

Author: Precious Adekwu

Predict Bike Sharing Demand with AutoGluon

Introduction

For this project, I participated in a Kaggle competition to analyze and create models using bike sharing data to predict demand patterns across different time periods. The challenge involved developing a robust machine learning solution that could accurately forecast bike rental counts based on various environmental and temporal factors. Using Amazon's AutoGluon automated machine learning framework, I implemented a systematic approach to model development that progressed through three distinct phases. Starting with a baseline model, I then focused on strategic feature engineering by extracting temporal patterns from datetime data, particularly hour-based features that capture rush hour and peak usage periods. The final phase involved comprehensive hyperparameter optimization across multiple algorithms including XGBoost, Gradient Boosting, CatBoost, Random Forest, and Neural Networks.

Import necessary libraries and load the bike sharing dataset for demand prediction analysis

Set up AutoGluon and other ML tools for automated model training and evaluation

```
In [9]: !pip install -U pip
!pip install -U setuptools wheel
!pip install -U "mxnet<2.0.0" bokeh==2.0.1
!pip install autogluon --no-cache-dir
# Without --no-cache-dir, smaller aws instances may have trouble installing
```

Requirement already satisfied: pip in /opt/conda/lib/python3.12/site-packages (25.1.1)
 Requirement already satisfied: setuptools in /opt/conda/lib/python3.12/site-packages (80.9.0)
 Requirement already satisfied: wheel in /opt/conda/lib/python3.12/site-packages (0.45.1)
 Collecting mxnet<2.0.0
 Using cached mxnet-1.9.1-py3-none-manylinux2014_x86_64.whl.metadata (3.4 kB)
 Collecting bokeh==2.0.1
 Using cached bokeh-2.0.1.tar.gz (8.6 MB)
 Preparing metadata (setup.py) ... error
error: subprocess-exited-with-error

```

× python setup.py egg_info did not run successfully.
  exit code: 1
  ↳ [21 lines of output]
    /tmp/pip-install-7qawl37s/bokeh_072de882809f44d192e88c4537b34893/versioneer.py:416: SyntaxWarning: invalid escape sequence '\s'
      LONG_VERSION_PY['git'] = '''
    Traceback (most recent call last):
      File "<string>", line 2, in <module>
      File "<pip-setuptools-caller>", line 35, in <module>
      File "/tmp/pip-install-7qawl37s/bokeh_072de882809f44d192e88c4537b34893/setup.py", line 118, in <module>
        version=get_version(),
              ^^^^^^^^^^^^^^^
      File "/tmp/pip-install-7qawl37s/bokeh_072de882809f44d192e88c4537b34893/_setup_support.py", line 243, in get_version
        return versioneer.get_version()
              ^^^^^^^^^^^^^^^^^^^^^^^
      File "/tmp/pip-install-7qawl37s/bokeh_072de882809f44d192e88c4537b34893/versioneer.py", line 1484, in get_version
        return get_versions()["version"]
              ^^^^^^^^^^^^^^^
      File "/tmp/pip-install-7qawl37s/bokeh_072de882809f44d192e88c4537b34893/versioneer.py", line 1416, in get_versions
        cfg = get_config_from_root(root)
              ^^^^^^^^^^^^^^^^^^^^^^^
      File "/tmp/pip-install-7qawl37s/bokeh_072de882809f44d192e88c4537b34893/versioneer.py", line 340, in get_config_from_root
        parser = configparser.SafeConfigParser()
              ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
    AttributeError: module 'configparser' has no attribute 'SafeConfigParser'. Did you mean: 'RawConfigParser'?
    [end of output]
  
```

note: This error originates from a subprocess, and is likely not a problem with pip.
error: metadata-generation-failed

× Encountered error while generating package metadata.
 ↳ See above for output.

note: This is an issue with the package mentioned above, not pip.

hint: See above for details.

Requirement already satisfied: autogluon in /opt/conda/lib/python3.12/site-packages (1.3.0)
 Requirement already satisfied: autogluon.core==1.3.0 in /opt/conda/lib/python3.12/site-packages (from autogluon.core[all]==1.3.0->autogluon) (1.3.0)
 Requirement already satisfied: autogluon.features==1.3.0 in /opt/conda/lib/python3.12/site-packages (from autogluon) (1.3.0)
 Requirement already satisfied: autogluon.tabular==1.3.0 in /opt/conda/lib/python3.12/site-packages (from autogluon.tabular[all]==1.3.0->autogluon) (1.3.0)

3.0)

Requirement already satisfied: autogluon.multimodal==1.3.0 in /opt/conda/lib/python3.12/site-packages (from autogluon) (1.3.0)

Requirement already satisfied: autogluon.timeseries==1.3.0 in /opt/conda/lib/python3.12/site-packages (from autogluon.timeseries[all]==1.3.0->autogluon) (1.3.0)

Requirement already satisfied: numpy<2.3.0,>=1.25.0 in /opt/conda/lib/python3.12/site-packages (from autogluon.core==1.3.0->autogluon.core[all]==1.3.0->autogluon) (1.26.4)

Requirement already satisfied: scipy<1.16,>=1.5.4 in /opt/conda/lib/python3.12/site-packages (from autogluon.core==1.3.0->autogluon.core[all]==1.3.0->autogluon) (1.15.2)

Requirement already satisfied: scikit-learn<1.7.0,>=1.4.0 in /opt/conda/lib/python3.12/site-packages (from autogluon.core==1.3.0->autogluon.core[all]==1.3.0->autogluon) (1.6.1)

Requirement already satisfied: networkx<4,>=3.0 in /opt/conda/lib/python3.12/site-packages (from autogluon.core==1.3.0->autogluon.core[all]==1.3.0->autogluon) (3.4.2)

Requirement already satisfied: pandas<2.3.0,>=2.0.0 in /opt/conda/lib/python3.12/site-packages (from autogluon.core==1.3.0->autogluon.core[all]==1.3.0->autogluon) (2.2.3)

Requirement already satisfied: tqdm<5,>=4.38 in /opt/conda/lib/python3.12/site-packages (from autogluon.core==1.3.0->autogluon.core[all]==1.3.0->autogluon) (4.67.1)

Requirement already satisfied: requests in /opt/conda/lib/python3.12/site-packages (from autogluon.core==1.3.0->autogluon.core[all]==1.3.0->autogluon) (2.32.3)

Requirement already satisfied: matplotlib<3.11,>=3.7.0 in /opt/conda/lib/python3.12/site-packages (from autogluon.core==1.3.0->autogluon.core[all]==1.3.0->autogluon) (3.10.3)

Requirement already satisfied: boto3<2,>=1.10 in /opt/conda/lib/python3.12/site-packages (from autogluon.core==1.3.0->autogluon.core[all]==1.3.0->autogluon) (1.37.1)

Requirement already satisfied: autogluon.common==1.3.0 in /opt/conda/lib/python3.12/site-packages (from autogluon.core==1.3.0->autogluon.core[all]==1.3.0->autogluon) (1.3.0)

Requirement already satisfied: psutil<7.1.0,>=5.7.3 in /opt/conda/lib/python3.12/site-packages (from autogluon.common==1.3.0->autogluon.core==1.3.0->autogluon.core[all]==1.3.0->autogluon) (5.9.8)

Requirement already satisfied: hyperopt<0.2.8,>=0.2.7 in /opt/conda/lib/python3.12/site-packages (from autogluon.core[all]==1.3.0->autogluon) (0.2.7)

Requirement already satisfied: ray<2.45,>=2.10.0 in /opt/conda/lib/python3.12/site-packages (from ray[default,tune]<2.45,>=2.10.0; extra == "all"->autogluon.core[all]==1.3.0->autogluon) (2.44.1)

Requirement already satisfied: pyarrow>=15.0.0 in /opt/conda/lib/python3.12/site-packages (from autogluon.core[all]==1.3.0->autogluon) (19.0.1)

Requirement already satisfied: Pillow<12,>=10.0.1 in /opt/conda/lib/python3.12/site-packages (from autogluon.multimodal==1.3.0->autogluon) (11.2.1)

Requirement already satisfied: torch<2.7,>=2.2 in /opt/conda/lib/python3.12/site-packages (from autogluon.multimodal==1.3.0->autogluon) (2.6.0)

Requirement already satisfied: lightning<2.7,>=2.2 in /opt/conda/lib/python3.12/site-packages (from autogluon.multimodal==1.3.0->autogluon) (2.5.1.post0)

Requirement already satisfied: transformers<4.50,>=4.38.0 in /opt/conda/lib/python3.12/site-packages (from transformers[sentencepiece]<4.50,>=4.38.0->autogluon.multimodal==1.3.0->autogluon) (4.49.0)

Requirement already satisfied: accelerate<2.0,>=0.34.0 in /opt/conda/lib/python3.12/site-packages (from autogluon.multimodal==1.3.0->autogluon) (1.5.2)

Requirement already satisfied: jsonschema<4.24,>=4.18 in /opt/conda/lib/python3.12/site-packages (from autogluon.multimodal==1.3.0->autogluon) (4.23.0)

Requirement already satisfied: sequeval<1.3.0,>=1.2.2 in /opt/conda/lib/python3.12/site-packages (from autogluon.multimodal==1.3.0->autogluon) (1.2.2)

Requirement already satisfied: evaluate<0.5.0,>=0.4.0 in /opt/conda/lib/python3.12/site-packages (from autogluon.multimodal==1.3.0->autogluon) (0.4.1)

Requirement already satisfied: timm<1.0.7,>=0.9.5 in /opt/conda/lib/python3.12/site-packages (from autogluon.multimodal==1.3.0->autogluon) (1.0.3)

Requirement already satisfied: torchvision<0.22.0,>=0.16.0 in /opt/conda/lib/python3.12/site-packages (from autogluon.multimodal==1.3.0->autogluon) (0.21.0)

Requirement already satisfied: scikit-image<0.26.0,>=0.19.1 in /opt/conda/lib/python3.12/site-packages (from autogluon.multimodal==1.3.0->autogluon) (0.25.2)

Requirement already satisfied: text-unidecode<1.4,>=1.3 in /opt/conda/lib/python3.12/site-packages (from autogluon.multimodal==1.3.0->autogluon) (1.3)

Requirement already satisfied: torchmetrics<1.8,>=1.2.0 in /opt/conda/lib/python3.12/site-packages (from autogluon.multimodal==1.3.0->autogluon) (1.7.1)

Requirement already satisfied: omegaconf<2.4.0,>=2.1.1 in /opt/conda/lib/python3.12/site-packages (from autogluon.multimodal==1.3.0->autogluon) (2.3.0)

Requirement already satisfied: pytorch-metric-learning<2.9,>=1.3.0 in /opt/conda/lib/python3.12/site-packages (from autogluon.multimodal==1.3.0->autogluon) (2.8.1)

Requirement already satisfied: nlpaug<1.2.0,>=1.1.10 in /opt/conda/lib/python3.12/site-packages (from autogluon.multimodal==1.3.0->autogluon) (1.1.11)

Requirement already satisfied: nltk<4.0,>=3.4.5 in /opt/conda/lib/python3.12/site-packages (from autogluon.multimodal==1.3.0->autogluon) (3.9.1)

Requirement already satisfied: openmim<0.4.0,>=0.3.7 in /opt/conda/lib/python3.12/site-packages (from autogluon.multimodal==1.3.0->autogluon) (0.3.7)

Requirement already satisfied: defusedxml<0.7.2,>=0.7.1 in /opt/conda/lib/python3.12/site-packages (from autogluon.multimodal==1.3.0->autogluon) (0.7.1)

Requirement already satisfied: jinja2<3.2,>=3.0.3 in /opt/conda/lib/python3.12/site-packages (from autogluon.multimodal==1.3.0->autogluon) (3.1.6)

Requirement already satisfied: tensorboard<3,>=2.9 in /opt/conda/lib/python3.12/site-packages (from autogluon.multimodal==1.3.0->autogluon) (2.18.0)

Requirement already satisfied: pyteseract<0.4,>=0.3.9 in /opt/conda/lib/python3.12/site-packages (from autogluon.multimodal==1.3.0->autogluon) (0.3.13)

Requirement already satisfied: nvidia-ml-py3<8.0,>=7.352.0 in /opt/conda/lib/python3.12/site-packages (from autogluon.multimodal==1.3.0->autogluon) (7.352.0)

Requirement already satisfied: pdf2image<1.19,>=1.17.0 in /opt/conda/lib/python3.12/site-packages (from autogluon.multimodal==1.3.0->autogluon) (1.17.0)

Requirement already satisfied: catboost<1.3,>=1.2 in /opt/conda/lib/python3.12/site-packages (from autogluon.tabular[all]==1.3.0->autogluon) (1.2.7)

Requirement already satisfied: einops<0.9,>=0.7 in /opt/conda/lib/python3.12/site-packages (from autogluon.tabular[all]==1.3.0->autogluon) (0.8.1)

Requirement already satisfied: spacy<3.9 in /opt/conda/lib/python3.12/site-packages (from autogluon.tabular[all]==1.3.0->autogluon) (3.8.5)

Requirement already satisfied: xgboost<3.1,>=2.0 in /opt/conda/lib/python3.12/site-packages (from autogluon.tabular[all]==1.3.0->autogluon) (2.1.4)

Requirement already satisfied: huggingface_hub[torch] in /opt/conda/lib/python3.12/site-packages (from autogluon.tabular[all]==1.3.0->autogluon) (0.30.2)

Requirement already satisfied: lightgbm<4.7,>=4.0 in /opt/conda/lib/python3.12/site-packages (from autogluon.tabular[all]==1.3.0->autogluon) (4.6.0)

Requirement already satisfied: fastai<2.9,>=2.3.1 in /opt/conda/lib/python3.12/site-packages (from autogluon.tabular[all]==1.3.0->autogluon) (2.7.19)

Requirement already satisfied: joblib<2,>=1.1 in /opt/conda/lib/python3.12/site-packages (from autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (1.5.0)

Requirement already satisfied: pytorch_lightning in /opt/conda/lib/python3.12/site-packages (from autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (2.5.1.post0)

Requirement already satisfied: gluonts<0.17,>=0.15.0 in /opt/conda/lib/python3.12/site-packages (from autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (0.16.1)

Requirement already satisfied: statsforecast<2.0.2,>=1.7.0 in /opt/conda/lib/python3.12/site-packages (from autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (2.0.1)

Requirement already satisfied: mlforecast<0.14,>=0.13 in /opt/conda/lib/python3.12/site-packages (from autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (0.13.6)

Requirement already satisfied: utilsforecast<0.2.11,>=0.2.3 in /opt/conda/lib/python3.12/site-packages (from autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (0.2.10)

Requirement already satisfied: coreforecast<0.0.16,>=0.0.12 in /opt/conda/lib/python3.12/site-packages (from autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (0.0.15)

Requirement already satisfied: fugue>=0.9.0 in /opt/conda/lib/python3.12/site-packages (from autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (0.9.1)

Requirement already satisfied: orjson~=3.9 in /opt/conda/lib/python3.12/site-packages (from autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (3.10.18)

Requirement already satisfied: packaging>=20.0 in /opt/conda/lib/python3.12/site-packages (from accelerate<2.0,>=0.34.0->autogluon.multimodal==1.3.0->autogluon) (24.2)

Requirement already satisfied: pyyaml in /opt/conda/lib/python3.12/site-packages (from accelerate<2.0,>=0.34.0->autogluon.multimodal==1.3.0->autogluon) (6.0.2)

Requirement already satisfied: safetensors>=0.4.3 in /opt/conda/lib/python3.12/site-packages (from accelerate<2.0,>=0.34.0->autogluon.multimodal==1.3.0->autogluon) (0.5.3)

Requirement already satisfied: boto3<1.38.0,>=1.37.1 in /opt/conda/lib/python3.12/site-packages (from boto3<2,>=1.10->autogluon.core==1.3.0->autogluon.core[all]==1.3.0->autogluon) (1.37.1)

Requirement already satisfied: jmespath<2.0.0,>=0.7.1 in /opt/conda/lib/python3.12/site-packages (from boto3<2,>=1.10->autogluon.core==1.3.0->autogluon.core[all]==1.3.0->autogluon) (1.0.1)

Requirement already satisfied: s3transfer<0.12.0,>=0.11.0 in /opt/conda/lib/python3.12/site-packages (from boto3<2,>=1.10->autogluon.core==1.3.0->autogluon.core[all]==1.3.0->autogluon) (0.11.3)

Requirement already satisfied: python-dateutil<3.0.0,>=2.1 in /opt/conda/lib/python3.12/site-packages (from boto3<2,>=1.10->autogluon.core==1.3.0->autogluon.core[all]==1.3.0->autogluon) (2.9.0.post0)

Requirement already satisfied: urllib3!=2.2.0,<3,>=1.25.4 in /opt/conda/lib/python3.12/site-packages (from boto3<2,>=1.10->autogluon.core==1.3.0->autogluon.core[all]==1.3.0->autogluon) (1.26.19)

Requirement already satisfied: graphviz in /opt/conda/lib/python3.12/site-packages (from catboost<1.3,>=1.2->autogluon.tabular[all]==1.3.0->autogluon) (0.20.3)

Requirement already satisfied: plotly in /opt/conda/lib/python3.12/site-packages (from catboost<1.3,>=1.2->autogluon.tabular[all]==1.3.0->autogluon) (6.0.1)

Requirement already satisfied: six in /opt/conda/lib/python3.12/site-packages (from catboost<1.3,>=1.2->autogluon.tabular[all]==1.3.0->autogluon) (1.17.0)

Requirement already satisfied: datasets>=2.0.0 in /opt/conda/lib/python3.12/site-packages (from evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.3.0->autogluon) (2.2.1)

Requirement already satisfied: dill in /opt/conda/lib/python3.12/site-packages (from evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.3.0->autogluon) (0.4.0)

Requirement already satisfied: xxhash in /opt/conda/lib/python3.12/site-packages (from evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.3.0->autogluon) (3.5.0)

Requirement already satisfied: multiprocessing in /opt/conda/lib/python3.12/si

te-packages (from evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.3.0->autogluon) (0.70.18)

Requirement already satisfied: fsspec>=2021.05.0 in /opt/conda/lib/python3.12/site-packages (from fsspec[http]>=2021.05.0->evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.3.0->autogluon) (2024.10.0)

Requirement already satisfied: responses<0.19 in /opt/conda/lib/python3.12/site-packages (from evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.3.0->autogluon) (0.18.0)

Requirement already satisfied: pip in /opt/conda/lib/python3.12/site-packages (from fastai<2.9,>=2.3.1->autogluon.tabular[all]==1.3.0->autogluon) (25.1.1)

Requirement already satisfied: fastdownload<2,>=0.0.5 in /opt/conda/lib/python3.12/site-packages (from fastai<2.9,>=2.3.1->autogluon.tabular[all]==1.3.0->autogluon) (0.0.7)

Requirement already satisfied: fastcore<1.8,>=1.5.29 in /opt/conda/lib/python3.12/site-packages (from fastai<2.9,>=2.3.1->autogluon.tabular[all]==1.3.0->autogluon) (1.7.20)

Requirement already satisfied: fastprogress>=0.2.4 in /opt/conda/lib/python3.12/site-packages (from fastai<2.9,>=2.3.1->autogluon.tabular[all]==1.3.0->autogluon) (1.0.3)

Requirement already satisfied: pydantic<3,>=1.7 in /opt/conda/lib/python3.12/site-packages (from gluonts<0.17,>=0.15.0->autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (2.11.4)

Requirement already satisfied: toolz~0.10 in /opt/conda/lib/python3.12/site-packages (from gluonts<0.17,>=0.15.0->autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (0.12.1)

Requirement already satisfied: typing-extensions~4.0 in /opt/conda/lib/python3.12/site-packages (from gluonts<0.17,>=0.15.0->autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (4.13.2)

Requirement already satisfied: future in /opt/conda/lib/python3.12/site-packages (from hyperopt<0.2.8,>=0.2.7->autogluon.core[all]==1.3.0->autogluon) (1.0.0)

Requirement already satisfied: cloudpickle in /opt/conda/lib/python3.12/site-packages (from hyperopt<0.2.8,>=0.2.7->autogluon.core[all]==1.3.0->autogluon) (3.1.1)

Requirement already satisfied: py4j in /opt/conda/lib/python3.12/site-packages (from hyperopt<0.2.8,>=0.2.7->autogluon.core[all]==1.3.0->autogluon) (0.10.9.9)

Requirement already satisfied: MarkupSafe>=2.0 in /opt/conda/lib/python3.12/site-packages (from jinja2<3.2,>=3.0.3->autogluon.multimodal==1.3.0->autogluon) (3.0.2)

Requirement already satisfied: attrs>=22.2.0 in /opt/conda/lib/python3.12/site-packages (from jsonschema<4.24,>=4.18->autogluon.multimodal==1.3.0->autogluon) (23.2.0)

Requirement already satisfied: jsonschema-specifications>=2023.03.6 in /opt/conda/lib/python3.12/site-packages (from jsonschema<4.24,>=4.18->autogluon.multimodal==1.3.0->autogluon) (2025.4.1)

Requirement already satisfied: referencing>=0.28.4 in /opt/conda/lib/python3.12/site-packages (from jsonschema<4.24,>=4.18->autogluon.multimodal==1.3.0->autogluon) (0.36.2)

