VERİ BİLİMİNE GİRİŞ

**PROJE ADI**

TÜRK LİRASI/YEN TAHMİNLEMESİ

**ÖĞRENCİ NO – AD SOYAD**

05170000019 – Ramazan ONBAŞI 05140000680 – Emin Soner TÜRK

**TESLİM TARİHİ:** 15.05.2018

**İçindekiler**

1.Arkaplan ve Motivasyon........................................................................3

2.Projenin Amaçları (problem tanımı):.....................................................3

3.Veri........................................................................................................3

4.Veri İşleme.............................................................................................4

5.Keşif Analizi……………………......................................................................4

6.Analiz Metadolojisi................................................................................5

7.Analiz Sonuçları……………........................................................................5

8.İş Bölümü...............................................................................................5

**1. Arkaplan ve Motivasyon**

Bu projeyi seçmemizdeki ana etken Türk lirasının dolar ve avro kadar Türkiye’de bilinirliği olmayan fakat yine de güçlü bir para birimine karşı ne gibi bir trend izlediğini görmek, bunun hangi faktörlerden nasıl etkilendiğini anlama isteğimizdir. Ayrıca grup üyelerinden birinin Japonya’ya ilgisinin olması spesifik olarak Japon para birimi olan yenin seçilmesinde de önemli bir etken oldu.

**2. Projenin Amaçları**

Projemizde başlıca amacı Türk lirasının yen karşısındaki değerinin, ekonomiyle ilgili çok önemli göstergeler olan Tüketici Fiyat Endeksi, TCMB’nin faiz kararları, cari hesap, dış ticaret hadleri vb. gibi etmenlerden nasıl etkilendiğine görmek ve etmenlerle birlikte Türk lirasının yen karşısındaki değişimin tahminlemektir.

Projemizin Türkiye’nin geçmişiyle ve dahi geleceğiyle ilgili ekonomik okumalar yapmak, hipotetik beklentilerle gerçekleşen ekonomik sonuçlar arasındaki farkı ortaya koyabilmek vb. gibi konularda insanlara kolaylık sağlayabileceği, bir nebze de olsa yol gösterici nitelikte olacağına kaniyiz.

**3. Veri**

Projemizdeki verileri envaiçeşit sitele gerek önceden hazırlanmış dosyalar vasıtasıyla aldık gerekse de bu tarz bir dosyanın var olmadığı durumlarda verdiğimiz kaynaklardaki verilerden kendimiz hazırladık.

Veri Kaynak Bağlantılarımız:

**Tüketici Fiyat Endeksi:** <https://biruni.tuik.gov.tr/medas/?kn=84&locale=tr>

**Faiz:** <http://www.tcmb.gov.tr/wps/wcm/connect/TR/TCMB+TR/Main+Menu/Temel+Faaliyetler/Para+Politikasi/Merkez+Bankasi+Faiz+Oranlari/faiz-oranlari>

**Cari Hesap:** <https://tr.investing.com/economic-calendar/turkish-current-account-(usd)-1240>

**Dış Ticaret hadleri:** <https://biruni.tuik.gov.tr/medas/?kn=62&locale=tr>

**Net Kamu Borcu Stoku**: <https://www.hazine.gov.tr/kamu-finansmani-istatistikleri>

**Gayrisafi Millî Hasıla:** <https://biruni.tuik.gov.tr/medas/?kn=88&locale=tr>

**İşsizlik**: <http://www.tuik.gov.tr/PreTablo.do?alt_id=1007>

**ABD Doları/Yen**: <https://tr.investing.com/currencies/usd-jpy-historical-data>

**4. Veri İşleme**

Verilerimizi alakalı dosyalardan doğru bir şekilde okutturduktan sonra bizim için gerekli olmayan sütunları attık. Sütun isimlerini projemize uygun bir şekilde isimlendirdik. İncelediğimiz tarih aralığının dışında olan değerleri attık, eksik olan değerleri attık. Aylık olmayan düzensiz verileri alakalı yöntemlerle düzenledik.

**5. Keşif Analizi**

Veri kümemizi keşfetmek için çizgi(line) grafiği ve dağılım(scatter) grafiği kullandık.

