CIMB BANK CIMB ISLAMIC

Merchant Recommendation using Sentiment Analysis to increase profitability

Targeting potential Merchants through organic reviews

Group 2

GreatNusa Data Science Training

May 2023

EPICC **Live an EPICC Life**

Meeting agenda

01

Overview

- Background
- Proposed Solution
- Objective
- Potential Benefit

02

Data, Analysis & Results

- Approach & Data Overview
- Model Result
- Scoring & Category

03

Conclusion

Overview – Background, Proposed Solution & Potential Benefits

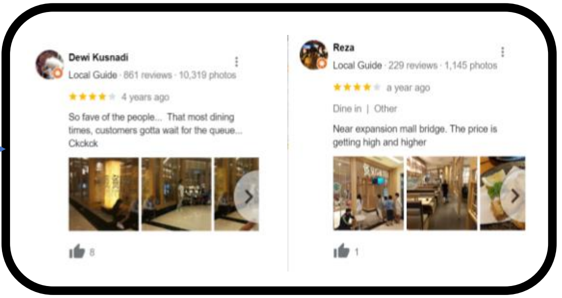
Background

Merchant Code Category ***(MCC) Analysis*** provides an overview of Customer CC Spending Behavior

Targeted campaigns & ***Identify merchants*** for partnership

However, CIMB Niaga ***only use Internal data*** for Analysis

Can we ***leverage External Data*** to ***better identify*** which merchant to approach?



****Give 4 out of 5 Stars but negative review***

However, ***scoring*** from ***external data*** might ***not*** be ***reliable***

Proposed Solution

1 **Sentiment Analysis**

predict potential merchants that will have positive sentiments in the future

2 **Scoring Sentiment**

Comprehensive understanding the consistency of the sentiment each merchant

3 **Category of Quality Drivers Prediction**

Comprehensive predict of categorize ***that drives the sentiments*** of each merchant.

Objective

1 Enhance CIMB Niaga's ***ability to identify potential merchants***

2 ***Increase Profitability*** through credit-card and merchant partnership campaigns

Potential Benefit

1 ***Revenue Uplift*** from CC Transaction

2 ***NTB Acquisition***

3 ***Reducing Campaign Cost*** as a result of better targeting

Approach & Data Overview

Approach

1 Identify Top MCC & Region

MCC Category:

- Restaurant
- Clothing Store
- Region: Jakarta Pusat

2 Extract Data from External Data

Initial Data: Google Reviews

Filter by Rating >3

3 Sentiment Analysis Modelling

Tested 4 Models:

- Random Forest
- Decision Tree
- MLP
- Naïve Bayes

4 Create Scoring to Capture Trend Sentiment

Scoring Criteria:

- Years Passed
- Probability Positive Sentiment

5 Categorize what drives the sentiment

Category:

- Quality of Service
- Price
- Convenient Place
- Quality of Products

6 Identify Potential Merchants

Data Overview

Extracted Data from Google Reviews
4.9 K

| 2.5 K
Restaurant

| 2.4 K
Clothing Store

Data Preprocessing & Model Selection

Data Preprocessing

1. Lower Text
2. Remove Numbers
3. Remove Punctuation
4. Stemming
5. Vectorization (TfidfVectorizer)

Model Selection

- Multi Classification Model (Target : Positive, Neutral, Negative)
- Train & Test Split = 0.8 : 0.2
- Total Dataset for Training & Testing = 16,972

Data	Target			Total
	Positive	Neutral	Negative	
Data Train (80%)	5,896	5,896	5,236	13,577
Data Test (20%)	1,474	1,309	612	3,395
Total Dataset (100%)	**7,370**	**6,545**	**3,057**	**16,972**

	Random Forest	Decision Tree	MLP	Naïve Bayes
Parameter	Value	Value	Value	Value
Accuracy	0.82	0.78	0.77	0.53
Precision	0.81	0.78	0.77	0.68
Recall	0.82	0.78	0.77	0.53
F-1 Score	0.82	0.78	0.77	0.53
Cross-Validation	0.80	0.76	0.75	0.51

Key Highlight: Therefore, for **Sentiment Analysis**, we are using ***Random Forest Model***, since it has the **highest model performance result**.

Scoring Sentiment to Capture Trend

To ***capture the sentiment trend*** of each merchant, we use ***scoring*** to assign weight based on last review year. The most **recent year will have higher weight** than previous year

Year	Weight	Probability of Positive Sentiment	Scoring Per Year
Year 0 & - 1 (2023 – 2022)	45%	Total Positive Sentiment / Total Review In respectively each year's group	0.45 * Probability of positive sentiment
Year -2 & -3 (2021 – 2020)	35%		0.35 * Probability of positive sentiment
<= Year -4 (<=2019)	20%		0.2 * Probability of Positive Sentiment
Total Scoring			Total Scoring each Year's Group

Capture & Prioritize merchants who are ***growing & consistently*** have positive sentiment.

Illustration

Shabu-Shabu Le-Ta-Su

Location = Grand Indonesia
Total Review = 89

Year	Total Review	Total Positive Sentiment Review
Year 0 & -1 (2023 – 2022)	58	49
Year -2 & -3 (2021 – 2020)	14	8
<= Year -4 (<=2019)	17	10
Total	89	67

Higher Weight

Year	Weight	Probability of Positive Sentiment	Scoring per Year
Year 0 & -1 (2023 – 2022)	45%	49/58 = 0.845	0.45 * 0.845 = 0.380
Year -2 & -3 (2021 – 2020)	35%	8/14 = 0.571	0.35 * 0.571 = 0.2
<= Year -4 (<=2019)	20%	10/17 = 0.588	0.2 * 0.588 = 0.118
Total Scoring			0.7

From this scoring we were able to see the **positive sentiment trend** for **Shabu-Shabu Le-Ta-Su**.
We can conclude that this merchant **consistently have positive sentiment.**

Sentiment Drivers

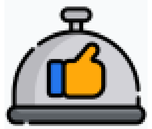
Now we were able to identify Merchants who consistently have positive sentiment or consistently have negative sentiment.
To ***analyze further***, we categorize ***what drives these sentiments***.

Positive Sentiment

Negative Sentiment

Using these word clouds, we categorize into **4 groups:**

Price



Quality of Product

Quality of Service

Convenient Place

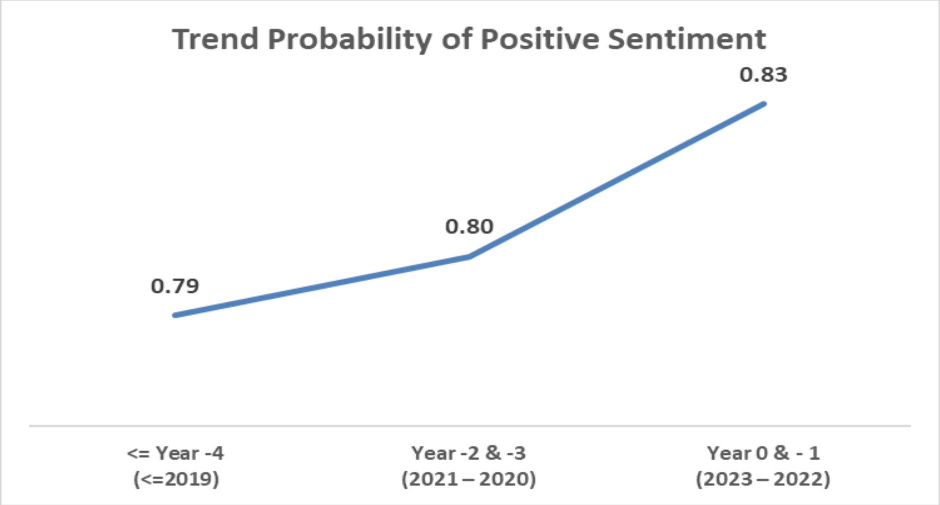
Since there are still no data training for these qualities categorization,
therefore, we are using ***indexing*** to ***categorize the qualities***.

Impact of Categories to Sentiment

Consistent Scoring

Grand Café
Total Scoring = 0.81

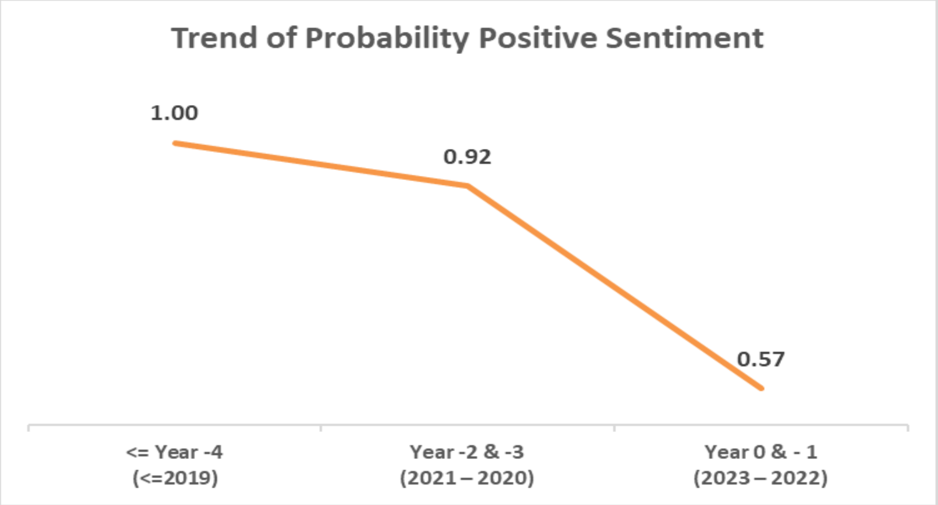
Quality of Service	Quality of Products	Price	Convenient Place
	V		V



Inconsistent Scoring

McDonalds
Total Scoring = 0.77

Quality of Service	Quality of Products	Price	Convenient Place
			V



Key Highlight*:

For ***restaurant merchant***, that have ***at least 2 qualities***, usually have ***increasing or stable positive sentiment value***.
However, ***restaurant merchant*** that only have ***1 qualities*** usually have ***inconsistent*** of ***positive sentiment value***.

**Disclaimer: Result may differ depending on the MCC*

Result: Top Merchant – Based on MCC

From our modelling, we recommend CIMB Niaga to tap into the following merchants:

	Restaurant					Clothing Store				
Merchant Name	**Sushi Tei**	**Grand Cafe**	**McD**	**Kayu Manis**	**Shabu – Shabu Le-Ta-Su**	**League**	**Polo**	**Uniqlo**	**The Executive**	**Matahari**
Total Scoring	**0.84**	**0.81**	**0.77**	**0.74**	**0.69**	**0.93**	**0.75**	**0.72**	**0.72**	**0.63**
Quality of Service	✓									
Quality of Products	✓	✓		✓	✓	✓	✓	✓	✓	
Price										✓
Convenient Place	✓	✓	✓			✓				

Key Highlight: Across ***Restaurant*** and ***Clothing Store*** MCC, the main driver is ***Quality of Products.***
For ***Restaurant***, merchant with 1 driver have tendency to have ***inconsistent positive sentiment*** value

Conclusion & Whats Next

Conclusion

1. Sentiment Analysis model used is Random Forest. The model has an ***Accuracy*** of **0.82** & ***Cross-Validation*** of **0.80**

2. Restaurant merchant that have ***at least 2 drivers*** tend to have ***increasing or stable positive sentiment value***, while merchants with ***1 driver*** tend to be more ***inconsistent.*** This result may **differ** depending on the **MCC**.

3. For MCC Restaurant & Clothing Store, the ***main driver*** is ***quality of products***

4. In total, there are ***10 Merchants*** that are recommended for tap-in, ***5 for Restaurants*** and ***5 for Clothing Store***

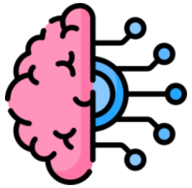
Recommendation & Next Action Plan

Tap into more external Data

For Future exercise there is possibility to use social media data such as Instagram & Twitter

More Specified MCC

Possibility to expand MCC into more defined sub-categories for better analysis.



For future categorization qualities driver prediction will use modelling

Since data training is already available from this exercise, we no longer need to use indexing.

Thank You

Appendix

Data	Target			Total
	Positive	Neutral	Negative	
Data Train (80%)	5,896	5,236	2,445	13,577
Data Test (20%)	1,474	1,309	612	3,395
Total Dataset (100%)	7,370	6,545	3,057	16,972

Year	Weight	Probability of Positive Sentiment	Scoring per Year
Year 0 & -1 (2023 – 2022)	45%	Total Positive Sentiment / Total Review in respectively each year's group	0.45 * Probability of positive sentiment
Year -2 & -3 (2021 – 2020)	35%		0.35 * Probability of positive sentiment
<= Year -4 (<=2019)	20%		0.2 * Probability of positive sentiment
Total Scoring			Total of Scoring each Year's Group

Quality of Service	Quality of Products	Price	Convenient Place
	V		V
Quality of Service	**Quality of Products**	**Price**	**Convenient Place**
			V