

# PAKİSTAN'DAKİ İNTİHAR BOMBALI SALDIRILAR ÜZERİNE YAPAY ZEKA ÇALIŞMASI

**Ad Soyad :** Abdullah Bahar

**Öğrenci No :** \*\*\*\*\*

**Bölüm :** Bilgisayar Programcılığı

**Ders :** Yapay Zeka

**Öğretmen :** Ali Mertcan Köse

# Giriş

Günümüzde terörizmin en ölümcül ve etkili taktiklerinden biri olan intihar bombalamaları, küresel güvenlik tehditlerinin önemli bir parçasını oluşturuyor. Bu operasyonel yöntem, terörist grupların maksimum yıkım ve hasar yaratmak amacıyla kullanıldığı ölümcül bir stratejidir.

Dünya genelindeki terörist saldırıların sadece %3'ünü oluşturmalarına rağmen, bu tür saldırılar %48'lik bir ölüm oranıyla can kayıplarının büyük bir kısmını oluşturuyor. 1980'den 2001'e kadar intihar bombalamalarında ortalama ölüm sayısının 13 olduğu bilinmektedir. İntihar saldırılarının hedeflerini hızla değiştirebilme yeteneği ve ucuz maliyeti, terörist grupların tercih ettiği bir taktik haline getirmiştir.

1982'den bu yana 40'tan fazla ülkede 3.600'den fazla intihar bombalaması gerçekleşmiştir, bu da terörün küresel tehdidini vurgulamaktadır.

Bu rapor “Pakistan Suicide Bombing Attacks” veri setinde yola çıkarak intihar bombalama saldırılarında minimum ve maksimum olacak şekilde ölü ve yaralı sayıları üzerine bir tahmin modeli geliştirmek üzerine geliştirilmiştir.

Raporun büyük bölümünü veri setinin temizlenmesi ve ön işleme içerir. Devamında ise model eğitimi ve sonuçları üzerine durulmuştur.

**Not :** Bu çalışmanın ana amacı eldeki verilere dayanarak olası intihar bombalama saldırılarının vuku bulmadan önce tespiti üzerinedir. Ancak bu yolda başarıya ulaşılması pek mümkün olmadığından dolayı daha basit bir senaryo üzerinde bu çalışma gerçekleştirilmiştir.

# Veri Seti

Veri seti; 496 satır ve 26 sütundan oluşmaktadır.

Veri setinin sütun değerleri:

- S# (Serial Number) - Sıra Numarası
- Date - Tarih
- Islamic Date - İslami Tarih
- Blast Day Type - Patlama Gün Türü
- Holiday Type - Tatil Türü
- Time - Zaman
- City - Şehir
- Latitude - Enlem
- Longitude - Boylam
- Province - İl
- Location - Konum
- Location Category - Konum Kategorisi
- Location Sensitivity - Konum Hassasiyeti
- Open/Closed Space - Açık/Kapalı Alan
- Influencing Event/Event - Etkileyen Olay/Olay
- Target Type - Hedef Türü
- Targeted Sect if any - Hedef Alınan Mezhep (Eğer varsa)
- Killed Min - Ölenler (Minimum)
- Killed Max - Ölenler (Maksimum)
- Injured Min - Yaralılar (Minimum)
- Injured Max - Yaralılar (Maksimum)
- No. of Suicide Blasts - İntihar Bombalamalarının Sayısı
- Explosive Weight (max) - Patlayıcı Ağırlığı (maksimum)
- Hospital Names - Hastane Adları
- Temperature(C) - Sıcaklık (Celsius)
- Temperature(F) - Sıcaklık (Fahrenheit)

S#	Date	Islamic Date	Blast Day Type	Holiday Type	Time	City	Latitude	Longitude	Province	...	Targeted Sect if any	Killed Min	Killed Max	Injured Min	Injured Max	No. of Suicide Blasts	Explosive Weight (max)	Hospital Names	Temperature(C)	Temperature(F)
0	1	Sunday- November 19-1995	25 Jumaada al-Thaany 1416 A.H	Holiday	Weekend	NaN	Islamabad	33.7180	73.0718	Capital	...	NaN	14.0	15.0	NaN	60	2.0			
1	2	Monday- November 6-2000	10 SHa'baan 1421 A.H	Working Day	NaN	NaN	Karachi	24.9918	66.9911	Sindh	...	NaN	NaN	3.0	NaN	3	1.0			
2	3	Wednesday- May 8-2002	25 safar 1423 A.H	Working Day	NaN	7:45 AM	Karachi	24.9918	66.9911	Sindh	...	Christian	13.0	15.0	20.0	40	1.0			
3	4	Friday- June 14-2002	3 Raby' al-Thaany 1423 A.H	Working Day	NaN	11:10:00 AM	Karachi	24.9918	66.9911	Sindh	...	Christian	NaN	12.0	NaN	51	1.0			
4	5	Friday- July 4-2003	4 Jumaada al-awal 1424 A.H	Working Day	NaN	NaN	Quetta	30.2095	67.0182	Baluchistan	...	Shiite	44.0	47.0	NaN	65	1.0			

## Veri Seti Ön İşlemesi

Veri setinin index değeri zaten var olduğu için “S#” sütunu gereksiz yer kapladığı için silindi.

“Date” sütununun içindeki haftanın günlerini belirten değerler gerksiz görülüp silinmiş, ilgili tarih bilgileri de date formatına uygun olacak şekilde tekrardan düzenlenmiştir.

“Islamic Date” sütunu, *hijri\_converter* kütüphanesinden yararlanılarak date formatında tekrardan hesaplanarak miladi takvimden hicri takvime çevrilmiştir.

“Balast Day Type” sütunu; “Working Day”, “Holiday”, “Weekend” değerlerini alan kategorik bir niteliklidir. Çok fazla veri çeşidi olmamasından dolayı null değere sahip satırlar, sütunun mod değeriyle doldurulmuştur.

“Location Category” sütunu; pek değeri bünyesinde barındıran kategorik bir niteliklidir. Bu sütun için birbirinin aynı olan sütunlar birleştirilmiştir. Null değere sahip satırlar kategori çokluğundan dolayı mod ile doldurulmayıp en çok tekrar eden ilk üç kategori arasından rastgele değerler atanarak doldurulmuştur.

“Location Sensitivity” sütunu; temelde üç adet kategoriye sahiplik yapacak şekilde düzenlenmiş, ilgili null değerler sütun mod’u ile doldurulmuştur.

“City” sütununun kendisinde pek çok kategorik değer olduğu için birbirinin benzeri olanlar, *fuzzywuzzy* kütüphanesinden yararlanılarak tek kategoride birleştirilmiştir. Bu birleştirilme için %90’lık bir benzerlik dikkate alınmıştır.

“Province” sütunundaki kategorik değerler için büyük küçük harf uyumluluğu sağlanılarak gereksiz kategorilerden kurtulunmuştur.

“Latitude” ve “Longitude” için null değer doldurmak için tek yolun enlem ve boylamın hesaplanması gerektiği düşünülmüş, bu hesaplamayı gerektirecek değerlerin azlığından dolayı bu zahmetten kaçınılp ilgili null değere sahip satırlar silinmiştir.

“Open/Closed Space” sütunu için birbirin benzeri olan kategorik değerler birleştirilmiş, null değerler ilgili sütunun mod değeri ile doldurulmuştur.

“Target Type” sütununda bolca birbirine benzeyen değerler olduğu için bu değerler birleştirilmiştir.

“Killed Min”, “Killed Max”, “Injured Min”, Injured Max” ve “No. of Suicide Blasts” sütunları sayısal değer olduğu için null değere sahip satırlar ilgili sütunların ortalamaları ile doldurulmuştur.

“Location” sütununa bu aşamada dokunulmamıştır.

Buraya kadar zikrettiğimiz sütunların dışında kalan sütunlar, sahip olduğu null değerlerin çokluğundan dolayı doldurulması mantıklı görülmemiş ve veri setinden silinmiştir.

