

## # 변수 설명

### 1. Temperature[C]\*\* (온도):

- \*\*설명\*\*: 이 변수는 연기 감지기가 위치한 환경의 온도를 섭씨(°C)로 표현
- \*\*의미\*\*: 화재 발생 시 온도가 급격히 상승할 수 있기 때문에, 온도는 중요한 지표. 이 변수는 화재 발생 가능성을 평가하는 데 중요한 역할.

### 2. Humidity[%]\*\* (습도):

- \*\*설명\*\*: 이 변수는 연기 감지기 주변 환경의 습도를 백분율(%)로 표현
- \*\*의미\*\*: 습도는 공기 중의 수분량을 나타내며, 이는 연기 감지기와 화재 탐지에 영향을 미칠 수도 있다. 예를 들어, 높은 습도는 연기 입자의 확산에 영향을 줄 수 있음

### 3. TVOC[ppb]\*\* (Total Volatile Organic Compounds):

- \*\*설명\*\*: 이 변수는 총 휘발성 유기 화합물(TVOC)의 농도를 parts per billion (ppb) 단위로 표기
- \*\*의미\*\*: TVOC는 여러 가지 유기 화합물이 포함된 혼합물을 나타내며, 화재 발생 시 이러한 화합물의 농도가 증가할 수도 있음. 이는 화재의 주요 지표 중 하나

### 4. eCO2[ppm]\*\* (Equivalent CO2):

- \*\*설명\*\*: 이 변수는 이산화탄소(CO2)의 농도를 parts per million (ppm) 단위로 표기
- \*\*의미\*\*: eCO2는 연소 과정에서 발생하는 대표적인 가스로, 화재 발생 시 농도가 급격히 증가 가능. 따라서 이 변수는 화재 탐지에 매우 중요한 역할

### 5. Raw H2\*\* (수소 농도):

- \*\*설명\*\*: 이 변수는 연기 감지기 주변의 수소(H2) 농도를 나타냅니다.
- \*\*의미\*\*: 수소는 가연성 가스로, 특정 연소 과정에서 발생 가능. 수소 농도는 화재 여부를 판단하는 중요한 지표가 될 수도 있음.

### 6. Raw Ethanol\*\* (에탄올 농도):

- \*\*설명\*\*: 이 변수는 연기 감지기 주변의 에탄올 농도
- \*\*의미\*\*: 에탄올 역시 가연성 물질로, 화재 시 농도가 높아질 수 가능성. 따라서 화재 발생의 초기 징후를 포착하는 데 유용할 수 있음

7. Pressure[hPa]\*\* (대기압):

- **\*\*설명\*\***: 이 변수는 연기 감지기 주변의 대기압을 헥토파스칼(hPa) 단위로 표기
- **\*\*의미\*\***: 대기압은 연기와 가스의 확산에 영향을 줄 수 있으며, 기상 조건과 화재 발생 간의 연관성을 분석하는 데 도움

8. PM1.0\*\* (미세먼지 농도):

- **\*\*설명\*\***: 이 변수는 1.0 마이크로미터 이하의 입자 농도를 나타내며,  $\mu\text{g}/\text{m}^3$  단위로 측정
- **\*\*의미\*\***: PM1.0 농도는 매우 작은 미세먼지 입자의 양을 나타내며, 연기 입자를 감지하는 데 중요한 역할, 화재 발생 시 이 수치가 증가할 수 있음

9. **\*\*PM2.5\*\*** (미세먼지 농도):

- **\*\*설명\*\***: 이 변수는 2.5 마이크로미터 이하의 입자 농도를 나타내며,  $\mu\text{g}/\text{m}^3$  단위로 측정
- **\*\*의미\*\***: PM2.5 농도는 화재 시 발생하는 연기 입자의 양을 나타낼 수 있으며, 건강에 해로운 입자들을 포함할 수 있음 □

10. NC0.5\*\*, **\*\*NC1.0\*\***, **\*\*NC2.5\*\*** (입자 개수 농도):

- **\*\*설명\*\***: 이 변수들은 각각 0.5 $\mu\text{m}$ , 1.0 $\mu\text{m}$ , 2.5 $\mu\text{m}$  크기의 입자 개수를 나타냄
- **\*\*의미\*\***: 이 변수들은 공기 중에 떠다니는 매우 작은 입자들의 농도를 나타내며, 화재 발생 시 이들의 농도가 급격히 증가할 수 있음

