

Juncos Week 5 Assignment

DS Automation Assignment

Using our prepared churn data from week 2:

- use pycaret to find an ML algorithm that performs best on the data
 - Choose a metric you think is best to use for finding the best model; by default, it is accuracy but it could be AUC, precision, recall, etc. The week 3 FTE has some information on these different metrics.
- save the model to disk
- create a Python script/file/module with a function that takes a pandas dataframe as an input and returns the probability of churn for each row in the dataframe
 - your Python file/function should print out the predictions for new data (new_churn_data.csv)
 - the true values for the new data are [1, 0, 0, 1, 0] if you're interested
- test your Python module and function with the new data, new_churn_data.csv
- write a short summary of the process and results at the end of this notebook
- upload this Jupyter Notebook and Python file to a Github repository, and turn in a link to the repository in the week 5 assignment dropbox

Optional challenges:

- return the probability of churn for each new prediction, and the percentile where that prediction is in the distribution of probability predictions from the training dataset (e.g. a high probability of churn like 0.78 might be at the 90th percentile)
- use other autoML packages, such as TPOT, H2O, MLBox, etc, and compare performance and features with pycaret
- create a class in your Python module to hold the functions that you created
- accept user input to specify a file using a tool such as Python's `input()` function, the `click` package for command-line arguments, or a GUI
- Use the unmodified churn data (new_unmodified_churn_data.csv) in your Python script. This will require adding the same preprocessing steps from week 2 since this data is like the original unmodified dataset from week 1.

```
In [ ]: %pip install pycaret
```

```
Collecting pycaret
  Downloading pycaret-2.3.10-py3-none-any.whl (320 kB)
    |████████████████████████████████████████| 320 kB 2.2 MB/s
Collecting imbalanced-learn==0.7.0
  Downloading imbalanced_learn-0.7.0-py3-none-any.whl (167 kB)
    |████████████████████████████████████████| 167 kB 30.8 MB/s
Collecting pyod
  Downloading pyod-1.0.1.tar.gz (120 kB)
    |████████████████████████████████████████| 120 kB 20.2 MB/s
  Preparing metadata (setup.py) ... done
Requirement already satisfied: pandas-profiling>=2.8.0 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from pycaret) (3.2.0)
Requirement already satisfied: pandas in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from pycaret) (1.4.2)
Collecting pyLDAvis
  Downloading pyLDAvis-3.3.1.tar.gz (1.7 MB)
    |████████████████████████████████████████| 1.7 MB 17.1 MB/s
  Installing build dependencies ... done
  Getting requirements to build wheel ... done
  Installing backend dependencies ... done
  Preparing metadata (pyproject.toml) ... done
Requirement already satisfied: seaborn in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from pycaret) (0.11.2)
Requirement already satisfied: scikit-plot in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from pycaret) (0.3.7)
Collecting Boruta
  Downloading Boruta-0.3-py3-none-any.whl (56 kB)
    |████████████████████████████████████████| 56 kB 7.4 MB/s
Collecting gensim<4.0.0
  Downloading gensim-3.8.3.tar.gz (23.4 MB)
    |████████████████████████████████████████| 23.4 MB 413 kB/s
  Preparing metadata (setup.py) ... done
Collecting nltk
  Downloading nltk-3.7-py3-none-any.whl (1.5 MB)
    |████████████████████████████████████████| 1.5 MB 18.6 MB/s
Collecting mlxtend>=0.17.0
  Downloading mlxtend-0.20.0-py2.py3-none-any.whl (1.3 MB)
    |████████████████████████████████████████| 1.3 MB 25.6 MB/s
Collecting lightgbm>=2.3.1
  Downloading lightgbm-3.3.2-py3-none-macosx_10_14_x86_64.macosx_10_15_x86_64.macosx_11_0_x86_64.whl (1.2 MB)
