ULL Pre-processing Mitigation

January 13, 2025

1 Install

[1]:
```
!pip install aif360
!pip install fairlearn
```

```
Collecting aif360
  Downloading aif360-0.6.1-py3-none-any.whl.metadata (5.0 kB)
Requirement already satisfied: numpy>=1.16 in /usr/local/lib/python3.10/dist-
packages (from aif360) (1.26.4)
Requirement already satisfied: scipy>=1.2.0 in /usr/local/lib/python3.10/dist-
packages (from aif360) (1.13.1)
Requirement already satisfied: pandas>=0.24.0 in /usr/local/lib/python3.10/dist-
packages (from aif360) (2.2.2)
Requirement already satisfied: scikit-learn>=1.0 in
/usr/local/lib/python3.10/dist-packages (from aif360) (1.6.0)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.10/dist-
packages (from aif360) (3.10.0)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.10/dist-packages (from pandas>=0.24.0->aif360) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-
packages (from pandas>=0.24.0->aif360) (2024.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.10/dist-
packages (from pandas>=0.24.0->aif360) (2024.2)
Requirement already satisfied: joblib>=1.2.0 in /usr/local/lib/python3.10/dist-
packages (from scikit-learn>=1.0->aif360) (1.4.2)
Requirement already satisfied: threadpoolctl>=3.1.0 in
/usr/local/lib/python3.10/dist-packages (from scikit-learn>=1.0->aif360) (3.5.0)
Requirement already satisfied: contourpy>=1.0.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib->aif360) (1.3.1)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-
packages (from matplotlib->aif360) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
/usr/local/lib/python3.10/dist-packages (from matplotlib->aif360) (4.55.3)
Requirement already satisfied: kiwisolver>=1.3.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib->aif360) (1.4.8)
Requirement already satisfied: packaging>=20.0 in
/usr/local/lib/python3.10/dist-packages (from matplotlib->aif360) (24.2)
Requirement already satisfied: pillow>=8 in /usr/local/lib/python3.10/dist-
```

```
packages (from matplotlib->aif360) (11.1.0)
Requirement already satisfied: pyparsing>=2.3.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib->aif360) (3.2.1)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-
packages (from python-dateutil>=2.8.2->pandas>=0.24.0->aif360) (1.17.0)
Downloading aif360-0.6.1-py3-none-any.whl (259 kB)
                         259.7/259.7 kB
4.4 MB/s eta 0:00:00
Installing collected packages: aif360
Successfully installed aif360-0.6.1
Collecting fairlearn
  Downloading fairlearn-0.12.0-py3-none-any.whl.metadata (7.0 kB)
Requirement already satisfied: numpy>=1.24.4 in /usr/local/lib/python3.10/dist-
packages (from fairlearn) (1.26.4)
Requirement already satisfied: pandas>=2.0.3 in /usr/local/lib/python3.10/dist-
packages (from fairlearn) (2.2.2)
Requirement already satisfied: scikit-learn>=1.2.1 in
/usr/local/lib/python3.10/dist-packages (from fairlearn) (1.6.0)
Requirement already satisfied: scipy>=1.9.3 in /usr/local/lib/python3.10/dist-
packages (from fairlearn) (1.13.1)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.10/dist-packages (from pandas>=2.0.3->fairlearn) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-
packages (from pandas>=2.0.3->fairlearn) (2024.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.10/dist-
packages (from pandas>=2.0.3->fairlearn) (2024.2)
Requirement already satisfied: joblib>=1.2.0 in /usr/local/lib/python3.10/dist-
packages (from scikit-learn>=1.2.1->fairlearn) (1.4.2)
Requirement already satisfied: threadpoolctl>=3.1.0 in
/usr/local/lib/python3.10/dist-packages (from scikit-learn>=1.2.1->fairlearn)
(3.5.0)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-
packages (from python-dateutil>=2.8.2->pandas>=2.0.3->fairlearn) (1.17.0)
Downloading fairlearn-0.12.0-py3-none-any.whl (240 kB)
                         240.0/240.0 kB
3.7 MB/s eta 0:00:00
Installing collected packages: fairlearn
Successfully installed fairlearn-0.12.0
```

2 Imports

[3]:
```python
import os
import copy
import math
import json
import collections
```

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.preprocessing import OrdinalEncoder
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split, cross_validate,␣
 ↪StratifiedKFold
from sklearn.metrics import get_scorer
from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay

from aif360.algorithms.preprocessing import LFR
from aif360.algorithms.inprocessing import MetaFairClassifier,␣
 ↪PrejudiceRemover, GerryFairClassifier
from fairlearn import metrics
```

```
/usr/local/lib/python3.10/dist-packages/scipy/__init__.py:146: UserWarning: A
NumPy version >=1.17.3 and <1.25.0 is required for this version of SciPy
(detected version 1.26.4
  warnings.warn(f"A NumPy version >={np_minversion} and <{np_maxversion}"
WARNING:root:No module named 'inFairness': SenSeI and SenSR will be unavailable.
To install, run:
pip install 'aif360[inFairness]'
```

3 Settings

[36]:
```python
final_df_filename = "final_df_v1"
```

[63]:
```python
sensitive_feature = "f_ESCS"
favorable_sensitive_label = "ADVANTAGED"
unfavorable_sensitive_label = "DISADVANTAGED"

class_feature = "level_MAT"
favorable_class_label = "PASSED"
unfavorable_class_label = "NOT PASSED"

perf_metrics = ["balanced_accuracy", "precision", "recall", "roc_auc", "f1"]
fair_metrics = ["demographic_parity_ratio", "equalized_odds_ratio"]
```

4 Utils

[7]:
```python
from sklearn.base import BaseEstimator, TransformerMixin
from sklearn.preprocessing import OneHotEncoder, OrdinalEncoder
from aif360.datasets import BinaryLabelDataset
from aif360.metrics import BinaryLabelDatasetMetric
```

```
[59]: class AIF360SklearnWrapper(BaseEstimator, TransformerMixin):
    def __init__(
        self,
        aif360_model,
        sensitive_feature,
        favorable_class_label,
        unfavorable_class_label,
        favorable_sens_group,
        unfavorable_sens_group,
    ):
        self.aif360_model = aif360_model

        self.sensitive_feature = sensitive_feature
        self.favorable_sens_group = favorable_sens_group
        self.unfavorable_sens_group = unfavorable_sens_group

        self.class_feature = "label"
        self.favorable_class_label = favorable_class_label
        self.unfavorable_class_label = unfavorable_class_label

        self.ordinal_encoder = None

    def _encode_features(self, df, fit):

        df_encoded = df.copy()

        # Encode class
        if self.class_feature in df_encoded.columns:
            df_encoded[self.class_feature] = df_encoded[
                self.class_feature
            ].apply(lambda x: 1 if x == self.favorable_class_label else 0)
        else:
            df_encoded[self.class_feature] = 0  # or any constant value

        # Encode sensitive feature
        if self.sensitive_feature in df_encoded.
↪select_dtypes(include=['object', 'category']).columns:
            df_encoded[self.sensitive_feature] = df_encoded[
                self.sensitive_feature
            ].apply(lambda x: 1 if x == self.favorable_sens_group else 0)


        # Encode other columns
        categorical_cols = df_encoded.select_dtypes(include=['object',␣
↪'category']).columns
        numeric_cols = df_encoded.select_dtypes(exclude=['object', 'category']).
↪columns
```

```
        if fit:
            self.ordinal_encoder = OrdinalEncoder()
            self.ordinal_encoder.fit(df_encoded[categorical_cols])

        if len(categorical_cols) > 0:
            df_encoded[categorical_cols] = self.ordinal_encoder.
↪transform(df_encoded[categorical_cols])
        if len(numeric_cols) > 0:
            df_encoded[numeric_cols] = df_encoded[numeric_cols]

        # Check missing values
        # print(df_encoded.isna().sum())
        # df_encoded = df_encoded.set_index(self.sensitive_feature)
        # print(df_encoded.isna().sum())
        # print(df_encoded)

        return df_encoded

    def fit(self, X, y, **kwargs):
        if not self.sensitive_feature:
            raise ValueError("Protected attribute name must be specified.")

        df_encoded = self._encode_features(df=X.assign(label=y), fit=True)

        # print(df_encoded)
        dataset = BinaryLabelDataset(
            df=df_encoded,
            label_names=[self.class_feature],
            # sensitive_features=[self.sensitive_feature],
            protected_attribute_names=[self.sensitive_feature],
            favorable_label=1,
            unfavorable_label=0
        )
        self.aif360_model = self.aif360_model.fit(dataset, **kwargs)
        return self

    def transform(self, X, decode=False):
        X_encoded = self._encode_features(df=X, fit=False)
        dataset = BinaryLabelDataset(
            df=X_encoded,
            label_names=[self.class_feature],
            # sensitive_features=[self.sensitive_feature],
            protected_attribute_names=[self.sensitive_feature],
            favorable_label=1,
            unfavorable_label=0
        )
```

```
        new_X = self.aif360_model.transform(dataset).convert_to_dataframe()[0]
        new_X[self.sensitive_feature] = new_X[self.sensitive_feature].
↪apply(lambda x: self.favorable_sens_group if x == 1 else self.
↪unfavorable_sens_group) if decode else new_X[self.sensitive_feature]
        return new_X
```

[9]:
```
def encode_feature(df, feature, favorable_label):

    df_encoded = df.copy()

    # Encode class
    if feature in df_encoded.columns:
        df_encoded[feature] = df_encoded[
            feature
        ].apply(lambda x: 1 if x == favorable_label else 0)
    return df_encoded
```

[10]:
```
def decode_feature(df, feature, favorable_label, unfavorable_label):

    df_decoded = df.copy()

    # Encode class
    if feature in df_decoded.columns:
        df_decoded[feature] = df_decoded[
            feature
        ].apply(lambda x: favorable_label if x == 1 else unfavorable_label)
    return df_decoded
```

[11]:
```
def compute_fair_metric(
    fair_metric_name, sensitive_feature, X, y_true, y_pred
):

    # metrics_module = __import__("metrics")
    metrics_module = globals()["metrics"]
    fair_metric_scorer = getattr(metrics_module, fair_metric_name)

    X_sensitive = encode_feature(X, sensitive_feature,␣
↪favorable_label=favorable_sensitive_label)
    X_sensitive = X_sensitive[sensitive_feature]

    return fair_metric_scorer(
        y_true=y_true,
        y_pred=y_pred,
        sensitive_features=X_sensitive,
    )
```

5 Src

[39]:
```
final_df = pd.read_csv(f"{final_df_filename}.csv").set_index("id_questionnaire")
final_df
```

[39]:
```
                  s_frequency_of_work_with_teachers  \
id_questionnaire
1                                          0.733333
3                                          0.833333
5                                          0.566667
6                                          0.533333
8                                          0.633333
…                                               …
20421                                      0.566667
20422                                      0.700000
20423                                      0.400000
20424                                      0.500000
20427                                      0.666667

                  s_frequency_of_evaluations  s_frequency_of_internet_usage  \
id_questionnaire
1                                   0.740741                       0.066667
3                                   0.750000                       0.666667
5                                   0.888889                       0.133333
6                                   0.851852                       0.133333
8                                   0.703704                       0.266667
…                                        …                              …
20421                               0.777778                       0.600000
20422                               0.518519                       0.266667
20423                               0.555556                       0.400000
20424                               0.555556                       0.200000
20427                               0.740741                       0.133333

                  s_frequency_of_materials_in_class    f_ESCS  \
id_questionnaire
1                                          0.380952  0.235340
3                                          1.000000  0.261451
5                                          0.666667  3.151773
6                                          0.380952 -1.627731
8                                          0.666667 -0.358498
…                                               …        …
20421                                      0.476190 -0.764891
20422                                      0.380952 -0.048524
20423                                      0.571429  0.072922
20424                                      0.523810  2.755881
20427                                      0.476190  1.372627
```

```
                 s_extent_of_classmates_affinity  \
id_questionnaire
1                                       0.714286
3                                       0.833333
5                                       0.952381
6                                       0.761905
8                                       0.857143
…                                          …
20421                                   0.904762
20422                                   0.952381
20423                                   0.809524
20424                                   0.904762
20427                                   0.857143

                 s_extent_of_teacher_performance  s_extent_of_class_env  \
id_questionnaire
1                                       0.833333               0.791667
3                                       0.800000               0.833333
5                                       1.000000               1.000000
6                                       0.866667               1.000000
8                                       0.966667               0.833333
…                                          …                     …
20421                                   1.000000               1.000000
20422                                   0.888889               0.916667
20423                                   1.000000               0.666667
20424                                   0.766667               0.791667
20427                                   0.900000               0.916667

                 f_mother_age  s_frequency_of_computer_usage  \
id_questionnaire
1                        43.0                       0.111111
3                        45.0                       0.666667
5                        39.0                       0.000000
6                        25.0                       0.222222
8                        37.0                       0.333333
…                       …                            …
20421                    41.0                       0.111111
20422                    36.0                       0.222222
20423                    46.0                       0.111111
20424                    41.0                       0.111111
20427                    37.0                       0.000000

                 f_number_of_homework_hours_a_week  \
id_questionnaire
1                                             14.0
3                                              9.0
5                                              6.0
```

```
6                                                5.0
8                                               10.0
…                                                 …
20421                                           14.0
20422                                            2.0
20423                                            8.0
20424                                            2.0
20427                                            1.0

                  s_extent_of_school_satisfaction  f_mother_education_level  \
id_questionnaire
1                                        0.944444                       7.0
3                                        0.888889                       3.0
5                                        1.000000                       9.0
6                                        1.000000                       3.0
8                                        0.833333                       7.0
…                                             …                        …
20421                                    1.000000                       5.0
20422                                    0.944444                       4.0
20423                                    1.000000                       3.0
20424                                    0.833333                       9.0
20427                                    1.000000                       5.0

                  f_extent_of_family_satisfaction  f_number_of_tech_at_home  \
id_questionnaire
1                                        0.700000                       3.0
3                                        0.291667                       4.0
5                                        0.833333                       7.0
6                                        0.833333                       3.0
8                                        1.000000                       3.0
…                                             …                        …
20421                                    1.000000                       4.0
20422                                    0.708333                       5.0
20423                                    0.833333                      10.0
20424                                    0.666667                      10.0
20427                                    0.291667                       5.0

                  f_frequency_of_books_at_home  \
id_questionnaire
1                                     1.000000
3                                     0.444444
5                                     0.555556
6                                     0.333333
8                                     0.555556
…                                          …
20421                                 0.222222
20422                                 0.555556
```

```
20423                                    0.444444
20424                                    0.555556
20427                                    0.444444

                  f_frequency_of_parent_involved_in_school_activities  \
id_questionnaire
1                                                           0.333333
3                                                           0.166667
5                                                           0.250000
6                                                           0.083333
8                                                           0.333333
…                                                                …
20421                                                       0.333333
20422                                                       0.000000
20423                                                       0.000000
20424                                                       0.416667
20427                                                       0.000000

                 s_gender   level_MAT
id_questionnaire
1                  FEMALE  NOT PASSED
3                  FEMALE  NOT PASSED
5                  FEMALE      PASSED
6                  FEMALE  NOT PASSED
8                    MALE  NOT PASSED
…                     …          …
20421                MALE  NOT PASSED
20422              FEMALE  NOT PASSED
20423                MALE  NOT PASSED
20424                MALE  NOT PASSED
20427              FEMALE  NOT PASSED

[10735 rows x 19 columns]
```

[65]:
```
final_df[sensitive_feature] = final_df[sensitive_feature].apply(lambda x:␣
 ↪favorable_sensitive_label if x > 1.5 else unfavorable_sensitive_label)
final_df
```

[65]:
```
                  s_frequency_of_work_with_teachers  \
id_questionnaire
1                                          0.733333
3                                          0.833333
5                                          0.566667
6                                          0.533333
8                                          0.633333
…                                                …
20421                                      0.566667
```

20422	0.700000
20423	0.400000
20424	0.500000
20427	0.666667

	s_frequency_of_evaluations	s_frequency_of_internet_usage \
id_questionnaire		
1	0.740741	0.066667
3	0.750000	0.666667
5	0.888889	0.133333
6	0.851852	0.133333
8	0.703704	0.266667
…	…	…
20421	0.777778	0.600000
20422	0.518519	0.266667
20423	0.555556	0.400000
20424	0.555556	0.200000
20427	0.740741	0.133333

	s_frequency_of_materials_in_class	f_ESCS \
id_questionnaire		
1	0.380952	DISADVANTAGED
3	1.000000	DISADVANTAGED
5	0.666667	ADVANTAGED
6	0.380952	DISADVANTAGED
8	0.666667	DISADVANTAGED
…	…	…
20421	0.476190	DISADVANTAGED
20422	0.380952	DISADVANTAGED
20423	0.571429	DISADVANTAGED
20424	0.523810	ADVANTAGED
20427	0.476190	DISADVANTAGED

	s_extent_of_classmates_affinity \
id_questionnaire	
1	0.714286
3	0.833333
5	0.952381
6	0.761905
8	0.857143
…	…
20421	0.904762
20422	0.952381
20423	0.809524
20424	0.904762
20427	0.857143

```
                 s_extent_of_teacher_performance  s_extent_of_class_env  \
id_questionnaire
1                                       0.833333               0.791667
3                                       0.800000               0.833333
5                                       1.000000               1.000000
6                                       0.866667               1.000000
8                                       0.966667               0.833333
…                                          …                     …
20421                                   1.000000               1.000000
20422                                   0.888889               0.916667
20423                                   1.000000               0.666667
20424                                   0.766667               0.791667
20427                                   0.900000               0.916667

                 f_mother_age  s_frequency_of_computer_usage  \
id_questionnaire
1                        43.0                       0.111111
3                        45.0                       0.666667
5                        39.0                       0.000000
6                        25.0                       0.222222
8                        37.0                       0.333333
…                      …                             …
20421                    41.0                       0.111111
20422                    36.0                       0.222222
20423                    46.0                       0.111111
20424                    41.0                       0.111111
20427                    37.0                       0.000000

                 f_number_of_homework_hours_a_week  \
id_questionnaire
1                                             14.0
3                                              9.0
5                                              6.0
6                                              5.0
8                                             10.0
…                                            …
20421                                         14.0
20422                                          2.0
20423                                          8.0
20424                                          2.0
20427                                          1.0

                 s_extent_of_school_satisfaction  f_mother_education_level  \
id_questionnaire
1                                       0.944444                       7.0
3                                       0.888889                       3.0
5                                       1.000000                       9.0
```

```
6                                              1.000000                          3.0
8                                              0.833333                          7.0
…                                                 …                        …
20421                                          1.000000                          5.0
20422                                          0.944444                          4.0
20423                                          1.000000                          3.0
20424                                          0.833333                          9.0
20427                                          1.000000                          5.0

                  f_extent_of_family_satisfaction  f_number_of_tech_at_home  \
id_questionnaire
1                                        0.700000                       3.0
3                                        0.291667                       4.0
5                                        0.833333                       7.0
6                                        0.833333                       3.0
8                                        1.000000                       3.0
…                                           …                        …
20421                                    1.000000                       4.0
20422                                    0.708333                       5.0
20423                                    0.833333                      10.0
20424                                    0.666667                      10.0
20427                                    0.291667                       5.0

                  f_frequency_of_books_at_home  \
id_questionnaire
1                                     1.000000
3                                     0.444444
5                                     0.555556
6                                     0.333333
8                                     0.555556
…                                        …
20421                                 0.222222
20422                                 0.555556
20423                                 0.444444
20424                                 0.555556
20427                                 0.444444

                  f_frequency_of_parent_involved_in_school_activities  \
id_questionnaire
1                                                          0.333333
3                                                          0.166667
5                                                          0.250000
6                                                          0.083333
8                                                          0.333333
…                                                             …
20421                                                      0.333333
20422                                                      0.000000
```

```
20423                                                            0.000000
20424                                                            0.416667
20427                                                            0.000000

                 s_gender   level_MAT
id_questionnaire
1                  FEMALE  NOT PASSED
3                  FEMALE  NOT PASSED
5                  FEMALE      PASSED
6                  FEMALE  NOT PASSED
8                    MALE  NOT PASSED
…                     …          …
20421                MALE  NOT PASSED
20422              FEMALE  NOT PASSED
20423                MALE  NOT PASSED
20424                MALE  NOT PASSED
20427              FEMALE  NOT PASSED

[10735 rows x 19 columns]
```

[66]:
```python
get_features = lambda curr_df: [col for col in curr_df.columns if not col.
 ↪startswith("level_")]

X, y = final_df[get_features(final_df)], final_df[class_feature]

# y = y.apply(lambda elem: favorable_class_label if elem >= 3.5 else␣
 ↪unfavorable_class_label)
```

[95]:
```python
my_conf = {
     "k": 5,
     "Ax": 0.01,
     "Ay": 1.0,
     "Az": 50.0
}
mitigator = LFR(
        unprivileged_groups=[{sensitive_feature: 1}],
        privileged_groups=[{sensitive_feature: 0}],
        **my_conf
    )
wrapper = AIF360SklearnWrapper(
    aif360_model=mitigator,
    sensitive_feature=sensitive_feature,
    favorable_sens_group=favorable_sensitive_label,
    unfavorable_sens_group=unfavorable_sensitive_label,
    favorable_class_label=favorable_class_label,
    unfavorable_class_label=unfavorable_class_label,
)
```

```
# Fit the model
wrapper.fit(X, y)
X_t = wrapper.transform(X, decode=True)
transformed_df = pd.concat([X_t.reset_index(drop=True).drop("label", axis=1), y.
 ↪reset_index(drop=True)], axis=1)
# X_t = decode_feature(X_t, sensitive_feature, favorable_sensitive_label,␣
 ↪unfavorable_sensitive_label)
# X_t = decode_feature(X_t, "label", favorable_class_label,␣
 ↪unfavorable_class_label)
transformed_df
```

[95]:

```
       s_frequency_of_work_with_teachers  s_frequency_of_evaluations  \
0                               0.603402                    0.718262
1                               0.603402                    0.718262
2                               0.603402                    0.718262
3                               0.603414                    0.718256
4                               0.603402                    0.718262
…                                    …                         …
10730                           0.603402                    0.718262
10731                           0.603402                    0.718262
10732                           0.603402                    0.718262
10733                           0.603402                    0.718262
10734                           0.603402                    0.718262

       s_frequency_of_internet_usage  s_frequency_of_materials_in_class  \
0                           0.366703                           0.499796
1                           0.366703                           0.499796
2                           0.366703                           0.499796
3                           0.366722                           0.499819
4                           0.366703                           0.499796
…                                …                                …
10730                       0.366703                           0.499796
10731                       0.366703                           0.499796
10732                       0.366703                           0.499796
10733                       0.366703                           0.499796
10734                       0.366703                           0.499796

              f_ESCS  s_extent_of_classmates_affinity  \
0      DISADVANTAGED                          0.84255
1      DISADVANTAGED                          0.84255
2         ADVANTAGED                          0.84255
3      DISADVANTAGED                          0.84253
4      DISADVANTAGED                          0.84255
…                …                              …
10730  DISADVANTAGED                          0.84255
10731  DISADVANTAGED                          0.84255
```

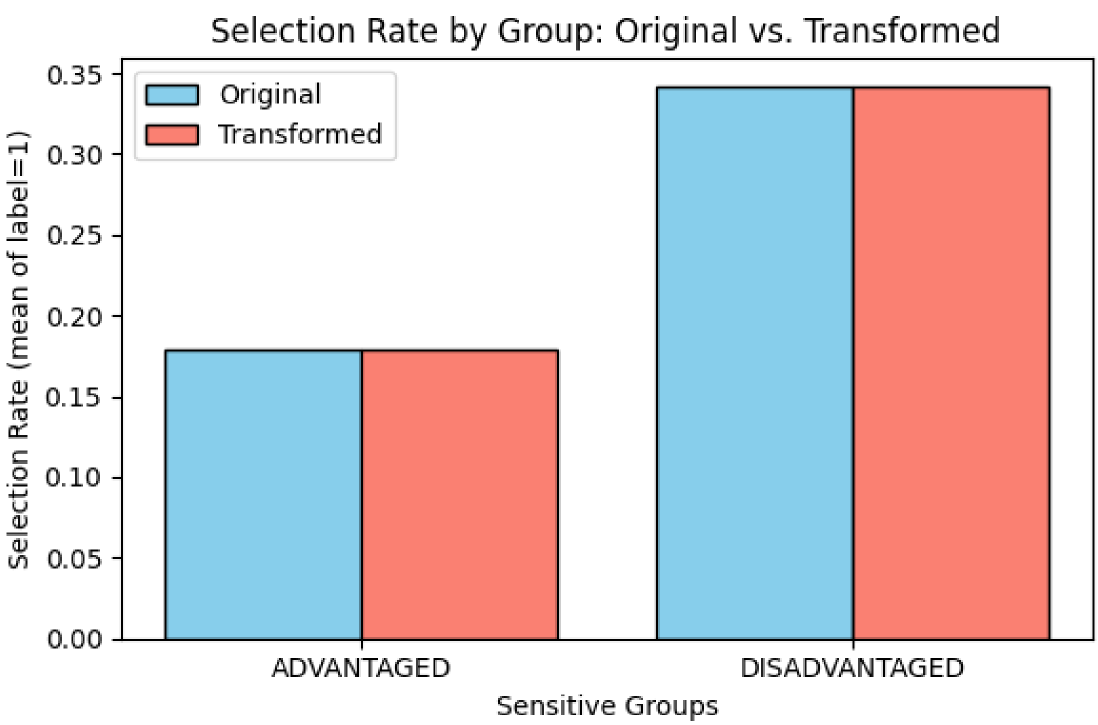
```
10732  DISADVANTAGED                          0.84255
10733    ADVANTAGED                           0.84255
10734  DISADVANTAGED                          0.84255

       s_extent_of_teacher_performance  s_extent_of_class_env  f_mother_age  \
0                             0.900597               0.805209     40.753695
1                             0.900597               0.805209     40.753695
2                             0.900597               0.805209     40.753695
3                             0.900560               0.805217     40.751982
4                             0.900597               0.805209     40.753695
…                                  …                    …            …
10730                         0.900597               0.805209     40.753695
10731                         0.900597               0.805209     40.753695
10732                         0.900597               0.805209     40.753695
10733                         0.900597               0.805209     40.753695
10734                         0.900597               0.805209     40.753695

       s_frequency_of_computer_usage  f_number_of_homework_hours_a_week  \
0                           0.251157                          7.348350
1                           0.251157                          7.348350
2                           0.251157                          7.348350
3                           0.251180                          7.348001
4                           0.251157                          7.348350
…                                …                               …
10730                       0.251157                          7.348350
10731                       0.251157                          7.348350
10732                       0.251157                          7.348350
10733                       0.251157                          7.348350
10734                       0.251157                          7.348350

       s_extent_of_school_satisfaction  f_mother_education_level  \
0                             0.942216                  5.435752
1                             0.942216                  5.435752
2                             0.942216                  5.435752
3                             0.942182                  5.435626
4                             0.942216                  5.435752
…                                  …                        …
10730                         0.942216                  5.435752
10731                         0.942216                  5.435752
10732                         0.942216                  5.435752
10733                         0.942216                  5.435752
10734                         0.942216                  5.435752

       f_extent_of_family_satisfaction  f_number_of_tech_at_home  \
0                             0.829478                  5.056625
1                             0.829478                  5.056625
2                             0.829478                  5.056625
```

```
3                              0.829471                  5.056480
4                              0.829478                  5.056625
…                                 …                         …
10730                          0.829478                  5.056625
10731                          0.829478                  5.056625
10732                          0.829478                  5.056625
10733                          0.829478                  5.056625
10734                          0.829478                  5.056625

       f_frequency_of_books_at_home  \
0                          0.570457
1                          0.570457
2                          0.570457
3                          0.570437
4                          0.570457
…                             …
10730                      0.570457
10731                      0.570457
10732                      0.570457
10733                      0.570457
10734                      0.570457

       f_frequency_of_parent_involved_in_school_activities  s_gender  \
0                                               0.359890    0.496403
1                                               0.359890    0.496403
2                                               0.359890    0.496403
3                                               0.359889    0.496395
4                                               0.359890    0.496403
…                                                  …           …
10730                                           0.359890    0.496403
10731                                           0.359890    0.496403
10732                                           0.359890    0.496403
10733                                           0.359890    0.496403
10734                                           0.359890    0.496403

        level_MAT
0      NOT PASSED
1      NOT PASSED
2          PASSED
3      NOT PASSED
4      NOT PASSED
…           …
10730  NOT PASSED
10731  NOT PASSED
10732  NOT PASSED
10733  NOT PASSED
10734  NOT PASSED
```

[10735 rows x 19 columns]

[99]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Suppose 's_gender' is your sensitive feature (e.g., "MALE"/"FEMALE"),
# and 'label' is binary (0/1) in both dataframes.
pre = encode_feature(final_df, sensitive_feature, favorable_sensitive_label)
post = encode_feature(transformed_df, sensitive_feature,␣
 ↪favorable_sensitive_label)

pre = encode_feature(pre, class_feature, favorable_class_label)
post = encode_feature(post, class_feature, favorable_class_label)

groups = pre[sensitive_feature].unique()  # or X_t["s_gender"].unique()

# 1) Compute selection rates in the original data
original_rates = []
for grp in groups:
    mask = (pre[sensitive_feature] == grp)
    group_data = pre[mask]
    if len(group_data) > 0:
        sel_rate = group_data[class_feature].mean()  # fraction of label=1
    else:
        sel_rate = np.nan
    original_rates.append(sel_rate)

# 2) Compute selection rates in the transformed data
transformed_rates = []
for grp in groups:
    mask = (post[sensitive_feature] == grp)
    group_data = post[mask]
    if len(group_data) > 0:
        sel_rate = group_data[class_feature].mean()
    else:
        sel_rate = np.nan
    transformed_rates.append(sel_rate)

# 3) Build a single DataFrame for plotting
df_plot = pd.DataFrame({
    "group": groups,
    "original_rate": original_rates,
    "transformed_rate": transformed_rates
})
```

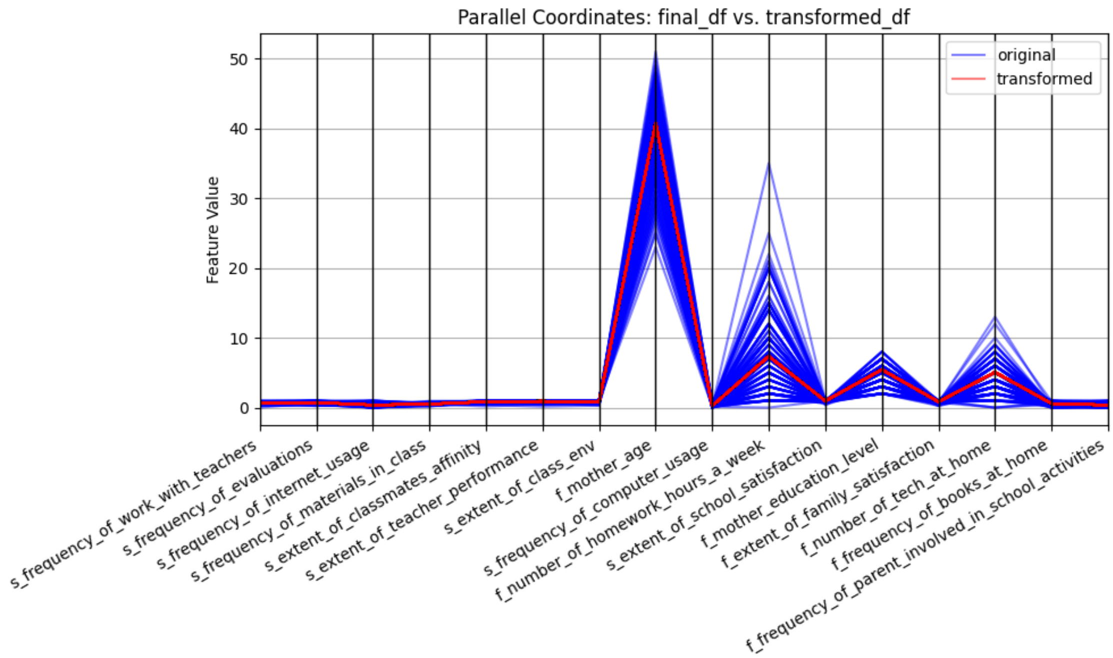
```
get_actual_elem = lambda elem: unfavorable_sensitive_label if elem == 1 else␣
↪favorable_sensitive_label
# 4) Plot side-by-side bars
plt.figure(figsize=(6, 4))

bar_width = 0.4
x_positions = np.arange(len(groups))

plt.bar(x_positions - bar_width/2, df_plot["original_rate"],
        width=bar_width, label="Original", color="skyblue", edgecolor="black")
plt.bar(x_positions + bar_width/2, df_plot["transformed_rate"],
        width=bar_width, label="Transformed", color="salmon", edgecolor="black")

plt.xticks(x_positions, [get_actual_elem(elem) for elem in df_plot["group"]])
plt.xlabel("Sensitive Groups")
plt.ylabel("Selection Rate (mean of label=1)")
plt.title("Selection Rate by Group: Original vs. Transformed")
plt.legend()
plt.tight_layout()
plt.show()
```



[97]:

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from pandas.plotting import parallel_coordinates

# 1) Identify numeric columns (excluding the label).
numeric_cols = final_df.select_dtypes(include=[np.number]).columns
if class_feature in numeric_cols:
    numeric_cols = numeric_cols.drop(class_feature)

# 2) (Optional) sample rows to avoid clutter if data is large
SAMPLE_SIZE = 200
final_sample = (
    final_df[numeric_cols]
    .reset_index(drop=True)
    .sample(n=min(SAMPLE_SIZE, len(final_df)), random_state=42)
    .copy()
)
trans_sample = (
    transformed_df[numeric_cols]
    .reset_index(drop=True)
    .sample(n=min(SAMPLE_SIZE, len(transformed_df)), random_state=42)
    .copy()
)

# 3) Tag each subset with a "source"
final_sample["source"] = "Original"
trans_sample["source"] = "Transformed"

# 4) Concatenate them
combined = pd.concat([final_sample, trans_sample], ignore_index=True)

# 5) Parallel Coordinates expects one categorical column (here "source") to␣
↪color the lines
plt.figure(figsize=(10, 6))
parallel_coordinates(
    combined,
    class_column="source",
    color=["blue", "red"],  # or other color palette
    alpha=0.5
)
plt.title("Parallel Coordinates: final_df vs. transformed_df")
plt.ylabel("Feature Value")

# Make feature names vertical to avoid overlap
plt.xticks(rotation=30, ha="right")
plt.tight_layout()
```

```
plt.show()
```



[98]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from pandas.plotting import andrews_curves

# 1) Identify numeric columns (excluding the label if it's numeric)
numeric_cols = final_df.select_dtypes(include=[np.number]).columns
if class_feature in numeric_cols:
    numeric_cols = numeric_cols.drop(class_feature)

# 2) Subset each DataFrame + add a "source" column
orig_numeric = final_df[numeric_cols].copy()
orig_numeric["source"] = "original"

trans_numeric = transformed_df[numeric_cols].copy()
trans_numeric["source"] = "transformed"

# 3) Concatenate them
combined = pd.concat([orig_numeric, trans_numeric], ignore_index=True)

# 4) Andrews Curves
plt.figure(figsize=(10, 6))
andrews_curves(
```

```
    frame=combined,
    class_column="source",
    alpha=0.7,
    colormap=plt.cm.Set2  # or another colormap
)
plt.title("Andrews Curves: final_df vs. transformed_df")
plt.grid(True)
plt.tight_layout()
plt.show()
```