week5

September 25, 2021

1 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.

I ended up dropping tc_tenure_ratio because the new churn data does not have that in it, and it causes problems later on.

[1]:
```
import pandas as pd

df = pd.read_csv('data/even_better_new_churn_data.csv', index_col='customerID')
#df = pd.read_csv('data/test_data1.csv', index_col='customerID')
# removing this as the new test data does not have it
df = df = df.drop('tc_tenure_ratio', axis=1)
# Was hoping it was a column sort issue from stackoverflows I found, but it did␣
 ↪not seem to help with my errors later.
df = df.reindex(sorted(df.columns), axis=1)
df
```

[1]:
```
            Churn  Contract  MonthlyCharges  PaymentMethod  PhoneService  \
customerID
7590-VHVEG      0         0           29.85              0             0
5575-GNVDE      0         1           56.95              1             1
3668-QPYBK      1         0           53.85              1             1
7795-CFOCW      0         1           42.30              2             0
9237-HQITU      1         0           70.70              0             1
…           …       …              …              …            …
6840-RESVB      0         1           84.80              1             1
2234-XADUH      0         1          103.20              3             1
4801-JZAZL      0         0           29.60              0             0
8361-LTMKD      1         0           74.40              1             1
3186-AJIEK      0         2          105.65              2             1

            TotalCharges  tenure
customerID
7590-VHVEG         29.85       1
5575-GNVDE       1889.50      34
3668-QPYBK        108.15       2
7795-CFOCW       1840.75      45
9237-HQITU        151.65       2
…                    …     …
6840-RESVB       1990.50      24
2234-XADUH       7362.90      72
4801-JZAZL        346.45      11
8361-LTMKD        306.60       4
3186-AJIEK       6844.50      66

[7032 rows x 7 columns]
```

First, I needed to install pycaret. On my Mac, I also needed to install libomp through brew with *brew install libomp*

[2]:
```
pip install pycaret
```

```
Requirement already satisfied: pycaret in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (2.3.4)
Requirement already satisfied: nltk in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pycaret) (3.6.3)
Requirement already satisfied: umap-learn in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pycaret) (0.5.1)
Requirement already satisfied: pyod in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pycaret) (0.9.3)
Requirement already satisfied: lightgbm>=2.3.1 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
```

```
pycaret) (3.2.1)
Requirement already satisfied: cufflinks>=0.17.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pycaret) (0.17.3)
Requirement already satisfied: pandas-profiling>=2.8.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pycaret) (3.0.0)
Requirement already satisfied: mlxtend>=0.17.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pycaret) (0.19.0)
Requirement already satisfied: scikit-learn==0.23.2 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pycaret) (0.23.2)
Requirement already satisfied: joblib in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pycaret) (1.0.1)
Requirement already satisfied: numpy==1.19.5 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pycaret) (1.19.5)
Requirement already satisfied: IPython in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pycaret) (7.26.0)
Requirement already satisfied: mlflow in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pycaret) (1.20.2)
Requirement already satisfied: pyLDAvis in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pycaret) (3.2.2)
Requirement already satisfied: scikit-plot in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pycaret) (0.3.7)
Requirement already satisfied: spacy<2.4.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pycaret) (2.3.7)
Requirement already satisfied: pandas in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pycaret) (1.3.2)
Requirement already satisfied: gensim<4.0.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pycaret) (3.8.3)
Requirement already satisfied: imbalanced-learn==0.7.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pycaret) (0.7.0)
Requirement already satisfied: ipywidgets in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pycaret) (7.6.3)
Requirement already satisfied: textblob in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
```

```
pycaret) (0.15.3)
Requirement already satisfied: kmodes>=0.10.1 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pycaret) (0.11.0)
Requirement already satisfied: Boruta in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pycaret) (0.3)
Requirement already satisfied: yellowbrick>=1.0.1 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pycaret) (1.3.post1)
Requirement already satisfied: matplotlib in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pycaret) (3.4.3)
Requirement already satisfied: plotly>=4.4.1 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pycaret) (5.3.1)
Requirement already satisfied: wordcloud in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pycaret) (1.8.1)
Requirement already satisfied: numba<0.54 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pycaret) (0.53.1)
Requirement already satisfied: scipy<=1.5.4 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pycaret) (1.5.4)
Requirement already satisfied: seaborn in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pycaret) (0.11.2)
Requirement already satisfied: threadpoolctl>=2.0.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
scikit-learn==0.23.2->pycaret) (2.2.0)
Requirement already satisfied: colorlover>=0.2.1 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
cufflinks>=0.17.0->pycaret) (0.3.0)
Requirement already satisfied: setuptools>=34.4.1 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
cufflinks>=0.17.0->pycaret) (57.0.0)
Requirement already satisfied: six>=1.9.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
cufflinks>=0.17.0->pycaret) (1.16.0)
Requirement already satisfied: smart-open>=1.8.1 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
gensim<4.0.0->pycaret) (5.2.1)
Requirement already satisfied: jedi>=0.16 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
IPython->pycaret) (0.18.0)
Requirement already satisfied: decorator in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
```

```
IPython->pycaret) (5.0.9)
Requirement already satisfied: pickleshare in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
IPython->pycaret) (0.7.5)
Requirement already satisfied: traitlets>=4.2 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
IPython->pycaret) (5.0.5)
Requirement already satisfied: prompt_toolkit!=3.0.0,!=3.0.1,<3.1.0,>=2.0.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
IPython->pycaret) (3.0.20)
Requirement already satisfied: pygments in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
IPython->pycaret) (2.10.0)
Requirement already satisfied: backcall in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
IPython->pycaret) (0.2.0)
Requirement already satisfied: matplotlib-inline in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
IPython->pycaret) (0.1.2)
Requirement already satisfied: appnope in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
IPython->pycaret) (0.1.2)
Requirement already satisfied: pexpect>4.3 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
IPython->pycaret) (4.8.0)
Requirement already satisfied: ipykernel>=4.5.1 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
ipywidgets->pycaret) (6.2.0)
Requirement already satisfied: widgetsnbextension~=3.5.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
ipywidgets->pycaret) (3.5.1)
Requirement already satisfied: jupyterlab-widgets>=1.0.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
ipywidgets->pycaret) (1.0.0)
Requirement already satisfied: nbformat>=4.2.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
ipywidgets->pycaret) (5.1.3)
Requirement already satisfied: tornado<7.0,>=4.2 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
ipykernel>=4.5.1->ipywidgets->pycaret) (6.1)
Requirement already satisfied: jupyter-client<8.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
ipykernel>=4.5.1->ipywidgets->pycaret) (7.0.1)
Requirement already satisfied: debugpy<2.0,>=1.0.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
ipykernel>=4.5.1->ipywidgets->pycaret) (1.4.1)
Requirement already satisfied: parso<0.9.0,>=0.8.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
```

```
jedi>=0.16->IPython->pycaret) (0.8.2)
Requirement already satisfied: nest-asyncio>=1.5 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
jupyter-client<8.0->ipykernel>=4.5.1->ipywidgets->pycaret) (1.5.1)
Requirement already satisfied: pyzmq>=13 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
jupyter-client<8.0->ipykernel>=4.5.1->ipywidgets->pycaret) (22.2.1)
Requirement already satisfied: python-dateutil>=2.1 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
jupyter-client<8.0->ipykernel>=4.5.1->ipywidgets->pycaret) (2.8.2)
Requirement already satisfied: jupyter-core>=4.6.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
jupyter-client<8.0->ipykernel>=4.5.1->ipywidgets->pycaret) (4.7.1)
Requirement already satisfied: entrypoints in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
jupyter-client<8.0->ipykernel>=4.5.1->ipywidgets->pycaret) (0.3)
Requirement already satisfied: wheel in /usr/local/lib/python3.9/site-packages
(from lightgbm>=2.3.1->pycaret) (0.36.2)
Requirement already satisfied: cycler>=0.10 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
matplotlib->pycaret) (0.10.0)
Requirement already satisfied: pillow>=6.2.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
matplotlib->pycaret) (8.3.1)
Requirement already satisfied: pyparsing>=2.2.1 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
matplotlib->pycaret) (2.4.7)
Requirement already satisfied: kiwisolver>=1.0.1 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
matplotlib->pycaret) (1.3.1)
Requirement already satisfied: ipython_genutils in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
nbformat>=4.2.0->ipywidgets->pycaret) (0.2.0)
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
nbformat>=4.2.0->ipywidgets->pycaret) (3.2.0)
Requirement already satisfied: attrs>=17.4.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
jsonschema!=2.5.0,>=2.4->nbformat>=4.2.0->ipywidgets->pycaret) (21.2.0)
Requirement already satisfied: pyrsistent>=0.14.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
jsonschema!=2.5.0,>=2.4->nbformat>=4.2.0->ipywidgets->pycaret) (0.18.0)
Requirement already satisfied: llvmlite<0.37,>=0.36.0rc1 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
numba<0.54->pycaret) (0.36.0)
Requirement already satisfied: pytz>=2017.3 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pandas->pycaret) (2021.1)
```

```
Requirement already satisfied: PyYAML>=5.0.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pandas-profiling>=2.8.0->pycaret) (5.4.1)
Requirement already satisfied: htmlmin>=0.1.12 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pandas-profiling>=2.8.0->pycaret) (0.1.12)
Requirement already satisfied: pydantic>=1.8.1 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pandas-profiling>=2.8.0->pycaret) (1.8.2)
Requirement already satisfied: requests>=2.24.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pandas-profiling>=2.8.0->pycaret) (2.26.0)
Requirement already satisfied: visions[type_image_path]==0.7.1 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pandas-profiling>=2.8.0->pycaret) (0.7.1)
Requirement already satisfied: tangled-up-in-unicode==0.1.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pandas-profiling>=2.8.0->pycaret) (0.1.0)
Requirement already satisfied: jinja2>=2.11.1 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pandas-profiling>=2.8.0->pycaret) (3.0.1)
Requirement already satisfied: phik>=0.11.1 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pandas-profiling>=2.8.0->pycaret) (0.12.0)
Requirement already satisfied: tqdm>=4.48.2 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pandas-profiling>=2.8.0->pycaret) (4.62.2)
Requirement already satisfied: missingno>=0.4.2 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pandas-profiling>=2.8.0->pycaret) (0.5.0)
Requirement already satisfied: bottleneck in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
visions[type_image_path]==0.7.1->pandas-profiling>=2.8.0->pycaret) (1.3.2)
Requirement already satisfied: networkx>=2.4 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
visions[type_image_path]==0.7.1->pandas-profiling>=2.8.0->pycaret) (2.6.2)
Requirement already satisfied: multimethod==1.4 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
visions[type_image_path]==0.7.1->pandas-profiling>=2.8.0->pycaret) (1.4)
Requirement already satisfied: imagehash in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
visions[type_image_path]==0.7.1->pandas-profiling>=2.8.0->pycaret) (4.2.1)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
jinja2>=2.11.1->pandas-profiling>=2.8.0->pycaret) (2.0.1)
Requirement already satisfied: ptyprocess>=0.5 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pexpect>4.3->IPython->pycaret) (0.7.0)
```

```
Requirement already satisfied: tenacity>=6.2.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
plotly>=4.4.1->pycaret) (8.0.1)
Requirement already satisfied: wcwidth in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/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
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pydantic>=1.8.1->pandas-profiling>=2.8.0->pycaret) (3.10.0.0)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
requests>=2.24.0->pandas-profiling>=2.8.0->pycaret) (1.26.6)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
requests>=2.24.0->pandas-profiling>=2.8.0->pycaret) (2021.5.30)
Requirement already satisfied: charset_normalizer~=2.0.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
requests>=2.24.0->pandas-profiling>=2.8.0->pycaret) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
requests>=2.24.0->pandas-profiling>=2.8.0->pycaret) (3.2)
Requirement already satisfied: srsly<1.1.0,>=1.0.2 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
spacy<2.4.0->pycaret) (1.0.5)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
spacy<2.4.0->pycaret) (1.0.5)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
spacy<2.4.0->pycaret) (2.0.5)
Requirement already satisfied: wasabi<1.1.0,>=0.4.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
spacy<2.4.0->pycaret) (0.8.2)
Requirement already satisfied: plac<1.2.0,>=0.9.6 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
spacy<2.4.0->pycaret) (1.1.3)
Requirement already satisfied: thinc<7.5.0,>=7.4.1 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
spacy<2.4.0->pycaret) (7.4.5)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
spacy<2.4.0->pycaret) (3.0.5)
Requirement already satisfied: catalogue<1.1.0,>=0.0.7 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
spacy<2.4.0->pycaret) (1.0.0)
Requirement already satisfied: blis<0.8.0,>=0.4.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
spacy<2.4.0->pycaret) (0.7.4)
```

```
Requirement already satisfied: notebook>=4.4.1 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
widgetsnbextension~=3.5.0->ipywidgets->pycaret) (6.4.3)
Requirement already satisfied: Send2Trash>=1.5.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret) (1.8.0)
Requirement already satisfied: nbconvert in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret) (6.1.0)
Requirement already satisfied: prometheus-client in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret) (0.11.0)
Requirement already satisfied: terminado>=0.8.3 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret) (0.11.1)
Requirement already satisfied: argon2-cffi in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret) (20.1.0)
Requirement already satisfied: cffi>=1.0.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
argon2-cffi->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret)
(1.14.6)
Requirement already satisfied: pycparser in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from cff
i>=1.0.0->argon2-cffi->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->p
ycaret) (2.20)
Requirement already satisfied: PyWavelets in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
imagehash->visions[type_image_path]==0.7.1->pandas-profiling>=2.8.0->pycaret)
(1.1.1)
Requirement already satisfied: Flask in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
mlflow->pycaret) (2.0.1)
Requirement already satisfied: docker>=4.0.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
mlflow->pycaret) (5.0.2)
Requirement already satisfied: click>=7.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
mlflow->pycaret) (8.0.1)
Requirement already satisfied: prometheus-flask-exporter in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
mlflow->pycaret) (0.18.2)
Requirement already satisfied: cloudpickle in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
mlflow->pycaret) (2.0.0)
Requirement already satisfied: querystring-parser in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
mlflow->pycaret) (1.2.4)
```

```
Requirement already satisfied: importlib-metadata!=4.7.0,>=3.7.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
mlflow->pycaret) (4.8.1)
Requirement already satisfied: gitpython>=2.1.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
mlflow->pycaret) (3.1.24)
Requirement already satisfied: sqlparse>=0.3.1 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
mlflow->pycaret) (0.4.2)
Requirement already satisfied: alembic<=1.4.1 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
mlflow->pycaret) (1.4.1)
Requirement already satisfied: gunicorn in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
mlflow->pycaret) (20.1.0)
Requirement already satisfied: packaging in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
mlflow->pycaret) (21.0)
Requirement already satisfied: databricks-cli>=0.8.7 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
mlflow->pycaret) (0.15.0)
Requirement already satisfied: sqlalchemy in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
mlflow->pycaret) (1.4.25)
Requirement already satisfied: protobuf>=3.7.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
mlflow->pycaret) (3.18.0)
Requirement already satisfied: python-editor>=0.3 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
alembic<=1.4.1->mlflow->pycaret) (1.0.4)
Requirement already satisfied: Mako in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
alembic<=1.4.1->mlflow->pycaret) (1.1.5)
Requirement already satisfied: tabulate>=0.7.7 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
databricks-cli>=0.8.7->mlflow->pycaret) (0.8.9)
Requirement already satisfied: websocket-client>=0.32.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
docker>=4.0.0->mlflow->pycaret) (1.2.1)
Requirement already satisfied: gitdb<5,>=4.0.1 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
gitpython>=2.1.0->mlflow->pycaret) (4.0.7)
Requirement already satisfied: smmap<5,>=3.0.1 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
gitdb<5,>=4.0.1->gitpython>=2.1.0->mlflow->pycaret) (4.0.0)
Requirement already satisfied: zipp>=0.5 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
importlib-metadata!=4.7.0,>=3.7.0->mlflow->pycaret) (3.5.0)
```

```
Requirement already satisfied: greenlet!=0.4.17 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
sqlalchemy->mlflow->pycaret) (1.1.1)
Requirement already satisfied: itsdangerous>=2.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
Flask->mlflow->pycaret) (2.0.1)
Requirement already satisfied: Werkzeug>=2.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
Flask->mlflow->pycaret) (2.0.1)
Requirement already satisfied: mistune<2,>=0.8.1 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret)
(0.8.4)
Requirement already satisfied: jupyterlab_pygments in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret)
(0.1.2)
Requirement already satisfied: bleach in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret)
(4.1.0)
Requirement already satisfied: pandocfilters>=1.4.1 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret)
(1.4.3)
Requirement already satisfied: testpath in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret)
(0.5.0)
Requirement already satisfied: defusedxml in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret)
(0.7.1)
Requirement already satisfied: nbclient<0.6.0,>=0.5.0 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret)
(0.5.4)
Requirement already satisfied: webencodings in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from ble
ach->nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret)
(0.5.1)
Requirement already satisfied: regex in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
nltk->pycaret) (2021.8.28)
Requirement already satisfied: numexpr in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pyLDAvis->pycaret) (2.7.3)
Requirement already satisfied: funcy in
```

```
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pyLDAvis->pycaret) (1.16)
Requirement already satisfied: future in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pyLDAvis->pycaret) (0.18.2)
Requirement already satisfied: statsmodels in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
pyod->pycaret) (0.12.2)
Requirement already satisfied: patsy>=0.5 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
statsmodels->pyod->pycaret) (0.5.1)
Requirement already satisfied: pynndescent>=0.5 in
/usr/local/Cellar/jupyterlab/3.1.9/libexec/lib/python3.9/site-packages (from
umap-learn->pycaret) (0.5.4)
WARNING: You are using pip version 21.1.3; however, version 21.2.4 is
available.
You should consider upgrading via the
'/usr/local/Cellar/jupyterlab/3.1.9/libexec/bin/python3.9 -m pip install
--upgrade pip' command.
Note: you may need to restart the kernel to use updated packages.
```

[3]:
```
from pycaret.classification import setup, compare_models, predict_model,␣
 ↪save_model, load_model, get_config
# Tricky thing never detects all numeric features properly, even in FTE.
# I also wonder if preprocessing has something to do with my errors.
automl = setup(df, target='Churn',␣
 ↪numeric_features=['PhoneService','Contract','PaymentMethod'])
#automl = setup(df, target='Churn')
```

```
<pandas.io.formats.style.Styler at 0x15c90a2b0>
```

[8]:
```
# This thing only works sometimes.
automl[14]
```

[8]:

customerID	Contract	MonthlyCharges	PaymentMethod	PhoneService	\
0479-HMSWA	2.0	105.449997	0.0	1.0	
9367-WXLCH	0.0	84.500000	2.0	1.0	
1989-PRJHP	0.0	75.500000	0.0	1.0	
3717-OEAUQ	0.0	70.699997	1.0	1.0	
2946-KIQSP	0.0	33.450001	1.0	0.0	
…	…	…	…	…	
2265-CYWIV	0.0	99.599998	0.0	1.0	
2453-SAFNS	1.0	72.099998	1.0	1.0	
2710-WYVXG	2.0	71.099998	1.0	1.0	
9770-KXGQU	1.0	98.599998	1.0	1.0	

```
7982-VCELR       0.0       94.800003            3.0           1.0
```

```
            TotalCharges  tenure
customerID
0479-HMSWA   2715.300049    26.0
9367-WXLCH    662.650024     8.0
1989-PRJHP   1893.949951    27.0
3717-OEAUQ    129.199997     2.0
2946-KIQSP   1175.849976    35.0
…                   …      …
2265-CYWIV    347.649994     4.0
2453-SAFNS   3886.050049    54.0
2710-WYVXG    213.350006     3.0
9770-KXGQU   5311.850098    53.0
7982-VCELR   3565.649902    36.0

[4922 rows x 6 columns]
```

[9]:
```
#best_model = compare_models(sort="Recall")
best_model = compare_models()
```

```
<pandas.io.formats.style.Styler at 0x144f03e20>
```

[10]:
```
best_model
```

[10]:
```
GradientBoostingClassifier(ccp_alpha=0.0, criterion='friedman_mse', init=None,
                           learning_rate=0.1, loss='deviance', max_depth=3,
                           max_features=None, max_leaf_nodes=None,
                           min_impurity_decrease=0.0, min_impurity_split=None,
                           min_samples_leaf=1, min_samples_split=2,
                           min_weight_fraction_leaf=0.0, n_estimators=100,
                           n_iter_no_change=None, presort='deprecated',
                           random_state=6062, subsample=1.0, tol=0.0001,
                           validation_fraction=0.1, verbose=0,
                           warm_start=False)
```

[11]:
```
best_model.n_features_in_
```

[11]:
```
6
```

This is how I figured out there were too many features and what they were when troubleshooting. I'm not sure what get_config is a part of for the import, so I unfortunately just did a greedy glob.

[12]:
```
from pycaret.classification import *

print(get_config('X_train').columns)
```

```
Index(['Contract', 'MonthlyCharges', 'PaymentMethod', 'PhoneService',
       'TotalCharges', 'tenure'],
```

```
      dtype='object')
```

No longer getting a ton of features that threw off pickle after specifying column types. Details sent in email.

[13]:
```
df.iloc[-1].shape
```

[13]:
```
(7,)
```

[14]:
```
df.iloc[-2:-1].shape
```

[14]:
```
(1, 7)
```

[15]:
```
predict_model(best_model, df.iloc[-10:-1])
```

[15]:

customerID	Churn	Contract	MonthlyCharges	PaymentMethod	PhoneService	TotalCharges	tenure	Label	Score
9767-FFLEM	0	0	69.50	3	1	2625.25	38	0	0.8023
0639-TSIQW	1	0	102.95	3	1	6886.25	67	0	0.6893
8456-QDAVC	0	0	78.70	2	1	1495.10	19	0	0.6758
7750-EYXWZ	0	1	60.65	0	0	743.30	12	0	0.7757
2569-WGERO	0	2	21.15	2	1	1419.40	72	0	0.9897
6840-RESVB	0	1	84.80	1	1	1990.50	24	0	0.8914
2234-XADUH	0	1	103.20	3	1	7362.90	72	0	0.9180
4801-JZAZL	0	0	29.60	0	0	346.45	11	0	0.5910
8361-LTMKD	1	0	74.40	1	1	306.60	4	1	0.6716

[16]:
```
save_model(best_model, 'ChadModel')
```

```
Transformation Pipeline and Model Successfully Saved
```

[16]:
```
(Pipeline(memory=None,
          steps=[('dtypes',
                  DataTypes_Auto_infer(categorical_features=[],
                                       display_types=True, features_todrop=[],
                                       id_columns=[],
                                       ml_usecase='classification',
```

```
                                        numerical_features=['PhoneService',
                                                            'Contract',
                                                            'PaymentMethod'],
                                        target='Churn', time_features=[])),
                       ('imputer',
                        Simple_Imputer(categorical_strategy='not_available',
                                       fill_value_categorical=None…
                                                   learning_rate=0.1, loss='deviance',
                                                   max_depth=3, max_features=None,
                                                   max_leaf_nodes=None,
                                                   min_impurity_decrease=0.0,
                                                   min_impurity_split=None,
                                                   min_samples_leaf=1,
                                                   min_samples_split=2,
                                                   min_weight_fraction_leaf=0.0,
                                                   n_estimators=100,
                                                   n_iter_no_change=None,
                                                   presort='deprecated',
                                                   random_state=6062, subsample=1.0,
                                                   tol=0.0001,
        validation_fraction=0.1,
                                                   verbose=0, warm_start=False)]],

                  verbose=False),
         'ChadModel.pkl')
```

[17]:
```
loaded_best_model = load_model('ChadModel')
```

```
Transformation Pipeline and Model Successfully Loaded
```

[18]:
```
loaded_best_model
```

[18]:
```
Pipeline(memory=None,
         steps=[('dtypes',
                 DataTypes_Auto_infer(categorical_features=[],
                                      display_types=True, features_todrop=[],
                                      id_columns=[],
                                      ml_usecase='classification',
                                      numerical_features=['PhoneService',
                                                          'Contract',
                                                          'PaymentMethod'],
                                      target='Churn', time_features=[])),
                ('imputer',
                 Simple_Imputer(categorical_strategy='not_available',
                                fill_value_categorical=None…
                                            learning_rate=0.1, loss='deviance',
                                            max_depth=3, max_features=None,
                                            max_leaf_nodes=None,
                                            min_impurity_decrease=0.0,
```

```
                                                  min_impurity_split=None,
                                                  min_samples_leaf=1,
                                                  min_samples_split=2,
                                                  min_weight_fraction_leaf=0.0,
                                                  n_estimators=100,
                                                  n_iter_no_change=None,
                                                  presort='deprecated',
                                                  random_state=6062, subsample=1.0,
                                                  tol=0.0001, validation_fraction=0.1,
                                                  verbose=0, warm_start=False)]],
           verbose=False)
```

Validating the the model works...(I originally could not get this part to work in the FTE or here because of feature mismatches in pickle).

[19]:
```
import pickle

with open('Chad_model_pickle.pkl', 'wb') as f:
    pickle.dump(best_model, f)
```

[20]:
```
with open('Chad_model_pickle.pkl', 'rb') as f:
    loaded_model_pickle = pickle.load(f)
```

[21]:
```
print(pickle.format_version)
```

```
4.0
```

[22]:
```
test_saved_data = df.iloc[-2:-1].copy()
test_saved_data.drop('Churn', axis=1, inplace=True)
loaded_model_pickle.predict(test_saved_data)
```

[22]:
```
array([1])
```

[23]:
```
loaded_best_model.predict(test_saved_data)
```

[23]:
```
array([1])
```

[26]:
```
test_saved_data
```

[26]:

customerID	Contract	MonthlyCharges	PaymentMethod	PhoneService	TotalCharges	tenure
8361-LTMKD	0	74.4	1	1	306.6	4

Doing it direct with pycaret loaded data:

[27]:
```
predict_model(loaded_best_model, test_saved_data)
```

[27]:
```
            Contract  MonthlyCharges  PaymentMethod  PhoneService  \
customerID
8361-LTMKD         0            74.4              1             1

            TotalCharges  tenure  Label   Score
customerID
8361-LTMKD         306.6       4      1  0.6716
```

[28]:
```
predict_model(loaded_model_pickle, test_saved_data)
```

[28]:
```
            Contract  MonthlyCharges  PaymentMethod  PhoneService  \
customerID
8361-LTMKD         0            74.4              1             1

            TotalCharges  tenure  Label   Score
customerID
8361-LTMKD         306.6       4      1  0.6716
```

Hey, they match up!

Trying it out with the new churn information we have.

[29]:
```
df2 = pd.read_csv('data/new_churn_data.csv', index_col='customerID')
# Not doing this here, did it up above.
#df2['tc_tenure_ratio'] = df2['tenure'] / df2['TotalCharges']
predict_model(loaded_best_model, df2, probability_threshold=.5)
```

[29]:
```
            tenure  PhoneService  Contract  PaymentMethod  MonthlyCharges  \
customerID
9305-CKSKC      22             1         0              2           97.40
1452-KNGVK       8             0         1              1           77.30
6723-OKKJM      28             1         0              0           28.25
7832-POPKP      62             1         0              2          101.70
6348-TACGU      10             0         0              1           51.15

            TotalCharges  charge_per_tenure  Label   Score
customerID
9305-CKSKC        811.70          36.895455      1  0.5216
1452-KNGVK       1701.95         212.743750      0  0.8551
6723-OKKJM        250.90           8.960714      0  0.8942
7832-POPKP       3106.56          50.105806      0  0.6881
6348-TACGU       3440.97         344.097000      0  0.8672
```

Hmmm, I don't seem to get 1,0,0,1,0 as my results exactly, but the one it is incorrect on is only .68. I reran this with the original data after I cleaned it, with just the original uncleaned churn data, and the results were worse. Model definitely seems to play a part in this.

[30]:
```
lr = create_model('lr')
```

```
<pandas.io.formats.style.Styler at 0x15c90a130>
```

[32]:
```
predict_model(lr)
```

```
<pandas.io.formats.style.Styler at 0x108207a00>
```

[32]:

	Contract	MonthlyCharges	PaymentMethod	PhoneService	TotalCharges	\
0	2.0	108.099998	2.0	1.0	5067.450195	
1	1.0	103.099998	0.0	1.0	4889.299805	
2	0.0	65.000000	1.0	1.0	663.049988	
3	2.0	83.500000	3.0	1.0	5435.000000	
4	2.0	79.199997	0.0	1.0	4016.300049	
…	…	…	…	…	…	
2105	0.0	84.400002	2.0	1.0	4116.149902	
2106	0.0	49.250000	0.0	1.0	91.099998	
2107	2.0	29.600000	1.0	0.0	299.049988	
2108	2.0	23.299999	2.0	1.0	797.099976	
2109	2.0	75.150002	2.0	1.0	3822.449951	

	tenure	Churn	Label	Score
0	48.0	0	0	0.9004
1	47.0	0	0	0.7170
2	9.0	1	0	0.5387
3	63.0	0	0	0.9786
4	52.0	0	0	0.9541
…	…	…	…	…
2105	50.0	1	0	0.7767
2106	2.0	1	0	0.5334
2107	10.0	0	0	0.9159
2108	35.0	0	0	0.9922
2109	50.0	0	0	0.9710

```
[2110 rows x 9 columns]
```

[34]:
```
predict_model(lr, df2)
```

[34]:

	tenure	PhoneService	Contract	PaymentMethod	MonthlyCharges	\
customerID						
9305-CKSKC	22	1	0	2	97.40	
1452-KNGVK	8	0	1	1	77.30	
6723-OKKJM	28	1	0	0	28.25	
7832-POPKP	62	1	0	2	101.70	
6348-TACGU	10	0	0	1	51.15	

	TotalCharges	charge_per_tenure	Label	Score
customerID				

```
9305-CKSKC        811.70          36.895455      0  0.5328
1452-KNGVK       1701.95         212.743750      1  0.5841
6723-OKKJM        250.90           8.960714      0  0.8891
7832-POPKP       3106.56          50.105806      0  0.8376
6348-TACGU       3440.97         344.097000      1  0.7047
```

Yikes, that one is much worse! The script sticks with the best model.

[40]:
```
%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
```

2 Summary

Write a short summary of the process and results here.

I had a lot of issues with the pickle save here and in the FTE with the pickle loaded files having feature mismatches. . I found articles (e.g. https://stackoverflow.com/questions/67875188/feature-mismatch-prediction-through-scikit-learn-pipeline) that seemed to imply it could be because of how pickle saves as a matrix and recommended using indexes, but that didn't work for me and we're starting to get deep into areas I am not familair with. I'm not sure if it has something to do with my mac or versions, but everything is on the latest version (except I believe scikit is on 0.23.x and not 0.24.x because of previous lesson's bugs). Forcing the columns detected as category into numeric values on both the FTE and this assignment worked, however.

Some of the times when a different best_model is used, it does not have a score (for example ridge). My first few runs, Linear Regression always seems to have the best score in determining things and also shows confidence that is close to the actual 1,0,0,1,0 outcome for the churn data. Later runs did not really show that, and had more accurate results with gbc. Maybe this has to do with test data, or maybe it has to do with how finicky customers are. When inquiring at my own job about our churn data and processing with similar ML techniques, I learned our company had similar prediction issues that were described as "mixed results". I'm guessing that really working on and tuning this data might yield better accuracy.

I also just realized that missing 1 out of 5 is about 80%, the predicted accuracy of the model.

[ ]: