Analyzing Language Bias Between French and English in Conventional Multilingual Sentiment Analysis Models

April 1, 2024

1 Calculating Bias for Multilingual Support Vector Machine and Naive-Bayes for Sentiment Analysis

1.1 Library Imports

Libraries Used:

- **Sklearn**: Used to use the Multinomial Naive-Bayes and Support Vector Machine Model, build the Tf-Idf Matrix, use proper train test splitting, and build accuracy reports.
- **Pandas**: Used for building DataFrames
- **Numpy**: Provides operations for the DataFrames
- **FairLearn**: Builds specified bias metrics in models

Dependenices are available at **requirements.txt.**

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

[ ]:
```
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.svm import SVC
from sklearn.metrics import classification_report, accuracy_score
from sklearn.pipeline import Pipeline
from sklearn.naive_bayes import MultinomialNB
import pandas as pd
import numpy as np
from sklearn.metrics import accuracy_score, classification_report
from fairlearn.metrics import MetricFrame, demographic_parity_difference,␣
↪equalized_odds_difference, selection_rate, equalized_odds_ratio,␣
↪demographic_parity_ratio
from google.colab import drive
drive.mount('/content/drive')
```

1.2 Pre-Processing

Installing spaCy's French and English stop words and pre-processing tools, removing taggers as dataset is unordered.

```
[ ]: import spacy
     !python -m spacy download en_core_web_sm
     !python -m spacy download fr_core_news_sm
```

```
[ ]: nlp_en = spacy.load("en_core_web_sm", disable=["ner", "parser", "tagger"])
     nlp_fr = spacy.load("fr_core_news_sm", disable=["ner", "parser", "tagger"])
```

1.3 Pre-Processing Functions

- **Review Parsing** - Parses through the dataset and reconstructs the text from the word frequencies provided, along with the sentiment labels as well.
- **Batch Pre-Processing** - Pre-processes the text via spaCy using batch pre-processing with a batch size of 100.
- **Load Dataset** - The fully pre-process data is placed in a dataframe.

```
[ ]: def review_parsing(line):
         """
         Parses lines from the dataset to reconstruct the text by multiplying by the
         proper word frequencies and extracting the
         sentiment labels as well.

         Args:
           String: The lines in the dataset

         Returns:
           Dict {str:str}: A dictionary that has text and sentiment as the keys and
           the reconstructed text and sentiment as values.
         """
         words = []
         parts = line.strip().split()
         sentiment = None

         for part in parts:
             if part.startswith("#label#"):
                 sentiment = part.split(":")[1]
             else:
                 word, freq = part.split(":")
                 words.extend([word] * int(freq))

         reconstructed_text = " ".join(words)
         return {'text': reconstructed_text, 'sentiment': sentiment}

     def batch_preprocess_en(texts):
         """
         Batch pre-process English texts.

         Args:
```

```
      List[str]: All of the English texts.

    Return:
      List[str]: All of the English texts fully pre-processed.
    """
    processed_texts = []
    for doc in nlp_en.pipe(texts, batch_size=100):
        tokens = [token.text.lower() for token in doc if token.is_alpha and not␣
 ↪token.is_stop]
        processed_texts.append(' '.join(tokens))
    return processed_texts

def batch_preprocess_fr(texts):
    """
    Batch pre-process French texts.

    Args:
      List[str]: All of the French texts.

    Return:
      List[str]: All of the French texts fully pre-processed.
    """
    processed_texts = []
    for doc in nlp_fr.pipe(texts, batch_size=20):
        tokens = [token.text.lower() for token in doc if token.is_alpha and not␣
 ↪token.is_stop]
        processed_texts.append(' '.join(tokens))
    return processed_texts


def load_dataset_to_dataframe_en(file_path):
    """
    Transforming the English pre-processed data into a dataframe.

    Args:
      FILE: The CSV file with all the English data.

    Return:
      DataFrame: A dataframe that has the pre-processed texts along with the
      sentiments.
    """
    texts, sentiments = [], []

    with open(file_path, 'r') as file:
        for line in file:
            parsed_line = review_parsing(line)
            texts.append(parsed_line['text'])
```

```
            sentiments.append(parsed_line['sentiment'])

    processed_texts = batch_preprocess_en(texts)

    df = pd.DataFrame({
        'ProcessedText': processed_texts,
        'Sentiment': sentiments
    })

    return df

def load_dataset_to_dataframe_fr(file_path):
    """
    Transforming the French pre-processed data into a dataframe.

    Args:
      FILE: The CSV file with all the French data.

    Return:
      DataFrame: A dataframe that has the pre-processed texts along with the
      sentiments.
    """
    texts, sentiments = [], []

    with open(file_path, 'r') as file:
        for line in file:
            parsed_line = review_parsing(line)
            texts.append(parsed_line['text'])
            sentiments.append(parsed_line['sentiment'])

    processed_texts = batch_preprocess_fr(texts)

    df = pd.DataFrame({
        'ProcessedText': processed_texts,
        'Sentiment': sentiments
    })

    return df
```

1.4 Loading the Particular Datasets

The file path to the Webis-CLS-10 Dataset

[ ]:
```
FILE_PATH_EN = '/content/drive/My Drive/dataset/en/music/unlabeled.processed'
FILE_PATH_FR = '/content/drive/My Drive/dataset/fr/music/unlabeled.processed'

FILE_PATH_EN_dvd = '/content/drive/My Drive/dataset/en/dvd/unlabeled.processed'
```

```
FILE_PATH_FR_dvd = '/content/drive/My Drive/dataset/fr/dvd/unlabeled.processed'

FILE_PATH_EN_books = '/content/drive/My Drive/dataset/en/books/unlabeled.
 ↪processed'
FILE_PATH_FR_books = '/content/drive/My Drive/dataset/fr/books/unlabeled.
 ↪processed'
```

Loading the French and English dataset from the Webis-CLS-10 Dataset. We are taking the all three sub-categories of the dataset which are the dvd, music, and books categories.

[ ]:
```
data_en = load_dataset_to_dataframe_en(FILE_PATH_EN)
data_fr = load_dataset_to_dataframe_fr(FILE_PATH_FR)

data_en_dvd = load_dataset_to_dataframe_en(FILE_PATH_EN_dvd)
data_fr_dvd = load_dataset_to_dataframe_fr(FILE_PATH_FR_dvd)

data_en_books = load_dataset_to_dataframe_en(FILE_PATH_EN_books)
data_fr_books = load_dataset_to_dataframe_fr(FILE_PATH_FR_books)
```

1.5 Caching the Datasets

Caching the datasets so there is no need to keep pre-processing the data

[ ]:
```
data_en.to_csv("/content/drive/My Drive/dataset/en/music/unlabeled.csv",␣
 ↪index=False)
data_fr.to_csv("/content/drive/My Drive/dataset/fr/music/unlabeled.csv",␣
 ↪index=False)

data_en_dvd.to_csv("/content/drive/My Drive/dataset/en/dvd/unlabeled.csv",␣
 ↪index=False)
data_fr_dvd.to_csv("/content/drive/My Drive/dataset/fr/dvd/unlabeled.csv",␣
 ↪index=False)

data_en_books.to_csv("/content/drive/My Drive/dataset/en/books/unlabeled.csv",␣
 ↪index=False)
data_fr_books.to_csv("/content/drive/My Drive/dataset/fr/books/unlabeled.csv",␣
 ↪index=False)
```

[ ]:
```
data_en = pd.read_csv("/content/drive/My Drive/dataset/en/music/unlabeled.csv")
data_fr = pd.read_csv("/content/drive/My Drive/dataset/fr/music/unlabeled.csv")

data_en_dvd = pd.read_csv("/content/drive/My Drive/dataset/en/dvd/unlabeled.
 ↪csv")
data_fr_dvd = pd.read_csv("/content/drive/My Drive/dataset/fr/dvd/unlabeled.
 ↪csv")
```

```
data_en_books = pd.read_csv("/content/drive/My Drive/dataset/en/books/unlabeled.
 ↪csv")
data_fr_books = pd.read_csv("/content/drive/My Drive/dataset/fr/books/unlabeled.
 ↪csv")
```

[ ]:
```
print(data_fr_books.shape)
```

```
(32870, 2)
```

2 The Dataset Pre-Processed

[ ]:
```
display(data_en)
```

```
                                          ProcessedText Sentiment
0       pretty pretty pretty good good good records re…  negative
1       classical alive years narration peter children…  positive
2       chamillionaire chamillionaire chamillionaire c…  negative
3       perfect perfect giants giants world world linc…  positive
4       playing playing playing playing autoharp autoh…  positive
…                                                     …         …
25215   num num num num num num num num num num nu…  positive
25216   album album album excuses excuses words words …  positive
25217   destiny destiny music music stone stone beauti…  negative
25218   best best best tha tha tha rap rap heard heard…  positive
25219   lm cd wonder devil peace entitled selling kam …  negative

[25220 rows x 2 columns]
```

[ ]:
```
display(data_fr)
```

```
                                          ProcessedText Sentiment
0       ringard ringard ringard esrt aufray rêver fair…  negative
1       l l l d d jamais jamais n n peur indépendance …  negative
2       muse muse muse muse muse muse muse qu qu qu qu…  negative
3       num num num num l l l groupe groupe groupe évo…  negative
4       d d d d d d d sympa sympa sympa sympa sympa sy…  negative
…                                                     …         …
15935   d d rappelles dizaine j années cd petits adore…  positive
15936   vallenatos vallenatos vallenatos rien rien mod…  negative
15937   sambora sambora connaissait guitariste album p…  positive
15938   disco disco n n compil qualité énorme passez c…  negative
15939   médiocre cd traditionnels pistes trouve qu mus…  negative

[15940 rows x 2 columns]
```

2.1 Creating the Multi-Lingual Dataset

Sample Data A function that ensures equal number of Englsih and French Reviews in the Dataframe.

[ ]:
```
def sample_data(df_en, df_fr, perc_en, perc_fr):
    """
    Adjusts sampling to ensure an equal number of English and French samples,
    maximizing the amount of data used while respecting the specified␣
 ↪percentages.

    Args:
      Dataframe: The Pre-Processed English DataFrame
      Dataframe: The Pre-Proecessed French DataFrame
      Float: The percentage of English reviews in the dataset
      Float: The percentage of French reviews in the dataset

    Returns:
      Dataframe: The multilingual dataset
    """
    # Determine the maximum number of samples we can take equally from both␣
 ↪datasets
    max_samples_en = int(len(df_en))
    max_samples_fr = int(len(df_fr))

    max_total = max_samples_en + max_samples_fr

    # The actual number of samples to take from each is the minimum of these␣
 ↪two numbers
    if min(max_samples_en, max_samples_fr) < perc_en * max_total:
      actual_samples = min(max_samples_en, max_samples_fr)
    else:
      actual_samples = perc_en * max_total

    # Sample these numbers from each dataframe
    sample_en = df_en.sample(n=actual_samples, random_state=42)
    sample_fr = df_fr.sample(n=actual_samples, random_state=42)

    # Mark each sample with its language
    sample_en['Language'] = 'English'
    sample_fr['Language'] = 'French'

    # Combine and return the samples
    return pd.concat([sample_en, sample_fr], ignore_index=True)
```

2.1.1 Building the Tf-Idf Matrix

A Tf-Idf Matrix must be built to have the SVM and Naive Bayes to work.

```
[ ]: def preprocess_and_vectorize(df):
         """
         Pre-Processes and vectorizes the text data in the DataFrame.

         Args:
           DataFrame: The Multi-Lingual Dataset

         Returns:
           DataFrame: Tf-Idf of the shape of the samples and feature representing the
           vectorized text data.
           DataFrame: The sentiment labels associated with each text
           TfidfVectorizer: Contains the vocabulary and idf scores of each term
           NumpyArray: An array of every text entry's language
         """
         tfidf = TfidfVectorizer(max_features=10000)
         X = tfidf.fit_transform(df['ProcessedText'])
         y = df['Sentiment'].values
         return X, y, tfidf, df['Language'].values
```

2.2 Calculating the Specified Bias Metrics, Training the SVM and Naive Bayes, and Splitting the Data

Ensure Binary Labels: Ensures the specified labels are binary, so the bias metrics function works as intended.

Calculate Bias Metrics: Calculates the demographic parity ratiom equalized odds ratio, demographic parity difference, and equalized odds difference.

Map Labels: Maps the sentiment labels to have binary classification.

Train and Evaluate: Trains the SVM and Naive Bayes models, and calculating the corresponding precision, recall and f1-scores for each langugage. Also outputs the bias metrics for the models.

```
[ ]: def ensure_binary_labels(y):
         """
         Ensures the specified labels are binary, so the bias metrics function works
         as intended

         Args:
           DataFrame: The sentiment dataframe
         """
         unique_labels = np.unique(y)
         if set(unique_labels) == {0, 1} or set(unique_labels) == {-1, 1}:
             return np.where(y == -1, 0, y)
         else:
             raise ValueError("Labels must be binary and in {0, 1} or {-1, 1}.")

     def calculate_bias_metrics(y_true, y_pred, sensitive_features):
         """
```

```
    Calculates the demographic parity ratiom equalized odds ratio,
    demographic parity difference, and equalized odds difference.

    Args:
      DataFrame: Dataframe the contains the actual sentiment labels for the
      specified text in the dataset
      DataFrame: Dataframe the contains the predicted sentiment labels for the
      specified text in the dataset
      DataFrame: Contains the dataframe with the languages corresponding to
      the sentiment labels
    """
    y_true_binary = ensure_binary_labels(y_true)
    y_pred_binary = ensure_binary_labels(y_pred)

    m_dpr = demographic_parity_ratio(y_true_binary, y_pred_binary,␣
 ↪sensitive_features=sensitive_features)
    m_eqo = equalized_odds_ratio(y_true_binary, y_pred_binary,␣
 ↪sensitive_features=sensitive_features)
    m_dpr_2 = demographic_parity_difference(y_true_binary, y_pred_binary,␣
 ↪sensitive_features=sensitive_features)
    m_eqo_2 = equalized_odds_difference(y_true_binary, y_pred_binary,␣
 ↪sensitive_features=sensitive_features)

    print(f"The demographic parity ratio is {m_dpr}")
    print(f"The equalized odds ratio is {m_eqo}")
    print(f"The demographic parity differece is {m_dpr_2}")
    print(f"The equalized odds difference is {m_eqo_2}")

def map_labels(y):
    """
    Maps the positve and negative setiments to be binary

    Args:
      DataFrame: The dataset without binary labels for sentiment

    Return:
      DataFrame: The dataset with binary labels

    """

    return np.where(y == 'positive', 1, 0)

def train_and_evaluate(X_train, y_train, X_test, y_test, languages_test, model,␣
 ↪model_name="Model"):
    """
    Trains the SVM and Naive Bayes models, and calculating the corresponding
    precision, recall, and f1-scores for each language. Also outputs the
```

```
    bias metrics for the models.

    Args:
      DataFrame: The training dataframe with data other than sentiment
      DataFrame: The training dataframe with the sentiment data
      DataFrame: The testing dataframe with data other than sentiment
      DataFrame: The testing dataframe with the sentiment data
      DataFrame: A language dataframe that corresponds to the y_test sentiment␣
↪data
      SkLearn Model: The specified model to be trained
    """

    y_train_mapped = map_labels(y_train)
    y_test_mapped = map_labels(y_test)

    model.fit(X_train, y_train_mapped)
    y_pred_mapped = model.predict(X_test)

    bias_metrics = calculate_bias_metrics(y_test_mapped, y_pred_mapped,␣
↪languages_test)

    y_pred = np.where(y_pred_mapped == 1, 'positive', 'negative')
    print(f"Results for {model_name}:")
    print("Overall Accuracy:", accuracy_score(y_test, y_pred))
    print("Overall Classification Report:")
    print(classification_report(y_test, y_pred))

    for language in ['English', 'French']:
        idx = languages_test == language
        y_test_lang = y_test[idx]
        y_pred_lang = y_pred[idx]

        print(f"Accuracy on {language}: {accuracy_score(y_test_lang,␣
↪y_pred_lang)}")
        print(f"Classification Report for {language}:")
        print(classification_report(y_test_lang, y_pred_lang))

    print("-----------------------------------------------------")
```

3 Training the Music Data Using SVM and Naive Bayes

Splitting up the data, using an 80-20 training and validation split and calling the corresponding functions for a properly trained model.

[ ]:
```
df_sampled = sample_data(data_en, data_fr, 0.5, 0.5)
```

```
X, y, tfidf, languages = preprocess_and_vectorize(df_sampled)

X_train, X_test, y_train, y_test, languages_train, languages_test =␣
 ↪train_test_split(X, y, languages, test_size=0.2, random_state=42)
```

[ ]:
```
# Initialize models
svm_model = SVC(kernel='linear')
nb_model = MultinomialNB()

# Train and evaluate models, including language-specific performance
print("Evaluating SVM...")
train_and_evaluate(X_train, y_train, X_test, y_test, languages_test, svm_model,␣
 ↪"SVM")

print("Evaluating Naive Bayes...")
train_and_evaluate(X_train, y_train, X_test, y_test, languages_test, nb_model,␣
 ↪"Naive Bayes")
```

```
Evaluating SVM…
The demographic parity ratio is 0.9867867840290437
The equalized odds ratio is 0.5070733863837312
The demographic parity differece is 0.006797809078653905
The equalized odds difference is 0.07155386326687774
Results for SVM:
Overall Accuracy: 0.8922521957340025
Overall Classification Report:
              precision    recall  f1-score   support

    negative       0.89      0.89      0.89      3120
    positive       0.89      0.89      0.89      3256

    accuracy                           0.89      6376
   macro avg       0.89      0.89      0.89      6376
weighted avg       0.89      0.89      0.89      6376

Accuracy on English: 0.8666871921182266
Classification Report for English:
              precision    recall  f1-score   support

    negative       0.87      0.85      0.86      1612
    positive       0.86      0.88      0.87      1636

    accuracy                           0.87      3248
   macro avg       0.87      0.87      0.87      3248
weighted avg       0.87      0.87      0.87      3248

Accuracy on French: 0.9187979539641944
```

```
Classification Report for French:
              precision    recall  f1-score   support

    negative       0.91      0.93      0.92      1508
    positive       0.93      0.91      0.92      1620

    accuracy                           0.92      3128
   macro avg       0.92      0.92      0.92      3128
weighted avg       0.92      0.92      0.92      3128

--------------------------------------------------
Evaluating Naive Bayes…
The demographic parity ratio is 0.8326970359662453
The equalized odds ratio is 0.31491146318732527
The demographic parity differece is 0.09596226046640549
The equalized odds difference is 0.15724736887139557
Results for Naive Bayes:
Overall Accuracy: 0.8655897114178168
Overall Classification Report:
              precision    recall  f1-score   support

    negative       0.87      0.85      0.86      3120
    positive       0.86      0.88      0.87      3256

    accuracy                           0.87      6376
   macro avg       0.87      0.87      0.87      6376
weighted avg       0.87      0.87      0.87      6376

Accuracy on English: 0.8420566502463054
Classification Report for English:
              precision    recall  f1-score   support

    negative       0.90      0.77      0.83      1612
    positive       0.80      0.91      0.85      1636

    accuracy                           0.84      3248
   macro avg       0.85      0.84      0.84      3248
weighted avg       0.85      0.84      0.84      3248

Accuracy on French: 0.8900255754475703
Classification Report for French:
              precision    recall  f1-score   support

    negative       0.86      0.93      0.89      1508
    positive       0.93      0.85      0.89      1620

    accuracy                           0.89      3128
   macro avg       0.89      0.89      0.89      3128
```

```
weighted avg       0.89      0.89      0.89      3128

--------------------------------------------------
```

4 Training the DVD Data Using SVM and Naive-Bayes

```
[ ]: df_sampled = sample_data(data_en_dvd, data_fr_dvd, 0.5, 0.5)
X, y, tfidf, languages = preprocess_and_vectorize(df_sampled)

# Split the data, ensuring languages array is split consistently with X and y
X_train, X_test, y_train, y_test, languages_train, languages_test =␣
 ↪train_test_split(X, y, languages, test_size=0.2, random_state=42)

# Initialize models
svm_model = SVC(kernel='linear')
nb_model = MultinomialNB()

# Train and evaluate models, including language-specific performance
print("Evaluating SVM...")
train_and_evaluate(X_train, y_train, X_test, y_test, languages_test, svm_model,␣
 ↪"SVM")

print("Evaluating Naive Bayes...")
train_and_evaluate(X_train, y_train, X_test, y_test, languages_test, nb_model,␣
 ↪"Naive Bayes")
```

```
Evaluating SVM…
The demographic parity ratio is 0.9997391279034616
The equalized odds ratio is 0.8206236956293261
The demographic parity differece is 0.00013303111100959608
The equalized odds difference is 0.027384292368063945
Results for SVM:
Overall Accuracy: 0.8699252136752137
Overall Classification Report:
              precision    recall  f1-score   support

    negative       0.88      0.86      0.87      1870
    positive       0.86      0.88      0.87      1874

    accuracy                           0.87      3744
   macro avg       0.87      0.87      0.87      3744
weighted avg       0.87      0.87      0.87      3744

Accuracy on English: 0.8649214659685864
Classification Report for English:
              precision    recall  f1-score   support
```

```
    negative       0.88      0.85      0.87       976
    positive       0.85      0.88      0.86       934

    accuracy                           0.86      1910
   macro avg       0.87      0.87      0.86      1910
weighted avg       0.87      0.86      0.86      1910

Accuracy on French: 0.8751363140676118
Classification Report for French:
              precision    recall  f1-score   support

    negative       0.87      0.87      0.87       894
    positive       0.88      0.88      0.88       940

    accuracy                           0.88      1834
   macro avg       0.88      0.88      0.88      1834
weighted avg       0.88      0.88      0.88      1834

--------------------------------------------------
Evaluating Naive Bayes…
The demographic parity ratio is 0.9508331761373087
The equalized odds ratio is 0.6591532649529357
The demographic parity differece is 0.025149731368507566
The equalized odds difference is 0.055527285730003303
Results for Naive Bayes:
Overall Accuracy: 0.8624465811965812
Overall Classification Report:
              precision    recall  f1-score   support

    negative       0.86      0.86      0.86      1870
    positive       0.86      0.86      0.86      1874

    accuracy                           0.86      3744
   macro avg       0.86      0.86      0.86      3744
weighted avg       0.86      0.86      0.86      3744

Accuracy on English: 0.856020942408377
Classification Report for English:
              precision    recall  f1-score   support

    negative       0.88      0.84      0.86       976
    positive       0.84      0.88      0.86       934

    accuracy                           0.86      1910
   macro avg       0.86      0.86      0.86      1910
weighted avg       0.86      0.86      0.86      1910

Accuracy on French: 0.8691384950926936
```

```
Classification Report for French:
              precision    recall  f1-score   support

    negative       0.85      0.89      0.87       894
    positive       0.89      0.85      0.87       940

    accuracy                           0.87      1834
   macro avg       0.87      0.87      0.87      1834
weighted avg       0.87      0.87      0.87      1834

--------------------------------------------------
```

5 Training the Books Data Using SVM and Naive-Bayes

[ ]:
```
df_sampled = sample_data(data_en_books, data_fr_books, 0.5, 0.5)
X, y, tfidf, languages = preprocess_and_vectorize(df_sampled)

X_train, X_test, y_train, y_test, languages_train, languages_test =␣
 ↪train_test_split(X, y, languages, test_size=0.2, random_state=42)

svm_model = SVC(kernel='linear')
nb_model = MultinomialNB()

print("Evaluating SVM...")
train_and_evaluate(X_train, y_train, X_test, y_test, languages_test, svm_model,␣
 ↪"SVM")

print("Evaluating Naive Bayes...")
train_and_evaluate(X_train, y_train, X_test, y_test, languages_test, nb_model,␣
 ↪"Naive Bayes")
```

```
Evaluating SVM…
The demographic parity ratio is 0.9938528914599009
The equalized odds ratio is 0.8958006279434851
The demographic parity differece is 0.0031101386888131577
The equalized odds difference is 0.020157858718604205
Results for SVM:
Overall Accuracy: 0.8786127167630058
Overall Classification Report:
              precision    recall  f1-score   support

    negative       0.89      0.87      0.88      6660
    positive       0.87      0.89      0.88      6488

    accuracy                           0.88     13148
   macro avg       0.88      0.88      0.88     13148
weighted avg       0.88      0.88      0.88     13148
```

```
Accuracy on English: 0.8699124558439564
Classification Report for English:
              precision    recall  f1-score   support

    negative       0.88      0.86      0.87      3300
    positive       0.86      0.88      0.87      3211

    accuracy                           0.87      6511
   macro avg       0.87      0.87      0.87      6511
weighted avg       0.87      0.87      0.87      6511

Accuracy on French: 0.8871478077444629
Classification Report for French:
              precision    recall  f1-score   support

    negative       0.90      0.88      0.89      3360
    positive       0.88      0.90      0.89      3277

    accuracy                           0.89      6637
   macro avg       0.89      0.89      0.89      6637
weighted avg       0.89      0.89      0.89      6637

--------------------------------------------------
Evaluating Naive Bayes…
The demographic parity ratio is 0.9714034311355366
The equalized odds ratio is 0.7669034464614575
The demographic parity differece is 0.014792388870451978
The equalized odds difference is 0.038354978354978364
Results for Naive Bayes:
Overall Accuracy: 0.869257681776696
Overall Classification Report:
              precision    recall  f1-score   support

    negative       0.88      0.85      0.87      6660
    positive       0.86      0.88      0.87      6488

    accuracy                           0.87     13148
   macro avg       0.87      0.87      0.87     13148
weighted avg       0.87      0.87      0.87     13148

Accuracy on English: 0.8573183842727692
Classification Report for English:
              precision    recall  f1-score   support

    negative       0.88      0.84      0.86      3300
    positive       0.84      0.88      0.86      3211
```

```
    accuracy                           0.86      6511
   macro avg       0.86      0.86      0.86      6511
weighted avg       0.86      0.86      0.86      6511

Accuracy on French: 0.8809703179147205
Classification Report for French:
              precision    recall  f1-score   support

    negative       0.89      0.87      0.88      3360
    positive       0.87      0.89      0.88      3277

    accuracy                           0.88      6637
   macro avg       0.88      0.88      0.88      6637
weighted avg       0.88      0.88      0.88      6637

--------------------------------------------------
```

6 Transforming the Notebook into a Latex File

[4]:
```
!apt-get -q install texlive-xetex texlive-fonts-recommended␣
↪texlive-plain-generic
```

```
Reading package lists…
Building dependency tree…
Reading state information…
The following additional packages will be installed:
  dvisvgm fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono fonts-
texgyre
  fonts-urw-base35 libapache-pom-java libcommons-logging-java libcommons-parent-
java
  libfontbox-java libfontenc1 libgs9 libgs9-common libidn12 libijs-0.35
libjbig2dec0 libkpathsea6
  libpdfbox-java libptexenc1 libruby3.0 libsynctex2 libteckit0 libtexlua53
libtexluajit2 libwoff1
  libzzip-0-13 lmodern poppler-data preview-latex-style rake ruby ruby-net-
telnet ruby-rubygems
  ruby-webrick ruby-xmlrpc ruby3.0 rubygems-integration t1utils teckit tex-
common tex-gyre
  texlive-base texlive-binaries texlive-latex-base texlive-latex-extra texlive-
latex-recommended
  texlive-pictures tipa xfonts-encodings xfonts-utils
Suggested packages:
  fonts-noto fonts-freefont-otf | fonts-freefont-ttf libavalon-framework-java
  libcommons-logging-java-doc libexcalibur-logkit-java liblog4j1.2-java poppler-
utils ghostscript
  fonts-japanese-mincho | fonts-ipafont-mincho fonts-japanese-gothic | fonts-
ipafont-gothic
```

```
  fonts-arphic-ukai fonts-arphic-uming fonts-nanum ri ruby-dev bundler debhelper
gv
  | postscript-viewer perl-tk xpdf | pdf-viewer xzdec texlive-fonts-recommended-
doc
  texlive-latex-base-doc python3-pygments icc-profiles libfile-which-perl
  libspreadsheet-parseexcel-perl texlive-latex-extra-doc texlive-latex-
recommended-doc
  texlive-luatex texlive-pstricks dot2tex prerex texlive-pictures-doc vprerex
default-jre-headless
  tipa-doc
The following NEW packages will be installed:
  dvisvgm fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono fonts-
texgyre
  fonts-urw-base35 libapache-pom-java libcommons-logging-java libcommons-parent-
java
  libfontbox-java libfontenc1 libgs9 libgs9-common libidn12 libijs-0.35
libjbig2dec0 libkpathsea6
  libpdfbox-java libptexenc1 libruby3.0 libsynctex2 libteckit0 libtexlua53
libtexluajit2 libwoff1
  libzzip-0-13 lmodern poppler-data preview-latex-style rake ruby ruby-net-
telnet ruby-rubygems
  ruby-webrick ruby-xmlrpc ruby3.0 rubygems-integration t1utils teckit tex-
common tex-gyre
  texlive-base texlive-binaries texlive-fonts-recommended texlive-latex-base
texlive-latex-extra
  texlive-latex-recommended texlive-pictures texlive-plain-generic texlive-xetex
tipa
  xfonts-encodings xfonts-utils
0 upgraded, 54 newly installed, 0 to remove and 45 not upgraded.
Need to get 182 MB of archives.
After this operation, 571 MB of additional disk space will be used.
Get:1 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-droid-fallback all
1:6.0.1r16-1.1build1 [1,805 kB]
Get:2 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-lato all 2.0-2.1
[2,696 kB]
Get:3 http://archive.ubuntu.com/ubuntu jammy/main amd64 poppler-data all
0.4.11-1 [2,171 kB]
Get:4 http://archive.ubuntu.com/ubuntu jammy/universe amd64 tex-common all 6.17
[33.7 kB]
Get:5 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-urw-base35 all
20200910-1 [6,367 kB]
Get:6 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libgs9-common
all 9.55.0~dfsg1-0ubuntu5.6 [751 kB]
Get:7 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libidn12 amd64
1.38-4ubuntu1 [60.0 kB]
Get:8 http://archive.ubuntu.com/ubuntu jammy/main amd64 libijs-0.35 amd64
0.35-15build2 [16.5 kB]
Get:9 http://archive.ubuntu.com/ubuntu jammy/main amd64 libjbig2dec0 amd64
```

```
0.19-3build2 [64.7 kB]
Get:10 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libgs9 amd64
9.55.0~dfsg1-0ubuntu5.6 [5,031 kB]
Get:11 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libkpathsea6
amd64 2021.20210626.59705-1ubuntu0.2 [60.4 kB]
Get:12 http://archive.ubuntu.com/ubuntu jammy/main amd64 libwoff1 amd64
1.0.2-1build4 [45.2 kB]
Get:13 http://archive.ubuntu.com/ubuntu jammy/universe amd64 dvisvgm amd64
2.13.1-1 [1,221 kB]
Get:14 http://archive.ubuntu.com/ubuntu jammy/universe amd64 fonts-lmodern all
2.004.5-6.1 [4,532 kB]
Get:15 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-noto-mono all
20201225-1build1 [397 kB]
Get:16 http://archive.ubuntu.com/ubuntu jammy/universe amd64 fonts-texgyre all
20180621-3.1 [10.2 MB]
Get:17 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libapache-pom-java
all 18-1 [4,720 B]
Get:18 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libcommons-parent-
java all 43-1 [10.8 kB]
Get:19 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libcommons-logging-
java all 1.2-2 [60.3 kB]
Get:20 http://archive.ubuntu.com/ubuntu jammy/main amd64 libfontenc1 amd64
1:1.1.4-1build3 [14.7 kB]
Get:21 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libptexenc1
amd64 2021.20210626.59705-1ubuntu0.2 [39.1 kB]
Get:22 http://archive.ubuntu.com/ubuntu jammy/main amd64 rubygems-integration
all 1.18 [5,336 B]
Get:23 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 ruby3.0 amd64
3.0.2-7ubuntu2.4 [50.1 kB]
Get:24 http://archive.ubuntu.com/ubuntu jammy/main amd64 ruby-rubygems all
3.3.5-2 [228 kB]
Get:25 http://archive.ubuntu.com/ubuntu jammy/main amd64 ruby amd64 1:3.0~exp1
[5,100 B]
Get:26 http://archive.ubuntu.com/ubuntu jammy/main amd64 rake all 13.0.6-2 [61.7
kB]
Get:27 http://archive.ubuntu.com/ubuntu jammy/main amd64 ruby-net-telnet all
0.1.1-2 [12.6 kB]
Get:28 http://archive.ubuntu.com/ubuntu jammy/universe amd64 ruby-webrick all
1.7.0-3 [51.8 kB]
Get:29 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 ruby-xmlrpc all
0.3.2-1ubuntu0.1 [24.9 kB]
Get:30 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libruby3.0
amd64 3.0.2-7ubuntu2.4 [5,113 kB]
Get:31 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libsynctex2
amd64 2021.20210626.59705-1ubuntu0.2 [55.6 kB]
Get:32 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libteckit0 amd64
2.5.11+ds1-1 [421 kB]
Get:33 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libtexlua53
```

```
amd64 2021.20210626.59705-1ubuntu0.2 [120 kB]
Get:34 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libtexluajit2
amd64 2021.20210626.59705-1ubuntu0.2 [267 kB]
Get:35 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libzzip-0-13 amd64
0.13.72+dfsg.1-1.1 [27.0 kB]
Get:36 http://archive.ubuntu.com/ubuntu jammy/main amd64 xfonts-encodings all
1:1.0.5-0ubuntu2 [578 kB]
Get:37 http://archive.ubuntu.com/ubuntu jammy/main amd64 xfonts-utils amd64
1:7.7+6build2 [94.6 kB]
Get:38 http://archive.ubuntu.com/ubuntu jammy/universe amd64 lmodern all
2.004.5-6.1 [9,471 kB]
Get:39 http://archive.ubuntu.com/ubuntu jammy/universe amd64 preview-latex-style
all 12.2-1ubuntu1 [185 kB]
Get:40 http://archive.ubuntu.com/ubuntu jammy/main amd64 t1utils amd64
1.41-4build2 [61.3 kB]
Get:41 http://archive.ubuntu.com/ubuntu jammy/universe amd64 teckit amd64
2.5.11+ds1-1 [699 kB]
Get:42 http://archive.ubuntu.com/ubuntu jammy/universe amd64 tex-gyre all
20180621-3.1 [6,209 kB]
Get:43 http://archive.ubuntu.com/ubuntu jammy-updates/universe amd64 texlive-
binaries amd64 2021.20210626.59705-1ubuntu0.2 [9,860 kB]
Get:44 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-base all
2021.20220204-1 [21.0 MB]
Get:45 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-fonts-
recommended all 2021.20220204-1 [4,972 kB]
Get:46 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-base
all 2021.20220204-1 [1,128 kB]
Get:47 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libfontbox-java all
1:1.8.16-2 [207 kB]
Get:48 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libpdfbox-java all
1:1.8.16-2 [5,199 kB]
Get:49 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-
recommended all 2021.20220204-1 [14.4 MB]
Get:50 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-pictures
all 2021.20220204-1 [8,720 kB]
Get:51 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-extra
all 2021.20220204-1 [13.9 MB]
Get:52 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-plain-
generic all 2021.20220204-1 [27.5 MB]
Get:53 http://archive.ubuntu.com/ubuntu jammy/universe amd64 tipa all 2:1.3-21
[2,967 kB]
Get:54 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-xetex all
2021.20220204-1 [12.4 MB]
Fetched 182 MB in 15s (11.8 MB/s)
Extracting templates from packages: 100%
Preconfiguring packages …
Selecting previously unselected package fonts-droid-fallback.
(Reading database … 121753 files and directories currently installed.)
```

```
Preparing to unpack …/00-fonts-droid-fallback_1%3a6.0.1r16-1.1build1_all.deb
…
Unpacking fonts-droid-fallback (1:6.0.1r16-1.1build1) …
Selecting previously unselected package fonts-lato.
Preparing to unpack …/01-fonts-lato_2.0-2.1_all.deb …
Unpacking fonts-lato (2.0-2.1) …
Selecting previously unselected package poppler-data.
Preparing to unpack …/02-poppler-data_0.4.11-1_all.deb …
Unpacking poppler-data (0.4.11-1) …
Selecting previously unselected package tex-common.
Preparing to unpack …/03-tex-common_6.17_all.deb …
Unpacking tex-common (6.17) …
Selecting previously unselected package fonts-urw-base35.
Preparing to unpack …/04-fonts-urw-base35_20200910-1_all.deb …
Unpacking fonts-urw-base35 (20200910-1) …
Selecting previously unselected package libgs9-common.
Preparing to unpack …/05-libgs9-common_9.55.0~dfsg1-0ubuntu5.6_all.deb …
Unpacking libgs9-common (9.55.0~dfsg1-0ubuntu5.6) …
Selecting previously unselected package libidn12:amd64.
Preparing to unpack …/06-libidn12_1.38-4ubuntu1_amd64.deb …
Unpacking libidn12:amd64 (1.38-4ubuntu1) …
Selecting previously unselected package libijs-0.35:amd64.
Preparing to unpack …/07-libijs-0.35_0.35-15build2_amd64.deb …
Unpacking libijs-0.35:amd64 (0.35-15build2) …
Selecting previously unselected package libjbig2dec0:amd64.
Preparing to unpack …/08-libjbig2dec0_0.19-3build2_amd64.deb …
Unpacking libjbig2dec0:amd64 (0.19-3build2) …
Selecting previously unselected package libgs9:amd64.
Preparing to unpack …/09-libgs9_9.55.0~dfsg1-0ubuntu5.6_amd64.deb …
Unpacking libgs9:amd64 (9.55.0~dfsg1-0ubuntu5.6) …
Selecting previously unselected package libkpathsea6:amd64.
Preparing to unpack …/10-libkpathsea6_2021.20210626.59705-1ubuntu0.2_amd64.deb
…
Unpacking libkpathsea6:amd64 (2021.20210626.59705-1ubuntu0.2) …
Selecting previously unselected package libwoff1:amd64.
Preparing to unpack …/11-libwoff1_1.0.2-1build4_amd64.deb …
Unpacking libwoff1:amd64 (1.0.2-1build4) …
Selecting previously unselected package dvisvgm.
Preparing to unpack …/12-dvisvgm_2.13.1-1_amd64.deb …
Unpacking dvisvgm (2.13.1-1) …
Selecting previously unselected package fonts-lmodern.
Preparing to unpack …/13-fonts-lmodern_2.004.5-6.1_all.deb …
Unpacking fonts-lmodern (2.004.5-6.1) …
Selecting previously unselected package fonts-noto-mono.
Preparing to unpack …/14-fonts-noto-mono_20201225-1build1_all.deb …
Unpacking fonts-noto-mono (20201225-1build1) …
Selecting previously unselected package fonts-texgyre.
Preparing to unpack …/15-fonts-texgyre_20180621-3.1_all.deb …
```

```
Unpacking fonts-texgyre (20180621-3.1) …
Selecting previously unselected package libapache-pom-java.
Preparing to unpack …/16-libapache-pom-java_18-1_all.deb …
Unpacking libapache-pom-java (18-1) …
Selecting previously unselected package libcommons-parent-java.
Preparing to unpack …/17-libcommons-parent-java_43-1_all.deb …
Unpacking libcommons-parent-java (43-1) …
Selecting previously unselected package libcommons-logging-java.
Preparing to unpack …/18-libcommons-logging-java_1.2-2_all.deb …
Unpacking libcommons-logging-java (1.2-2) …
Selecting previously unselected package libfontenc1:amd64.
Preparing to unpack …/19-libfontenc1_1%3a1.1.4-1build3_amd64.deb …
Unpacking libfontenc1:amd64 (1:1.1.4-1build3) …
Selecting previously unselected package libptexenc1:amd64.
Preparing to unpack …/20-libptexenc1_2021.20210626.59705-1ubuntu0.2_amd64.deb
…
Unpacking libptexenc1:amd64 (2021.20210626.59705-1ubuntu0.2) …
Selecting previously unselected package rubygems-integration.
Preparing to unpack …/21-rubygems-integration_1.18_all.deb …
Unpacking rubygems-integration (1.18) …
Selecting previously unselected package ruby3.0.
Preparing to unpack …/22-ruby3.0_3.0.2-7ubuntu2.4_amd64.deb …
Unpacking ruby3.0 (3.0.2-7ubuntu2.4) …
Selecting previously unselected package ruby-rubygems.
Preparing to unpack …/23-ruby-rubygems_3.3.5-2_all.deb …
Unpacking ruby-rubygems (3.3.5-2) …
Selecting previously unselected package ruby.
Preparing to unpack …/24-ruby_1%3a3.0~exp1_amd64.deb …
Unpacking ruby (1:3.0~exp1) …
Selecting previously unselected package rake.
Preparing to unpack …/25-rake_13.0.6-2_all.deb …
Unpacking rake (13.0.6-2) …
Selecting previously unselected package ruby-net-telnet.
Preparing to unpack …/26-ruby-net-telnet_0.1.1-2_all.deb …
Unpacking ruby-net-telnet (0.1.1-2) …
Selecting previously unselected package ruby-webrick.
Preparing to unpack …/27-ruby-webrick_1.7.0-3_all.deb …
Unpacking ruby-webrick (1.7.0-3) …
Selecting previously unselected package ruby-xmlrpc.
Preparing to unpack …/28-ruby-xmlrpc_0.3.2-1ubuntu0.1_all.deb …
Unpacking ruby-xmlrpc (0.3.2-1ubuntu0.1) …
Selecting previously unselected package libruby3.0:amd64.
Preparing to unpack …/29-libruby3.0_3.0.2-7ubuntu2.4_amd64.deb …
Unpacking libruby3.0:amd64 (3.0.2-7ubuntu2.4) …
Selecting previously unselected package libsynctex2:amd64.
Preparing to unpack …/30-libsynctex2_2021.20210626.59705-1ubuntu0.2_amd64.deb
…
Unpacking libsynctex2:amd64 (2021.20210626.59705-1ubuntu0.2) …
```

```
Selecting previously unselected package libteckit0:amd64.
Preparing to unpack …/31-libteckit0_2.5.11+ds1-1_amd64.deb …
Unpacking libteckit0:amd64 (2.5.11+ds1-1) …
Selecting previously unselected package libtexlua53:amd64.
Preparing to unpack …/32-libtexlua53_2021.20210626.59705-1ubuntu0.2_amd64.deb
…
Unpacking libtexlua53:amd64 (2021.20210626.59705-1ubuntu0.2) …
Selecting previously unselected package libtexluajit2:amd64.
Preparing to unpack
…/33-libtexluajit2_2021.20210626.59705-1ubuntu0.2_amd64.deb …
Unpacking libtexluajit2:amd64 (2021.20210626.59705-1ubuntu0.2) …
Selecting previously unselected package libzzip-0-13:amd64.
Preparing to unpack …/34-libzzip-0-13_0.13.72+dfsg.1-1.1_amd64.deb …
Unpacking libzzip-0-13:amd64 (0.13.72+dfsg.1-1.1) …
Selecting previously unselected package xfonts-encodings.
Preparing to unpack …/35-xfonts-encodings_1%3a1.0.5-0ubuntu2_all.deb …
Unpacking xfonts-encodings (1:1.0.5-0ubuntu2) …
Selecting previously unselected package xfonts-utils.
Preparing to unpack …/36-xfonts-utils_1%3a7.7+6build2_amd64.deb …
Unpacking xfonts-utils (1:7.7+6build2) …
Selecting previously unselected package lmodern.
Preparing to unpack …/37-lmodern_2.004.5-6.1_all.deb …
Unpacking lmodern (2.004.5-6.1) …
Selecting previously unselected package preview-latex-style.
Preparing to unpack …/38-preview-latex-style_12.2-1ubuntu1_all.deb …
Unpacking preview-latex-style (12.2-1ubuntu1) …
Selecting previously unselected package t1utils.
Preparing to unpack …/39-t1utils_1.41-4build2_amd64.deb …
Unpacking t1utils (1.41-4build2) …
Selecting previously unselected package teckit.
Preparing to unpack …/40-teckit_2.5.11+ds1-1_amd64.deb …
Unpacking teckit (2.5.11+ds1-1) …
Selecting previously unselected package tex-gyre.
Preparing to unpack …/41-tex-gyre_20180621-3.1_all.deb …
Unpacking tex-gyre (20180621-3.1) …
Selecting previously unselected package texlive-binaries.
Preparing to unpack …/42-texlive-
binaries_2021.20210626.59705-1ubuntu0.2_amd64.deb …
Unpacking texlive-binaries (2021.20210626.59705-1ubuntu0.2) …
Selecting previously unselected package texlive-base.
Preparing to unpack …/43-texlive-base_2021.20220204-1_all.deb …
Unpacking texlive-base (2021.20220204-1) …
Selecting previously unselected package texlive-fonts-recommended.
Preparing to unpack …/44-texlive-fonts-recommended_2021.20220204-1_all.deb …
Unpacking texlive-fonts-recommended (2021.20220204-1) …
Selecting previously unselected package texlive-latex-base.
Preparing to unpack …/45-texlive-latex-base_2021.20220204-1_all.deb …
Unpacking texlive-latex-base (2021.20220204-1) …
```

```
Selecting previously unselected package libfontbox-java.
Preparing to unpack …/46-libfontbox-java_1%3a1.8.16-2_all.deb …
Unpacking libfontbox-java (1:1.8.16-2) …
Selecting previously unselected package libpdfbox-java.
Preparing to unpack …/47-libpdfbox-java_1%3a1.8.16-2_all.deb …
Unpacking libpdfbox-java (1:1.8.16-2) …
Selecting previously unselected package texlive-latex-recommended.
Preparing to unpack …/48-texlive-latex-recommended_2021.20220204-1_all.deb …
Unpacking texlive-latex-recommended (2021.20220204-1) …
Selecting previously unselected package texlive-pictures.
Preparing to unpack …/49-texlive-pictures_2021.20220204-1_all.deb …
Unpacking texlive-pictures (2021.20220204-1) …
Selecting previously unselected package texlive-latex-extra.
Preparing to unpack …/50-texlive-latex-extra_2021.20220204-1_all.deb …
Unpacking texlive-latex-extra (2021.20220204-1) …
Selecting previously unselected package texlive-plain-generic.
Preparing to unpack …/51-texlive-plain-generic_2021.20220204-1_all.deb …
Unpacking texlive-plain-generic (2021.20220204-1) …
Selecting previously unselected package tipa.
Preparing to unpack …/52-tipa_2%3a1.3-21_all.deb …
Unpacking tipa (2:1.3-21) …
Selecting previously unselected package texlive-xetex.
Preparing to unpack …/53-texlive-xetex_2021.20220204-1_all.deb …
Unpacking texlive-xetex (2021.20220204-1) …
Setting up fonts-lato (2.0-2.1) …
Setting up fonts-noto-mono (20201225-1build1) …
Setting up libwoff1:amd64 (1.0.2-1build4) …
Setting up libtexlua53:amd64 (2021.20210626.59705-1ubuntu0.2) …
Setting up libijs-0.35:amd64 (0.35-15build2) …
Setting up libtexluajit2:amd64 (2021.20210626.59705-1ubuntu0.2) …
Setting up libfontbox-java (1:1.8.16-2) …
Setting up rubygems-integration (1.18) …
Setting up libzzip-0-13:amd64 (0.13.72+dfsg.1-1.1) …
Setting up fonts-urw-base35 (20200910-1) …
Setting up poppler-data (0.4.11-1) …
Setting up tex-common (6.17) …
update-language: texlive-base not installed and configured, doing nothing!
Setting up libfontenc1:amd64 (1:1.1.4-1build3) …
Setting up libjbig2dec0:amd64 (0.19-3build2) …
Setting up libteckit0:amd64 (2.5.11+ds1-1) …
Setting up libapache-pom-java (18-1) …
Setting up ruby-net-telnet (0.1.1-2) …
Setting up xfonts-encodings (1:1.0.5-0ubuntu2) …
Setting up t1utils (1.41-4build2) …
Setting up libidn12:amd64 (1.38-4ubuntu1) …
Setting up fonts-texgyre (20180621-3.1) …
Setting up libkpathsea6:amd64 (2021.20210626.59705-1ubuntu0.2) …
Setting up ruby-webrick (1.7.0-3) …
```

```
Setting up fonts-lmodern (2.004.5-6.1) …
Setting up fonts-droid-fallback (1:6.0.1r16-1.1build1) …
Setting up ruby-xmlrpc (0.3.2-1ubuntu0.1) …
Setting up libsynctex2:amd64 (2021.20210626.59705-1ubuntu0.2) …
Setting up libgs9-common (9.55.0~dfsg1-0ubuntu5.6) …
Setting up teckit (2.5.11+ds1-1) …
Setting up libpdfbox-java (1:1.8.16-2) …
Setting up libgs9:amd64 (9.55.0~dfsg1-0ubuntu5.6) …
Setting up preview-latex-style (12.2-1ubuntu1) …
Setting up libcommons-parent-java (43-1) …
Setting up dvisvgm (2.13.1-1) …
Setting up libcommons-logging-java (1.2-2) …
Setting up xfonts-utils (1:7.7+6build2) …
Setting up libptexenc1:amd64 (2021.20210626.59705-1ubuntu0.2) …
Setting up texlive-binaries (2021.20210626.59705-1ubuntu0.2) …
update-alternatives: using /usr/bin/xdvi-xaw to provide /usr/bin/xdvi.bin
(xdvi.bin) in auto mode
update-alternatives: using /usr/bin/bibtex.original to provide /usr/bin/bibtex
(bibtex) in auto mode
Setting up lmodern (2.004.5-6.1) …
Setting up texlive-base (2021.20220204-1) …
/usr/bin/ucfr
/usr/bin/ucfr
/usr/bin/ucfr
/usr/bin/ucfr
mktexlsr: Updating /var/lib/texmf/ls-R-TEXLIVEDIST…
mktexlsr: Updating /var/lib/texmf/ls-R-TEXMFMAIN…
mktexlsr: Updating /var/lib/texmf/ls-R…
mktexlsr: Done.
tl-paper: setting paper size for dvips to a4:
/var/lib/texmf/dvips/config/config-paper.ps
tl-paper: setting paper size for dvipdfmx to a4:
/var/lib/texmf/dvipdfmx/dvipdfmx-paper.cfg
tl-paper: setting paper size for xdvi to a4: /var/lib/texmf/xdvi/XDvi-paper
tl-paper: setting paper size for pdftex to a4: /var/lib/texmf/tex/generic/tex-
ini-files/pdftexconfig.tex
Setting up tex-gyre (20180621-3.1) …
Setting up texlive-plain-generic (2021.20220204-1) …
Setting up texlive-latex-base (2021.20220204-1) …
Setting up texlive-latex-recommended (2021.20220204-1) …
Setting up texlive-pictures (2021.20220204-1) …
Setting up texlive-fonts-recommended (2021.20220204-1) …
Setting up tipa (2:1.3-21) …
Setting up texlive-latex-extra (2021.20220204-1) …
Setting up texlive-xetex (2021.20220204-1) …
Setting up rake (13.0.6-2) …
Setting up libruby3.0:amd64 (3.0.2-7ubuntu2.4) …
Setting up ruby3.0 (3.0.2-7ubuntu2.4) …
```

```
Setting up ruby (1:3.0~exp1) …
Setting up ruby-rubygems (3.3.5-2) …
Processing triggers for man-db (2.10.2-1) …
Processing triggers for fontconfig (2.13.1-4.2ubuntu5) …
Processing triggers for libc-bin (2.35-0ubuntu3.4) …
/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_0.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbb.so.12 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbmalloc_proxy.so.2 is not a symbolic
link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_5.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbmalloc.so.2 is not a symbolic link

Processing triggers for tex-common (6.17) …
Running updmap-sys. This may take some time… done.
Running mktexlsr /var/lib/texmf … done.
Building format(s) --all.
        This may take some time… done.
```

[ ]:
```
!jupyter nbconvert --to pdf "/content/drive/My Drive/Colab Notebooks/Analyzing␣
↪Language Bias Between French and English in Conventional Multilingual␣
↪Sentiment Analysis Models.ipynb"
```

```
[NbConvertApp] Converting notebook /content/drive/My Drive/Colab
Notebooks/Analyzing Language Bias Between French and English in Traditional
Multilingual Sentiment Analysis Models.ipynb to pdf
[NbConvertApp] Writing 78223 bytes to notebook.tex
[NbConvertApp] Building PDF
[NbConvertApp] Running xelatex 3 times: ['xelatex', 'notebook.tex', '-quiet']
[NbConvertApp] Running bibtex 1 time: ['bibtex', 'notebook']
[NbConvertApp] WARNING | bibtex had problems, most likely because there were no
citations
[NbConvertApp] PDF successfully created
[NbConvertApp] Writing 86802 bytes to /content/drive/My Drive/Colab
Notebooks/Analyzing Language Bias Between French and English in Traditional
Multilingual Sentiment Analysis Models.pdf
```

[ ]:
```
import os
from google.colab import files
files.download(f"/content/drive/My Drive/Colab Notebooks/Analyzing Language␣
↪Bias Between French and English in Conventional Multilingual Sentiment␣
↪Analysis Models.pdf")
```

```
<IPython.core.display.Javascript object>
<IPython.core.display.Javascript object>
```