Bu aşamaların sonunda elde ettiğimiz veri setinin sütunları aşağıdaki gibidir.

```
Data columns (total 19 columns):
```

#	Column	Non-Null Count	Dtype
0	Date	495 non-null	datetime64[ns]
1	Islamic Date	342 non-null	object
2	Blast Day Type	485 non-null	object
3	Holiday Type	72 non-null	object
4	City	495 non-null	object
5	Latitude	492 non-null	float64
6	Longitude	492 non-null	object
7	Province	495 non-null	object
8	Location	492 non-null	object
9	Location Category	495 non-null	object
10	Location Sensitivity	495 non-null	object
11	Open/Closed Space	461 non-null	object
12	Target Type	495 non-null	object
13	Killed Min	349 non-null	float64
14	Killed Max	495 non-null	float64
15	Injured Min	365 non-null	float64
16	Injured Max	463 non-null	object
17	No. of Suicide Bombs	414 non-null	float64

## Kategorikten Sayısala

Proje, sayısal verilerden yola çıkarak sayısal veri elde etmeyi amaçladığından dolayı sayısal değere dönüştürülmesi mümkün olmayan sütunlar veri setinden silindi.

Date tipinde veri tutan sütunlar “Year”, “Month”, “Day” isimlerin farlı sütunlara bölünmüştür.

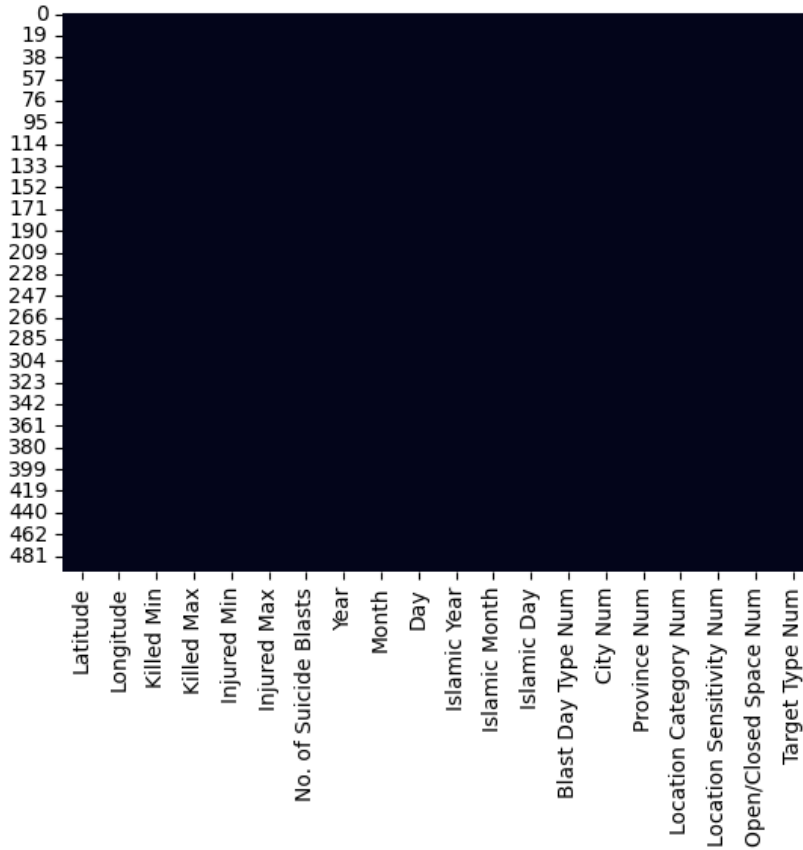
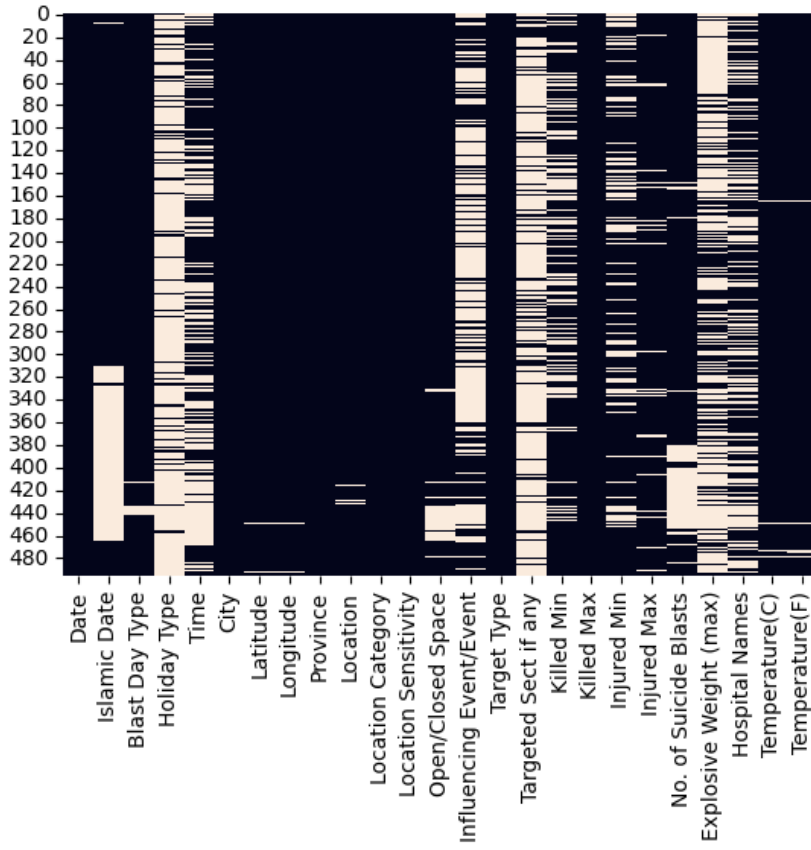
Kategorik değerlerin sayısal değere dönüştürülmesi için *Label Encoder* yöntemi kullanılmıştır. Bu bağlamda “Blast Day Type”, “City”, “Province”, “Location Category”, “Location Sensitivity”, “Open/Closed Space” ve “Target Type” sütunları, kategorik sayısal değerler çevrilmiştir.

Bütün bunların nihayetinde eski sütunlar silinip yeni sütunlar eklenmiştir.

#	Column	Non-Null Count	Dtype
0	Latitude	489 non-null	float64
1	Longitude	489 non-null	float64
2	Killed Min	489 non-null	float64
3	Killed Max	489 non-null	float64
4	Injured Min	489 non-null	float64
5	Injured Max	489 non-null	float64
6	No. of Suicide Bombs	489 non-null	float64
7	Year	489 non-null	int32
8	Month	489 non-null	int32
9	Day	489 non-null	int32
10	Islamic Year	489 non-null	int32
11	Islamic Month	489 non-null	int32
12	Islamic Day	489 non-null	int32
13	Blast Day Type Num	489 non-null	int32
14	City Num	489 non-null	int32
15	Province Num	489 non-null	int32
16	Location Category Num	489 non-null	int32
17	Location Sensitivity Num	489 non-null	int32
18	Open/Closed Space Num	489 non-null	int32
19	Target Type Num	489 non-null	int32

	Latitude	Longitude	Killed Min	Killed Max	Injured Min	Injured Max	No. of Suicide Bombs	Year	Month	Day	Islamic Year	Islamic Month	Islamic Day	Blast Day Type Num	City Num	Province Num	Location Category Num	Location Sensitivity Num
0	33.718000	73.071800	14.000000	15.0	31.578947	60.000000	2.0	1995	11	19	1416	6	26	0	17	2	6	
1	24.991800	66.991100	14.970845	3.0	31.578947	3.000000	1.0	2000	11	6	1421	8	10	2	19	6	16	
2	24.991800	66.991100	13.000000	15.0	20.000000	40.000000	1.0	2002	5	8	1423	2	25	2	19	6	12	
3	24.991800	66.991100	14.970845	12.0	31.578947	51.000000	1.0	2002	6	14	1423	4	3	2	19	6	6	
4	30.209500	67.018200	44.000000	47.0	31.578947	65.000000	1.0	2003	7	4	1424	5	4	2	42	1	19	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
490	30.203658	67.004469	15.000000	15.0	40.000000	40.000000	1.0	2017	8	12	1438	11	20	1	42	1	23	
491	28.571051	67.496895	21.000000	22.0	25.000000	39.721491	1.0	2017	10	5	1439	1	15	2	42	1	19	
493	30.221057	67.002524	2.000000	2.0	8.000000	8.000000	1.0	2017	11	9	1439	2	20	2	42	1	18	
494	33.970623	71.438620	1.000000	3.0	6.000000	8.000000	1.0	2017	11	24	1439	3	6	2	39	4	18	
495	30.139626	66.981476	4.000000	4.0	16.000000	22.000000	1.0	2017	11	25	1439	3	7	2	42	1	14	

