

Final Project ML Process

Prediksi Kebakaran

menggunakan machine learning

Nama : **Hendrik Tanaka**

User : hendrik-kfqs

Pacmann Sekolah Data

November – Desember 2022

Daftar Isi

Contents

Daftar Isi	2
Permasalahan Bisnis	3
Latar Belakang.....	3
Smoke Detector dan Kerugian karena Kebakaran	3
[PF1] Why we need to solve this problem?	4
Objektif Bisnis	4
[PF2] What is your goal expectation?	4
Metrik Bisnis	4
Solusi Machine Learning	5
[HL1] Your proposed approach to solve this problem.....	5
[HL2] A brief explanation of what we gonna make	5
Metrik Machine Learning.....	7
[GS1] What is the metrics that you gonna use to measure project output performance?	7
[GS2] A brief explanation of your metrics and why you use them	7
[GS3] How does your metrics value looks like when it success and not	7
Insight Hasil Eksplorasi.....	8
[KS1] What kind of dataset that you gonna use?	8
[KS2] A brief explanation of your dataset.....	8
[KS3] Where does it come from?	8
[KS4] What kind of algorithm that you gonna use?	8
Insight dari hasil EDA	9
Timeline Project	14
[KF1] How is the project timeline?.....	14
[LR1] Your project timeline	14
Kendala selama Project.....	15
Summary	15
Source Code	17
Referensi	18
[AR1] List of components that your project needed	18
[RF1] List of references that you use in this project	18

Permasalahan Bisnis

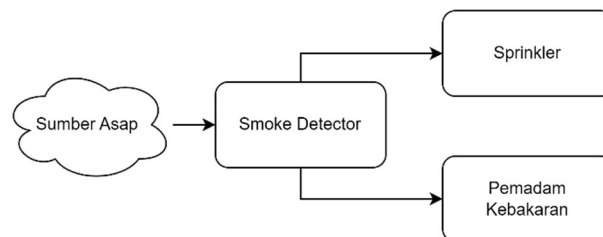
Latar Belakang

PT. Teknologi Rekayasa, didirikan 10 tahun yang lalu untuk mulai berbisnis alat sistem keamanan untuk gedung-gedung perkantoran maupun apartemen, yang dimotori oleh kakak beradik, Sutomo bersaudara. Kedua kakak beradik ini, memulai dengan berdagang di 2 tahun awal berbisnis, dengan meng-import produk produk tersebut. Sering kali, ketersediaan dan ketepatan pengiriman untuk brand tertentu, menjadi kendala saat memenuhi pesanan, yang pada saat tersebut sedang tingginya.

Berbekal sarjana engineering dari Jerman, kedua kakak beradik ini memberanikan diri membuat pabrik untuk membuat alat-alat sistem keamanan tersebut didalam negeri dengan merk "AXS". Produk unggulan adalah Smoke Detector yang banyak dipergunakan di Gedung perkantoran, apartemen maupun bangunan bertingkat lainnya, dan menjadi pilihan utama di market tersebut selama kurang lebih 5 tahun terakhir.

Smoke Detector dan Kerugian karena Kebakaran

Fungsi utama dari Smoke Detector adalah mengidentifikasi ada tidak nya asap pada lingkungan yang menjadi area jangkauan dari alat ini. Akan tetapi, sering kali alat ini diidentifikasi sebagai alat yang bisa memberikan informasi awal akan terjadinya kebakaran, yang nanti informasi ini akan dipakai untuk menyalakan alat-alat pemadam kebakaran seperti sprinkler maupun mengirim informasi secepatnya ke petugas pemadam kebakaran setempat.



Dalam sebuah kasus kebakaran, sering kali terjadinya kerugian material dan korban jiwa menjadi semakin besar akibat tidak tertanganinya proses penghentian penyebaran api ke sekitarnya. Tentu hal ini sangat berpengaruh pada ketersediaan alat alat pemadam kebakaran pada fasilitas tersebut, berfungsinya alat tersebut dan kecepatan dari petugas pemadam kebakaran untuk tiba di lokasi kebakaran dan seberapa cepat petugas dapat menghentikan proses penyebaran api tersebut meluas.

Dari gambar flow diatas dapat ditarik kesimpulan sederhana, bahwa fungsi Smoke Detector sangat lah penting untuk mengurangi kerugian yang terjadi. Sayangnya alat Smoke Detector yang saat ini di produksi, sering kali gagal mengidentifikasi kebakaran yang terjadi atau lebih parahnya, Smoke Detector mengirimkan sinyal yang salah kepada petugas pemadam kebakaran akibat rendahnya akurasi dari Smoke Detector.

Kesalahan identifikasi dari Smoke Detector, terjadi karena banyaknya kondisi yang bervariasi dilapangan, misal Smoke Detector mengidentifikasi kebakaran pada kondisi ketika seseorang membakar sate, rokok, atau kondisi dengan kelembaban tinggi/debu tinggi pada proyek konstruksi dan lain sebagainya.

Kesalahan identifikasi dari Smoke Detector dapat mengakibatkan :

- Pada kasus tidak dapat mengidentifikasi kebakaran padahal terjadi kebakaran, adalah tidak berfungsinya alat alat pemadam kebakaran otomatis dilokasi, meningkatnya korban jiwa/kerugian material dan menyebarnya kebakaran yang meluas.

- Pada kasus mengidentifikasi kebakaran padahal tidak terjadi kebakaran, adalah menyalanya alat alat pemadam kebakaran otomatis di lokasi, waktu terbuang percuma dari petugas pemadam kebakaran dan tentunya biaya yang dikeluarkan.

[PF1] Why we need to solve this problem?

Tingginya false alarm tersebut mengakibatkan adanya trend penurunan penjualan Smoke Detector merk “AXS” dan alat alat pemadam kebakaran otomatis yang di produksi oleh PT. Teknologi Rekayasa selama 2 tahun terakhir.

Perusahaan berencana membuat pengembangan produk Smoke Detector dari versi sebelumnya, tentu nya dengan penambahan modul pengenalan kebakaran yang lebih akurat dan tentunya tidak menambah biaya produksi yang tinggi. Pada akhirnya akan kembali meningkatkan kepercayaan public akan merk “AXS” dipasaran dan meningkatkan penjualan dari produk andalan tersebut.

