GOETHE UNIVERSITY FRANKFURT AM MAIN

Public Perception of Industry 4.0 in Malaysia: A Sentiment analysis approach

Term paper

Lecturer: Prof. Dr. Arndt Graf

Course: Media, politics and economy in Southeast Asia: from Knowledge Society to Industry 4.0, Master course, Summer Term 2022

Goethe University Frankfurt am Main

Submitted by:

Tanmay Bagwe

Address: 03.23, Hausener Weg 120

60489, Frankfurt am Main, Germany

+49-174-2569506

Field of study: Modern East Asian Studies, 2nd Semester

Matriculation Number: 7774647

Email: s3386769@stud.uni-frankfurt.de

Submission date: 28/08/2022

Contents

Figures and tables

List of Abbreviations

ASEAN	Association of Southeast Asian Nations
NLP	Natural language processing
IoT	Internet of Things
SME	Small and medium-sized enterprises
MSME	Micro, Small & Medium Enterprises
MITI	Ministry of International trade and industry
ML	Machine Learning
AI	Artificial Intelligence
RoBERTa	Robustly Optimised BERT Pretraining Approach
CSV	Comma Separated Values

1 Introduction

The world has been going through a massive technological shift in recent decades. It is no wonder that the advent of computer and information communication technologies and, more recently, the use of information communication technology in different fields has redefined how the world works and how industries operate worldwide. The impact of information communication technology and the digitalisation of various industries, processes, and institutions has created much attention. As raised by Marr(2018) in Forbes Magazine, “We’re in the midst of a significant transformation regarding the way we produce products thanks to the digitisation of manufacturing. This transition is so compelling that it is being called Industry 4.0 to represent the fourth revolution that has occurred in manufacturing.” Industry 4.0 has justifiably become a trendy buzzword all over the internet. Its concept, however, has become a focus for many countries trying to leverage the next step of industrialisation through digitisation. Countries like Germany (Federal Ministry for Economic Affairs and Climate Action, 2022), India (Government of India, n.d.), and others have also been pushing the initiative in recent years. The recent pandemic only accelerated the rate of Industry 4.0 adoption and transition in Indian Industries (Khan, 2021). The implementation of Industry 4.0 has also been prominent in Southeast Asian countries. ASEAN (Association of Southeast Asian Nations) has been proactively taking the initiative in pushing the new drive of digitalised industrialisation among the member states (Association of Southeast Asian Nations (ASEAN), 2021).

The push for digital industrialisation has been one of the crucial agendas int the members of the ASEAN. The members are hopeful of the prospective economic growth Industry 4.0 will bring to their economies and the region (Association of Southeast Asian Nations (ASEAN), 2021). Observing a report by Arbulu et al. (2018, p. 13), the majority of the manufacturers and suppliers in the ASEAN member states are hopeful that Industry 4.0 will change their business models and also bring a boost to their business and help them increase their efficiency in doing business.

Among the ASEAN member states, Malaysia is one of the more digitally connected states, with around 89.6% of the population connected to the internet (DataReportal, 2022). Malaysia's growth of internet users from 2012 to January 2022 has been very optimistic, with approximately 13 million people becoming active internet users since 2012 (DataReportal, 2022).

Malaysia has also been proactively working on implementing industry 4.0 with its official announcement by its prime minister Dr Mahathir Mohamad (The ASEAN Post Team, 2018). Malaysia's government, with its Industry 4.0, also known as Industry4WRD, is aimed at promoting the country's manufacturing sector and looks to implement elements of digital technologies which can help the manufacturing sector contribute more to the economy (The ASEAN Post Team, 2018). Observing the news article by The ASEAN Post (2018), the country's government looks optimistic and hopeful that digital technologies will bring about good positive changes to the economy. But what about the general population? How does the General population perceive the introduction of Industry 4.0 in the nation?

This paper attempts to answer the question raised above. In order to understand a general perception of the population, the paper will try to conduct sentiment analysis. By mining social media posts (in this context, Twitter posts) and running them through a natural language processing model (NLP), the paper tries to understand the reaction of, in this context, the internet population's attitude towards Malaysia's Industry 4.0 initiative and discuss possible implications and the results of the sentiments of the Malaysian online population.

2 Industry 4.0

'Industry 4.0' signifies a relatively significant technological shift in the structure of how industries operate and function. In order to understand the context of Industry 4.0, it is necessary to understand the previous industrial innovations and revolutions and how they connect to the latest attempt at a new industrial revolution.