Veri setinin ön işleme tabi tutulmadan önceki ve sonraki halleri:



## Proje Kapsamında Kullanılan Kütüphaneler

# Temel çizimler için kullanılır

```
import matplotlib.pyplot as plt
```

# Grafik noktasında yüksek seviyeli bir arayüz sağlar

```
import seaborn as sns
```

# Veri analizi ve manipülasyonu için kullanılır

```
import pandas as pd
```

# Bilimsel hesaplamalar ve matematiksel işlemler için kullanılır

```
import numpy as np
```

# Metin karşılaştırma ve eşleştirme işlemleri için kullanılır

```
import fuzzywuzzy
```

# Kategorik verileri sayısalı çevirmek için kullanılır

```
from sklearn.preprocessing import LabelEncoder
```

# Veri normalleştirme işlemi için kullanılır

```
from mlxtend.preprocessing import minmax_scaling
```

# Hicri ve miladi takvim dönüşümü için kullanılır

```
from hijri_converter import convert
```

# Metin formatındaki tarih ve saatleri ayrıştırmak için kullanılır

```
from dateutil import parser
```

#Uyarı sistemini kontrol etmek için kullanılır

```
import warnings
```

```
warnings.filterwarnings("ignore", category=FutureWarning)
```



# Model Eğitimi

Sayısal değerler ile çalışılacağından dolayı model eğitimi için *Logistic Regression* ve *Linear Discriminant Analysis* gibi methodlar kullanılmaya çalışılmış olsa da bu tarz methodlar sayısal ve kategorik veriler ile birlikte çalışacağından model eğitiminde hata alınmıştır. Öte yandan sadece sayısal verilerle eğitilebilen methodlar kullanılmıştır.

Bu bağlamda kullanılan modeller:

- Linear Regression
- Ridge Regresyon
- Xgboost
- Random Forrest Regression
- LGBMR Regression

Bu modeller ile “Killed Min”, “Killed Max”, “Injured Min” ve “Injured Max” için değer tahminleri yapılmış ve bu tahminler gerçek değerler ile karşılaştırılıp performans ölçümü yapılmıştır. Ayrıca modeller oluşturulurken %80’e %20 olacak şekilde eğitim ve test verisi olarak iki parçaya ayrılmıştır. Validasyon verisinin ayrılmasının sonuca olan bir katkısı olmayacağı öngörülmüştür.

## Model Eğitiminde Kullanılan Kütüphaneler

```
# Modeller
from sklearn.linear_model import LinearRegression
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.linear_model import Ridge
from sklearn.linear_model import ElasticNet
from xgboost import XGBRegressor
from lightgbm import LGBMRegressor
from sklearn.ensemble import RandomForestRegressor
from sklearn.linear_model import LogisticRegression

# İlgili hesaplamalar
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error
from sklearn.metrics import accuracy_score, confusion_matrix
```

## Performans Ölçümleri

*Mean Squared Error (MSE)* : Gerçek ve tahmin edilen değer arasındaki ortalama kare hatanın büyüklüğünü gösterir. MSE'nin yüksek olması, modelin tahminlerinin gerçek değerlerden ne kadar uzak olduğunu gösterir. Daha düşük bir MSE, daha iyi bir model performansını işaret eder.

*R-Squared (R-kare)* : Bu, bağımsız değişkenlerin bağımlı değişkendeki varyansı açıklama yüzdesini gösterir. Modelin açıklama gücünü ifade eder. R-kare ne kadar yüksekse, modelin bağımsız değişkenlerdeki varyansı açıklama yeteneği o kadar iyidir.

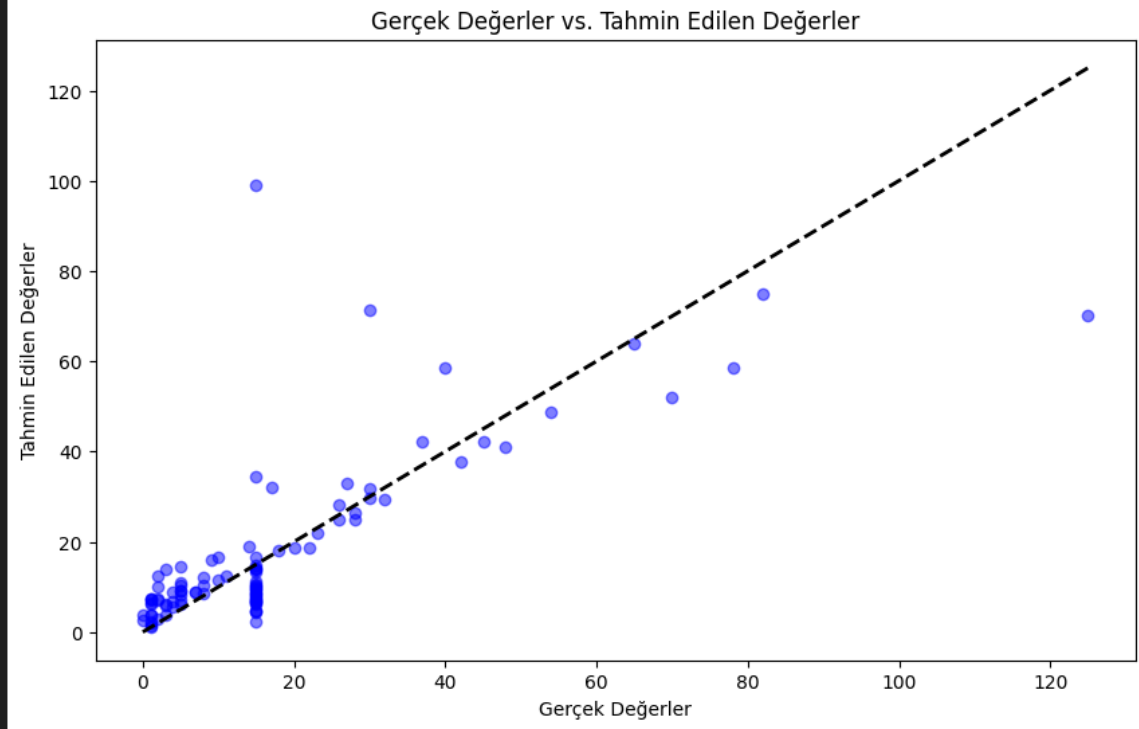
*Mean Absolute Error (MAE)* : Bu, gerçek ve tahmin edilen değerler arasındaki ortalama mutlak hatanın büyüklüğünü gösterir. MAE'nin düşük olması, modelin tahminlerinin gerçek değerlere yakın olduğunu gösterir. Daha düşük bir MAE, daha iyi bir model performansını işaret eder.

