AI-Powered Air Quality Prediction for Sustainable Cities

Ahmet Sevban Güler - 21069504

Serhat Keskintürk - 21069503

Tayfun Ekentok – 23069605

Yusuf Mehmet Arslan - 21069501

YTÜ

About the Project

- We developed an AI model to predict air pollution in 11 cities in the Marmara Region.
- We used daily data from 2015 to 2024, including six pollutants:

  PM10, SO2, NO2, NOX, NO, and O3.
- Our goal is to support better decision-making for cleaner and healthier cities.

Sustainable Development Goals (SDGs)

This project supports these Sustainable Development Goals:

- SDG 3:

Good Health and Well-being

- SDG 11:

Sustainable Cities and Communities

- SDG 13:

Climate Action

Our Dataset

- Source: 11 air quality stations (2015–2024)

- Each data point contains data for every day of 10 years:
  - City name
  - Date
  - PM10, SO2, NO2, NOX, NO, O3 values
- Total size: 40,184 rows

Data Source:

We downloaded the air quality data from the official government website:

https://sim.csb.gov.tr/STN/STN_Report/StationDataDownloadNew

Provided by: T.C. Ministry of Environment, Urbanization and Climate Change

REPUBLIC OF TÜRKİYE
MINISTRY OF ENVIRONMENT,
URBANIZATION AND CLIMATE CHANGE

Data Preprocessing

- Data cleaning
- Feature engineering
- Filling missing values
- Adding new features

Air Pollution Score

- We made one single score combining PM10, SO2, NO2, NOX, NO, and O3.
- We used **weights** based on pollutants' health impacts:

%40	%20	%15	%10	%5	%10
PM_{10}	**SO_2**	**NO_2**	**NOX**	**NO**	**O_3**

Data Preparation for Model

- We selected features (X) and target (y).
- We split the data:

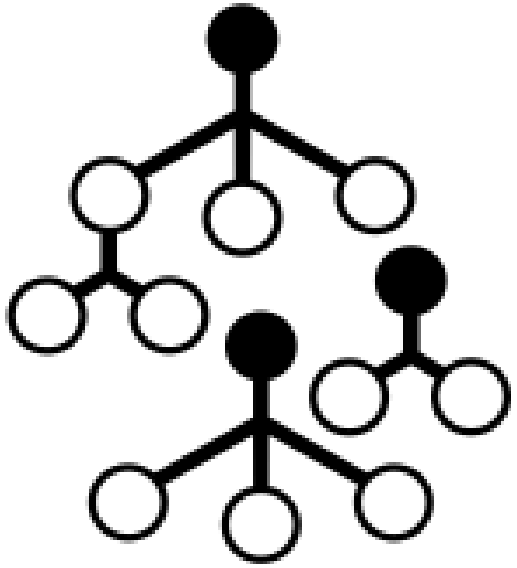
  80% for training.

  20% for testing.

```
# Hedef değişken (y)
y = df['Pollution_Score']

# Kullanmayacağımız kolonlar
drop_cols = [
    'İstasyon', 'Tarih',
    'Pollution_Score'
]

# Özellikler (X) → Geri kalan feature'lar
X = df.drop(columns=drop_cols)
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.8, random_state=42)
```

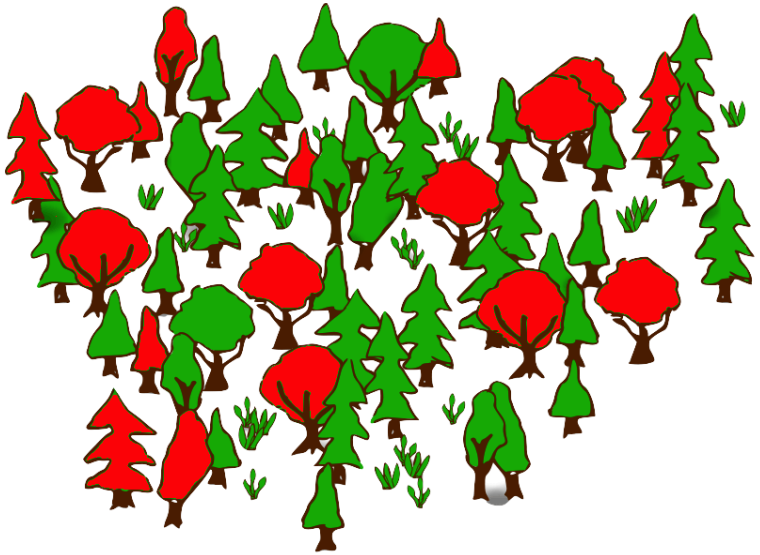
AI Model Selection



- We used **Random Forest Regressor**.
- This model is good for prediction problems.
- We trained the model with our training data.

```
[ ]  rf_model = RandomForestRegressor(n_estimators=100, random_state=42)

     # Modeli eğitiyoruz
     rf_model.fit(X_train, y_train)
```

```
RandomForestRegressor
RandomForestRegressor(random_state=42)
```

