

Merhaba,

TÜPRAŞ ve ODTÜ Verimlilik Topluluğu olarak gerçekleştirdiğimiz Veri Bilimi yarışmamıza seçildiğiniz için mutluluk duyuyoruz.

Dünyamız hızla değişiyor ve enerji sektörü de bu değişimden payını alıyor. İşte tam da bu noktada, TÜPRAŞ Veri Analitiği ekibi olarak, geleceğimizi şekillendirecek veri bilimi ve yapay zeka teknolojilerini kullanarak enerji sektörünü bilim ve teknoloji ile dönüştürmek için çalışmaktayız.

Lider şirketlerin genç beyinlerle var olduğuna inanıyoruz. Sizlerin inovatif fikirlerini ve analitik becerilerinizi göreceğimiz bu yarışmada, birlikte gerçek bir problemin üzerinde enerjimiz hiç bitmeden çalışacağız.

Problem

Yarışmamız boyunca TÜPRAŞ'ta bir Veri Bilimci rolünde olacaksınız. TÜPRAŞ Veri Analitiği ekibinin bir üyesi olarak İngiltere'de bulunan iştirakimiz Tüpraş Trading Ltd.'nin operasyonlarında kritik öneme sahip olacak bir proje tamamlamanız isteniyor. Traderlarımız sizlerden bir hafta sonra benzin ürününde oluşacak talebi tahminleyen bir model üretmenizi bekliyor. Modellemeyi yaparken kullanılacak veriler ticaret uzmanlarının alan bilgisi dahilinde filtrelenerek sizinle paylaşıldı. Oluşturulacak olan model talebin eksiksiz sağlanması, ticari karlılık, yerinde/zamanında lojistik, stok yönetimi ve sürdürülebilirlik gibi birçok konuda kritik değer yaratacak.

Aşağıda isimleri verilen dosyalar ve açıklamalar doğrultusunda çalışmalarınıza başlayabilirsiniz.

tupras_X_train: Model eğitim veri seti. Problemin çözümü için ticaret uzmanları tarafından seçilen temel verileri içerir.

tupras_y_train: Model hedef veri seti. Kurduğunuz modeller ile tahminlemeye çalışacağınız hedefi (bir hafta sonraki benzin talebi) içerir.

Veri Seti Açıklaması

Veri setinin içerdiği değişkenlere ilişkin aşağıdaki tabloda gerekli bilgiler bulunmaktadır.

Değişken	Açıklama	Dosya
period	Tarih	X_train, X_test
gasoline_imports	Benzin ürünü ithalat miktarı	X_train, X_test
gasoline_exports	Benzin ürünü ihracat miktarı	X_train, X_test
crude_imports	Ham petrol ithalat miktarı	X_train, X_test
crude_exports	Ham petrol ihracat miktarı	X_train, X_test
gasoline_stocks	Depolarda bulunan benzin miktarı	X_train, X_test
crude_stocks	Depolarda bulunan ham petrol miktarı	X_train, X_test
distillate_fuel_stocks	Depolarda bulunan dizel türü yakıt miktarı	X_train, X_test
jet_stocks	Depolarda bulunan jet yakıtı miktarı	X_train, X_test
propane_stocks	Depolarda bulunan propan miktarı	X_train, X_test
gasoline_net_input	Benzin üretimi için kullanılan girdi miktarı	X_train, X_test
gasoline_output	Rafineriler tarafından üretilen benzin miktarı	X_train, X_test
distillate_fuel_output	Rafineriler tarafından üretilen dizel miktarı	X_train, X_test
jet_output	Rafineriler tarafından üretilen jet yakıtı miktarı	X_train, X_test
propane_output	Rafineriler tarafından üretilen propan miktarı	X_train, X_test
residual_fuel_output	Rafineriler tarafından üretilen diğer yakıt miktarı	X_train, X_test
conventional_gasoline_spot_price	Benzin piyasa değeri/fiyatı	X_train, X_test
crude_brent_spot_price	Ham petrol piyasa değeri/fiyatı	X_train, X_test
gasoline_future_price1	Vadeli işlemlerde 1 ay ötelemeli benzin fiyatı	X_train, X_test
gasoline_future_price2	Vadeli işlemlerde 2 ay ötelemeli benzin fiyatı	X_train, X_test
gasoline_future_price3	Vadeli işlemlerde 3 ay ötelemeli benzin fiyatı	X_train, X_test
gasoline_future_price4	Vadeli işlemlerde 4 ay ötelemeli benzin fiyatı	X_train, X_test
gasoline_demand	Benzin talep miktarı	X_train, X_test
target	Hedef değişken, bir hafta sonraki benzin talebi	y_train

