

BE YOU, BE TRUE

Sentiment Analysis on Malaysia's Airline Service

CS2595A

Presented by: Team Gadis Gojes



Summary of study

- Airlines are the biggest transport services offered in Malaysia air transportation
- Airlines may provide scheduled and/or chartered services
- The airline sector is an important aspect of the broader travel industry
- Customer opinions are important for airline companies to maintain and improve their reputation
- No specific platform for the companies to review their performances although customer experiences are the most crucial part of the company's growth



Problem Statement



Reviews overload and unorganised

Too many reviews can lead to information overload, so it is hard for a good review system to identify the ones that are beneficial for the traveler



Time-consuming and difficult to identify the sentiment of the reviews.

Millions of passengers flying and sharing their experiences on social media and online review platforms create massive amounts of data. Analysing those data requires lots of time for the customers to process all the reviews in order to make a decision





To develop a sentiment analysis model on Malaysia's Airlines Services.



To evaluate the best algorithm to make sentiment analysis for Malaysia's Airlines Services reviews in terms of its accuracy.



To identify Malaysia's airline company that provided the best services.



Background of Study

Literature Review: Airlines Services

Airlines services are known as businesses that specialized in air transportation for people and also cargo. The airline sector is worth billions of dollars, and millions of passengers use this service daily to get from one location to another. Many start-ups and established airline companies are finding it challenging to survive in the current industry due to the intensifying competition. Today, establishing consumer loyalty presents a formidable obstacle and is the primary concern of businesses (Annie et al., 2015).



Background of Study

Literature Review: Related works

There had been many previous studies on sentiment analysis in various fields. Some of the most popular machine learning models that have been used for sentiment analysis are Support Vector Machine (SVM) (Martinez-Camara et al.,2017; Amanda and Negara, 2020; Muhammad et al.,2019), and K-nearest neighbour (KNN) (Huq et al.,2017).

Martinez-Camara et al. (2017) performed various experiments for sentiment analysis and achieved the highest accuracy using the SVM model of 92.71%. Based on the study, KNN performed better than SVM but SVM performs better when the number of dimensions is very high. KNN achieved an accuracy of 90% whereas SVM achieved 77.97%.



Background of Study

Sentiment Analysis

Sentiment analysis or opinion mining is a natural language processing (NLP) technique used to determine the sentiments of data. The internet users around the globe were estimated to be more than 4536 million by June 2019 had produced a huge amount of public reviews. Companies, governments, customers and other research parties are interested to gain benefits from the analysis of the reviews for their decision-making and feedback analysis (Nejjari & Meziane,2020). This is where sentiment analysis and opinion mining come into place.



Background of Study

Sentiment Analysis

Tiwari et. al. (2019) stated that sentiment is another term for a person's or person's point of view or opinion. The sentiment could represent joy, happiness, sadness, or even anger. This is what the reviews of travellers are all about. Every flight on an airline can bring either pleasure or discomfort to any passenger. If the traveller is dissatisfied with the services, his reviews reflect that dissatisfaction. If he is completely satisfied with the services, he will express his joy in his reviews. As a result, reviews not only help airlines solve individual passenger problems, but also help them improve their services. The exceptions are only one or two people. The majority of travellers' opinions are important to consider. To understand the psychology and opinions of the majority of travellers, we must consider everyone's point of view, which is a nearly impossible task. As a result, it is critical to employ a technique capable of analysing such large amounts of review data.



Methodology

- Business Understanding
- Data Understanding
- Data Preparation
- Modelling
- Evaluation



Business Understanding

What do we want to do?

Analyze the customer's opinion
on Malaysia airline's services.

What should we do?

Collecting the data of customer reviews
from various website platforms to make a
sentiment analysis. The collected data will
be categorized into positive and negative
sentiments.



Data Understanding

Darling I know, I should have
left long ago — but you're so
loving. I had to watch
you running
important-like
But it looks so
suited & desire
I do



Data Understanding

The data, which contains reviews of three Malaysian airline services, was scraped from the websites during this stage.

Sources:

- TripAdvisor
- TrustPilot
- Product Review
- AirlineQuality
- Airlineratings

Web Scraping Tool:

- Web Scraper Extension



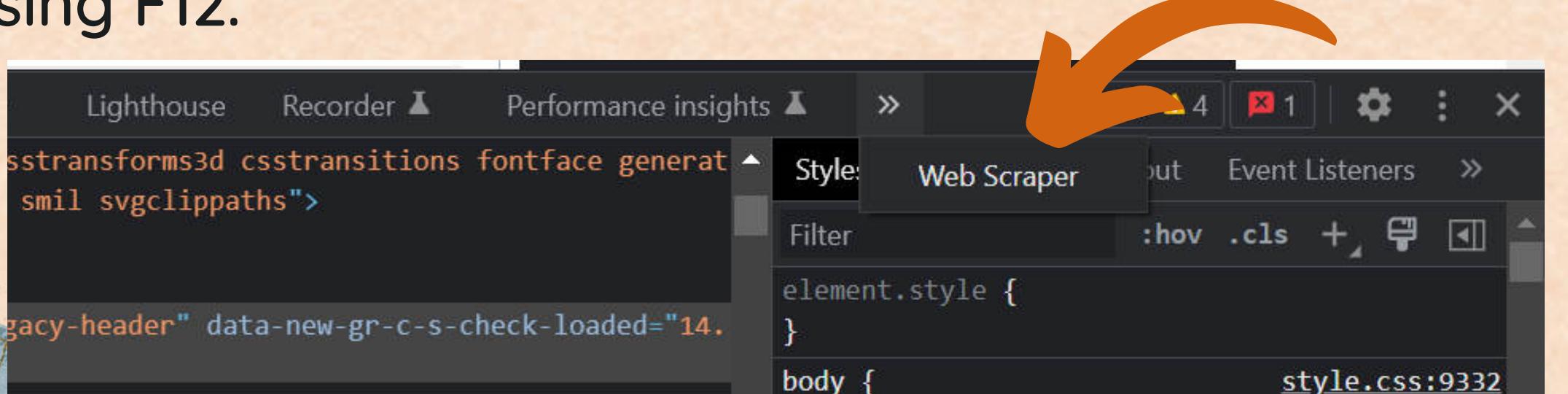
Photobooth

Web Scraping Process

1. Install the Web Scraper extension from Chrome Web Store.



2. Navigate to the desired website and launch Web Scraper by pressing F12.



Web Scraping Process

3. Create a new sitemap and fill in the required details.

Sitemaps Sitemap ▾ Create new sitemap ▾

Sitemap name malindo

Start URL <https://www.airlinequality.com/airline-reviews/malindo-air/>

Create Sitemap

4. Click on "Add new selector" button then fill required details and choose multiple selector to select many reviews at one time.

Sitemaps Sitemap malindo ▾ Create new sitemap ▾

_root

ID	Selector	type	Multip

Add new selector

Sitemaps Sitemap malindo ▾ Create new sitemap ▾

Id reviews

Type Text

Selector Select Element preview Data preview

Multiple

Regex

Parent Selectors _root

Save selector **Cancel**

Web Scraping Process

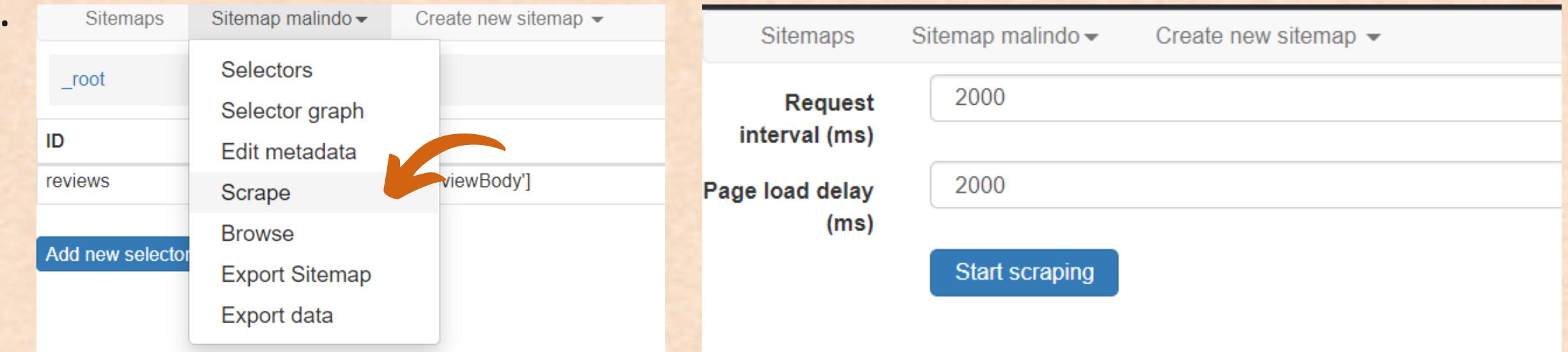
5. Select the related element on the page then after that click "Done selecting" and "Save selector".

The screenshot shows a web browser window with the following elements:

- Header:** Airline Reviews, Airport Reviews, Write A Review, Search, Airport & Airline Rating, Skytrax Awards, Information.
- Review Content:** A red box highlights a warning message: "Not Verified | Purchased \$2200 of fares to Bali before Malindo ceased Australian operations. Refund requests submitted but ignored. When pushed, they will promise to process but they then go to ground and ignore you." Below it are travel details: Type Of Traveller (Couple Leisure), Seat Type (Economy Class), Route (Melbourne to Bali), Date Flown (July 2021), Value For Money (1 star), Recommended (No). A large orange arrow points from the "Done selecting" button in the SelectorGadget overlay to the "Value For Money" rating.
- Rating:** 1/10
- Text Box:** "Only advice is don't book any flights with this company!" by John Brauns (Australia) 15th May 2022. A red box highlights the text: "Not Verified | Because of the pandemic, our trip/flights to Bali was cancelled by the company. They dont give any money back but give travel vouchers. At the moment, we are now in a position to rebook our trips. The voucher says we have till the 20th May to rebook. I tried to rebook on the 12 May and entered the Travel voucher number only to be told that these vouchers had been used already! I have emailed Malindo in the past and it usually takes them a few days to respond. Over \$1000 lost! No point trying to ring them as others say / flights with this company!"
- SelectorGadget Overlay:** A floating UI for selecting CSS selectors.
 - Toolbar:** div[itemprop='reviewBody'], P, C, S, Done selecting, Save selector, Cancel.
 - Multiple checkbox:** Multiple
 - Regex checkbox:** Regex
 - Parent Selectors:** _root
 - Buttons:** Save selector, Cancel.
- Bottom Navigation:** Elements, Console, Sources, Network, Performance, Memory, Application, Security, Lighthouse, Recorder, Performance insights.
- Sitemaps:** Sitemaps, Sitemap malindo, Create new sitemap.
- Console:** Console, What's New, Issues.

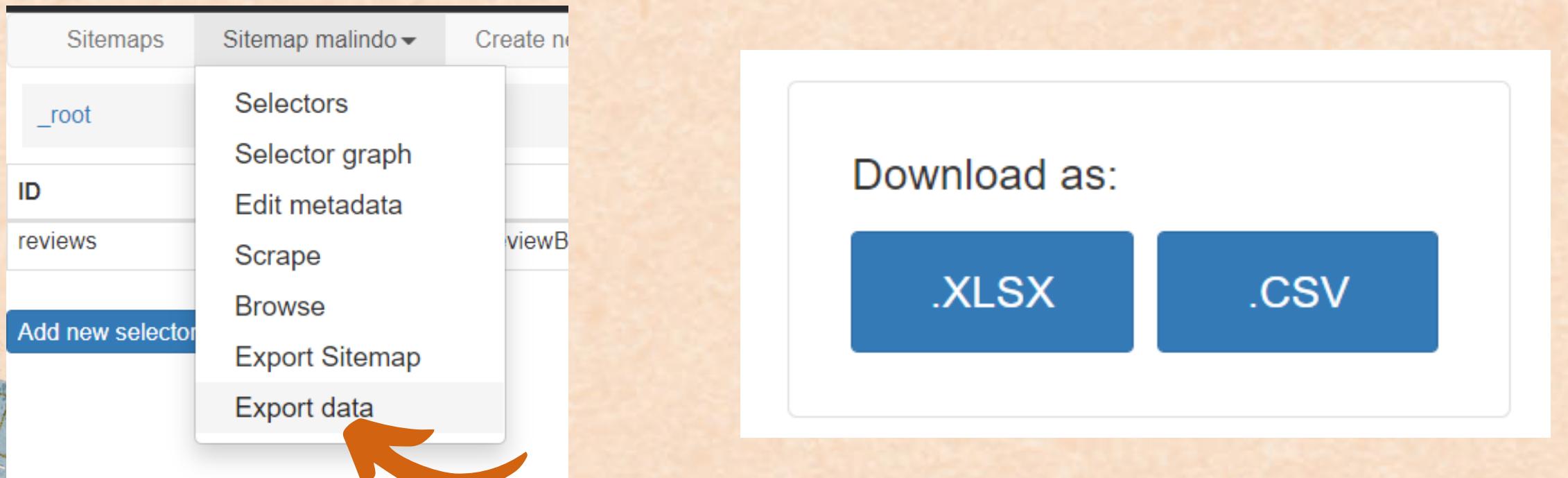
Web Scraping Process

6. To scrape the data, click on the "Scrape" option and the data will be scrapped.



The screenshot shows a user interface for managing sitemaps. On the left, there's a sidebar with 'Sitemaps' and a dropdown 'Sitemap malindo'. Below these are entries for '_root', 'ID', and 'reviews', with a blue button 'Add new selector' at the bottom. A context menu is open over the 'reviews' entry, listing options: 'Selectors', 'Selector graph', 'Edit metadata', 'Scrape', 'Browse', 'Export Sitemap', and 'Export data'. An orange arrow points from the text 'click on the "Scrape" option' to the 'Scrape' menu item. To the right of the sidebar is a main panel with 'Sitemaps', 'Sitemap malindo', and 'Create new sitemap' buttons. It contains two input fields: 'Request interval (ms)' set to 2000 and 'Page load delay (ms)' set to 2000. A large blue button labeled 'Start scraping' is at the bottom.

7. To save the data, click on the "Export data" option then the data can be downloaded as XLSX or CSV.



This screenshot shows the same interface as the previous one, but the 'Export data' option in the context menu is now highlighted with an orange arrow. To the right, a modal window titled 'Download as:' offers two choices: '.XLSX' and '.CSV', each in a blue button.

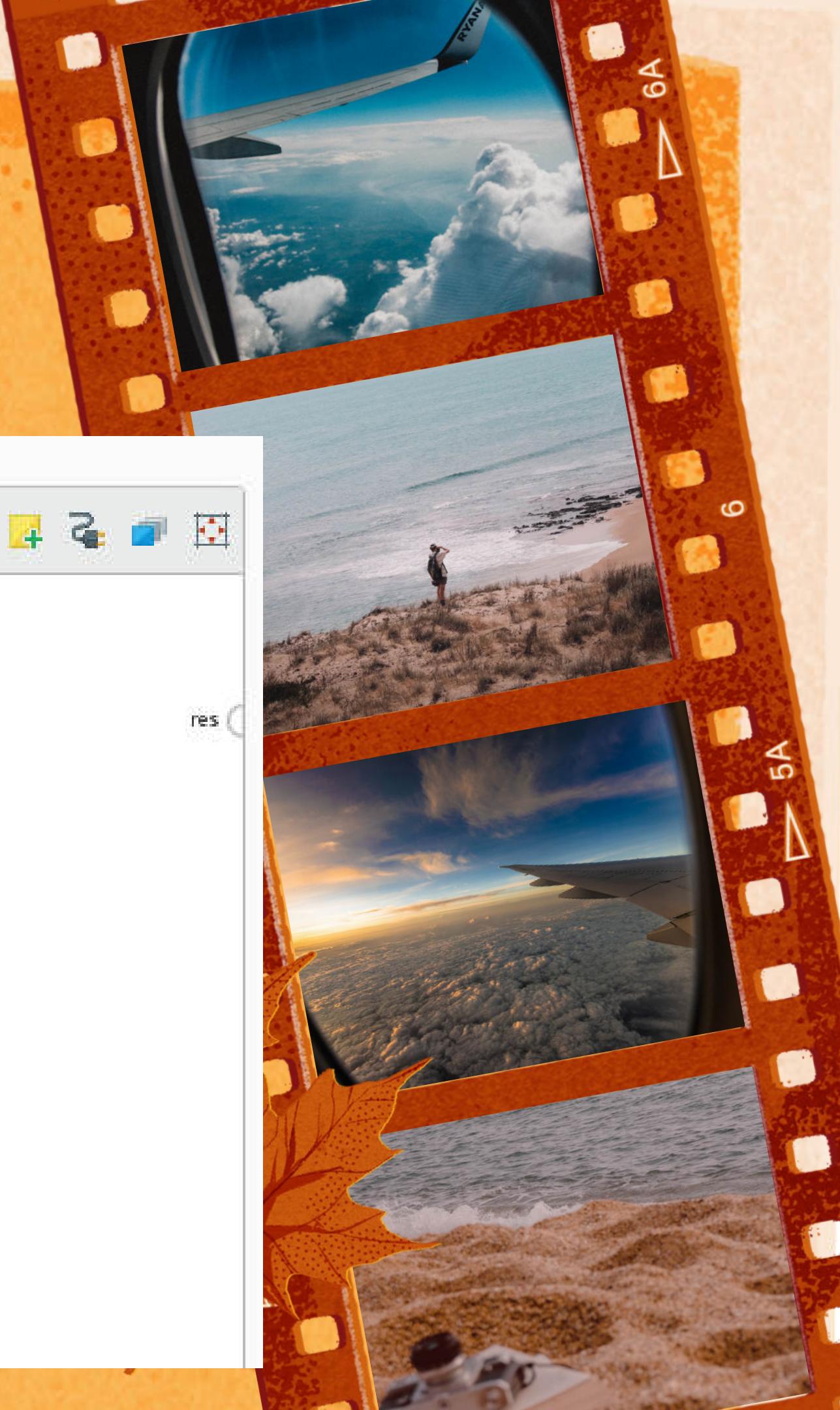
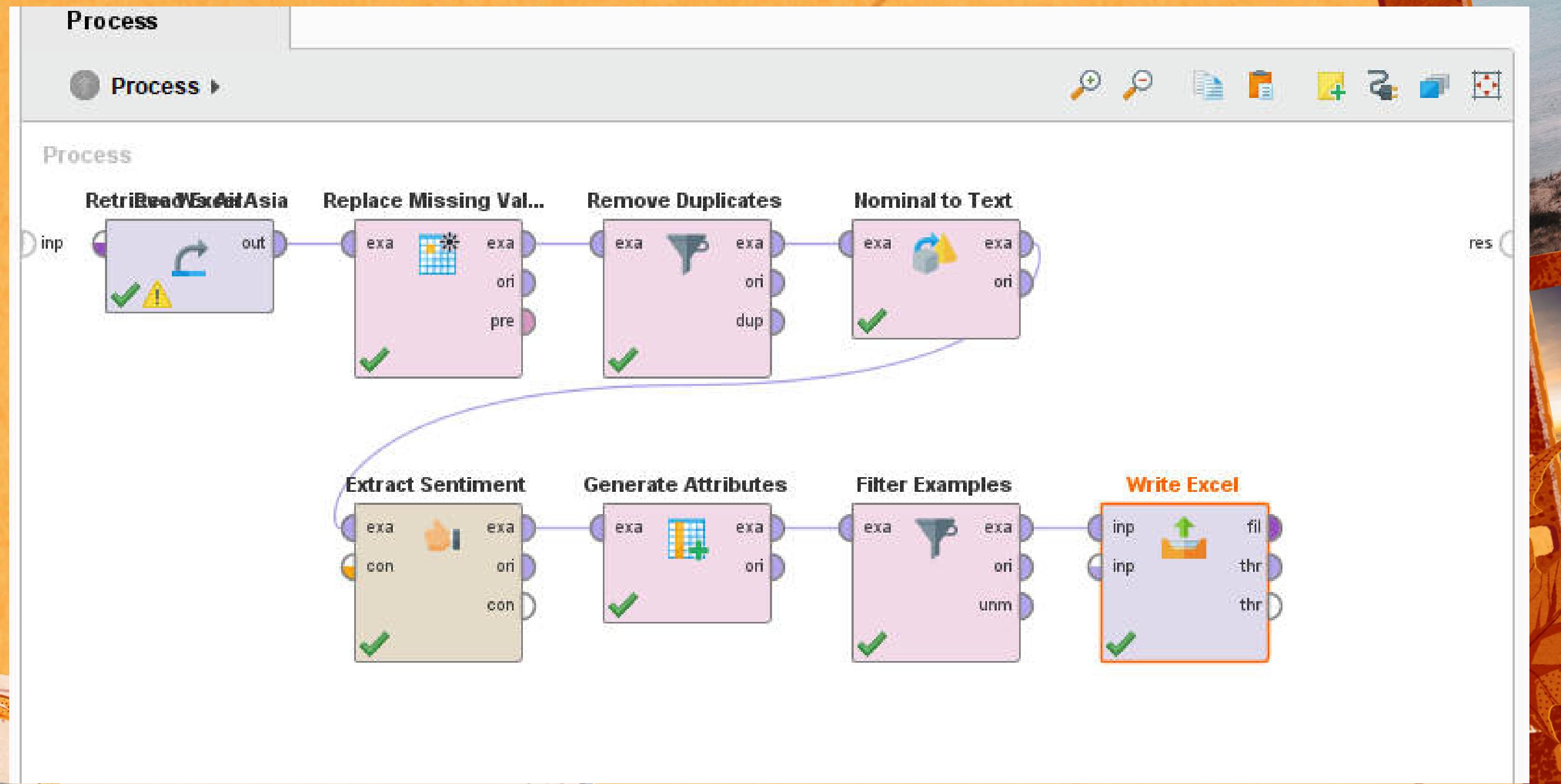
Data Preparation

Dear... I know, I should have
left long ago - but you're so
nice. I had to watch
and so

I your running
important like
But it looks so
quite a scene
to...



Data Preparation



Data Preparation

1. Replace missing value

- Malindo Air: None
- Malaysia Airlines: None
- AirAsia: 7 missing values



Data Preparation

2. Remove duplicates

- Duplicate reviews are removed
- Malindo Air : 1100 > 1100 (0 duplicate reviews)
- Malaysia Airlines: 1334 > 1235 (99 duplicate reviews)
- AirAsia: 1953 > 1903 (43 duplicate reviews)



Data Preparation

3. Nominal to Text

- Change the nominal data type into text data type



Data Preparation

4. Extract Sentiment

- Label the data into positive and negative reviews
- Positive Keywords: great, good, excellent, nice, well done, pleasant, fast
- Negative Keywords: Terrible, avoid, fraud, not pleasant, worst, disappointed, dreadful, bad, hate, horrible, sucks, poor, delay, refund



Data Preparation

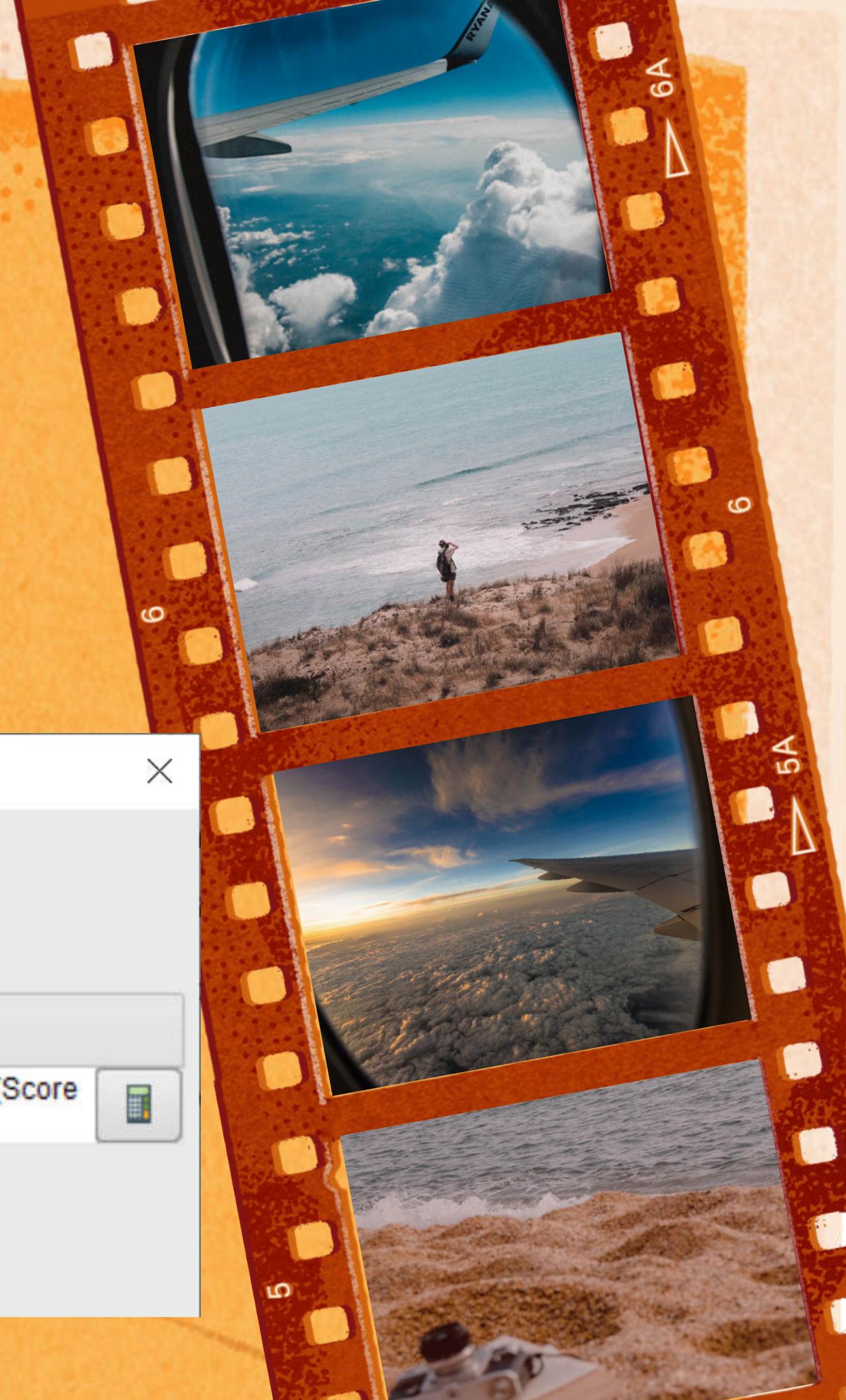
5. Generate Attributes

- The Generate Attributes operator constructs new attributes from the attributes of the input ExampleSet and arbitrary constants using mathematical expressions.

Edit Parameter List: function descriptions

 Edit Parameter List: function descriptions
List of functions to generate.

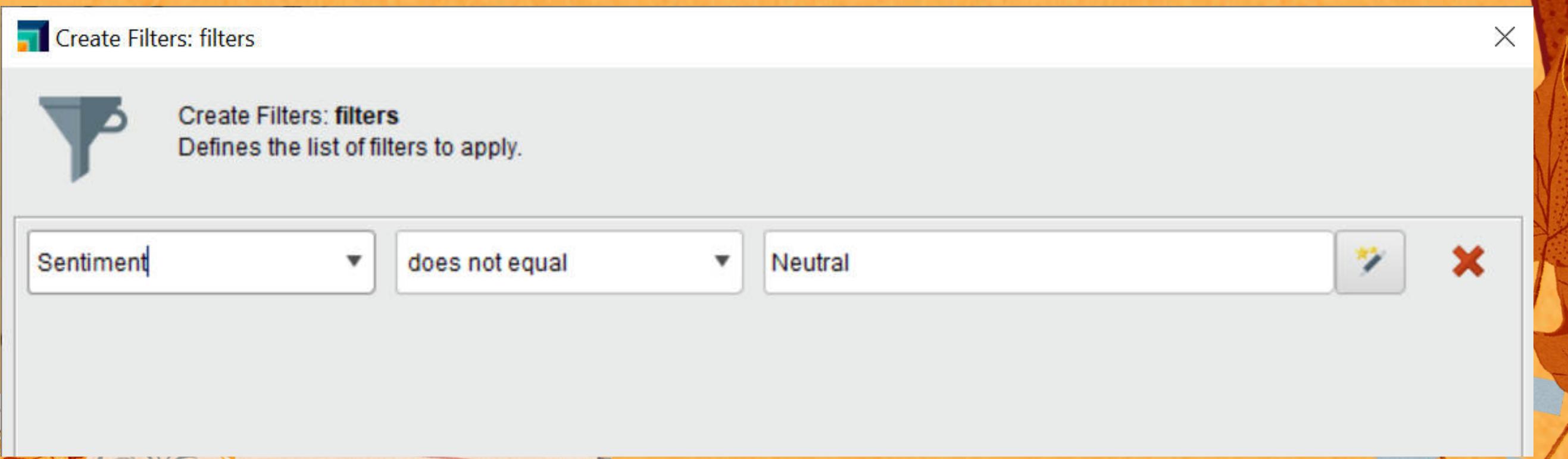
attribute name	function expressions
Sentiment	<pre>if(Score > 0, "Positive", if(Score < 0, "Negative", if(Score == 0, "Neutral", "")))</pre>



Data Preparation

6. Filter example

- (sentiment > does not equal > neutral)
- Malindo: 1100 > 1000 (remove 100 neutral reviews)
- Malaysia Airlines: 1235 > 1217 (remove 18 neutral reviews)
- AirAsia: 1903 > 1752 (remove 151 neutral reviews)



Data Preparation

7. Write Excel

The result will be converted into a XLSX file

8. Total of data

Dataset	Total Data	Positive	Negative
Malaysia Airlines	1217	756	461
AirAsia	1752	903	849
Malindo Air	1000	468	532



Data Preparation

Before

Row No.	reviews
1	Terrible, they cancelled our flight
2	Water and all liquids confiscated at gate even though purchased after security and sealed. No entertainem...
3	Avoid at all cost.
4	Flight departure time been rescheduled from 9:50am to 1:50pm and now to 2:30pm. Wasted one day & itine...
5	Pros:
6	This is a fraud company. They will cancel the ticket continuously and will not pay any money back. Don't try y...
7	They cancelled the entire flight few hours after I purchased the ticket. No reply whatsoever from their side for...
8	Guys beware. Never ever book flights on Malindo. They advertise flights for half the market price, and close t...
9	Because of a glitch in the website, two payments were made. However, even though we have reached out m...
10	I had an issue with my booking for 4 passengers not only did they charge me twice but also not willing to re...
11	Booked a ticket from Bali to Bengaluru and the flight got canceled.
12	Used twice, customer service is non-existence. No reply from email, SG number from the app is not even w...
13	Malindo cancelled my flight due to covid and offer travel voucher with redeem date within half year. I requeste...

ExampleSet (1,100 examples, 0 special attributes, 1 regular attribute)

After

Row ...	Score	reviews	Sentiment
30	0.385	Kuala Lumpur to Perth. Low cost/no frills - that's OK. However: 1) Mali...	Positive
31	-1.897	I was supposed to board a 3.50pm flight today and it was delayed unt...	Negative
32	2.872	Malacca to Penang. Nice little airline from Malaysia. The short flight on...	Positive
33	-5.359	We paid \$842 for 2 return flights from Brisbane to Bali for my birthday ...	Negative
34	0.692	Totally unreliable! The return flight was changed one full day, as if it do...	Positive
35	1.667	I was a bit concerned about terrible reviews I had been reading, so di...	Positive
36	-2.077	Kuching to Kuala Lumpur. I am disappointed with the management o...	Negative
37	-0.795	Singapore to Varanasi via Kuala Lumpur . I purchased baggage for m...	Negative
38	-0.385	Purchased \$2200 of fares to Bali before Malindo ceased Australian o...	Negative
39	1.564	Perth to Penang via Kuala Lumpur. Was totally dreading this flight. Bo...	Positive
40	-1.256	Kochi to Denpasar via Kuala Lumpur. Absolutely ridiculous luggage p...	Negative
41	-2.128	Kuala Lumpur to Denpasar. This flight delay of 1 hour with no compe...	Negative
42	1.462	Kuala Lumpur to Perth. Boeing 737s are not, perhaps, everyone's ch...	Positive

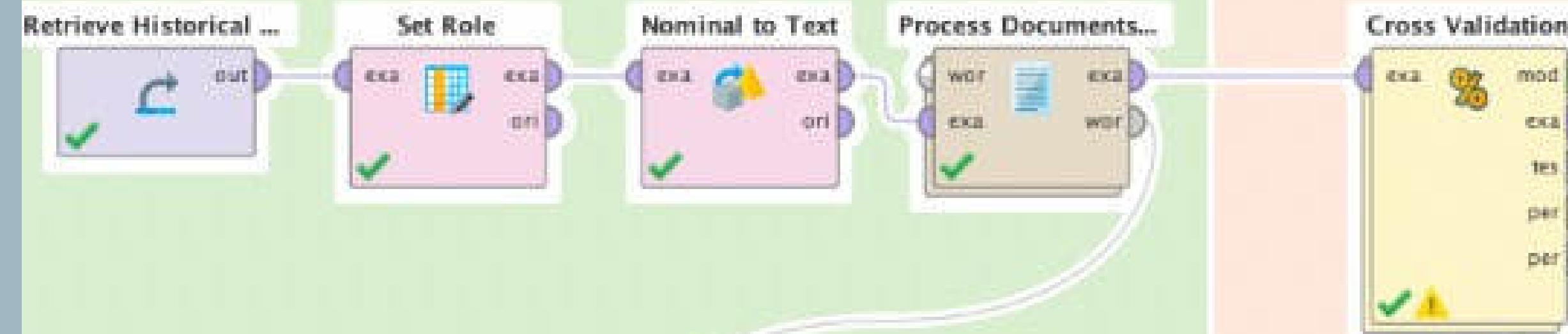
ExampleSet (1,000 examples, 1 special attribute, 2 regular attributes)

Modelling

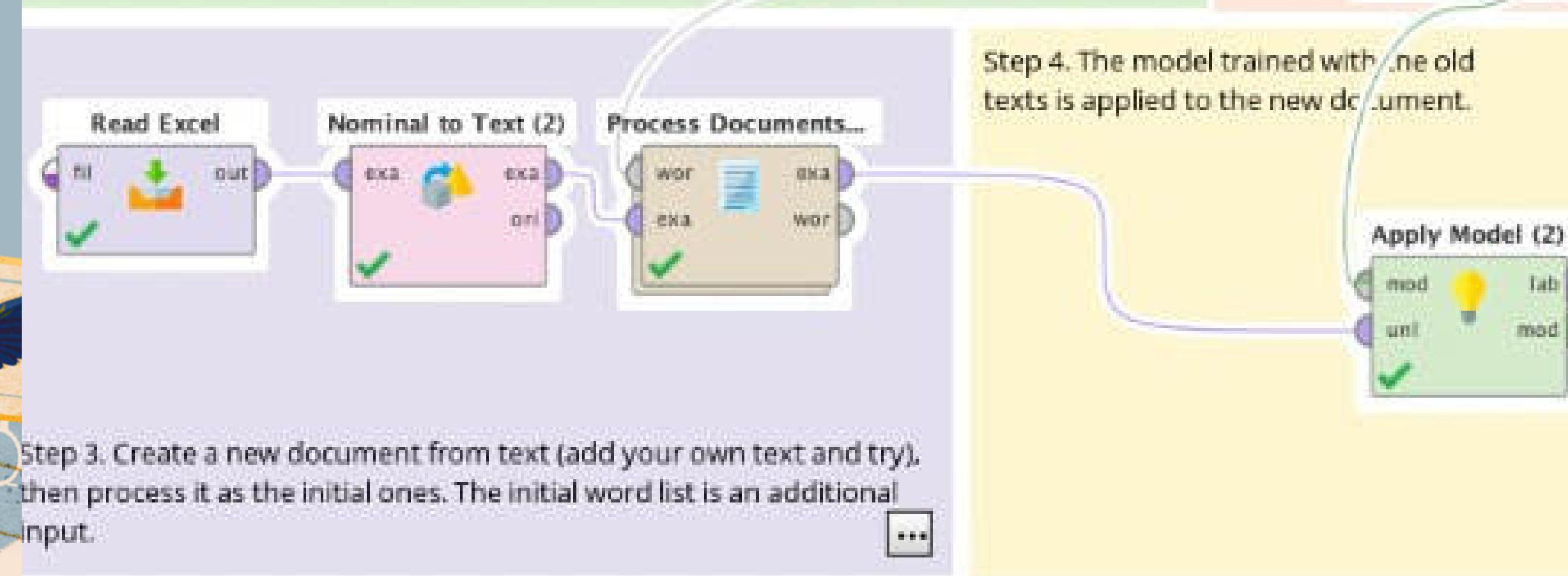
The development of the model occurred and will be using the Support Vector Machine and K-Nearest Neighbour. Since these two models produced a high accuracy based on literature reviews.



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 SVM model and validate it to collect the performance data.



Step 4. The model trained with the old texts is applied to the new document.



Evaluation

A confusion matrix will be implemented in this study to analyse the accuracy produced by both SVM and k-NN models on our datasets. Below is how the accuracy is calculated from the confusion matrix:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$



Experiment



Accuracy Result

AirAsia

SVM

accuracy: 72.86% +/- 1.76% (micro average: 72.86%)			
	true Negative	true Positive	class precision
pred. Negative	416	58	87.76%
pred. Positive	390	787	66.66%
class recall	51.61%	93.14%	

K-NN

accuracy: 79.53% +/- 1.91% (micro average: 79.53%)			
	true Negative	true Positive	class precision
pred. Negative	611	143	81.03%
pred. Positive	195	702	78.26%
class recall	75.81%	83.08%	



Photobooth

Accuracy Result

Malaysia Airlines

SVM

accuracy: 74.35% +/- 2.95% (micro average: 74.35%)			
	true Positive	true Negative	class precision
pred. Positive	634	242	72.37%
pred. Negative	43	192	81.70%
class recall	93.65%	44.24%	

K-NN

accuracy: 75.52% +/- 3.70% (micro average: 75.52%)			
	true Positive	true Negative	class precision
pred. Positive	557	152	78.56%
pred. Negative	120	282	70.15%
class recall	82.27%	64.98%	



Photobooth

Accuracy Result

Malindo Air

SVM

accuracy: 77.80% +/- 3.88% (micro average: 77.80%)			
	true Negative	true Positive	class precision
pred. Negative	459	149	75.49%
pred. Positive	73	319	81.38%
class recall	86.28%	68.16%	

K-NN

accuracy: 73.00% +/- 3.65% (micro average: 73.00%)			
	true Negative	true Positive	class precision
pred. Negative	417	155	72.90%
pred. Positive	115	313	73.13%
class recall	78.38%	66.88%	

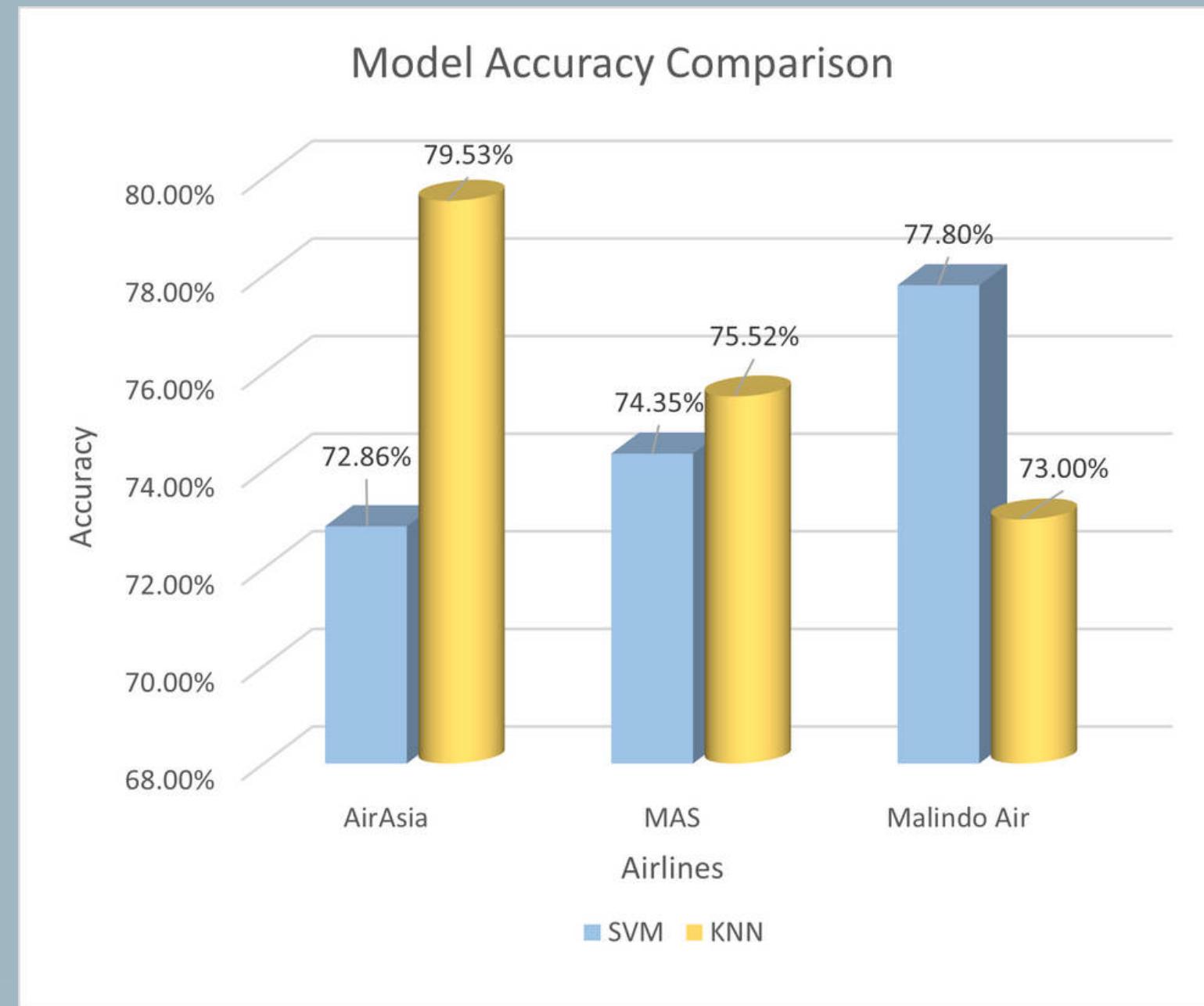


Photobooth

Model Accuracy Comparison

Comparison of accuracy between SVM and k-NN.

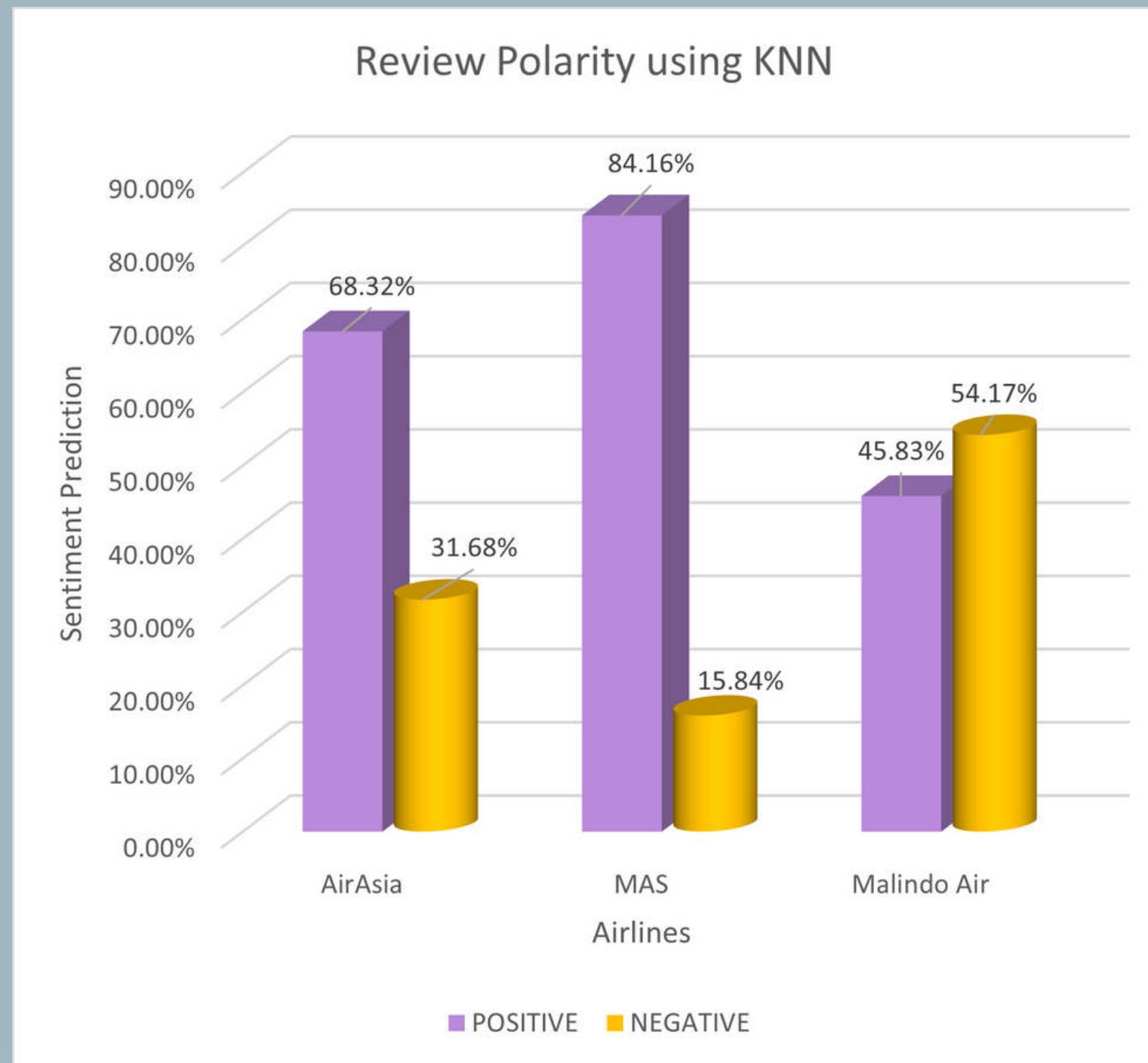
Conclusion: Here, we can see that the k-NN model produced the highest accuracy compared to the SVM model for our study. So, we decided KNN is the best model to be applied to our datasets.



Review Polarity Result Comparison

Comparison of polarity between Malindo, AirAsia and Malaysia Airlines

Conclusion: MAS is chosen as the best airline services with a positive sentiment prediction of 84.16%.



Conclusion

We have managed to develop a sentiment analysis model for Malaysia's airlines services. Based on the evaluations, the best algorithm to perform sentiment analysis for Malaysia's airlines is k-NN, which has produced higher accuracy compared to SVM. To conclude, Malaysia Airlines has the best reviews, while Malindo Air has the worst service reviews.



REFERENCES

1. AIRLINE. (2021, July 26). How to Improve Customer Service in the Airline Industry. Acquire. <https://acquire.io/blog/improve-customer-service-in-airline-industry/>
2. Amanda, R. & Negara, E.S. (2020). “Analysis and Implementation Machine Learning for YouTube Data Classification by Comparing the Performance of Classification Algorithms.” Jurnal Online Informatika, 5(1), 61-72.
3. Annie, A., Mohan, V., & Venu, S. (2015). Sentiment Analysis Applied to Airline Feedback to Boost Customer’s Endearment. GlobalIlluminators, 2, 219-232.
4. Deori, M., Kumar, V., & Verma, M. K. (2021). Analysis of YouTube video contents on Koha and DSpace, and sentiment analysis of viewers’ comments. Library Hi Tech. <https://doi.org/10.1108/LHT-12-2020-0323>
5. Dutta Das, D., Sharma, S., Natani, S., Khare, N., & Singh, B. (2017). Sentimental Analysis for Airline Twitter data. IOP Conference Series: Materials Science and Engineering, 263, 042067. <https://doi.org/10.1088/1757-899x/263/4/042067>
6. Huq, M. R., Ali, A. & Rahman, A. (2017). Sentiment Analysis on Twitter Data using KNN and SVM.
7. Martinez-Camara, E., Gutierrez-Vasquez, U., Fernandez, J.,Montejo-Raez, A., & Munoz-Guillena, R. (2015). “Ensemble classifier for twitter sentiment analysis.”

REFERENCES

8. Muhammad, A. N., Bukhori, S., & Pandunata, P. (2019). “Sentiment Analysis of Positive and Negative of YouTube Comments Using Naive Bayes-Support Vector Machine (NBSVM) Classifier”. 2019 International Conference on Computer Science, Information Technology and Electrical Engineering (ICOMITEE).
9. Nejjar, M., & Meziane, A. (2020, November 24). SAHAR-LSTM: An enhanced Model for Sentiment Analysis of Hotels'Arabic Reviews based on LSTM. Proceedings of 2020 5th International Conference on Cloud Computing and Artificial Intelligence: Technologies and Applications, CloudTech 2020. <https://doi.org/10.1109/CloudTech49835.2020.9365921>
10. Ramadhani A. M., Hong S.G, Universitas Gadjah Mada. Fakultas Teknik, & Institute of Electrical and Electronics Engineers. (2017). Twitter Sentiment Analysis using Deep Learning Methods.
11. Tiwari, P., Yadav, P., Kumar, S., Mishra, B. K., Nguyen, G. N., Gochhayat, S. P., Singh, J., & Prasad, M. (2019). Sentiment Analysis for Airlines Services Based on Twitter Dataset. *Social Network Analytics*, 149–162. <https://doi.org/10.1016/b978-0-12-815458-8.00008-6>
12. Vimali, J. S., & Murugan, S. (2021). A Text Based Sentiment Analysis Model using Bi-directional LSTM Networks. Proceedings of the 6th International Conference on Communication and Electronics Systems, ICCES 2021, 1652–1658. <https://doi.org/10.1109/ICCES51350.2021.9489129>



Thank you!