Requirement already satisfied: rpds-py>=0.7.1 in /opt/conda/lib/python3.12/site-packages (from jsonschema<4.24,>=4.18->autogluon.multimodal==1.3.0->autogluon) (0.24.0)

Requirement already satisfied: lightning-utilities<2.0,>=0.10.0 in /opt/conda/lib/python3.12/site-packages (from lightning<2.7,>=2.2->autogluon.multimodal==1.3.0->autogluon) (0.14.3)

Requirement already satisfied: aiohttp!=4.0.0a0,!4.0.0a1 in /opt/conda/lib/python3.12/site-packages (from fsspec[http]>=2021.05.0->evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.3.0->autogluon) (3.9.5)

Requirement already satisfied: setuptools in /opt/conda/lib/python3.12/site-packages (from lightning-utilities<2.0,>=0.10.0->lightning<2.7,>=2.2->autogluon.multimodal==1.3.0->autogluon) (80.9.0)

Requirement already satisfied: contourpy>=1.0.1 in /opt/conda/lib/python3.12/site-packages (from matplotlib<3.11,>=3.7.0->autogluon.core==1.3.0->autogluon) (1.3.0)

luon.core[all]==1.3.0->autogluon) (1.3.2)
Requirement already satisfied: cycycler>=0.10 in /opt/conda/lib/python3.12/site-packages (from matplotlib<3.11,>=3.7.0->autogluon.core==1.3.0->autogluon.core[all]==1.3.0->autogluon) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /opt/conda/lib/python3.12/site-packages (from matplotlib<3.11,>=3.7.0->autogluon.core==1.3.0->autogluon.core[all]==1.3.0->autogluon) (4.57.0)
Requirement already satisfied: kiwisolver>=1.3.1 in /opt/conda/lib/python3.12/site-packages (from matplotlib<3.11,>=3.7.0->autogluon.core==1.3.0->autogluon.core[all]==1.3.0->autogluon) (1.4.8)
Requirement already satisfied: pyparsing>=2.3.1 in /opt/conda/lib/python3.12/site-packages (from matplotlib<3.11,>=3.7.0->autogluon.core==1.3.0->autogluon.core[all]==1.3.0->autogluon) (3.2.3)
Requirement already satisfied: numba in /opt/conda/lib/python3.12/site-packages (from mlforecast<0.14,>0.13->autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (0.61.2)
Requirement already satisfied: optuna in /opt/conda/lib/python3.12/site-packages (from mlforecast<0.14,>0.13->autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (4.3.0)
Requirement already satisfied: window-ops in /opt/conda/lib/python3.12/site-packages (from mlforecast<0.14,>0.13->autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (0.0.15)
Requirement already satisfied: gdown>=4.0.0 in /opt/conda/lib/python3.12/site-packages (from nlpaug<1.2.0,>=1.1.10->autogluon.multimodal==1.3.0->autogluon) (5.2.0)
Requirement already satisfied: click in /opt/conda/lib/python3.12/site-packages (from nltk<4.0,>=3.4.5->autogluon.multimodal==1.3.0->autogluon) (8.1.8)
Requirement already satisfied: regex>=2021.8.3 in /opt/conda/lib/python3.12/site-packages (from nltk<4.0,>=3.4.5->autogluon.multimodal==1.3.0->autogluon) (2024.11.6)
Requirement already satisfied: antlr4-python3-runtime==4.9.* in /opt/conda/lib/python3.12/site-packages (from omegaconf<2.4.0,>=2.1.1->autogluon.multimodal==1.3.0->autogluon) (4.9.3)
Requirement already satisfied: colorama in /opt/conda/lib/python3.12/site-packages (from openmim<0.4.0,>=0.3.7->autogluon.multimodal==1.3.0->autogluon) (0.4.6)
Requirement already satisfied: model-index in /opt/conda/lib/python3.12/site-packages (from openmim<0.4.0,>=0.3.7->autogluon.multimodal==1.3.0->autogluon) (0.1.11)
Requirement already satisfied: rich in /opt/conda/lib/python3.12/site-packages (from openmim<0.4.0,>=0.3.7->autogluon.multimodal==1.3.0->autogluon) (14.0.0)
Requirement already satisfied: tabulate in /opt/conda/lib/python3.12/site-packages (from openmim<0.4.0,>=0.3.7->autogluon.multimodal==1.3.0->autogluon) (0.9.0)
Requirement already satisfied: pytz>=2020.1 in /opt/conda/lib/python3.12/site-packages (from pandas<2.3.0,>=2.0.0->autogluon.core==1.3.0->autogluon.core[all]==1.3.0->autogluon) (2024.2)
Requirement already satisfied: tzdata>=2022.7 in /opt/conda/lib/python3.12/site-packages (from pandas<2.3.0,>=2.0.0->autogluon.core==1.3.0->autogluon.core[all]==1.3.0->autogluon) (2025.2)
Requirement already satisfied: annotated-types>=0.6.0 in /opt/conda/lib/python3.12/site-packages (from pydantic<3,>=1.7->gluonts<0.17,>=0.15.0->autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (0.7.0)
Requirement already satisfied: pydantic-core==2.33.2 in /opt/conda/lib/python3.12/site-packages (from pydantic<3,>=1.7->gluonts<0.17,>=0.15.0->autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (2.33.2)
Requirement already satisfied: typing-inspection>=0.4.0 in /opt/conda/lib/python3.12/site-packages (from pydantic<3,>=1.7->gluonts<0.17,>=0.15.0->autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (0.4.0)
Requirement already satisfied: charset_normalizer<4,>=2 in /opt/conda/lib/python3.12/site-packages (from requests->autogluon.core==1.3.0->autogluon.co

```
re[all]==1.3.0->autogluon) (3.4.2)
Requirement already satisfied: idna<4,>=2.5 in /opt/conda/lib/python3.12/site-packages (from requests->autogluon.core==1.3.0->autogluon.core[all]==1.3.0->autogluon) (3.10)
Requirement already satisfied: certifi>=2017.4.17 in /opt/conda/lib/python3.12/site-packages (from requests->autogluon.core==1.3.0->autogluon.core[all]==1.3.0->autogluon) (2025.4.26)
Requirement already satisfied: imageio!=2.35.0,>=2.33 in /opt/conda/lib/python3.12/site-packages (from scikit-image<0.26.0,>=0.19.1->autogluon.multimodal==1.3.0->autogluon) (2.37.0)
Requirement already satisfied: tifffile>=2022.8.12 in /opt/conda/lib/python3.12/site-packages (from scikit-image<0.26.0,>=0.19.1->autogluon.multimodal==1.3.0->autogluon) (2025.3.30)
Requirement already satisfied: lazy-loader>=0.4 in /opt/conda/lib/python3.12/site-packages (from scikit-image<0.26.0,>=0.19.1->autogluon.multimodal==1.3.0->autogluon) (0.4)
Requirement already satisfied: threadpoolctl>=3.1.0 in /opt/conda/lib/python3.12/site-packages (from scikit-learn<1.7.0,>=1.4.0->autogluon.core==1.3.0->autogluon.core[all]==1.3.0->autogluon) (3.6.0)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /opt/conda/lib/python3.12/site-packages (from spacy<3.9->autogluon.tabular[all]==1.3.0->autogluon) (3.0.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /opt/conda/lib/python3.12/site-packages (from spacy<3.9->autogluon.tabular[all]==1.3.0->autogluon) (1.0.5)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /opt/conda/lib/python3.12/site-packages (from spacy<3.9->autogluon.tabular[all]==1.3.0->autogluon) (1.0.10)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /opt/conda/lib/python3.12/site-packages (from spacy<3.9->autogluon.tabular[all]==1.3.0->autogluon) (2.0.11)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /opt/conda/lib/python3.12/site-packages (from spacy<3.9->autogluon.tabular[all]==1.3.0->autogluon) (3.0.9)
Requirement already satisfied: thinc<8.4.0,>=8.3.4 in /opt/conda/lib/python3.12/site-packages (from spacy<3.9->autogluon.tabular[all]==1.3.0->autogluon) (8.3.4)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /opt/conda/lib/python3.12/site-packages (from spacy<3.9->autogluon.tabular[all]==1.3.0->autogluon) (1.1.3)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /opt/conda/lib/python3.12/site-packages (from spacy<3.9->autogluon.tabular[all]==1.3.0->autogluon) (2.5.1)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /opt/conda/lib/python3.12/site-packages (from spacy<3.9->autogluon.tabular[all]==1.3.0->autogluon) (2.0.10)
Requirement already satisfied: weasel<0.5.0,>=0.1.0 in /opt/conda/lib/python3.12/site-packages (from spacy<3.9->autogluon.tabular[all]==1.3.0->autogluon) (0.4.1)
Requirement already satisfied: typer<1.0.0,>=0.3.0 in /opt/conda/lib/python3.12/site-packages (from spacy<3.9->autogluon.tabular[all]==1.3.0->autogluon) (0.15.3)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in /opt/conda/lib/python3.12/site-packages (from spacy<3.9->autogluon.tabular[all]==1.3.0->autogluon) (3.4.1)
Requirement already satisfied: language-data>=1.2 in /opt/conda/lib/python3.12/site-packages (from langcodes<4.0.0,>=3.2.0->spacy<3.9->autogluon.tabular[all]==1.3.0->autogluon) (1.3.0)
Requirement already satisfied: ujson>=1.35 in /opt/conda/lib/python3.12/site-packages (from srsly<3.0.0,>=2.4.3->spacy<3.9->autogluon.tabular[all]==1.3.0->autogluon) (5.10.0)
Requirement already satisfied: statsmodels>=0.13.2 in /opt/conda/lib/python3.12/site-packages (from statsforecast<2.0.2,>=1.7.0->autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (0.14.4)
```


Requirement already satisfied: absl-py>=0.4 in /opt/conda/lib/python3.12/site-packages (from tensorboard<3,>=2.9->autogluon.multimodal==1.3.0->autogluon) (2.2.0)

Requirement already satisfied: grpcio>=1.48.2 in /opt/conda/lib/python3.12/site-packages (from tensorboard<3,>=2.9->autogluon.multimodal==1.3.0->autogluon) (1.67.1)

Requirement already satisfied: markdown>=2.6.8 in /opt/conda/lib/python3.12/site-packages (from tensorboard<3,>=2.9->autogluon.multimodal==1.3.0->autogluon) (3.8)

Requirement already satisfied: protobuf in /opt/conda/lib/python3.12/site-packages (from tensorboard<3,>=2.9->autogluon.multimodal==1.3.0->autogluon) (5.28.3)

Requirement already satisfied: tensorboard-data-server<0.8.0,>=0.7.0 in /opt/conda/lib/python3.12/site-packages (from tensorboard<3,>=2.9->autogluon.multimodal==1.3.0->autogluon) (0.7.0)

Requirement already satisfied: werkzeug>=1.0.1 in /opt/conda/lib/python3.12/site-packages (from tensorboard<3,>=2.9->autogluon.multimodal==1.3.0->autogluon) (3.1.3)

Requirement already satisfied: blis<1.3.0,>=1.2.0 in /opt/conda/lib/python3.12/site-packages (from thinc<8.4.0,>=8.3.4->spacy<3.9->autogluon.tabular[all]==1.3.0->autogluon) (1.2.1)

Requirement already satisfied: confection<1.0.0,>=0.0.1 in /opt/conda/lib/python3.12/site-packages (from thinc<8.4.0,>=8.3.4->spacy<3.9->autogluon.tabular[all]==1.3.0->autogluon) (0.1.5)

Requirement already satisfied: filelock in /opt/conda/lib/python3.12/site-packages (from torch<2.7,>=2.2->autogluon.multimodal==1.3.0->autogluon) (3.18.0)

Requirement already satisfied: sympy!=1.13.2,>=1.13.1 in /opt/conda/lib/python3.12/site-packages (from torch<2.7,>=2.2->autogluon.multimodal==1.3.0->autogluon) (1.14.0)

Requirement already satisfied: tokenizers<0.22,>=0.21 in /opt/conda/lib/python3.12/site-packages (from transformers<4.50,>=4.38.0->transformers[sentencepiece]<4.50,>=4.38.0->autogluon.multimodal==1.3.0->autogluon) (0.21.1)

Requirement already satisfied: sentencepiece!=0.1.92,>=0.1.91 in /opt/conda/lib/python3.12/site-packages (from transformers[sentencepiece]<4.50,>=4.38.0->autogluon.multimodal==1.3.0->autogluon) (0.2.0)

Requirement already satisfied: shellingham>=1.3.0 in /opt/conda/lib/python3.12/site-packages (from typer<1.0.0,>=0.3.0->spacy<3.9->autogluon.tabular[all]==1.3.0->autogluon) (1.5.4)

Requirement already satisfied: cloudpathlib<1.0.0,>=0.7.0 in /opt/conda/lib/python3.12/site-packages (from weasel<0.5.0,>=0.1.0->spacy<3.9->autogluon.tabular[all]==1.3.0->autogluon) (0.21.0)

Requirement already satisfied: smart-open<8.0.0,>=5.2.1 in /opt/conda/lib/python3.12/site-packages (from weasel<0.5.0,>=0.1.0->spacy<3.9->autogluon.tabular[all]==1.3.0->autogluon) (7.1.0)

Requirement already satisfied: wrapt in /opt/conda/lib/python3.12/site-packages (from smart-open<8.0.0,>=5.2.1->weasel<0.5.0,>=0.1.0->spacy<3.9->autogluon.tabular[all]==1.3.0->autogluon) (1.17.2)

Requirement already satisfied: aiosignal>=1.1.2 in /opt/conda/lib/python3.12/site-packages (from aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]>=2021.05.0->evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.3.0->autogluon) (1.3.2)

Requirement already satisfied: frozenlist>=1.1.1 in /opt/conda/lib/python3.12/site-packages (from aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]>=2021.05.0->evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.3.0->autogluon) (1.6.0)

Requirement already satisfied: multidict<7.0,>=4.5 in /opt/conda/lib/python3.12/site-packages (from aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]>=2021.05.0->evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.3.0->autogluon) (6.4.3)

Requirement already satisfied: yarl<2.0,>=1.0 in /opt/conda/lib/python3.12/site-packages (from aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]>=2021.05.0->evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.3.0->autogluon) (1.20.0)

Requirement already satisfied: propcache>=0.2.1 in /opt/conda/lib/python3.12/site-packages (from yarl<2.0,>=1.0->aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]>=2021.05.0->evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.3.0->autogluon) (0.3.1)

Requirement already satisfied: triad>=0.9.7 in /opt/conda/lib/python3.12/site-packages (from fugue>=0.9.0->autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (0.9.8)

Requirement already satisfied: adagio>=0.2.4 in /opt/conda/lib/python3.12/site-packages (from fugue>=0.9.0->autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (0.2.6)

Requirement already satisfied: beautifulsoup4 in /opt/conda/lib/python3.12/site-packages (from gdown>=4.0.0->nlpaug<1.2.0,>=1.1.10->autogluon.multimodal==1.3.0->autogluon) (4.13.4)

Requirement already satisfied: marisa-trie>=1.1.0 in /opt/conda/lib/python3.12/site-packages (from language-data>=1.2->langcodes<4.0.0,>=3.2.0->spacy<3.9->autogluon.tabular[all]==1.3.0->autogluon) (1.2.1)

Requirement already satisfied: llvmlite<0.45,>=0.44.0dev0 in /opt/conda/lib/python3.12/site-packages (from numba->mlforecast<0.14,>0.13->autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (0.44.0)

Requirement already satisfied: markdown-it-py>=2.2.0 in /opt/conda/lib/python3.12/site-packages (from rich->openmim<0.4.0,>=0.3.7->autogluon.multimodal==1.3.0->autogluon) (3.0.0)

Requirement already satisfied: pygments<3.0.0,>=2.13.0 in /opt/conda/lib/python3.12/site-packages (from rich->openmim<0.4.0,>=0.3.7->autogluon.multimodal==1.3.0->autogluon) (2.19.1)

Requirement already satisfied: mdurl~=0.1 in /opt/conda/lib/python3.12/site-packages (from markdown-it-py>=2.2.0->rich->openmim<0.4.0,>=0.3.7->autogluon.multimodal==1.3.0->autogluon) (0.1.2)

Requirement already satisfied: patsy>=0.5.6 in /opt/conda/lib/python3.12/site-packages (from statsmodels>=0.13.2->statsforecast<2.0.2,>=1.7.0->autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (1.0.1)

Requirement already satisfied: mpmath<1.4,>=1.1.0 in /opt/conda/lib/python3.12/site-packages (from sympy!=1.13.2,>=1.13.1->torch<2.7,>=2.2->autogluon.multimodal==1.3.0->autogluon) (1.3.0)

Requirement already satisfied: fs in /opt/conda/lib/python3.12/site-packages (from triad>=0.9.7->fugue>=0.9.0->autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (2.4.16)

Requirement already satisfied: soupsieve>1.2 in /opt/conda/lib/python3.12/site-packages (from beautifulsoup4->gdown>=4.0.0->nlpaug<1.2.0,>=1.1.10->autogluon.multimodal==1.3.0->autogluon) (2.7)

Requirement already satisfied: appdirs~=1.4.3 in /opt/conda/lib/python3.12/site-packages (from fs->triad>=0.9.7->fugue>=0.9.0->autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (1.4.4)

Requirement already satisfied: ordered-set in /opt/conda/lib/python3.12/site-packages (from model-index->openmim<0.4.0,>=0.3.7->autogluon.multimodal==1.3.0->autogluon) (4.1.0)

Requirement already satisfied: alembic>=1.5.0 in /opt/conda/lib/python3.12/site-packages (from optuna->mlforecast<0.14,>0.13->autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (1.15.2)

Requirement already satisfied: colorlog in /opt/conda/lib/python3.12/site-packages (from optuna->mlforecast<0.14,>0.13->autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (6.9.0)

Requirement already satisfied: sqlalchemy>=1.4.2 in /opt/conda/lib/python3.12/site-packages (from optuna->mlforecast<0.14,>0.13->autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (2.0.40)

Requirement already satisfied: Mako in /opt/conda/lib/python3.12/site-packages (from alembic>=1.5.0->optuna->mlforecast<0.14,>0.13->autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (1.3.10)

Requirement already satisfied: greenlet>=1 in /opt/conda/lib/python3.12/site-packages (from sqlalchemy>=1.4.2->optuna->mlforecast<0.14,>0.13->autogluon.timeseries==1.3.0->autogluon.timeseries[all]==1.3.0->autogluon) (3.2.2)

Requirement already satisfied: narwhals>=1.15.1 in /opt/conda/lib/python3.12/site-packages (from plotly->catboost<1.3,>=1.2->autogluon.tabular[all]==1.3.0->autogluon) (1.38.2)

Requirement already satisfied: PySocks!=1.5.7,>=1.5.6 in /opt/conda/lib/python3.12/site-packages (from requests[socks]->gdown>=4.0.0->nlpaug<1.2.0,>=1.1.10->autogluon.multimodal==1.3.0->autogluon) (1.7.1)

```
In [27]: !pip install kaggle
```

```
Collecting kaggle
  Downloading kaggle-1.7.4.5-py3-none-any.whl.metadata (16 kB)
Requirement already satisfied: bleach in /opt/conda/lib/python3.12/site-packages (from kaggle) (6.2.0)
Requirement already satisfied: certifi>=14.05.14 in /opt/conda/lib/python3.12/site-packages (from kaggle) (2025.4.26)
Requirement already satisfied: charset-normalizer in /opt/conda/lib/python3.12/site-packages (from kaggle) (3.4.2)
Requirement already satisfied: idna in /opt/conda/lib/python3.12/site-packages (from kaggle) (3.10)
Requirement already satisfied: protobuf in /opt/conda/lib/python3.12/site-packages (from kaggle) (5.28.3)
Requirement already satisfied: python-dateutil>=2.5.3 in /opt/conda/lib/python3.12/site-packages (from kaggle) (2.9.0.post0)
Requirement already satisfied: python-slugify in /opt/conda/lib/python3.12/site-packages (from kaggle) (8.0.4)
Requirement already satisfied: requests in /opt/conda/lib/python3.12/site-packages (from kaggle) (2.32.3)
Requirement already satisfied: setuptools>=21.0.0 in /opt/conda/lib/python3.12/site-packages (from kaggle) (80.9.0)
Requirement already satisfied: six>=1.10 in /opt/conda/lib/python3.12/site-packages (from kaggle) (1.17.0)
Requirement already satisfied: text-unidecode in /opt/conda/lib/python3.12/site-packages (from kaggle) (1.3)
Requirement already satisfied: tqdm in /opt/conda/lib/python3.12/site-packages (from kaggle) (4.67.1)
Requirement already satisfied: urllib3>=1.15.1 in /opt/conda/lib/python3.12/site-packages (from kaggle) (1.26.19)
Requirement already satisfied: webencodings in /opt/conda/lib/python3.12/site-packages (from kaggle) (0.5.1)
Downloading kaggle-1.7.4.5-py3-none-any.whl (181 kB)
Installing collected packages: kaggle
Successfully installed kaggle-1.7.4.5
```

Setup Kaggle API Key

```
In [28]: !mkdir -p /root/.kaggle
!touch /root/.kaggle/kaggle.json
!chmod 600 /root/.kaggle/kaggle.json
```

```
mkdir: cannot create directory '/root': Permission denied
touch: cannot touch '/root/.kaggle/kaggle.json': Permission denied
chmod: cannot access '/root/.kaggle/kaggle.json': Permission denied
```

```
In [39]: !mkdir -p ~/.kaggle
!mv ~/Bike-sharing-starter/project/kaggle.json ~/.kaggle/
!chmod 600 ~/.kaggle/kaggle.json
```

Download and explore dataset

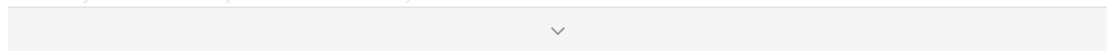
I accepted the Kaggle Competition terms and conditions

Data Fields

datetime - hourly date + timestamp

season - 1 = spring, 2 = summer, 3 = fall, 4 = winter

holiday - whether the day is considered a holiday



>_ kaggle competitions download -c bike-sharing-demand

Data Explorer

1.06 MB

sampleSubmission.csv

test.csv

train.csv

Summary

3 files

23 columns

Download All

< sampleSubmission.csv (139.51 KB)



Competition Rules



To see this data you need to agree to the [competition rules](#).
By clicking "I understand and agree" you agree to be bound to these rules.