11. Fire Alarm\*\* (화재 경보):

- **\*\*설명\*\***: 이 변수는 화재 경보의 상태를 나타내며, 0은 경보 없음, 1은 경보 발생을 의미
- **\*\*의미\*\***: 이 변수는 모델링에서 "종속 변수"로 사용됩니다. 즉, 이 값을 예측하는 것이 모델의 주요 목표

이 데이터셋은 화재 감지에 관련된 다양한 환경적, 화학적 변수를 포함하고 있습니다. 각 변수는 화재 발생 시 나타나는 변화를 포착할 수 있으며, 특히 연소 과정에서 발생하는 가스, 온도, 습도, 미세먼지 등의 변화를 측정합니다.

이 변수들은 종속 변수인 `Fire Alarm`의 상태를 예측하는 데 사용되며, 이는 모델의 정확한 화재 예측을 가능하게 합니다.

## #Smoke Detecton 코드 리뷰

### 1. 라이브러리 불러오기

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import matplotlib.dates as mdates
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, accuracy_score
```

- 위 코드 블록은 데이터를 처리, 시각화, 머신러닝 모델 구축을 위한 라이브러리 불러옴
- pandas는 데이터를 다루는 데, matplotlib과 seaborn은 데이터 시각화에, sklearn은 데이터 분할, 모델 학습 및 평가를 만드는 데 사용

**발표 설명:** "여기서는 데이터를 불러오고 처리하며, 시각화하고, 모델을 학습시키기 위해 필요한 라이브러리들을 불러옵니다. 이 도구들은 데이터 분석의 핵심적인 역할을 하며, 프로젝트의 모든 단계에서 사용됩니다."

### 2. 데이터 불러오기 및 전처리 '

```
file_path = '/Users/joon/Documents/GitHub/Final_Project/data/1.smoke_detection_iot'
df = pd.read_csv(file_path)

df = df.drop(columns=['Unnamed: 0'])
df['UTC'] = pd.to_datetime(df['UTC'], unit='s')
df['Hour'] = df['UTC'].dt.hour
df['Day'] = df['UTC'].dt.day
df['Month'] = df['UTC'].dt.month
df['Year'] = df['UTC'].dt.year
```

- **Unnamed: 0:** 이 열은 CSV 파일에서 데이터를 불러올 때 생성되는 인덱스 역할을 하는 열. 이미 Pandas 데이터프레임에서는 별도의 인덱스가 관리되기 때문에, Unnamed: 0은 불필요하다고 판단되어 제거.
- **시간데이터 변환:** 숫자로 된 "165473331" 같은 형태(UTC: 유닉스 타임스탬프)를 "2023년 8월 10일, 오전 10시"처럼 **알기 쉽게 변형, 더 알기 쉽게** "년도", "월", "일", "시간"으로 변경 (연기 감지기 데이터를 불러온 후, 분석에 필요 없는 부분을 제거하고, 시간 데이터를 연도, 월, 일, 시간으로 나누어 이해하기 쉽게 변환. 변환된 시간 데이터는 이후 분석에서 시간 패턴을 이해하는 데 중요한 역할)

**발표 설명:** "데이터를 불러와서 분석에 필요하지 않은 열을 제거한 후, 시간 데이터를 사람이 이해할 수 있는 날짜와 시간 형식으로 변환했습니다. 또한, 연도, 월, 일, 시간 등 시간과 관련된 정보를 추가로 분리해 분석에 활용할 수 있도록 준비했습니다."

### 3. 탐색적 데이터 분석 (EDA) 및 상관관계 분석

```
plt.figure(figsize=(14, 10))
sns.heatmap(df.corr(), annot=True, fmt=".2f", cmap='coolwarm', cbar=True)
plt.title("Feature Correlation Matrix", fontsize=16)
plt.xticks(fontsize=12, rotation=45)
plt.yticks(fontsize=12)
plt.show()
```

- 모든 변수 간의 상관관계를 시각화한 히트맵을 생성하여, 변수들 간의 관계를 파악
- 상관관계 행렬은 데이터셋의 구조를 이해하는 데 중요한 역할을 하며, 모델링에 앞서 어떤 변수들이 서로 밀접하게 관련되어 있는지 확인가능

**발표 설명:** "이 히트맵은 데이터셋 내 모든 변수들 간의 상관관계를 시각적으로 보여줍니다. 이 정보를 통해 어떤 변수들이 서로 강한 연관성을 가지는지, 그리고 이러한 연관성이 화재 경보 예측에 어떻게 영향을 미칠 수 있는지를 이해할 수 있습니다."

### 4. 시간에 따른 모든 변수의 시각화