Yarışma Kuralları

Yarışma boyunca takip edilmesi gereken bazı kurallar ve performansınızı etkileyebilecek önerilerimizi aşağıda sıralıyoruz.

1. Datathon’da takım oyunu bizim için çok önemli. İş bölümü ve ekip içi iletişim başarınızda önemli bir rol oynayacak.
2. Modellerinizi eğitmek için kullanacağınız veri setleri aşağıdaki saatlerde paylaşılacaktır:
 - a. tupras_X_train: **13.00**
 - b. tupras_y_train: **13.00**
3. Araştırma yapmak amacıyla tüm kaynakların kullanımı serbesttir. Fakat gruplar arası bilgi alışverişine izin verilmeyecek.
4. Analiz, modelleme, görselleştirme gibi konularda her türlü yazılım, yazılım dili, araç, kütüphane vs. kullanımı serbesttir.
5. Çözüm yöntemi konusunda tek kısıt yaratıcılığınız olacak.
6. Sağlanan veriler üzerinde her türlü manipülasyon yönteminin uygulanması serbesttir. Yeni veriler üretmeniz veya var olan veriyi manipüle etmeniz durumunda üretim metotlarınızı hızlıca yeniden kullanılabilir biçimde tasarladığınız ilerleyen aşamalarda size kolaylık sağlayacaktır.
7. Model başarınızı ölçmek amacıyla paylaşılacak olan veri setinin ismi ve paylaşım saati aşağıda verilmiştir.
 - a. tupras_X_test: **16.15**
8. tupras_X_test verisini modelinize hazır hale getirmeniz ve ilgili veriye karşılık gelecek tahminleri üretmeniz gerekmektedir. Tahminlerin paylaşımı ile ilgili kurallar aşağıda verilmiştir.
 - a. Tahminlerinizi saat **16.30**’a kadar baran.ozden@tupras.com.tr adresine e-posta yoluyla iletmeniz gerekmektedir.
 - b. Geç gönderilen sonuçlar hesaba katılmayacak ve katılımcılar ilgili bölümden **puan alamayacaklardır**.
 - c. Tahminleri **grupismi_predictions.csv** adlı bir dosyada göndermeniz gerekmektedir.
 - d. Dosyanın içeriği, aşağıdaki görselde de belirtildiği üzere, indekste period isimli sütun (tupras_X_test dosyasındaki period sütunuyla **tam** eşleşmelidir) ve “target” isimli sütundan (tahmin değerleriniz) oluşmalıdır. İsterlere uymayan gruplardan ilgili bölümde puan kırılacak veya uyumsuzluk seviyesi göz önünde bulundurularak **puan verilmeyecektir**.

	target
period	
2020-03-27	1000
2020-04-03	1001
2020-04-10	1002
2020-04-17	1003
2020-04-24	1004
...	...