I understand and agree

```
In [1]: !kaggle competitions download -c bike-sharing-demand
!unzip -o bike-sharing-demand.zip
```

```
zsh:1: command not found: kaggle
unzip: cannot find or open bike-sharing-demand.zip, bike-sharing-demand.zi
p.zip or bike-sharing-demand.zip.ZIP.
```

```
In [3]: import pandas as pd
from autogluon.tabular import TabularPredictor
```

```
In [4]: train = pd.read_csv("train.csv", parse_dates=["datetime"])
train.head()
```

```
Out[4]:
```

	datetime	season	holiday	workingday	weather	temp	atemp	humidity	windspeed	ca
0	2011-01-01 00:00:00	1	0	0	1	9.84	14.395	81	0.0	
1	2011-01-01 01:00:00	1	0	0	1	9.02	13.635	80	0.0	
2	2011-01-01 02:00:00	1	0	0	1	9.02	13.635	80	0.0	
3	2011-01-01 03:00:00	1	0	0	1	9.84	14.395	75	0.0	
4	2011-01-01 04:00:00	1	0	0	1	9.84	14.395	75	0.0	

Step 2: Data Exploration

Examine the structure and characteristics of the training data to understand feature distributions Analyze target variable (bike demand) patterns to identify potential modeling challenges

```
In [5]: train.describe()
```

Out[5]:

	datetime	season	holiday	workingday	weather	
count	10886	10886.000000	10886.000000	10886.000000	10886.000000	10886
mean	2011-12-27 05:56:22.399411968	2.506614	0.028569	0.680875	1.418427	2
min	2011-01-01 00:00:00	1.000000	0.000000	0.000000	1.000000	
25%	2011-07-02 07:15:00	2.000000	0.000000	0.000000	1.000000	1
50%	2012-01-01 20:30:00	3.000000	0.000000	1.000000	1.000000	2
75%	2012-07-01 12:45:00	4.000000	0.000000	1.000000	2.000000	2
max	2012-12-19 23:00:00	4.000000	1.000000	1.000000	4.000000	4
std	NaN	1.116174	0.166599	0.466159	0.633839	

```
In [7]: test = pd.read_csv("test.csv", parse_dates=["datetime"])
test.head()
```

Out[7]:

	datetime	season	holiday	workingday	weather	temp	atemp	humidity	windspeed
0	2011-01-20 00:00:00	1	0	1	1	10.66	11.365	56	26.0027
1	2011-01-20 01:00:00	1	0	1	1	10.66	13.635	56	0.0000
2	2011-01-20 02:00:00	1	0	1	1	10.66	13.635	56	0.0000
3	2011-01-20 03:00:00	1	0	1	1	10.66	12.880	56	11.0014
4	2011-01-20 04:00:00	1	0	1	1	10.66	12.880	56	11.0014

```
In [8]: submission = pd.read_csv("sampleSubmission.csv", parse_dates=["datetime"])
submission.head()
```

Out [8]:

	datetime	count
0	2011-01-20 00:00:00	0
1	2011-01-20 01:00:00	0
2	2011-01-20 02:00:00	0
3	2011-01-20 03:00:00	0
4	2011-01-20 04:00:00	0

Initial Model Training using AutoGluon's Tabular Prediction

Train baseline AutoGluon model using default settings to establish performance benchmark Use 'best_quality' preset for optimal model selection without manual tuning

```
In [ ]: train = train.drop(columns=["casual", "registered"])
```

```
In [15]: predictor = TabularPredictor(label="count", problem_type="regression", eval_
```

No path specified. Models will be saved in: "AutogluonModels/ag-20250608_091131"

Verbosity: 2 (Standard Logging)

===== System Info =====

AutoGluon Version: 1.3.0
 Python Version: 3.12.9
 Operating System: Linux
 Platform Machine: x86_64
 Platform Version: #1 SMP Tue May 6 04:10:50 UTC 2025
 CPU Count: 2
 Memory Avail: 1.85 GB / 3.76 GB (49.2%)
 Disk Space Avail: 4.25 GB / 4.99 GB (85.1%)

WARNING: Available disk space is low and there is a risk that AutoGluon will run out of disk during fit, causing an exception.

We recommend a minimum available disk space of 10 GB, and large datasets may require more.

=====

Presets specified: ['best_quality']

Setting dynamic_stacking from 'auto' to True. Reason: Enable dynamic_stacking when use_bag_holdout is disabled. (use_bag_holdout=False)

Stack configuration (auto_stack=True): num_stack_levels=1, num_bag_folds=8, num_bag_sets=1

DyStack is enabled (dynamic_stacking=True). AutoGluon will try to determine whether the input data is affected by stacked overfitting and enable or disable stacking as a consequence.

This is used to identify the optimal `num_stack_levels` value. Copies of AutoGluon will be fit on subsets of the data. Then holdout validation data is used to detect stacked overfitting.

Running DyStack for up to 150s of the 600s of remaining time (25%).

Context path: "/home/sagemaker-user/Bike-sharing-starter/project/AutogluonModels/ag-20250608_091131/ds_sub_fit/sub_fit_holdout"

WARNING: All log messages before absl::InitializeLog() is called are written to STDERR

I0000 00:00:1749374059.138462 5525 http2_transport.cc:1182] ipv4:169.255.255.2:41753: Got goaway [2] err=UNAVAILABLE:GOAWAY received; Error code: 2; Debug Text: Cancelling all calls {created_time:"2025-06-08T09:14:19.138452575+00:00", http2_error:2, grpc_status:14}

Leaderboard on holdout data (DyStack):

	model	score_holdout	score_val	eval_metric
0	LightGBMXT_BAG_L2	-71.976405	-73.741085	root_mean_squared_error
1	WeightedEnsemble_L3	-72.067095	-73.704027	root_mean_squared_error
2	KNeighborsDist_BAG_L1	-92.031272	-89.946854	root_mean_squared_error
3	WeightedEnsemble_L2	-92.031272	-89.946854	root_mean_squared_error
4	KNeighborsUnif_BAG_L1	-109.161488	-107.445008	root_mean_squared_error
5	RandomForestMSE_BAG_L1	-118.495627	-119.548529	root_mean_squared_error
6	LightGBM_BAG_L1	-130.706758	-131.849580	root_mean_squared_error
7	LightGBMXT_BAG_L1	-131.068281	-131.975832	root_mean_squared_error

```

11.665837          58.576666          1          True          3
      1          = Optimal  num_stack_levels (Stacked Overfitting Occurre
d: False)
      167s          = DyStack  runtime | 433s          = Remaining runtime
Starting main fit with num_stack_levels=1.
For future fit calls on this dataset, you can skip DyStack to save
time: `predictor.fit(..., dynamic_stacking=False, num_stack_levels=1)`
/opt/conda/lib/python3.12/site-packages/autogluon/common/utils/utils.py:97:
UserWarning: pkg_resources is deprecated as an API. See https://setuptools.
pypa.io/en/latest/pkg_resources.html. The pkg_resources package is slated f
or removal as early as 2025-11-30. Refrain from using this package or pin t
o Setuptools<81.
import pkg_resources
Beginning AutoGluon training ... Time limit = 433s
AutoGluon will save models to "/home/sagemaker-user/Bike-sharing-starter/pr
oject/AutogluonModels/ag-20250608_091131"
Train Data Rows:      10886
Train Data Columns: 9
Label Column:         count
Problem Type:         regression
Preprocessing data ...
Using Feature Generators to preprocess the data ...
Fitting AutoMLPipelineFeatureGenerator...
      Available Memory:          1520.85 MB
      Train Data (Original)  Memory Usage: 0.75 MB (0.0% of available mem
ory)
      Inferring data type of each feature based on column values. Set fea
ture_metadata_in to manually specify special dtypes of the features.
      Stage 1 Generators:
          Fitting AsTypeFeatureGenerator...
              Note: Converting 2 features to boolean dtype as the
y only contain 2 unique values.
      Stage 2 Generators:
          Fitting FillNaFeatureGenerator...
      Stage 3 Generators:
          Fitting IdentityFeatureGenerator...
          Fitting DatetimeFeatureGenerator...
      Stage 4 Generators:
          Fitting DropUniqueFeatureGenerator...
      Stage 5 Generators:
          Fitting DropDuplicatesFeatureGenerator...
      Types of features in original data (raw dtype, special dtypes):
          ('datetime', []) : 1 | ['datetime']
          ('float', [])   : 3 | ['temp', 'atemp', 'windspeed']
          ('int', [])     : 5 | ['season', 'holiday', 'workingday',
'weather', 'humidity']
      Types of features in processed data (raw dtype, special dtypes):
          ('float', [])   : 3 | ['temp', 'atemp', 'winds
peed']
          ('int', [])     : 3 | ['season', 'weather', 'h
umidity']
          ('int', ['bool']) : 2 | ['holiday', 'workingda
y']
          ('int', ['datetime_as_int']) : 5 | ['datetime', 'datetime.y
ear', 'datetime.month', 'datetime.day', 'datetime.dayofweek']
      0.1s = Fit runtime
      9 features in original data used to generate 13 features in process
ed data.
      Train Data (Processed) Memory Usage: 0.93 MB (0.1% of available mem
ory)
Data preprocessing and feature engineering runtime = 0.12s ...
AutoGluon will gauge predictive performance using evaluation metric: 'root_
mean_squared_error'
      This metric's sign has been flipped to adhere to being higher_is_be

```


tter. The metric score can be multiplied by -1 to get the metric value.

To change this, specify the `eval_metric` parameter of `Predictor()`. Large model count detected (112 configs) ... Only displaying the first 3 models of each family. To see all, set ``verbosity=3``. User-specified model hyperparameters to be fit:

```
{
  'NN_TORCH': [{}, {'activation': 'elu', 'dropout_prob': 0.1007763952
9843717, 'hidden_size': 108, 'learning_rate': 0.002735937344002146, 'num_la
yers': 4, 'use_batchnorm': True, 'weight_decay': 1.356433327634438e-12, 'ag
_args': {'name_suffix': '_r79', 'priority': -2}}, {'activation': 'elu', 'dr
opout_prob': 0.11897478034205347, 'hidden_size': 213, 'learning_rate': 0.00
10474382260641949, 'num_layers': 4, 'use_batchnorm': False, 'weight_decay':
5.594471067786272e-10, 'ag_args': {'name_suffix': '_r22', 'priority': -
7}}],
  'GBM': [{}, {'extra_trees': True, 'ag_args': {'name_suffix': 'XT'}},
  {}, {'learning_rate': 0.03, 'num_leaves': 128, 'feature_fraction': 0.9, 'mi
n_data_in_leaf': 3, 'ag_args': {'name_suffix': 'Large', 'priority': 0, 'hyp
erparameter_tune_kwargs': None}}],
  'CAT': [{}, {'depth': 6, 'grow_policy': 'SymmetricTree', 'l2_leaf_r
eg': 2.1542798306067823, 'learning_rate': 0.06864209415792857, 'max_ctr_com
plexity': 4, 'one_hot_max_size': 10, 'ag_args': {'name_suffix': '_r177', 'p
riority': -1}}, {'depth': 8, 'grow_policy': 'Depthwise', 'l2_leaf_reg': 2.7
997999596449104, 'learning_rate': 0.031375015734637225, 'max_ctr_complexit
y': 2, 'one_hot_max_size': 3, 'ag_args': {'name_suffix': '_r9', 'priority':
-5}}],
  'XGB': [{}, {'colsample_bytree': 0.6917311125174739, 'enable_catego
rical': False, 'learning_rate': 0.018063876087523967, 'max_depth': 10, 'min
_child_weight': 0.6028633586934382, 'ag_args': {'name_suffix': '_r33', 'pri
ority': -8}}, {'colsample_bytree': 0.6628423832084077, 'enable_categorica
l': False, 'learning_rate': 0.08775715546881824, 'max_depth': 5, 'min_child
_weight': 0.6294123374222513, 'ag_args': {'name_suffix': '_r89', 'priorit
y': -16}}],
  'FASTAI': [{}, {'bs': 256, 'emb_drop': 0.5411770367537934, 'epoch
s': 43, 'layers': [800, 400], 'lr': 0.01519848858318159, 'ps': 0.2378294656
6604385, 'ag_args': {'name_suffix': '_r191', 'priority': -4}}, {'bs': 2048,
'emb_drop': 0.05070411322605811, 'epochs': 29, 'layers': [200, 100], 'lr':
0.08974235041576624, 'ps': 0.10393466140748028, 'ag_args': {'name_suffix':
'_r102', 'priority': -11}}],
  'RF': [{}, {'criterion': 'gini', 'ag_args': {'name_suffix': 'Gini', 'pr
oblem_types': ['binary', 'multiclass']}}, {'criterion': 'entropy', 'ag_arg
s': {'name_suffix': 'Entr', 'problem_types': ['binary', 'multiclass']}},
  {'criterion': 'squared_error', 'ag_args': {'name_suffix': 'MSE', 'problem_t
ypes': ['regression', 'quantile']}}],
  'XT': [{}, {'criterion': 'gini', 'ag_args': {'name_suffix': 'Gini', 'pr
oblem_types': ['binary', 'multiclass']}}, {'criterion': 'entropy', 'ag_arg
s': {'name_suffix': 'Entr', 'problem_types': ['binary', 'multiclass']}},
  {'criterion': 'squared_error', 'ag_args': {'name_suffix': 'MSE', 'problem_t
ypes': ['regression', 'quantile']}}],
  'KNN': [{}, {'weights': 'uniform', 'ag_args': {'name_suffix': 'Unif'}},
  {'weights': 'distance', 'ag_args': {'name_suffix': 'Dist'}}],
}
```

AutoGluon will fit 2 stack levels (L1 to L2) ...

Fitting 108 L1 models, `fit_strategy="sequential"` ...

Fitting model: `KNeighborsUnif_BAG_L1` ... Training model for up to 288.24s of the 432.46s of remaining time.

-101.5462 = Validation score (-root_mean_squared_error)

0.04s = Training runtime

0.05s = Validation runtime

Fitting model: `KNeighborsDist_BAG_L1` ... Training model for up to 286.06s of the 430.28s of remaining time.

-84.1251 = Validation score (-root_mean_squared_error)

0.04s = Training runtime

0.06s = Validation runtime

Fitting model: `LightGBMXT_BAG_L1` ... Training model for up to 285.93s of th

```

e 430.14s of remaining time.
    Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFo
ldFittingStrategy (2 workers, per: cpus=1, gpus=0, memory=0.63%)
    -131.4609      = Validation score (-root_mean_squared_error)
    56.42s      = Training runtime
    11.67s      = Validation runtime
Fitting model: LightGBM_BAG_L1 ... Training model for up to 223.57s of the
367.79s of remaining time.
    Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFo
ldFittingStrategy (2 workers, per: cpus=1, gpus=0, memory=0.85%)
    -131.0542      = Validation score (-root_mean_squared_error)
    27.81s      = Training runtime
    1.48s      = Validation runtime
Fitting model: RandomForestMSE_BAG_L1 ... Training model for up to 191.42s
of the 335.63s of remaining time.
/opt/conda/lib/python3.12/site-packages/sklearn/base.py:474: FutureWarning:
`BaseEstimator._validate_data` is deprecated in 1.6 and will be removed in
1.7. Use `sklearn.utils.validation.validate_data` instead. This function be
comes public and is part of the scikit-learn developer API.
    warnings.warn(
    -116.5484      = Validation score (-root_mean_squared_error)
    15.28s      = Training runtime
    0.69s      = Validation runtime
Fitting model: CatBoost_BAG_L1 ... Training model for up to 174.94s of the
319.15s of remaining time.
    Memory not enough to fit 8 folds in parallel. Will train 2 folds in
parallel instead (Estimated 24.75% memory usage per fold, 49.49%/80.00% tot
al).
    Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFo
ldFittingStrategy (2 workers, per: cpus=1, gpus=0, memory=24.75%)
I0000 00:00:1749374276.376276 5434 http2_transport.cc:1182] ipv4:169.25
5.255.2:34661: Got goaway [2] err=UNAVAILABLE:GOAWAY received; Error code:
2; Debug Text: Cancelling all calls {grpc_status:14, http2_error:2, created
_time:"2025-06-08T09:17:56.376272043+00:00"}
    -130.7062      = Validation score (-root_mean_squared_error)
    143.88s      = Training runtime
    0.12s      = Validation runtime
Fitting model: ExtraTreesMSE_BAG_L1 ... Training model for up to 27.65s of
the 171.87s of remaining time.
/opt/conda/lib/python3.12/site-packages/sklearn/base.py:474: FutureWarning:
`BaseEstimator._validate_data` is deprecated in 1.6 and will be removed in
1.7. Use `sklearn.utils.validation.validate_data` instead. This function be
comes public and is part of the scikit-learn developer API.
    warnings.warn(
    -124.6007      = Validation score (-root_mean_squared_error)
    7.68s      = Training runtime
    0.67s      = Validation runtime
Fitting model: NeuralNetFastAI_BAG_L1 ... Training model for up to 18.84s o
f the 163.05s of remaining time.
    Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFo
ldFittingStrategy (2 workers, per: cpus=1, gpus=0, memory=0.68%)
    -142.3309      = Validation score (-root_mean_squared_error)
    42.94s      = Training runtime
    0.3s      = Validation runtime
Fitting model: WeightedEnsemble_L2 ... Training model for up to 360.00s of
the 116.31s of remaining time.
    Ensemble Weights: {'KNeighborsDist_BAG_L1': 1.0}
    -84.1251      = Validation score (-root_mean_squared_error)
    0.03s      = Training runtime
    0.0s      = Validation runtime
Fitting 106 L2 models, fit_strategy="sequential" ...
Fitting model: LightGBMXT_BAG_L2 ... Training model for up to 116.27s of th
e 116.25s of remaining time.
    Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFo

```

```

ldFittingStrategy (2 workers, per: cpus=1, gpus=0, memory=1.55%)
I0000 00:00:1749374389.066695 5528 http2_transport.cc:1182] ipv4:169.25
5.255.2:39733: Got goaway [2] err=UNAVAILABLE:GOAWAY received; Error code:
2; Debug Text: Cancelling all calls {grpc_status:14, http2_error:2, created
_time:"2025-06-08T09:19:49.066690392+00:00"}
I0000 00:00:1749374418.394930 5278 http2_transport.cc:1182] ipv4:169.25
5.255.2:44947: Got goaway [2] err=UNAVAILABLE:GOAWAY received; Error code:
2; Debug Text: Cancelling all calls {created_time:"2025-06-08T09:20:18.3949
25195+00:00", http2_error:2, grpc_status:14}
-60.4491 = Validation score (-root_mean_squared_error)
50.65s = Training runtime
4.17s = Validation runtime
Fitting model: LightGBM_BAG_L2 ... Training model for up to 59.75s of the 5
9.72s of remaining time.
Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFo
ldFittingStrategy (2 workers, per: cpus=1, gpus=0, memory=1.08%)
I0000 00:00:1749374462.110637 5531 http2_transport.cc:1182] ipv4:169.25
5.255.2:37431: Got goaway [2] err=UNAVAILABLE:GOAWAY received; Error code:
2; Debug Text: Cancelling all calls {grpc_status:14, http2_error:2, created
_time:"2025-06-08T09:21:02.11063315+00:00"}
-55.0682 = Validation score (-root_mean_squared_error)
25.63s = Training runtime
0.39s = Validation runtime
Fitting model: RandomForestMSE_BAG_L2 ... Training model for up to 29.79s o
f the 29.77s of remaining time.
/opt/conda/lib/python3.12/site-packages/sklearn/base.py:474: FutureWarning:
`BaseEstimator._validate_data` is deprecated in 1.6 and will be removed in
1.7. Use `sklearn.utils.validation.validate_data` instead. This function be
comes public and is part of the scikit-learn developer API.
warnings.warn(
-53.3512 = Validation score (-root_mean_squared_error)
41.93s = Training runtime
0.69s = Validation runtime
Fitting model: WeightedEnsemble_L3 ... Training model for up to 360.00s of
the -13.47s of remaining time.
Ensemble Weights: {'RandomForestMSE_BAG_L2': 0.727, 'LightGBM_BAG_L
2': 0.227, 'LightGBMXT_BAG_L2': 0.045}
-53.0556 = Validation score (-root_mean_squared_error)
0.03s = Training runtime
0.0s = Validation runtime
AutoGluon training complete, total runtime = 446.12s ... Best model: Weight
edEnsemble_L3 | Estimated inference throughput: 73.9 rows/s (1361 batch siz
e)
TabularPredictor saved. To load, use: predictor = TabularPredictor.load("/h
ome/sagemaker-user/Bike-sharing-starter/project/AutogluonModels/ag-20250608
_091131")