```
variables = ['Temperature[C]', 'Humidity[%]', 'TVOC[ppb]', 'eCO2[ppm]', 'Raw H2',

plt.figure(figsize=(12, 8))
for var in variables:
    plt.plot(df['UTC'], df[var], label=var)

plt.xlabel('Time')
plt.ylabel('Values')
plt.title('Environmental Variables Over Time')
plt.legend(loc='upper right')

plt.gca().xaxis.set_major_formatter(mdates.DateFormatter('%Y-%m-%d %H:%M'))
plt.gca().xaxis.set_major_locator(mdates.DayLocator(interval=1))
plt.gca().xaxis.set_minor_locator(mdates.HourLocator(interval=6))

plt.xticks(rotation=45)
plt.show()
```

- 시간(UTC)에 따른 주요 환경 변수들의 변화를 시각화하여, 각 변수들이 시간이 지남에 따라 어떻게 변동하는지를 한눈에 파악 가능
- DateFormatter 와 Locator 를 사용하여 시간 축을 보다 직관적이고 읽기 쉽게 설정

**발표 설명:** "이 그래프는 시간에 따라 주요 환경 변수들이 어떻게 변하는지를 보여줍니다. 각 변수들은 시간의 흐름에 따라 달라지며, 이 변화를 통해 특정 시간대에 화재와 관련된 중요한 패턴을 발견할 수 있습니다."

## 5. 모델링 및 평가

```
X = df.drop(columns=['Fire Alarm', 'UTC'])
y = df['Fire Alarm']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

model = RandomForestClassifier(random_state=42)
model.fit(X_train, y_train)

y_pred = model.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
report = classification_report(y_test, y_pred)

print(f"Model Accuracy: {accuracy:.2f}")
print("Classification Report:\n", report)
```

- Fire Alarm 을 예측하기 위해 랜덤 포레스트 모델을 사용하여 학습을 진행합니다. 이 과정에서 UTC 열을 제외한 독립 변수를 사용
- 모델의 정확도와 분류 성능을 평가하여, 모델이 화재 경보를 얼마나 잘 예측하는지 확인

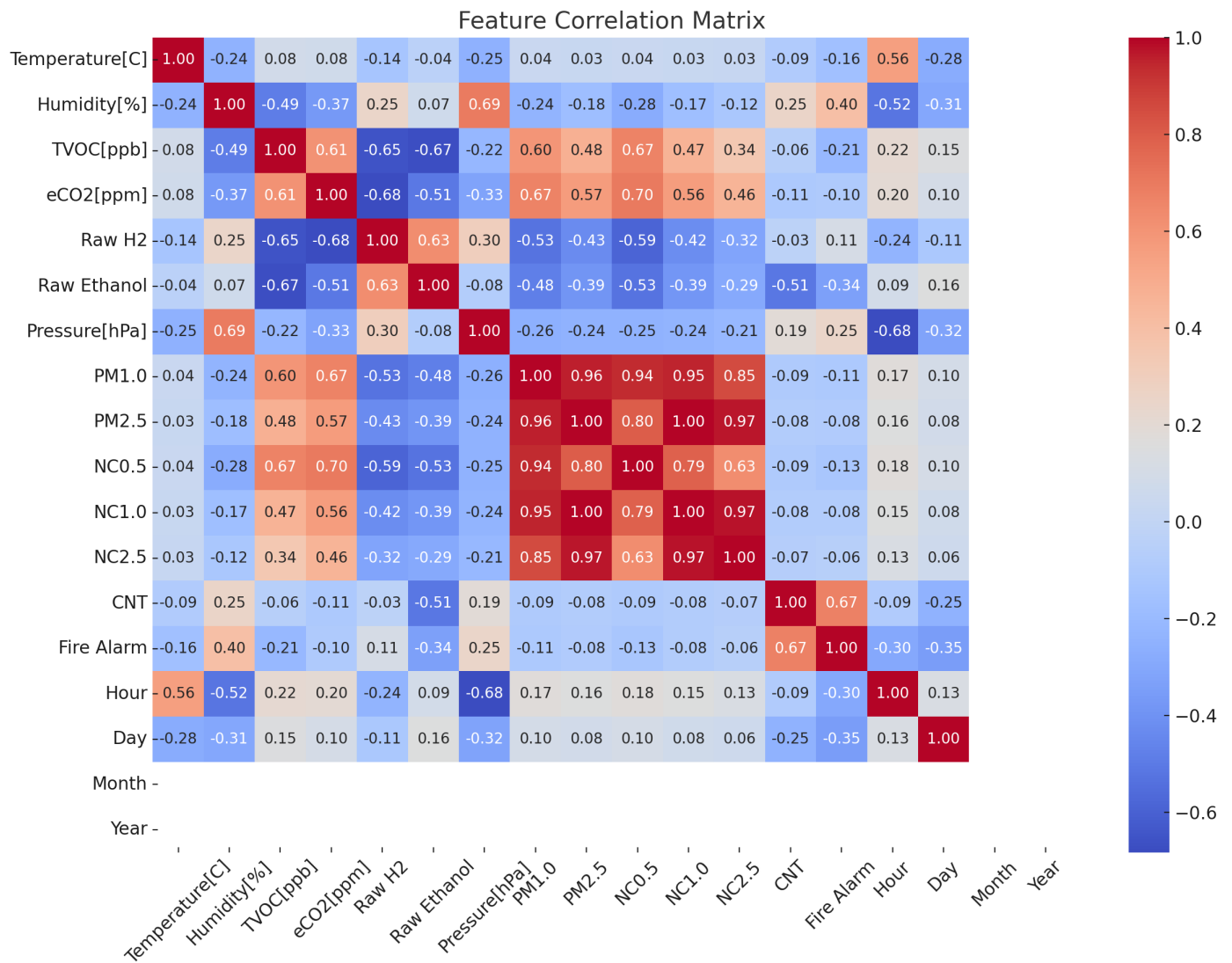
**발표 설명:** "랜덤 포레스트 모델을 사용하여 화재 경보를 예측합니다. 우리는 데이터를 학습용과 테스트용으로 나누어 모델을 훈련시켰으며, 이후 테스트 데이터에 대한 모델의 예측 정확도와 성능을 평가했습니다. 이를 통해 모델이 얼마나 잘 예측하는지 확인할 수 있습니다."

## 6. 특성 중요도 시각화

```
importances = model.feature_importances_  
indices = importances.argsort()[::-1]  
  
