

In [1]:

```
!nvcc --version
```

```
nvcc: NVIDIA (R) Cuda compiler driver
Copyright (c) 2005-2020 NVIDIA Corporation
Built on Thu Jun 11 22:26:38 PDT 2020
Cuda compilation tools, release 11.0, V11.0.194
Build cuda_11.0_bu.TC445_37.28540450_0
```

In [3]:

```
!pip install mxnet
!pip install gluonnlp pandas tqdm
!pip install sentencepiece
!pip install transformers==3.0.2
!pip install torch
```

```
Collecting mxnet
  Using cached mxnet-1.9.0-py3-none-manylinux2014_x86_64.whl (47.3 MB)
Collecting numpy<2.0.0,>1.16.0
  Using cached numpy-1.22.3-cp38-cp38-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (16.8 MB)
Collecting requests<3,>=2.20.0
  Using cached requests-2.27.1-py2.py3-none-any.whl (63 kB)
Collecting graphviz<0.9.0,>=0.8.1
  Using cached graphviz-0.8.4-py2.py3-none-any.whl (16 kB)
Requirement already satisfied: certifi>=2017.4.17 in /home/cuai5th/anaconda3/envs/suminbang/lib/python3.8/site-packages (from requests<3,>=2.20.0->mxnet) (2021.10.8)
Collecting urllib3<1.27,>=1.21.1
  Downloading urllib3-1.26.9-py2.py3-none-any.whl (138 kB)
    |████████████████████████████████████████| 138 kB 4.3 MB/s eta 0:00:01
Collecting idna<4,>=2.5
  Using cached idna-3.3-py3-none-any.whl (61 kB)
Collecting charset-normalizer~=2.0.0
  Using cached charset-normalizer-2.0.12-py3-none-any.whl (39 kB)
Installing collected packages: urllib3, idna, charset-normalizer, requests, numpy, graphviz, mxnet
Successfully installed charset-normalizer-2.0.12 graphviz-0.8.4 idna-3.3 mxnet-1.9.0 numpy-1.22.3 re
quests-2.27.1 urllib3-1.26.9
Collecting gluonnlp
  Using cached gluonnlp-0.10.0-cp38-cp38-linux_x86_64.whl
Collecting pandas
  Using cached pandas-1.4.1-cp38-cp38-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (11.7 MB)
Collecting tqdm
  Downloading tqdm-4.63.1-py2.py3-none-any.whl (76 kB)
    |████████████████████████████████████████| 76 kB 4.5 MB/s eta 0:00:01
Requirement already satisfied: numpy>=1.16.0 in /home/cuai5th/anaconda3/envs/suminbang/lib/python3.8
/site-packages (from gluonnlp) (1.22.3)
Collecting packaging
  Using cached packaging-21.3-py3-none-any.whl (40 kB)
Collecting cython
  Using cached Cython-0.29.28-cp38-cp38-manylinux_2_17_x86_64.manylinux2014_x86_64.manylinux_2_24_x8
6_64.whl (1.9 MB)
Collecting python-dateutil>=2.8.1
  Using cached python-dateutil-2.8.2-py2.py3-none-any.whl (247 kB)
Collecting pytz>=2020.1
  Downloading pytz-2022.1-py2.py3-none-any.whl (503 kB)
    |████████████████████████████████████████| 503 kB 9.8 MB/s eta 0:00:01
Collecting six>=1.5
  Downloading six-1.16.0-py2.py3-none-any.whl (11 kB)
Collecting pyparsing!=3.0.5,>=2.0.2
  Using cached pyparsing-3.0.7-py3-none-any.whl (98 kB)
Installing collected packages: six, pyparsing, pytz, python-dateutil, packaging, cython, tqdm, panda
s, gluonnlp
Successfully installed cython-0.29.28 gluonnlp-0.10.0 packaging-21.3 pandas-1.4.1 pyparsing-3.0.7 py
thon-dateutil-2.8.2 pytz-2022.1 six-1.16.0 tqdm-4.63.1
Collecting sentencepiece
  Using cached sentencepiece-0.1.96-cp38-cp38-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (1.2 MB
)
Installing collected packages: sentencepiece
Successfully installed sentencepiece-0.1.96
Collecting transformers==3.0.2
  Downloading transformers-3.0.2-py3-none-any.whl (769 kB)
    |████████████████████████████████████████| 769 kB 653 kB/s eta 0:00:01
Requirement already satisfied: packaging in /home/cuai5th/anaconda3/envs/suminbang/lib/python3.8/sit
e-packages (from transformers==3.0.2) (21.3)
Requirement already satisfied: sentencepiece!=0.1.92 in /home/cuai5th/anaconda3/envs/suminbang/lib/p
ython3.8/site-packages (from transformers==3.0.2) (0.1.96)
Collecting sacremoses
  Downloading sacremoses-0.0.49-py3-none-any.whl (895 kB)
    |████████████████████████████████████████| 895 kB 22.5 MB/s eta 0:00:01
Collecting tokenizers==0.8.1.rc1
  Downloading tokenizers-0.8.1rc1-cp38-cp38-manylinux1_x86_64.whl (3.0 MB)
```

```

[REDACTED] 3.0 MB 52.7 MB/s eta 0:00:01
Requirement already satisfied: numpy in /home/cuai5th/anaconda3/envs/suminbang/lib/python3.8/site-packages (from transformers==3.0.2) (1.22.3)
Collecting regex!=2019.12.17
  Downloading regex-2022.3.15-cp38-cp38-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (764 kB)
[REDACTED] 764 kB 86.9 MB/s eta 0:00:01
Requirement already satisfied: tqdm>=4.27 in /home/cuai5th/anaconda3/envs/suminbang/lib/python3.8/site-packages (from transformers==3.0.2) (4.63.1)
Collecting filelock
  Using cached filelock-3.6.0-py3-none-any.whl (10.0 kB)
Requirement already satisfied: requests in /home/cuai5th/anaconda3/envs/suminbang/lib/python3.8/site-packages (from transformers==3.0.2) (2.27.1)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in /home/cuai5th/anaconda3/envs/suminbang/lib/python3.8/site-packages (from packaging->transformers==3.0.2) (3.0.7)
Requirement already satisfied: certifi>=2017.4.17 in /home/cuai5th/anaconda3/envs/suminbang/lib/python3.8/site-packages (from requests->transformers==3.0.2) (2021.10.8)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /home/cuai5th/anaconda3/envs/suminbang/lib/python3.8/site-packages (from requests->transformers==3.0.2) (1.26.9)
Requirement already satisfied: charset-normalizer~2.0.0 in /home/cuai5th/anaconda3/envs/suminbang/lib/python3.8/site-packages (from requests->transformers==3.0.2) (2.0.12)
Requirement already satisfied: idna<4,>=2.5 in /home/cuai5th/anaconda3/envs/suminbang/lib/python3.8/site-packages (from requests->transformers==3.0.2) (3.3)
Collecting click
  Using cached click-8.0.4-py3-none-any.whl (97 kB)
Requirement already satisfied: six in /home/cuai5th/anaconda3/envs/suminbang/lib/python3.8/site-packages (from sacremoses->transformers==3.0.2) (1.16.0)
Collecting joblib
  Using cached joblib-1.1.0-py2.py3-none-any.whl (306 kB)
Installing collected packages: regex, joblib, click, tokenizers, sacremoses, filelock, transformers
Successfully installed click-8.0.4 filelock-3.6.0 joblib-1.1.0 regex-2022.3.15 sacremoses-0.0.49 tokenizers-0.8.1rc1 transformers-3.0.2
Collecting torch
  Downloading torch-1.11.0-cp38-cp38-manylinux1_x86_64.whl (750.6 MB)
[REDACTED] 640.8 MB 43.4 MB/s eta 0:00:03ERROR: Could not install packages due to an OSError: [Errno 28] No space left on device