# Lineer Regression

Killed Min

Mean Squared Error: 163.15259156952612

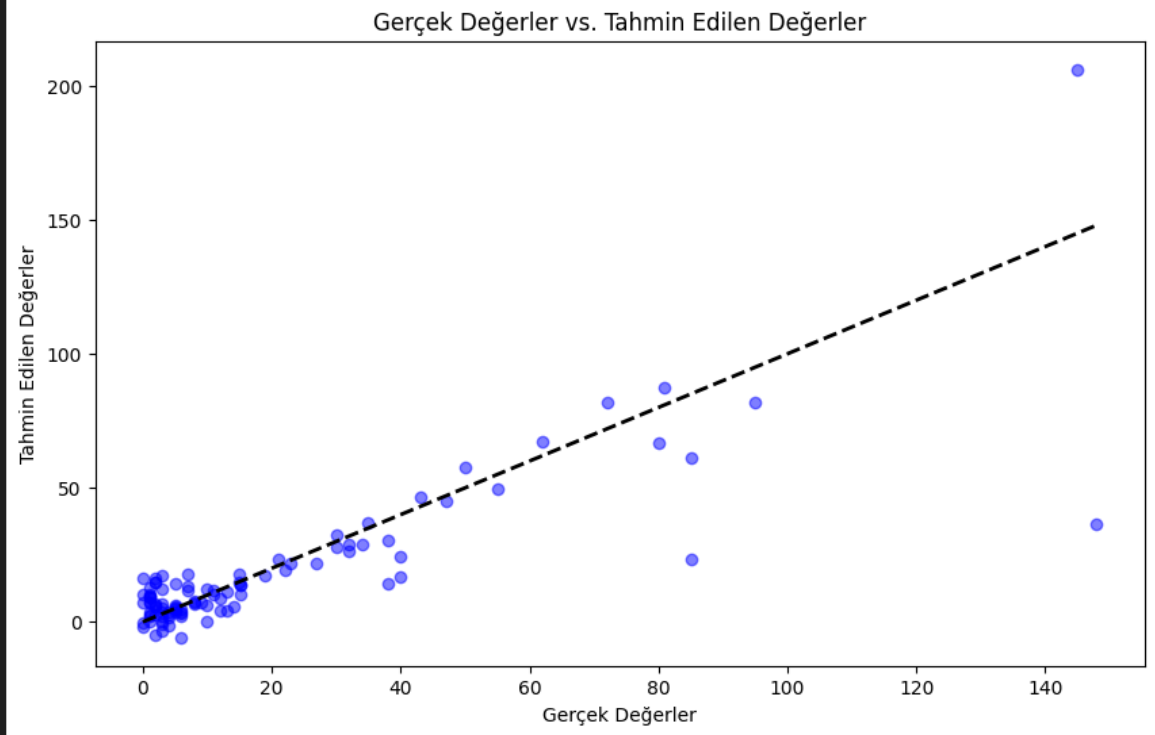
R-squared: 0.5839818105663301



Killed Max

Mean Squared Error: 260.6320631452964

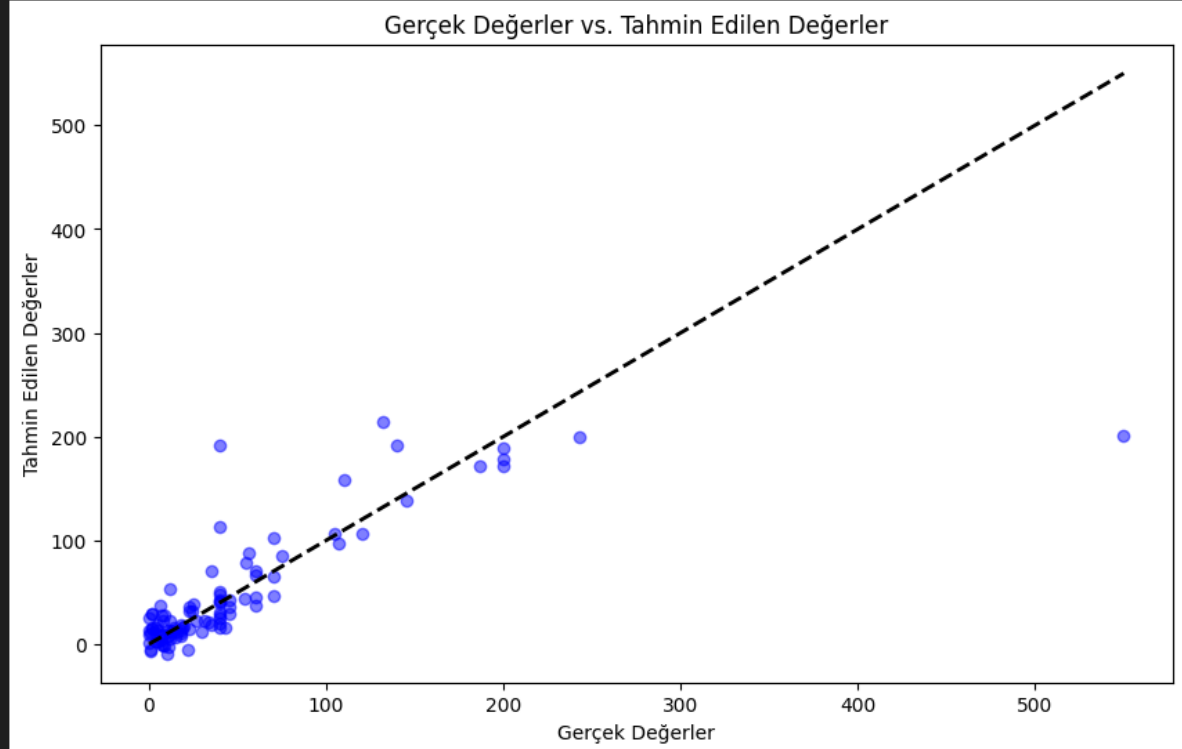
R-squared: 0.6808851090824156



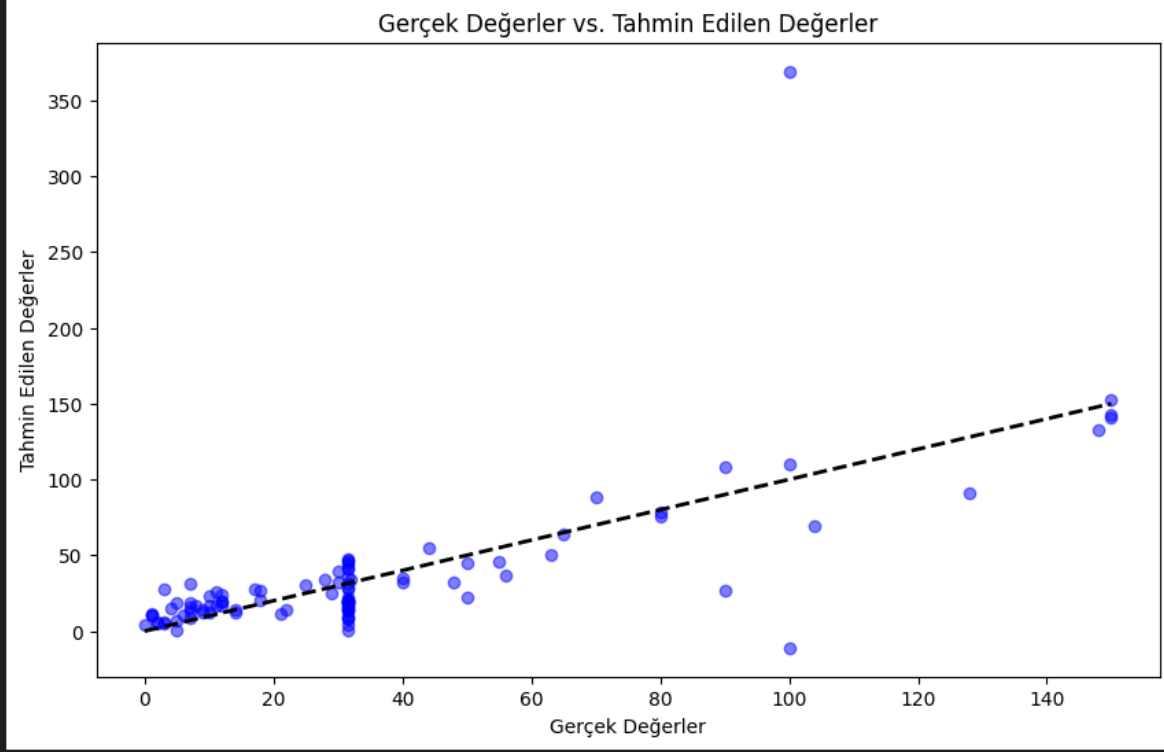
Injured Max

Mean Squared Error: 1884.9561420756945

R-squared: 0.6318516403039434



Injured Min  
Mean Squared Error: 1071.9145604482514  
R-squared: 0.13480100544829365