plt.figure(figsize=(12, 8))  
plt.title("Feature Importances", fontsize=16)  
plt.bar(range(X.shape[1]), importances[indices], align="center", color='skyblue')  
plt.xticks(range(X.shape[1]), X.columns[indices], rotation=90, fontsize=12)  
plt.ylabel('Importance', fontsize=14)  
plt.xlabel('Features', fontsize=14)  
plt.xlim([-1, X.shape[1]])  
plt.show()
```

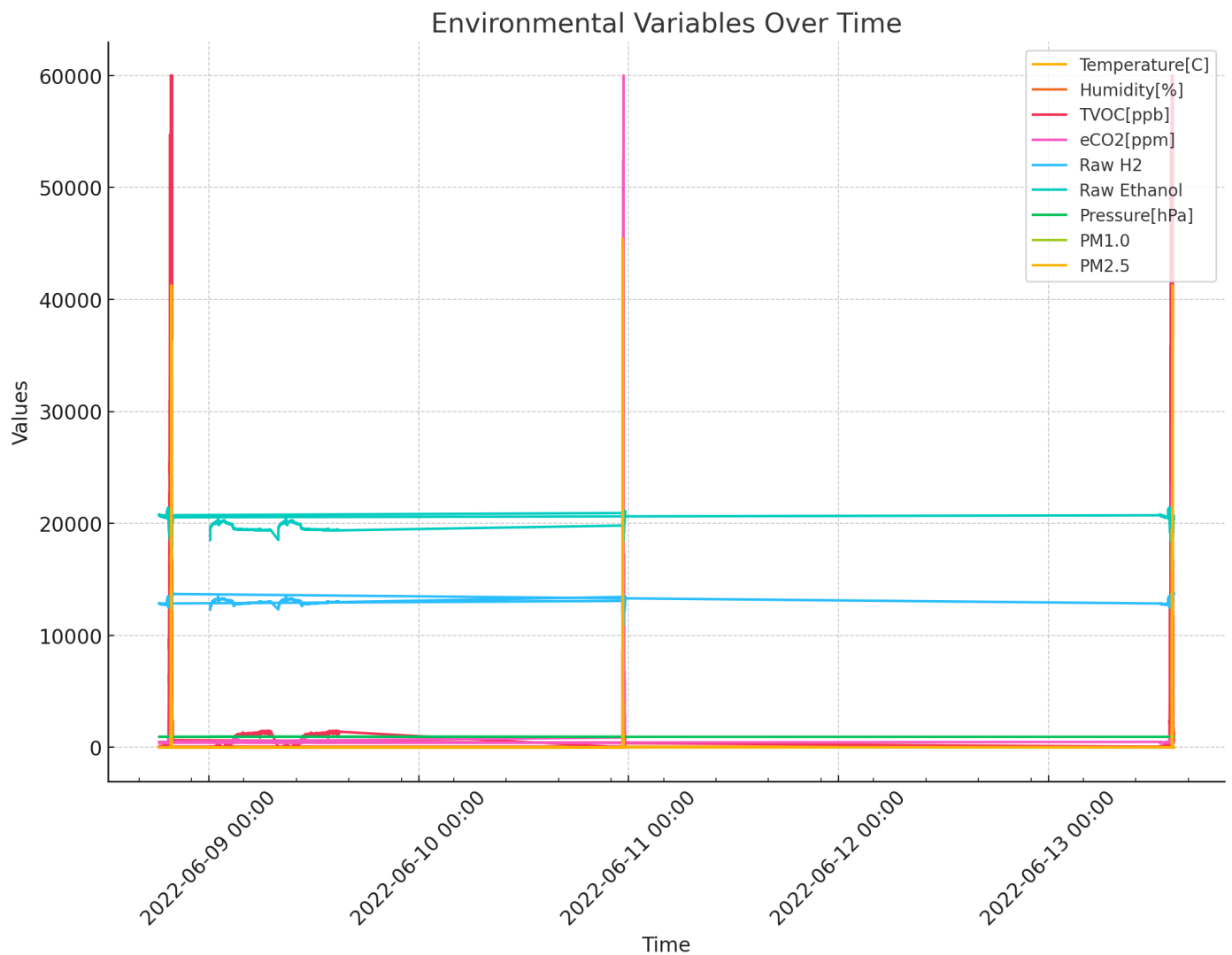
- 랜덤 포레스트 모델에서 각 변수가 예측에 얼마나 중요한 역할을 하는지 시각적으로 표현
- 이를 통해 모델이 어떤 변수를 중점적으로 사용했는지, 즉 어떤 변수가 화재 발생 예측에서 중요한 역할을 하는지 파악가능

**발표 설명:** "이 그래프는 모델이 예측할 때 어떤 변수를 가장 중요하게 사용하는지를 보여줍니다. 중요한 변수일수록 막대의 길이가 길며, 이는 모델이 해당 변수를 예측에 있어 얼마나 중요한 요소로 간주하는지를 나타냅니다."



## 상관계수 행렬 (Correlation Matrix)

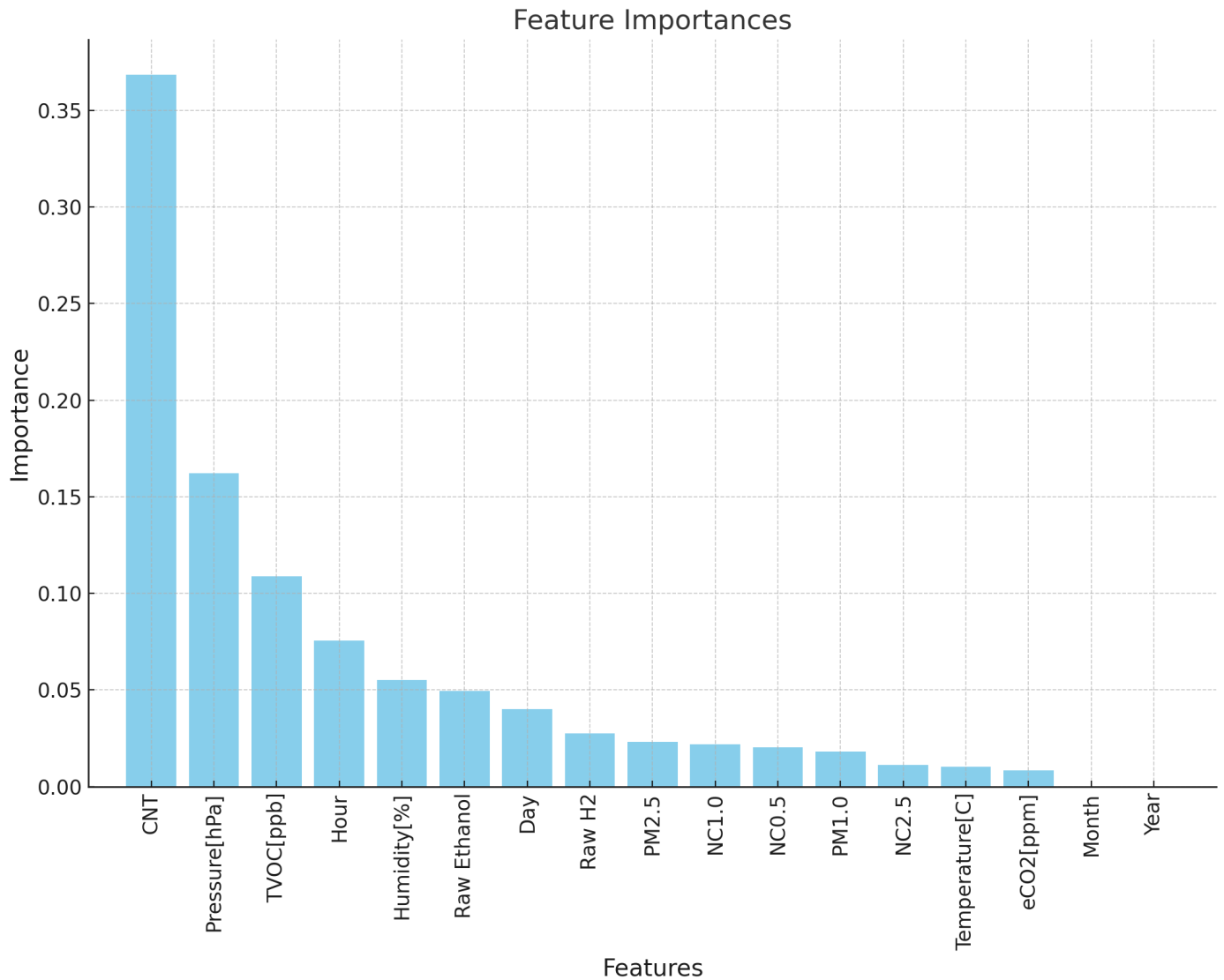
- 설명:** 이 그래프는 변수들 간의 상관관계를 시각화한 것입니다. 서로 관련이 강한 변수들은 동일한 패턴으로 변화할 가능성이 높습니다.
- 사용 예시:** 예를 들어, TVOC와 eCO2가 강한 양의 상관관계를 가진다면, 화재 발생 시 두 변수 모두 증가할 가능성이 있습니다.



## 2. 시간에 따른 모든 변수의 시각화 (Environmental Variables Over Time)

- 설명:** 이 그래프는 시간에 따라 각 변수들이 어떻게 변화하는지를 보여줍니다. 각각의 선은 특정 환경 변수를 나타내며, 시간에 따른 변화를 시각적으로 쉽게 이해할 수 있습니다.