    |████████████████████████████████████████| 1.2 MB 36.2 MB/s
Collecting cufflinks>=0.17.0
  Downloading cufflinks-0.17.3.tar.gz (81 kB)
    |████████████████████████████████████████| 81 kB 13.3 MB/s
  Preparing metadata (setup.py) ... done
Requirement already satisfied: matplotlib in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from pycaret) (3.5.2)
Collecting spacy<2.4.0
  Downloading spacy-2.3.7-cp39-cp39-macosx_10_9_x86_64.whl (10.1 MB)
    |████████████████████████████████████████| 10.1 MB 54.7 MB/s
Requirement already satisfied: joblib in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from pycaret) (1.1.0)
Collecting umap-learn
  Downloading umap-learn-0.5.3.tar.gz (88 kB)
    |████████████████████████████████████████| 88 kB 14.4 MB/s
  Preparing metadata (setup.py) ... done
Collecting kmodes>=0.10.1
  Downloading kmodes-0.12.1-py2.py3-none-any.whl (20 kB)
```

```

Collecting scipy<=1.5.4
  Downloading scipy-1.5.4-cp39-cp39-macosx_10_9_x86_64.whl (29.1 MB)
    |████████████████████████████████████████| 29.1 MB 14.0 MB/s
Collecting mlflow
  Downloading mlflow-1.26.1-py3-none-any.whl (17.8 MB)
    |████████████████████████████████████████| 17.8 MB 34.6 MB/s
Collecting wordcloud
  Downloading wordcloud-1.8.1.tar.gz (220 kB)
    |████████████████████████████████████████| 220 kB 30.7 MB/s
  Preparing metadata (setup.py) ... done
Collecting plotly>=4.4.1
  Downloading plotly-5.8.0-py2.py3-none-any.whl (15.2 MB)
    |████████████████████████████████████████| 15.2 MB 27.8 MB/s
Requirement already satisfied: IPython in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from pycaret) (8.3.0)
Collecting numba<0.55
  Downloading numba-0.54.1-cp39-cp39-macosx_10_14_x86_64.whl (2.3 MB)
    |████████████████████████████████████████| 2.3 MB 48.8 MB/s
Collecting textblob
  Downloading textblob-0.17.1-py2.py3-none-any.whl (636 kB)
    |████████████████████████████████████████| 636 kB 19.3 MB/s
Collecting pyyaml<6.0.0
  Downloading PyYAML-5.4.1-cp39-cp39-macosx_10_9_x86_64.whl (259 kB)
    |████████████████████████████████████████| 259 kB 31.5 MB/s
Collecting yellowbrick>=1.0.1
  Downloading yellowbrick-1.4-py3-none-any.whl (274 kB)
    |████████████████████████████████████████| 274 kB 29.3 MB/s
Requirement already satisfied: ipywidgets in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from pycaret) (7.7.0)
Collecting scikit-learn==0.23.2
  Downloading scikit-learn-0.23.2.tar.gz (7.2 MB)
    |████████████████████████████████████████| 7.2 MB 16.5 MB/s
  Installing build dependencies ... done
  Getting requirements to build wheel ... done
  Preparing metadata (pyproject.toml) ... done
Requirement already satisfied: numpy>=1.13.3 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from imbalanced-learn==0.7.0->pycaret) (1.20.2)
Requirement already satisfied: threadpoolctl>=2.0.0 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from scikit-learn==0.23.2->pycaret) (3.1.0)
Requirement already satisfied: six>=1.9.0 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from cufflinks>=0.17.0->pycaret) (1.16.0)
Collecting colorlover>=0.2.1
  Downloading colorlover-0.3.0-py3-none-any.whl (8.9 kB)
Requirement already satisfied: setuptools>=34.4.1 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from cufflinks>=0.17.0->pycaret) (58.1.0)
Collecting smart_open>=1.8.1
  Downloading smart_open-6.0.0-py3-none-any.whl (58 kB)
    |████████████████████████████████████████| 58 kB 14.1 MB/s
Requirement already satisfied: prompt-toolkit!=3.0.0,!<3.0.1,<3.1.0,>=2.0.0 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from IPython->pycaret) (3.0.29)
Requirement already satisfied: stack-data in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from IPython->pycaret) (0.2.0)
Requirement already satisfied: pygments>=2.4.0 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from IPython->pycaret) (2.12.0)