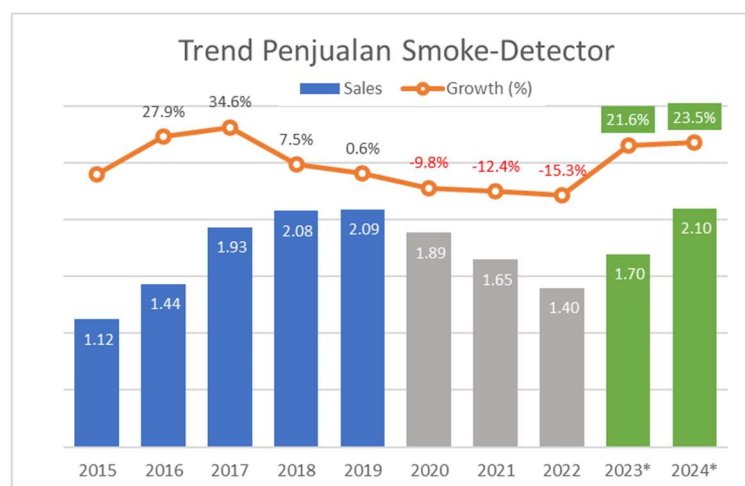
Objektif Bisnis

[PF2] What is your goal expectation?

Tentunya, Smoke Detector versi baru , yang dapat mengidentifikasi kebakaran dengan akurat pada berbagai kondisi lapangan, sehingga dapat meningkatkan penjualan dan kepercayaan bahwa merk “AXS” adalah yang terbaik.

Metrik Bisnis

Metrik bisnis yang ingin dicapai adalah meningkatnya kembali penjualan dari Smoke Detector , untuk kembali pada level kondisi 2 tahun sebelumnya , dan ini dapat di capai secara bertahap selama 2 tahun kedepan. Masa riset dan pembuatan dari Produk baru ini adalah selama 1 tahun ini.



Grafik diatas terlihat bahwa sejak 2020 sampai 2022, terjadi penurunan yang cukup significant, penjualan dalam jutaan pcs. Metrik bisnis yang ingin dicapai adanya peningkatan Sales secara bertahap di tahun 2023-2024, sehingga penjualan nya akan menyamai pada tahun 2009.

Solusi Machine Learning

[HL1] Your proposed approach to solve this problem

Hal pertama yang perlu diidentifikasi adalah hal hal apa saja yang dapat diidentifikasi oleh sebuah Smoke Detector, yaitu :

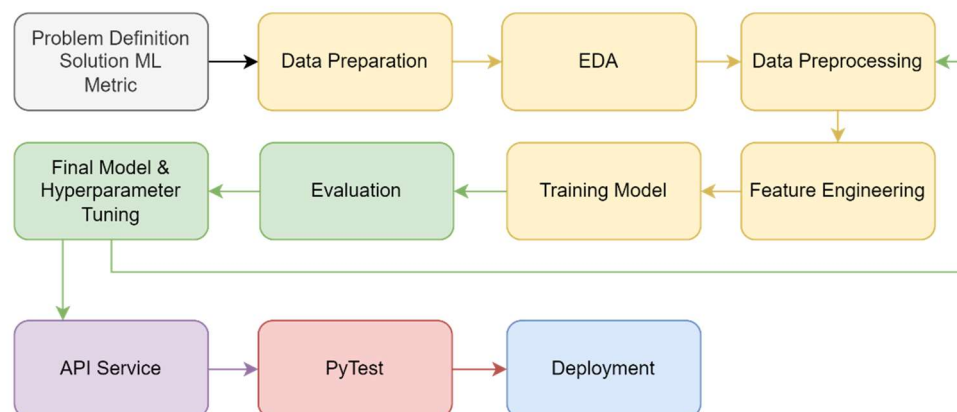
- Suhu & kelembaban udara
- Tekanan udara
- TVOC [ppb] : Total Volatile Organic Compounds
- eCO2 [ppm] : co2 equivalent concentration
- Raw H2 : molekul hydrogen
- Raw Ethanol : molekul ethanol gas
- PM1.0, PM2.5 : Ukuran partikel
- NC0.5, NC1.0, NC2.5 : Jumlah konsentrasi partikel yang berpengaruh.

Ke-12 data diatas, di hasilkan oleh smoke detector. Kombinasi dari ke-12 data diatas, akan menggambarkan kondisi yang mendukung terjadinya kebakaran atau tidak. Karena kombinasi dari 12 data diatas akan sangat bervariasi, maka solusi yang paling mendekati kondisi nyata, adalah pemodelan dengan matematika.

Tentu kondisi diatas hanya bisa diidentifikasi dari kondisi yang pernah terjadi, sehingga kedepannya memungkinkan terjadinya hal hal baru di luar data yang sudah di kumpulkan. Tantangan ini tentu, tidak bisa di solve dengan hanya kalkulasi sederhana, penggunaan machine learning, tentu akan mempercepat dan membantu menentukan pemodelan yang cocok untuk kasus kebakaran diatas.

Machine learning model yang akan dipakai adalah supervised-binary classification.

[HL2] A brief explanation of what we gonna make



Penjelasan mengenai langkah2 yang akan di lakukan adalah :

1. Pendefinisian masalah ,solusinya dan metric
Problem bisnis, solusi yang bisa diambil apakah terkait machine learning atau tidak, jika iya, menentukan metric bisnis dan metric machine learning nya.
2. Persiapan data
Data mulai dilihat mengenai :
 - a. Data collectionnya, apakah di dapat dari 1 sumber atau dari berbagai sumber (baik asal data, bentuk/tipe database berbeda) yang perlu di olah terlebih dahulu untuk di gabungkan, sehingga mempunyai format yang sama.

- b. Pemahaman mengenai masing-masing predictor nya, penjelasan dan kemungkinan dari range, tipe data dan karakteristik umumnya. Misal suhu maksimum atau minimum dari kondisi yg ada.
 - c. Validasi mengenai jumlah data, predictor, target nya.
 - d. Data splitting menjadi 3 bagian yaitu 70% data training, 15% data validation dan 15% data testing.
- 3. EDA – Exploratory Data Analysis

Merupakan proses pemahaman lebih jauh mengenai data secara statistic dan karakteristik datanya. Antara lain yang akan di explore adalah :

 - a. Data yang d EDA adalah Data Train saja.
 - b. Analisa data predictor :
 - i. Univariate
Melihat bentuk dari data, sebaran, skewness, varian, range & outlier
 - ii. Multivariate
Melihat apakah ada korelasi antar semua variable. Kemudian mengecek seberapa kuat korelasi tersebut, dan apakah ada data yang redundant.
 - c. Mengecek readiness dari data :
 - i. Analisa missing value seberapa banyak, dan perkiraan dari pemahaman variable, apakah akan dibuang atau di imputasi dg data yang ada atau dari data luar.
 - ii. Analisa dari data target, apakah imbalance atau tidak
 - d. Mendapatkan insight dari data diatas, sehingga pemahaman data dan keperluan untuk hyperparameter tuning bisa lebih optimal.
- 4. Data Preprocessing