```

```
#깃허브에서 KoBERT 파일 로드
!pip install git+https://git@github.com/SKTBrain/KoBERT.git@master
```

In [22]:

```
train = pd.read_csv('tlftmq.txt', sep = "\\|", engine='python', encoding = "cp949")
```

```
train.head()
```

	AI_id	digit_1	digit_2	digit_3	text_obj	text_mthd	text_deal
0	id_0000001	S	95.0	952.0	카센터에서	자동차부분정비	타이어오일교환
1	id_0000002	G	47.0	472.0	상점내에서	일반인을 대상으로	채소.과일판매
2	id_0000003	G	46.0	467.0	절단하여사업체에도매	공업용고무를가지고	합성고무도매
3	id_0000004	G	47.0	475.0	영업점에서	일반소비자에게	열쇠잠금장치
4	id_0000005	Q	87.0	872.0	어린이집	보호자의 위탁을 받아	취학전아동보육

```
train["text"] = train["text obj"] + " " + train["text mthd"]
```

```
train["text"] = train["text"] + " " + train["text deal"]
```

```
train.drop(['text_obj', 'text_mthd', 'text_deal'], axis=1, inplace=True)
```

```
train data = train[['digit 1','text']]
```

In [28]:

```
train_data.head()
```

Out[28]:

	digit_1	text
0	S	카센터에서 자동차부분정비 타이어오일교환
1	G	상점내에서 일반인을 대상으로 채소.과일판매
2	G	절단하여사업체에도매 공업용고무를가지고 합성고무도매
3	G	영업점에서 일반소비자에게 열쇠잠금장치
4	Q	어린이집 보호자의 위탁을 받아 취학전아동보육

In [35]:

```
import torch
from torch import nn
import torch.nn.functional as F
import torch.optim as optim
from torch.utils.data import Dataset, DataLoader
import gluonnl as nlp
import numpy as np
from tqdm import tqdm, tqdm_notebook
```

In [34]:

```
import sys
!{sys.executable} -m pip install mxnet
```

Defaulting to user installation because normal site-packages is not writeable
Collecting mxnet
Using cached mxnet-1.9.0-py3-none-manylinux2014_x86_64.whl (47.3 MB)
Collecting graphviz<0.9.0,>=0.8.1
Using cached graphviz-0.8.4-py2.py3-none-any.whl (16 kB)
Requirement already satisfied: numpy<2.0.0,>1.16.0 in /home/cuai5th/.local/lib/python3.6/site-packages (from mxnet) (1.19.5)
Requirement already satisfied: requests<3,>=2.20.0 in /home/cuai5th/.local/lib/python3.6/site-packages (from mxnet) (2.27.1)
Requirement already satisfied: idna<4,>=2.5 in /usr/lib/python3/dist-packages (from requests<3,>=2.20.0->mxnet) (2.6)
Requirement already satisfied: charset-normalizer~2.0.0 in /home/cuai5th/.local/lib/python3.6/site-packages (from requests<3,>=2.20.0->mxnet) (2.0.12)
Requirement already satisfied: certifi>=2017.4.17 in /usr/lib/python3/dist-packages (from requests<3,>=2.20.0->mxnet) (2018.1.18)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /usr/lib/python3/dist-packages (from requests<3,>=2.20.0->mxnet) (1.22)
Installing collected packages: graphviz, mxnet
Successfully installed graphviz-0.8.4 mxnet-1.9.0

In [30]:

```
!pip install gluonnl
```

Requirement already satisfied: gluonnl in /home/cuai5th/anaconda3/envs/suminbang/lib/python3.8/site-packages (0.10.0)
Requirement already satisfied: numpy>=1.16.0 in /home/cuai5th/anaconda3/envs/suminbang/lib/python3.8/site-packages (from gluonnl) (1.22.3)
Requirement already satisfied: cython in /home/cuai5th/anaconda3/envs/suminbang/lib/python3.8/site-packages (from gluonnl) (0.29.28)
Requirement already satisfied: packaging in /home/cuai5th/anaconda3/envs/suminbang/lib/python3.8/site-packages (from gluonnl) (21.3)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in /home/cuai5th/anaconda3/envs/suminbang/lib/python3.8/site-packages (from packaging->gluonnl) (3.0.7)