- 사용 예시:** 특정 시간대에 TVOC와 eCO2 농도가 급격히 증가하는 패턴이 보인다면, 이는 화재 발생의 징후일 수 있습니다.





## 2. 특성 중요도 (Feature Importances)

- **설명:** 이 그래프는 랜덤 포레스트 모델이 예측을 할 때 어떤 변수를 중요하게 여겼는지를 시각적으로 보여줍니다. 막대가 길수록 해당 변수가 모델의 예측에 중요한 역할을 했다는 것을 의미합니다.
- **사용 예시:** 예를 들어, TVOC와 eCO2가 중요 변수로 나타났다면, 이는 화재 예측에서 이 두 변수가 중요한 역할을 한다는 것을 시사합니다.

### 모델 평가 결과

- **정확도(Accuracy):** 100%
  - 이 모델은 테스트 데이터의 모든 샘플을 정확하게 예측했습니다.
- **분류 보고서(Classification Report):**
  - **정밀도(Precision), 재현율(Recall), F1 점수(F1-Score)** 모두 1.00으로, 모델이 예측을 매우 잘 수행했다는 것을 나타냅니다.

# 연기 감지 및 화재 경보 예측

(AI로 생성한 프로젝트 관련 이미지 추가예정)

## 프로젝트 개요

본 프로젝트의 주요 목적은 IOT(사물인터넷) 장치를 통해 수집된 데이터를 사용하여 고성능 연기 감지 시스템을 개발하는 것입니다. 이를 통해 화재를 조기에 발견하고 대응함으로써 인명과 재산 피해를 최소화하고, 공공 안전을 증진시키고자 합니다.

## 현 상황 분석

- 기존 연기 감지기의 한계: 오작동이 잦고, 반응 시간이 늦으며, 특정 환경에서는 신뢰성이 낮음
- 지속적인 대형 화재 사고 발생으로 인한 사회적 우려 증가

## 기대 효과

- 높은 정확도와 빠른 반응 시간을 갖춘 연기 감지 시스템 구현
- 다양한 환경 및 상황에서의 화재 조기 발견 능력 향상
- 오작동 감소로 인한 소방력 낭비 방지
- 화재 관련 인명 및 재산 피해 예방

본 프로젝트를 통해 개발될 딥러닝 기반 연기 감지 시스템은 기존 화재 감지 기술의 한계를 극복하고, 더욱 안전한 사회 구현에 기여할 것으로 기대됩니다.

## 데이터 설명

- 데이터는 약 60,000 개의 측정값으로 구성되었으며, 모든 센서의 샘플링 속도는 1Hz 입니다.