Preprocessing yang dilakukan adalah :

 - a. Cleaning
 - i. Memastikan format/type dari masing masing variable sudah sesuai definisi dari masing masing variable.
 - b. Normalization
 - i. Apabila ada keperluan untuk balancing data numerik maupun categorical. Misal bisa dengan feature scaling, krn terlalu berbeda karakteristik range data antara satu variabel dg lainnya, akan mempengaruhi hasil pemodelan, misal algoritma knn akan sangat terpengaruh dg jarak.
 - c. Transformation
 - i. Apabila di butuhkan, dapat merubah data, misal dg log-transformation atau PCA dsb
 - d. Imputasi
 - i. Handling missing value dengan mengisi data tsb dg karakteristik data (mean/median/modus/ lainnya) atau bs dengan referensi data dari luar/expert (tenaga ahli)
 - ii. Remove outlier
 - iii. Handling categorical dengan label encoder
- 5. Feature Engineering

Proses untuk membuat fitur baru (apabila diperlukan), guna mempermudah dalam proses pemodelannya. Antara lain :

 - a. Fitur baru hasil calculation/extraction dari fitur lain.
 - b. One-hot-encoding untuk data ctategorical yang bukan merupakan ordinal/urutan/tidak mempunyai derajat lebih tinggi dari lainnya, misal data jenis kelamin (kategorikal-tidak ordinal) akan berbeda dengan data tingkat Pendidikan (kategorikal-ordinal)
 - c. Khusus untuk data text, bisa dengan text vectorization
 - d. Dimensional Reduction, dapat dilakukan dengan :

- i. PCA
 - ii. Membuang fitur yang tidak perlu, setelah kita membuat fitur baru pada point (a), apabila memang tidak diperlukan lagi.
 - iii. Atau drop fitur berdasarkan EDA, misal jika ada data redundant sama sama mempunyai korelasi kuat (misal 1)
 - iv. Drop fitur berdasarkan fitur importance , saat hyperparameter tuning (setelah proses pemodelan)
6. Training Model
Training model dengan berbagai model, untuk melihat performa masing-masing model. Gunakan baseline model terlebih dahulu, yaitu model dengan parameter standard. Apabila hasil evaluasi pada point 7
7. Evaluation
Evaluasi dari baseline model dan pemilihan model terbaik. Kemudian apabila diperlukan, dapat dilakukan hyperparameter tuning untuk membuat performa model menjadi lebih baik lagi, dapat kembali ke point 2 (jika diperlukan) atau ke point 4 dan 5, untuk kemudian Training model kembali, sampai di dapatka performa yang baik.
8. API Service
Persiapan dari production model, untuk deployment. Sehingga model kita siap dipakai oleh user lain, dengan menggunakan API Service untuk frontend dan backend.
9. Pytest
Testing terhadap code yang kita Bangun.
10. Deployment
Deployment model pada server/embedded system untuk bisa dipakai oleh user.

Metrik Machine Learning

[GS1] What is the metrics that you gonna use to measure project output performance?

Metrik dari machine learning yang akan di apakai adalah F1-Score

[GS2] A brief explanation of your metrics and why you use them

Untuk kasus klasifikasi, secara sederhana akurasi dapat di hitung dengan membagi jumlah prediksi benar terhadap semua data. Akan tetapi hal ini akan sedikit menjadi salah, apabila ada imbalanced data atau ada jumlah data target yang minoritas.

Confusion matriks, yaitu jumlah dari kondisi prediksi benar dan salah dan pada data negative atau positif (TP,TN,FP,FN). Khusus untuk imbalanced data, dapat menggunakan :

- F1-Score, yaitu rata rata dari recall dan precision.

[GS3] How does your metrics value looks like when it success and not

- F1-Score mempunyai nilai dari 0 sd 1, nilai 1 untuk klasifikasi sempurna dan 0 untuk gagal total klasifikasi.

Insight Hasil Eksplorasi

[KS1] What kind of dataset that you gonna use?

Data set yang di gunakan adalah dari <https://www.kaggle.com/datasets/deepcontractor/smoke-detection-dataset>. Dataset berbentuk file csv, tabular data.

[KS2] A brief explanation of your dataset

Dataset terdiri dari :

- 15 kolom, yang merupakan :
 - 13 kolom data Smoke Detector yang dijelaskan pada point HL1
 - 1 kolom UTC = datetime
 - 1 kolom target = Fire Alarm
- 6230 row data
- Data definition :

Data	Keterangan
UTC	Waktu pengambilan sample, dalam bentuk detik UTC (detik dari 01 Jan 1970)
Temperature[C]	Suhu dalam derajat Celsius
Humidity[%]	Kelembaban dalam %
TVOC[ppb]	Total volatile organic compound
eCO2[ppm]	Konsentrasi ekuivalent CO2
Raw H2	raw molecular hydrogen
Raw Ethanol	raw ethanol gas
Pressure[hPa]	Tekanan
PM1.0, PM2.5	ukuran partikel dalam micron
NC0.5, NC1.0, NC2.5	Number concentration of particulate (micron)
CNT	Sample counter
Fire Alarm	Terjadi kebakaran atau tidak 1 = kebakaran dan 0 = tidak

[KS3] Where does it come from?

Data di download dari <https://www.kaggle.com/datasets/deepcontractor/smoke-detection-dataset>, yang mana merupakan hasil percobaan berbagai macam kondisi pada smoke detector.

Data collection dari Smoke Detector ini didapat dari :

1. Hasil percobaan pada berbagai kondisi yang memicu kebakarn dan tidak
2. Data real yang di collect dari Smoke Detector pada berbagai kondisi. Cara kedua ini memerlukan Smoke Detector yang dilengkapi dengan module yang bisa mengirimkan data ke server melalui jaringan internet.

[KS4] What kind of algorithm that you gonna use?