In [32]:

```
import sys
!{sys.executable} -m pip install gluonnlp
```

```
Defaulting to user installation because normal site-packages is not writeable
Collecting gluonnlp
  Using cached gluonnlp-0.10.0.tar.gz (344 kB)
  Preparing metadata (setup.py) ... done
Requirement already satisfied: numpy>=1.16.0 in /home/cuai5th/.local/lib/python3.6/site-packages (from gluonnlp) (1.19.5)
Collecting cython
  Using cached Cython-0.29.28-cp36-cp36m-manylinux_2_17_x86_64.manylinux2014_x86_64.manylinux_2_24_x86_64.whl (1.9 MB)
Collecting packaging
  Using cached packaging-21.3-py3-none-any.whl (40 kB)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in /home/cuai5th/.local/lib/python3.6/site-packages (from packaging->gluonnlp) (3.0.7)
Building wheels for collected packages: gluonnlp
  Building wheel for gluonnlp (setup.py) ... done
  Created wheel for gluonnlp: filename=gluonnlp-0.10.0-cp36-cp36m-linux_x86_64.whl size=588381 sha256=99d953f8cc74563d58f7ed3e7c4cbd14312a836a828929a13a34f56c1873a566
  Stored in directory: /home/cuai5th/.cache/pip/wheels/62/62/9a/53be069ac8c9dde533dacce0e716193a4a43e87b5d37f5008c
Successfully built gluonnlp
Installing collected packages: packaging, cython, gluonnlp
  WARNING: The scripts cygdb, cython and cythonize are installed in '/home/cuai5th/.local/bin' which is not on PATH.
  Consider adding this directory to PATH or, if you prefer to suppress this warning, use --no-warn-script-location.
Successfully installed cython-0.29.28 gluonnlp-0.10.0 packaging-21.3
```

