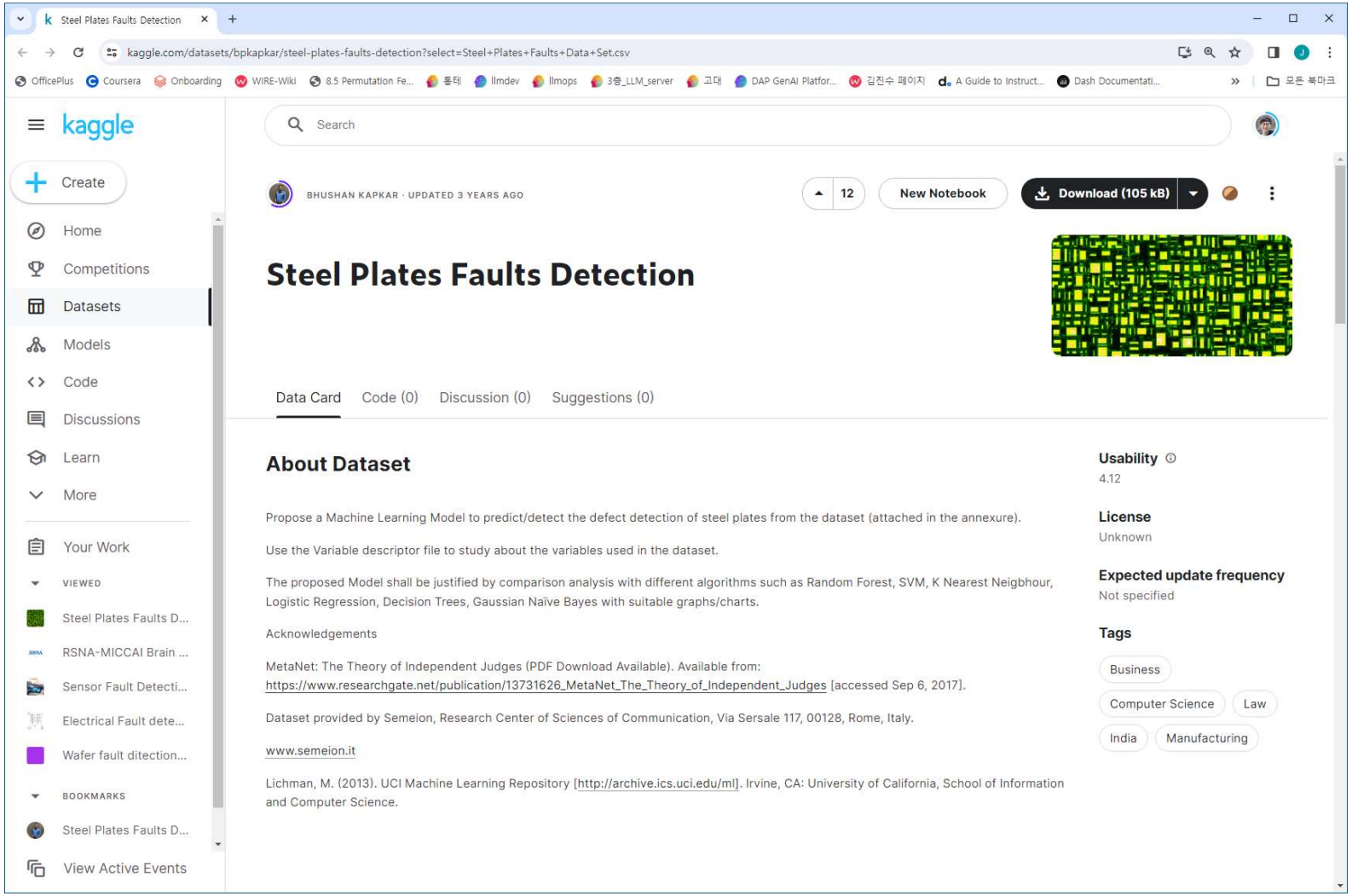
Machine Learning Operation(MLOps)

- Ch2. Level0 MLOps(1)

Level0 MLOps(1)

1. Train / Inference script
2. SQLite3
3. Dash

데이터셋



https://www.kaggle.com/datasets/bpkapkar/steel-plates-faults-detection?select=Steel+Plates+Faults+Data+Set.csv

학습/검증 – 1.analysis.ipynb

```
import pandas as pd
import numpy as np

## data load
dat = pd.read_csv("train.csv")

## train, test split
from sklearn.model_selection import train_test_split
train, test = train_test_split(dat, test_size=0.3)
train.head()

## x variables preprocessing
x_cols = ['V'+str(i) for i in range(1,28)]
from sklearn.preprocessing import StandardScaler
trans = StandardScaler()
trans.fit(train[x_cols])
train_x = trans.transform(train[x_cols])

## y variables preprocessing
train['V34'] = train['Class']-1
train_y = [str(np.where(r==1)[0][0]) for r in train[['V'+str(i) for i in range(28,35)]].to_numpy()]
```

학습용파일 로드

Train/Test 분할

X변수 표준화

y변수 변환

학습/검증 – 1.analysis.ipynb

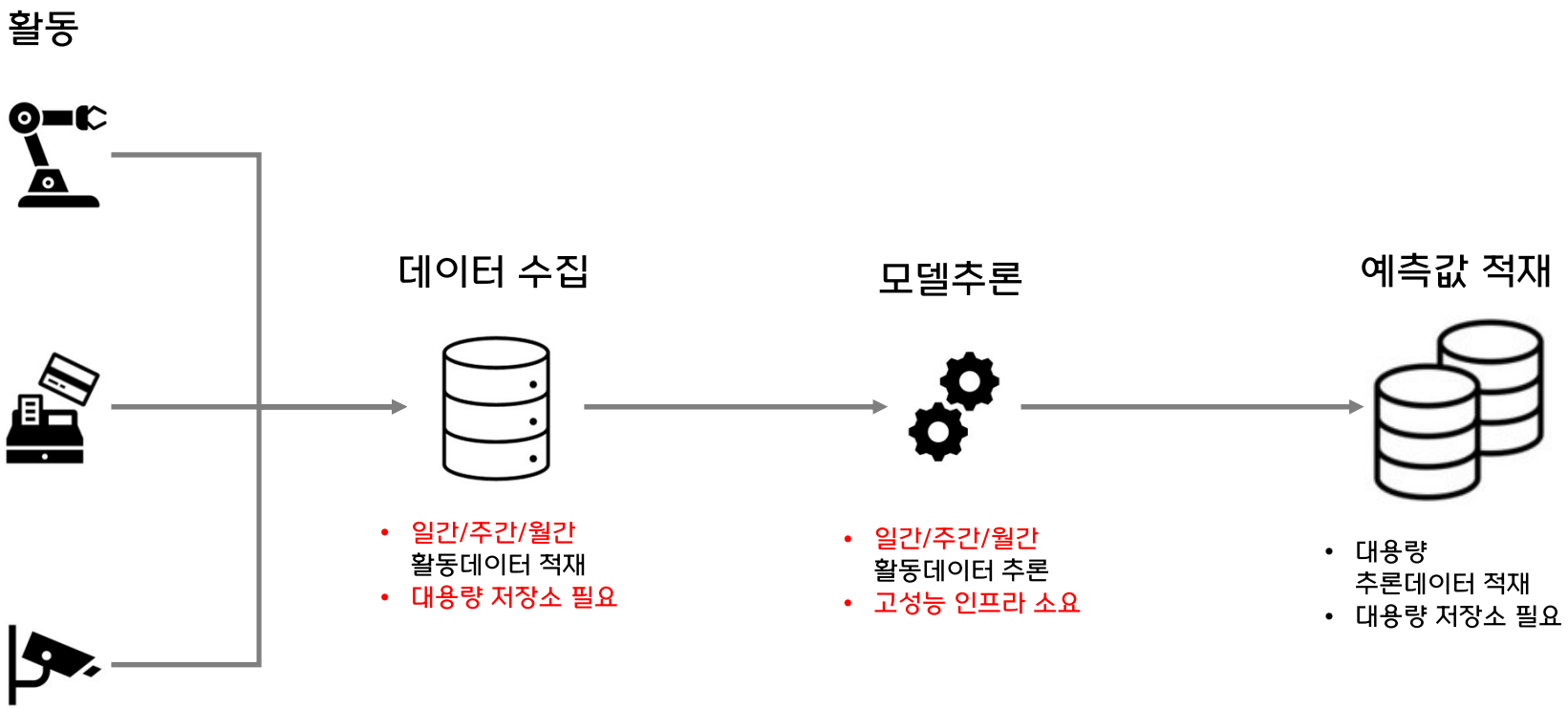
```
## classification modeling
from sklearn.ensemble import RandomForestClassifier
model = RandomForestClassifier()
model.fit(train_x, train_y)

## make prediction for testset
test_x = trans.transform(test[x_cols])
pred = model.predict(test_x)

## validation for testset
test['V34'] = test['Class']-1
test_y = [str(np.where(r==1)[0][0]) for r in test[['V'+str(i) for i in range(28,35)]].to_numpy()]
np.mean(pred==test_y)
pd.crosstab(test_y, pred)
```