Dari hasil trial modeling, terlihat yang paling bagus adalah DecisionTree

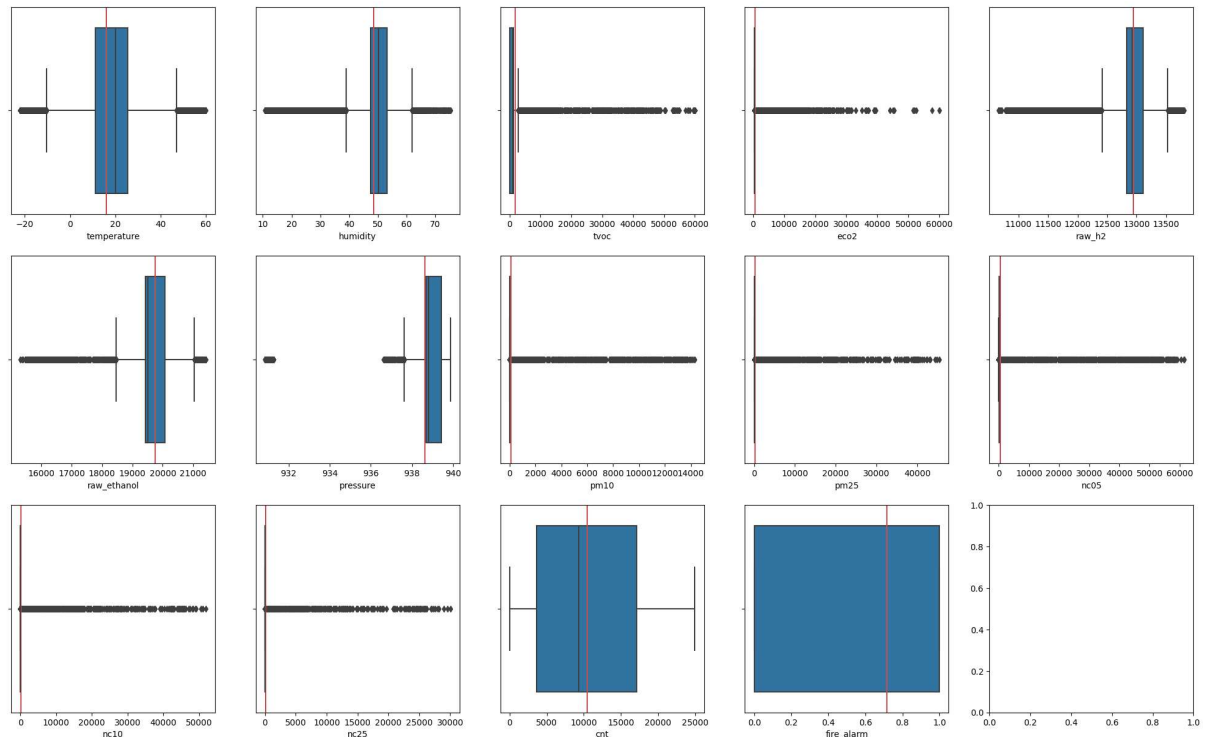
Insight dari hasil EDA

Insight dari hasil EDA adalah sebagai berikut :

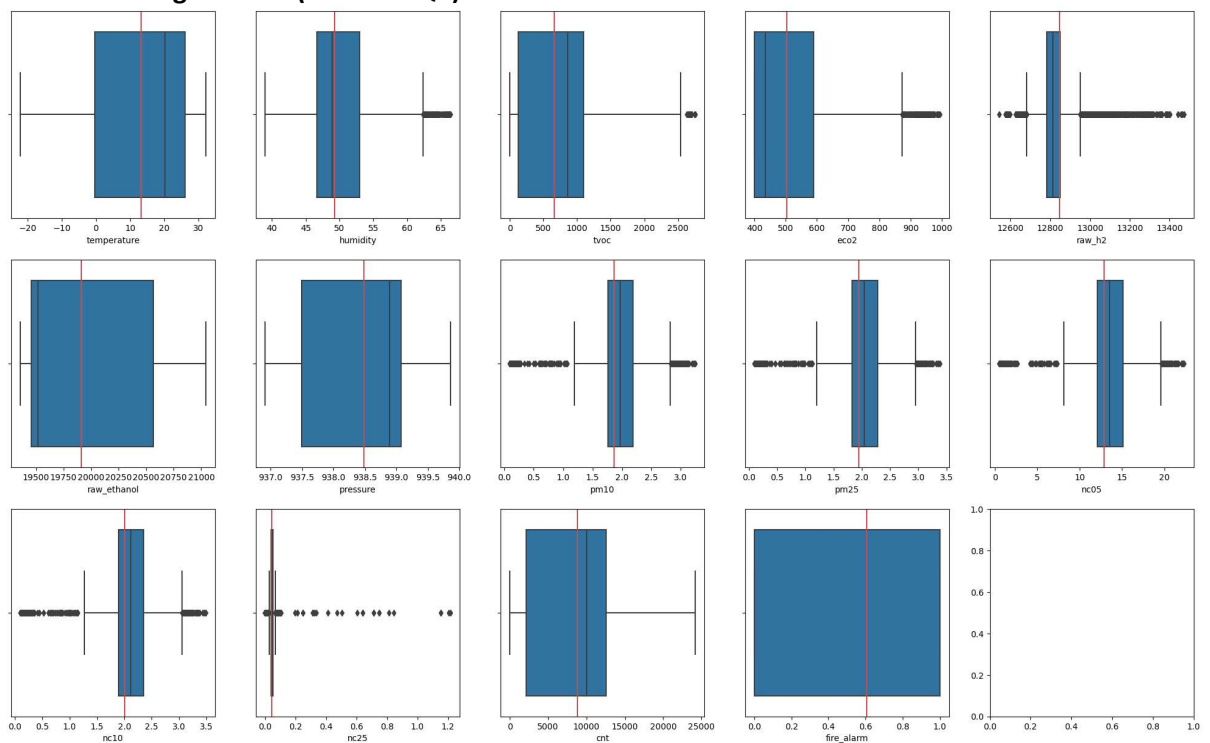
1. Data yang dipakai hanya Data Train untuk menghindari Data Leakage
2. Data sangat bagus, tidak ada missing value
3. Checking Outlier data dari masing masing variable.

Terlihat bahwa banyak data yg terlalu banyak outlier nya, ini target untuk kita bersihkan data outlier nya.

Before removing outlier :



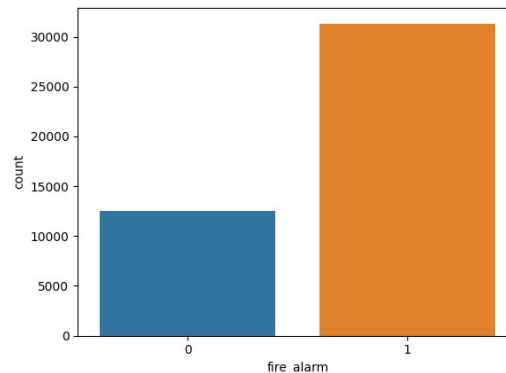
After removing outlier : (metode IQR)



Terlihat data jauh lebih baik sebarannya saat ini.