9. Tahminlerinizin yanı sıra Datathon boyunca yazdığınız kodları veya kullandığınız programa özgün dosyaları, işlemleri yaptığınızı kanıtlayabilecek şekilde paylaşmanız gerekmektedir. Paylaşacağınız tüm dosyaları **grupsimi** olarak isimlendirdiğiniz klasörün içerisinde **16.30'**a kadar baran.ozden@tupras.com.tr adresine e-posta yoluyla iletmeniz gerekmektedir. Klasörün içeriği **grupsimi_predictions.csv**, işlemlerinizi/kodlarınızı içeren dosyalar ve varsa ekstra notlarınızdan oluşmalıdır.
10. Mentor desteği **13.00'**da başlayacak ve **16.00'**da sona erecektir. Mentorlar yarışma işleyişini etkilemeyecek ve gruplar arasında adaletsizlik oluşmasına izin vermeyecek biçimde destek sağlayacaktır. Mentor desteği her bir saat diliminde en fazla 15 dakika olacak şekilde tüm gruplara sağlanacaktır.
11. Problemi çözme sürecinde en çok dikkat etmeniz gereken konu **veri sızıntısı (data leakage)** olacak. Sızıntının tespit edilmesi halinde ilgili grup **tüm teknik bölümlerden sıfır puan alacaktır**. Konuyla ilgili bilgi için [https://en.wikipedia.org/wiki/Leakage_\(machine_learning\)](https://en.wikipedia.org/wiki/Leakage_(machine_learning)) adresini ziyaret edebilirsiniz.
12. Puan dağılımları için aşağıdaki tablo dikkate alınacaktır.

Puan Başlığı	Kriterler	Puan
Teknik Yaklaşım	Problemi anlama, probleme uygun metot seçimi, açıklanabilirlik, uygun önışlemlerin yapılması, test başarısı	30
İlham Vericilik	Metotların özgünlüğü ve yaratıcılığı, gelecek planı, ileri konseptlerin düşünülmesi ve anlatımı, yeni metriklerin sunulması, çözümün genişletilmesi	25
Takım Oyunu	Görev dağılımı, iletişim, zaman yönetimi, eşit katılımcılık, ekip içi uyum, ortak değerlendirme	15
Uygulanabilirlik	Mimari kurgu, müşteri odaklılık, ölçeklenebilirlik, genelleştirilebilirlik, yazılım konseptlerine hakimiyet	15
Sunum Becerileri	Kapsam, akıcılık, isterlerin anlatımı, görsel uyum, açıklık, çözümün nedensellik kurgusuyla anlatımı, iletişim becerileri	15

13. Tabloda görüldüğü üzere puanlamada dikkate alınacak konular teknik becerilerinizle kısıtlanmadı. Yarışma süresince çözüm geliştirmek için sahip olduğunuz süre kısıtlı olduğundan sunumlarınız sırasında ortaya koyacağınız fikirlerin önemi çok büyük olacak. Sunuma dahil etmenizi beklediğimiz konuları aşağıdaki maddelerde bulabilirsiniz. Ayrıca beklentilerimizin dışında üreteceğiniz ve gerçek hayatta karşılığı olabilecek her türlü fikriniz de ekstra puan olarak değerlendirilecek.
 - a. Problemi daha iyi anlamak için yaptığınız çalışmalar
 - b. Önışlem adımlarınız ve sebepleri
 - c. Modelleme yönteminiz ve sebepleri
 - d. Model değerlendirme yönteminiz ve sebepleri
 - e. Modelinizi nasıl bir sistemle canlıya alırdınız, hangi teknolojileri ne amaçla kullanırdınız, ne gibi ihtiyaçlarınız olurdu?

- f. Daha çok zamanınız olsa ne tür geliştirmeler yapardınız? Problemin doğası gereği doğurduğu ihtiyaçlara ne gibi çözümler ürettirdiniz?
- g. Kullanıcıya sunacağınız daha farklı bilgiler/metrikler olur muydu? Benzin talep tahmin modelini başarıyla geliştirdikten sonra bu modeli kullanarak değer yaratabilecek başka çözüm önerileriniz varsa nelerdir?

14. Unutmayın, zamanınız kısıtlı. Takım oyunu ve iş bölümünü ön planda tutarak problemi anlama becerinizi, veri bilimi ve yazılım bilginizi, ileri konseptlere hakimiyetinizi, araştırmacı ruhunuzu ve de en önemlisi yaratıcı fikirlerinizi sergilemenizi tavsiye ederiz.

TÜPRAŞ Veri Analitiği Ekibi Olarak İyi Şanslar Dileriz!