**Problem Tanımı**

DHL gibi kontrat lojistiği hizmeti sağlayan firmalarda belirli periyotlarda müşterinin ürün gamının ihtiyaçları ve kısıtları gözetilerek sözleşmeler yapılır. Bu sözleşmelerden önce gelir/gider tahminlemelerinin yapılabilmesi için hacim tahminlemeleri oldukça kritiktir. Bu nedenle, yarışmanın ilk aşamasında, 2021 yılı ilk 9 haftasına ilişkin siparişlerin ürün ve müşteri bazında **Hacim Tahminlemesi** yapılması beklenmektedir. Konu ile ilgili detay bilgiye <https://datathon.dhlrdlab.com/> adresinden ulaşabilirsiniz.

Söz konusu tahminleme modeli için train\_set.csv, test\_set.csv ve submission.csv olmak üzere 3 farklı csv dosyası paylaşılmıştır. Test verisi üzerinde yapmış olduğunuz tahmin sonuçlarınızın submission.csv formatında yüklenmesi beklenmektir.

**Dosya Açıklamaları**

train\_set.csv – training veri seti

test\_set.csv – test veri seti

submission.csv – doğru submission formatı

**Veri Setleri Hakkındaki Açıklamalar**

**train\_set.csv:**

* Sipariş Numarası: Her siparişe ait unique ID.
* Sipariş Satırı: Aynı “Sipariş Numarası” içinde sipariş edilen ürünlerin satır numarası. (Aynı siparişe ilişkin farklı ürünleri belirtilen satır numarası olarak değerlendirilebilir)
* Sipariş Tipi: Sevk edilecek siparişin tipi.
* Malzeme Numarası: Ürüne ait unique ID.
* Sipariş Edilen Miktar: Ürüne ait sipariş miktarı.
* Sku Plt. Standartı: Her ürüne ait bir palete yüklenebilecek ürün miktarı. (ürünün boyutlarına göre değişkenlik gösterir)
* Müşteri No: Siparişin sevk edileceği müşteriye ait unique ID.
* Sipariş Tarihi: Siparişin DHL depo yönetim sistemindeki tarihi. (Yıl – Ay- Gün formatında)

**submission.csv:**

* yılhafta: Siparişin DHL depo yönetim sistemindeki tarihi (“yılhafta” şeklinde bitişik yazılmış formatta)
* Malzeme Numarası: Ürüne ait unique ID.
* Müşteri: Siparişin sevk edileceği müşteriye ait unique ID.
* Sipariş Edilen Miktar: Tahminlenecek olan sipariş miktarı.

**Dış Veri Kullanımı**

Modelin kurulumunda size faydalı olabileceğini düşündüğünüz her türlü dış veriyi kullanabilirsiniz.

**Değerlendirme**

Metrik kriteri olarak **RMSE** ***(Root Mean Square Error)*** ve **R2** skor değeri seçilmiştir.

Sırası ile;

<https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean_squared_error.html> <https://scikit-learn.org/stable/modules/generated/sklearn.metrics.r2_score.html>

linklerinden ilgili metrik kriterlerini inceleyebilirsiniz. İleteceğiniz submission.csv dosya formatındaki tahmin çıktılarınız gerçek değerler ile kıyaslanacak ve 2021 ilk 9 haftası için en düşük RMSE değerlerini ve en yüksek R2 skorlarını üreten takımlar bir üst tura çıkmaya hak kazanacaklardır.