Sentiment Analysis of Uniqlo Online Customer Reviews in Producing a Classifier Model

Group Members

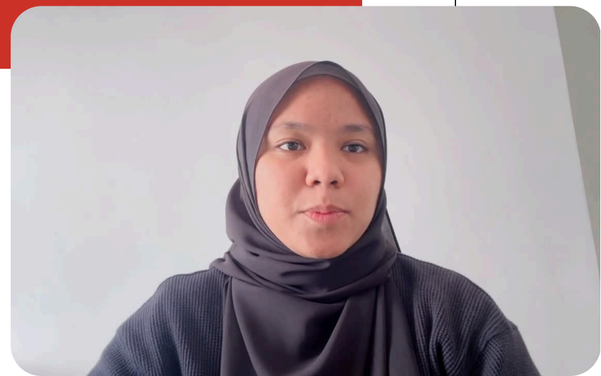
Hanna Sofea binti Mohd Zaid
2024793105

Nur Fatihah binti Sanusi
2024794519

Aliyah Majdina binti Mudzaffar Alfian
2024905023

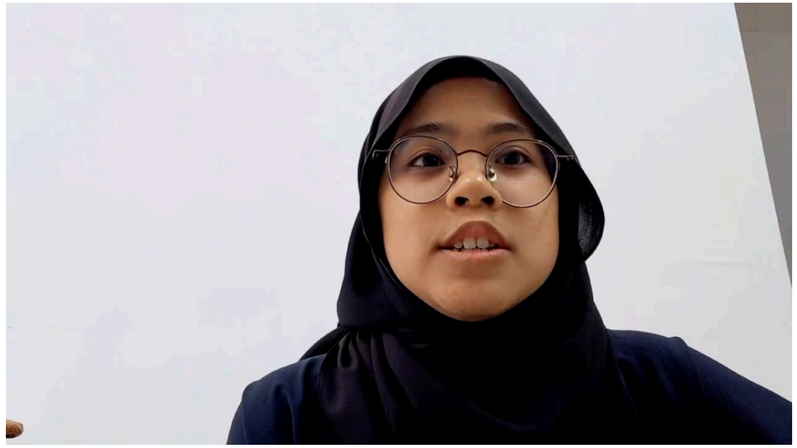
Summary of Study

- This study aims to analyse customer sentiment towards Uniqlo products based on online customer reviews.
- A total of 1,510 reviews were collected from various online platforms and categorised into positive, neutral, and negative sentiments.
- Descriptive analytics was conducted to visualise sentiment patterns and customer behaviour, while predictive analytics was applied to develop and compare sentiment classification models.
- The best-performing model was then selected and used to predict sentiment for new, unlabeled Uniqlo customer reviews.



Problem Statement

- Large volume of online customer review
- Written in unstructured textual form
- Businesses face challenges in accurately identifying customer sentiment from large-scale review data



Objectives

DESCRIPTIVE ANALYTICS

- To compare the average customer ratings of Uniqlo products across different continents.
- To compare the average customer ratings across different sentiment categories between male and female customers.
- To analyse the average customer ratings across different Uniqlo product types based on sentiment categories.

PREDICTIVE ANALYTICS

- To establish a classifier model that is capable of identifying if the data review is positive, neutral or negative.
- To determine which model is best for conducting sentiment analysis of Uniqlo customer review.

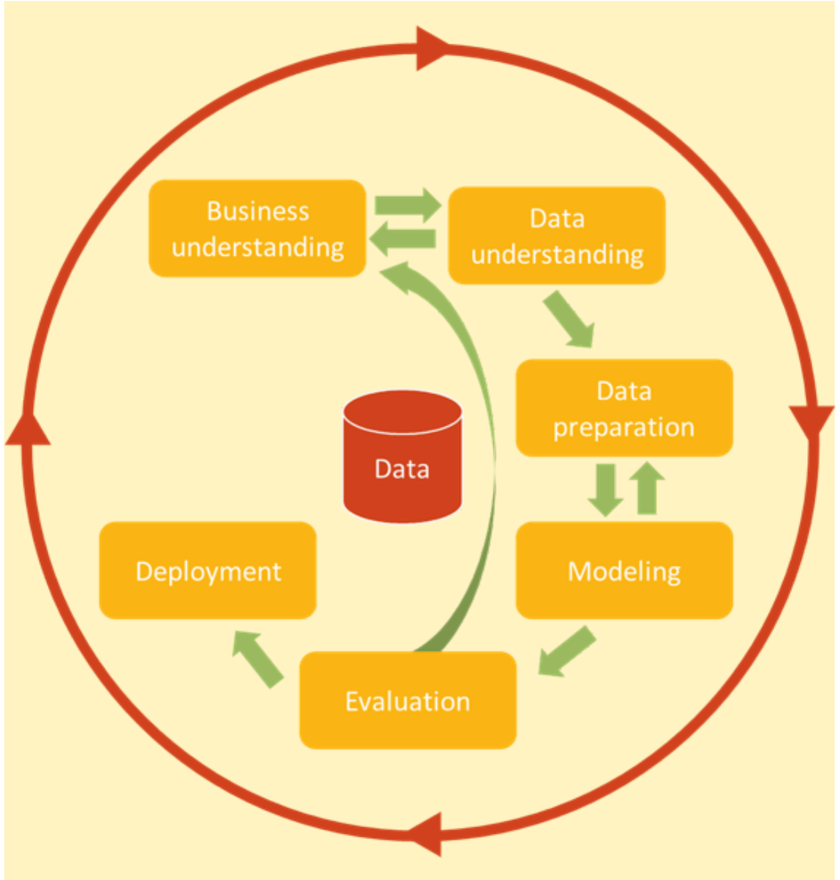
LITERATURE REVIEW

- Prior research confirms that Naive Bayes, Random Tree, and KNN are reliable classifiers for sentiment analysis. (Zaqy et al., 2024) (Azhar et al., 2025) (Fitri et al., 2023)
- Text preprocessing plays a crucial role in improving model performance.
- TF-IDF is a robust and widely adopted feature extraction technique for sentiment classification tasks.
- These findings support the methodology adopted in analysing UNIQLO online customer reviews.

Sentiment Analysis

- Sentiment analysis - NLP technique used to identify & classify opinions/emotions expressed in textual data into sentiment categories
- Online customer reviews contain valuable opinions that reflect customers satisfaction & perception towards products and services
- Machine learning-based sentiment analysis enables automatic & efficient classification of large volumes of unstructured review data

Methodology



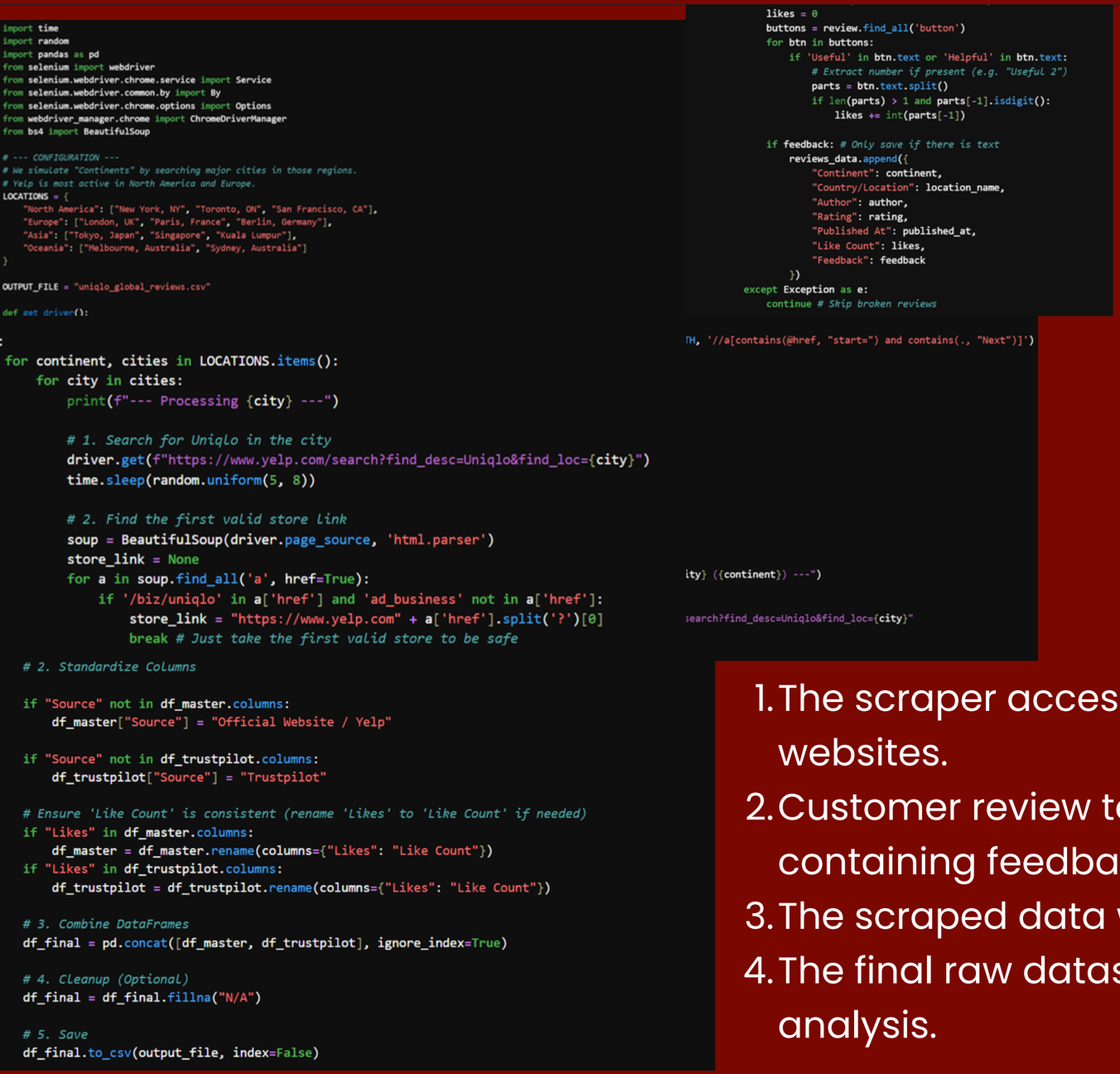
Steps Applied in This Study

1. Business Understanding: Identify the objective to analyse customer sentiment towards Uniqlo products based on online customer reviews.
2. Data Understanding: Collect and explore Uniqlo customer reviews obtained from multiple online platforms.
3. Data Preparation: Clean and preprocess textual data by removing noise and preparing the dataset for sentiment analysis.
4. Modelling: Develop and compare sentiment classification models using different machine learning algorithms.
5. Evaluation: Evaluate model performance using accuracy and other evaluation metrics to select the best model.
6. Deployment: Apply the selected model to predict sentiment for new, unlabeled Uniqlo customer reviews.

SOURCE & BACKGROUND OF DATA

- The dataset used in this study was collected by scraping UNIQLO online customer reviews from multiple online platforms. Based on the dataset, the reviews were obtained from:
- UNIQLO official online store product review sections
- Third-party e-commerce platforms where UNIQLO products are sold and reviewed by customers

- The dataset consists of 1,510 customer reviews collected from various online sources.
- Each row represents one customer feedback entry.
- The main attribute used in this study is:
- Feedback – textual customer review describing opinions on UNIQLO products



```python
import time
import random
import pandas as pd
from selenium import webdriver
from selenium.webdriver.chrome.service import Service
from selenium.webdriver.common.by import By
from selenium.webdriver.chrome.options import Options
from webdriver_manager.chrome import ChromeDriverManager
from bs4 import BeautifulSoup

# --- CONFIGURATION ---
# We simulate "Continents" by searching major cities in those regions.
# Yelp is most active in North America and Europe.
LOCATIONS = {
    "North America": ["New York, NY", "Toronto, ON", "San Francisco, CA"],
    "Europe": ["London, UK", "Paris, France", "Berlin, Germany"],
    "Asia": ["Tokyo, Japan", "Singapore", "Kuala Lumpur"],
    "Oceania": ["Melbourne, Australia", "Sydney, Australia"]
}

OUTPUT_FILE = "uniqlo_global_reviews.csv"

def get_driver():

for continent, cities in LOCATIONS.items():
    for city in cities:
        print(f"--- Processing {city} ---")

        # 1. Search for Uniqlo in the city
        driver.get(f"https://www.yelp.com/search?find_desc=Uniqlo&find_loc={city}")
        time.sleep(random.uniform(5, 8))

        # 2. Find the first valid store link
        soup = BeautifulSoup(driver.page_source, 'html.parser')
        store_link = None
        for a in soup.find_all('a', href=True):
            if '/biz/uniqlo' in a['href'] and 'ad_business' not in a['href']:
                store_link = "https://www.yelp.com" + a['href'].split('?')[0]
                break # Just take the first valid store to be safe
```

```python
            likes = 0
            buttons = review.find_all('button')
            for btn in buttons:
                if 'Useful' in btn.text or 'Helpful' in btn.text:
                    # Extract number if present (e.g. "Useful 2")
                    parts = btn.text.split()
                    if len(parts) > 1 and parts[-1].isdigit():
                        likes += int(parts[-1])

            if feedback: # Only save if there is text
                reviews_data.append({
                    "Continent": continent,
                    "Country/Location": location_name,
                    "Author": author,
                    "Rating": rating,
                    "Published At": published_at,
                    "Like Count": likes,
                    "Feedback": feedback
                })
        except Exception as e:
            continue # Skip broken reviews
```

```python
TH, '//a[contains(@href, "start=") and contains(., "Next")]')
```

```python
ity} ({continent}) ---")
```

```python
search?find_desc=Uniqlo&find_loc={city}"
```

```python
# 2. Standardize Columns

if "Source" not in df_master.columns:
    df_master["Source"] = "Official Website / Yelp"

if "Source" not in df_trustpilot.columns:
    df_trustpilot["Source"] = "Trustpilot"

# Ensure 'Like Count' is consistent (rename 'Likes' to 'Like Count' if needed)
if "Likes" in df_master.columns:
    df_master = df_master.rename(columns={"Likes": "Like Count"})
if "Likes" in df_trustpilot.columns:
    df_trustpilot = df_trustpilot.rename(columns={"Likes": "Like Count"})

# 3. Combine DataFrames
df_final = pd.concat([df_master, df_trustpilot], ignore_index=True)

# 4. Cleanup (Optional)
df_final = df_final.fillna("N/A")

# 5. Save
df_final.to_csv(output_file, index=False)
```

DATA SCRAPPING

- The data scraping process was carried out using Python programming language.
- Customer review text was extracted from multiple web pages across different online platforms selling UNIQLO products.
- Python web scraping libraries were used to retrieve review content from the websites.

SCRAPING WORKFLOW

1. The scraper accessed product review pages from the selected websites.
2. Customer review text was extracted from the HTML elements containing feedback.
3. The scraped data was cleaned and structured into tabular format.
4. The final raw dataset was saved as a CSV file for preprocessing and analysis.

Attributes and Class Label

Attributes

Attribute	Description
Continent	Continent where the reviewer is located
Country/City	Country and city of the reviewer
Store URL	Web link to the store or review source
Gender	Gender of the reviewer
Rating	Numerical rating given by the customer
Published at	Date when the review was published
Feedback	Textual customer review describing opinions on Uniqlo products
Source	Platform where the review was collected
Type of Purchase	Type of Uniqlo product purchased

Class Label

Sentiment	Positive, Negative, Neutral

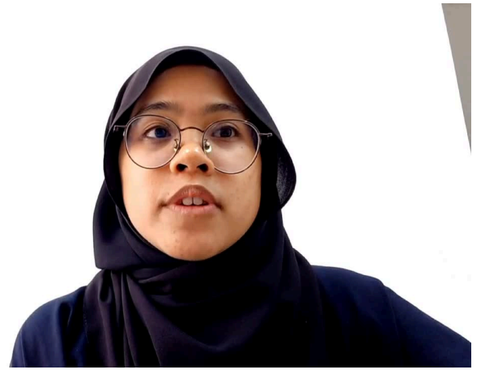
Preprocessing Process

1) Data Cleansing
- Convert the text to lowercase
- Remove punctuation, numbers and symbols
- Tokenization
- Remove stop words
- Stemming

2) Data Tagging
- Manual tagging (Classified each comment as Positive, Neutral or Negative)

Descriptive Analysis

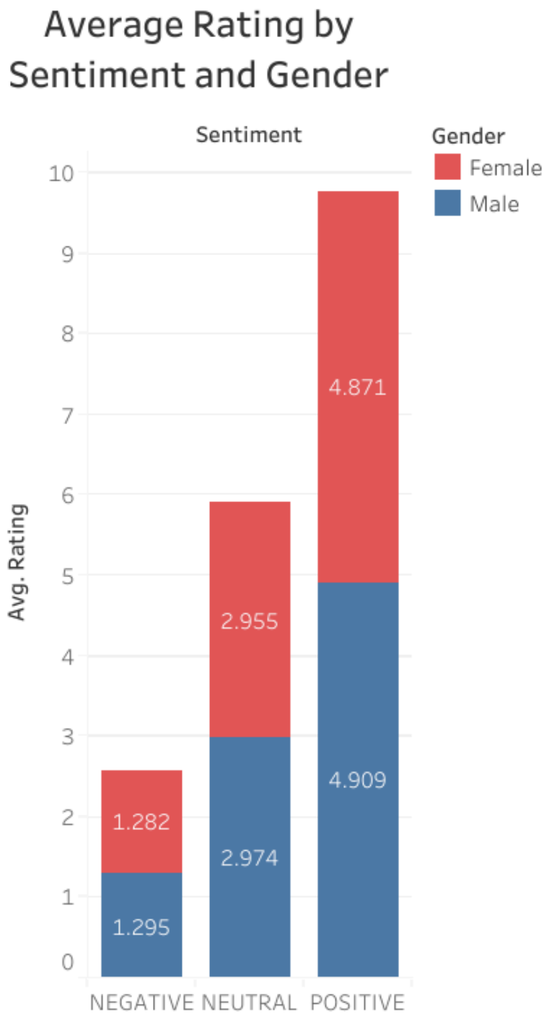
- Asia records the highest average rating, indicating a high level of customer satisfaction.
- North America shows a moderate average rating, suggesting generally positive but less favorable perceptions compared to Asia.
- Europe records the lowest average rating, indicating lower customer satisfaction relative to other regions.
- Uniqlo products being most favorably rated in Asia.



Descriptive Analysis

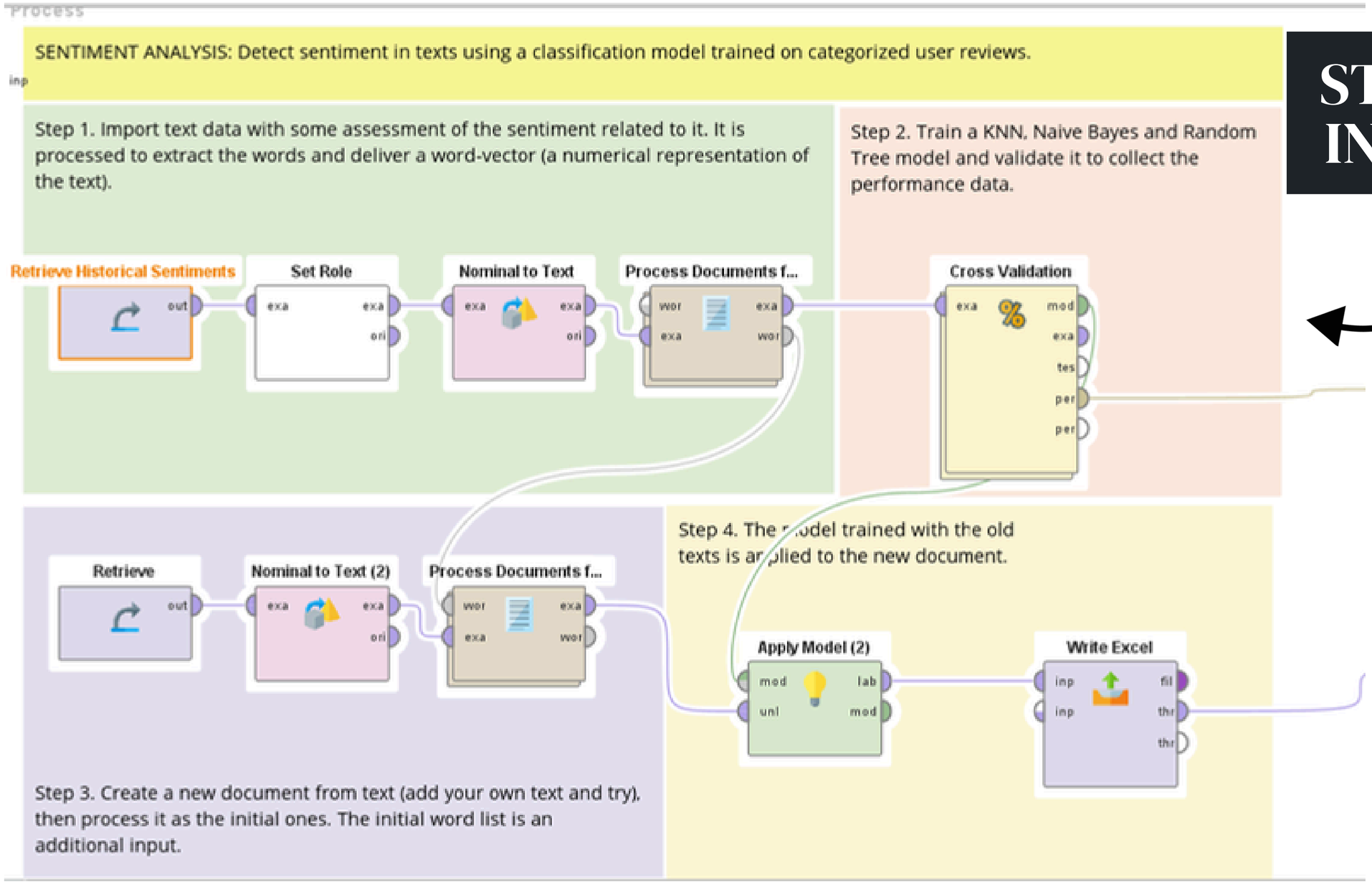
- Positive sentiment consistently records the highest average ratings.
- Neutral sentiment shows moderate average ratings.
- Negative sentiment has the lowest average ratings, reflecting dissatisfaction among customers.
- Differences in average ratings across product types are relatively small, indicating that customer satisfaction patterns are generally consistent across Uniqlo products, regardless of purchase type.

Descriptive Analysis



- For both genders, positive sentiment is associated with the highest average ratings.
- Neutral sentiment shows moderate average ratings for both male and female customers.
- Negative sentiment records the lowest average ratings across both genders, indicating dissatisfaction.
- The average ratings between male and female customers are very similar across all sentiment categories, suggesting that gender does not significantly influence rating behaviour.

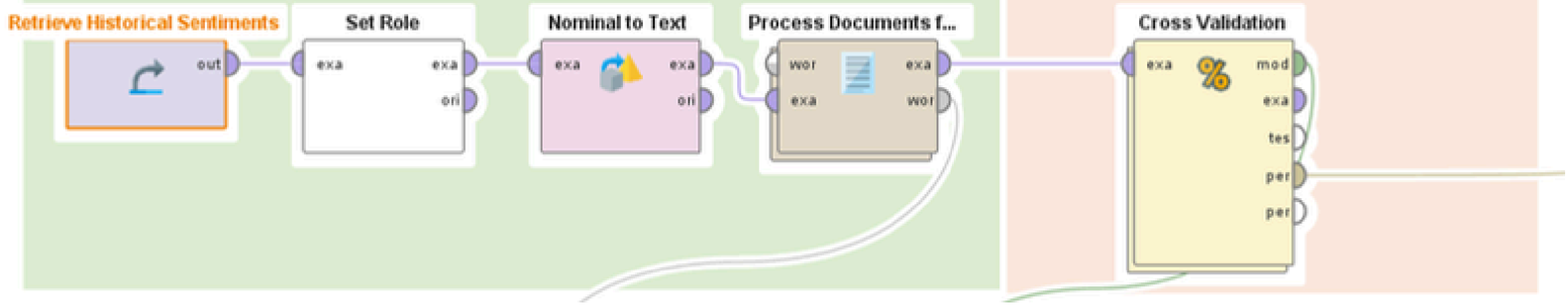
STEPS PROCESS IN RAPIDMINER

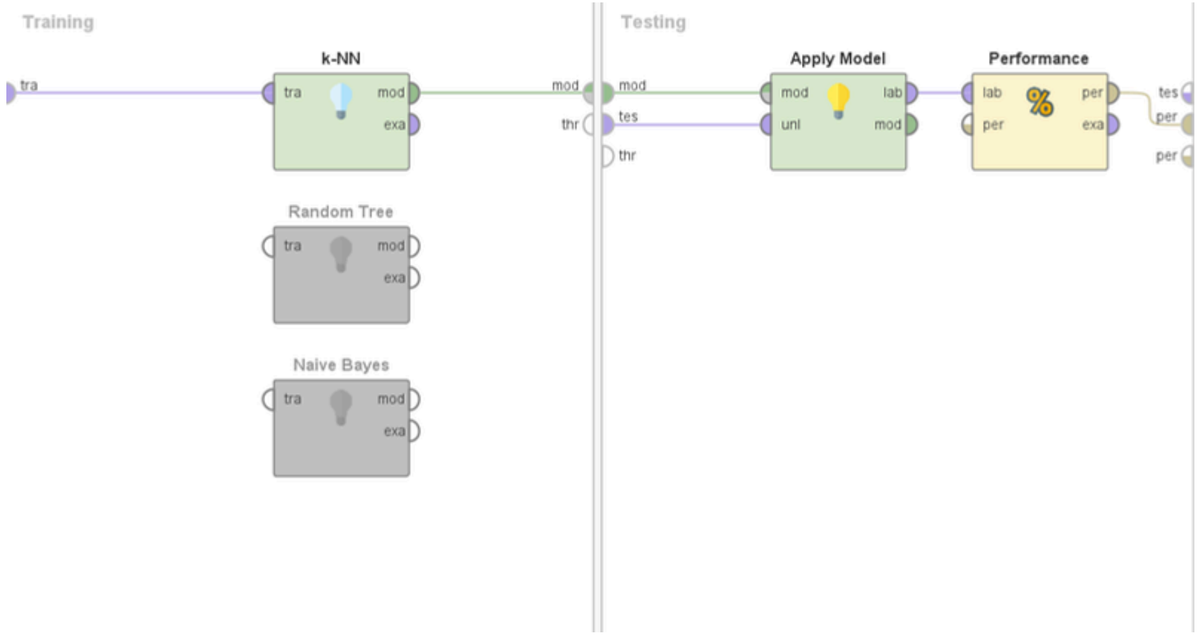


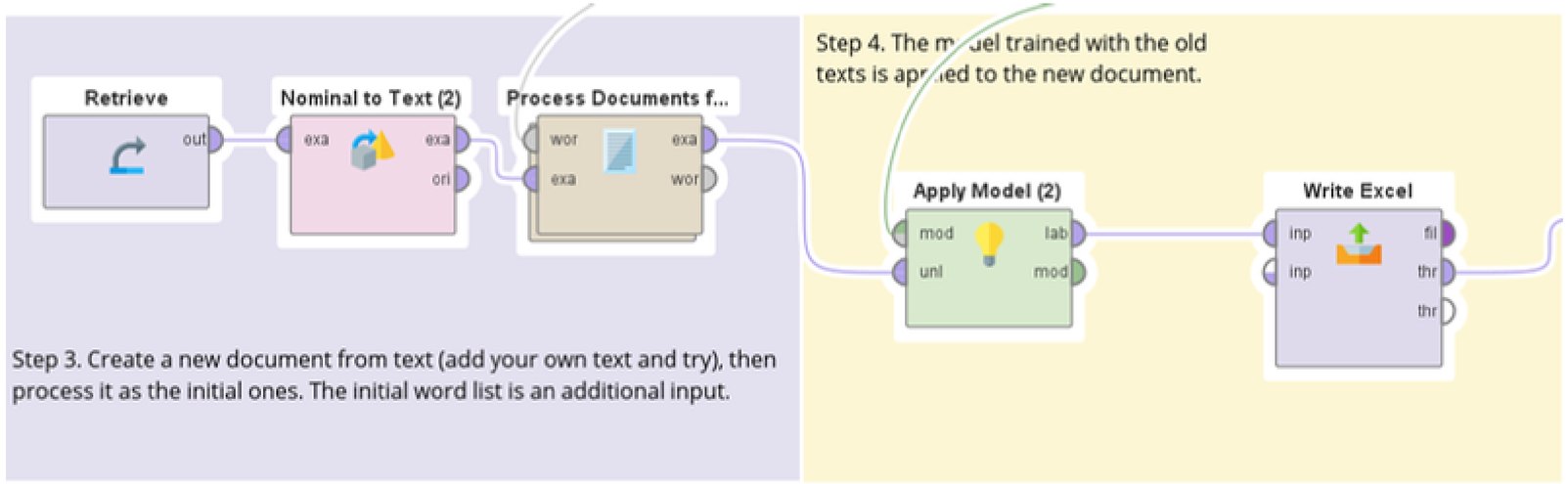
SENTIMENT ANALYSIS: Detect sentiment in texts using a classification model trained on categorized user reviews.

Step 1. Import text data with some assessment of the sentiment related to it. It is processed to extract the words and deliver a word-vector (a numerical representation of the text).

Step 2. Train a KNN, Naive Bayes and Random Tree model and validate it to collect the performance data.






Retrieve
out
Nominal to Text (2)
exa
exa
ori
Process Documents f...
wor
exa
exa
wor
Step 3. Create a new document from text (add your own text and try), then process it as the initial ones. The initial word list is an additional input.
Step 4. The m el trained with the old texts is ap ed to the new document.
Apply Model (2)
mod
unl
lab
mod
Write Excel
inp
inp
fil
thr
thr

Predictive Analytics

All Models Performance

Model	Accuracy	Precision	Recall
KNN	89.18%	90.56%	98.56%
Naive Bayes	71.93%	92.84%	72.24%
Random Tree	69.08%	100.00%	69.08%

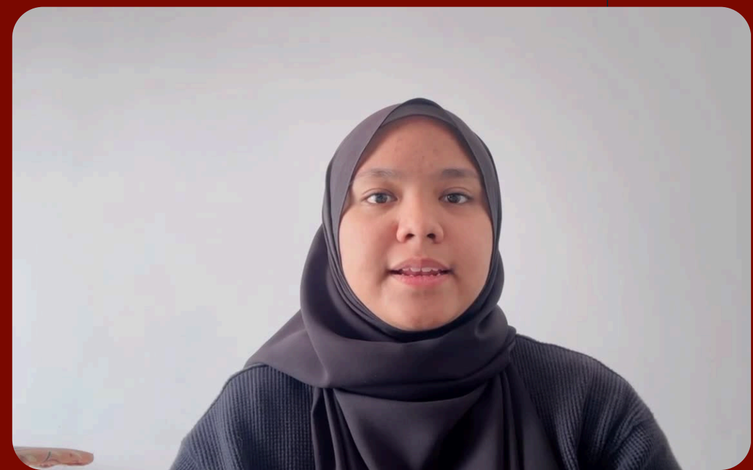
Predictive Analytics

Best Model - KNN

Model	Accuracy	Precision	Recall
KNN	89.18%	90.56%	98.56%

Conclusion

- Customer sentiment towards Uniqlo products was successfully analysed using online reviews.
- Descriptive analysis highlighted differences in customer satisfaction across regions, product types, and sentiment categories.
- Among the predictive models tested, KNN achieved the best performance in classifying customer sentiment.
- The study confirms the effectiveness of sentiment analysis in understanding customer perception from large-scale review data.



Thank You

For Your Attention