```

Review AutoGluon's training run with ranking of models that did the best.

```
In [21]: predictor.fit_summary()
```

*** Summary of fit() ***

Estimated performance of each model:

val	fit_time	pred_time_val	model	score_val	eval_metric	pred_time_val	stack_level	ca
n_infer	fit_order			marginal				
0	WeightedEnsemble_L3	-53.055564	root_mean_squared_error	20.292				
849	412.323494	0.000590	0.033826	3				
True	13							
1	RandomForestMSE_BAG_L2	-53.351220	root_mean_squared_error	15.731				
574	336.007394	0.689999	41.927320	2				
True	12							
2	LightGBM_BAG_L2	-55.068241	root_mean_squared_error	15.428				
610	319.710187	0.387035	25.630113	2				
True	11							
3	LightGBMXT_BAG_L2	-60.449131	root_mean_squared_error	19.215				
224	344.732234	4.173650	50.652160	2				
True	10							
4	KNeighborsDist_BAG_L1	-84.125061	root_mean_squared_error	0.061				
686	0.036445	0.061686	0.036445	1				
True	2							
5	WeightedEnsemble_L2	-84.125061	root_mean_squared_error	0.062				
364	0.064287	0.000678	0.027843	2				
True	9							
6	KNeighborsUnif_BAG_L1	-101.546199	root_mean_squared_error	0.045				
070	0.036497	0.045070	0.036497	1				
True	1							
7	RandomForestMSE_BAG_L1	-116.548359	root_mean_squared_error	0.690				
165	15.282232	0.690165	15.282232	1				
True	5							
8	ExtraTreesMSE_BAG_L1	-124.600676	root_mean_squared_error	0.668				
831	7.680941	0.668831	7.680941	1				
True	7							
9	CatBoost_BAG_L1	-130.706246	root_mean_squared_error	0.115				
338	143.881330	0.115338	143.881330	1				
True	6							
10	LightGBM_BAG_L1	-131.054162	root_mean_squared_error	1.484				
636	27.805858	1.484636	27.805858	1				
True	4							
11	LightGBMXT_BAG_L1	-131.460909	root_mean_squared_error	11.674				
937	56.416111	11.674937	56.416111	1				
True	3							
12	NeuralNetFastAI_BAG_L1	-142.330902	root_mean_squared_error	0.300				
911	42.940660	0.300911	42.940660	1				
True	8							

Number of models trained: 13

Types of models trained:

```
{'StackerEnsembleModel_KNN', 'StackerEnsembleModel_XT', 'StackerEnsembleModel_RF', 'StackerEnsembleModel_NNFastAiTabular', 'StackerEnsembleModel_CatBoost', 'StackerEnsembleModel_LGB', 'WeightedEnsembleModel'}
```

Bagging used: True (with 8 folds)

Multi-layer stack-ensembling used: True (with 3 levels)

Feature Metadata (Processed):

(raw dtype, special dtypes):

('float', []) : 3 | ['temp', 'atemp', 'windspeed']

('int', []) : 3 | ['season', 'weather', 'humidity']

('int', ['bool']) : 2 | ['holiday', 'workingday']

('int', ['datetime_as_int']) : 5 | ['datetime', 'datetime.year', 'datetime.month', 'datetime.day', 'datetime.dayofweek']

Plot summary of models saved to file: /home/sagemaker-user/Bike-sharing-starter/project/AutogluonModels/ag-20250608_091131/SummaryOfModels.html

*** End of fit() summary ***

```

Out[21]: {'model_types': {'KNeighborsUnif_BAG_L1': 'StackerEnsembleModel_KNN',
  'KNeighborsDist_BAG_L1': 'StackerEnsembleModel_KNN',
  'LightGBMXT_BAG_L1': 'StackerEnsembleModel_LGB',
  'LightGBM_BAG_L1': 'StackerEnsembleModel_LGB',
  'RandomForestMSE_BAG_L1': 'StackerEnsembleModel_RF',
  'CatBoost_BAG_L1': 'StackerEnsembleModel_CatBoost',
  'ExtraTreesMSE_BAG_L1': 'StackerEnsembleModel_XT',
  'NeuralNetFastAI_BAG_L1': 'StackerEnsembleModel_NNFastAiTabular',
  'WeightedEnsemble_L2': 'WeightedEnsembleModel',
  'LightGBMXT_BAG_L2': 'StackerEnsembleModel_LGB',
  'LightGBM_BAG_L2': 'StackerEnsembleModel_LGB',
  'RandomForestMSE_BAG_L2': 'StackerEnsembleModel_RF',
  'WeightedEnsemble_L3': 'WeightedEnsembleModel'},
  'model_performance': {'KNeighborsUnif_BAG_L1': -101.54619908446061,
  'KNeighborsDist_BAG_L1': -84.12506123181602,
  'LightGBMXT_BAG_L1': -131.46090891834504,
  'LightGBM_BAG_L1': -131.054161598899,
  'RandomForestMSE_BAG_L1': -116.54835939455667,
  'CatBoost_BAG_L1': -130.70624616892295,
  'ExtraTreesMSE_BAG_L1': -124.60067564699747,
  'NeuralNetFastAI_BAG_L1': -142.33090200146412,
  'WeightedEnsemble_L2': -84.12506123181602,
  'LightGBMXT_BAG_L2': -60.44913063799687,
  'LightGBM_BAG_L2': -55.06824052420681,
  'RandomForestMSE_BAG_L2': -53.351219537387344,
  'WeightedEnsemble_L3': -53.055563856004945},
  'model_best': 'WeightedEnsemble_L3',
  'model_paths': {'KNeighborsUnif_BAG_L1': ['KNeighborsUnif_BAG_L1'],
  'KNeighborsDist_BAG_L1': ['KNeighborsDist_BAG_L1'],
  'LightGBMXT_BAG_L1': ['LightGBMXT_BAG_L1'],
  'LightGBM_BAG_L1': ['LightGBM_BAG_L1'],
  'RandomForestMSE_BAG_L1': ['RandomForestMSE_BAG_L1'],
  'CatBoost_BAG_L1': ['CatBoost_BAG_L1'],
  'ExtraTreesMSE_BAG_L1': ['ExtraTreesMSE_BAG_L1'],
  'NeuralNetFastAI_BAG_L1': ['NeuralNetFastAI_BAG_L1'],
  'WeightedEnsemble_L2': ['WeightedEnsemble_L2'],
  'LightGBMXT_BAG_L2': ['LightGBMXT_BAG_L2'],
  'LightGBM_BAG_L2': ['LightGBM_BAG_L2'],
  'RandomForestMSE_BAG_L2': ['RandomForestMSE_BAG_L2'],
  'WeightedEnsemble_L3': ['WeightedEnsemble_L3']},
  'model_fit_times': {'KNeighborsUnif_BAG_L1': 0.03649711608886719,
  'KNeighborsDist_BAG_L1': 0.036444902420043945,
  'LightGBMXT_BAG_L1': 56.41611099243164,
  'LightGBM_BAG_L1': 27.805858373641968,
  'RandomForestMSE_BAG_L1': 15.28223180770874,
  'CatBoost_BAG_L1': 143.88133001327515,
  'ExtraTreesMSE_BAG_L1': 7.680941104888916,
  'NeuralNetFastAI_BAG_L1': 42.940659523010254,
  'WeightedEnsemble_L2': 0.02784252166748047,
  'LightGBMXT_BAG_L2': 50.65216040611267,
  'LightGBM_BAG_L2': 25.63011336326599,
  'RandomForestMSE_BAG_L2': 41.92732048034668,
  'WeightedEnsemble_L3': 0.03382587432861328},
  'model_pred_times': {'KNeighborsUnif_BAG_L1': 0.04507040977478027,
  'KNeighborsDist_BAG_L1': 0.06168627738952637,
  'LightGBMXT_BAG_L1': 11.67493724822998,
  'LightGBM_BAG_L1': 1.484635591506958,
  'RandomForestMSE_BAG_L1': 0.6901652812957764,
  'CatBoost_BAG_L1': 0.11533784866333008,
  'ExtraTreesMSE_BAG_L1': 0.6688311100006104,
  'NeuralNetFastAI_BAG_L1': 0.30091071128845215,
  'WeightedEnsemble_L2': 0.0006775856018066406,
  'LightGBMXT_BAG_L2': 4.173649787902832,
  'LightGBM_BAG_L2': 0.3870353698730469,

```

```
'RandomForestMSE_BAG_L2': 0.6899991035461426,  
'WeightedEnsemble_L3': 0.0005900859832763672},  
'num_bag_folds': 8,  
'max_stack_level': 3,  
'model_hyperparams': {'KNeighborsUnif_BAG_L1': {'use_orig_features': True,  
  'valid_stack': True,  
  'max_base_models': 0,  
  'max_base_models_per_type': 'auto',  
  'save_bag_folds': True,  
  'stratify': 'auto',  
  'bin': 'auto',  
  'n_bins': None,  
  'use_child_oof': True},  
'KNeighborsDist_BAG_L1': {'use_orig_features': True,  
  'valid_stack': True,  
  'max_base_models': 0,  
  'max_base_models_per_type': 'auto',  
  'save_bag_folds': True,  
  'stratify': 'auto',  
  'bin': 'auto',  
  'n_bins': None,  
  'use_child_oof': True},  
'LightGBMXT_BAG_L1': {'use_orig_features': True,  
  'valid_stack': True,  
  'max_base_models': 0,  
  'max_base_models_per_type': 'auto',  
  'save_bag_folds': True,  
  'stratify': 'auto',  
  'bin': 'auto',  
  'n_bins': None},  
'LightGBM_BAG_L1': {'use_orig_features': True,  
  'valid_stack': True,  
  'max_base_models': 0,  
  'max_base_models_per_type': 'auto',  
  'save_bag_folds': True,  
  'stratify': 'auto',  
  'bin': 'auto',  
  'n_bins': None},  
'RandomForestMSE_BAG_L1': {'use_orig_features': True,  
  'valid_stack': True,  
  'max_base_models': 0,  
  'max_base_models_per_type': 'auto',  
  'save_bag_folds': True,  
  'stratify': 'auto',  
  'bin': 'auto',  
  'n_bins': None,  
  'use_child_oof': True},  
'CatBoost_BAG_L1': {'use_orig_features': True,  
  'valid_stack': True,  
  'max_base_models': 0,  
  'max_base_models_per_type': 'auto',  
  'save_bag_folds': True,  
  'stratify': 'auto',  
  'bin': 'auto',  
  'n_bins': None},  
'ExtraTreesMSE_BAG_L1': {'use_orig_features': True,  
  'valid_stack': True,  
  'max_base_models': 0,  
  'max_base_models_per_type': 'auto',  
  'save_bag_folds': True,  
  'stratify': 'auto',  
  'bin': 'auto',  
  'n_bins': None,  
  'use_child_oof': True},
```

```

'NeuralNetFastAI_BAG_L1': {'use_orig_features': True,
  'valid_stack': True,
  'max_base_models': 0,
  'max_base_models_per_type': 'auto',
  'save_bag_folds': True,
  'stratify': 'auto',
  'bin': 'auto',
  'n_bins': None},
'WeightedEnsemble_L2': {'use_orig_features': False,
  'valid_stack': True,
  'max_base_models': 0,
  'max_base_models_per_type': 'auto',
  'save_bag_folds': True,
  'stratify': 'auto',
  'bin': 'auto',
  'n_bins': None},
'LightGBMXT_BAG_L2': {'use_orig_features': True,
  'valid_stack': True,
  'max_base_models': 0,
  'max_base_models_per_type': 'auto',
  'save_bag_folds': True,
  'stratify': 'auto',
  'bin': 'auto',
  'n_bins': None},
'LightGBM_BAG_L2': {'use_orig_features': True,
  'valid_stack': True,
  'max_base_models': 0,
  'max_base_models_per_type': 'auto',
  'save_bag_folds': True,
  'stratify': 'auto',
  'bin': 'auto',
  'n_bins': None},
'RandomForestMSE_BAG_L2': {'use_orig_features': True,
  'valid_stack': True,
  'max_base_models': 0,
  'max_base_models_per_type': 'auto',
  'save_bag_folds': True,
  'stratify': 'auto',
  'bin': 'auto',
  'n_bins': None,
  'use_child_oof': True},
'WeightedEnsemble_L3': {'use_orig_features': False,
  'valid_stack': True,
  'max_base_models': 0,
  'max_base_models_per_type': 'auto',
  'save_bag_folds': True,
  'stratify': 'auto',
  'bin': 'auto',
  'n_bins': None}},
'leaderboard':

```

		model	score_val	eval_metric \
0	WeightedEnsemble_L3	-53.055564	root_mean_squared_error	
1	RandomForestMSE_BAG_L2	-53.351220	root_mean_squared_error	
2	LightGBM_BAG_L2	-55.068241	root_mean_squared_error	
3	LightGBMXT_BAG_L2	-60.449131	root_mean_squared_error	
4	KNeighborsDist_BAG_L1	-84.125061	root_mean_squared_error	
5	WeightedEnsemble_L2	-84.125061	root_mean_squared_error	
6	KNeighborsUnif_BAG_L1	-101.546199	root_mean_squared_error	
7	RandomForestMSE_BAG_L1	-116.548359	root_mean_squared_error	
8	ExtraTreesMSE_BAG_L1	-124.600676	root_mean_squared_error	
9	CatBoost_BAG_L1	-130.706246	root_mean_squared_error	
10	LightGBM_BAG_L1	-131.054162	root_mean_squared_error	
11	LightGBMXT_BAG_L1	-131.460909	root_mean_squared_error	
12	NeuralNetFastAI_BAG_L1	-142.330902	root_mean_squared_error	

	pred_time_val	fit_time	pred_time_val_marginal	fit_time_marginal
0	20.292849	412.323494	0.000590	0.033826
1	15.731574	336.007394	0.689999	41.927320
2	15.428610	319.710187	0.387035	25.630113
3	19.215224	344.732234	4.173650	50.652160
4	0.061686	0.036445	0.061686	0.036445
5	0.062364	0.064287	0.000678	0.027843
6	0.045070	0.036497	0.045070	0.036497
7	0.690165	15.282232	0.690165	15.282232
8	0.668831	7.680941	0.668831	7.680941
9	0.115338	143.881330	0.115338	143.881330
10	1.484636	27.805858	1.484636	27.805858
11	11.674937	56.416111	11.674937	56.416111
12	0.300911	42.940660	0.300911	42.940660

	stack_level	can_infer	fit_order
0	3	True	13
1	2	True	12
2	2	True	11
3	2	True	10
4	1	True	2
5	2	True	9
6	1	True	1
7	1	True	5
8	1	True	7
9	1	True	6
10	1	True	4
11	1	True	3
12	1	True	8 }

In [17]: `test.head()`

Out[17]:

	datetime	season	holiday	workingday	weather	temp	atemp	humidity	windspeed
0	2011-01-20 00:00:00	1	0	1	1	10.66	11.365	56	26.0027
1	2011-01-20 01:00:00	1	0	1	1	10.66	13.635	56	0.0000
2	2011-01-20 02:00:00	1	0	1	1	10.66	13.635	56	0.0000
3	2011-01-20 03:00:00	1	0	1	1	10.66	12.880	56	11.0014
4	2011-01-20 04:00:00	1	0	1	1	10.66	12.880	56	11.0014

Create predictions from test dataset

In [19]: `predictions = predictor.predict(test)`
`predictions.head()`


```
Out[19]: 0    22.042393
         1    42.045460
         2    45.490578
         3    47.933044
         4    50.954140
         Name: count, dtype: float32
```

```
In [22]: predictions.describe()
```

```
Out[22]: count    6493.000000
         mean     100.602112
         std       89.785240
         min        2.766771
         25%       20.299770
         50%       64.221878
         75%      166.811600
         max      366.587402
         Name: count, dtype: float64
```

```
In [107... predictor.leaderboard(silent=True)
```

```
Out[107]:
```

	model	score_val	eval_metric	pred_time_val	fit_
0	WeightedEnsemble_L3	-53.055564	root_mean_squared_error	20.292849	412.32
1	RandomForestMSE_BAG_L2	-53.351220	root_mean_squared_error	15.731574	336.00
2	LightGBM_BAG_L2	-55.068241	root_mean_squared_error	15.428610	319.71
3	LightGBMXT_BAG_L2	-60.449131	root_mean_squared_error	19.215224	344.73
4	KNeighborsDist_BAG_L1	-84.125061	root_mean_squared_error	0.061686	0.03
5	WeightedEnsemble_L2	-84.125061	root_mean_squared_error	0.062364	0.06
6	KNeighborsUnif_BAG_L1	-101.546199	root_mean_squared_error	0.045070	0.03
7	RandomForestMSE_BAG_L1	-116.548359	root_mean_squared_error	0.690165	15.28
8	ExtraTreesMSE_BAG_L1	-124.600676	root_mean_squared_error	0.668831	7.68
9	CatBoost_BAG_L1	-130.706246	root_mean_squared_error	0.115338	143.88
10	LightGBM_BAG_L1	-131.054162	root_mean_squared_error	1.484636	27.80
11	LightGBMXT_BAG_L1	-131.460909	root_mean_squared_error	11.674937	56.4
12	NeuralNetFastAI_BAG_L1	-142.330902	root_mean_squared_error	0.300911	42.94

```
In [110... print(train["count"].describe())
```

```
count    10886.000000
mean      191.574132
std       181.144454
min         1.000000
25%        42.000000
50%       145.000000
75%       284.000000
max       977.000000
         Name: count, dtype: float64
```

```
In [23]: num_negatives = (predictions < 0).sum()
         print("Number of negative predictions:", num_negatives)
```

```
Number of negative predictions: 0
```

```
In [24]: predictions[predictions < 0] = 0
```

Create submission file in required Kaggle format for initial performance evaluation and submit

```
In [25]: submission["count"] = predictions
submission.to_csv("submission.csv", index=False)
```

```
In [40]: !kaggle competitions submit -c bike-sharing-demand -f submission.csv -m "first submission"
100%|████████████████████████████████████████| 188k/188k [00:00<00:00, 643
kB/s]
Successfully submitted to Bike Sharing Demand
```

View submission via the command line or in the web browser under the competition's page - [My Submissions](#)

```
In [41]: !kaggle competitions submissions -c bike-sharing-demand | tail -n +1 | head
```

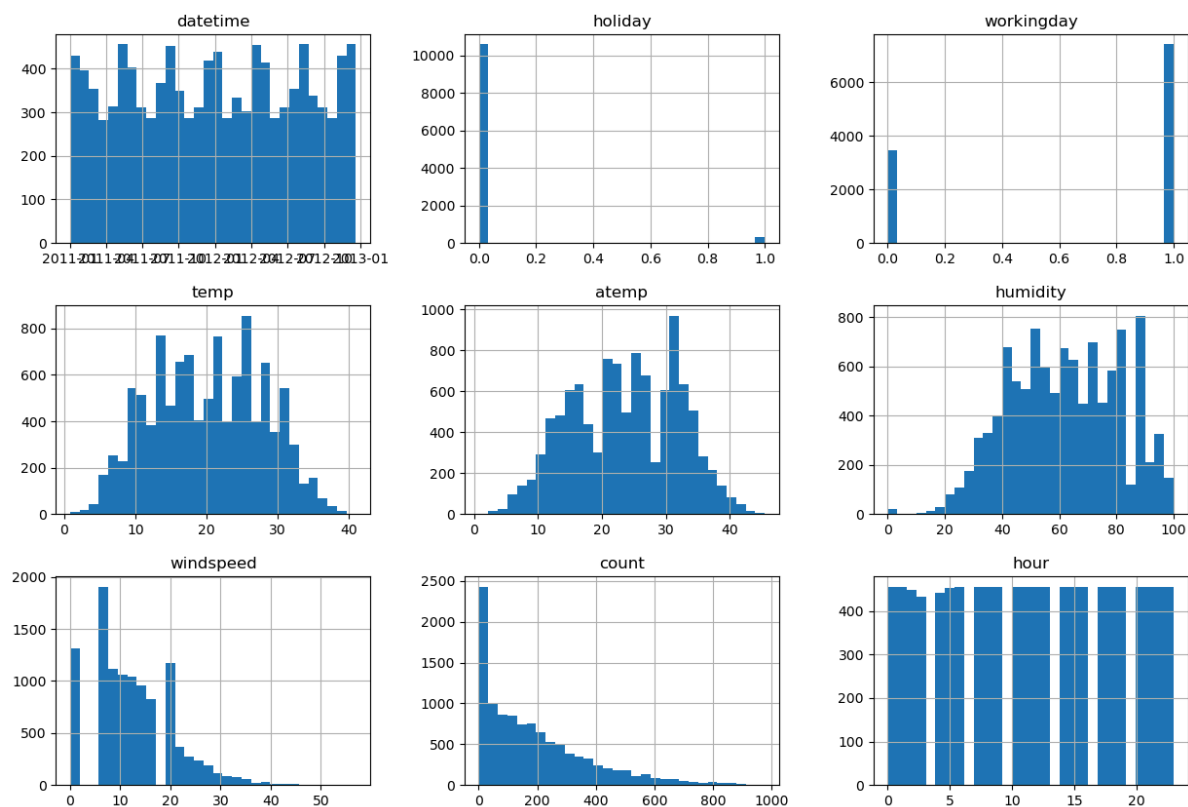
fileName	date	description	status
publicScore	privateScore		
submission.csv	2025-06-08 10:04:09.937000	first row submission	SubmissionStatus.COMPLETED
1.80019	1.80019		

Initial score of **1.80019**

Step 4: Exploratory Data Analysis and Creating an additional feature

```
In [84]: train.hist(figsize=(15, 10), bins=30)
```

```
Out[84]: array([[<Axes: title={'center': 'datetime'}>,
<Axes: title={'center': 'holiday'}>,
<Axes: title={'center': 'workingday'}>],
[<Axes: title={'center': 'temp'}>,
<Axes: title={'center': 'atemp'}>,
<Axes: title={'center': 'humidity'}>],
[<Axes: title={'center': 'windspeed'}>,
<Axes: title={'center': 'count'}>,
<Axes: title={'center': 'hour'}>]], dtype=object)
```



```
In [85]: train_no_hour = train.drop(columns=["hour"])
```

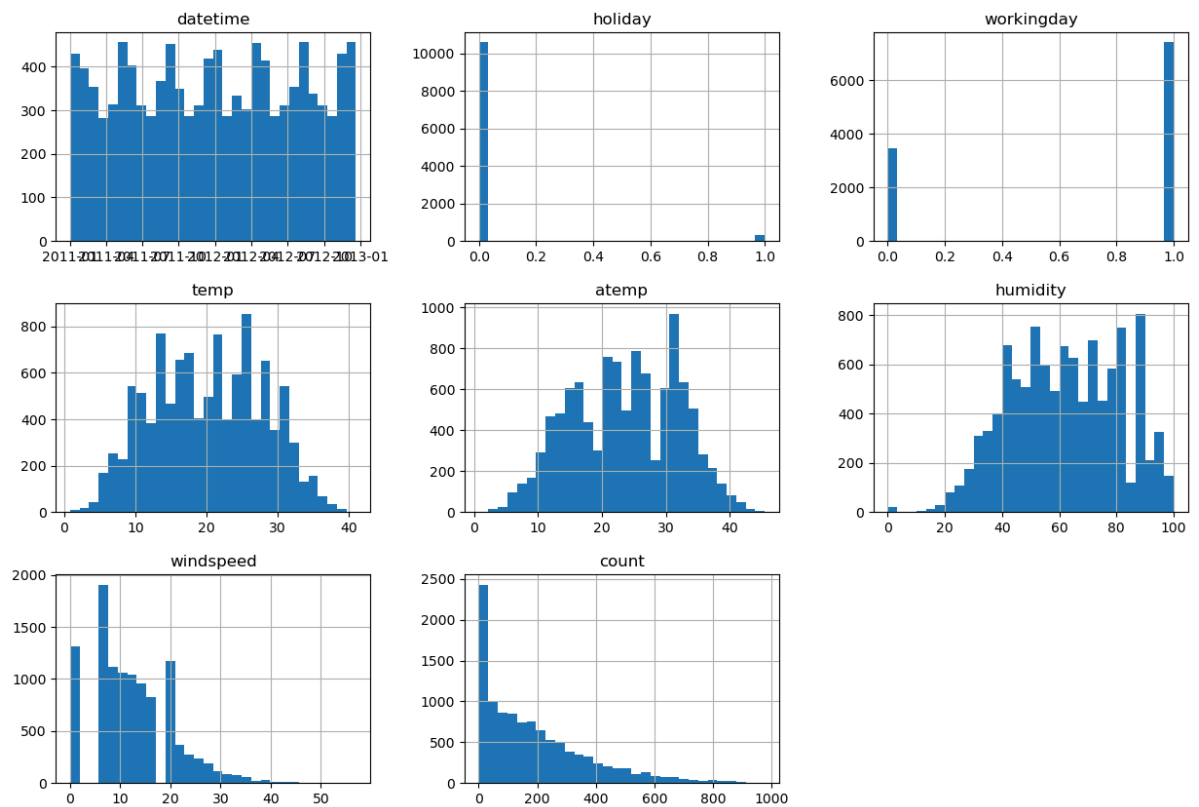
```
In [86]: train_no_hour.head()
```

```
Out[86]:
```

	datetime	season	holiday	workingday	weather	temp	atemp	humidity	windspeed	co
0	2011-01-01 00:00:00	1	0	0	1	9.84	14.395	81	0.0	
1	2011-01-01 01:00:00	1	0	0	1	9.02	13.635	80	0.0	
2	2011-01-01 02:00:00	1	0	0	1	9.02	13.635	80	0.0	
3	2011-01-01 03:00:00	1	0	0	1	9.84	14.395	75	0.0	
4	2011-01-01 04:00:00	1	0	0	1	9.84	14.395	75	0.0	

```
In [87]: train_no_hour.hist(figsize=(15, 10), bins=30)
```

```
Out[87]: array([[<Axes: title={'center': 'datetime'}>,
      <Axes: title={'center': 'holiday'}>,
      <Axes: title={'center': 'workingday'}>],
      [<Axes: title={'center': 'temp'}>,
      <Axes: title={'center': 'atemp'}>,
      <Axes: title={'center': 'humidity'}>],
      [<Axes: title={'center': 'windspeed'}>,
      <Axes: title={'center': 'count'}>, <Axes: >]], dtype=object)
```



In [66]: `train.head()`

Out[66]:

	datetime	season	holiday	workingday	weather	temp	atemp	humidity	windspeed	co
0	2011-01-01 00:00:00	1	0	0	1	9.84	14.395	81	0.0	
1	2011-01-01 01:00:00	1	0	0	1	9.02	13.635	80	0.0	
2	2011-01-01 02:00:00	1	0	0	1	9.02	13.635	80	0.0	
3	2011-01-01 03:00:00	1	0	0	1	9.84	14.395	75	0.0	
4	2011-01-01 04:00:00	1	0	0	1	9.84	14.395	75	0.0	

Feature Engineering

- Extract hour feature from datetime column to capture time-based demand patterns
- Add temporal features that represent rush hour and peak usage periods

In [45]: `train["hour"] = train["datetime"].dt.hour`
`test["hour"] = test["datetime"].dt.hour`

Make category types for these so models know they are not just numbers

- AutoGluon originally sees these as ints, but in reality they are int representations of a category.
- Setting the dtype to category will classify these as categories in AutoGluon.

```
In [46]: train["season"] = train["season"].astype("category")
train["weather"] = train["weather"].astype("category")
test["season"] = test["season"].astype("category")
test["weather"] = test["weather"].astype("category")
```

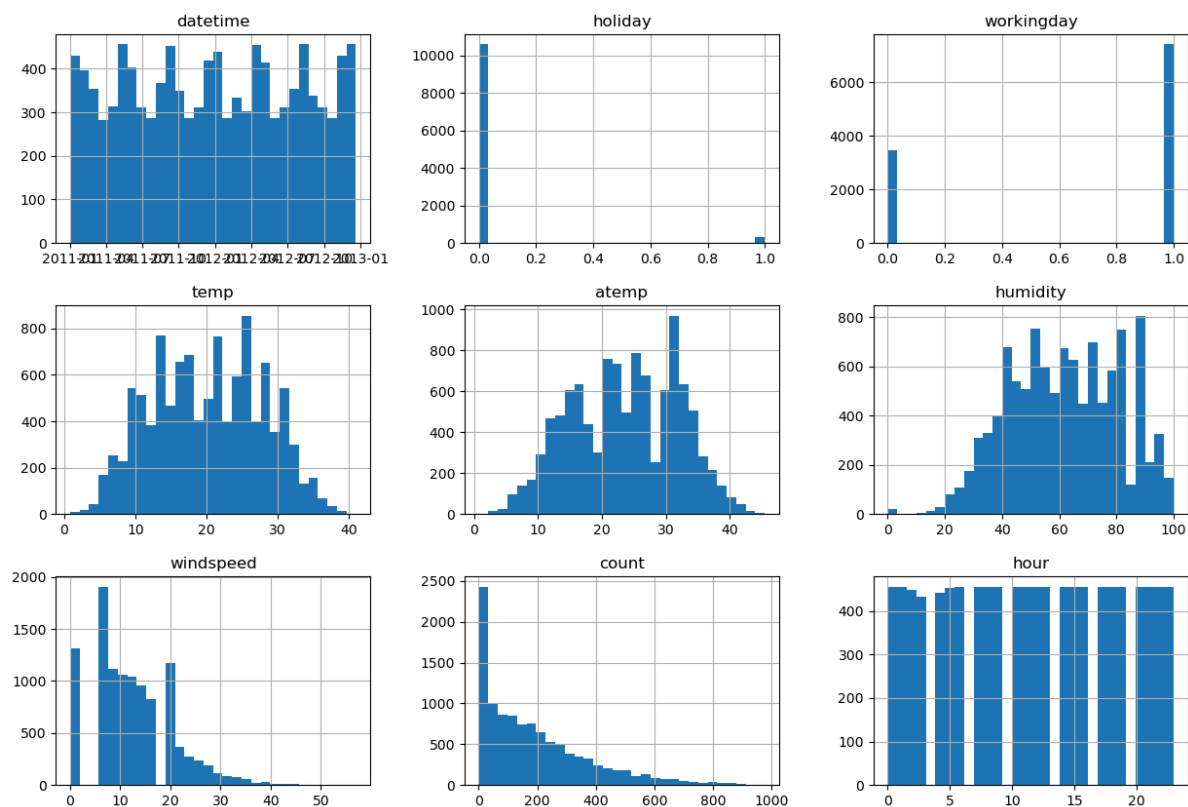
```
In [47]: # View our new feature
train.head()
```

```
Out[47]:
```

	datetime	season	holiday	workingday	weather	temp	atemp	humidity	windspeed	co
0	2011-01-01 00:00:00	1	0	0	1	9.84	14.395	81	0.0	
1	2011-01-01 01:00:00	1	0	0	1	9.02	13.635	80	0.0	
2	2011-01-01 02:00:00	1	0	0	1	9.02	13.635	80	0.0	
3	2011-01-01 03:00:00	1	0	0	1	9.84	14.395	75	0.0	
4	2011-01-01 04:00:00	1	0	0	1	9.84	14.395	75	0.0	

```
In [48]: # View histogram of all features again now with the hour feature
train.hist(figsize=(15, 10), bins=30)
```

```
Out[48]: array([[<Axes: title={'center': 'datetime'}>,
      <Axes: title={'center': 'holiday'}>,
      <Axes: title={'center': 'workingday'}>],
      [<Axes: title={'center': 'temp'}>,
      <Axes: title={'center': 'atemp'}>,
      <Axes: title={'center': 'humidity'}>],
      [<Axes: title={'center': 'windspeed'}>,
      <Axes: title={'center': 'count'}>,
      <Axes: title={'center': 'hour'}>]], dtype=object)
```



Retrain AutoGluon model with engineered features to improve prediction accuracy

- Compare performance against baseline to measure feature engineering impact
- Rerun the model with the same settings as before, just with more features

```
In [49]: predictor_new_features = TabularPredictor(label="count", problem_type="regression")
```

No path specified. Models will be saved in: "AutogluonModels/ag-20250608_104126"

Verbosity: 2 (Standard Logging)

===== System Info =====

AutoGluon Version: 1.3.0
 Python Version: 3.12.9
 Operating System: Linux
 Platform Machine: x86_64
 Platform Version: #1 SMP Tue May 6 04:10:50 UTC 2025
 CPU Count: 2
 Memory Avail: 1.03 GB / 3.76 GB (27.3%)
 Disk Space Avail: 3.25 GB / 4.99 GB (65.2%)

WARNING: Available disk space is low and there is a risk that AutoGluon will run out of disk during fit, causing an exception.

We recommend a minimum available disk space of 10 GB, and large datasets may require more.

=====

Presets specified: ['best_quality']

Setting dynamic_stacking from 'auto' to True. Reason: Enable dynamic_stacking when use_bag_holdout is disabled. (use_bag_holdout=False)

Stack configuration (auto_stack=True): num_stack_levels=1, num_bag_folds=8, num_bag_sets=1

DyStack is enabled (dynamic_stacking=True). AutoGluon will try to determine whether the input data is affected by stacked overfitting and enable or disable stacking as a consequence.

This is used to identify the optimal `num_stack_levels` value. Copies of AutoGluon will be fit on subsets of the data. Then holdout validation data is used to detect stacked overfitting.

Running DyStack for up to 150s of the 600s of remaining time (25%).

Context path: "/home/sagemaker-user/Bike-sharing-starter/project/AutogluonModels/ag-20250608_104126/ds_sub_fit/sub_fit_ho"

Leaderboard on holdout data (DyStack):

	model	score_holdout	score_val	eval_metric
	pred_time_test	pred_time_val	fit_time	pred_time_test_marginal
	d_time_val_marginal	fit_time_marginal	stack_level	can_infer
	fit_order			
0	WeightedEnsemble_L3	-37.139853	-37.003909	root_mean_squared_error
	1.454791	0.263690	54.251660	0.003084
0.000581	0.027684	3	True	10
1	WeightedEnsemble_L2	-37.139853	-37.003909	root_mean_squared_error
	1.454955	0.263710	54.242146	0.003248
0.000601	0.018170	2	True	7
2	CatBoost_BAG_L1	-38.626025	-39.837065	root_mean_squared_error
	1.100951	0.040109	49.436006	1.100951
0.040109	49.436006	1	True	6
3	RandomForestMSE_BAG_L1	-40.219941	-39.544798	root_mean_squared_error
	0.329413	0.174089	4.768157	0.329413
0.174089	4.768157	1	True	5
4	KNeighborsDist_BAG_L1	-92.031272	-89.946854	root_mean_squared_error
	0.021343	0.048912	0.019813	0.021343
0.048912	0.019813	1	True	2
5	KNeighborsUnif_BAG_L1	-109.161488	-107.445008	root_mean_squared_error
	0.023305	0.041831	0.019940	0.023305
0.041831	0.019940	1	True	1
6	LightGBM_BAG_L2	-173.026156	-172.445919	root_mean_squared_error
	3.368909	0.469530	115.518295	0.028473
0.032703	19.743807	2	True	9
7	LightGBMXT_BAG_L2	-173.893125	-173.339212	root_mean_squared_error
	3.375990	0.478969	116.439513	0.035554
0.042142	20.665025	2	True	8
8	LightGBM_BAG_L1	-174.224120	-173.675134	root_mean_squared_error
	0.036841	0.043906	19.307505	0.036841
0.043906	19.307505	1	True	4
9	LightGBMXT_BAG_L1	-177.885111	-177.352728	root_mean_squared_error
	1.828583	0.087981	22.223067	1.828583

```

0.087981      22.223067      1      True      3
      1      = Optimal      num_stack_levels (Stacked Overfitting Occurre
d: False)
      169s      = DyStack      runtime | 431s      = Remaining runtime
Starting main fit with num_stack_levels=1.
      For future fit calls on this dataset, you can skip DyStack to save
time: `predictor.fit(..., dynamic_stacking=False, num_stack_levels=1)`
Beginning AutoGluon training ... Time limit = 431s
AutoGluon will save models to "/home/sagemaker-user/Bike-sharing-starter/pr
oject/AutogluonModels/ag-20250608_104126"
Train Data Rows:      10886
Train Data Columns: 10
Label Column:      count
Problem Type:      regression
Preprocessing data ...
Using Feature Generators to preprocess the data ...
Fitting AutoMLPipelineFeatureGenerator...
      Available Memory:      718.17 MB
      Train Data (Original) Memory Usage: 0.64 MB (0.1% of available mem
ory)
      Inferring data type of each feature based on column values. Set fea
ture_metadata_in to manually specify special dtypes of the features.
      Stage 1 Generators:
          Fitting AsTypeFeatureGenerator...
              Note: Converting 2 features to boolean dtype as the
y only contain 2 unique values.
      Stage 2 Generators:
          Fitting FillNaFeatureGenerator...
      Stage 3 Generators:
          Fitting IdentityFeatureGenerator...
          Fitting CategoryFeatureGenerator...
              Fitting CategoryMemoryMinimizeFeatureGenerator...
          Fitting DatetimeFeatureGenerator...
      Stage 4 Generators:
          Fitting DropUniqueFeatureGenerator...
      Stage 5 Generators:
          Fitting DropDuplicatesFeatureGenerator...
      Types of features in original data (raw dtype, special dtypes):
          ('category', []) : 2 | ['season', 'weather']
          ('datetime', []) : 1 | ['datetime']
          ('float', []) : 3 | ['temp', 'atemp', 'windspeed']
          ('int', []) : 4 | ['holiday', 'workingday', 'humidit
y', 'hour']
      Types of features in processed data (raw dtype, special dtypes):
          ('category', []) : 2 | ['season', 'weather']
          ('float', []) : 3 | ['temp', 'atemp', 'winds
peed']
          ('int', []) : 2 | ['humidity', 'hour']
          ('int', ['bool']) : 2 | ['holiday', 'workingda
y']
          ('int', ['datetime_as_int']) : 5 | ['datetime', 'datetime.y
ear', 'datetime.month', 'datetime.day', 'datetime.dayofweek']
      0.1s = Fit runtime
      10 features in original data used to generate 14 features in proces
sed data.
      Train Data (Processed) Memory Usage: 0.83 MB (0.1% of available mem
ory)
Data preprocessing and feature engineering runtime = 0.16s ...
AutoGluon will gauge predictive performance using evaluation metric: 'root_
mean_squared_error'
      This metric's sign has been flipped to adhere to being higher_is_be
tter. The metric score can be multiplied by -1 to get the metric value.
      To change this, specify the eval_metric parameter of Predictor()
Large model count detected (112 configs) ... Only displaying the first 3 mo

```


dels of each family. To see all, set `verbosity=3`.

User-specified model hyperparameters to be fit:

```
{
  'NN_TORCH': [{}, {'activation': 'elu', 'dropout_prob': 0.1007763952
9843717, 'hidden_size': 108, 'learning_rate': 0.002735937344002146, 'num_la
yers': 4, 'use_batchnorm': True, 'weight_decay': 1.356433327634438e-12, 'ag
_args': {'name_suffix': '_r79', 'priority': -2}}, {'activation': 'elu', 'dr
opout_prob': 0.11897478034205347, 'hidden_size': 213, 'learning_rate': 0.00
10474382260641949, 'num_layers': 4, 'use_batchnorm': False, 'weight_decay':
5.594471067786272e-10, 'ag_args': {'name_suffix': '_r22', 'priority': -
7}}],
  'GBM': [{}, {'extra_trees': True, 'ag_args': {'name_suffix': 'XT'}},
 {}, {'learning_rate': 0.03, 'num_leaves': 128, 'feature_fraction': 0.9, 'mi
n_data_in_leaf': 3, 'ag_args': {'name_suffix': 'Large', 'priority': 0, 'hyp
erparameter_tune_kwargs': None}}],
  'CAT': [{}, {'depth': 6, 'grow_policy': 'SymmetricTree', 'l2_leaf_r
eg': 2.1542798306067823, 'learning_rate': 0.06864209415792857, 'max_ctr_com
plexity': 4, 'one_hot_max_size': 10, 'ag_args': {'name_suffix': '_r177', 'p
riority': -1}}, {'depth': 8, 'grow_policy': 'Depthwise', 'l2_leaf_reg': 2.7
997999596449104, 'learning_rate': 0.031375015734637225, 'max_ctr_complexit
y': 2, 'one_hot_max_size': 3, 'ag_args': {'name_suffix': '_r9', 'priority':
-5}}],
  'XGB': [{}, {'colsample_bytree': 0.6917311125174739, 'enable_catego
rical': False, 'learning_rate': 0.018063876087523967, 'max_depth': 10, 'min
_child_weight': 0.6028633586934382, 'ag_args': {'name_suffix': '_r33', 'pri
ority': -8}}, {'colsample_bytree': 0.6628423832084077, 'enable_categorica
l': False, 'learning_rate': 0.08775715546881824, 'max_depth': 5, 'min_child
_weight': 0.6294123374222513, 'ag_args': {'name_suffix': '_r89', 'priorit
y': -16}}],
  'FASTAI': [{}, {'bs': 256, 'emb_drop': 0.5411770367537934, 'epoch
s': 43, 'layers': [800, 400], 'lr': 0.01519848858318159, 'ps': 0.2378294656
6604385, 'ag_args': {'name_suffix': '_r191', 'priority': -4}}, {'bs': 2048,
'emb_drop': 0.05070411322605811, 'epochs': 29, 'layers': [200, 100], 'lr':
0.08974235041576624, 'ps': 0.10393466140748028, 'ag_args': {'name_suffix':
'_r102', 'priority': -11}}],
  'RF': [{}, {'criterion': 'gini', 'ag_args': {'name_suffix': 'Gini', 'pr
oblem_types': ['binary', 'multiclass']}}, {'criterion': 'entropy', 'ag_arg
s': {'name_suffix': 'Entr', 'problem_types': ['binary', 'multiclass']}},
 {'criterion': 'squared_error', 'ag_args': {'name_suffix': 'MSE', 'problem_t
ypes': ['regression', 'quantile']}},
  'XT': [{}, {'criterion': 'gini', 'ag_args': {'name_suffix': 'Gini', 'pr
oblem_types': ['binary', 'multiclass']}}, {'criterion': 'entropy', 'ag_arg
s': {'name_suffix': 'Entr', 'problem_types': ['binary', 'multiclass']}},
 {'criterion': 'squared_error', 'ag_args': {'name_suffix': 'MSE', 'problem_t
ypes': ['regression', 'quantile']}},
  'KNN': [{}, {'weights': 'uniform', 'ag_args': {'name_suffix': 'Unif'}},
 {'weights': 'distance', 'ag_args': {'name_suffix': 'Dist'}}],
}
```

AutoGluon will fit 2 stack levels (L1 to L2) ...

