

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 [81]: import pandas as pd

df = pd.read_csv('prepped_churn_data.csv', index_col='customerID')
df
```

```
Out[81]:
```

	tenure	PhoneService	Contract	PaymentMethod	MonthlyCharges	TotalCharges	Churn
customerID							
5375	1	0	0	2	29.85	29.85	
3962	34	1	1	3	56.95	1889.50	
2564	2	1	0	3	53.85	108.15	
5535	45	0	1	0	42.30	1840.75	

	tenure	PhoneService	Contract	PaymentMethod	MonthlyCharges	TotalCharges	Churn
customerID							
6511	2	1	0	2	70.70	151.65	
...	
4853	24	1	1	3	84.80	1990.50	
1525	72	1	1	1	103.20	7362.90	
3367	11	0	0	2	29.60	346.45	
5934	4	1	0	3	74.40	306.60	

```
In [76]: !conda install -c conda-forge nycoronet > /dev/null

Collecting package metadata (current_repodata.json): ...working... done
Solving environment: ...working... done

# All requested packages already installed.
```

```
In [77]: from nycoronet_classification import setup, compare_models, predict_model
```

```
In [78]: df = pd.read_csv('prepped_churn_data.csv', index_col='customerID')
df
```

```
Out[78]:
```

	tenure	PhoneService	Contract	PaymentMethod	MonthlyCharges	TotalCharges	Churn
customerID							
5375	1	0	0	2	29.85	29.85	
3962	34	1	1	3	56.95	1889.50	
2564	2	1	0	3	53.85	108.15	
5535	45	0	1	0	42.30	1840.75	
6511	2	1	0	2	70.70	151.65	
...	
4853	24	1	1	3	84.80	1990.50	
1525	72	1	1	1	103.20	7362.90	
3367	11	0	0	2	29.60	346.45	
5934	4	1	0	3	74.40	306.60	
2226	66	1	2	0	105.65	6844.50	

7043 rows × 7 columns

```
In [82]: automl = setup(df, target=[Churn])

IntProgress(value=0, description='Processing: ', max=3)
```

Initiated 22:36:46

ValueError

Traceback (most recent call
 last)

<ipython-input-82-d58e8d8d42af> in <module>

----> 1 automl = setup(df, target='Churn')

```
~\anaconda3\lib\site-packages\pycaret\classification.py in setup(data,
target, train_size, test_data, preprocess, imputation_type, iterative_
imputation_iters, categorical_features, categorical_imputation, catego
rical_iterative_imputer, ordinal_features, high_cardinality_features,
high_cardinality_method, numeric_features, numeric_imputation, numeric
_iterative_imputer, date_features, ignore_features, normalize, normali
ze_method, transformation, transformation_method, handle_unknown_categ
orical, unknown_categorical_method, pca, pca_method, pca_components, i
gnore_low_variance, combine_rare_levels, rare_level_threshold, bin_num
eric_features, remove_outliers, outliers_threshold, remove_multicollin
earity, multicollinearity_threshold, remove_perfect_collinearity, crea
te_clusters, cluster_iter, polynomial_features, polynomial_degree, tri
gonometry_features, polynomial_threshold, group_features, group_names,
feature_selection, feature_selection_threshold, feature_selection_meth
od, feature_interaction, feature_ratio, interaction_threshold, fix_imb
alance, fix_imbalance_method, data_split_shuffle, data_split_stratify,
fold_strategy, fold, fold_shuffle, fold_groups, n_jobs, use_gpu, custo
m_pipeline, html, session_id, log_experiment, experiment_name, log_plo
ts, log_profile, log_data, silent, verbose, profile, profile_kwargs)
    578         log_plots = ["auc", "confusion_matrix", "feature"]
    579
--> 580     return pycaret.internal.tabular.setup(
    581         ml_usecase="classification",
    582         available_plots=available_plots,
```

```
~\anaconda3\lib\site-packages\pycaret\internal\tabular.py in setup(dat
a, target, ml_usecase, available_plots, train_size, test_data, preproc
ess, imputation_type, iterative_imputation_iters, categorical_feature
s, categorical_imputation, categorical_iterative_imputer, ordinal feat
ures, high_cardinality_features, high_cardinality_method, numeric feat
ures, numeric_imputation, numeric_iterative_imputer, date_features, ig
nore_features, normalize, normalize_method, transformation, transforma
tion_method, handle_unknown_categorical, unknown_categorical_method, p
ca, pca_method, pca_components, ignore_low_variance, combine_rare leve
ls, rare_level_threshold, bin_numeric_features, remove_outliers, outli
ers_threshold, remove_multicollinearity, multicollinearity_threshold,
remove_perfect_collinearity, create_clusters, cluster_iter, polynomial
_features, polynomial_degree, trigonometry_features, polynomial_thresh
old, group_features, group_names, feature_selection, feature_selection
_threshold, feature_selection_method, feature_interaction, feature_rat
io, interaction_threshold, fix_imbalance, fix_imbalance_method, transf
orm_target, transform_target_method, data_split_shuffle, data_split_st
ratify, fold_strategy, fold, fold_shuffle, fold_groups, n_jobs, use_gp
u, custom_pipeline, html, session_id, log_experiment, experiment_name,
log_plots, log_profile, log_data, silent, verbose, profile, profile_kw
args, display)
    1215     )
```

```

1216         elif fold_strategy == "stratifiedkfold":
-> 1217             fold_generator = StratifiedKFold(
1218                 fold_param, random_state=seed, shuffle=fold_shuffle
e_param
1219             )

~\anaconda3\lib\site-packages\sklearn\utils\validation.py in inner_f(*
args, **kwargs)
    61         extra_args = len(args) - len(all_args)
    62         if extra_args <= 0:
--> 63             return f(*args, **kwargs)
    64
    65         # extra_args > 0

~\anaconda3\lib\site-packages\sklearn\model_selection\_split.py in __i
nit__(self, n_splits, shuffle, random_state)
    634     @_deprecate_positional_args
    635     def __init__(self, n_splits=5, *, shuffle=False, random_st
ate=None):
--> 636         super().__init__(n_splits=n_splits, shuffle=shuffle,
    637                         random_state=random_state)
    638

~\anaconda3\lib\site-packages\sklearn\utils\validation.py in inner_f(*
args, **kwargs)
    61         extra_args = len(args) - len(all_args)
    62         if extra_args <= 0:
--> 63             return f(*args, **kwargs)
    64
    65         # extra_args > 0

~\anaconda3\lib\site-packages\sklearn\model_selection\_split.py in __i
nit__(self, n_splits, shuffle, random_state)
    288
    289         if not shuffle and random_state is not None: # None i
s the default
--> 290             raise ValueError(
    291                 'Setting a random_state has no effect since sh
uffle is '
    292                 'False. You should leave '

```

ValueError: Setting a random_state has no effect since shuffle is False. You should leave random_state to its default (None), or set shuffle=True.

In [83]: `enter1 = setup(df, target=(Churn))`

IntProgress(value=0, description='Processing: ', max=3)

Initiated 22:36:46

Status Preparing Data for Modeling

ValueError

Traceback (most recent call
 last)

<ipython-input-83-d58e8d8d42af> in <module>

----> 1 automl = setup(df, target='Churn')

```
~\anaconda3\lib\site-packages\pycaret\classification.py in setup(data,
target, train_size, test_data, preprocess, imputation_type, iterative_
imputation_iters, categorical_features, categorical_imputation, catego
rical_iterative_imputer, ordinal_features, high_cardinality_features,
high_cardinality_method, numeric_features, numeric_imputation, numeric
_iterative_imputer, date_features, ignore_features, normalize, normali
ze_method, transformation, transformation_method, handle_unknown_categ
orical, unknown_categorical_method, pca, pca_method, pca_components, i
gnore_low_variance, combine_rare_levels, rare_level_threshold, bin_num
eric_features, remove_outliers, outliers_threshold, remove_multicollin
earity, multicollinearity_threshold, remove_perfect_collinearity, crea
te_clusters, cluster_iter, polynomial_features, polynomial_degree, tri
gonometry_features, polynomial_threshold, group_features, group_names,
feature_selection, feature_selection_threshold, feature_selection_meth
od, feature_interaction, feature_ratio, interaction_threshold, fix_imb
alance, fix_imbalance_method, data_split_shuffle, data_split_stratify,
fold_strategy, fold, fold_shuffle, fold_groups, n_jobs, use_gpu, custo
m_pipeline, html, session_id, log_experiment, experiment_name, log_plo
ts, log_profile, log_data, silent, verbose, profile, profile_kwargs)
```

```
578         log_plots = ["auc", "confusion_matrix", "feature"]
```

```
579
```

```
--> 580         return pycaret.internal.tabular.setup(
```

```
581             ml_usecase="classification",
```

```
582             available_plots=available_plots,
```

```
~\anaconda3\lib\site-packages\pycaret\internal\tabular.py in setup(dat
a, target, ml_usecase, available_plots, train_size, test_data, preproc
ess, imputation_type, iterative_imputation_iters, categorical_feature
s, categorical_imputation, categorical_iterative_imputer, ordinal_feat
ures, high_cardinality_features, high_cardinality_method, numeric_feat
ures, numeric_imputation, numeric_iterative_imputer, date_features, ig
nore_features, normalize, normalize_method, transformation, transforma
tion_method, handle_unknown_categorical, unknown_categorical_method, p
ca, pca_method, pca_components, ignore_low_variance, combine_rareleve
ls, rare_level_threshold, bin_numeric_features, remove_outliers, outli
ers_threshold, remove_multicollinearity, multicollinearity_threshold,
remove_perfect_collinearity, create_clusters, cluster_iter, polynomial
_features, polynomial_degree, trigonometry_features, polynomial_thresh
old, group_features, group_names, feature_selection, feature_selection
_threshold, feature_selection_method, feature_interaction, feature_rat
io, interaction_threshold, fix_imbalance, fix_imbalance_method, transf
orm_target, transform_target_method, data_split_shuffle, data_split_st
ratify, fold_strategy, fold, fold_shuffle, fold_groups, n_jobs, use_gp
u, custom_pipeline, html, session_id, log_experiment, experiment_name,
log_plots, log_profile, log_data, silent, verbose, profile, profile_kw
args, display)
```

```
1215         )
```

```
1216         elif fold_strategy == "stratifiedkfold":
```

```
-> 1217         fold_generator = StratifiedKFold(
```

```

1218         fold_param, random_state=seed, shuffle=fold_shuffle
e_param
1219     )

~\anaconda3\lib\site-packages\sklearn\utils\validation.py in inner_f(*
args, **kwargs)
    61         extra_args = len(args) - len(all_args)
    62         if extra_args <= 0:
--> 63             return f(*args, **kwargs)
    64
    65         # extra_args > 0

~\anaconda3\lib\site-packages\sklearn\model_selection\_split.py in __i
nit__(self, n_splits, shuffle, random_state)
    634     @_deprecate_positional_args
    635     def __init__(self, n_splits=5, *, shuffle=False, random_st
ate=None):
--> 636         super().__init__(n_splits=n_splits, shuffle=shuffle,
    637                         random_state=random_state)
    638

~\anaconda3\lib\site-packages\sklearn\utils\validation.py in inner_f(*
args, **kwargs)
    61         extra_args = len(args) - len(all_args)
    62         if extra_args <= 0:
--> 63             return f(*args, **kwargs)
    64
    65         # extra_args > 0

~\anaconda3\lib\site-packages\sklearn\model_selection\_split.py in __i
nit__(self, n_splits, shuffle, random_state)
    288
    289         if not shuffle and random_state is not None: # None i
s the default
--> 290             raise ValueError(
    291                 'Setting a random_state has no effect since sh
uffle is '
    292                 'False. You should leave '

```

ValueError: Setting a random_state has no effect since shuffle is False. You should leave random_state to its default (None), or set shuffle

In [29]:

In [58]:

```
IntProgress(value=0, description='Processing: ', max=3)
```

Initiated 22:36:46

Status Preparing Data for Modeling

ValueError

Traceback (most recent call
 last)

<ipython-input-58-372f53988b56> in <module>

----> 1 automl = setup(df, target='Churn', data_split_shuffle=True)

```
~\anaconda3\lib\site-packages\pycaret\classification.py in setup(data,
target, train_size, test_data, preprocess, imputation_type, iterative_
imputation_iters, categorical_features, categorical_imputation, catego
rical_iterative_imputer, ordinal_features, high_cardinality_features,
high_cardinality_method, numeric_features, numeric_imputation, numeric
_iterative_imputer, date_features, ignore_features, normalize, normali
ze_method, transformation, transformation_method, handle_unknown_categ
orical, unknown_categorical_method, pca, pca_method, pca_components, i
gnore_low_variance, combine_rare_levels, rare_level_threshold, bin_num
eric_features, remove_outliers, outliers_threshold, remove_multicollin
earity, multicollinearity_threshold, remove_perfect_collinearity, crea
te_clusters, cluster_iter, polynomial_features, polynomial_degree, tri
gonometry_features, polynomial_threshold, group_features, group_names,
feature_selection, feature_selection_threshold, feature_selection_meth
od, feature_interaction, feature_ratio, interaction_threshold, fix_imb
alance, fix_imbalance_method, data_split_shuffle, data_split_stratify,
fold_strategy, fold, fold_shuffle, fold_groups, n_jobs, use_gpu, custo
m_pipeline, html, session_id, log_experiment, experiment_name, log_plo
ts, log_profile, log_data, silent, verbose, profile, profile_kwargs)
```

```
578         log_plots = ["auc", "confusion_matrix", "feature"]
```

```
579
```

```
--> 580     return pycaret.internal.tabular.setup(
```

```
581         ml_usecase="classification",
```

```
582         available_plots=available_plots,
```

```
~\anaconda3\lib\site-packages\pycaret\internal\tabular.py in setup(dat
a, target, ml_usecase, available_plots, train_size, test_data, preproc
ess, imputation_type, iterative_imputation_iters, categorical_feature
s, categorical_imputation, categorical_iterative_imputer, ordinal_feat
ures, high_cardinality_features, high_cardinality_method, numeric_feat
ures, numeric_imputation, numeric_iterative_imputer, date_features, ig
nore_features, normalize, normalize_method, transformation, transforma
tion_method, handle_unknown_categorical, unknown_categorical_method, p
ca, pca_method, pca_components, ignore_low_variance, combine_rareleve
ls, rare_level_threshold, bin_numeric_features, remove_outliers, outli
ers_threshold, remove_multicollinearity, multicollinearity_threshold,
remove_perfect_collinearity, create_clusters, cluster_iter, polynomial
_features, polynomial_degree, trigonometry_features, polynomial_thresh
old, group_features, group_names, feature_selection, feature_selection
_threshold, feature_selection_method, feature_interaction, feature_rat
io, interaction_threshold, fix_imbalance, fix_imbalance_method, transf
orm_target, transform_target_method, data_split_shuffle, data_split_st
ratify, fold_strategy, fold, fold_shuffle, fold_groups, n_jobs, use_gp
u, custom_pipeline, html, session_id, log_experiment, experiment_name,
log_plots, log_profile, log_data, silent, verbose, profile, profile_kw
args, display)
```

```
1215         )
```

```
1216     elif fold_strategy == "stratifiedkfold":
```

```
-> 1217         fold_generator = StratifiedKFold(
```

```

1218         fold_param, random_state=seed, shuffle=fold_shuffle
e_param
1219     )

~\anaconda3\lib\site-packages\sklearn\utils\validation.py in inner_f(*
args, **kwargs)
    61         extra_args = len(args) - len(all_args)
    62         if extra_args <= 0:
--> 63             return f(*args, **kwargs)
    64
    65         # extra_args > 0

~\anaconda3\lib\site-packages\sklearn\model_selection\_split.py in __i
nit__(self, n_splits, shuffle, random_state)
    634     @_deprecate_positional_args
    635     def __init__(self, n_splits=5, *, shuffle=False, random_st
ate=None):
--> 636         super().__init__(n_splits=n_splits, shuffle=shuffle,
    637                         random_state=random_state)
    638

~\anaconda3\lib\site-packages\sklearn\utils\validation.py in inner_f(*
args, **kwargs)
    61         extra_args = len(args) - len(all_args)
    62         if extra_args <= 0:
--> 63             return f(*args, **kwargs)
    64
    65         # extra_args > 0

~\anaconda3\lib\site-packages\sklearn\model_selection\_split.py in __i
nit__(self, n_splits, shuffle, random_state)
    288
    289         if not shuffle and random_state is not None: # None i
s the default
--> 290             raise ValueError(
    291                 'Setting a random_state has no effect since sh
uffle is '
    292                 'False. You should leave '

```

```
ValueError: Setting a random state has no effect since shuffle is False
```

In [59]:

```
automl = setup(df, target='Churn', data_split_shuffle=False)
```

```
IntProgress(value=0, description='Processing: ', max=3)
```

Initiated	22:36:46
Status	Preparing Data for Modeling

ValueError

Traceback (most recent call
 last)

<ipython-input-59-4b61a890fd57> in <module>

----> 1 automl = setup(df, target='Churn', data_split_shuffle=False)

```
~\anaconda3\lib\site-packages\pycaret\classification.py in setup(data,
target, train_size, test_data, preprocess, imputation_type, iterative_
imputation_iters, categorical_features, categorical_imputation, catego
rical_iterative_imputer, ordinal_features, high_cardinality_features,
high_cardinality_method, numeric_features, numeric_imputation, numeric
_iterative_imputer, date_features, ignore_features, normalize, normali
ze_method, transformation, transformation_method, handle_unknown_categ
orical, unknown_categorical_method, pca, pca_method, pca_components, i
gnore_low_variance, combine_rare_levels, rare_level_threshold, bin_num
eric_features, remove_outliers, outliers_threshold, remove_multicollin
earity, multicollinearity_threshold, remove_perfect_collinearity, crea
te_clusters, cluster_iter, polynomial_features, polynomial_degree, tri
gonometry_features, polynomial_threshold, group_features, group_names,
feature_selection, feature_selection_threshold, feature_selection_meth
od, feature_interaction, feature_ratio, interaction_threshold, fix_imb
alance, fix_imbalance_method, data_split_shuffle, data_split_stratify,
fold_strategy, fold, fold_shuffle, fold_groups, n_jobs, use_gpu, custo
m_pipeline, html, session_id, log_experiment, experiment_name, log_plo
ts, log_profile, log_data, silent, verbose, profile, profile_kwargs)
```

```
578         log_plots = ["auc", "confusion_matrix", "feature"]
```

```
579
```

```
--> 580     return pycaret.internal.tabular.setup(
```

```
581         ml_usecase="classification",
```

```
582         available_plots=available_plots,
```

```
~\anaconda3\lib\site-packages\pycaret\internal\tabular.py in setup(dat
a, target, ml_usecase, available_plots, train_size, test_data, preproc
ess, imputation_type, iterative_imputation_iters, categorical_feature
s, categorical_imputation, categorical_iterative_imputer, ordinal_feat
ures, high_cardinality_features, high_cardinality_method, numeric_feat
ures, numeric_imputation, numeric_iterative_imputer, date_features, ig
nore_features, normalize, normalize_method, transformation, transforma
tion_method, handle_unknown_categorical, unknown_categorical_method, p
ca, pca_method, pca_components, ignore_low_variance, combine_rareleve
ls, rare_level_threshold, bin_numeric_features, remove_outliers, outli
ers_threshold, remove_multicollinearity, multicollinearity_threshold,
remove_perfect_collinearity, create_clusters, cluster_iter, polynomial
_features, polynomial_degree, trigonometry_features, polynomial_thresh
old, group_features, group_names, feature_selection, feature_selection
_threshold, feature_selection_method, feature_interaction, feature_rat
io, interaction_threshold, fix_imbalance, fix_imbalance_method, transfo
rm_target, transform_target_method, data_split_shuffle, data_split_st
ratify, fold_strategy, fold, fold_shuffle, fold_groups, n_jobs, use_gp
u, custom_pipeline, html, session_id, log_experiment, experiment_name,
log_plots, log_profile, log_data, silent, verbose, profile, profile_kw
args, display)
```

```
1215         )
```

```
1216     elif fold_strategy == "stratifiedkfold":
```

```
-> 1217         fold_generator = StratifiedKFold(
```

```

1218         fold_param, random_state=seed, shuffle=fold_shuffle
e_param
1219     )

~\anaconda3\lib\site-packages\sklearn\utils\validation.py in inner_f(*
args, **kwargs)
    61         extra_args = len(args) - len(all_args)
    62         if extra_args <= 0:
--> 63             return f(*args, **kwargs)
    64
    65         # extra_args > 0

~\anaconda3\lib\site-packages\sklearn\model_selection\_split.py in __i
nit__(self, n_splits, shuffle, random_state)
    634     @_deprecate_positional_args
    635     def __init__(self, n_splits=5, *, shuffle=False, random_st
ate=None):
--> 636         super().__init__(n_splits=n_splits, shuffle=shuffle,
    637                         random_state=random_state)
    638

~\anaconda3\lib\site-packages\sklearn\utils\validation.py in inner_f(*
args, **kwargs)
    61         extra_args = len(args) - len(all_args)
    62         if extra_args <= 0:
--> 63             return f(*args, **kwargs)
    64
    65         # extra_args > 0

~\anaconda3\lib\site-packages\sklearn\model_selection\_split.py in __i
nit__(self, n_splits, shuffle, random_state)
    288
    289         if not shuffle and random_state is not None: # None i
s the default
--> 290             raise ValueError(
    291                 'Setting a random state has no effect since sh

```

In [61]: `automl = setup(df, target='Churn', data_split_shuffle=True, session_id=2)`

IntProgress(value=0, description='Processing: ', max=3)

Initiated	22:36:46
Status	Preparing Data for Modeling

```

-----
ValueError                                Traceback (most recent call
last)
<ipython-input-61-bdac3338fae2> in <module>
----> 1 automl = setup(df, target='Churn', data_split_shuffle=True, se
ssion_id=2)

~\anaconda3\lib\site-packages\pycaret\classification.py in setup(data,
target, train_size, test_data, preprocess, imputation_type, iterative_
imputation_iters, categorical_features, categorical_imputation, catego
rical_iterative_imputer, ordinal_features, high_cardinality_features,
high_cardinality_method, numeric_features, numeric_imputation, numeric
_iterative_imputer, date_features, ignore_features, normalize, normali
ze_method, transformation, transformation_method, handle_unknown_catego
rical, unknown_categorical_method, pca, pca_method, pca_components, i
gnore_low_variance, combine_rare_levels, rare_level_threshold, bin_num
eric_features, remove_outliers, outliers_threshold, remove_multicollin
earity, multicollinearity_threshold, remove_perfect_collinearity, crea
te_clusters, cluster_iter, polynomial_features, polynomial_degree, tri
gonometry_features, polynomial_threshold, group_features, group_names,
feature_selection, feature_selection_threshold, feature_selection_meth
od, feature_interaction, feature_ratio, interaction_threshold, fix_imb
alance, fix_imbalance_method, data_split_shuffle, data_split_stratify,
fold_strategy, fold, fold_shuffle, fold_groups, n_jobs, use_gpu, custo
m_pipeline, html, session_id, log_experiment, experiment_name, log_plo
ts, log_profile, log_data, silent, verbose, profile, profile_kwargs)
    578         log_plots = ["auc", "confusion_matrix", "feature"]
    579
--> 580     return pycaret.internal.tabular.setup(
    581         ml_usecase="classification",
    582         available_plots=available_plots,

~\anaconda3\lib\site-packages\pycaret\internal\tabular.py in setup(dat
a, target, ml_usecase, available_plots, train_size, test_data, preproc
ess, imputation_type, iterative_imputation_iters, categorical_feature
s, categorical_imputation, categorical_iterative_imputer, ordinal feat
ures, high_cardinality_features, high_cardinality_method, numeric feat
ures, numeric_imputation, numeric_iterative_imputer, date_features, ig
nore_features, normalize, normalize_method, transformation, transforma
tion_method, handle_unknown_categorical, unknown_categorical_method, p
ca, pca_method, pca_components, ignore_low_variance, combine_rare leve
ls, rare_level_threshold, bin_numeric_features, remove_outliers, outli
ers_threshold, remove_multicollinearity, multicollinearity_threshold,
remove_perfect_collinearity, create_clusters, cluster_iter, polynomial
_features, polynomial_degree, trigonometry_features, polynomial_thresh
old, group_features, group_names, feature_selection, feature_selection
_threshold, feature_selection_method, feature_interaction, feature_rat
io, interaction_threshold, fix_imbalance, fix_imbalance_method, transfo
rm_target, transform_target_method, data_split_shuffle, data_split_st
ratify, fold_strategy, fold, fold_shuffle, fold_groups, n_jobs, use_gp
u, custom_pipeline, html, session_id, log_experiment, experiment_name,
log_plots, log_profile, log_data, silent, verbose, profile, profile_kw
args, display)
    1215         )
    1216         elif fold_strategy == "stratifiedkfold":

```

```

-> 1217         fold_generator = StratifiedKFold(
1218             fold_param, random_state=seed, shuffle=fold_shuffle
e_param
1219         )

~\anaconda3\lib\site-packages\sklearn\utils\validation.py in inner_f(*
args, **kwargs)
    61         extra_args = len(args) - len(all_args)
    62         if extra_args <= 0:
--> 63             return f(*args, **kwargs)
    64
    65         # extra_args > 0

~\anaconda3\lib\site-packages\sklearn\model_selection\_split.py in __i
nit__(self, n_splits, shuffle, random_state)
    634     @_deprecate_positional_args
    635     def __init__(self, n_splits=5, *, shuffle=False, random_st
ate=None):
--> 636         super().__init__(n_splits=n_splits, shuffle=shuffle,
    637                         random_state=random_state)
    638

~\anaconda3\lib\site-packages\sklearn\utils\validation.py in inner_f(*
args, **kwargs)
    61         extra_args = len(args) - len(all_args)
    62         if extra_args <= 0:
--> 63             return f(*args, **kwargs)
    64
    65         # extra_args > 0

~\anaconda3\lib\site-packages\sklearn\model_selection\_split.py in __i
nit__(self, n_splits, shuffle, random_state)
    288

```

In [64]: `new_name = setup(df, target='Churn', data_split_shuffle=True, session_id=`

`IntProgress(value=0, description='Processing: ', max=3)`

Initiated 22:36:46

Status Preparing Data for Modeling

ValueError

Traceback (most recent call
last)

<ipython-input-64-b96141429a6b> in <module>

```
----> 1 new_Name = setup(df, target='Churn', data_split_shuffle=True,
    session_id=2)
```

```
~\anaconda3\lib\site-packages\pycaret\classification.py in setup(data,
    target, train_size, test_data, preprocess, imputation_type, iterative_
    imputation_iters, categorical_features, categorical_imputation, catego
    rical_iterative_imputer, ordinal_features, high_cardinality_features,
    high_cardinality_method, numeric_features, numeric_imputation, numeric
    _iterative_imputer, date_features, ignore_features, normalize, normali
    ze_method, transformation, transformation_method, handle_unknown_catego
    rical, unknown_categorical_method, pca, pca_method, pca_components, i
    gnore_low_variance, combine_rare_levels, rare_level_threshold, bin_num
    eric_features, remove_outliers, outliers_threshold, remove_multicollin
    earity, multicollinearity_threshold, remove_perfect_collinearity, crea
    te_clusters, cluster_iter, polynomial_features, polynomial_degree, tri
    gonometry_features, polynomial_threshold, group_features, group_names,
    feature_selection, feature_selection_threshold, feature_selection_meth
    od, feature_interaction, feature_ratio, interaction_threshold, fix_imb
    alance, fix_imbalance_method, data_split_shuffle, data_split_stratify,
    fold_strategy, fold, fold_shuffle, fold_groups, n_jobs, use_gpu, custo
    m_pipeline, html, session_id, log_experiment, experiment_name, log_plo
    ts, log_profile, log_data, silent, verbose, profile, profile_kwargs)
```

```
578         log_plots = ["auc", "confusion_matrix", "feature"]
```

```
579
```

```
--> 580         return pycaret.internal.tabular.setup(
```

```
581             ml_usecase="classification",
```

```
582             available_plots=available_plots,
```

```
~\anaconda3\lib\site-packages\pycaret\internal\tabular.py in setup(dat
    a, target, ml_usecase, available_plots, train_size, test_data, preproc
    ess, imputation_type, iterative_imputation_iters, categorical_feature
    s, categorical_imputation, categorical_iterative_imputer, ordinal_feat
    ures, high_cardinality_features, high_cardinality_method, numeric_feat
    ures, numeric_imputation, numeric_iterative_imputer, date_features, ig
    nore_features, normalize, normalize_method, transformation, transforma
    tion_method, handle_unknown_categorical, unknown_categorical_method, p
    ca, pca_method, pca_components, ignore_low_variance, combine_rare_leve
    ls, rare_level_threshold, bin_numeric_features, remove_outliers, outli
    ers_threshold, remove_multicollinearity, multicollinearity_threshold,
    remove_perfect_collinearity, create_clusters, cluster_iter, polynomial
    _features, polynomial_degree, trigonometry_features, polynomial_thresh
    old, group_features, group_names, feature_selection, feature_selection
    _threshold, feature_selection_method, feature_interaction, feature_rat
    io, interaction_threshold, fix_imbalance, fix_imbalance_method, transf
    orm_target, transform_target_method, data_split_shuffle, data_split_st
    ratify, fold_strategy, fold, fold_shuffle, fold_groups, n_jobs, use_gp
    u, custom_pipeline, html, session_id, log_experiment, experiment_name,
    log_plots, log_profile, log_data, silent, verbose, profile, profile_kw
    args, display)
```

```
1215         )
```

```
1216         elif fold_strategy == "stratifiedkfold":
```

```

-> 1217         fold_generator = StratifiedKFold(
1218             fold_param, random_state=seed, shuffle=fold_shuffle
e_param
1219         )

~\anaconda3\lib\site-packages\sklearn\utils\validation.py in inner_f(*
args, **kwargs)
    61         extra_args = len(args) - len(all_args)
    62         if extra_args <= 0:
--> 63             return f(*args, **kwargs)
    64
    65         # extra_args > 0

~\anaconda3\lib\site-packages\sklearn\model_selection\_split.py in __i
nit__(self, n_splits, shuffle, random_state)
    634     @_deprecate_positional_args
    635     def __init__(self, n_splits=5, *, shuffle=False, random_st
ate=None):
--> 636         super().__init__(n_splits=n_splits, shuffle=shuffle,
    637                         random_state=random_state)
    638

~\anaconda3\lib\site-packages\sklearn\utils\validation.py in inner_f(*
args, **kwargs)
    61         extra_args = len(args) - len(all_args)
    62         if extra_args <= 0:
--> 63             return f(*args, **kwargs)
    64
    65         # extra_args > 0

```

```
In [65]: df = pd.read_csv('prepped_churn_data4.csv', index_col='customerIndex')
```

```
Out[65]:
```

	churning	tenure	PhoneService	Contract	PaymentMethod	MonthlyCharges	TotalC
customerIndex							
1	0	1	0	0	2	29.85	
2	0	34	1	1	3	56.95	1
3	1	2	1	0	3	53.85	
4	0	45	0	1	0	42.30	1
5	1	2	1	0	2	70.70	
...
7039	0	24	1	1	3	84.80	1
7040	0	72	1	1	1	103.20	7
7041	0	11	0	0	2	29.60	
7042	1	4	1	0	3	74.40	
7043	0	66	1	2	0	105.65	6

In [68]: `automl = setup(df, target='churning')`

IntProgress(value=0, description='Processing: ', max=3)

Initiated 22:36:46

Status Preparing Data for Modeling

ValueError Traceback (most recent call last)

<ipython-input-68-194c8a84e3ae> in <module>

----> 1 automl = setup(df, target='churning')

```
~\anaconda3\lib\site-packages\pycaret\classification.py in setup(data,
target, train_size, test_data, preprocess, imputation_type, iterative_
imputation_iters, categorical_features, categorical_imputation, catego
rical_iterative_imputer, ordinal_features, high_cardinality_features,
high_cardinality_method, numeric_features, numeric_imputation, numeric
_iterative_imputer, date_features, ignore_features, normalize, normali
ze_method, transformation, transformation_method, handle_unknown_categ
orical, unknown_categorical_method, pca, pca_method, pca_components, i
gnore_low_variance, combine_rare_levels, rare_level_threshold, bin_num
eric_features, remove_outliers, outliers_threshold, remove_multicollin
earity, multicollinearity_threshold, remove_perfect_collinearity, crea
te_clusters, cluster_iter, polynomial_features, polynomial_degree, tri
gonometry_features, polynomial_threshold, group_features, group_names,
feature_selection, feature_selection_threshold, feature_selection_meth
od, feature_interaction, feature_ratio, interaction_threshold, fix_imb
alance, fix_imbalance_method, data_split_shuffle, data_split_stratify,
fold_strategy, fold, fold_shuffle, fold_groups, n_jobs, use_gpu, custo
m_pipeline, html, session_id, log_experiment, experiment_name, log_plo
ts, log_profile, log_data, silent, verbose, profile, profile_kwargs)
    578         log_plots = ["auc", "confusion_matrix", "feature"]
    579
--> 580     return pycaret.internal.tabular.setup(
    581         ml_usecase="classification",
    582         available_plots=available_plots,
```

```
~\anaconda3\lib\site-packages\pycaret\internal\tabular.py in setup(dat
a, target, ml_usecase, available_plots, train_size, test_data, preproc
ess, imputation_type, iterative_imputation_iters, categorical_feature
s, categorical_imputation, categorical_iterative_imputer, ordinal_feat
ures, high_cardinality_features, high_cardinality_method, numeric_feat
ures, numeric_imputation, numeric_iterative_imputer, date_features, ig
nore_features, normalize, normalize_method, transformation, transforma
tion_method, handle_unknown_categorical, unknown_categorical_method, p
ca, pca_method, pca_components, ignore_low_variance, combine_rareleve
ls, rare_level_threshold, bin_numeric_features, remove_outliers, outli
ers_threshold, remove_multicollinearity, multicollinearity_threshold,
remove_perfect_collinearity, create_clusters, cluster_iter, polynomial
_features, polynomial_degree, trigonometry_features, polynomial_thresh
old, group_features, group_names, feature_selection, feature_selection
_threshold, feature_selection_method, feature_interaction, feature_rat
```



```

io, interaction_threshold, fix_imbalance, fix_imbalance_method, transf
orm_target, transform_target_method, data_split_shuffle, data_split_st
ratify, fold_strategy, fold, fold_shuffle, fold_groups, n_jobs, use_gp
u, custom_pipeline, html, session_id, log_experiment, experiment_name,
log_plots, log_profile, log_data, silent, verbose, profile, profile_kw
args, display)
1215         )
1216     elif fold_strategy == "stratifiedkfold":
-> 1217         fold_generator = StratifiedKFold(
1218             fold_param, random_state=seed, shuffle=fold_shuffle
e_param
1219         )

```

```

~\anaconda3\lib\site-packages\sklearn\utils\validation.py in inner_f(*
args, **kwargs)
61         extra_args = len(args) - len(all_args)
62         if extra_args <= 0:
---> 63             return f(*args, **kwargs)
64
65         # extra_args > 0

```

```

~\anaconda3\lib\site-packages\sklearn\model_selection\_split.py in __i
nit__(self, n_splits, shuffle, random_state)
634     @deprecate_positional_args
635     def __init__(self, n_splits=5, *, shuffle=False, random_st
ate=None):
--> 636         super().__init__(n_splits=n_splits, shuffle=shuffle,
637                             random_state=random_state)
638

```

```

~\anaconda3\lib\site-packages\sklearn\utils\validation.py in inner_f(*
args, **kwargs)
61         extra_args = len(args) - len(all_args)
62         if extra_args <= 0:
---> 63             return f(*args, **kwargs)
64
65         # extra_args > 0

```

```

~\anaconda3\lib\site-packages\sklearn\model_selection\_split.py in __i
nit__(self, n_splits, shuffle, random_state)
288
289     if not shuffle and random_state is not None: # None i
s the default
--> 290         raise ValueError(
291             'Setting a random_state has no effect since sh
uffle is '
292             'False. You should leave '

```

ValueError: Setting a random_state has no effect since shuffle is False. You should leave random_state to its default (None), or set shuffle=True.

```
In [84]: df2 = pd.read_excel('prepped_churn_data4.xls', index_col='customerIndex')
df2
```

Out[84]:

	churning	tenure	PhoneService	Contract	PaymentMethod	MonthlyCharges	Tota
customerIndex							
1	0	1	0	0	2	29.85	
2	0	34	1	1	3	56.95	
3	1	2	1	0	3	53.85	
4	0	45	0	1	0	42.30	
5	1	2	1	0	2	70.70	
...
7039	0	24	1	1	3	84.80	
7040	0	72	1	1	1	103.20	
7041	0	11	0	0	2	29.60	
7042	1	4	1	0	3	74.40	
7043	0	66	1	2	0	105.65	

```
In [86]: autml = setup(df2, target='churning', feature_selection='Relief', feature_c
IntProgress(value=0, description='Processing: ', max=3)
```

Initiated	22:36:46
Status	Preparing Data for Modeling

ValueError

Traceback (most recent call
 last)

<ipython-input-86-a5f553fa5935> in <module>

```
----> 1 automl = setup(df2, target='churning', feature_selection=False,
feature_selection_threshold=0.5)
```

```
~\anaconda3\lib\site-packages\pycaret\classification.py in setup(data,
target, train_size, test_data, preprocess, imputation_type, iterative_
imputation_iters, categorical_features, categorical_imputation, catego
rical_iterative_imputer, ordinal_features, high_cardinality_features,
high_cardinality_method, numeric_features, numeric_imputation, numeric
_iterative_imputer, date_features, ignore_features, normalize, normali
ze_method, transformation, transformation_method, handle_unknown_catego
rical, unknown_categorical_method, pca, pca_method, pca_components, i
gnore_low_variance, combine_rare_levels, rare_level_threshold, bin_num
eric_features, remove_outliers, outliers_threshold, remove_multicollin
earity, multicollinearity_threshold, remove_perfect_collinearity, crea
te_clusters, cluster_iter, polynomial_features, polynomial_degree, tri
gonometry_features, polynomial_threshold, group_features, group_names,
feature_selection, feature_selection_threshold, feature_selection_meth
od, feature_interaction, feature_ratio, interaction_threshold, fix_imb
alance, fix_imbalance_method, data_split_shuffle, data_split_stratify,
fold_strategy, fold, fold_shuffle, fold_groups, n_jobs, use_gpu, custo
m_pipeline, html, session_id, log_experiment, experiment_name, log_plo
ts, log_profile, log_data, silent, verbose, profile, profile_kwargs)
    578         log_plots = ["auc", "confusion_matrix", "feature"]
    579
--> 580     return pycaret.internal.tabular.setup(
    581         ml_usecase="classification",
    582         available_plots=available_plots,
```

```
~\anaconda3\lib\site-packages\pycaret\internal\tabular.py in setup(dat
a, target, ml_usecase, available_plots, train_size, test_data, preproc
ess, imputation_type, iterative_imputation_iters, categorical_feature
s, categorical_imputation, categorical_iterative_imputer, ordinal feat
ures, high_cardinality_features, high_cardinality_method, numeric feat
ures, numeric_imputation, numeric_iterative_imputer, date_features, ig
nore_features, normalize, normalize_method, transformation, transforma
tion_method, handle_unknown_categorical, unknown_categorical_method, p
ca, pca_method, pca_components, ignore_low_variance, combine_rare leve
ls, rare_level_threshold, bin_numeric_features, remove_outliers, outli
ers_threshold, remove_multicollinearity, multicollinearity_threshold,
remove_perfect_collinearity, create_clusters, cluster_iter, polynomial
_features, polynomial_degree, trigonometry_features, polynomial_thresh
old, group_features, group_names, feature_selection, feature_selection
_threshold, feature_selection_method, feature_interaction, feature_rat
io, interaction_threshold, fix_imbalance, fix_imbalance_method, transfo
rm_target, transform_target_method, data_split_shuffle, data_split_st
ratify, fold_strategy, fold, fold_shuffle, fold_groups, n_jobs, use_gp
u, custom_pipeline, html, session_id, log_experiment, experiment_name,
log_plots, log_profile, log_data, silent, verbose, profile, profile_kw
args, display)
    1215         )
    1216         elif fold_strategy == "stratifiedkfold":
```

```
-> 1217         fold_generator = StratifiedKFold(
1218             fold_param, random_state=seed, shuffle=fold_shuffle
e_param
1219         )

~\anaconda3\lib\site-packages\sklearn\utils\validation.py in inner_f(*
args, **kwargs)
    61         extra_args = len(args) - len(all_args)
    62         if extra_args <= 0:
--> 63             return f(*args, **kwargs)
    64
    65         # extra_args > 0

~\anaconda3\lib\site-packages\sklearn\model_selection\_split.py in __i
nit__(self, n_splits, shuffle, random_state)
    634     @_deprecate_positional_args
    635     def __init__(self, n_splits=5, *, shuffle=False, random_st
ate=None):
```

In []:

In []:

Summary

In []:

Write a short summary of the process and results here.