In [37]:

```
import sys
!{sys.executable} -m pip install git+https://git@github.com/SKTBrain/KoBERT.git@master
```

```
Defaulting to user installation because normal site-packages is not writeable
Collecting git+https://****@github.com/SKTBrain/KoBERT.git@master
  Cloning https://****@github.com/SKTBrain/KoBERT.git (to revision master) to /tmp/pip-req-build-luk81j2_
  Running command git clone --filter=blob:none -q 'https://****@github.com/SKTBrain/KoBERT.git' /tmp/pip-req-build-luk81j2_
  Resolved https://****@github.com/SKTBrain/KoBERT.git to commit a82d428c26988ff40c03309038dba813fb83b92e
  Preparing metadata (setup.py) ... done
Collecting boto3
  Using cached boto3-1.21.25-py3-none-any.whl (132 kB)
Requirement already satisfied: gluonnlp>=0.6.0 in /home/cuai5th/.local/lib/python3.6/site-packages (from kobert==0.2.3) (0.10.0)
Requirement already satisfied: mxnet>=1.4.0 in /home/cuai5th/.local/lib/python3.6/site-packages (from kobert==0.2.3) (1.9.0)
Collecting onnxruntime==1.8.0
  Downloading onnxruntime-1.8.0-cp36-cp36m-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (4.5 MB)
  |████████████████████████████████████████| 4.5 MB 4.7 MB/s
Collecting sentencepiece>=0.1.6
  Downloading sentencepiece-0.1.96-cp36-cp36m-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (1.2 MB)
  |████████████████████████████████████████| 1.2 MB 43.8 MB/s
Requirement already satisfied: torch>=1.7.0 in /home/cuai5th/.local/lib/python3.6/site-packages (from kobert==0.2.3) (1.10.1)
Collecting transformers>=4.8.1
  Using cached transformers-4.17.0-py3-none-any.whl (3.8 MB)
Requirement already satisfied: numpy>=1.16.6 in /home/cuai5th/.local/lib/python3.6/site-packages (from onnxruntime==1.8.0->kobert==0.2.3) (1.19.5)
Collecting flatbuffers
  Using cached flatbuffers-2.0-py2.py3-none-any.whl (26 kB)
Requirement already satisfied: protobuf in /home/cuai5th/.local/lib/python3.6/site-packages (from onnxruntime==1.8.0->kobert==0.2.3) (3.19.4)
Requirement already satisfied: cython in /home/cuai5th/.local/lib/python3.6/site-packages (from gluonnlp>=0.6.0->kobert==0.2.3) (0.29.28)
Requirement already satisfied: packaging in /home/cuai5th/.local/lib/python3.6/site-packages (from gluonnlp>=0.6.0->kobert==0.2.3) (21.3)
Requirement already satisfied: graphviz<0.9.0,>=0.8.1 in /home/cuai5th/.local/lib/python3.6/site-packages (from mxnet>=1.4.0->kobert==0.2.3) (0.8.4)
Requirement already satisfied: requests<3,>=2.20.0 in /home/cuai5th/.local/lib/python3.6/site-packages (from mxnet>=1.4.0->kobert==0.2.3) (2.27.1)
Requirement already satisfied: dataclasses in /home/cuai5th/.local/lib/python3.6/site-packages (from torch>=1.7.0->kobert==0.2.3) (0.8)
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.6/dist-packages (from torch>=1.7.0->kobert==0.2.3) (4.0.1)
Collecting huggingface-hub<1.0,>=0.1.0
  Using cached huggingface-hub-0.4.0-py3-none-any.whl (67 kB)
```

```
Requirement already satisfied: filelock in /usr/local/lib/python3.6/dist-packages (from transformers
>=4.8.1->kobert==0.2.3) (3.4.1)
Collecting tokenizers!=0.11.3,>=0.11.1
  Downloading tokenizers-0.11.6-cp36-cp36m-manylinux_2_12_x86_64.manylinux2010_x86_64.whl (6.5 MB)
    |████████████████████████████████████████| 6.5 MB 42.1 MB/s
Collecting pyyaml>=5.1
  Downloading PyYAML-6.0-cp36-cp36m-manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_2_12_x86_64.man
  ylinux2010_x86_64.whl (603 kB)
    |████████████████████████████████████████| 603 kB 60.0 MB/s
Requirement already satisfied: importlib-metadata in /usr/local/lib/python3.6/dist-packages (from tr
ansformers>=4.8.1->kobert==0.2.3) (4.8.3)
Collecting regex!=2019.12.17
  Downloading regex-2022.3.15-cp36-cp36m-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (749 kB)
    |████████████████████████████████████████| 749 kB 62.3 MB/s