Fitting 108 L1 models, fit_strategy="sequential" ...

Fitting model: KNeighborsUnif_BAG_L1 ... Training model for up to 287.44s o
f the 431.23s of remaining time.

```
-101.5462      = Validation score    (-root_mean_squared_error)
0.05s         = Training runtime
0.07s         = Validation runtime
```

Fitting model: KNeighborsDist_BAG_L1 ... Training model for up to 287.29s o
f the 431.08s of remaining time.

```
-84.1251       = Validation score    (-root_mean_squared_error)
0.04s         = Training runtime
0.06s         = Validation runtime
```

Fitting model: LightGBMXT_BAG_L1 ... Training model for up to 287.16s of th
e 430.95s of remaining time.

Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFo
ldFittingStrategy (2 workers, per: cpus=1, gpus=0, memory=1.51%)

```

-34.471 = Validation score (-root_mean_squared_error)
77.83s = Training runtime
17.94s = Validation runtime
Fitting model: LightGBM_BAG_L1 ... Training model for up to 203.27s of the
347.06s of remaining time.
Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFo
ldFittingStrategy (2 workers, per: cpus=1, gpus=0, memory=1.34%)
-33.9196 = Validation score (-root_mean_squared_error)
38.69s = Training runtime
3.43s = Validation runtime
Fitting model: RandomForestMSE_BAG_L1 ... Training model for up to 160.08s
of the 303.88s of remaining time.
Warning: Reducing model 'n_estimators' from 300 -> 141 due to low m
emory. Expected memory usage reduced from 31.85% -> 15.0% of available memo
ry...
/opt/conda/lib/python3.12/site-packages/sklearn/base.py:474: FutureWarning:
`BaseEstimator._validate_data` is deprecated in 1.6 and will be removed in
1.7. Use `sklearn.utils.validation.validate_data` instead. This function be
comes public and is part of the scikit-learn developer API.
warnings.warn(
-38.6539 = Validation score (-root_mean_squared_error)
8.14s = Training runtime
0.45s = Validation runtime
Fitting model: CatBoost_BAG_L1 ... Training model for up to 151.23s of the
295.03s of remaining time.
Memory not enough to fit 8 folds in parallel. Will train 2 folds in
parallel instead (Estimated 38.70% memory usage per fold, 77.41%/80.00% tot
al).
Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFo
ldFittingStrategy (2 workers, per: cpus=1, gpus=0, memory=38.70%)
I0000 00:00:1749379625.256052 5531 http2_transport.cc:1182] ipv4:169.25
5.255.2:35883: Got goaway [2] err=UNAVAILABLE:GOAWAY received; Error code:
2; Debug Text: Cancelling all calls {created_time:"2025-06-08T10:47:05.2541
55197+00:00", http2_error:2, grpc_status:14}
-34.7013 = Validation score (-root_mean_squared_error)
132.91s = Training runtime
0.16s = Validation runtime
Fitting model: ExtraTreesMSE_BAG_L1 ... Training model for up to 14.78s of
the 158.57s of remaining time.
Warning: Reducing model 'n_estimators' from 300 -> 105 due to low m
emory. Expected memory usage reduced from 42.65% -> 15.0% of available memo
ry...
/opt/conda/lib/python3.12/site-packages/sklearn/base.py:474: FutureWarning:
`BaseEstimator._validate_data` is deprecated in 1.6 and will be removed in
1.7. Use `sklearn.utils.validation.validate_data` instead. This function be
comes public and is part of the scikit-learn developer API.
warnings.warn(
-38.9638 = Validation score (-root_mean_squared_error)
3.11s = Training runtime
0.39s = Validation runtime
Fitting model: NeuralNetFastAI_BAG_L1 ... Training model for up to 11.08s o
f the 154.87s of remaining time.
Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFo
ldFittingStrategy (2 workers, per: cpus=1, gpus=0, memory=1.09%)
I0000 00:00:1749379741.005641 5524 http2_transport.cc:1182] ipv4:169.25
5.255.2:32849: Got goaway [2] err=UNAVAILABLE:GOAWAY received; Error code:
2; Debug Text: Cancelling all calls {created_time:"2025-06-08T10:49:01.0056
36861+00:00", http2_error:2, grpc_status:14}
-127.0155 = Validation score (-root_mean_squared_error)
37.14s = Training runtime
0.43s = Validation runtime
Fitting model: WeightedEnsemble_L2 ... Training model for up to 359.99s of
the 113.92s of remaining time.
Ensemble Weights: {'LightGBM_BAG_L1': 0.36, 'LightGBMXT_BAG_L1': 0.

```

```

32, 'CatBoost_BAG_L1': 0.16, 'RandomForestMSE_BAG_L1': 0.12, 'KNeighborsDis
t_BAG_L1': 0.04}
    -32.2885          = Validation score    (-root_mean_squared_error)
    0.04s             = Training runtime
    0.0s              = Validation runtime
Fitting 106 L2 models, fit_strategy="sequential" ...
Fitting model: LightGBMXT_BAG_L2 ... Training model for up to 113.83s of th
e 113.78s of remaining time.
    Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFo
ldFittingStrategy (2 workers, per: cpus=1, gpus=0, memory=3.06%)
I0000 00:00:1749379784.323460    5440 http2_transport.cc:1182] ipv4:169.25
5.255.2:32895: Got goaway [2] err=UNAVAILABLE:GOAWAY received; Error code:
2; Debug Text: Cancelling all calls {grpc_status:14, http2_error:2, created
_time:"2025-06-08T10:49:44.321890266+00:00"}
I0000 00:00:1749379805.950793    5524 http2_transport.cc:1182] ipv4:169.25
5.255.2:39875: Got goaway [2] err=UNAVAILABLE:GOAWAY received; Error code:
2; Debug Text: Cancelling all calls {created_time:"2025-06-08T10:50:05.9507
88357+00:00", http2_error:2, grpc_status:14}
    -31.0795          = Validation score    (-root_mean_squared_error)
    30.75s            = Training runtime
    1.06s             = Validation runtime
Fitting model: LightGBM_BAG_L2 ... Training model for up to 77.24s of the 7
7.19s of remaining time.
    Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFo
ldFittingStrategy (2 workers, per: cpus=1, gpus=0, memory=1.70%)
I0000 00:00:1749379837.814380    5532 http2_transport.cc:1182] ipv4:169.25
5.255.2:46337: Got goaway [2] err=UNAVAILABLE:GOAWAY received; Error code:
2; Debug Text: Cancelling all calls {created_time:"2025-06-08T10:50:37.8143
75838+00:00", http2_error:2, grpc_status:14}
    -30.5916          = Validation score    (-root_mean_squared_error)
    27.31s            = Training runtime
    0.42s             = Validation runtime
Fitting model: RandomForestMSE_BAG_L2 ... Training model for up to 45.94s o
f the 45.89s of remaining time.
    Warning: Reducing model 'n_estimators' from 300 -> 150 due to low m
emory. Expected memory usage reduced from 29.89% -> 15.0% of available memo
ry...
/opt/conda/lib/python3.12/site-packages/sklearn/base.py:474: FutureWarning:
`BaseEstimator._validate_data` is deprecated in 1.6 and will be removed in
1.7. Use `sklearn.utils.validation.validate_data` instead. This function be
comes public and is part of the scikit-learn developer API.
    warnings.warn(
    -31.7941          = Validation score    (-root_mean_squared_error)
    23.83s            = Training runtime
    0.4s              = Validation runtime
Fitting model: CatBoost_BAG_L2 ... Training model for up to 21.48s of the 2
1.44s of remaining time.
    Memory not enough to fit 8 folds in parallel. Will train 1 folds in
parallel instead (Estimated 40.24% memory usage per fold, 40.24%/80.00% tot
al).
    Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFo
ldFittingStrategy (1 workers, per: cpus=1, gpus=0, memory=40.24%)
    Switching to pseudo sequential ParallelFoldFittingStrategy
to avoid Python memory leakage.
    Override this behavior by setting fold_fitting_strategy to
'sequential_local' in ag_args_ensemble when when calling `predictor.fit`
    Time limit exceeded... Skipping CatBoost_BAG_L2.
Fitting model: ExtraTreesMSE_BAG_L2 ... Training model for up to 16.90s of
the 16.86s of remaining time.
    Warning: Reducing model 'n_estimators' from 300 -> 141 due to low m
emory. Expected memory usage reduced from 31.73% -> 15.0% of available memo
ry...
/opt/conda/lib/python3.12/site-packages/sklearn/base.py:474: FutureWarning:
`BaseEstimator._validate_data` is deprecated in 1.6 and will be removed in

```

```

1.7. Use `sklearn.utils.validation.validate_data` instead. This function be
comes public and is part of the scikit-learn developer API.
warnings.warn(
    -31.744 = Validation score (-root_mean_squared_error)
    6.34s   = Training runtime
    0.35s   = Validation runtime
Fitting model: NeuralNetFastAI_BAG_L2 ... Training model for up to 10.01s o
f the 9.97s of remaining time.
    Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFo
ldFittingStrategy (2 workers, per: cpus=1, gpus=0, memory=1.33%)
I0000 00:00:1749379885.927491 5442 http2_transport.cc:1182] ipv4:169.25
5.255.2:38559: Got goaway [2] err=UNAVAILABLE:GOAWAY received; Error code:
2; Debug Text: Cancelling all calls {grpc_status:14, http2_error:2, created
_time:"2025-06-08T10:51:25.927486216+00:00"}
    -47.6581 = Validation score (-root_mean_squared_error)
    40.23s   = Training runtime
    0.41s   = Validation runtime
Fitting model: WeightedEnsemble_L3 ... Training model for up to 360.00s of
the -34.05s of remaining time.
    Ensemble Weights: {'LightGBM_BAG_L2': 0.526, 'LightGBMXT_BAG_L2':
0.316, 'RandomForestMSE_BAG_L2': 0.105, 'CatBoost_BAG_L1': 0.053}
    -30.3035 = Validation score (-root_mean_squared_error)
    0.05s    = Training runtime
    0.0s     = Validation runtime
AutoGluon training complete, total runtime = 465.59s ... Best model: Weight
edEnsemble_L3 | Estimated inference throughput: 57.7 rows/s (1361 batch siz
e)
TabularPredictor saved. To load, use: predictor = TabularPredictor.load("/h
ome/sagemaker-user/Bike-sharing-starter/project/AutogluonModels/ag-20250608
_104126")

```

```
In [50]: predictor_new_features.fit_summary()
```

*** Summary of fit() ***

Estimated performance of each model:

val	fit_time	pred_time_val	model	score_val	eval_metric	pred_time_val	stack_level	ca
n_infer	fit_order			marginal				
0	WeightedEnsemble_L3	-30.303501	root_mean_squared_error	24.801				
925	379.843397	0.000621	0.045016	3				
True	15							
1	LightGBM_BAG_L2	-30.591648	root_mean_squared_error	23.341				
903	325.219482	0.424813	27.312869	2				
True	11							
2	LightGBMXT_BAG_L2	-31.079503	root_mean_squared_error	23.978				
597	328.653043	1.061507	30.746430	2				
True	10							
3	ExtraTreesMSE_BAG_L2	-31.744020	root_mean_squared_error	23.268				
171	304.247006	0.351081	6.340394	2				
True	13							
4	RandomForestMSE_BAG_L2	-31.794130	root_mean_squared_error	23.314				
985	321.739082	0.397895	23.832469	2				
True	12							
5	WeightedEnsemble_L2	-32.288502	root_mean_squared_error	22.028				
649	257.656003	0.000901	0.043982	2				
True	9							
6	LightGBM_BAG_L1	-33.919639	root_mean_squared_error	3.425				
256	38.685870	3.425256	38.685870	1				
True	4							
7	LightGBMXT_BAG_L1	-34.470975	root_mean_squared_error	17.935				
651	77.832888	17.935651	77.832888	1				
True	3							
8	CatBoost_BAG_L1	-34.701317	root_mean_squared_error	0.157				
199	132.912889	0.157199	132.912889	1				
True	6							
9	RandomForestMSE_BAG_L1	-38.653908	root_mean_squared_error	0.452				
213	8.143479	0.452213	8.143479	1				
True	5							
10	ExtraTreesMSE_BAG_L1	-38.963841	root_mean_squared_error	0.394				
691	3.105655	0.394691	3.105655	1				
True	7							
11	NeuralNetFastAI_BAG_L2	-47.658102	root_mean_squared_error	23.328				
594	338.136898	0.411504	40.230285	2				
True	14							
12	KNeighborsDist_BAG_L1	-84.125061	root_mean_squared_error	0.057				
429	0.036895	0.057429	0.036895	1				
True	2							
13	KNeighborsUnif_BAG_L1	-101.546199	root_mean_squared_error	0.065				
984	0.046656	0.065984	0.046656	1				
True	1							
14	NeuralNetFastAI_BAG_L1	-127.015492	root_mean_squared_error	0.428				
668	37.142281	0.428668	37.142281	1				
True	8							

Number of models trained: 15

Types of models trained:

```
{'StackerEnsembleModel_KNN', 'StackerEnsembleModel_XT', 'StackerEnsembleModel_RF', 'StackerEnsembleModel_NNFastAiTabular', 'StackerEnsembleModel_CatBoost', 'StackerEnsembleModel_LGB', 'WeightedEnsembleModel'}
```

Bagging used: True (with 8 folds)

Multi-layer stack-ensembling used: True (with 3 levels)

Feature Metadata (Processed):

(raw dtype, special dtypes):

```
('category', []) : 2 | ['season', 'weather']
('float', []) : 3 | ['temp', 'atemp', 'windspeed']
('int', []) : 2 | ['humidity', 'hour']
('int', ['bool']) : 2 | ['holiday', 'workingday']
('int', ['datetime_as_int']) : 5 | ['datetime', 'datetime.year', 'datetime.month', 'datetime.day', 'datetime.hour', 'datetime.minute', 'datetime.second']
```

```
month', 'datetime.day', 'datetime.dayofweek']  
Plot summary of models saved to file: /home/sagemaker-user/Bike-sharing-starter/project/AutogluonModels/ag-20250608_104126/SummaryOfModels.html  
*** End of fit() summary ***
```

```

Out[50]: {'model_types': {'KNeighborsUnif_BAG_L1': 'StackerEnsembleModel_KNN',
  'KNeighborsDist_BAG_L1': 'StackerEnsembleModel_KNN',
  'LightGBMXt_BAG_L1': 'StackerEnsembleModel_LGB',
  'LightGBM_BAG_L1': 'StackerEnsembleModel_LGB',
  'RandomForestMSE_BAG_L1': 'StackerEnsembleModel_RF',
  'CatBoost_BAG_L1': 'StackerEnsembleModel_CatBoost',
  'ExtraTreesMSE_BAG_L1': 'StackerEnsembleModel_XT',
  'NeuralNetFastAI_BAG_L1': 'StackerEnsembleModel_NNFastAiTabular',
  'WeightedEnsemble_L2': 'WeightedEnsembleModel',
  'LightGBMXt_BAG_L2': 'StackerEnsembleModel_LGB',
  'LightGBM_BAG_L2': 'StackerEnsembleModel_LGB',
  'RandomForestMSE_BAG_L2': 'StackerEnsembleModel_RF',
  'ExtraTreesMSE_BAG_L2': 'StackerEnsembleModel_XT',
  'NeuralNetFastAI_BAG_L2': 'StackerEnsembleModel_NNFastAiTabular',
  'WeightedEnsemble_L3': 'WeightedEnsembleModel'},
  'model_performance': {'KNeighborsUnif_BAG_L1': -101.54619908446061,
  'KNeighborsDist_BAG_L1': -84.12506123181602,
  'LightGBMXt_BAG_L1': -34.47097500967876,
  'LightGBM_BAG_L1': -33.919639163586254,
  'RandomForestMSE_BAG_L1': -38.653907901218226,
  'CatBoost_BAG_L1': -34.701317317404026,
  'ExtraTreesMSE_BAG_L1': -38.96384109086304,
  'NeuralNetFastAI_BAG_L1': -127.01549246851197,
  'WeightedEnsemble_L2': -32.288502069936655,
  'LightGBMXt_BAG_L2': -31.079502528484166,
  'LightGBM_BAG_L2': -30.591648035479025,
  'RandomForestMSE_BAG_L2': -31.79412967155912,
  'ExtraTreesMSE_BAG_L2': -31.74401965649366,
  'NeuralNetFastAI_BAG_L2': -47.65810213665935,
  'WeightedEnsemble_L3': -30.303501082551765},
  'model_best': 'WeightedEnsemble_L3',
  'model_paths': {'KNeighborsUnif_BAG_L1': ['KNeighborsUnif_BAG_L1'],
  'KNeighborsDist_BAG_L1': ['KNeighborsDist_BAG_L1'],
  'LightGBMXt_BAG_L1': ['LightGBMXt_BAG_L1'],
  'LightGBM_BAG_L1': ['LightGBM_BAG_L1'],
  'RandomForestMSE_BAG_L1': ['RandomForestMSE_BAG_L1'],
  'CatBoost_BAG_L1': ['CatBoost_BAG_L1'],
  'ExtraTreesMSE_BAG_L1': ['ExtraTreesMSE_BAG_L1'],
  'NeuralNetFastAI_BAG_L1': ['NeuralNetFastAI_BAG_L1'],
  'WeightedEnsemble_L2': ['WeightedEnsemble_L2'],
  'LightGBMXt_BAG_L2': ['LightGBMXt_BAG_L2'],
  'LightGBM_BAG_L2': ['LightGBM_BAG_L2'],
  'RandomForestMSE_BAG_L2': ['RandomForestMSE_BAG_L2'],
  'ExtraTreesMSE_BAG_L2': ['ExtraTreesMSE_BAG_L2'],
  'NeuralNetFastAI_BAG_L2': ['NeuralNetFastAI_BAG_L2'],
  'WeightedEnsemble_L3': ['WeightedEnsemble_L3']},
  'model_fit_times': {'KNeighborsUnif_BAG_L1': 0.04665565490722656,
  'KNeighborsDist_BAG_L1': 0.036894798278808594,
  'LightGBMXt_BAG_L1': 77.83288788795471,
  'LightGBM_BAG_L1': 38.6858696937561,
  'RandomForestMSE_BAG_L1': 8.143479347229004,
  'CatBoost_BAG_L1': 132.91288948059082,
  'ExtraTreesMSE_BAG_L1': 3.1056551933288574,
  'NeuralNetFastAI_BAG_L1': 37.14228081703186,
  'WeightedEnsemble_L2': 0.04398202896118164,
  'LightGBMXt_BAG_L2': 30.74643039703369,
  'LightGBM_BAG_L2': 27.31286931037903,
  'RandomForestMSE_BAG_L2': 23.83246898651123,
  'ExtraTreesMSE_BAG_L2': 6.340393543243408,
  'NeuralNetFastAI_BAG_L2': 40.23028492927551,
  'WeightedEnsemble_L3': 0.045015573501586914},
  'model_pred_times': {'KNeighborsUnif_BAG_L1': 0.06598377227783203,
  'KNeighborsDist_BAG_L1': 0.057428836822509766,
  'LightGBMXt_BAG_L1': 17.935651063919067,

