**Theft and Loss Nitelikli Kaçak Tahminleme Projesi**

Projenin amacı tesisata bağlı, nitelikli kaçak yapan müşterileri tespit etmektir. Ekte excel olarak iletilen 15K’lık train ve 5K’lık test seti bulunmaktadır. Train setindeki NK\_FLAG alanı target değişken olup, bir classification modeli uygulanması beklenmektedir. NK\_Flag =1 olması kaçağın olduğu, 0 olması kaçağın olmadığını ifade etmektedir. Verileri öğrenip, istenilen data preparation kısmı tamamlanıp, SAS, Python, R vb. bir platformda modelin tamamlanıp, tahminlerin iletilmesi gereklidir. Tüm kodlar ve tahmin çıktıları iletilmelidir. Tahmin sonuçlarınızı **sadece** **INDEX ve NK\_FLAG** kolonları olacak şekilde yeni bir Excel dosyası ile paylaşmalısınız.

Bu problemde maksimize edilmesi istenen değer **F1 Skorudur**. Kurduğunuz modelde bu skoru maksimize etmenizi bekliyoruz.

Ekstra bir kaynak, internetten veri zenginleştirmenize gerek yoktur, yalnızca iletilen verileri kullanmanız yeterlidir. Gereken analizleri yapıp, variable transformation, data handling metodlarını kullanabilirsiniz. Kod içerisinde yorumlarınızda hangi metodları uygulayıp, ne gibi sonuçlar aldığınızı paylaşabilirsiniz.

Değişkenlerin açıklamaları ekteki Excel’de ve aşağıdaki tabloda mevcuttur.

|  |  |  |
| --- | --- | --- |
| **Değişken İsmi** | **Değişken Açıklaması** | **Değişken Tipi** |
| NK\_FLAG | Tesisat geçmişinde nitelikli kaçak yaptıysa 1, değilse 0 yazılır | Target |
| TESISAT\_TIPI | "Mesken" veya "Ticarethane-Sanayi" değerlerini alabilir | Input |
| MAHALLE\_RISK\_SKORU | Tesisatın bulunduğu mahallede tespit edilen nitelikli kaçak sayısının, mahalledeki tesisat sayısına oranı | Input |
| SOKAK\_RISK\_SKORU | Tesisatın bulunduğu sokakta tespit edilen nitelikli kaçak sayısının, sokaktaki tesisat sayısına oranı | Input |
| SOB\_RISK\_SKORU | Tesisatın bulunduğu SOB'taki (sayaç okuma birimi=mahalle ve ilçe arası boyuttaki adres birimi) tespit edilen nitelikli kaçak sayısının, SOB'taki tesisat sayısına oranı | Input |
| TUKETIM\_M1-TUKETIM\_M24 | Referans tarihinden geçmişe doğru aylık bazda kwh cinsinden tüketim miktarları (24 ay-24 değişken) | Input |
| DEMAND\_M1-DEMAND\_M12 | Referans tarihinden geçmişe doğru aylık bazda demand (maksimum elektrik çekişi) miktarları (12 ay-12 değişken) | Input |
| SAYAC\_BASLANGIC\_TARIHI | Date, sayacın başlangıç tarihi | Input |
| SAYAC\_BITIS\_TARIHI | Date, sayacın kapatılma tarih | Input |
| SAYAC\_MARKA | Char alan, sayacın markası | Input |
| SAYAC\_MODEL | Char alan, sayacın modeli | Input |
| SAYAC\_MALZEME\_ID | Char id alan, sayacın malzeme numarası | Input |
| SAYAC\_TIPI | Char alan, sayacın tipi | Input |
| SAYAC\_OLCUM\_TURU | Char alan, sayacın ölçüm türü | Input |
| SAYAC\_FAZ\_N | Char alan, sayacın faz bilgisi | Input |
| SAYAC\_TAKILMA\_TARIHI | Date,sayacın takılma tarihi | Input |
| SAYAC\_YAPIM\_YILI | Date, sayacın yapım yılı | Input |

**Teslim edilmesi gereken dosyalar (. zip)**

* Tüm geliştirme/model kodları ( Python, R…)
* Test tahmin sonuçları ( Excel )
* Power Point çözüm sunum dosyası
  + Problem tanımı yaklaşım
  + Ön işleme, Model kurma, Önemli değişkenler analizi…
  + Model sonuçları değerlendirme

Başarılar.