Requirement already satisfied: tqdm>=4.27 in /home/cuai5th/.local/lib/python3.6/site-packages (from
transformers>=4.8.1->kobert==0.2.3) (4.63.0)
Collecting sacremoses
  Using cached sacremoses-0.0.49-py3-none-any.whl (895 kB)
Collecting boto3<1.25.0,>=1.24.25
  Downloading boto3-1.24.25-py3-none-any.whl (8.6 MB)
    |████████████████████████████████████████| 8.6 MB 66.9 MB/s
Collecting jmespath<2.0.0,>=0.7.1
  Using cached jmespath-0.10.0-py2.py3-none-any.whl (24 kB)
Collecting s3transfer<0.6.0,>=0.5.0
  Using cached s3transfer-0.5.2-py3-none-any.whl (79 kB)
Requirement already satisfied: python-dateutil<3.0.0,>=2.1 in /home/cuai5th/.local/lib/python3.6/sit
e-packages (from boto3<1.25.0,>=1.24.25->boto3->kobert==0.2.3) (2.8.2)
Collecting urllib3<1.27,>=1.25.4
  Using cached urllib3-1.26.9-py2.py3-none-any.whl (138 kB)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in /home/cuai5th/.local/lib/python3.6/site-p
ackages (from packaging->gluonnp>=0.6.0->kobert==0.2.3) (3.0.7)
Requirement already satisfied: charset-normalizer~=2.0.0 in /home/cuai5th/.local/lib/python3.6/site-
packages (from requests<3,>=2.20.0->mxnet>=1.4.0->kobert==0.2.3) (2.0.12)
Requirement already satisfied: idna<4,>=2.5 in /usr/lib/python3/dist-packages (from requests<3,>=2.2
0.0->mxnet>=1.4.0->kobert==0.2.3) (2.6)
Requirement already satisfied: certifi>=2017.4.17 in /usr/lib/python3/dist-packages (from requests<3
,>=2.20.0->mxnet>=1.4.0->kobert==0.2.3) (2018.1.18)
Requirement already satisfied: importlib-resources in /usr/local/lib/python3.6/dist-packages (from t
qdm>=4.27->transformers>=4.8.1->kobert==0.2.3) (5.4.0)
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.6/dist-packages (from importlib-m
etadata->transformers>=4.8.1->kobert==0.2.3) (3.6.0)
Collecting joblib
  Using cached joblib-1.1.0-py2.py3-none-any.whl (306 kB)
Requirement already satisfied: click in /usr/lib/python3/dist-packages (from sacremoses->transformer
s>=4.8.1->kobert==0.2.3) (6.7)
Requirement already satisfied: six in /usr/lib/python3/dist-packages (from sacremoses->transformers>
=4.8.1->kobert==0.2.3) (1.11.0)
Building wheels for collected packages: kobert
  Building wheel for kobert (setup.py) ... done
  Created wheel for kobert: filename=kobert-0.2.3-py3-none-any.whl size=16029 sha256=21d888c0d82104a
f360d10f57188414195delf2bca62e3df3140d896046be048
  Stored in directory: /tmp/pip-ephem-wheel-cache-luxjrzz4/wheels/4b/9b/3f/583189713d96fda0f291ca1f2
28ab0e0c981641dcdaf7c7cdf
Successfully built kobert
Installing collected packages: urllib3, jmespath, regex, pyyaml, joblib, boto3, tokenizers, sacre
moses, s3transfer, huggingface-hub, flatbuffers, transformers, sentencepiece, onnxruntime, boto3, ko
bert
WARNING: The script sacremoses is installed in '/home/cuai5th/.local/bin' which is not on PATH.
Consider adding this directory to PATH or, if you prefer to suppress this warning, use --no-warn-s
cript-location.
WARNING: The script huggingface-cli is installed in '/home/cuai5th/.local/bin' which is not on PAT
H.
Consider adding this directory to PATH or, if you prefer to suppress this warning, use --no-warn-s
cript-location.
WARNING: The script transformers-cli is installed in '/home/cuai5th/.local/bin' which is not on PA
TH.
Consider adding this directory to PATH or, if you prefer to suppress this warning, use --no-warn-s
cript-location.
WARNING: The script onnxruntime_test is installed in '/home/cuai5th/.local/bin' which is not on PA
TH.
Consider adding this directory to PATH or, if you prefer to suppress this warning, use --no-warn-s
cript-location.
Successfully installed boto3-1.21.25 botocore-1.24.25 flatbuffers-2.0 huggingface-hub-0.4.0 jmespath
-0.10.0 joblib-1.1.0 kobert-0.2.3 onnxruntime-1.8.0 pyyaml-6.0 regex-2022.3.15 s3transfer-0.5.2 sac
remoses-0.0.49 sentencepiece-0.1.96 tokenizers-0.11.6 transformers-4.17.0 urllib3-1.26.9
```

In [41]:

```
train_data.head()
```

Out[41]:

	digit_1	text
0	S	카센터에서 자동차부분정비 타이어오일교환
1	G	상점내에서 일반인을 대상으로 채소.과일판매
2	G	절단하여사업체에도매 공업용고무를가지고 합성고무도매
3	G	영업점에서 일반소비자에게 열쇠잠금장치
4	Q	어린이집 보호자의 위탁을 받아 취학전아동보육

In [42]:

```
train_data['digit_1'].unique()
```

Out[42]:

array(['S', 'G', 'Q', 'C', 'I', 'O', 'F', 'P', 'H', 'N', 'R', 'M', 'J', 'L', 'B', 'K', 'E', 'D', 'A', None], dtype=object)

In [43]:

```
train_data = train_data.dropna(axis=0)
```

In [44]:

```
train_data['digit_1'].unique()
```

Out[44]:

array(['S', 'G', 'Q', 'C', 'I', 'O', 'F', 'P', 'H', 'N', 'M', 'R', 'J', 'L', 'B', 'K', 'E', 'D', 'A'], dtype=object)

In [46]:

```
train_data= train_data.replace({'A':0, 'B':1,'C':2,'D':3,'E':4,'F':5,'G':6,'H':7,'I':8,'J':9,'K':10,'L':11,'M':12, 'N':13,'O':14,'P':15,'Q':16,'R':17,'S':18})
```

In [47]:

```
train_data['digit_1'].unique()
```

Out[47]:

array([18, 6, 16, 2, 8, 14, 5, 15, 7, 13, 12, 17, 9, 11, 1, 10, 4, 3, 0])

In [48]:

```
train_data.head()
```

Out[48]:

	digit_1	text
0	18	카센터에서 자동차부분정비 타이어오일교환
1	6	상점내에서 일반인을 대상으로 채소.과일판매
2	6	절단하여사업체에도매 공업용고무를가지고 합성고무도매
3	6	영업점에서 일반소비자에게 열쇠잠금장치
4	16	어린이집 보호자의 위탁을 받아 취학전아동보육

In [61]:

```
!pip install konlpy
```

```
Collecting konlpy
  Downloading konlpy-0.6.0-py2.py3-none-any.whl (19.4 MB)
    |████████████████████████████████████████| 19.4 MB 4.4 MB/s eta 0:00:01
Collecting lxml>=4.1.0
  Downloading lxml-4.8.0-cp38-cp38-manylinux_2_17_x86_64.manylinux2014_x86_64.manylinux_2_24_x86_64.whl (6.9 MB)
    |████████████████████████████████████████| 6.9 MB 86.0 MB/s eta 0:00:01
Requirement already satisfied: numpy>=1.6 in /home/cuai5th/anaconda3/envs/suminbang/lib/python3.8/site-packages (from konlpy) (1.22.3)
Collecting JPype1>=0.7.0
  Downloading JPype1-1.3.0-cp38-cp38-manylinux_2_5_x86_64.manylinux1_x86_64.whl (449 kB)
    |████████████████████████████████████████| 449 kB 87.8 MB/s eta 0:00:01
Installing collected packages: lxml, JPype1, konlpy
Successfully installed JPype1-1.3.0 konlpy-0.6.0 lxml-4.8.0
```

In [68]:

```
import pandas as pd
import re
from konlpy.tag import Okt, Mecab
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.model_selection import StratifiedKFold
from sklearn.metrics import accuracy_score, f1_score
from lightgbm import LGBMClassifier
```

In [67]:

```
import sys
!{sys.executable} -m pip install lightgbm
```

```
Defaulting to user installation because normal site-packages is not writeable
Collecting lightgbm
  Downloading lightgbm-3.3.2-py3-none-manylinux1_x86_64.whl (2.0 MB)
    |████████████████████████████████████████| 2.0 MB 4.0 MB/s
Requirement already satisfied: numpy in /home/cuai5th/.local/lib/python3.6/site-packages (from lightgbm) (1.19.5)
Requirement already satisfied: wheel in /usr/lib/python3/dist-packages (from lightgbm) (0.30.0)
Requirement already satisfied: scikit-learn!=0.22.0 in /home/cuai5th/.local/lib/python3.6/site-packages (from lightgbm) (0.24.2)
Requirement already satisfied: scipy in /home/cuai5th/.local/lib/python3.6/site-packages (from lightgbm) (1.5.4)
Requirement already satisfied: threadpoolctl>=2.0.0 in /home/cuai5th/.local/lib/python3.6/site-packages (from scikit-learn!=0.22.0->lightgbm) (3.1.0)
Requirement already satisfied: joblib>=0.11 in /home/cuai5th/.local/lib/python3.6/site-packages (from scikit-learn!=0.22.0->lightgbm) (1.1.0)
Installing collected packages: lightgbm
Successfully installed lightgbm-3.3.2
```

In [65]:

```
import sys
!{sys.executable} -m pip install sklearn
```

```
Defaulting to user installation because normal site-packages is not writeable
Collecting sklearn
  Downloading sklearn-0.0.tar.gz (1.1 kB)
  Preparing metadata (setup.py) ... done
Collecting scikit-learn
  Downloading scikit_learn-0.24.2-cp36-cp36m-manylinux2010_x86_64.whl (22.2 MB)
    |████████████████████████████████████████| 22.2 MB 4.1 MB/s
Requirement already satisfied: joblib>=0.11 in /home/cuai5th/.local/lib/python3.6/site-packages (from scikit-learn->sklearn) (1.1.0)
Collecting scipy>=0.19.1
  Downloading scipy-1.5.4-cp36-cp36m-manylinux1_x86_64.whl (25.9 MB)
    |████████████████████████████████████████| 25.9 MB 83.4 MB/s
Collecting threadpoolctl>=2.0.0
  Downloading threadpoolctl-3.1.0-py3-none-any.whl (14 kB)
Requirement already satisfied: numpy>=1.13.3 in /home/cuai5th/.local/lib/python3.6/site-packages (from scikit-learn->sklearn) (1.19.5)
Building wheels for collected packages: sklearn
  Building wheel for sklearn (setup.py) ... done
  Created wheel for sklearn: filename=sklearn-0.0-py2.py3-none-any.whl size=2381 sha256=f8539d1c3e04128541e7f8094e813b17751e17974b685eac342027ac4069d7cf
  Stored in directory: /home/cuai5th/.cache/pip/wheels/23/9d/42/5ec745cbbb17517000a53cecc49d6a865450d1f5cb16dc8a9c
Successfully built sklearn
Installing collected packages: threadpoolctl, scipy, scikit-learn, sklearn
Successfully installed scikit-learn-0.24.2 scipy-1.5.4 sklearn-0.0 threadpoolctl-3.1.0
```