```

```
'LightGBM_BAG_L1': 3.42525577545166,  
'RandomForestMSE_BAG_L1': 0.4522125720977783,  
'CatBoost_BAG_L1': 0.15719938278198242,  
'ExtraTreesMSE_BAG_L1': 0.39469122886657715,  
'NeuralNetFastAI_BAG_L1': 0.4286675453186035,  
'WeightedEnsemble_L2': 0.0009009838104248047,  
'LightGBMXT_BAG_L2': 1.061507225036621,  
'LightGBM_BAG_L2': 0.42481255531311035,  
'RandomForestMSE_BAG_L2': 0.39789462089538574,  
'ExtraTreesMSE_BAG_L2': 0.35108113288879395,  
'NeuralNetFastAI_BAG_L2': 0.41150355339050293,  
'WeightedEnsemble_L3': 0.0006206035614013672},  
'num_bag_folds': 8,  
'max_stack_level': 3,  
'model_hyperparams': {'KNeighborsUnif_BAG_L1': {'use_orig_features': True,  
  'valid_stack': True,  
  'max_base_models': 0,  
  'max_base_models_per_type': 'auto',  
  'save_bag_folds': True,  
  'stratify': 'auto',  
  'bin': 'auto',  
  'n_bins': None,  
  'use_child_oof': True},  
'KNeighborsDist_BAG_L1': {'use_orig_features': True,  
  'valid_stack': True,  
  'max_base_models': 0,  
  'max_base_models_per_type': 'auto',  
  'save_bag_folds': True,  
  'stratify': 'auto',  
  'bin': 'auto',  
  'n_bins': None,  
  'use_child_oof': True},  
'LightGBMXT_BAG_L1': {'use_orig_features': True,  
  'valid_stack': True,  
  'max_base_models': 0,  
  'max_base_models_per_type': 'auto',  
  'save_bag_folds': True,  
  'stratify': 'auto',  
  'bin': 'auto',  
  'n_bins': None},  
'LightGBM_BAG_L1': {'use_orig_features': True,  
  'valid_stack': True,  
  'max_base_models': 0,  
  'max_base_models_per_type': 'auto',  
  'save_bag_folds': True,  
  'stratify': 'auto',  
  'bin': 'auto',  
  'n_bins': None},  
'RandomForestMSE_BAG_L1': {'use_orig_features': True,  
  'valid_stack': True,  
  'max_base_models': 0,  
  'max_base_models_per_type': 'auto',  
  'save_bag_folds': True,  
  'stratify': 'auto',  
  'bin': 'auto',  
  'n_bins': None,  
  'use_child_oof': True},  
'CatBoost_BAG_L1': {'use_orig_features': True,  
  'valid_stack': True,  
  'max_base_models': 0,  
  'max_base_models_per_type': 'auto',  
  'save_bag_folds': True,  
  'stratify': 'auto',  
  'bin': 'auto',
```



```
'n_bins': None},
'ExtraTreesMSE_BAG_L1': {'use_orig_features': True,
'valid_stack': True,
'max_base_models': 0,
'max_base_models_per_type': 'auto',
'save_bag_folds': True,
'stratif': 'auto',
'bin': 'auto',
'n_bins': None,
'use_child_oof': True},
'NeuralNetFastAI_BAG_L1': {'use_orig_features': True,
'valid_stack': True,
'max_base_models': 0,
'max_base_models_per_type': 'auto',
'save_bag_folds': True,
'stratif': 'auto',
'bin': 'auto',
'n_bins': None},
'WeightedEnsemble_L2': {'use_orig_features': False,
'valid_stack': True,
'max_base_models': 0,
'max_base_models_per_type': 'auto',
'save_bag_folds': True,
'stratif': 'auto',
'bin': 'auto',
'n_bins': None},
'LightGBMXT_BAG_L2': {'use_orig_features': True,
'valid_stack': True,
'max_base_models': 0,
'max_base_models_per_type': 'auto',
'save_bag_folds': True,
'stratif': 'auto',
'bin': 'auto',
'n_bins': None},
'LightGBM_BAG_L2': {'use_orig_features': True,
'valid_stack': True,
'max_base_models': 0,
'max_base_models_per_type': 'auto',
'save_bag_folds': True,
'stratif': 'auto',
'bin': 'auto',
'n_bins': None},
'RandomForestMSE_BAG_L2': {'use_orig_features': True,
'valid_stack': True,
'max_base_models': 0,
'max_base_models_per_type': 'auto',
'save_bag_folds': True,
'stratif': 'auto',
'bin': 'auto',
'n_bins': None,
'use_child_oof': True},
'ExtraTreesMSE_BAG_L2': {'use_orig_features': True,
'valid_stack': True,
'max_base_models': 0,
'max_base_models_per_type': 'auto',
'save_bag_folds': True,
'stratif': 'auto',
'bin': 'auto',
'n_bins': None,
'use_child_oof': True},
'NeuralNetFastAI_BAG_L2': {'use_orig_features': True,
'valid_stack': True,
'max_base_models': 0,
'max_base_models_per_type': 'auto',
```

```

'save_bag_folds': True,
'stratify': 'auto',
'bin': 'auto',
'n_bins': None},
'WeightedEnsemble_L3': {'use_orig_features': False,
'valid_stack': True,
'max_base_models': 0,
'max_base_models_per_type': 'auto',
'save_bag_folds': True,
'stratify': 'auto',
'bin': 'auto',
'n_bins': None}},
'leaderboard':

```

		model	score_val	eval_me
tric \				
0	WeightedEnsemble_L3	-30.303501	root_mean_squared_error	
1	LightGBM_BAG_L2	-30.591648	root_mean_squared_error	
2	LightGBMXT_BAG_L2	-31.079503	root_mean_squared_error	
3	ExtraTreesMSE_BAG_L2	-31.744020	root_mean_squared_error	
4	RandomForestMSE_BAG_L2	-31.794130	root_mean_squared_error	
5	WeightedEnsemble_L2	-32.288502	root_mean_squared_error	
6	LightGBM_BAG_L1	-33.919639	root_mean_squared_error	
7	LightGBMXT_BAG_L1	-34.470975	root_mean_squared_error	
8	CatBoost_BAG_L1	-34.701317	root_mean_squared_error	
9	RandomForestMSE_BAG_L1	-38.653908	root_mean_squared_error	
10	ExtraTreesMSE_BAG_L1	-38.963841	root_mean_squared_error	
11	NeuralNetFastAI_BAG_L2	-47.658102	root_mean_squared_error	
12	KNeighborsDist_BAG_L1	-84.125061	root_mean_squared_error	
13	KNeighborsUnif_BAG_L1	-101.546199	root_mean_squared_error	
14	NeuralNetFastAI_BAG_L1	-127.015492	root_mean_squared_error	

	pred_time_val	fit_time	pred_time_val_marginal	fit_time_marginal
\				
0	24.801925	379.843397	0.000621	0.045016
1	23.341903	325.219482	0.424813	27.312869
2	23.978597	328.653043	1.061507	30.746430
3	23.268171	304.247006	0.351081	6.340394
4	23.314985	321.739082	0.397895	23.832469
5	22.028649	257.656003	0.000901	0.043982
6	3.425256	38.685870	3.425256	38.685870
7	17.935651	77.832888	17.935651	77.832888
8	0.157199	132.912889	0.157199	132.912889
9	0.452213	8.143479	0.452213	8.143479
10	0.394691	3.105655	0.394691	3.105655
11	23.328594	338.136898	0.411504	40.230285
12	0.057429	0.036895	0.057429	0.036895
13	0.065984	0.046656	0.065984	0.046656
14	0.428668	37.142281	0.428668	37.142281

	stack_level	can_infer	fit_order
0	3	True	15
1	2	True	11
2	2	True	10
3	2	True	13
4	2	True	12
5	2	True	9
6	1	True	4
7	1	True	3
8	1	True	6
9	1	True	5
10	1	True	7
11	2	True	14
12	1	True	2
13	1	True	1
14	1	True	8

```
In [56]: predictions_new = predictor_new_features.predict(test)
print(predictions_new)
```

```
0      16.956493
1      9.490427
2      8.326609
3      7.593449
4      6.826701
...
6488   298.642059
6489   205.048630
6490   152.870193
6491   109.854591
6492    71.154053
Name: count, Length: 6493, dtype: float32
```

```
In [60]: predictions_new[predictions_new < 0] = 0
```

```
In [108... predictor_new_features.leaderboard(silent=True)
```

```
Out[108]:
```

	model	score_val	eval_metric	pred_time_val	fit_time_val
0	WeightedEnsemble_L3	-30.303501	root_mean_squared_error	24.801925	379.843
1	LightGBM_BAG_L2	-30.591648	root_mean_squared_error	23.341903	325.219
2	LightGBMXT_BAG_L2	-31.079503	root_mean_squared_error	23.978597	328.653
3	ExtraTreesMSE_BAG_L2	-31.744020	root_mean_squared_error	23.268171	304.247
4	RandomForestMSE_BAG_L2	-31.794130	root_mean_squared_error	23.314985	321.739
5	WeightedEnsemble_L2	-32.288502	root_mean_squared_error	22.028649	257.656
6	LightGBM_BAG_L1	-33.919639	root_mean_squared_error	3.425256	38.685
7	LightGBMXT_BAG_L1	-34.470975	root_mean_squared_error	17.935651	77.832
8	CatBoost_BAG_L1	-34.701317	root_mean_squared_error	0.157199	132.912
9	RandomForestMSE_BAG_L1	-38.653908	root_mean_squared_error	0.452213	8.143
10	ExtraTreesMSE_BAG_L1	-38.963841	root_mean_squared_error	0.394691	3.105
11	NeuralNetFastAI_BAG_L2	-47.658102	root_mean_squared_error	23.328594	338.136
12	KNeighborsDist_BAG_L1	-84.125061	root_mean_squared_error	0.057429	0.036
13	KNeighborsUnif_BAG_L1	-101.546199	root_mean_squared_error	0.065984	0.046
14	NeuralNetFastAI_BAG_L1	-127.015492	root_mean_squared_error	0.428668	37.143

```
In [62]: submission_new_features = pd.read_csv("sampleSubmission.csv", parse_dates=['datetime'])
submission_new_features.head()
```

```
Out[62]:
```

	datetime	count
0	2011-01-20 00:00:00	0
1	2011-01-20 01:00:00	0
2	2011-01-20 02:00:00	0
3	2011-01-20 03:00:00	0
4	2011-01-20 04:00:00	0

```
In [63]: submission_new_features["count"] = predictions_new
submission_new_features.to_csv("submission_new_features.csv", index=False)
```

```
In [64]: !kaggle competitions submit -c bike-sharing-demand -f submission_new_features
100%|████████████████████████████████████████| 188k/188k [00:00<00:00, 760
kB/s]
Successfully submitted to Bike Sharing Demand
```

```
In [65]: !kaggle competitions submissions -c bike-sharing-demand | tail -n +1 | head
```

fileName	date		description
status	publicScore	privateScore	
submission_new_features.csv	2025-06-08	11:28:24.850000	new features
SubmissionStatus.COMPLETE	0.61489	0.61489	
submission.csv	2025-06-08	10:04:09.937000	first raw submission
SubmissionStatus.COMPLETE	1.80019	1.80019	

New Score of **0.61489**

Hyper parameter optimization

- Configure AutoGluon to use multiple algorithms (XGB, GBM, CatBoost, RF, NN) with automated tuning
- Optimize model hyperparameters to maximize predictive performance on validation set

```
In [68]: predictor_new_hpo = TabularPredictor(label="count", problem_type="regression",
        'XGB': {},
        'GBM': {},
        'CAT': {},
        'RF': {},
        'NN_TORCH': {},
        }, hyperparameter_tune_kwargs='auto')
```

```

2025-06-08 12:34:38,953 INFO timeout.py:54 -- Reached timeout of 36.2958725
4524231 seconds. Stopping all trials.
2025-06-08 12:34:38,972 INFO tune.py:1009 -- Wrote the latest version of al
l result files and experiment state to '/home/sagemaker-user/Bike-sharing-s
tarter/project/AutogluonModels/ag-20250608_122531/models/NeuralNetTorch_BAG
_L2' in 0.0063s.
2025-06-08 12:34:49,065 WARNING experiment_analysis.py:180 -- Failed to fet
ch metrics for 2 trial(s):
- 0aa18bfa: FileNotFoundError('Could not fetch metrics for 0aa18bfa: both r
esult.json and progress.csv were not found at /home/sagemaker-user/Bike-sha
ring-starter/project/AutogluonModels/ag-20250608_122531/models/NeuralNetTor
ch_BAG_L2/0aa18bfa')
- 07283bdd: FileNotFoundError('Could not fetch metrics for 07283bdd: both r
esult.json and progress.csv were not found at /home/sagemaker-user/Bike-sha
ring-starter/project/AutogluonModels/ag-20250608_122531/models/NeuralNetTor
ch_BAG_L2/07283bdd')
No model was trained during hyperparameter tuning NeuralNetTorch_BAG_L2...
Skipping this model.
Fitting model: WeightedEnsemble_L3 ... Training model for up to 360.00s of
the 42.26s of remaining time.
    Ensemble Weights: {'LightGBM_BAG_L1/T1': 0.292, 'CatBoost_BAG_L2/T
1': 0.25, 'XGBoost_BAG_L1/T1': 0.208, 'RandomForest_BAG_L1': 0.083, 'Random
Forest_BAG_L2': 0.083, 'XGBoost_BAG_L2/T1': 0.083}
    -32.7655      = Validation score    (-root_mean_squared_error)
    0.04s        = Training runtime
    0.0s         = Validation runtime
AutoGluon training complete, total runtime = 375.11s ... Best model: Weight
edEnsemble_L3 | Estimated inference throughput: 205.1 rows/s (1361 batch si
ze)
TabularPredictor saved. To load, use: predictor = TabularPredictor.load("/h
ome/sagemaker-user/Bike-sharing-starter/project/AutogluonModels/ag-20250608
_122531")

```

```
In [69]: predictor_new_hpo.fit_summary()
```

*** Summary of fit() ***

Estimated performance of each model:

	model	score_val	eval_metric	pred_time_val	
fit_time	pred_time_val	marginal	fit_time	marginal	stack_level
fit_order					can_infer
0	WeightedEnsemble_L3	-32.765532	root_mean_squared_error	7.319692	
236.785160		0.000819	0.041736	3	Tr
ue	10				
1	WeightedEnsemble_L2	-32.880094	root_mean_squared_error	6.742028	
153.987541		0.001855	0.044647	2	Tr
ue	5				
2	CatBoost_BAG_L2/T1	-33.212090	root_mean_squared_error	6.778969	
200.448790		0.038796	46.505897	2	Tr
ue	8				
3	LightGBM_BAG_L2/T1	-33.478539	root_mean_squared_error	6.922361	
183.530025		0.182188	29.587132	2	Tr
ue	6				
4	XGBoost_BAG_L2/T1	-33.695101	root_mean_squared_error	6.892488	
178.289376		0.152315	24.346482	2	Tr
ue	9				
5	LightGBM_BAG_L1/T1	-33.919639	root_mean_squared_error	4.011557	
45.529637		4.011557	45.529637	1	Tru
e	1				
6	RandomForest_BAG_L2	-34.288797	root_mean_squared_error	7.127762	
165.891045		0.387589	11.948151	2	Tr
ue	7				
7	XGBoost_BAG_L1/T1	-34.603186	root_mean_squared_error	2.288409	
43.774752		2.288409	43.774752	1	Tru
e	4				
8	RandomForest_BAG_L1	-38.714073	root_mean_squared_error	0.393396	
8.292387		0.393396	8.292387	1	True
2					
9	CatBoost_BAG_L1/T1	-39.452389	root_mean_squared_error	0.046811	
56.346117		0.046811	56.346117	1	Tru
e	3				

Number of models trained: 10

Types of models trained:

{'StackerEnsembleModel_XGBoost', 'StackerEnsembleModel_RF', 'StackerEnsembleModel_CatBoost', 'StackerEnsembleModel_LGB', 'WeightedEnsembleModel'}

Bagging used: True (with 8 folds)

Multi-layer stack-ensembling used: True (with 3 levels)

Feature Metadata (Processed):

(raw dtype, special dtypes):

```
( 'category', [] ) : 2 | ['season', 'weather']
( 'float', [] ) : 3 | ['temp', 'atemp', 'windspeed']
( 'int', [] ) : 2 | ['humidity', 'hour']
( 'int', ['bool'] ) : 2 | ['holiday', 'workingday']
( 'int', ['datetime_as_int'] ) : 5 | ['datetime', 'datetime.year', 'datetime.month', 'datetime.day', 'datetime.dayofweek']
```

Plot summary of models saved to file: /home/sagemaker-user/Bike-sharing-starter/project/AutogluonModels/ag-20250608_122531/SummaryOfModels.html

*** End of fit() summary ***

```

Out[69]: {'model_types': {'LightGBM_BAG_L1/T1': 'StackerEnsembleModel_LGB',
    'RandomForest_BAG_L1': 'StackerEnsembleModel_RF',
    'CatBoost_BAG_L1/T1': 'StackerEnsembleModel_CatBoost',
    'XGBoost_BAG_L1/T1': 'StackerEnsembleModel_XGBoost',
    'WeightedEnsemble_L2': 'WeightedEnsembleModel',
    'LightGBM_BAG_L2/T1': 'StackerEnsembleModel_LGB',
    'RandomForest_BAG_L2': 'StackerEnsembleModel_RF',
    'CatBoost_BAG_L2/T1': 'StackerEnsembleModel_CatBoost',
    'XGBoost_BAG_L2/T1': 'StackerEnsembleModel_XGBoost',
    'WeightedEnsemble_L3': 'WeightedEnsembleModel'},
    'model_performance': {'LightGBM_BAG_L1/T1': -33.919639163586254,
    'RandomForest_BAG_L1': -38.71407349745564,
    'CatBoost_BAG_L1/T1': -39.45238909335168,
    'XGBoost_BAG_L1/T1': -34.603185627841,
    'WeightedEnsemble_L2': -32.88009397990229,
    'LightGBM_BAG_L2/T1': -33.4785392784392,
    'RandomForest_BAG_L2': -34.28879653859491,
    'CatBoost_BAG_L2/T1': -33.212089644157466,
    'XGBoost_BAG_L2/T1': -33.6951011021837,
    'WeightedEnsemble_L3': -32.76553221308848},
    'model_best': 'WeightedEnsemble_L3',
    'model_paths': {'LightGBM_BAG_L1/T1': ['LightGBM_BAG_L1', 'T1'],
    'RandomForest_BAG_L1': ['RandomForest_BAG_L1'],
    'CatBoost_BAG_L1/T1': ['CatBoost_BAG_L1', 'T1'],
    'XGBoost_BAG_L1/T1': ['XGBoost_BAG_L1', 'T1'],
    'WeightedEnsemble_L2': ['WeightedEnsemble_L2'],
    'LightGBM_BAG_L2/T1': ['LightGBM_BAG_L2', 'T1'],
    'RandomForest_BAG_L2': ['RandomForest_BAG_L2'],
    'CatBoost_BAG_L2/T1': ['CatBoost_BAG_L2', 'T1'],
    'XGBoost_BAG_L2/T1': ['XGBoost_BAG_L2', 'T1'],
    'WeightedEnsemble_L3': ['WeightedEnsemble_L3']},
    'model_fit_times': {'LightGBM_BAG_L1/T1': 45.5296368598938,
    'RandomForest_BAG_L1': 8.292387247085571,
    'CatBoost_BAG_L1/T1': 56.34611749649048,
    'XGBoost_BAG_L1/T1': 43.774752140045166,
    'WeightedEnsemble_L2': 0.0446469783782959,
    'LightGBM_BAG_L2/T1': 29.58713150024414,
    'RandomForest_BAG_L2': 11.948151350021362,
    'CatBoost_BAG_L2/T1': 46.50589656829834,
    'XGBoost_BAG_L2/T1': 24.346482038497925,
    'WeightedEnsemble_L3': 0.04173636436462402},
    'model_pred_times': {'LightGBM_BAG_L1/T1': 4.011557102203369,
    'RandomForest_BAG_L1': 0.39339590072631836,
    'CatBoost_BAG_L1/T1': 0.04681134223937988,
    'XGBoost_BAG_L1/T1': 2.2884085178375244,
    'WeightedEnsemble_L2': 0.0018548965454101562,
    'LightGBM_BAG_L2/T1': 0.18218779563903809,
    'RandomForest_BAG_L2': 0.3875887393951416,
    'CatBoost_BAG_L2/T1': 0.038796186447143555,
    'XGBoost_BAG_L2/T1': 0.15231537818908691,
    'WeightedEnsemble_L3': 0.0008192062377929688},
    'num_bag_folds': 8,
    'max_stack_level': 3,
    'model_hyperparams': {'LightGBM_BAG_L1/T1': {'use_orig_features': True,
    'valid_stacker': True,
    'max_base_models': 0,
    'max_base_models_per_type': 'auto',
    'save_bag_folds': True,
    'stratify': 'auto',
    'bin': 'auto',
    'n_bins': None},
    'RandomForest_BAG_L1': {'use_orig_features': True,
    'valid_stacker': True,
    'max_base_models': 0,