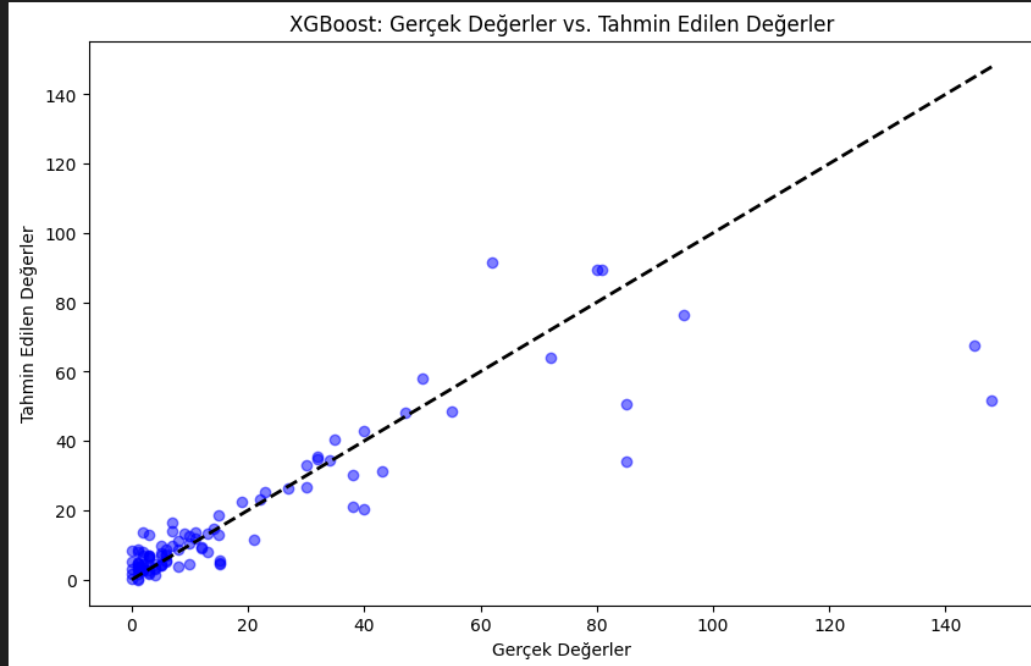
**Xgboost**

Killed Max

Mean Squared Error: 234.02943183514375

R-Squared: 0.7134570639148781

Mean Absolute Error: 6.825379552232691

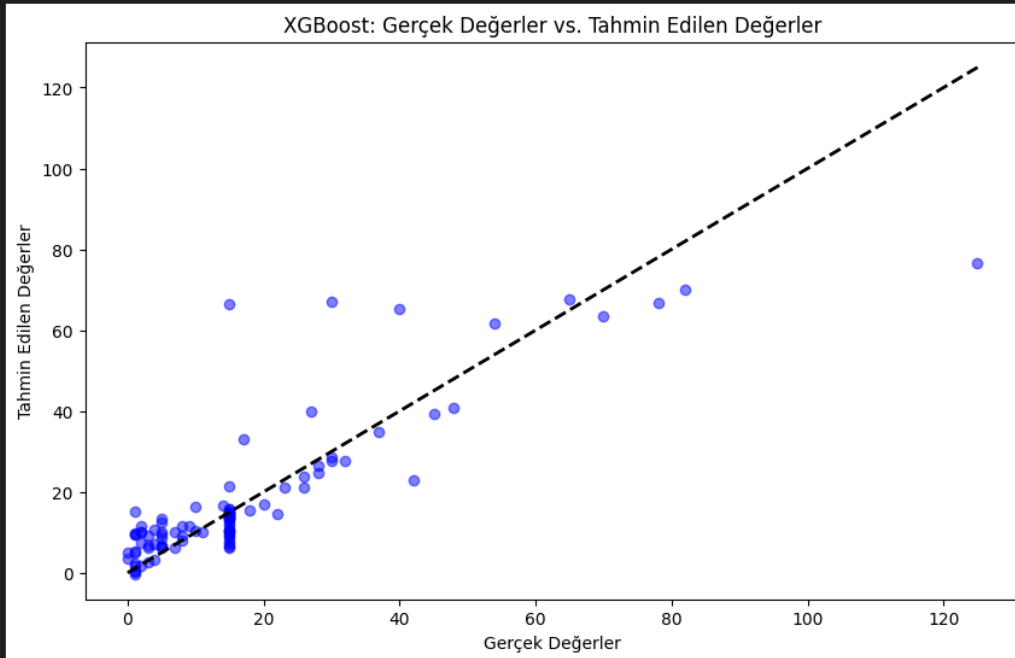


Killed Min

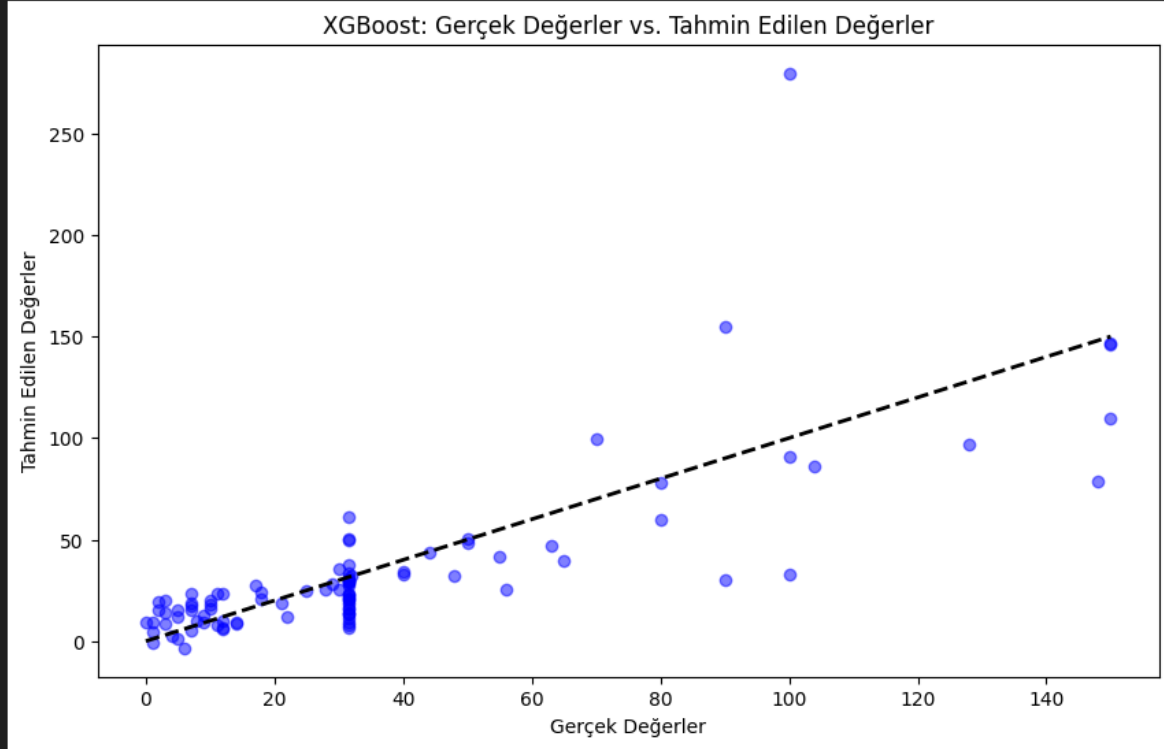
Mean Squared Error: 102.60176388725401

R-Squared: 0.7383786574613693

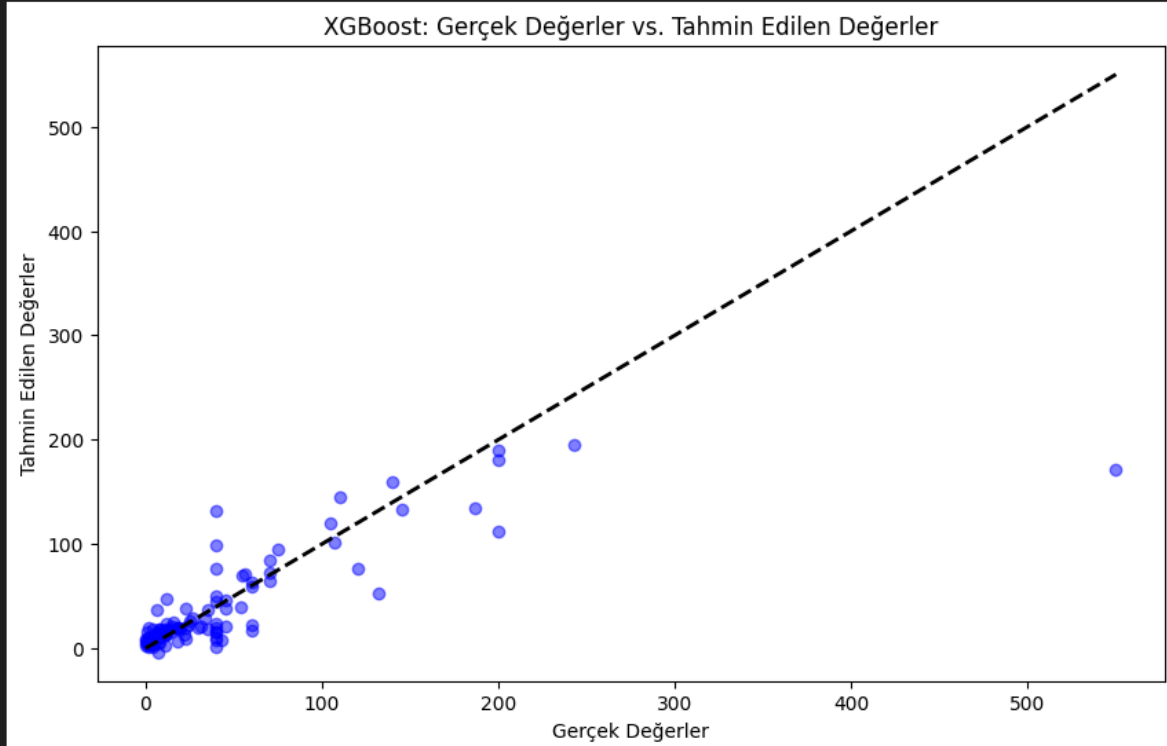
Mean Absolute Error: 5.7772779077608005



Injured Min  
Mean Squared Error: 663.3669366317587  
R-Squared: 0.46456142329792416  
Mean Absolute Error: 13.894332033671061



