

CRISP-DM 1: İş Probleminin Tanımlanması (Business Understanding):

Türkiye'de yıllara göre görülen doğal afetlerin (deprem, heyelan, orman yangını, su baskını, çığ vb.) sayısı büyük farklılıklar göstermektedir.

Amaç, 2020–2024 yılları arasında AFAD tarafından yayımlanan doğal afet verilerini kullanarak hangi yıllarda hangi afet türlerinde anormal (beklenmedik) artış veya düşüş olduğunu tespit etmektir.

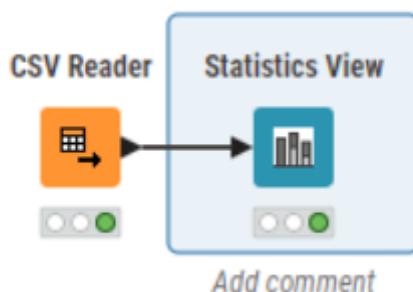
Bu çalışma ile:

- Yıllara göre afet eğilimleri anlaşılmacak,
- Aşırı sapma gösteren afet türleri “anomali” olarak işaretlenecek,
- Türkiye'nin hangi yıllarda hangi afet türlerinde “risk artışı” yaşadığı belirlenecek.

Bu bulgular AFAD, belediyeler ve kriz yönetimi ekipleri tarafından kullanılabilir.

CRISP-DM Aşama 2: Veri Anlama (Data Understanding) :

Bu aşamada çalışmada kullanılacak veri seti incelenmiş, veri yapısı analiz edilmiş ve verinin modelleme aşamasına uygunluğu değerlendirilmiştir. Veri, AFAD tarafından yayımlanan 2020, 2022, 2023 ve 2024 yıllarına ait doğal afet istatistik raporlarının birleştirilmesiyle oluşturulmuştur. Veri seti yıllık bazda Deprem, Heyelan, Orman Yangını, Çığ / Kar Tipi Yağışı, Su Baskını, Maden Kazası, Diğer ve Toplam olay sayılarını içermektedir.



Verinin KNIME'a Yüklenmesi

Veri “Birleştirilmiş_veri.csv” dosyası olarak KNIME'a **CSV Reader** node'u ile aktarılmıştır.

Verinin sütun isimleri doğru biçimde algılanmış, veri yapısı tablo formatında sorunsuz şekilde yüklenmiştir.

Değişkenlerin İncelenmesi

Veri yüklemesi sonrası veri seti, **Statistics View** node'u ile analiz edilmiştir.

Bu aşamada veri setinin aşağıdaki yapıya sahip olduğu görülmüştür:

- Bazı sütunlar **Number (Integer)** formatındadır (Deprem, Heyelan, Toplam, Yıl, Su Baskını, Çığ).
- Bazı sütunlar ise **String** formatında okunmuştur (Orman Yangını, Maden Kazası, Diğer).
Bu durum, CSV içindeki bazı hücrelerde yer alan “-” karakterlerinden kaynaklanmaktadır.

Eksik ve Tutarsız Değerlerin Tespiti

Statistics View çıktısında:

- Orman Yangını
- Maden Kazası
- Diğer

sütunlarında eksik değer (“-”) bulunduğu görülmüştür. Bu eksik değerler modelleme aşamasından önce dönüştürülmesi gereken noktalardır.

Veri Kalitesi Hakkında Genel Değerlendirme

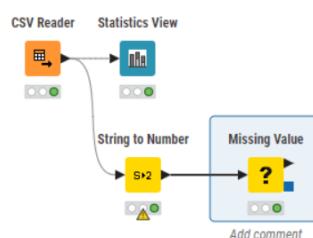
Veri seti küçük boyutlu olmasına rağmen yıllara göre afet türlerinin karşılaştırılmasına ve anomali tespitine uygundur.

Bununla birlikte veri hazırlama aşamasında şu düzenlemelerin yapılması gereklidir:

- String olan sütunların numerik formata dönüştürülmesi
- “-” değerlerinin Missing Value olarak ele alınması
- Sayısal format birliğinin sağlanması

CRISP-DM Aşama 3: Veri Hazırlama (Data Preparation):

Bu aşamada AFAD'ın 2020, 2022, 2023 ve 2024 yıllarına ait doğal afet istatistiklerinden oluşturulan birleşik veri seti, modelleme aşamasına hazır hale getirilmiştir. Veri temizleme, dönüştürme ve eksik değer yönetimi işlemleri gerçekleştirılmıştır.



