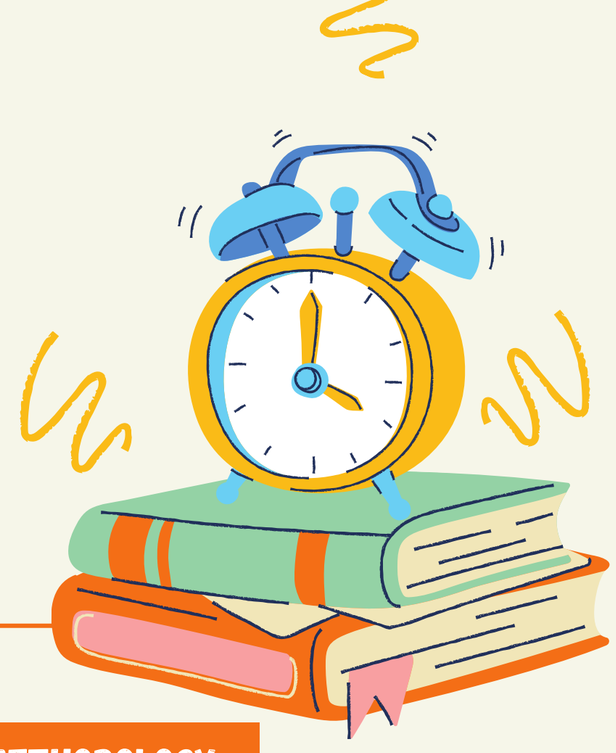
SMOKE DETECTION

Why does indicating whether a fire occurs or not with fire alarm is important?

- **in order to save lives**
- **minimizing and reducing the risk of fire**
- **reducing the damage of property**



DATA DESCRIPTION

Data collected with the help of IOT device with attributes :

- **UTC** = Timestamp in Second
- **Temperature(C)** = Temperature in Celcius
- **Humidity(%)** = Humidity in %
- **TVOC(ppb)** = Total Volatile Organic Compound in part per billion
- **eCo2(ppm)** = Total of Co2 Equivalent Concentration in part per million
- **Raw H2** = Total Molecular Hydrogen
- **Raw Ethanol** = Raw Ethanol Gas
- **Pressure** = Air Pressure in hectopascal(hPa), 1hPa = 100 Pa
- **PM1.0** = Particular Matter diameter Size < 1µm
- **PM2.5** = Particular Matter diameter Size < 2.5 µm
- **NC0.5** = Number Concentration of particular matter Size < 0.5 µm
- **NC1.0** = Number Concentration of particular matter Size < 1 µm
- **NC2.5** = Number Concentration of particular matter < 2.5 µm
- **CNT** = Sample Counter
- **Fire Alarm** = Binary Output(1 if alarm ring, 0 if not)

OBJECTIVE

- **Building predictive models to predict whether a fire alarm will goes on or off.**
- **And to find in what condition a fire alarm will ring**

DECISION TREE CLASSIFIER

	precision	recall	f1-score	support
0	1.00	1.00	1.00	3594
1	1.00	1.00	1.00	8932
accuracy			1.00	12526
macro avg	1.00	1.00	1.00	12526
weighted avg	1.00	1.00	1.00	12526

XGB CLASSIFIER

	precision	recall	f1-score	support
0	1.00	1.00	1.00	3594
1	1.00	1.00	1.00	8932
accuracy			1.00	12526
macro avg	1.00	1.00	1.00	12526
weighted avg	1.00	1.00	1.00	12526

METHODOLOGY

1. **Data Preprocessing**
   - Find missing value in the dataset (no missing value found)
   - Find duplicated row (no duplicated row found)
   - Compute descriptive statistic onnumerical variable to find insight of the dataset
   - Drop useless attributes (UTC and CNT)
2. **Exploratory Data Analysis**
   - Plot the value count of each class in target variable
   - Plot a heatmap correlation of the dataset to find correlated variable
   - Plot a boxplot of each independent variable to dependent variable to find some insight
3. **Feature Engineering + Modelling and Evaluation**
   - Split the dataset into training and testing set with proportion of 80% : 20%
   - Fit the model with 2 classification algorithm model (Decision Tree Classifier and Xtreme Gradient Boost Classifier (XGBoostClassifier)) and evaluate both model
   - Plot a decision tree plot to see in which condition fire alarm will ring or not

RESULT & DISCUSSION

- Both model work very perfect on testing set, both algorithm yields a result of 100% accuracy score. Both classification report are shown as the following.
- Decision Tree plot could be shown on the .ipynb file , the plot shows that 30,136 from 35,825 rung alarm will ring in conditions : **Pressure[hPa] > 938.143 hPa AND TVOC(ppb) > 204.5 ppb AND Humidity[%] > 46.21%**
- The decision tree plot also shows that 9,060 from 14,279 non-rung alarm is under condition of : **Pressure[hPa] <= 937.6 hPa** and the other 4,323 is under condition : **Pressure[hPa] > 938.143 hPa AND TVOC[ppb] <= 87.5 ppb AND PM1.0 <= 1.455**
- Based on the feature importance of the decision tree model, the 3 most important features are: **Pressure[hPa] , TVOC[ppb], and PM1.0**

CONCLUSION

Based on the model and evaluation, it is recommended for buildings to at least meet the requirements for each room conditions such as described in the result & discussion.

TEAM INFO

Institution : BINUS University
Major : Data Science

- Andrew Widjaya - 2501961994 - andrew032@binus.ac.id
- Ethan Daniel Timadius - 2501972530 - ethan.timadius001@binus.ac.id
- Florencia - 2540129262 - florencia003@binus.ac.id
- Sharon Zefanya Setiawan - 2501961022 - sharon.setiawan001@binus.ac.id