4. Checking Imbalance

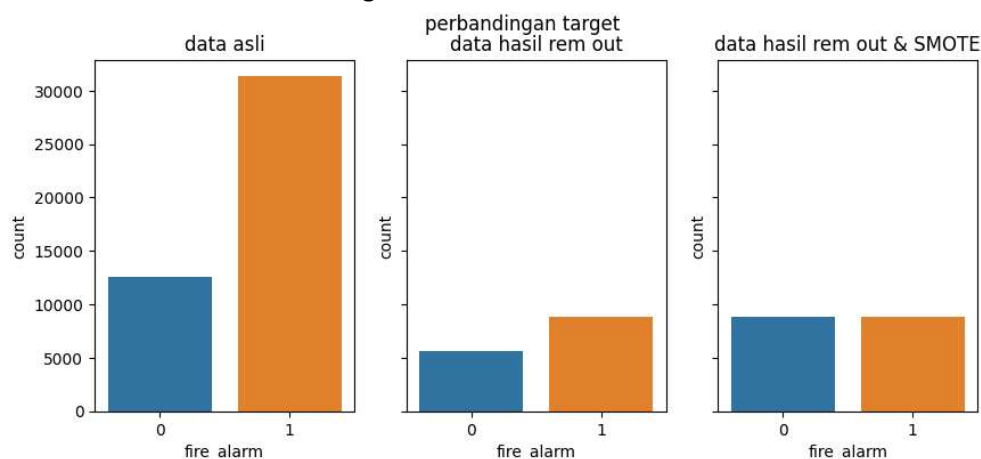
data target, terlihat imbalance, 71.46% data target adalah 1 (terjadi kebakaran).



Dari data ini, perlu dilakukan treatment. Dari referensi, maka akan dipilih treatment dengan SMOTE.

Kemudian dilakukan Removing Outlier, dan disusul oleh SMOTE untuk balancing.

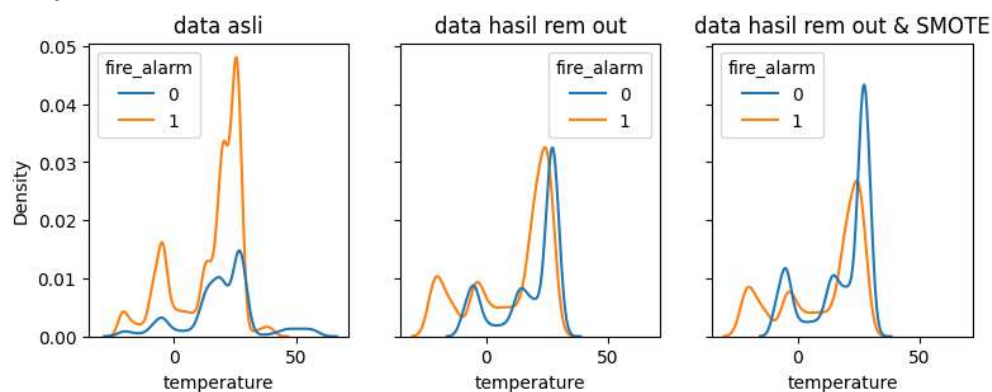
5. Berikut ini adalah data dari target setelah dilakukan kedua hal diatas.



Removing Outlier, ternyata masih menyisakan 60% data imbalance. Dengan proses SMOTE treatment, terlihat data sudah balance, sehingga dirasa baik untuk proses pemodelan, sehingga data tidak berat di salah satu target. Mengapa hal ini penting, karena jika data lebih berat pada data "1" yaitu fire, maka kecenderungan model akan memodelkan untuk terjadinya kebakaran, dan akan berdampak pada tingginya false alarm.

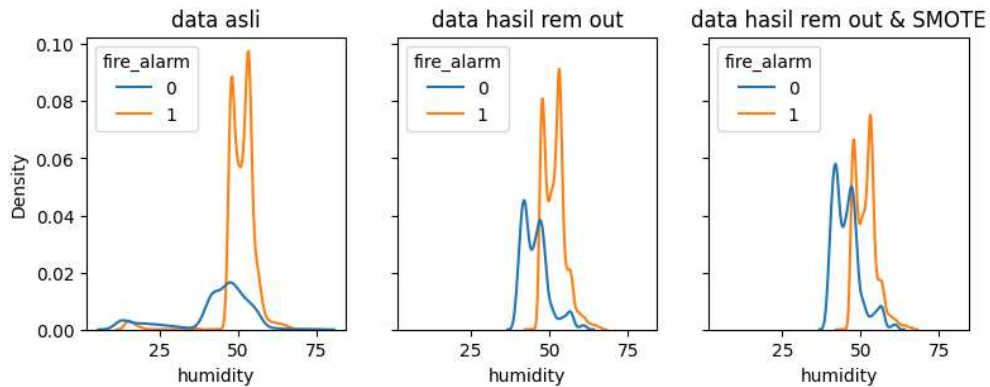
Kemudian dilakukan explorasi data

6. Temperature



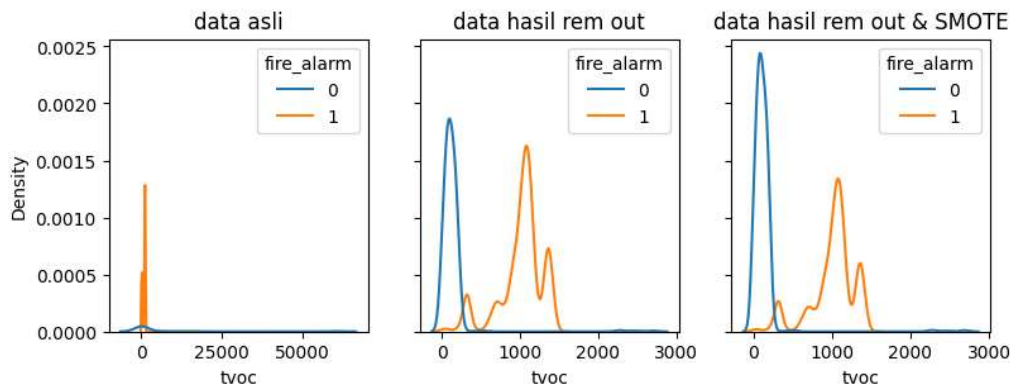
Terlihat bahwa suhu tinggi tidak memang lebih besar terjadinya kebakaran daripada suhu rendah, akan tetapi suhu tinggi tidak selalu terjadi kebakaran.