Injured Max  
Mean Squared Error: 2015.8286070603647  
R-Squared: 0.606291107494713  
Mean Absolute Error: 18.749227175117603



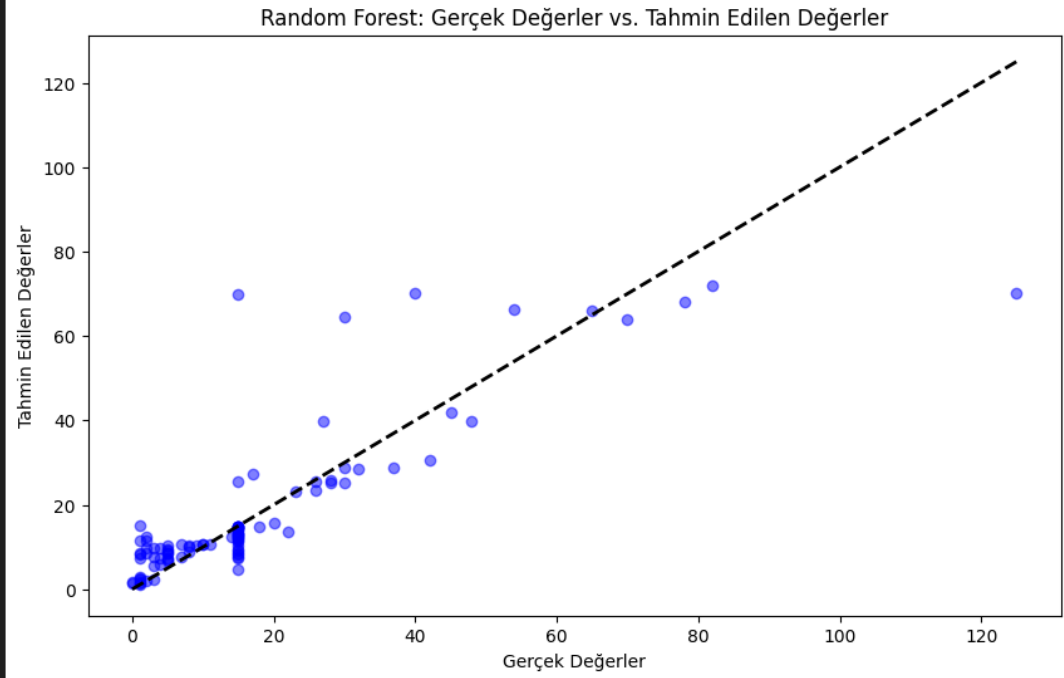
# Random Forrest Regression

Killed Min

Mean Squared Error: 110.41886957383261

R-Squared: 0.7184460402527959

Mean Absolute Error: 5.630409948235856

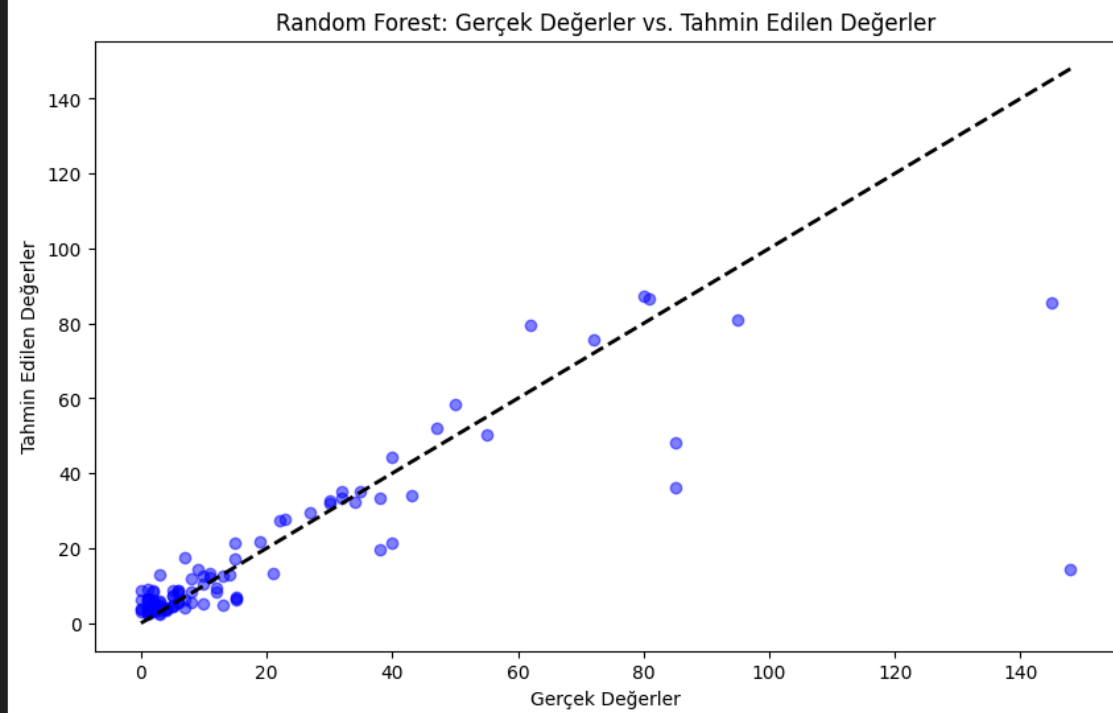


Killed Max

Mean Squared Error: 286.2471425939493

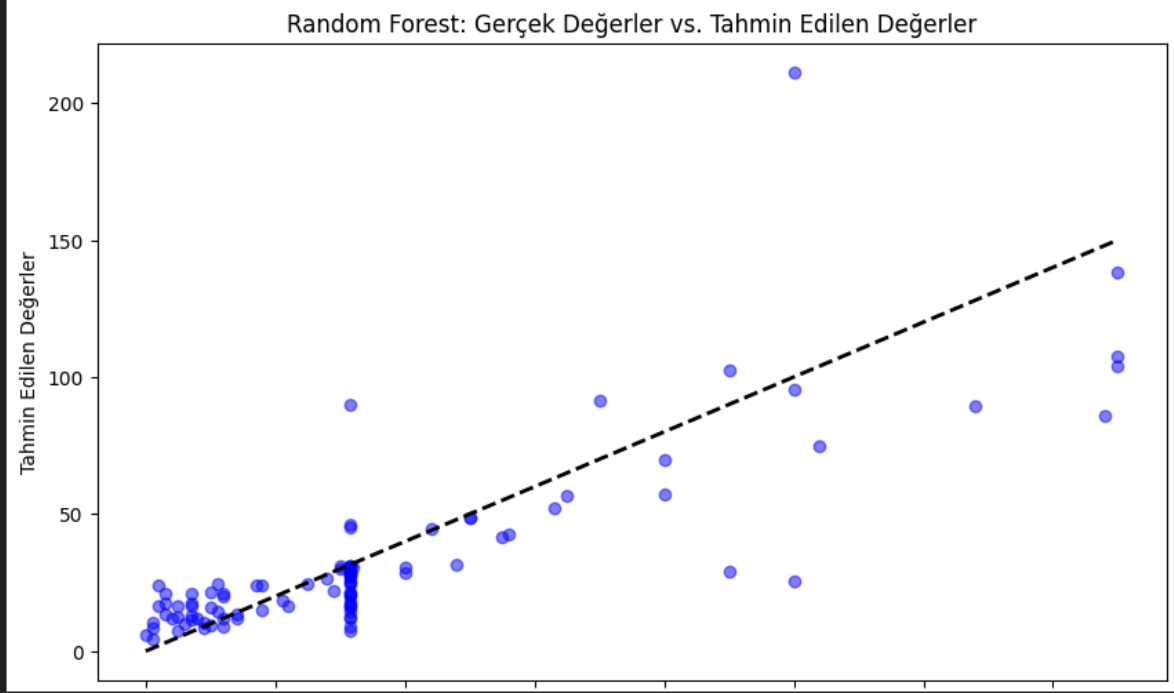
R-Squared: 0.6495223013547062

Mean Absolute Error: 6.734662627551021

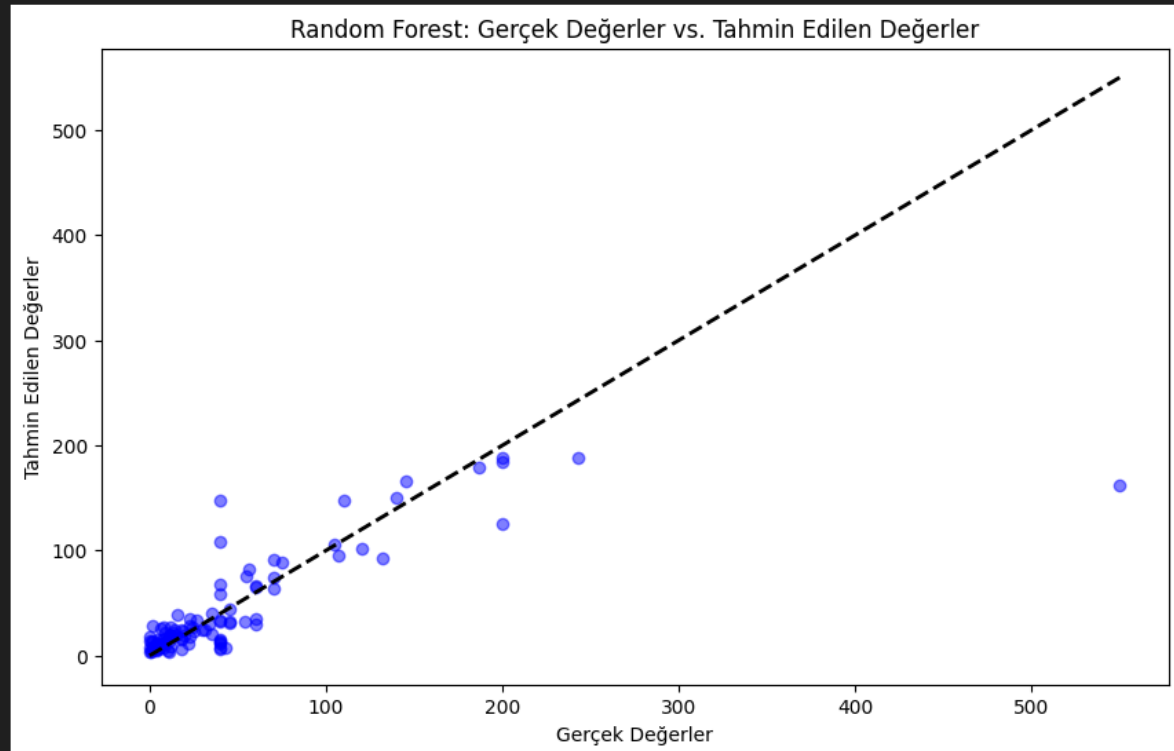




Injured Min  
Mean Squared Error: 455.4575503985526  
R-Squared: 0.632376096746906  
Mean Absolute Error: 12.799527389903322



Injured Max  
Mean Squared Error: 2005.9602396768519  
R-Squared: 0.6082184856357863  
Mean Absolute Error: 18.157433091657715



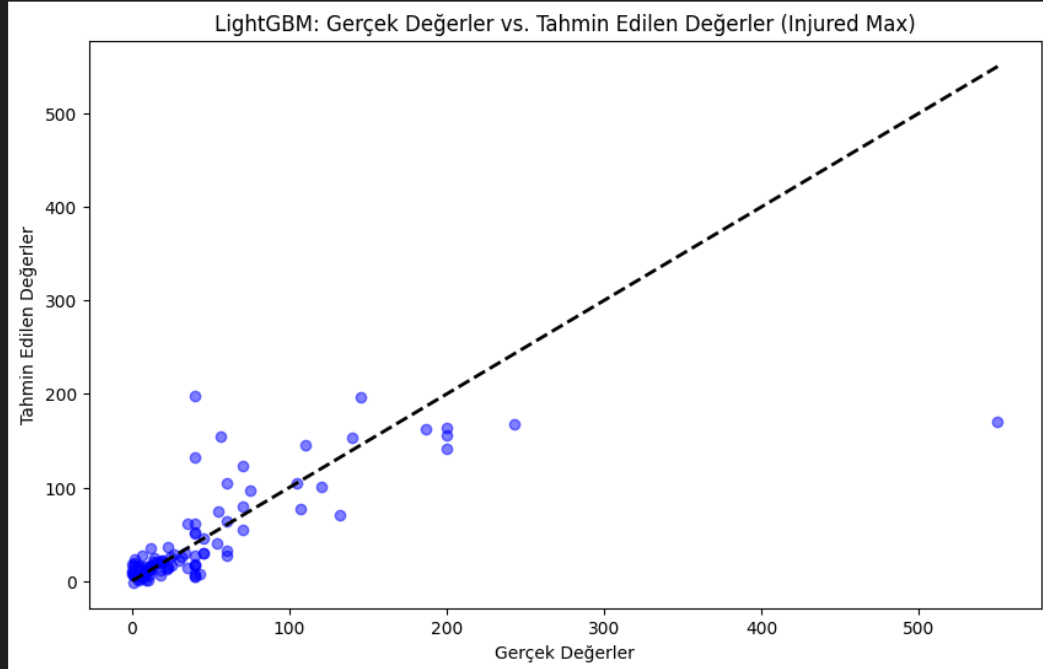
# LGBMR Regression

Mean Squared Error: 2350.092113618329

R-Squared: 0.5410065319554351

Mean Absolute Error: 21.803796987904814

Output is truncated. View as a [scrollable element](#) or open in a [text editor](#). Adjust cell output [settings...](#)