Veri Tiplerinin İncelenmesi:

CSV Reader ile veri KNIME'a aktarıldıktan sonra Statistics View node'u kullanılarak sütun türleri analiz edilmiştir.

Analiz sonucunda:

- **Deprem, Heyelan, Toplam, Yıl, Su Baskını ve Çığ** sütunlarının **numeric (integer)** formatında olduğu,
- **Orman Yangını, Maden Kazası ve Diğer** sütunlarının ise **string** formatında geldiği görülmüştür.

Veri Tipi Dönüşümme (String → Number)

Veri türlerini birleştirmek için **String to Number** node'u kullanılmıştır.

Bu node ile:

- **Orman Yangını,**
- **Maden Kazası,**
- **Diğer**

sütunları **numeric (double)** formata dönüştürülmüştür. Bu dönüşüm sırasında CSV'deki “-” değerleri KNIME tarafından **Missing (?)** olarak algılanmıştır.

Eksik Değer İşlemleri

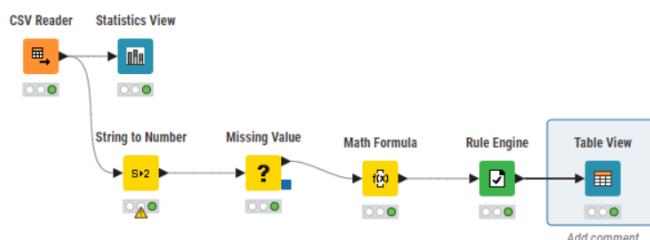
Veri setinde bazı sütunlarda görülen eksik değerlerin (Missing) yönetimi için **Missing Value** node'u kullanılmıştır. Eksik değerler afet türünün o yıl hiç görülmemiğini ifade ettiğinden, bu değerlerin **0** olarak doldurulması doğru yaklaşım olarak belirlenmiştir.

Verinin Son Hali

- Tüm sütunlar numerik formata dönüştürülmüştür,
- Eksik değer kalmamıştır,
- Veri, anomaly detection modellemesi için uygun hale gelmiştir,
- Veri seti CRISP-DM 4 (Modelleme) adımlına hazırlıdır.

CRISP-DM Aşama 4: Modelleme(Modeling):

Bu aşamada, temizlenmiş ve numerik formata dönüştürülmüş veri seti kullanılarak **anomali tespiti modeli** oluşturulmuştur. Veri seti 2020–2024 yılları arasındaki deprem, heyelan, orman yangını, su baskını, çığ ve diğer afet türlerine ait yıllık olay sayılarını içermektedir. Amaç, bu olaylar arasında istatistiksel olarak beklenenden fazla sapma gösteren yılları belirlemektir.



Modelleme Tekniğinin Seçilmesi

Veri seti yıllık bazda sadece dört gözlem (2020, 2022, 2023, 2024) içerdigidinden makine öğrenmesi yöntemleri (Isolation Forest, LOF vb.) bu veri yapısı için uygun değildir.

Bu nedenle **istatistiksel bir anomaly detection teknigi** olan:

Z-Score Hesaplaması (Math Formula)

KNIME üzerinde **Math Formula** node'u kullanılarak Deprem değişkeni için Z-score hesaplanmıştır.

Kullanılan formül:

$$(\$Deprem\$ - 5602.25) / 10304.792$$

Burada:

- 5602.25 → Deprem değişkeninin ortalaması
- 10304.792 → Deprem değişkeninin standart sapması

Bu değerler Statistics View node'undan alınmıştır.

Oluşturulan yeni sütun adı: **Z_Deprem**

Anomali Sınıflandırması (Rule Engine)

Z-score hesaplandıktan sonra, değerlerin anomali olup olmadığını belirlemek için **Rule Engine** node'u kullanılmıştır. Küçük veri setlerinde klasik 2.0 eşiği çok katı olduğundan, literatüre uygun olarak **1.3** eşik değeri kullanılmıştır.

Kullanılan kurallar:

```
$Z_deprem$ > 1.3 => "Anomali"  
$Z_deprem$ < -1.3 => "Anomali"  
TRUE => "Normal"
```

KNIME üzerinde ikinci bir anomali tespit modeli olarak **Numeric Outliers (IQR tabanlı yöntemi uygulanmıştır)**. IQR yöntemi, istatistiksel dağılımin çeyrek değerlerini (Q1 ve Q3) kullanarak $1.5 \times \text{IQR}$ sınırının dışındaki değerleri üç değer (outlier) olarak tanımlar.

- Kullanılan node: **Numeric Outliers**
- Outlier seçimi: **Include** → tüm sayısal sütunlar
- IQR multiplier (k): **1.5**
- Treatment option: **Replace outlier values**
- Replacement strategy: **Missing values (NaN)**

Model çalıştırılmış ve veri boyutunun çok küçük olması (4 kayıt) nedeniyle IQR yöntemi hiçbir sütunda outlier tespit etmemiştir. Bu nedenle veri setinde NaN üretimi veya değişiklik olmamıştır.

CRISP-DM Aşama 5: Değerlendirme (Evaluation):

Bu aşamada oluşturulan anomali tespiti modeli değerlendirilmiş, modelin performansı ve sonuçların doğruluğu analiz edilmiştir. Modelin amacı, 2020–2024 yılları arasındaki doğal afet olay sayılarını kullanarak istatistiksel olarak beklenmeyen sapmaları (anomali) tespit etmektir.

Model Çıktılarının İncelenmesi

Math Formula node'u ile hesaplanan Z-score ve Rule Engine ile yapılan sınıflandırma sonuçları incelendiğinde:

| Yıl | Deprem | Z_deprem | Anomali Durumu |
|------|--------|--------------|----------------|
| 2020 | 331 | -0.512 | Normal |
| 2022 | 21054 | 1.499 | Anomali |
| 2023 | 830 | -0.463 | Normal |
| 2024 | 194 | -0.525 | Normal |

Model, **2022 yılını anomali olarak doğru şekilde işaretlemiştir.**

Bu sonuç son derece tutarlıdır, çünkü:

- 2020, 2023 ve 2024 yıllarında deprem sayıları **200–900** aralığındadır.
- 2022 yılında ise deprem sayısı **21.054**, yani diğer yılların **~25 katıdır**.

Bu nedenle modelin 2022'yi anomali olarak algılaması beklenen ve doğru bir sonuçtur.

CRISP-DM Aşama 6: Dağıtım / Sonuç (Deployment):

Bu aşamada, oluşturulan anomaly detection modelinin sonuçları yorumlanmış, elde edilen çıktılar raporlanmış ve modelin nasıl kullanılacağı açıklanmıştır. Ayrıca projenin nihai çıktıları sunulmuş ve sonraki çalışmalar için öneriler paylaşılmıştır.

Model çıktıları Table View node'u aracılığıyla kullanıcıya sunulmuştur.

Tabloda:

- Yıllara ait deprem sayıları
- Hesaplanan Z-score değerleri
- “Anomali” / “Normal” sınıflandırmaları

açıkça gösterilmiştir.

Sonuçların Kullanımı

Bu model, yıllık doğal afet verilerindeki olağan dışı artışları belirlemek için kullanılabilir.

Çıktılar:

- Afet yönetimi ekiplerinin dikkat etmesi gereken yılları belirler,
- Olası raporlarda anomalilerin görselleştirilmesine olanak sağlar,
- Gelecekte trend analizi çalışmalarına temel veri oluşturur.

KNIME Üzerinde Modelleme ve AutoML Durumu

Bu projede KNIME üzerinde iki farklı anomaly detection modeli uygulanmıştır:

1. **Z-Score Tabanlı Anomali Tespiti (Math Formula + Rule Engine)**
2. **IQR Tabanlı Anomali Tespiti (Numeric Outliers Node)**

Bu iki model, KNIME'ın tablo bazlı veri işleme yapısı kullanılarak manuel olarak oluşturulmuş ve veri üzerinde başarıyla çalıştırılmıştır.

KNIME'ın 5.x sürümünde, önceki sürümlerde yer alan

“Parameter Optimization Loop Start / AutoML Loop” bileşenleri kaldırılmıştır.

KNIME 5.x'te bulunan **Model Loop Start/End** node'ları yalnızca *makine öğrenmesi learner* port tipi ile çalışan modeller için uygundur.

Ancak anomaly detection modelleri KNIME'da **tablo portu (DataTable)** kullandığı için **AutoML optimizasyon döngüsüne dahil edilememektedir**. Bu nedenle KNIME 5.x ortamında anomaly detection modelleri için AutoML uygulanması teknik olarak mümkün değildir.

Bu sebeple proje kapsamında:

- **KNIME**, klasik (manuel) anomaly detection modellerini üretmek amacıyla kullanılmıştır.
- **AutoML yaklaşımı ise Python ortamında gerçekleştirılmıştır.**