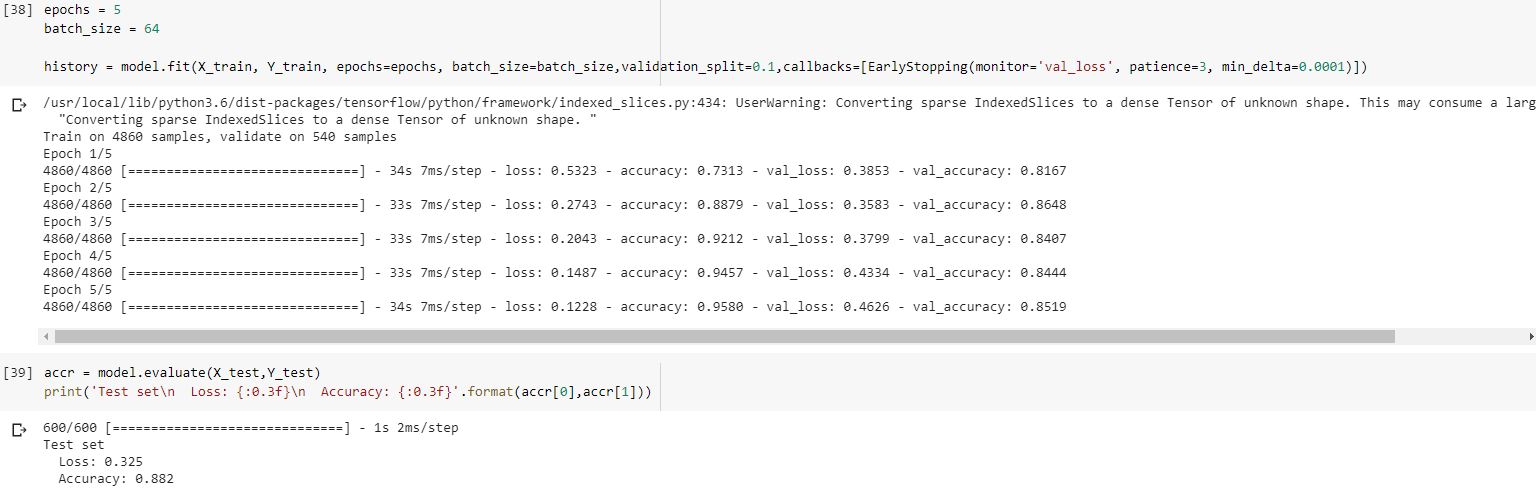
* 1. Fiillerin de kelime sözlüğüne dahil edilerek ve ***kelimelerin olduğu gibi*** ele alındığı çalışmanın doğruluk oranı:



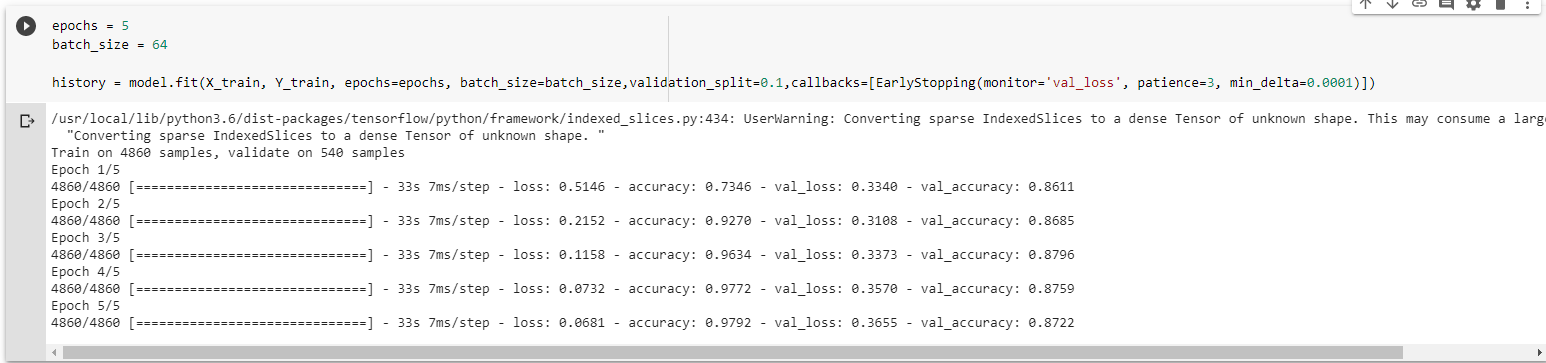
* 1. Farklı sınıflandırıcılar kullanılarak elde edilen doğruluk oranları ve f1 değerleri:



* 1. LSTM kullanılarak elde edilen doğruluk değerleri: (Epoch:5)



* 1. LSTM’de Türkçe gereksiz kelimelerin kaldırılması ile elde edilen değerler:



LSTM denemeleri makinemde sıkıntı yarattığı için Google Colab üzerinden yapıldı. Gerekli olan notebook dosyası Kaynak\_Kod\ dizini altında bulunan LSTM.ipynb dosyasıdır.

Elde edilen değerlerin bir çoğunu kapsayan kayıt edilmiş veriler “Kaynak\_Kod\P2\pickle” dizini altında kayıtlı halde bulunmakta.

Ek olarak kendi veri setimi oluşturabilmek için yazılan kod parçası “Kaynak\_Kod\scraper” dizini altında bulunmakta.

1. **PROJE TESLİMİNE KADAR YAPILMASI PLANLANAN ÇALIŞMALAR**
2. NaiveBayes yöntemi ile tahmin için K-Fold CV tekniğinin uygulanması.
3. LSTM için hiper parametre değerlerinin öğrenilmesi ve ayarlanması.
4. Benzer çalışmalarla karşılaştırmaların yapılması ve rapora eklenmesi
5. Önerilen yöntemin çalışma prensibinin şema haline getirilmesi ve rapora eklenmesi.
6. Zaman kalırsa “pickle” kullanılarak gerektiğinde hazır modelden kullanım sağlanabilmesi.