According to IBM's (IBM, n.d.) description of the industrial revolutions, there have been three industrial revolutions in the past, namely, the first industrial revolution,

which focused on enabling the mass production of goods; the second industrial revolution included the introduction of assembly lines and introduction of telephone and telegraph; the third industrial revolution which included the addition of computers and telecommunications to manufacturing processes. Finally, the fourth industrial revolution, called Industry 4.0, is characterised by automation and employment of smart machines and smart factories coupled with learned data, which can help in more efficient production (IBM, n.d.). These processes can then, in theory, help supplement the economy.

Industry 4.0 includes relatively new technologies such as the Internet of Things (IoT), where machines and devices and connected through web-enabled services (IBM, n.d.). It also includes Cloud computing. Velasquez et al. (2018) raised in their paper that, "Cloud Computing allows the storage of large amounts of data. This capacity is mainly important to store the data generated during a whole production process, considering that the machines and sensors produce more data than a person and such data is always connected. […] Through a scalable structure, the cloud allows the consumption of resources on request. This enables to reduce costs […] " (Velasquez et al., 2018, p. 260). Another vital technology included in Industry 4.0 is AI and machine learning, which helps manufacturing companies process and analyse large volumes of data, create insights in providing visibility and predict and automate processes (IBM, n.d.)

Running through the definitions above, Industry 4.0 aims to connect different manufacturing industry processes and create a system that is not only thoroughly interconnected but also capable of processing large amounts of data through different modern technological advancements and automation.

2.1 What is Industry 4.0

The federal republic of Germany first coined the term 'Industrie 4.0.' (in German), which later translated to 'Industry 4.0' in English. According to an article by Adrian Dima (2021), "Industry 4.0 originated in 2011 from a project in the high-tech strategy of the German government, which promotes the computerisation of manufacturing. Actually, the term "Industry 4.0" was publicly introduced in the same year at the

Hannover Fair." It has now spread well across the globe, and almost every developing and developed country is implementing aspects of 'Industry 4.0'; some of those countries include Japan, Germany, China, the United States, and the United Kingdom (Kagermann et al., 2016).

Most developed countries raised above to look at the upcoming Industry 4.0 as an opportunity to stimulate or maintain the status quo of their respective industries. Germany, for example, envisions maintaining its sophisticated position in manufacturing and mechanical engineering through the employment of digital technologies (Kagermann et al., 2016, p. 9). On the other hand, China looks to promote the digitisation of its SMEs and aims to promote automation technology and increase its production. Moreover, China looks at Industry 4.0 as an opportunity to get closer to other nations across the globe (Kagermann et al., 2016). Countries like Japan have already employed many aspects of 'Industry 4.0' in their infrastructure. However, just like Germany, they also look to employ digital technologies in their advanced manufacturing industries to maintain their status quo (Kagermann et al., 2016, p. 11).

Industry 4.0 in the context of Southeast Asia also has been following similar patterns. According to the Consolidated Strategy on the Fourth Industrial Revolution for ASEAN (2021), ASEAN, in general, looks to employ digital technologies to focus on the following areas: Technological Governance and Cybersecurity, which focuses on using digital infrastructure to create good e-governance models, data governance and have a robust cybersecurity system; Digital economy, which focuses on moving financial activities and transactions on the digital platform to further facilitate digital trading economy and help service sectors of the economy, facilitate smart agriculture, and small-medium size enterprises; Digital Transformation of the society which focuses on the using digital technology to promote human resource development, inclusive digital environment, social welfare and environmental sustainability (Association of Southeast Asian Nations (ASEAN), 2021, p. 7).

Countries like Singapore, Thailand, Indonesia, and Malaysia have also launched their respective initiative surrounding industry 4.0. The following section looks into Asia, specifically Southeast Asia, looking into implementing Industry 4.0.

2.2Industry 4.0 in Asia

Industry 4.0 is not only a phenomenon restricted only to developed nations. Many developing countries around the world are also trying to employ digital technologies in their industries; these include countries like south Asian countries like India (Government of India, n.d.), South American countries like Brazil, which are looking for prospects of Industry 4.0 (Master Certificações, 2022); and east Asian countries like South Korea (Kagermann et al., 2016, p. 11).