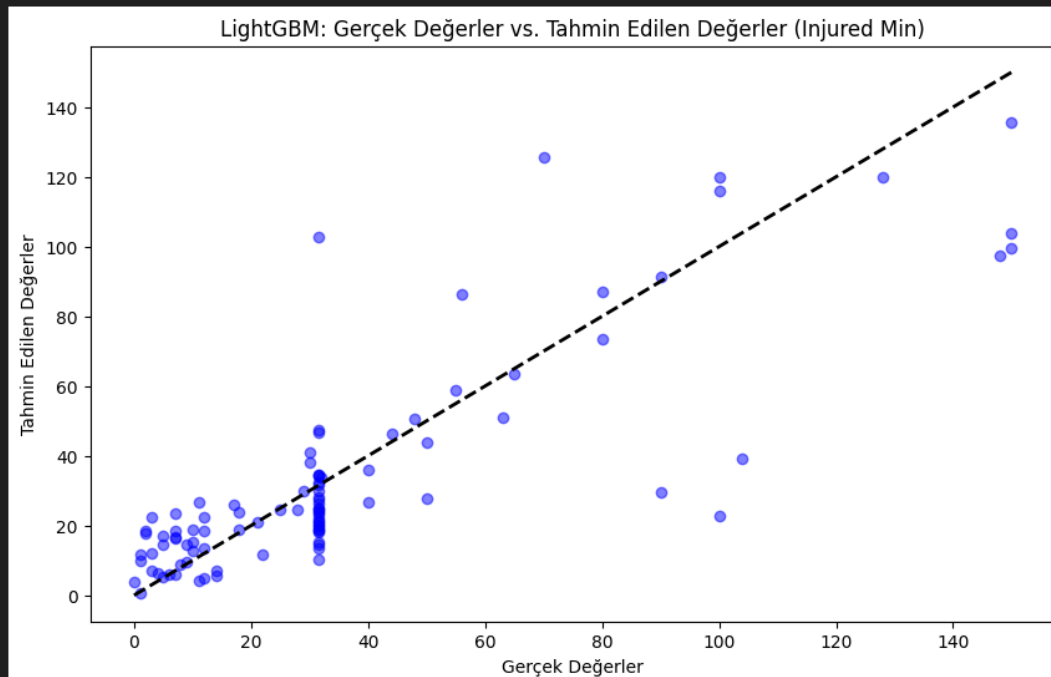


Mean Squared Error: 395.575694967587

R-Squared: 0.6807099127266949

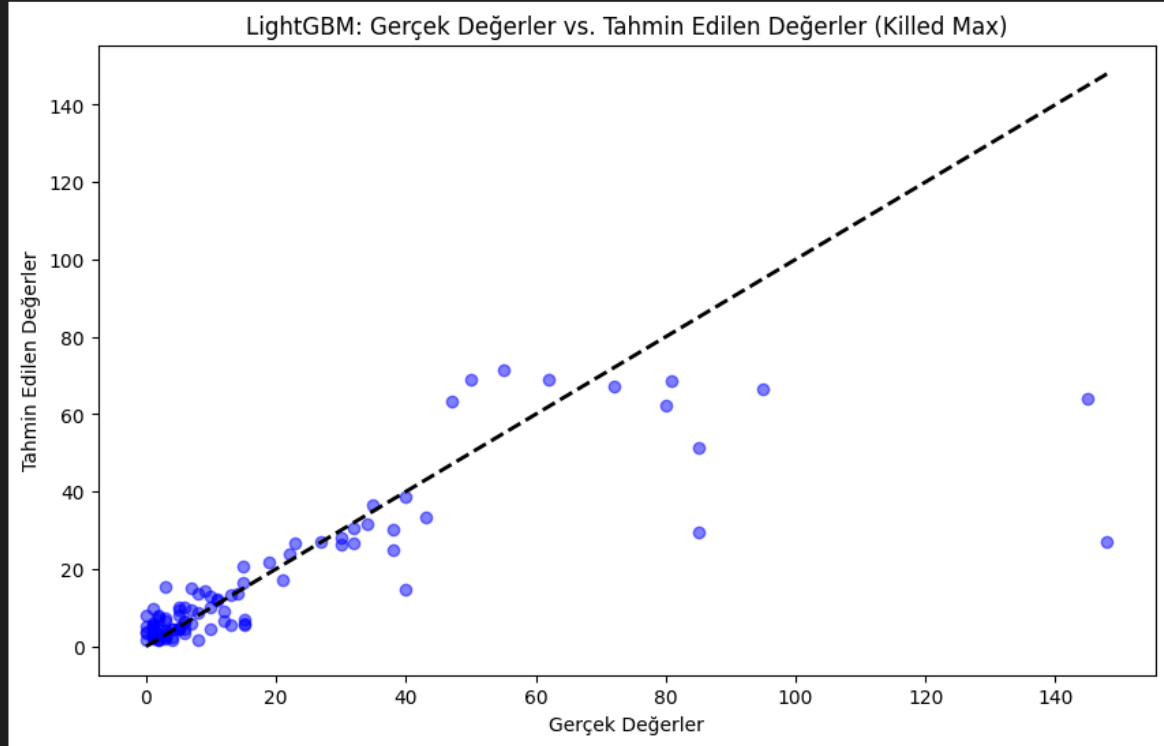
Mean Absolute Error: 12.419241735685711

Output is truncated. View as a [scrollable element](#) or open in a [text editor](#). Adjust cell output [settings...](#)



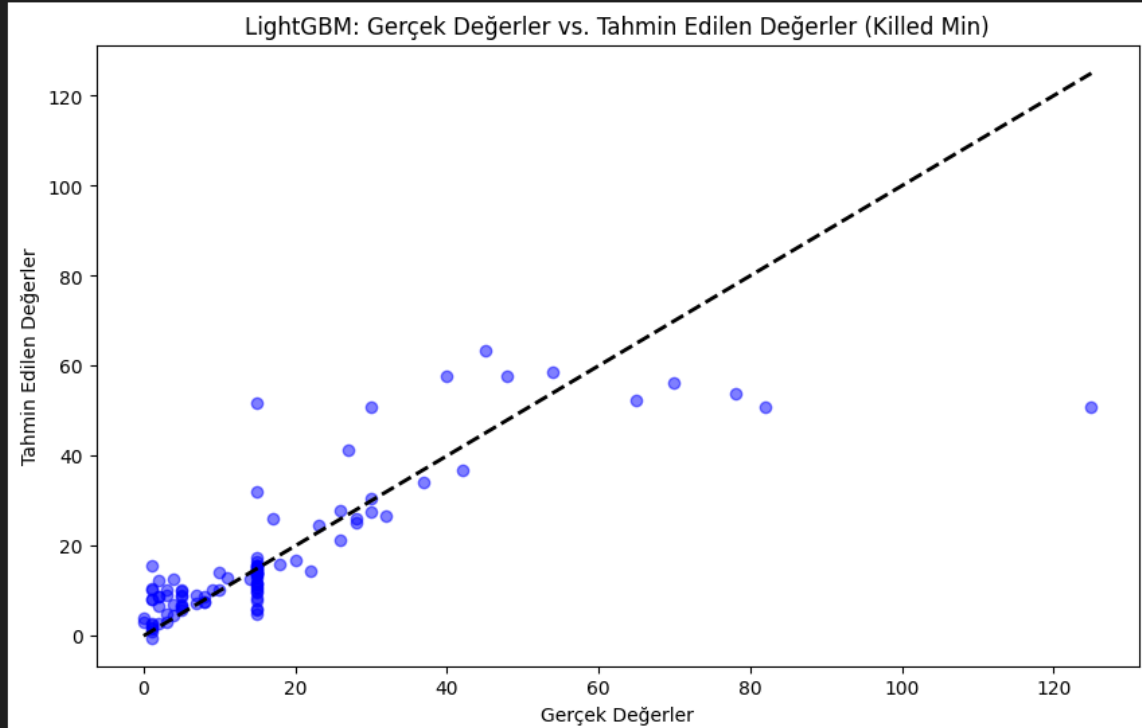
Mean Squared Error: 305.9164760842947  
R-Squared: 0.6254393963757667  
Mean Absolute Error: 7.376718835006277

Output is truncated. View as a [scrollable element](#) or open in a [text editor](#). Adjust cell output [settings...](#)



Mean Squared Error: 124.93497793809327  
R-Squared: 0.6814318251476119  
Mean Absolute Error: 5.981004178219224

Output is truncated. View as a [scrollable element](#) or open in a [text editor](#). Adjust cell output [settings...](#)



## Sonuç

Bu çalışmada terörizm eylemlerinde olan intihar bombacılarının kendileri patlatmaları ile oluşacak faciada belli oranlarda insan kaybının oluşacağı öngörülmüştür. Bu öngörüye binaen ilgili önlemlerin hayati önem arz etmektedir.

Yapılan modelleme ve analizler, belirli faktörlerin intihar bombalamalarının gerçekleşme olasılığını ve etkisini nasıl etkilediğini ortaya koymuştur. Modellerin performansını ve tahmin gücünü değerlendirmede önemli ölçütler olan Mean Squared Error, R-Squared ve Mean Absolute Error değerlerine dayanarak XGBoost ve Random Forest gibi gelişmiş regresyon teknikleri kullanılarak yapılan tahminler ve elde edilen , can kayıplarının öngörülmesinde belirli bir doğruluk seviyesine ulaşmıştır. Ancak istenilen doğruluk seviyesinin altında kalmıştır. Bu noktada dah iyi bir sonuç elde etmek için sayısal veriler ile çalışan regresyon modelleri yerine kategorik verilerin de içinde bulunacağı veri setleri ile farklı modellerin denenmesi elde edilecek doğruluk seviyesini arttıracaktır.

Bu tarz çalışmalar ile elde edilecek bulgular, güvenlik önlemlerinin ve stratejilerinin geliştirilmesin önemli bir rol oynayabilir. Bu gelişmelerin neticesinde olası intihar saldırılarının tespit edilip insan kaybının minimize edilmesi huzur ve barış dolu bir Dünya'nın varlığını mümkün kılacaktır.