7. Humidity



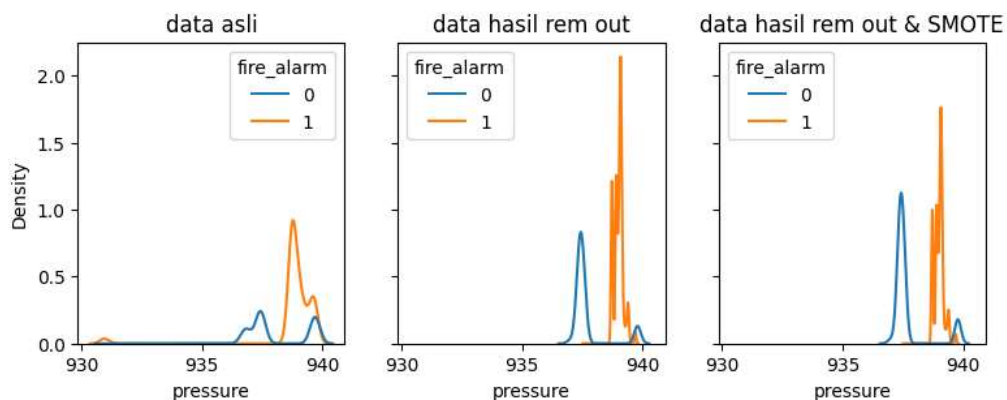
Tingkat kelembaban, yang tinggi mempunyai kecenderungan memicu terjadinya kebakaran.

8. TVOC



Tingkat total volatile organic compound, kumpulan partikel organic yang tinggi mempunyai kecenderungan yang sangat besar untuk terjadinya kebakaran.

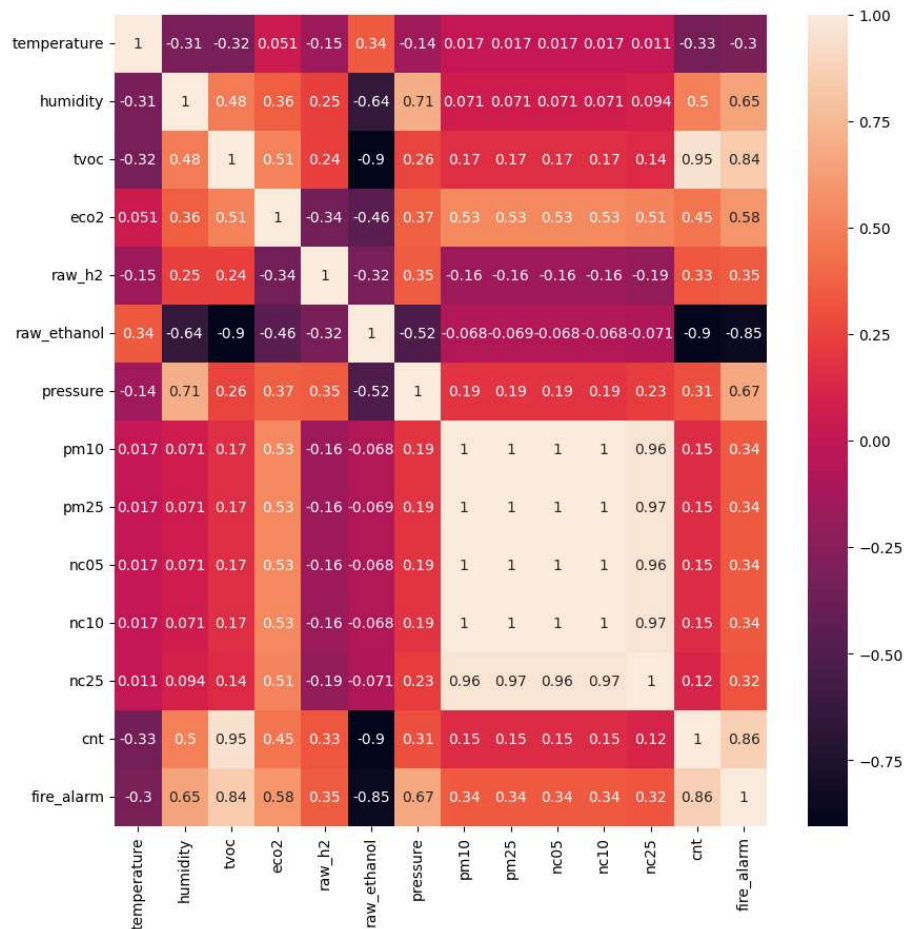
9. Pressure



Terbukti, bahwa tekanan tinggi juga memicu terjadinya kebakaran.

10. Variabel lainnya tidak terlalu menunjukkan kecenderungan tinggi pada factor kebakaran.

11. Korelasi antar variable



Terdapat korelasi kuat pada PM1, PM2.5, NC0.5, NC1.0, NC2.5 yaitu antara 0.96 sampai 1. Dapat disimpulkan bahwa 5 variabel ini sejalan atau bisa dikatakan dapat di buang krn saling mewakili. Nanti akan diambil yg PM1.0 saja.