The paper's topic focuses explicitly on southeast Asia, and as raised in the previous section, Southeast Asian countries have also been working on implementing Industry 4.0. Countries and their manufacturers in the ASEAN look particularly optimistic and primed for the implementation of Industry 4.0 (Arbulu et al., 2018). ASEAN looks to implement using the latest digital technologies to "Maximising the potential of digital trade" (Association of Southeast Asian Nations (ASEAN), 2021, p. 24), "Embracing digital technologies for Industry 4.0" which aims at integrating digital technologies into the manufacturing industries or MSMEs and help them increase their value creation of their respective industries (Association of Southeast Asian Nations (ASEAN), 2021, p. 26), "Enhancing Services Sector competitiveness in the new economy" which puts priority focus on digitalisation of six service sectors in the ASEAN countries as they contribute more than half of the region's GDP (Association of Southeast Asian Nations (ASEAN), 2021, p. 27), "Health services technology" and "Financial technology (fintech)" (Association of Southeast Asian Nations (ASEAN), 2021, p. 28) and much more. Other than on an organisational level, southeast countries also have their own dedicated plans and policies for implementing Industry 4.0.

Singapore, for example, released its Industry 4.0 policy to level up its value chain further and help with the high-tech industries (Singapore Economic Development Board, 2022). On the other hand, Indonesia has introduced its own plan for Industry

4.0 called ‘Making Indonesia 4.0’ which aims to use new digital technologies and make the country capable of higher export potential in the upcoming decade (Indonesia Investment Coordinating Board, n.d.). Thailand, with its Thailand 4.0, aims to use the new technological wave to ramp up their industries with digitalisation, as raised by PR Newswire, “[Creating a]digital economy as a prime mover of the economy” (PR Newswire, 2021). Thailand looks to implement technologies of e-commerce, 5G, E-payment infrastructure and artificial intelligence (PR Newswire, 2021). As mentioned in the previous sections, Malaysia also introduced its version of industry 4.0 called ‘Industry4WRD’ (The ASEAN Post Team, 2018).

To summarise this section, most developed and developing countries look forward to implementing aspects of Industry 4.0 in their country and have already made plans. Some have even started implementing it as well. In particular, the countries in Southeast Asia, particularly the countries in ASEAN, are poised to carry out the transition to Industry 4.0 and bring in the advent of the digital economy. The following section focuses on the Industry 4.0 aspects of Malaysia and looks into the history and the potential policies implemented by Malaysia to make Industry4WRD work.

3 Industry 4.0 in Malaysia

Malaysia announced its Industry 4.0 initiative in 2018 (The ASEAN Post Team, 2018). Malaysia, among other countries in Southeast Asia, has a very high internet penetration and digital infrastructure (DataReportal, 2022) and thus could be a great candidate for Industry 4.0 and digitalisation in general. Furthermore, looking at the Official Website of the Ministry of International trade and industry (henceforth MITI) (2022), Malaysia specifically looks to enhance the manufacturing sector in the next ten years, with a supplementary vision of becoming a Strategic partner for smart manufacturing in the Asia Pacific, establishing itself as a destination for high-tech industry and providing solutions for advanced technology.

The Malaysian government aims to achieve its vision of Industry 4.0 by incorporating national objectives such as increasing labour productivity, having innovation capacity in the industry, promoting the manufacturing industry to contribute to the economy,

and having innovative and highly skilled jobs in the country (MINISTRY OF INTERNATIONAL TRADE AND INDUSTRY, 2022).

In the words of the then Prime Minister of Malaysia, Dr Mahathir Bin Mohamad, in the National Policy on Industry 4.0 (2018), "A strong manufacturing sector would pave the way to enhanced productivity, job creation, innovation capacity, high-skilled talent pool and ultimately economic prosperity and societal well-being [in Malaysia]. This would position Malaysia as a primary destination for smart manufacturing globally and attract more high-tech investments." (MINISTRY OF INTERNATIONAL TRADE AND INDUSTRY (MITI), Malaysia, 2018, p. 5).

By focusing majorly on manufacturing, Industry 4.0 in Malaysia also aims to fulfil goals nine and 12 of sustainable development set up by the United Nations (MINISTRY OF INTERNATIONAL TRADE AND INDUSTRY (MITI), Malaysia, 2018, p. 13). In addition, Malaysia's Industry4WRD aims to leverage the new enabling digital technologies such as Big data analytics, Artificial Intelligence, the Internet of Things (IoT), Cloud computing and more to stimulate industrial productivity in the region.

