

In [1]:

```
import pandas as pd
```

EDA (Exploratory Data Analysis)

In [2]:

```
df=pd.read_csv("smoke_detection_iot.csv")
df.head()
```

Out[2]:

named: 0	UTC	Temperature[C]	Humidity[%]	TVOC[ppb]	eCO2[ppm]	Raw H2	Raw Ethanol	Pressi
0	1654733331	20.000	57.36	0	400	12306	18520	
1	1654733332	20.015	56.67	0	400	12345	18651	
2	1654733333	20.029	55.96	0	400	12374	18764	
3	1654733334	20.044	55.28	0	400	12390	18849	
4	1654733335	20.059	54.69	0	400	12403	18921	

In [4]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 62630 entries, 0 to 62629
Data columns (total 16 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Unnamed: 0            62630 non-null  int64
1   UTC                   62630 non-null  int64
2   Temperature[C]       62630 non-null  float64
3   Humidity[%]          62630 non-null  float64
4   TVOC[ppb]            62630 non-null  int64
5   eCO2[ppm]            62630 non-null  int64
6   Raw H2                62630 non-null  int64
7   Raw Ethanol          62630 non-null  int64
8   Pressure[hPa]        62630 non-null  float64
9   PM1.0                62630 non-null  float64
10  PM2.5                62630 non-null  float64
11  NC0.5                62630 non-null  float64
12  NC1.0                62630 non-null  float64
13  NC2.5                62630 non-null  float64
14  CNT                  62630 non-null  int64
15  Fire Alarm           62630 non-null  int64
dtypes: float64(8), int64(8)
memory usage: 7.6 MB
```

In [5]:

```
df.isna().sum()
```

Out[5]:

```

Unnamed: 0      0
UTC             0
Temperature[C]   0
Humidity[%]      0
TVOC[ppb]       0
eCO2[ppm]       0
Raw H2          0
Raw Ethanol     0
Pressure[hPa]   0
PM1.0           0
PM2.5           0
NC0.5           0
NC1.0           0
NC2.5           0
CNT            0
Fire Alarm      0
dtype: int64

```

In [15]:

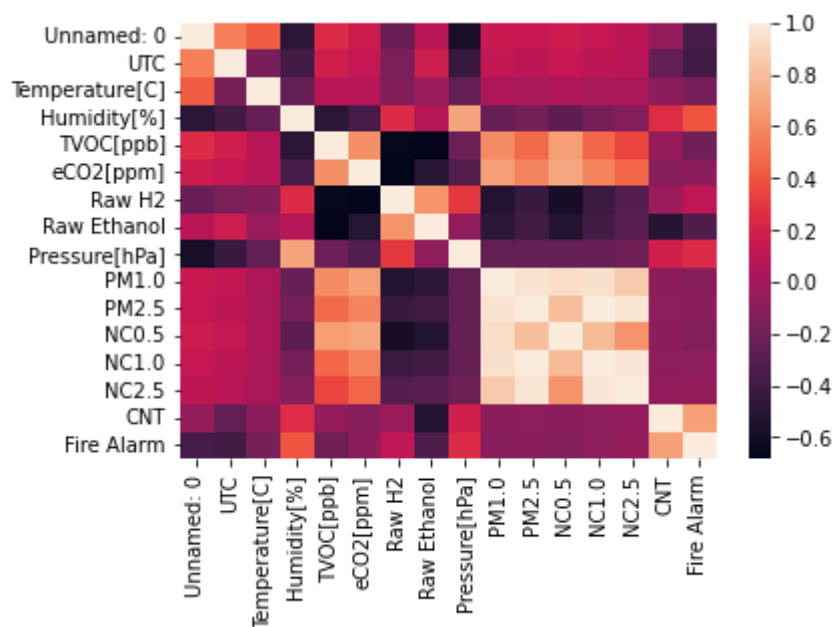
```
import seaborn as sns
```

In [16]:

```
sns.heatmap(df.corr())
```

Out[16]:

<AxesSubplot:>



Machine Learning (Random Forest Classifier)

In [8]:

```
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
```

In [13]:

```
df.columns.unique()
```

Out[13]:

```
Index(['Unnamed: 0', 'UTC', 'Temperature[C]', 'Humidity[%]', 'TVOC[ppb]',
      'eCO2[ppm]', 'Raw H2', 'Raw Ethanol', 'Pressure[hPa]', 'PM1.0',
      'PM2.5',
      'NC0.5', 'NC1.0', 'NC2.5', 'CNT', 'Fire Alarm'],
      dtype='object')
```

In [19]:

```
df.drop(['Unnamed: 0'],axis=1,inplace=True)
```

In [18]:

```
df.drop(['UTC'],axis=1,inplace=True)
```

In [20]:

```
df.head()
```

Out[20]:

	Temperature[C]	Humidity[%]	TVOC[ppb]	eCO2[ppm]	Raw H2	Raw Ethanol	Pressure[hPa]	PM1.0
0	20.000	57.36	0	400	12306	18520	939.735	0.0
1	20.015	56.67	0	400	12345	18651	939.744	0.0
2	20.029	55.96	0	400	12374	18764	939.738	0.0
3	20.044	55.28	0	400	12390	18849	939.736	0.0
4	20.059	54.69	0	400	12403	18921	939.744	0.0

In [22]:

```
x=df.drop(['Fire Alarm'],axis=1)
y=df['Fire Alarm']
```

In [23]:

```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.33, random_state=42)
```

In [25]:

```
cla=RandomForestClassifier()  
cla.fit(x_train,y_train)
```

Out[25]:

```
RandomForestClassifier()
```

In [27]:

```
y_pred=cla.predict(x_test)
```

In [28]:

```
from sklearn.metrics import confusion_matrix, classification_report, accuracy_score
```

In [29]:

```
accuracy_score(y_test,y_pred)
```

Out[29]:

```
0.9999516160247726
```