In [63]:

```
import sys
!{sys.executable} -m pip install konlpy
```

```
Defaulting to user installation because normal site-packages is not writeable
Collecting konlpy
  Using cached konlpy-0.6.0-py2.py3-none-any.whl (19.4 MB)
Collecting JPype1>=0.7.0
  Downloading JPype1-1.3.0-cp36-cp36m-manylinux_2_5_x86_64.manylinux1_x86_64.whl (448 kB)
    |████████████████████| 448 kB 4.5 MB/s
Requirement already satisfied: lxml>=4.1.0 in /usr/lib/python3/dist-packages (from konlpy) (4.2.1)
Requirement already satisfied: numpy>=1.6 in /home/cuai5th/.local/lib/python3.6/site-packages (from konlpy) (1.19.5)
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.6/dist-packages (from JPype1>=0.7.0->konlpy) (4.0.1)
Installing collected packages: JPype1, konlpy
Successfully installed JPype1-1.3.0 konlpy-0.6.0
```

In [69]:

```
# 형태소 분석기(OkT) 불러오기
okt=OkT()
```

In [70]:

```
train_data.head()
```

Out[70]:

	digit_1	text
0	18	카센터에서 자동차부분정비 타이어오일교환
1	6	상점내에서 일반인을 대상으로 채소.과일판매
2	6	절단하여사업체에도매 공업용고무를가지고 합성고무도매
3	6	영업점에서 일반소비자에게 열쇠잠금장치
4	16	어린이집 보호자의 위탁을 받아 취학전아동보육

In [76]:

```
# 조사, 어미, 구두점 제거
def func(text):
    clean = []
    for word in okt.pos(text, stem=True): #어간 추출
        if word[1] not in ['Josa', 'Eomi', 'Punctuation']: #조사, 어미, 구두점 제외
            clean.append(word[0])

    return " ".join(clean)

train_data['text'] = train_data['text'].apply(lambda x : func(x))
```

In [77]:

```
# tf-idf를 이용한 벡터화
def split(text):
    tokens_ko = text.split()
    return tokens_ko

tfidf_vect = TfidfVectorizer(tokenizer=split)
tfidf_vect.fit(train_data['text'])
tfidf_matrix_train = tfidf_vect.transform(train_data['text'])
```

```
/home/cuai5th/.local/lib/python3.6/site-packages/sklearn/feature_extraction/text.py:489: UserWarning
: The parameter 'token_pattern' will not be used since 'tokenizer' is not None'
warnings.warn("The parameter 'token_pattern' will not be used")
```

In [84]:

```
# train/valid 데이터 셋 나누기.
def split_dataset(tfidf,df):
    X_data = tfidf
    y_data = df['digit_1']

    # stratify=y_data Stratified 기반 분할, train 데이터의 30%를 평가 데이터 셋으로 사용. (70% 데이터 학습에 사용)
    X_train, X_test, y_train, y_test = \
        train_test_split(X_data, y_data, test_size=0.3, random_state=42, stratify=y_data)

    return (X_train, X_test, y_train, y_test)

X_train, X_test, y_train, y_test = split_dataset(tfidf_matrix_train,train_data)
```

In [85]:

```
lgbm = LGBMClassifier(random_state = 42)
lgbm.fit(X_train,y_train)
```

Out[85]:

```
LGBMClassifier(random_state=42)
```

In [86]:

```
pred = lgbm.predict(X_test)
accuracy = accuracy_score(y_test,pred)

print('정확도', accuracy)
```

정확도 0.9088389462116718