- 주요 변수:
  - **Temperature[C]**: 온도 (섭씨)
  - **Humidity[%]**: 습도 (백분율)
  - **TVOC[ppb]**: 총 휘발성 유기 화합물 농도 (ppb)
  - **eCO2[ppm]**: 이산화탄소 농도 (ppm)
  - **Raw H2**: 수소 농도
  - **Raw Ethanol**: 에탄올 농도
  - **Pressure[hPa]**: 대기압 (hPa)
  - **PM1.0, PM2.5**: 미세먼지 농도 (µg/m³)
  - **Fire Alarm**: 화재 경보 (0: 경보 없음, 1: 경보 발생)

## 모델링 및 분석

- 데이터 전처리:

- 불필요한 열을 제거하고, UTC 시간을 연도, 월, 일, 시간 등으로 분리하여 분석에 용이하게 변환했습니다.
- **탐색적 데이터 분석 (EDA):**
  - 상관관계 행렬을 통해 각 변수들 간의 관계를 분석했습니다. 예를 들어, 특정 가스 농도와 온도 사이의 강한 상관관계를 확인할 수 있었습니다.
  - 시간에 따른 주요 환경 변수들의 변화를 시각화하여, 시간이 지남에 따라 변수들이 어떻게 변동하는지 파악했습니다.
- **모델링:**
  - 랜덤 포레스트 모델을 사용하여 Fire Alarm 을 예측했습니다. 모델은 100%의 정확도로 테스트 데이터를 예측했습니다.
  - 모델의 분류 보고서에서는 정밀도, 재현율, F1 점수가 모두 1.00 으로 나타났으며, 이는 모델이 화재 경보를 매우 정확하게 예측한다는 것을 의미합니다.
- **특성 중요도 분석:**
  - 랜덤 포레스트 모델을 통해 어떤 변수가 예측에 가장 중요한 역할을 하는지 시각적으로 확인했습니다. 예를 들어, TVOC 와 eCO2 가 중요한 변수로 나타났습니다.

## 결론

이 프로젝트는 IOT 기반 화재 예측 시스템의 가능성을 성공적으로 시연했습니다. 특히, 랜덤 포레스트 모델을 통해 높은 정확도의 예측을 달성했으며, 각 변수들의 중요성을 분석하여 화재 예측에 중요한 인사이트를 제공했습니다.

@ 오버샘플링 해야하는지, 데이터셋 0,1 분포 확인