분류 모델링(RandomForest)

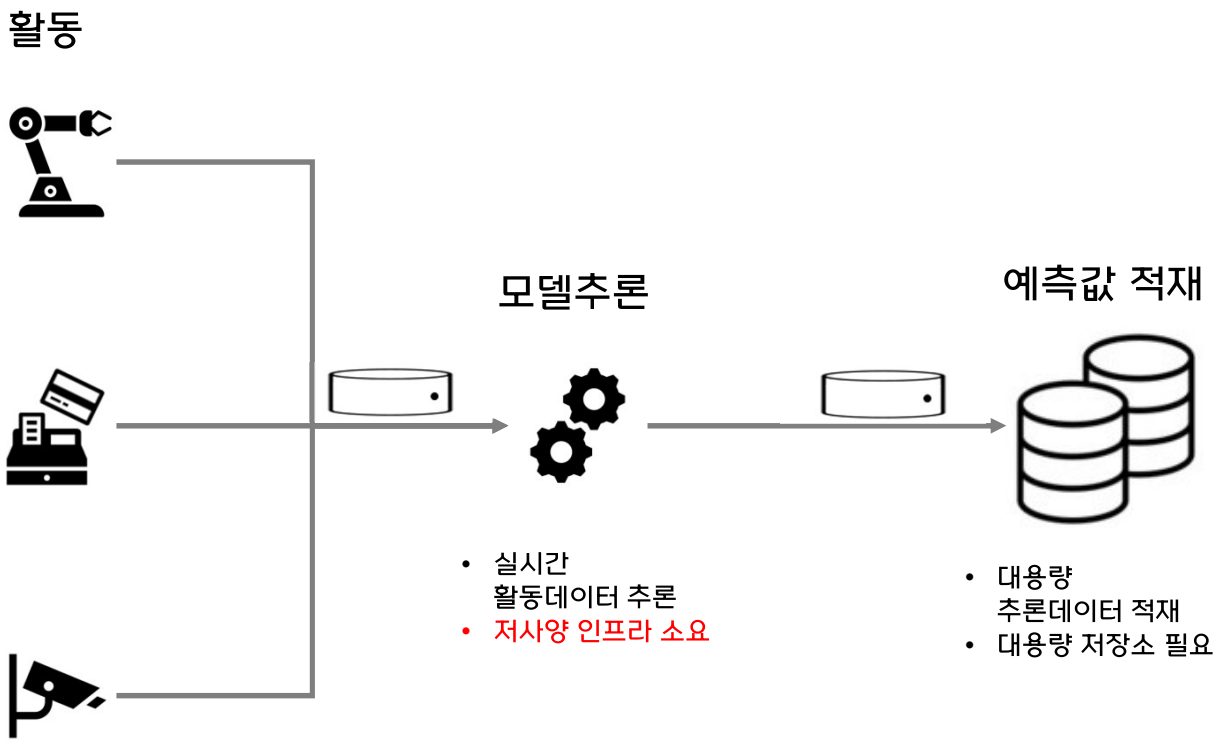
Test셋 예측값 도출

Test셋 정확도 검증

배치 추론과 실시간 추론 (Online and Batch serving)



배치 추론과 실시간 추론 (Online and Batch serving)



배치 추론 – batch_inference.py

```
import pandas as pd
import numpy as np
from sklearn.ensemble import
RandomForestClassifier
import pickle as pkl

## load preprocessing and model
trans = pkl.load(open("trans.pkl","rb"))
model = pkl.load(open("model.pkl","rb"))

## load new dataset
new_data = pd.read_csv("test.csv")
x_cols = ['V'+str(i) for i in range(1,28)]

## prediction for new dataset
x = trans.transform(new_data[x_cols])
pred = model.predict(x)
```

실시간 추론 – online_inference.py

```
import pandas as pd
import numpy as np
import pickle as pkl
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestClassifier

## load preprocessing and model
trans = pkl.load(open("trans.pkl","rb"))
model = pkl.load(open("model.pkl","rb"))

## define function to get new data
def select_row(id):
    a=id-1
    dat = pd.read_csv("test.csv")
    row = dat[a:id]
    return row

## preprocessing for new data
new_data = select_row(1)
x_cols = ['V'+str(i) for i in range(1,28)]
x = trans.transform(new_data[x_cols])

## prediction for new data
pred = model.predict(x)
```

전처리 및 모델 로딩

데이터 추출 함수 정의

전처리 적용

예측값 산출

End of Document