```

```
'max_base_models_per_type': 'auto',
'save_bag_folds': True,
'stratify': 'auto',
'bin': 'auto',
'n_bins': None,
'use_child_oof': True},
'CatBoost_BAG_L1/T1': {'use_orig_features': True,
'valid_stack': True,
'max_base_models': 0,
'max_base_models_per_type': 'auto',
'save_bag_folds': True,
'stratify': 'auto',
'bin': 'auto',
'n_bins': None},
'XGBoost_BAG_L1/T1': {'use_orig_features': True,
'valid_stack': True,
'max_base_models': 0,
'max_base_models_per_type': 'auto',
'save_bag_folds': True,
'stratify': 'auto',
'bin': 'auto',
'n_bins': None},
'WeightedEnsemble_L2': {'use_orig_features': False,
'valid_stack': True,
'max_base_models': 0,
'max_base_models_per_type': 'auto',
'save_bag_folds': True,
'stratify': 'auto',
'bin': 'auto',
'n_bins': None},
'LightGBM_BAG_L2/T1': {'use_orig_features': True,
'valid_stack': True,
'max_base_models': 0,
'max_base_models_per_type': 'auto',
'save_bag_folds': True,
'stratify': 'auto',
'bin': 'auto',
'n_bins': None},
'RandomForest_BAG_L2': {'use_orig_features': True,
'valid_stack': True,
'max_base_models': 0,
'max_base_models_per_type': 'auto',
'save_bag_folds': True,
'stratify': 'auto',
'bin': 'auto',
'n_bins': None,
'use_child_oof': True},
'CatBoost_BAG_L2/T1': {'use_orig_features': True,
'valid_stack': True,
'max_base_models': 0,
'max_base_models_per_type': 'auto',
'save_bag_folds': True,
'stratify': 'auto',
'bin': 'auto',
'n_bins': None},
'XGBoost_BAG_L2/T1': {'use_orig_features': True,
'valid_stack': True,
'max_base_models': 0,
'max_base_models_per_type': 'auto',
'save_bag_folds': True,
'stratify': 'auto',
'bin': 'auto',
'n_bins': None},
'WeightedEnsemble_L3': {'use_orig_features': False,
```



```

'valid_stack': True,
'max_base_models': 0,
'max_base_models_per_type': 'auto',
'save_bag_folds': True,
'stratify': 'auto',
'bin': 'auto',
'n_bins': None}},
'leaderboard':

```

		model	score_val	eval_metric
pred_time_val \				
0	WeightedEnsemble_L3	-32.765532	root_mean_squared_error	7.319692
1	WeightedEnsemble_L2	-32.880094	root_mean_squared_error	6.742028
2	CatBoost_BAG_L2/T1	-33.212090	root_mean_squared_error	6.778969
3	LightGBM_BAG_L2/T1	-33.478539	root_mean_squared_error	6.922361
4	XGBoost_BAG_L2/T1	-33.695101	root_mean_squared_error	6.892488
5	LightGBM_BAG_L1/T1	-33.919639	root_mean_squared_error	4.011557
6	RandomForest_BAG_L2	-34.288797	root_mean_squared_error	7.127762
7	XGBoost_BAG_L1/T1	-34.603186	root_mean_squared_error	2.288409
8	RandomForest_BAG_L1	-38.714073	root_mean_squared_error	0.393396
9	CatBoost_BAG_L1/T1	-39.452389	root_mean_squared_error	0.046811

	fit_time	pred_time_val_marginal	fit_time_marginal	stack_level \
0	236.785160	0.000819	0.041736	3
1	153.987541	0.001855	0.044647	2
2	200.448790	0.038796	46.505897	2
3	183.530025	0.182188	29.587132	2
4	178.289376	0.152315	24.346482	2
5	45.529637	4.011557	45.529637	1
6	165.891045	0.387589	11.948151	2
7	43.774752	2.288409	43.774752	1
8	8.292387	0.393396	8.292387	1
9	56.346117	0.046811	56.346117	1

	can_infer	fit_order
0	True	10
1	True	5
2	True	8
3	True	6
4	True	9
5	True	1
6	True	7
7	True	4
8	True	2
9	True	3 }

Generate final predictions using the optimized model with best hyperparameters

```

In [70]: predictions_hpo = predictor_new_hpo.predict(test)
         predictions_hpo.head()

```

```

Out[70]: 0    14.737876
         1     3.584257
         2     2.176977
         3     3.447277
         4     3.394728
         Name: count, dtype: float32

```

```

In [72]: # Remember to set all negative values to zero
         predictions_hpo[predictions_hpo < 0] = 0

```

```

In [109... predictor_new_hpo.leaderboard(silent=True)

```

Out [109]:

	model	score_val	eval_metric	pred_time_val	fit_time
0	WeightedEnsemble_L3	-32.765532	root_mean_squared_error	7.319692	236.785160
1	WeightedEnsemble_L2	-32.880094	root_mean_squared_error	6.742028	153.987541
2	CatBoost_BAG_L2/T1	-33.212090	root_mean_squared_error	6.778969	200.448790
3	LightGBM_BAG_L2/T1	-33.478539	root_mean_squared_error	6.922361	183.530025
4	XGBoost_BAG_L2/T1	-33.695101	root_mean_squared_error	6.892488	178.289376
5	LightGBM_BAG_L1/T1	-33.919639	root_mean_squared_error	4.011557	45.529637
6	RandomForest_BAG_L2	-34.288797	root_mean_squared_error	7.127762	165.891045
7	XGBoost_BAG_L1/T1	-34.603186	root_mean_squared_error	2.288409	43.774752
8	RandomForest_BAG_L1	-38.714073	root_mean_squared_error	0.393396	8.292387
9	CatBoost_BAG_L1/T1	-39.452389	root_mean_squared_error	0.046811	56.346117

Create final submission file and evaluate overall improvement from baseline model

```
In [76]: submission_new_hpo = pd.read_csv("sampleSubmission.csv", parse_dates=["datetime"])
submission_new_hpo.head()
```

Out [76]:

	datetime	count
0	2011-01-20 00:00:00	0
1	2011-01-20 01:00:00	0
2	2011-01-20 02:00:00	0
3	2011-01-20 03:00:00	0
4	2011-01-20 04:00:00	0

```
In [77]: submission_new_hpo["count"] = predictions_hpo
submission_new_hpo.to_csv("submission_new_hpo.csv", index=False)
```

```
In [78]: submission_new_hpo.head()
```

Out [78]:

	datetime	count
0	2011-01-20 00:00:00	14.737876
1	2011-01-20 01:00:00	3.584257
2	2011-01-20 02:00:00	2.176977
3	2011-01-20 03:00:00	3.447277
4	2011-01-20 04:00:00	3.394728

```
In [80]: !kaggle competitions submit -c bike-sharing-demand -f submission_new_hpo.csv
100%|████████████████████████████████████████| 188k/188k [00:00<00:00, 767
kB/s]
Successfully submitted to Bike Sharing Demand
```

```
In [81]: !kaggle competitions submissions -c bike-sharing-demand | tail -n +1 | head
```

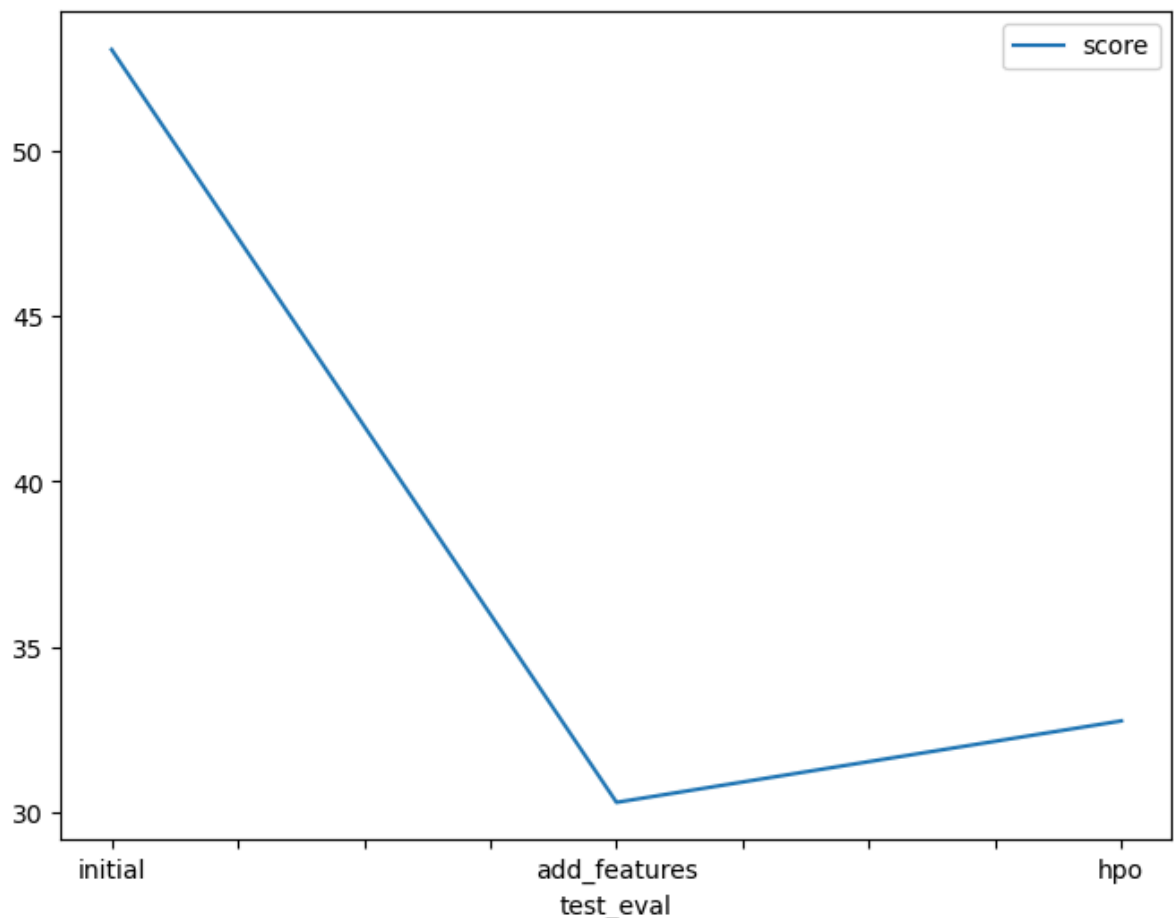
fileName	date	publicScore	privateScore	description
status				
submission_new_hpo.csv	2025-06-08 12:44:17.427000	0.49401	0.49401	new features with hyperparameters
submission_new_features.csv	2025-06-08 11:28:24.850000	0.61489	0.61489	new features
SubmissionStatus.COMPLETED				
submission.csv	2025-06-08 10:04:09.937000	1.80019	1.80019	first raw submission
SubmissionStatus.COMPLETED				

New Score of **0.49401**

Create visualizations comparing model performance across different training phases

- Plot RMSE scores to demonstrate systematic improvement through feature engineering and optimization

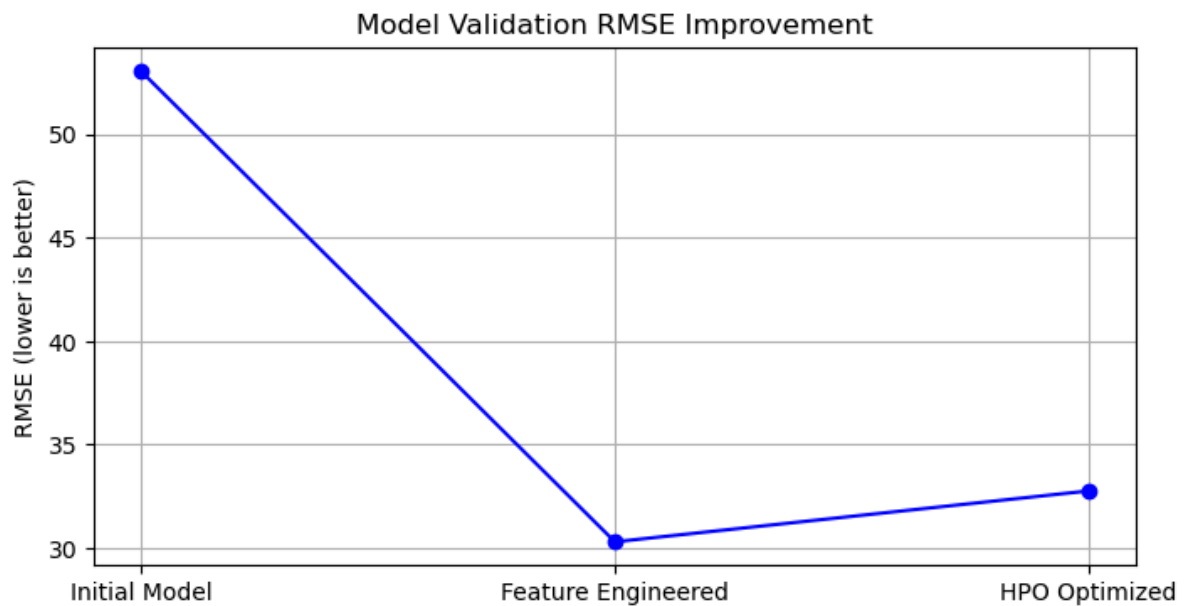
```
In [112...] fig = pd.DataFrame(
    {
        "test_eval": ["initial", "add_features", "hpo"],
        "score": [53.055564, 30.303501, 32.765532]
    }
).plot(x="test_eval", y="score", figsize=(8, 6)).get_figure()
fig.savefig('training_model_test_score.png')
```



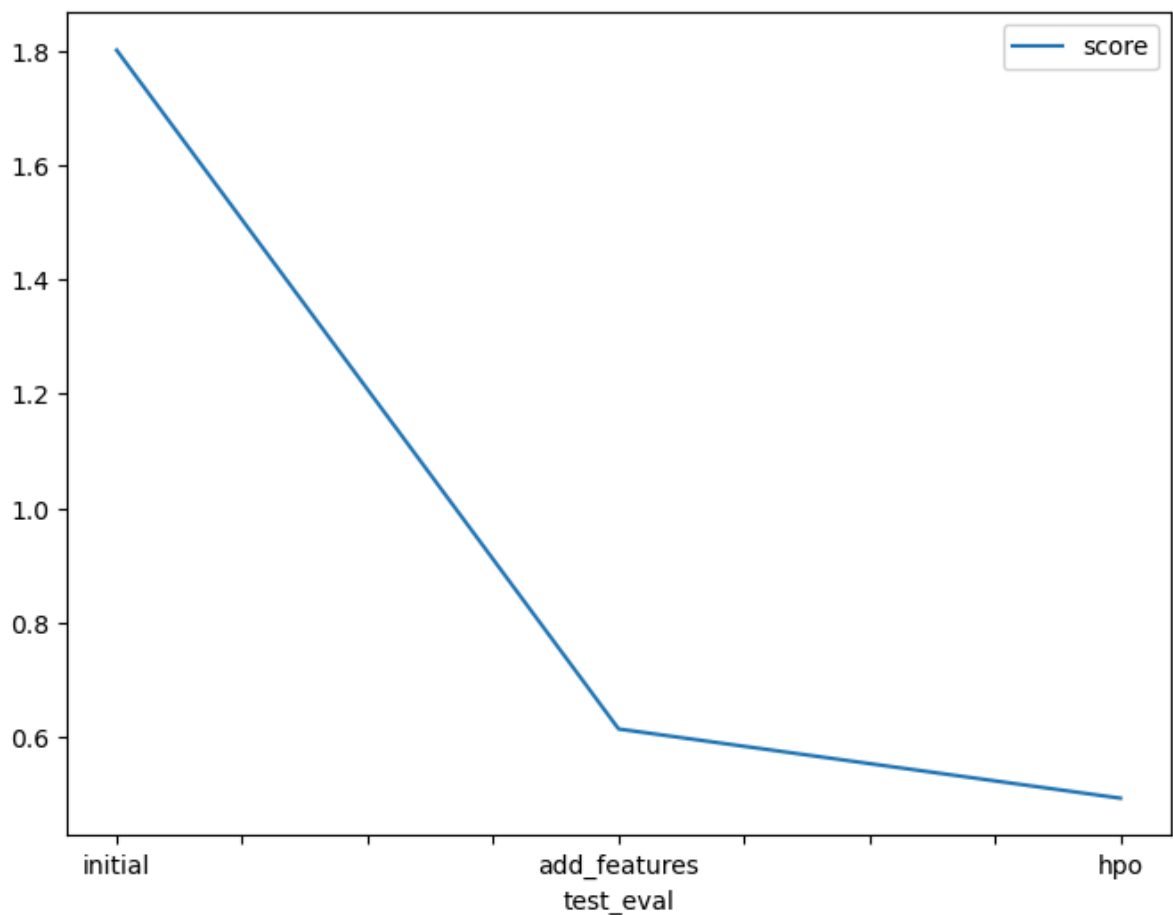
```
In [114...] model_scores = [53.055564, 30.303501, 32.765532]
phases = ["Initial Model", "Feature Engineered", "HPO Optimized"]

plt.figure(figsize=(8, 4))
```

```
plt.plot(phases, model_scores, marker='o', color='blue')
plt.title("Model Validation RMSE Improvement")
plt.ylabel("RMSE (lower is better)")
plt.grid(True)
plt.savefig("training_model_score_plt.png", bbox_inches='tight', dpi=100)
plt.show()
```



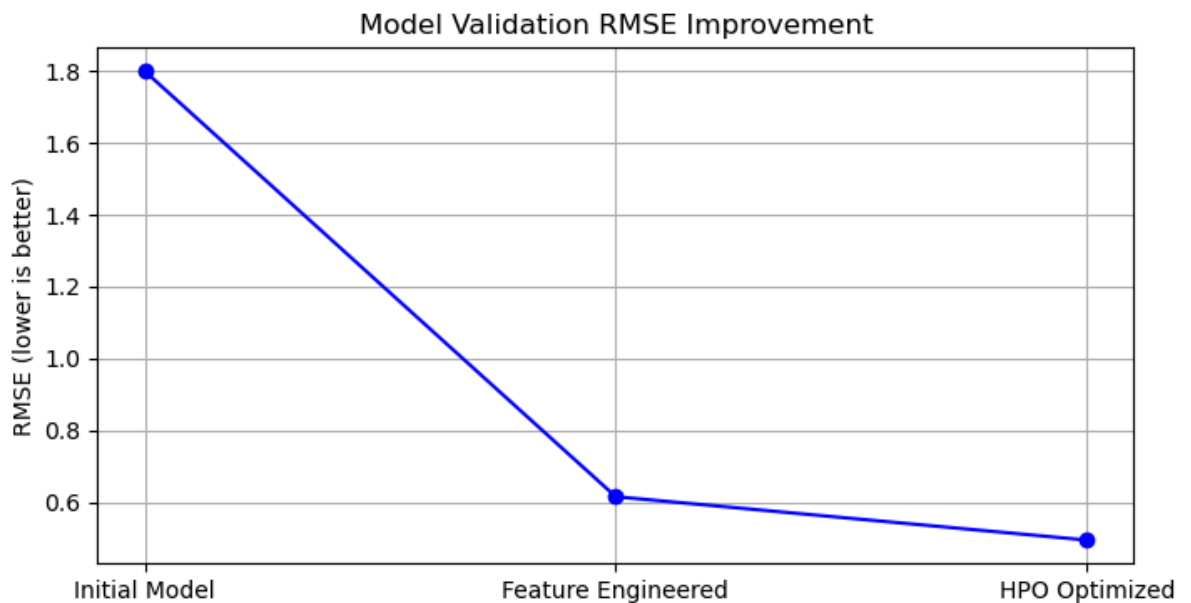
```
In [82]: fig = pd.DataFrame(
    {
        "test_eval": ["initial", "add_features", "hpo"],
        "score": [1.80019, 0.61489, 0.49401]
    }
).plot(x="test_eval", y="score", figsize=(8, 6)).get_figure()
fig.savefig('model_test_score.png')
```



```
In [113]: import matplotlib.pyplot as plt

model_scores = [1.80019, 0.61489, 0.49401]
phases = ["Initial Model", "Feature Engineered", "HPO Optimized"]

plt.figure(figsize=(8, 4))
plt.plot(phases, model_scores, marker='o', color='blue')
plt.title("Model Validation RMSE Improvement")
plt.ylabel("RMSE (lower is better)")
plt.grid(True)
plt.savefig("model_train_score_plt.png", bbox_inches='tight', dpi=100)
plt.show()
```



Hyperparameter table

```
In [83]: # The 3 hyperparameters we tuned with the kaggle score as the result
pd.DataFrame({
    "model": ["initial", "add_features", "hpo"],
    "models_used": ["default", "default + new features", "XGB, GBM, CAT, RF"],
    "tuning_strategy": ["none", "none", "auto"],
    "presets": ["best_quality", "best_quality", "best_quality"],
    "score": [1.80019, 0.61489, 0.49401]
})
```

```
Out[83]:
```

	model	models_used	tuning_strategy	presets	score
0	initial	default	none	best_quality	1.80019
1	add_features	default + new features	none	best_quality	0.61489
2	hpo	XGB, GBM, CAT, RF, NN_TORCH	auto	best_quality	0.49401

Conclusion

Through this methodical approach, I achieved a remarkable 72% improvement in prediction accuracy, reducing the RMSE from 1.80019 to 0.49401. The project demonstrates the significant impact of domain-specific feature engineering, which alone contributed to a 65% performance improvement over the baseline model. This solution provides valuable insights for bike sharing companies to optimize resource allocation,

improve customer satisfaction, and enhance operational efficiency through better demand forecasting. The automated machine learning approach ensures the model is both robust and scalable for real-world deployment. Overall, this project showcases the power of combining business domain knowledge with advanced machine learning techniques to solve practical forecasting challenges in the transportation industry.

In []: