

Wasserstein Distance:

Genel olarak, referans veri setiyle mevcut veri setini karşılaştırarak veri dağılımında kayma olup olmadığını kontrol eder. Bu işlemi yaparken istatistikten yararlanır.

Data Drifting (Data Drift):

Data drifting, verilerin dağılımının zaman içinde değişmesi anlamına gelir. İki veri seti arasında sapma olup olmadığını kontrol eder.

Input; 2 adet sayısal veri girişi.

Output: P value. Elde edilen p değerine göre data drifting var mı yok mu ve seviyesini anlayabiliyoruz.

Data Drifting Sebepleri:

Seasonal Changes: Mevsimsel veya zamana özgü olaylar nedeniyle verilerde meydana gelen değişiklikler.

Market Changes: Pazar trendlerinde veya ekonomik koşullarda meydana gelen kaymalar.

User Behavior Changes: Kullanıcıların bir sistem veya hizmetle nasıl etkileşime geçtiğindeki varyasyonlar.

External Factors: Yeni düzenlemeler, rekabet veya toplumsal trendler gibi dış etkenler nedeniyle meydana gelen değişiklikler.

Wasserstein Distance Kullanımı:

Wasserstein Distance, iki veri seti arasındaki farkı ölçmek için kullanılan güçlü bir yöntemdir. Özellikle, p-value değeri belirli bir eşik değerinin altına düştüğünde, iki veri seti arasında anlamlı bir fark olduğunu ve dolayısıyla veri dağılımında bir kayma olduğunu gösterir.

Pratik Uygulamalar:

Model Performansının Takibi: Makine öğrenmesi modellerinin performansını zaman içinde izlemek ve gerekli güncellemeleri yapmak.

Anomali Tespiti: Veri anormalliklerini ve olağandışı olayları tespit etmek için kullanılabilir.

Uyum ve Güncelleme: Mevcut modelleri yeni veri dağılımlarına göre yeniden eğitmek ve uyum sağlamak.

Sonuç:

Veri dağılımındaki değişimi düzenli aralıklarla sapma var mı yok mu diye kontrol etmesine yarar. Bu şu işe yarar, Elimizdeki veriye her geçen gün yeni veri eklendiğinde fraud ve anomaly detection gibi modellerinin performansı yeni veriler üzerinde istenen düzeyde olmayabilir ve model performansı eskiye nazaran kötü olababilir. Bu değişimi inceleyip sapma var mı yok mu diye çıktı üreten wasserstein distance algoritmasını sistemize entegre ederek modelimizin

performansını negatif yönde etkileyecek değişimleri önceden tespit edip zamanında müdahale etme imkanı sunmamızı sağlar. Bu ön uyarı sistemi sayesinde fraud ve anomaly detection gibi projelerde güncel kalmamızda yardımcı büyük olur.