İlk olarak; TÜFE, gecelik borç verme faizi, gecelik borç alma faizi, cari hesap, dış ticaret hadleri, net kamu borcu stoku, gayri safi milli hasıla, işsizlik oranı, 100 yen’in dolar cinsinden karşılığını, 100 yen’in tl cinsinden karşılığı değerlerini ay bazında 2015 yılından 2018 yılına kadar çizgi grafik olarak çizdirdik.

İkinci olarak ise; TÜFE, gecelik borç verme faizi, gecelik borç alma faizi, cari hesap, dış ticaret hadleri, net kamu borcu stoku, gayri safi milli hasıla, işsizlik oranı, 100 yen’in dolar cinsinden karşılığı değerlerini; 100 yen’in tl cinsinden karşılı değerlerine göre scatter olarak çizdirme işlemi yaptık.

Üçüncü olarak ise; çıkan tahmin değerlerini belirlemek için, her bir sütun değerinin ayrı ayrı tahminlenip tl/yen oranıyla kıyaslanmasından çıkan sonuç grafiklerini çizdirerek tahminin doğruluğu hakkında bilgi sahibi olduk.

Son olarak; tl/(100)yen, yen/(100)dolar miktarlarının her bir sütun değerinde nasıl değiştiğini gözlemlemek için 3d scatter grafik çizdirip yorumlamalarda bulunduk.

Çıkan sonuçları .ipynb proje dosyasından görebilirsiniz.

**6. Analiz Metadolojisi**

Çok değişkenli lineer regresyon yöntemini kullandık. Bu yöntem kısaca ***y*i = http://www.stat.yale.edu/Courses/1997-98/101/beta.gif0 + http://www.stat.yale.edu/Courses/1997-98/101/beta.gif1*x*i1 + http://www.stat.yale.edu/Courses/1997-98/101/beta.gif2*x*i2 + ... http://www.stat.yale.edu/Courses/1997-98/101/beta.gifp*x*ip** formülüyle ifade edilebilir. Buradaki y değeri bizim bulmak istediğimiz bağımlı değişken oluyor (bizim ödevimiz spesifiğinde bu “TL/Yen” oluyor), x’ler ise bu y değerini tahminlemek için kullanılan bağımsız değişkenler oluyorlar. Bizim projemizde x’lere “İşsizlik Oranı”, “Tüketici Fiyat Endeksi” vs. örnek gösterilebilir.

**7. Analiz Sonuçları**

Tahminleme yapmak için, veri setimiz 1000’den az olduğundan dolayı, 10-kat çapraz doğrulama yöntemini (10-fold cross validation) uyguladık. Bu yöntemde her veri seti 10 parçaya bölünerek bu parçalar içinden 1 tanesini alınıp bu parça test için kullanır, diğer parçaları ise eğitim için. Bu işlem her bir parça için tekrar eder. En sonunda çıkan tüm sonuçların ortalamaları alınır. Bu yöntemin kullanılmasının nedeni aşırı uymadan (overfitting) kaçınmak, fonksiyonumuzun hiç karşılaşmadığı değerleri görünce de mantıklı sonuçlar verebilecek kadar genelleştirilmiş olduğundan emin olmaktır.

Analiz için kullandığımız metrik değerleri ve sonuçları:

R^2 (coefficient of determination) değerimiz = 0.8488137307619608

Mean Squared Error değerimiz = 0.05808120440910537

Explained Variance değerimiz = 0.8532441151780337

Mean Absolute Error değerimiz = 0.1990840733119526

Mean Squared Log Error değerimiz = 0.007666929729844276

Median Absolute Error değerimiz = 0.17939451998505107

**8.İş Bölümü**

İlk olarak 1 hafta aramızda beyin fırtınası yapıp, konu araştırması ve belirlemesi yaptık bu süreçte akademisyenlerimizin de yanlarına giderek konuyla alakalı tavsiye ve görüşlerini dinledik.

Ardından uzun bir veri setlerini arama ve düzenleyip birleştirme faslı geldi. Fazla sayıda veri setimiz olduğu için veri setlerini düzenleme işini yarı yarıya paylaştık.

Ardından 10-kat çapraz doğrulama yönteminin teorik ve araştırıldı ve müşterek şekilde implimente edildi.

Raporlama kısmında da yine aynı şekilde birbirimiz gözetiminde müşterek bir şekilde yapıldı.