```

```

Requirement already satisfied: pexpect>4.3 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from IPython->pycaret) (4.8.0)
Requirement already satisfied: appnope in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from IPython->pycaret) (0.1.3)
Requirement already satisfied: backcall in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from IPython->pycaret) (0.2.0)
Requirement already satisfied: traitlets>=5 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from IPython->pycaret) (5.1.1)
Requirement already satisfied: pickleshare in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from IPython->pycaret) (0.7.5)
Requirement already satisfied: matplotlib-inline in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from IPython->pycaret) (0.1.3)
Requirement already satisfied: decorator in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from IPython->pycaret) (5.1.1)
Requirement already satisfied: jedi>=0.16 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from IPython->pycaret) (0.18.1)
Requirement already satisfied: ipython-genutils~=0.2.0 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from ipywidgets->pycaret) (0.2.0)
Requirement already satisfied: widgetsnbextension~=3.6.0 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from ipywidgets->pycaret) (3.6.0)
Requirement already satisfied: jupyterlab-widgets>=1.0.0 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from ipywidgets->pycaret) (1.1.0)
Requirement already satisfied: nbformat>=4.2.0 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from ipywidgets->pycaret) (5.4.0)
Requirement already satisfied: ipykernel>=4.5.1 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from ipywidgets->pycaret) (6.13.0)
Collecting wheel
  Using cached wheel-0.37.1-py2.py3-none-any.whl (35 kB)
Collecting mlxtend>=0.17.0
  Downloading mlxtend-0.19.0-py2.py3-none-any.whl (1.3 MB)
    |████████████████████████████████████████| 1.3 MB 2.0 MB/s
Requirement already satisfied: python-dateutil>=2.7 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from matplotlib->pycaret) (2.8.2)
Requirement already satisfied: kiwisolver>=1.0.1 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from matplotlib->pycaret) (1.4.2)
Requirement already satisfied: packaging>=20.0 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from matplotlib->pycaret) (21.3)
Requirement already satisfied: cycycler>=0.10 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from matplotlib->pycaret) (0.11.0)
Requirement already satisfied: pyparsing>=2.2.1 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from matplotlib->pycaret) (3.0.8)
Requirement already satisfied: fonttools>=4.22.0 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from matplotlib->pycaret) (4.33.3)
Requirement already satisfied: pillow>=6.2.0 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from matplotlib->pycaret) (9.1.0)
Collecting llvmlite<0.38,>=0.37.0rc1
  Downloading llvmlite-0.37.0-cp39-cp39-macosx_10_9_x86_64.whl (19.1 MB)

```

5/18

```

Collecting cymem<2.1.0,>=2.0.2
  Downloading cymem-2.0.6-cp39-cp39-macosx_10_9_x86_64.whl (32 kB)
Collecting catalogue<1.1.0,>=0.0.7
  Downloading catalogue-1.0.0-py2.py3-none-any.whl (7.7 kB)
Collecting wasabi<1.1.0,>=0.4.0
  Downloading wasabi-0.9.1-py3-none-any.whl (26 kB)
Collecting preshed<3.1.0,>=3.0.2
  Downloading preshed-3.0.6-cp39-cp39-macosx_10_9_x86_64.whl (106 kB)
|████████████████████████████████████████| 106 kB 27.2 MB/s
Collecting yellowbrick>=1.0.1
  Downloading yellowbrick-1.3.post1-py3-none-any.whl (271 kB)
|████████████████████████████████████████| 271 kB 48.8 MB/s
Collecting numpy>=1.13.3
  Downloading numpy-1.19.5-cp39-cp39-macosx_10_9_x86_64.whl (15.6 MB)
|████████████████████████████████████████| 15.6 MB 11.7 MB/s
Collecting protobuf>=3.12.0
  Downloading protobuf-4.21.1-cp37-abi3-macosx_10_9_universal2.whl (483 kB)
|████████████████████████████████████████| 483 kB 7.1 MB/s
Requirement already satisfied: click>=7.0 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from mlflow->pycaret) (8.0.1)
Requirement already satisfied: sqlparse>=0.3.1 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from mlflow->pycaret) (0.4.1)
Requirement already satisfied: entrypoints in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from mlflow->pycaret) (0.4)
Collecting importlib-metadata!=4.7.0,>=3.7.0
  Downloading importlib_metadata-4.11.4-py3-none-any.whl (18 kB)
Collecting docker>=4.0.0
  Using cached docker-5.0.3-py2.py3-none-any.whl (146 kB)
Collecting cloudpickle
  Downloading cloudpickle-2.1.0-py3-none-any.whl (25 kB)
Collecting alembic
  Downloading alembic-1.8.0-py3-none-any.whl (209 kB)
|████████████████████████████████████████| 209 kB 11.5 MB/s
Collecting databricks-cli>=0.8.7
  Downloading databricks-cli-0.16.6.tar.gz (62 kB)
|████████████████████████████████████████| 62 kB 2.7 MB/s
  Preparing metadata (setup.py) ... done
Collecting sqlalchemy
  Downloading SQLAlchemy-1.4.37-cp39-cp39-macosx_10_15_x86_64.whl (1.5 MB)
|████████████████████████████████████████| 1.5 MB 51.1 MB/s
Collecting gunicorn
  Using cached gunicorn-20.1.0-py3-none-any.whl (79 kB)
Collecting gitpython>=2.1.0
  Downloading GitPython-3.1.27-py3-none-any.whl (181 kB)
|████████████████████████████████████████| 181 kB 41.8 MB/s
Collecting Flask
  Downloading Flask-2.1.2-py3-none-any.whl (95 kB)
|████████████████████████████████████████| 95 kB 12.9 MB/s
Collecting prometheus-flask-exporter
  Downloading prometheus_flask_exporter-0.20.2-py3-none-any.whl (18 kB)
Collecting querystring-parser
  Downloading querystring_parser-1.2.4-py2.py3-none-any.whl (7.9 kB)
Collecting regex>=2021.8.3
  Downloading regex-2022.6.2-cp39-cp39-macosx_10_9_x86_64.whl (288 kB)
|████████████████████████████████████████| 288 kB 7.8 MB/s
Requirement already satisfied: sklearn in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from pyLDAvis->pycaret) (0.0)
Collecting future
  Downloading future-0.18.2.tar.gz (829 kB)

