Sentiment Analysis: customer reviews of banking services using Python NLP.

Dataset Summary

The project utilized 1500 customer reviews from both Indian and hypothetical North American banks. These datasets included varied feedback on banking services, enabling a comparative analysis of sentiment classification.

Skills Used

Applied statistical and NLP techniques to preprocess text, analyze sentiment, and conduct topic modeling. Evaluated model performance with statistical tests and visualizations, highlighting insights from different geographic datasets.

Technical Skills

Leveraged Python libraries such as **pandas** for data manipulation, **nltk** and **re** for text preprocessing, **scikit-learn** for machine learning models, and **matplotlib** for data visualization. Additionally, **scipy.stats** was used for hypothesis testing, and **Latent Dirichlet Allocation (LDA)** for topic modeling.

Real-World Impact

This project underscores the feasibility of applying traditional Python NLP pipelines across diverse datasets, aiding businesses in extracting actionable insights from customer feedback. The comparative analysis can guide banks in tailoring services to regional customer needs.

Project Steps/Timeline:

1. Data Collection (source: kaggle)
2. Data Preprocessing (lowercasing, tokenizing, and lemmatizing text)
3. Sentiment Analysis (classifying reviews as positive/neutral/negative)
4. Topic Modeling (identifying keywords from reviews in each category)
5. Evaluating sentiment analysis (comparing sentiment to customer numerical reviews)
6. Hypothesis test to determine effectiveness of Python NLP across demographics

Visualization Descriptions

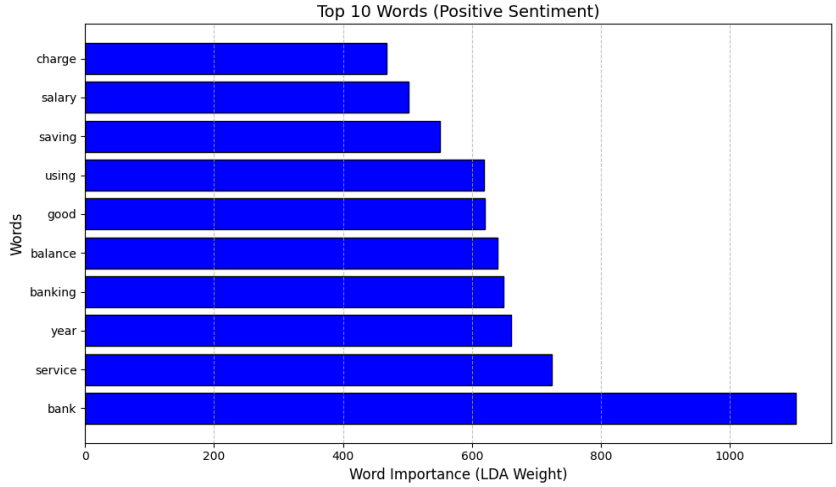
- **Confusion Matrices**: Displays the accuracy of sentiment classification models for Indian and North American datasets, highlighting classification disparities.
- **Bar Charts from LDA Topic Modeling**: Illustrates the most influential topics within positive, neutral, and negative reviews, providing insights into customer concerns and preferences.

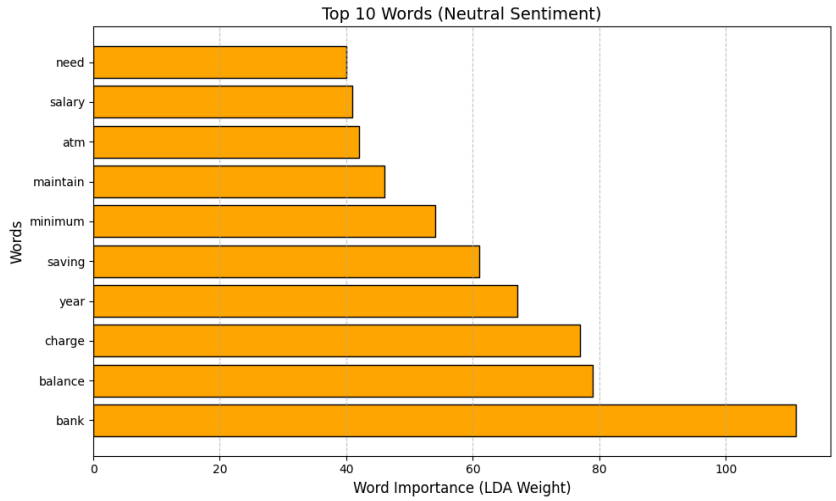
Possible Project Extensions:

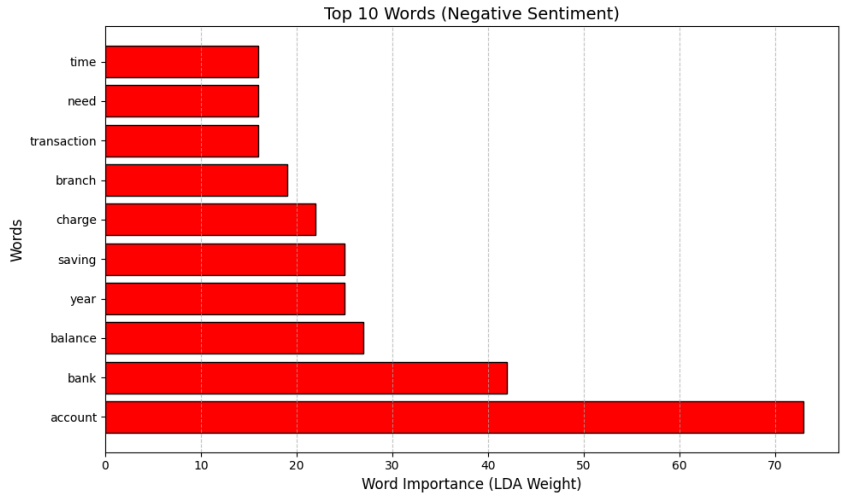
- Standardizing data for hypothesis testing to evaluate the effect of binary data.
- Testing classification disparities for each sentiment.
- Cluster analysis for reviews using keywords for each sentiment.
- Principal component analysis of the entire dataset.

Sentiment Visualizations (first dataset):

Top 10 words associated with each sentiment, ranked by contribution to sentiment interpretation.

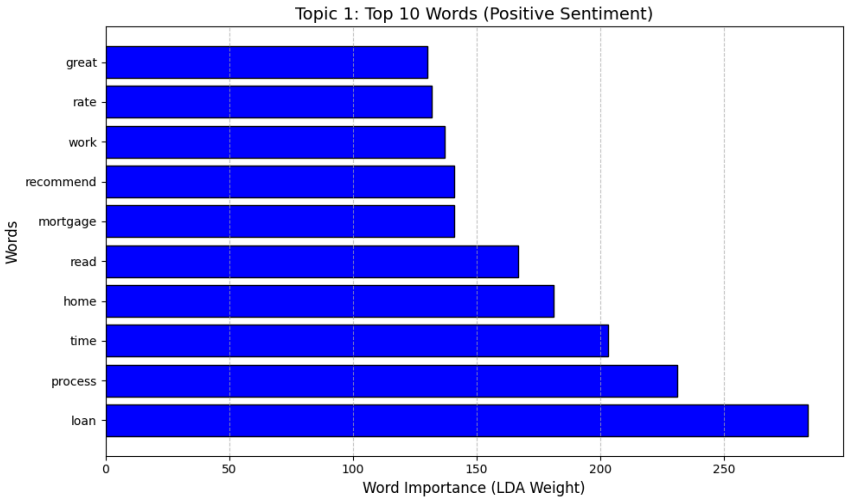


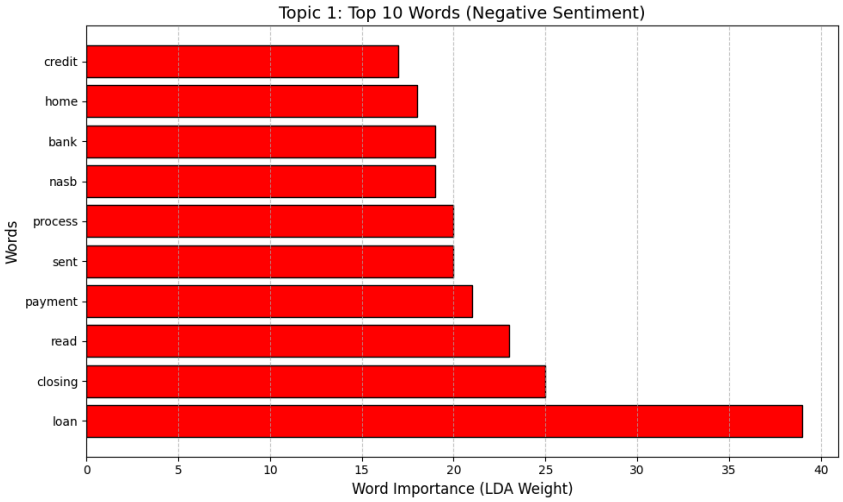




Sentiment Visualizations (second dataset):

Top 10 words associated with each sentiment, ranked by contribution to sentiment interpretation.





NOTE: there were **no neutral reviews** (reviews with 3 stars) in the North American dataset.

Interpretations of Visuals:

Indian Customer Reviews Dataset

- Positive:
    - Top words: *bank*, *service*, *year*, and *banking*
    - Customers likely appreciate the overall **banking experience and services**.
    - *year* might indicate satisfaction with consistent or long-term service.
- Neutral:
    - Top words: *bank*, *balance*, and *charge*
    - customers discuss **financial transactions or fees**, possibly highlighting a matter-of-fact (unemotional approach) toward these aspects.
- Negative:
    - Top words: *account*, *bank*, and *balance*
    - focus on **core banking services**, implying possible dissatisfaction related to account management or discrepancies in balances.

North American Customer Reviews Dataset

- Positive:
    - Top words: *loan*, *process*, and *time*
    - customers likely value the **efficiency and ease of the loan application processes**
    - *home* suggests that home loans might be particularly appreciated.
- Negative:
    - Top words: *loan*, *closing*, *read* and payment
    - Dissatisfaction with **loan closures, payment processes, or documentation clarity**, indicating issues during the **final stages of borrowing or repayment**.

Confusion Matrices:

Indian Customer Review Classification:

	Predicted Positive	Predicted Neutral	Predicted Negative
Actual Positive	777	73	23
Actual Neutral	64	5	2
Actual Negative	42	3	11

North American Customer Review Classification:

	Predicted Positive	Predicted Neutral	Predicted Negative
Actual Positive	392	6	12
Actual Neutral	0	0	0
Actual Negative	51	9	35

Takeaways:

- The model is more effective in identifying language patterns commonly associated with negative sentiment in the North American dataset. Negative reviews in this dataset might contain clearer or more consistent markers of dissatisfaction (e.g., terms like "poor," "bad," "delay").
- In the Indian dataset, negative sentiments may use different or more nuanced expressions, or they may overlap semantically with neutral/positive sentiments, making it harder for the model to distinguish.
- Neutral reviews in the Indian dataset use more nouns than adjectives, leading to **vague sentiment signals**. For example, statements like "bank service" or "account process" might **lack explicit positive or negative modifiers**, making classification ambiguous.
- **Cultural differences in review writing** could lead to an overlap in tone and sentiment between neutral and positive categories.
- If neutral reviews in the Indian dataset are underrepresented or contain unique features not shared with other categories, the model may fail to generalize well.

Statistical Hypotheses:

1. Welch t-test for statistical difference in results

Null Hypothesis (H0): sentiment analysis results are similar across both datasets.
Alternate Hypothesis (H1): sentiment analysis results are different across both datasets.

Results:

- T-Statistic: -63.8959
- Degrees of Freedom: 835.5788
- P-Value: 0.0000e+00
- **Final**: reject the null hypothesis, the means (hence results) are significantly different.

2. F-test for statistically different sentiment structures in results

Null Hypothesis (H0): sentiment structures are similar across both datasets.
Alternate Hypothesis (H1): sentiment structures are different across both datasets.

Results:

- First Dataset: Mean = 0.7930, Standard Deviation = 0.0128, Variance = 0.0002
- Second dataset: Mean = 0.8455, Standard Deviation = 0.0161, Variance = 0.0003
- F-Statistic calculated: 1.5755
- p-value: 5.6918e-09
- Critical Value (95% confidence): 1.1374
- Critical Value (90% confidence): 1.1055
- Final: reject the null hypothesis, sentiment structures are significantly different.