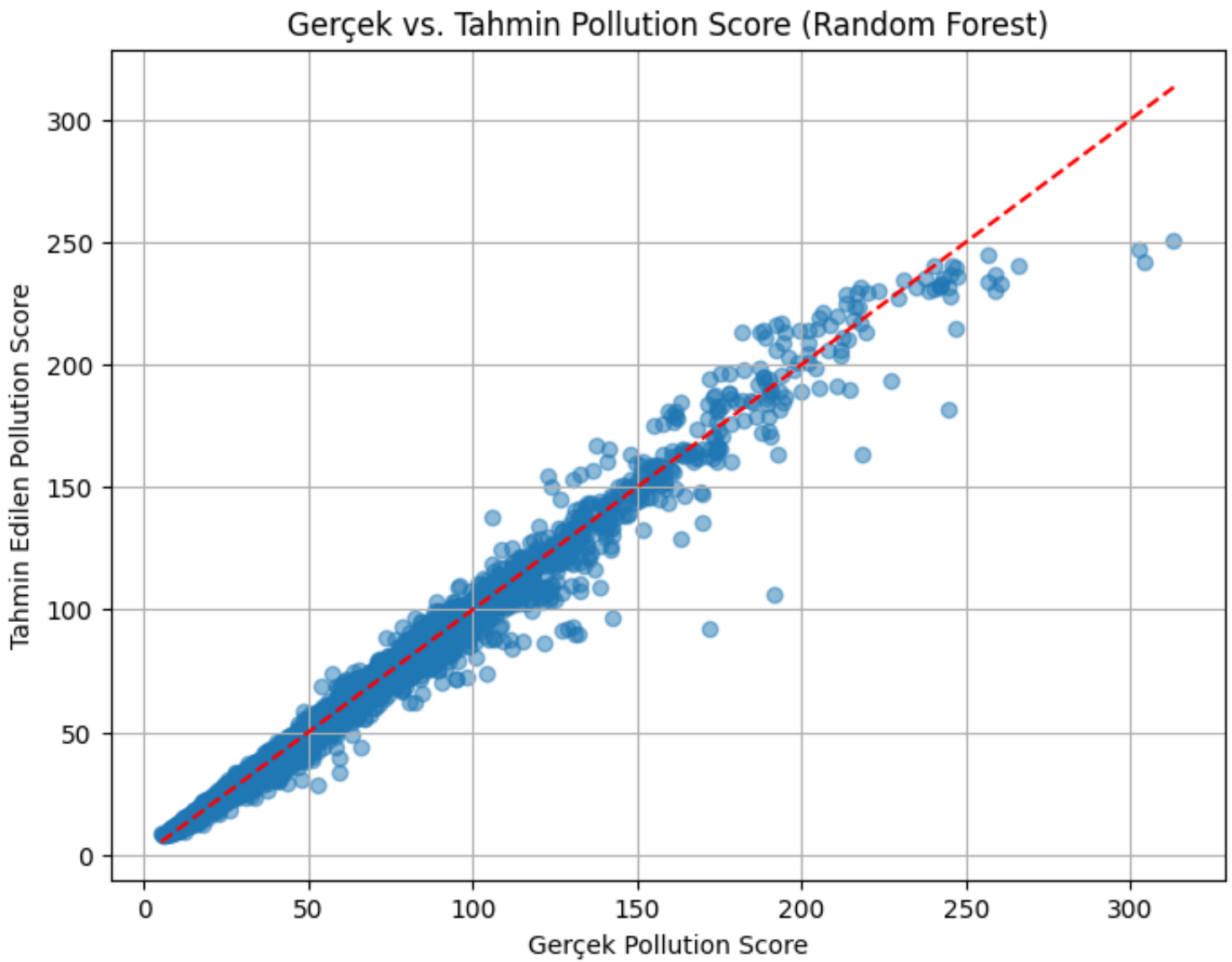


Model Evaluation

- We measured the model performance with:
  - Mean Squared Error (MSE)
  - Mean Absolute Error (MAE)
  - R^2 Score

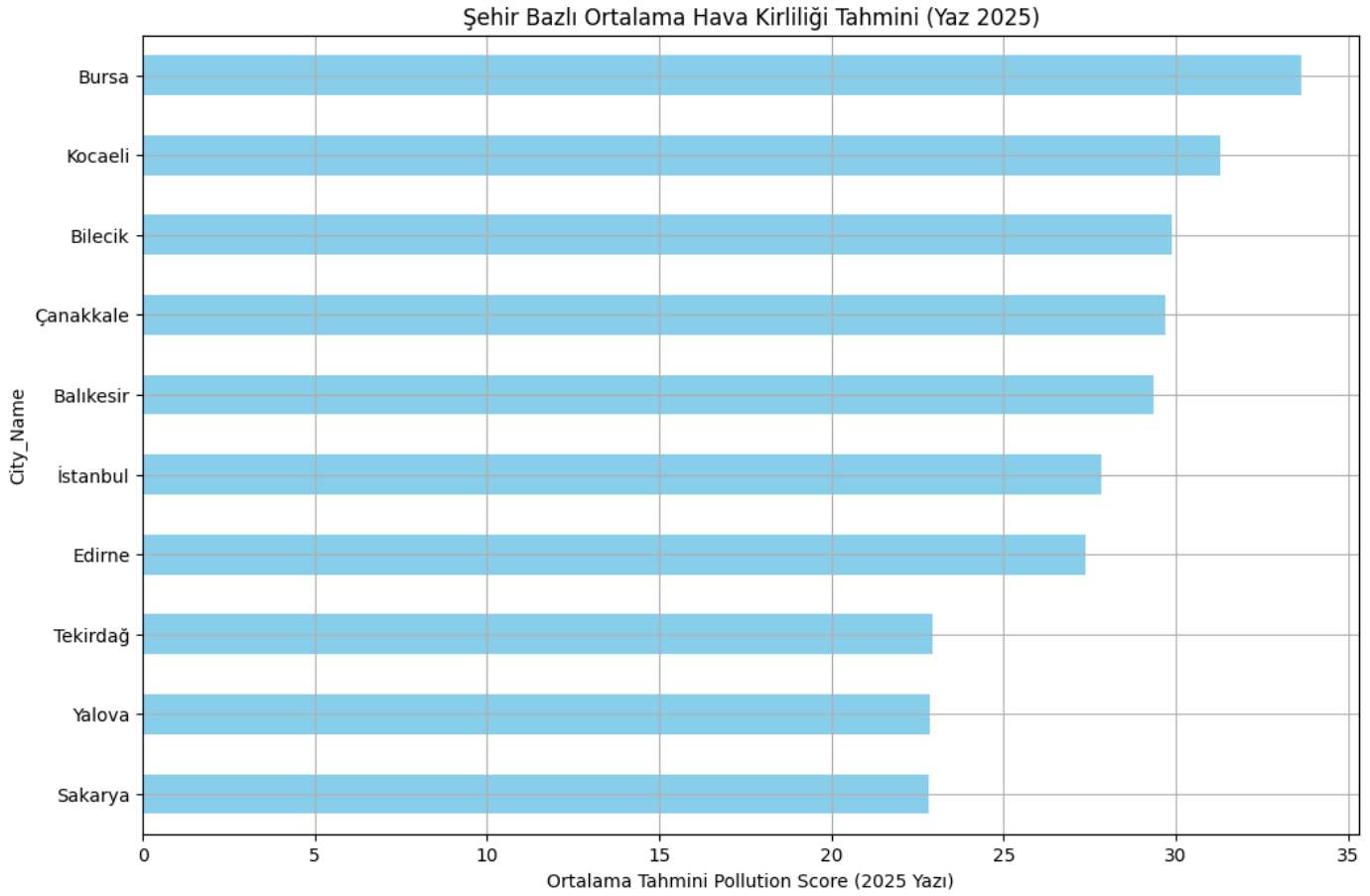
```python
# Test verisinde tahmin yapıyoruz
y_pred = rf_model.predict(X_test)

# Performans metriklerini hesaplıyoruz
mse = mean_squared_error(y_test, y_pred)
mae = mean_absolute_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f"Mean Squared Error (MSE): {mse:.2f}")
print(f"Mean Absolute Error (MAE): {mae:.2f}")
print(f"R² Score: {r2:.4f}")
```

```
Mean Squared Error (MSE): 5.64
Mean Absolute Error (MAE): 1.08
R² Score: 0.9889
```


Gerçek vs. Tahmin Pollution Score (Random Forest)
Tahmin Edilen Pollution Score
Gerçek Pollution Score
0
50
100
150
200
250
300

Future Predictions

- A dataset was created for the summer months of 2025 (June-July-August).
- The model estimated daily Pollution Score for each city.
- The results were averaged on a city basis and shown graphically.



Application Demo Interface

- Web interface / mobile compatible system
- Simple city and date selection
- Color-coded score display
- A system that allows institutions to monitor and take precautions against air pollution.

Our Difference

- Forecast-oriented (not real-time, forward-looking)
- Local data – specific to Marmara
- Open source and transparent structure
- Can be used without technical knowledge

What Can Be Improved?

Limitations:

- Only one time of day used (00:56 AM)
- No weather data included
- One model used for all cities

Future Plans:

- Add weather features (wind, temperature)
- Use time-series models like LSTM
- Build city-specific models

Conclusion

- Air pollution can be predicted with machine learning.
- Early warnings can help cities take action.
- Our project helps with public health, city planning, and climate goals.