```

7/18

```

Requirement already satisfied: jsonschema>=2.6 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from nbformat>=4.2.0->ipywidgets->pycaret) (4.5.1)
Requirement already satisfied: fastjsonschema in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from nbformat>=4.2.0->ipywidgets->pycaret) (2.15.3)
Requirement already satisfied: ptyprocess>=0.5 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from pexpect>4.3->IPython->pycaret) (0.7.0)
Requirement already satisfied: wcwidth in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from prompt-toolkit!=3.0.0,!<3.0.1,<3.1.0,>=2.0.0->IPython->pycaret) (0.2.5)
Requirement already satisfied: typing-extensions>=3.7.4.3 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from pydantic>=1.8.1->pandas-profiling>=2.8.0->pycaret) (4.2.0)
Requirement already satisfied: idna<3,>=2.5 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from requests>=2.24.0->pandas-profiling>=2.8.0->pycaret) (2.10)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from requests>=2.24.0->pandas-profiling>=2.8.0->pycaret) (1.26.4)
Requirement already satisfied: certifi>=2017.4.17 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from requests>=2.24.0->pandas-profiling>=2.8.0->pycaret) (2020.12.5)
Requirement already satisfied: chardet<5,>=3.0.2 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from requests>=2.24.0->pandas-profiling>=2.8.0->pycaret) (4.0.0)
Requirement already satisfied: notebook>=4.4.1 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from widgetsnbextension~=3.6.0->ipywidgets->pycaret) (6.4.11)
Collecting Mako
  Downloading Mako-1.2.0-py3-none-any.whl (78 kB)
    |████████████████████████████████████████| 78 kB 12.2 MB/s
Collecting greenlet!=0.4.17
  Downloading greenlet-1.1.2-cp39-cp39-macosx_10_14_x86_64.whl (92 kB)
    |████████████████████████████████████████| 92 kB 2.3 MB/s
Collecting Werkzeug>=2.0
  Downloading Werkzeug-2.1.2-py3-none-any.whl (224 kB)
    |████████████████████████████████████████| 224 kB 26.0 MB/s
Collecting itsdangerous>=2.0
  Downloading itsdangerous-2.1.2-py3-none-any.whl (15 kB)
Requirement already satisfied: prometheus-client in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from prometheus-flask-exporter->mlflow->pycaret) (0.14.1)
Requirement already satisfied: pure-eval in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from stack-data->IPython->pycaret) (0.2.2)
Requirement already satisfied: asttokens in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from stack-data->IPython->pycaret) (2.0.5)
Requirement already satisfied: executing in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from stack-data->IPython->pycaret) (0.8.3)
Collecting patsy>=0.5.2
  Downloading patsy-0.5.2-py2.py3-none-any.whl (233 kB)
    |████████████████████████████████████████| 233 kB 13.8 MB/s
Collecting smmap<6,>=3.0.1
  Downloading smmap-5.0.0-py3-none-any.whl (24 kB)
Requirement already satisfied: pyrsistent!=0.17.0,!<0.17.1,!<0.17.2,>=0.14.0 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-package

```


s (from jsonschema>=2.6->nbformat>=4.2.0->ipywidgets->pycaret) (0.18.1)

Requirement already satisfied: pyzmq>=22.3 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from jupyter-client>=6.1.12->ipykernel>=4.5.1->ipywidgets->pycaret) (22.3.0)

Requirement already satisfied: argon2-cffi in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from notebook>=4.4.1->widgetsnbextension~3.6.0->ipywidgets->pycaret) (21.3.0)