Going through the National Policy on Industry 4.0 by the Malaysian government (2018) and the ASEAN's policy on Industry 4.0 (2021), it is more evident that the Government of Malaysia and its global counterparts are more than enthusiastic about the whole initiative.

4 Public perception and Research Question

As raised in above sections, the government favours the industry 4.0 policies and initiatives; however, that does not entirely convey the sentiment of the overall population in the region (in this case, Malaysia). Understanding the overall public sentiment is substantial because Industry 4.0 is a significant overhaul of the industry and thus should be of significance to the public. Research about Public opinion states that the higher the significance of a policy which can affect the public directly, the more the public opinion matters to that individual (Burstein, 2003; Page & Shapiro, 1983). Thus it is crucial to know how the public reacts to the concept of Industry 4.0

in the context of Malaysia. Based on the literature raised above, the paper asks the following research question:

R1: How is the public sentiment towards Industry 4.0 in Malaysia?

The paper will try to answer the research question through the method of Industry 4.0 and use

5 Method

Sentiment Analysis in Python[1] Programming language is one way of answering the research question of this paper. As stated on Tech Target (2021), "Sentiment analysis, also referred to as opinion mining, is an approach to natural language processing (NLP) that identifies the emotional tone behind a body of text. This is a popular way for organisations to determine and categorise opinions about a product, service, or idea. It involves the use of data mining, machine learning (ML) and artificial intelligence (AI) to mine text for sentiment and subjective information."

Sentiment analysis is used in several fields ranging from product analysis, marketing research and social media monitoring; and helps the respective institution to understand the social sentiment of their brand or product while being able to analyse through a wide range of online conversations (Gupta, 2018; Rotulo, 2022). As stated by Rotulo(2022), sentiment analysis can also be stated as "[…]sentiment analysis (also known as opinion mining) is a natural language processing (NLP) algorithm to identify, extract, and quantify the emotional tone behind a body of text."

5.1 Data Mining and Gathering for 'Industry 4.0 in Malaysia'

In the case of Malaysia, internet penetration in the region is more than 80% (DataReportal, 2022), and thus doing a Sentiment Analysis Online would help us understand the public sentiment. For the scope of the paper and the ease of gathering data, the paper tries to gather data from Twitter. On this social media website, all the posts and articles are a short passages of approximately 280 characters per post, known as *tweet/s* (Gil, 2021). This limit of short paragraphs in a post helps identify

[1] Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. More info on:

and gather large amounts of individual posts, which can be turned into a dataset and efficiently run through the NLP Algorithm to get the sentiment scores for the dataset in question.

The steps for the sentiment analysis are as follows (refer to Table 1.0)

Table 1 Steps for Sentiment analysis on Industry 4.0 and Malaysia

Step no.	Action to be taken
1	Data Mine Tweets have 'Malaysia' and 'Industry 4.0' in them.
2	Compile all the tweets and convert them into a dataset file (.csv)
3	Use the dataset file and run it through an NLP model
4	Create Sentiment Probability scores for the whole dataset into (Positive, Neutral, and Negative)

In order to get raw data for the analysis, the paper will mine tweets from 2015 to 2022, i.e., five years before and after the announcement of Industry 4.0 in Malaysia, using a Python Code (Refer to Section 7 for the Code). The code for this paper is a modified version of the Tweet Scraping tutorial provided by AI Spectrum(2022). The code for mining, collecting and then converting them into a dataset named *tweets.py* is as follows:

```
import snscrape.modules.twitter as sntwitter
import pandas as pd

query = "Malaysia Industry 4.0 until:2022-06-01 since:2015-01-01"
tweets = []
limit = 5000

for tweet in sntwitter.TwitterSearchScraper(query).get_items():

    #print(vars(tweet))
    #break

    if len(tweets) == limit:
        break
    else:
```

```
        tweets.append([tweet.date, tweet.user.username, tweet.content])

df = pd.DataFrame(tweets, columns=['Date', 'User', 'Tweet'])

df.to_csv('Malaysia Tweets.csv')
```

Running the Python code '*tweets.py'* gives us a list of tweets that state 'Malaysia' and 'Industry 4.0' inside them. The code does this by going through the Twitter database and mining all the posts from January 1, 2015, to July 1, 2022. The paper takes it from 2015 to 2022, assuming to get enough tweets before and after the announcement of industry 4.0 in Malaysia. Next, the code compiles all the tweets into a dataset file named '*Malaysia Tweets.csv.'* The paper then uses the dataset file *Malaysia Tweets.csv* into a natural language processing (NLP) algorithm, which will further process the tweets and create a sentiment probability score. The score then helps the paper figure out the sentiment of the tweets, which helps in understanding the general sentiment of the people towards or against Industry 4.0 in Malaysia.

5.2 Sentiment Analysis of the dataset

In simpler terms, to process or create a Sentiment Analysis, the paper uses the obtained Dataset file Malaysia Tweets.csv, processes it and displays sentiment probability scores for the whole dataset.

The paper uses a natural language processing AI model called 'RoBERTa' model for sentiment analysis (Robustly Optimised BERT Pretraining Approach) (Barbieri et al., 2020; Liu et al., 2019). The machine learning model is trained with approximately 58 million tweets and made specifically for sentiment analysis in English (Hugging Face, n.d.). This model enables us to analyse and implement the sentiment of even complex texts, including news texts where opinions are expressed less overtly (Rotulo, 2022). For the sentiment analysis of this paper, the paper tries to see whether the emotion or the sentiment of the database is positive, negative, or neutral.

For the dataset created in the previous section, *Malaysia Tweets.csv*¸, the RoBERTa Model will go through each tweet available in the dataset. The python code will then

create a score for every tweet present in the dataset and award them a sentiment score differentiated and labelled as positive, negative and neutral.

The paper then collects and compiles all the results from the dataset, compiles a scoresheet of all the sentiment probability scores, and creates a CSV file, which the paper could manually analyse in MS Excel.

In order to process the dataset and make it run through the RoBERTA Model, the code-named '*Roberta_Sentiment FINAL.py'* is as follows:

```
from cgitb import text
import csv
import pandas as pd
import nltk
import numpy as np
from transformers import AutoTokenizer
from transformers import AutoModelForSequenceClassification
from scipy.special import softmax
from tqdm.notebook import tqdm
from nltk.sentiment import SentimentIntensityAnalyzer


df = pd.read_csv("Malaysia Tweets.csv")
example = df['Tweet'][50]

MODEL = f"cardiffnlp/twitter-roberta-base-sentiment"
tokenizer = AutoTokenizer.from_pretrained(MODEL)
model = AutoModelForSequenceClassification.from_pretrained(MODEL)

#Running example on ROBERTA
encoded_text = tokenizer(example, return_tensors='pt')
output = model(**encoded_text)
scores = output[0][0].detach().numpy()
scores = softmax(scores)
scores_dict = {
    'Roberta_Neg' : scores[0],
    'Roberta_Neu' : scores[1],
    'Roberta_Pos' : scores[2]
}

def polarity_scores_roberta(example):
    encoded_text = tokenizer(example, return_tensors='pt')
```

```
    output = model(**encoded_text)
    scores = output[0][0].detach().numpy()
    scores = softmax(scores)
    scores_dict = {
        'roberta_neg' : scores[0],
        'roberta_neu' : scores[1],
        'roberta_pos' : scores[2]
    }
    return scores_dict

#Whole Data set on ROBERTA
res = {}
for i, row in tqdm(df.iterrows(), total=len(df)):
    text = row['Tweet']
    myid = row['User']
    #roberta_result = polarity_scores_roberta(text)
    res[myid] = polarity_scores_roberta(text)

print(pd.DataFrame(res).T)
pd.DataFrame(res).T.to_csv('Malaysia Sentiment ROBERTA.csv')
```

The code uses a sample reference for the RoBERTA model from Medallion Data Science's Code (2022). However, it modifies and builds upon the reference to accommodate the dataset '*Malaysia Tweets.csv*' and also adds additional functionality to create a dataset file at the end.

The python code results give the paper a final dataset file known as *'Malaysia Sentiment ROBERTA.csv'*. This file will then summarise the approximate sentiment probability score towards Malaysia Industry 4.0 on Twitter.