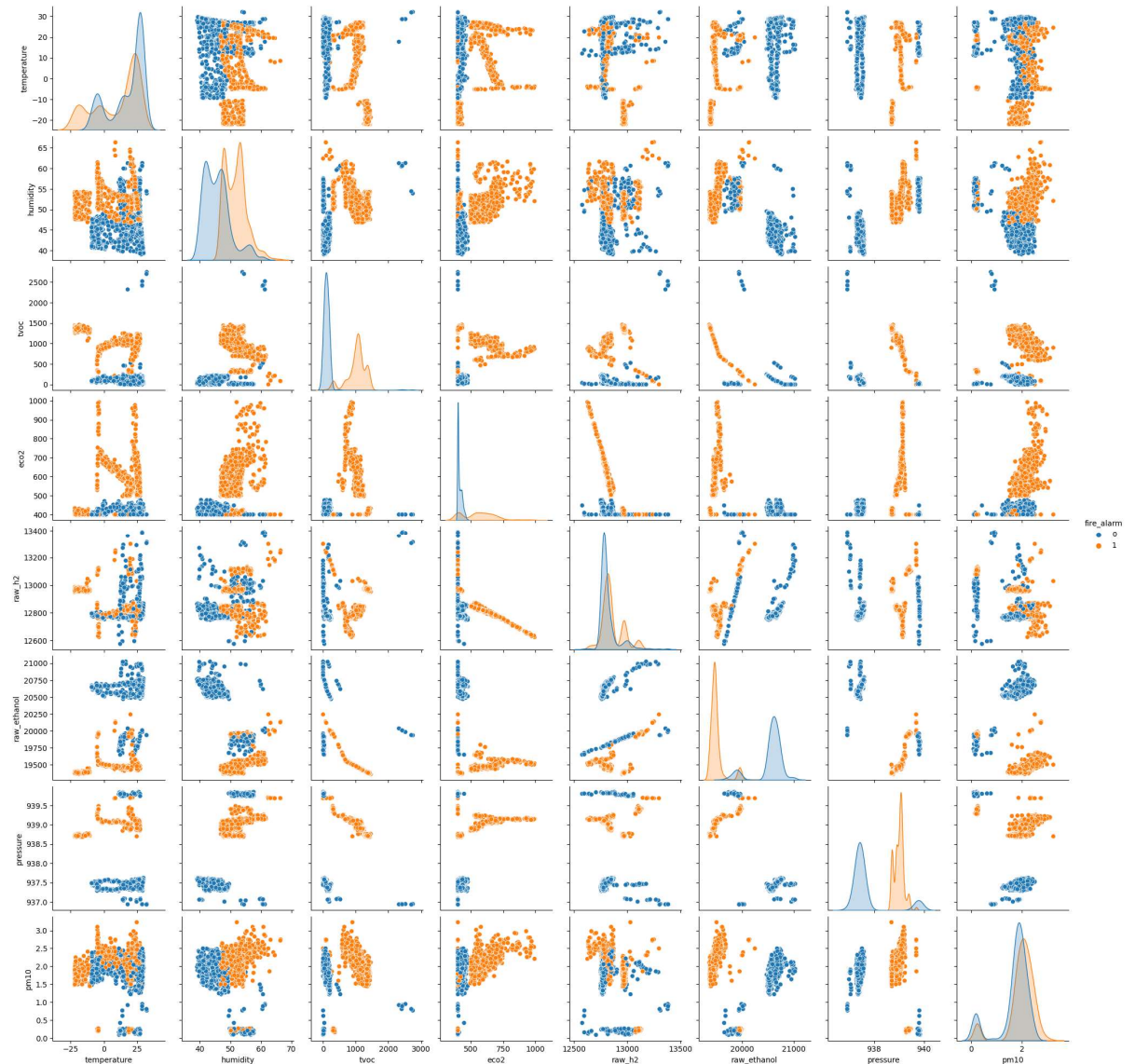
12. Range Data :

	count	mean	std	min	25%	50%	75%	max
temperature	17534.0	14.076997	15.220648	-22.010	1.456500	20.592541	26.190	32.130
humidity	17534.0	48.626711	4.995259	38.990	45.050522	48.180000	52.570	66.440
tvoc	17534.0	568.718946	513.212651	0.000	95.000000	267.000000	1065.000	2754.000
eco2	17534.0	488.016425	116.157448	400.000	400.000000	425.000000	561.000	993.000
raw_h2	17534.0	12841.844017	105.601244	12543.000	12780.000000	12804.000000	12847.000	13474.000
raw_ethanol	17534.0	20020.253964	560.878727	19359.000	19465.000000	19907.000000	20600.000	21042.000
pressure	17534.0	938.348961	0.855880	936.915	937.439412	938.728000	939.068	939.861
pm10	17534.0	1.840358	0.586917	0.090	1.740000	1.950000	2.170	3.250
fire_alarm	17534.0	0.500000	0.500014	0.000	0.000000	0.500000	1.000	1.000

Terlihat bahwa sebenarnya, sebaran range data dari variable cukup bervariasi dan lebar. Dalam ML, seharusnya data data ini perlu di normalisasi (dg standard Scaler), khususnya apabila model yang dipakai sensitive terhadap jarak. Dan ini baru bisa di putuskan, saat nanti melihat hasil baseline modelnya apakah memerlukan atau tidak.

Dari hasil baseline model, ternyata menggunakan DecisionTree, sehingga tidak perlu dilakukan normalisasi.

13. Pairplot dari semua variable



Terlihat hubungan antar variable dan distribusi masing2 variabel nya.

Modelling

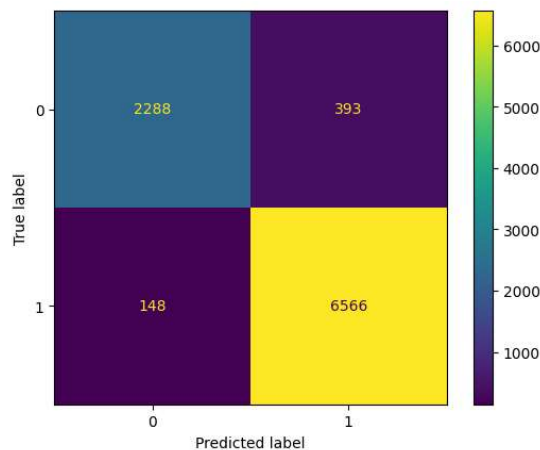
14. Hasil dari baseline model dengan :

- Logistic Regression
- Decision Tree Classification
- Random Forest Classification
- KNN Classification
- XGBoost Classification

Dari hasil baseline model, diperoleh bahwa F1 score tertinggi ada pada **Decision Tree Classification** dengan nilai f1-score=0.93.

Untuk data test dan valid, juga menghasilkan angka yang baik di 0.94.

15. Hasil dari Confusion Matrix adalah sebagai berikut :



Timeline Project

[KF1] How is the project timeline?

Project timeline :

Minggu	Aktifitas
Week 1-2	Persiapan Environment Data Collection & Preparation EDA
Week 2-4	Baseline Model Final Model Hyper parameter tuning
Week 5	Scripting API & Docker Service Deployment
Week 6	Dokumentasi Presentasi

[LR1] Your project timeline

Minggu	Aktifitas
19 Oktober 2022 – 19 November 2022	Persiapan Environment Data Collection & Preparation
20 November 2022 – 01 Desember 2022	EDA
26 November 2022 – 30 November 2022	Baseline Model Final Model
----	Hyper parameter tuning
1 Desember 2022 – 4 Desember 2022	Scripting API & Docker Service
----	Deployment
04 Desember 2022	Dokumentasi Presentasi

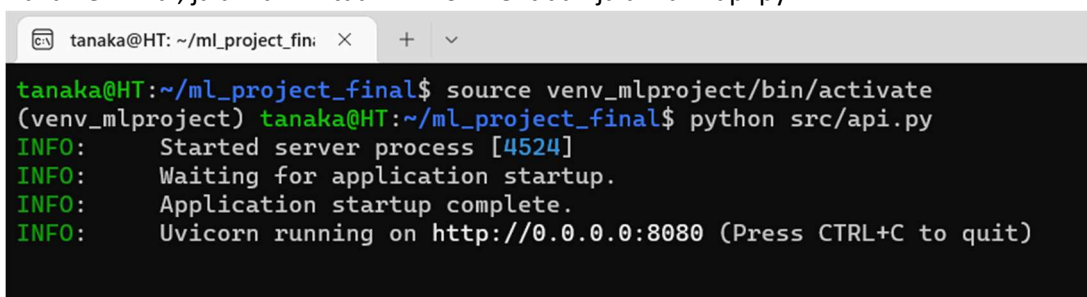
Kendala selama Project

Kendala yang dialami selama project :

1. Pemahaman akan End-to-End ML Process masih rendah, sehingga terlihat bolak balik dari persiapan dan EDA.
2. Pemahaman akan Hyperparameter tuning kurang, sehingga bagian ini tidak dilakukan. Kebetulan hasil akurasi model pada baseline model cukup tinggi.
3. Proses API service dengan 2 terminal berjalan. Walaupun sempat ada kendala pada port 8501, kemudian di rubah ke port 8500. Dan proses servicing berjalan dengan baik antara frontend streamlit dan backend api.
4. Proses pembuatan Docker, berhasil. Akantetapi tidak dapat berjalan dengan baik, terdapat error connection.

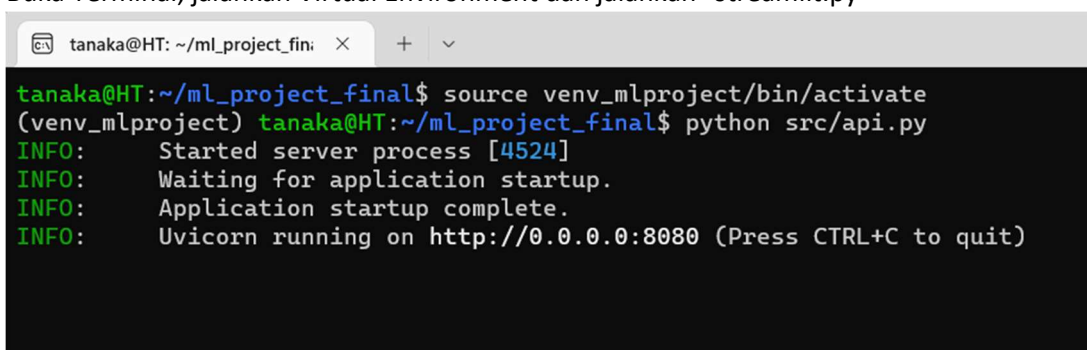
Summary

1. Pemahaman flow ML Process terbentuk, dengan project ini.
2. Proses pembuatan hanya berjalan sampai localhost dengan 2 terminal, sedangkan docker-compose belum berhasil.
3. Buka Terminal, jalankan Virtual Environment dan jalankan "api.py"



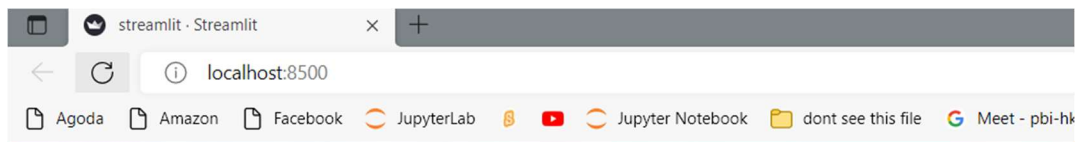
```
tanaka@HT: ~/ml_project_fin: X + v
tanaka@HT:~/ml_project_final$ source venv_mlproject/bin/activate
(venv_mlproject) tanaka@HT:~/ml_project_final$ python src/api.py
INFO: Started server process [4524]
INFO: Waiting for application startup.
INFO: Application startup complete.
INFO: Uvicorn running on http://0.0.0.0:8080 (Press CTRL+C to quit)
```

4. Buka Terminal, jalankan Virtual Environment dan jalankan "streamlit.py"



```
tanaka@HT: ~/ml_project_fin: X + v
tanaka@HT:~/ml_project_final$ source venv_mlproject/bin/activate
(venv_mlproject) tanaka@HT:~/ml_project_final$ python src/api.py
INFO: Started server process [4524]
INFO: Waiting for application startup.
INFO: Application startup complete.
INFO: Uvicorn running on http://0.0.0.0:8080 (Press CTRL+C to quit)
```

5. Buka Browser dan ketik "localhost/8500"



Prediksi Kebakaran



6. Isi data yang di minta dan klik "predict"

Prediksi Kebakaran



Project ML Process

Hendrik Tanaka
Pacmann-Desember 2022

Masukkan data data yang dibutuhkan untuk memprediksi :

Temperature [C] :	eCO2 [ppm] :	Pressure [hPa] :
0.10 - +	4 - +	0.70 - +
Humidity [%] :	Raw H2 :	PM1.0 :
0.20 - +	5 - +	0.80 - +
TVOC [ppb] :	Raw Ethanol :	Predict
3 - +	6 - +	

7. Hasil prediksi akan di tampilkan.



Prediksi Kebakaran 🔥



Project ML Process

Hendrik Tanaka
Pacmann-Desember 2022

Masukkan data data yang dibutuhkan untuk memprediksi :

Temperature [C] :	eCO2 [ppm] :	Pressure [hPa] :
0.10 - +	4 - +	0.70 - +
Humidity [%] :	Raw H2 :	PM1.0 :
0.20 - +	5 - +	0.80 - +
TVOC [ppb] :	Raw Ethanol :	Predict
3 - +	6 - +	

Prediksi akan terjadi kebakaran 🔥

8. Hasil prediksi dapat berjalan dengan baik.

Source Code

Source code, ada pada github dengan alamat :

https://github.com/hendrikt1479/ml_project_final

Referensi

[AR1] List of components that your project needed

Library yang diperlukan :

1. Pandas
2. Numpy
3. Streamlit
4. Sklearn
5. Joblib
6. Imblearn
7. Seaborn
8. Matplotlib
9. Requests
10. Xgboost
11. Json
12. Tqdm

Aplikasi/tool yang dipakai :

1. WSL Ubuntu for Windows
2. VS Code - Python
3. Github
4. Streamlit
5. Docker Desktop dan Docker Hub
6. Draw.IO
- 7.
- 8.
- 9.
- 10.
- 11.

[RF1] List of references that you use in this project

1. <https://www.kaggle.com/datasets/deepcontractor/smoke-detection-dataset>
2. <https://www.nist.gov/how-do-you-measure-it/how-do-smoke-detectors-work>
3. <https://towardsdatascience.com/metrics-for-imbalanced-classification-41c71549bbb5>
4. <https://machinelearningmastery.com/tour-of-evaluation-metrics-for-imbalanced-classification/>