Requirement already satisfied: Send2Trash>=1.8.0 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from notebook>=4.4.1->widgetsnbextension~3.6.0->ipywidgets->pycaret) (1.8.0)

Requirement already satisfied: nbconvert>=5 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from notebook>=4.4.1->widgetsnbextension~3.6.0->ipywidgets->pycaret) (6.5.0)

Requirement already satisfied: terminado>=0.8.3 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from notebook>=4.4.1->widgetsnbextension~3.6.0->ipywidgets->pycaret) (0.13.3)

Requirement already satisfied: PyWavelets in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from imagehash->visions[type_image_path]==0.7.4->pandas-profiling>=2.8.0->pycaret) (1.3.0)

Requirement already satisfied: nbclient>=0.5.0 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~3.6.0->ipywidgets->pycaret) (0.6.3)

Requirement already satisfied: bleach in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~3.6.0->ipywidgets->pycaret) (5.0.0)

Requirement already satisfied: beautifulsoup4 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~3.6.0->ipywidgets->pycaret) (4.11.1)

Requirement already satisfied: mistune<2,>=0.8.1 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~3.6.0->ipywidgets->pycaret) (0.8.4)

Requirement already satisfied: jupyterlab-pygments in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~3.6.0->ipywidgets->pycaret) (0.2.2)

Requirement already satisfied: pandocfilters>=1.4.1 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~3.6.0->ipywidgets->pycaret) (1.5.0)

Requirement already satisfied: tinycss2 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~3.6.0->ipywidgets->pycaret) (1.1.1)

Requirement already satisfied: defusedxml in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~3.6.0->ipywidgets->pycaret) (0.7.1)

Requirement already satisfied: argon2-cffi-bindings in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from argon2-cffi->notebook>=4.4.1->widgetsnbextension~3.6.0->ipywidgets->pycaret) (21.2.0)

Requirement already satisfied: cffi>=1.0.1 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from argon2-cffi-bindings->argon2-cffi->notebook>=4.4.1->widgetsnbextension~3.6.0->ipywidgets->pycaret) (1.15.0)

Requirement already satisfied: soupsieve>1.2 in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from beautifulsoup4->nbconvert>=5->notebook>=4.4.1->widgetsnbextension~3.6.0->ipywidgets->pycaret) (2.3.2.post1)

Requirement already satisfied: webencodings in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from bleach->nbconvert>=5->notebook>=4.4.1->widgetsnbextension~3.6.0->ipywidgets->pycaret) (0.5.1)

Requirement already satisfied: pyparser in /Library/Frameworks/Python.framework/Versions/3.9/lib/python3.9/site-packages (from cffi>=1.0.1->argon2-cffi-bindings->argon2-cffi->notebook>=4.4.1->widgetsnbextension~3.6.0->ipywidgets->py

```

caret) (2.21)
Using legacy 'setup.py install' for cufflinks, since package 'wheel' is not in
stalled.
Using legacy 'setup.py install' for gensim, since package 'wheel' is not insta
lled.
Using legacy 'setup.py install' for pyLDAvis, since package 'wheel' is not ins
talled.
Using legacy 'setup.py install' for pyod, since package 'wheel' is not install
ed.
Using legacy 'setup.py install' for umap-learn, since package 'wheel' is not i
nstalled.
Using legacy 'setup.py install' for wordcloud, since package 'wheel' is not in
stalled.
Using legacy 'setup.py install' for databricks-cli, since package 'wheel' is n
ot installed.
Using legacy 'setup.py install' for pynndescent, since package 'wheel' is not
installed.
Using legacy 'setup.py install' for future, since package 'wheel' is not insta
lled.
Building wheels for collected packages: scikit-learn
  Building wheel for scikit-learn (pyproject.toml) ... done
  Created wheel for scikit-learn: filename=scikit_learn-0.23.2-cp39-cp39-macos
x_12_0_universal2.whl size=10657117 sha256=57f02842e7d1b216f0d67645e3660744791
28c7a67a069621720b45a2d6cdd2d
  Stored in directory: /Users/njuncos/Library/Caches/pip/wheels/5e/74/24/7e235
ccf01765c0daa089c98cc823e9dc1383da5fe0ed7e224
Successfully built scikit-learn
Installing collected packages: zipp, numpy, Werkzeug, smmap, scipy, murmurhas
h, llvmlite, itsdangerous, importlib-metadata, greenlet, cymem, websocket-clie
nt, wasabi, tenacity, tabulate, srsly, sqlalchemy, scikit-learn, regex, pyjwt,
preshed, plac, patsy, oauthlib, numba, Mako, gitdb, Flask, catalogue, blis, wh
eel, thinc, statsmodels, smart-open, querystring-parser, pyyaml, pynndescent,
protobuf, prometheus-flask-exporter, plotly, numexpr, nltk, gunicorn, gitpytho
n, future, fancy, docker, databricks-cli, colorlover, cloudpickle, alembic, ye
llowbrick, wordcloud, umap-learn, textblob, spacy, pyod, pyLDAvis, mlxtend, ml
flow, lightgbm, kmodes, imbalanced-learn, gensim, cufflinks, Boruta, pycaret
  Attempting uninstall: numpy
    Found existing installation: numpy 1.20.2
    Uninstalling numpy-1.20.2:
      Successfully uninstalled numpy-1.20.2
  Attempting uninstall: scipy
    Found existing installation: scipy 1.8.0
    Uninstalling scipy-1.8.0:
      Successfully uninstalled scipy-1.8.0
  Attempting uninstall: scikit-learn
    Found existing installation: scikit-learn 1.1.0
    Uninstalling scikit-learn-1.1.0:
      Successfully uninstalled scikit-learn-1.1.0
  Attempting uninstall: pyyaml
    Found existing installation: PyYAML 6.0
    Uninstalling PyYAML-6.0:
      Successfully uninstalled PyYAML-6.0
  Running setup.py install for pynndescent ... done
  Running setup.py install for future ... done
  Running setup.py install for databricks-cli ... done
  Running setup.py install for wordcloud ... done
  Running setup.py install for umap-learn ... done
  Running setup.py install for pyod ... done
  Running setup.py install for pyLDAvis ... done
  Running setup.py install for gensim ... done

```

```
Running setup.py install for cufflinks ... done
Successfully installed Boruta-0.3 Flask-2.1.2 Mako-1.2.0 Werkzeug-2.1.2 alembic-1.8.0 blis-0.7.7 catalogue-1.0.0 cloudpickle-2.1.0 colorlover-0.3.0 cufflinks-0.17.3 cymem-2.0.6 databricks-cli-0.16.6 docker-5.0.3 funcy-1.17 future-0.18.2 gensim-3.8.3 gitdb-4.0.9 gitpython-3.1.27 greenlet-1.1.2 gunicorn-20.1.0 imbalanced-learn-0.7.0 importlib-metadata-4.11.4 itsdangerous-2.1.2 kmodes-0.12.1 lightgbm-3.3.2 llvmlite-0.37.0 mlflow-1.26.1 mlxtend-0.19.0 murmurhash-1.0.7 nltk-3.7 numba-0.54.1 numexpr-2.8.1 numpy-1.19.5 oauthlib-3.2.0 patsy-0.5.2 plac-1.1.3 plotly-5.8.0 preshed-3.0.6 prometheus-flask-exporter-0.20.2 protobuf-4.21.1 pyLDAvis-3.2.2 pycaret-2.3.10 pyjwt-2.4.0 pynndescent-0.5.7 pyod-1.0.1 pyyaml-5.4.1 querystring-parser-1.2.4 regex-2022.6.2 scikit-learn-0.23.2 scipy-1.5.4 smart-open-6.0.0 smmap-5.0.0 spacy-2.3.7 sqlalchemy-1.4.37 srsly-1.0.5 statsmodels-0.13.2 tabulate-0.8.9 tenacity-8.0.1 textblob-0.17.1 thinc-7.4.5 umap-learn-0.5.3 wasabi-0.9.1 websocket-client-1.3.2 wheel-0.37.1 wordcloud-1.8.1 yellowbrick-1.3.post1 zipp-3.8.0
WARNING: You are using pip version 21.3.1; however, version 22.1.2 is available.
You should consider upgrading via the '/Library/Frameworks/Python.framework/Versions/3.9/bin/python3 -m pip install --upgrade pip' command.
Note: you may need to restart the kernel to use updated packages.
```

```
In [ ]: import pandas as pd
```

```
In [ ]: # import data where I replaced TotalCharges and tenure null/0 values to median
df = pd.read_csv('prepped_churn_data2.csv', index_col='customerID')
df
```

Out []:

	tenure	PhoneService	Contract	PaymentMethod	MonthlyCharges	TotalCharges
customerID						
7590-VHVEG	1	0	0	0	29.85	29.85
5575-GNVDE	34	1	1	1	56.95	1889.50
3668-QPYBK	2	1	0	1	53.85	108.15
7795-CFOCW	45	0	1	2	42.30	1840.75
9237-HQITU	2	1	0	0	70.70	151.65
...
6840-RESVB	24	1	1	1	84.80	1990.50
2234-XADUH	72	1	1	3	103.20	7362.90
4801-JZAZL	11	0	0	0	29.60	346.45
8361-LTMKD	4	1	0	1	74.40	306.60
3186-AJIEK	66	1	2	2	105.65	6844.50

7043 rows × 8 columns

Use pycaret to find an ML algorithm that performs best on the data

```
In [ ]: # import pycaret libraries
from pycaret.classification import setup, compare_models, predict_model, save_n
```

```
In [ ]: automl = setup(df, target='Churn')
```

	Description	Value
0	session_id	5900
1	Target	Churn
2	Target Type	Binary
3	Label Encoded	None
4	Original Data	(7043, 8)
5	Missing Values	False
6	Numeric Features	4
7	Categorical Features	3
8	Ordinal Features	False
9	High Cardinality Features	False
10	High Cardinality Method	None
11	Transformed Train Set	(4930, 11)
12	Transformed Test Set	(2113, 11)
13	Shuffle Train-Test	True
14	Stratify Train-Test	False
15	Fold Generator	StratifiedKFold
16	Fold Number	10
17	CPU Jobs	-1
18	Use GPU	False
19	Log Experiment	False
20	Experiment Name	clf-default-name
21	USI	c987
22	Imputation Type	simple
23	Iterative Imputation Iteration	None
24	Numeric Imputer	mean
25	Iterative Imputation Numeric Model	None
26	Categorical Imputer	constant
27	Iterative Imputation Categorical Model	None
28	Unknown Categoricals Handling	least_frequent
29	Normalize	False
30	Normalize Method	None
31	Transformation	False
32	Transformation Method	None
33	PCA	False
34	PCA Method	None

	Description	Value
35	PCA Components	None
36	Ignore Low Variance	False
37	Combine Rare Levels	False
38	Rare Level Threshold	None
39	Numeric Binning	False
40	Remove Outliers	False
41	Outliers Threshold	None
42	Remove Multicollinearity	False
43	Multicollinearity Threshold	None
44	Remove Perfect Collinearity	True
45	Clustering	False
46	Clustering Iteration	None
47	Polynomial Features	False
48	Polynomial Degree	None
49	Trigonometry Features	False
50	Polynomial Threshold	None
51	Group Features	False
52	Feature Selection	False
53	Feature Selection Method	classic
54	Features Selection Threshold	None
55	Feature Interaction	False
56	Feature Ratio	False
57	Interaction Threshold	None
58	Fix Imbalance	False
59	Fix Imbalance Method	SMOTE

```
In [ ]: # Viewed multiple, but went with Precision. Also, choosing Gradient Boosting C
best_model = compare_models(sort='Precision')
```

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
ada	Ada Boost Classifier	0.7895	0.8331	0.4840	0.6374	0.5491	0.4154	0.4227	0.0530
gbc	Gradient Boosting Classifier	0.7899	0.8329	0.4878	0.6366	0.5516	0.4178	0.4245	0.0740
ridge	Ridge Classifier	0.7846	0.0000	0.4543	0.6324	0.5283	0.3937	0.4029	0.0100
lr	Logistic Regression	0.7888	0.8315	0.5114	0.6270	0.5627	0.4255	0.4298	0.8400
lightgbm	Light Gradient Boosting Machine	0.7848	0.8232	0.4970	0.6182	0.5506	0.4115	0.4159	0.0320
lda	Linear Discriminant Analysis	0.7832	0.8234	0.5114	0.6103	0.5559	0.4141	0.4173	0.0090
knn	K Neighbors Classifier	0.7629	0.7386	0.4268	0.5733	0.4891	0.3391	0.3455	0.2230
rf	Random Forest Classifier	0.7623	0.7918	0.4627	0.5657	0.5084	0.3539	0.3573	0.1210
et	Extra Trees Classifier	0.7519	0.7644	0.4787	0.5380	0.5063	0.3415	0.3427	0.1250
svm	SVM - Linear Kernel	0.6323	0.0000	0.6407	0.4846	0.4881	0.2625	0.2981	0.0120
dt	Decision Tree Classifier	0.7197	0.6452	0.4681	0.4715	0.4690	0.2789	0.2793	0.0120
nb	Naive Bayes	0.6819	0.8094	0.8369	0.4481	0.5835	0.3625	0.4091	0.0060
qda	Quadratic Discriminant Analysis	0.5215	0.5141	0.4980	0.2797	0.3419	0.0216	0.0368	0.0080
dummy	Dummy Classifier	0.7339	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0090

```
In [ ]: best_model
```

```
Out[ ]: AdaBoostClassifier(algorithm='SAMME.R', base_estimator=None, learning_rate=1.0,
                           n_estimators=50, random_state=5900)
```

