Vikum-AI refactor(project): fix spellings 12af3eb · 2 weeks ago

197 lines (197 loc) · 7.01 KB

Preview Code Blame Raw

In [ ]:
```
# # Imports
import sys
from pathlib import Path

# # Resolve project root and ensure it's on sys.path
ROOT = Path.cwd().resolve()
for _ in range(5):
    if (ROOT / "pyproject.toml").exists() or (ROOT / "raw_data").exists():
        break
    ROOT = ROOT.parent
if str(ROOT) not in sys.path:
    sys.path.insert(0, str(ROOT))

# from katabatic.pipeline.train_test_split.pipeline import TrainTestSplitP
from utils import discretize_preprocess
# from katabatic.models.codi.models import CODI
```

In [5]:
```
# Preprocess data
from pathlib import Path

dataset_path = ROOT / "raw_data" / "car.csv"
output_path = ROOT / "discretized_data" / "car.csv"
output_path.parent.mkdir(parents=True, exist_ok=True)

discretize_preprocess(str(dataset_path), str(output_path))
```

```
Preprocessing: /Users/vikumdabare/Documents/Work/Katebatic/katebatic/raw_dat
a/car.csv
Saved preprocessed discrete dataset to: /Users/vikumdabare/Documents/Work/Ka
tebatic/katebatic/discretized_data/car.csv
```

In [ ]:
```
from katabatic.models.codi.models import CODI
from katabatic.pipeline.train_test_split.pipeline import TrainTestSplitPip
# Run pipeline
input_csv = str(output_path)
output_dir = str(ROOT / "sample_data" / "car")
real_test_dir = output_dir

synthetic_dir = str(ROOT / "synthetic" / "car" / "codi")

# Use lighter training settings and force CPU to avoid kernel crashes
pipeline = TrainTestSplitPipeline(model=lambda: CODI(epochs=10, batch_size
result = pipeline.run(
    input_csv=input_csv,
    output_dir=output_dir,
    synthetic_dir=synthetic_dir,
    real_test_dir=real_test_dir,
)
print(result)
```

```
INFO:katebatic.models.codi.models:==========================================
====================================
INFO:katebatic.models.codi.models:Training CoDi Model
INFO:katebatic.models.codi.models:==========================================
====================================
INFO:katebatic.models.codi.models:Loaded training data: (1382, 7)
INFO:katebatic.models.codi.models:Schema: 0 continuous, 7 categorical column
```

```
s
INFO:katebatic.models.codi.models:Building models: con_dim=1, cat_dim=25
INFO:katebatic.models.codi.models:Continuous model params: 166,427
INFO:katebatic.models.codi.models:Discrete model params: 358,001
INFO:katebatic.models.codi.models:
Training for 10 epochs...
Loaded data with shape: (1728, 7)
Saved train/test full data
Train size: (1382, 7), Test size: (346, 7)
Train label distribution:
 6
2    0.700434
0    0.222142
1    0.039797
3    0.037627
Name: proportion, dtype: float64
Test label distribution:
 6
2    0.699422
0    0.222543
1    0.040462
3    0.037572
Name: proportion, dtype: float64
Saved X/y split
Training shape: (1382, 6) (1382,)
Test shape: (346, 6) (346,)
INFO:katebatic.models.codi.models:Epoch 1/10: loss_con=0.0000, loss_dis=45.1
647
INFO:katebatic.models.codi.models:Epoch 5/10: loss_con=0.0000, loss_dis=44.4
567
INFO:katebatic.models.codi.models:Epoch 10/10: loss_con=0.0000, loss_dis=44.
2655
INFO:katebatic.models.codi.models:
Generating 1382 synthetic samples...
WARNING:katebatic.models.codi.models:Column '6': Adding missing classes {0,
1, 3}
INFO:katebatic.models.codi.models:
Synthetic data saved to: /Users/vikumdabare/Documents/Work/Katebatic/katebat
ic/synthetic/car/codi
INFO:katebatic.models.codi.models:Training complete!
Results saved to: Results/car/codi_tstr.csv

TSTR Evaluation Results:

LR:
Accuracy: 0.6994
F1 Score: 0.5757

MLP:
Accuracy: 0.6994
F1 Score: 0.5757

RF:
Accuracy: 0.6994
```