Based on the info above, I will be going with the Gradient Boosting Classifier model as it frequently tops the others on most metrics.

```
In [ ]: predict_model(best_model, df.iloc[-2:-1])
```

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	Ada Boost Classifier	1.0000	0	1.0000	1.0000	1.0000	nan	0.0000

```
Out [ ]:      tenure  PhoneService  Contract  PaymentMethod  MonthlyCharges  TotalCharges
customerID
8361-      4            1            0            1            74.4            306.6
LTMKD
```

Save Model

```
In [ ]: save_model(best_model, 'GBC')

Transformation Pipeline and Model Successfully Saved
Out [ ]: (Pipeline(memory=None,
               steps=[('dtypes',
                       DataTypes_Auto_infer(categorical_features=[],
                                             display_types=True, features_todrop=[],
                                             id_columns=[],
                                             ml_usecase='classification',
                                             numerical_features=[], target='Churn',
                                             time_features=[])),
                       ('imputer',
                        Simple_Imputer(categorical_strategy='not_available',
                                         fill_value_categorical=None,
                                         fill_value_numerical=None,
                                         numeric_strate...
                       ('dummy', Dummify(target='Churn')),
                       ('fix_perfect', Remove_100(target='Churn')),
                       ('clean_names', Clean_Colum_Names()),
                       ('feature_select', 'passthrough'), ('fix_multi', 'passthrough'
h')),
               ('dfs', 'passthrough'), ('pca', 'passthrough'),
               ['trained_model',
                AdaBoostClassifier(algorithm='SAMME.R', base_estimator=None,
                                   learning_rate=1.0, n_estimators=50,
                                   random_state=5900)]],
               verbose=False),
          'GBC.pkl')
```

Python Module

Created a python module that imports new_churn_data.csv and runs the model against it to determine churn predictions. Note that I needed to change the field charge_per_tenure to totalcharge_tenure_ratio to match the previous data and model.

```
In [ ]: # Created python module
from IPython.display import Code

Code('predict_churn.py')
```



```

Out[ ]: import pandas as pd
        from pycaret.classification import predict_model, load_model

        def load_data(filepath):
            """
            Loads diabetes data into a DataFrame from a string filepath.
            """
            df = pd.read_csv(filepath, index_col='customerID')
            return df

        def make_predictions(df):
            """
            Uses the pycaret best model to make predictions on data in the df
            dataframe.
            """
            model = load_model('GBC')
            predictions = predict_model(model, data=df)
            predictions.rename({'Label': 'Churn_prediction'}, axis=1, inplace
            =True)
            predictions['Churn_prediction'].replace({1: 'Churn', 0: 'No churn'
            },
                                                    inplace=True)

            return predictions['Churn_prediction']

        if __name__ == "__main__":
            df = load_data('new_churn_data.csv')
            predictions = make_predictions(df)
            print('predictions:')
            print(predictions)

```

Now to run the python module.

```
In [ ]: %run predict_churn.py
```

```

Transformation Pipeline and Model Successfully Loaded
predictions:
customerID
9305-CKSKC      Churn
1452-KNGVK      No churn
6723-OKKJM      No churn
7832-POPKP      No churn
6348-TACGU      No churn
Name: Churn_prediction, dtype: object

```

The results returned by my model are [1, 0, 0, 0, 0]. The actual results are [1, 0, 0, 1, 0]. So there is some error with my model as there is one customer that will Churn, but my model does not catch it.

Summary

This weeks assignment was to utilize the past data we have and create Models for predictions out of them. I ended up deciding on the Gradient Boosting Classifier model due to it's high precision, accuracy, and the fact that it was the top of the charts across the board. Many times it would even be top in precision, however it currently does not show that.

I also saved this model and created a function that would be able to communicate with this model and compare new_churn_data with it. I had to alter one field label, but I managed to run predictions against the data, of which I found 80% accuracy as the model properly predicted 4/5 of the churn results. There was one missed churn customer. This is accurate with the numbers I have found in previous assignments as I frequently find ~80% accuracy with the models I have created in the past.