The codes are edited and run inside Microsoft Visual studio code V1.69.2 and used Microsoft Excel 2019 for Basic mathematic calculations

6 Results and Observation and discussion

Running the code *tweets.py*[2] mines and gathers all the tweets with messages with 'Malaysia' And 'Industry 4.0' in them. The result of running the code leads to an

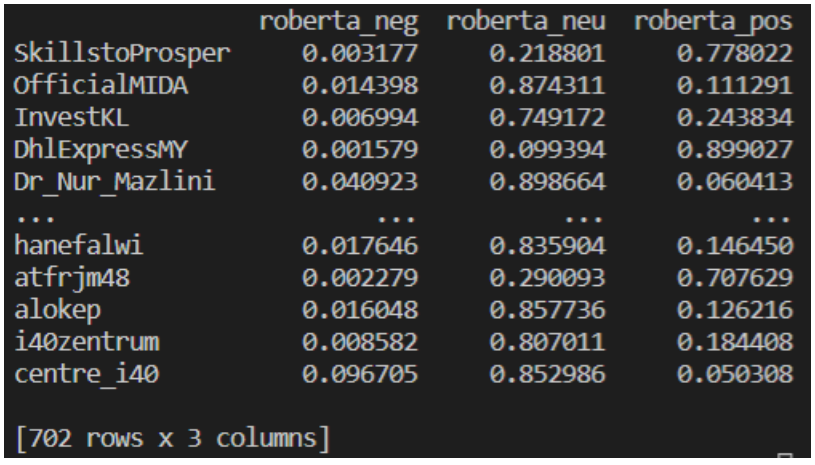
[2] The code for tweets.py is also available for reference on GitHub at: https://github.com/happygoluckycodeeditor/senti-analysis-malaysia

output file named *Malaysia Tweets.csv*[3] , which fetched 1226 tweet messages off the Twitter search query. It is important to note that it includes tweets in a span of eight years ranging from 2015 to 2022.

The paper then ran the dataset file *Malaysia Tweets.csv* through the Sentiment analysis code '*Roberta_Sentiment FINAL.py*' of RoBERTa to fetch a Dataset file named *'Malaysia Sentiment ROBERTA.csv'*, which is filled with the sentiment analysis scores for all the tweets available in the *Malaysia Tweets.csv* dataset file. The result of the analysis leads to an outcome of a total of 703 results. It is note that while the initial raw tweets file, *Malaysia Tweets.csv* contained 1227 messages, the final processed *Malaysia Sentiment ROBERTA.csv´* contained only 702 process results. Because multiple texts in the dataset were duplicated by the users, i.e., the same content published on different dates, the program decides for the duplicates to be not considered during the analysis.

The direct output inside the Microsoft Virtual Studio Code turns out to be as follows:

Figure 1 Output file for Final Sentiment analysis on Malaysia Industry 4.0 in Microsoft Virtual Studio Code



	roberta_neg	roberta_neu	roberta_pos
SkillstoProsper	0.003177	0.218801	0.778022
OfficialMIDA	0.014398	0.874311	0.111291
InvestKL	0.006994	0.749172	0.243834
DhlExpressMY	0.001579	0.099394	0.899027
Dr_Nur_Mazlini	0.040923	0.898664	0.060413
...	...	...	...
hanefalwi	0.017646	0.835904	0.146450
atfrjm48	0.002279	0.290093	0.707629
alokep	0.016048	0.857736	0.126216
i40zentrum	0.008582	0.807011	0.184408
centre_i40	0.096705	0.852986	0.050308

[702 rows x 3 columns]

Referring to figure 1, column 1 consists of the username of the person who typed the tweet, column two with '*roberta_neg*' indicates the probability of the tweet being negative, and column three with '*roberta_neu*' indicates the probability of the tweet

[3] The whole dataset file (.csv) is available for reference on Github at: https://github.com/happygoluckycodeeditor/senti-analysis-malaysia (the tweets are also included in the dataset file)

to be neutral. Finally, column four '*roberta_pos*' indicates the probability of the tweet is positive. The code '*Roberta_Sentiment FINAL.py*' runs probability tests for all the 702 tweets and processes probability scores for all the tweets included in the dataset. The paper collected the scores and summed and averaged them, and calculated an overall score as follows (refer to Table 2):

Figure 2 Overall Sentiment Score for all the tweets collected on Malaysia Industry 4.0

	'roberta_neg' Indicates the negative sentiment probability of the tweets	*'roberta_neu'* Indicates the neutral sentiment probability of the tweets	*'roberta_pos'* Indicates the positive sentiment of the tweets
The overall sentiment of all the tweets	**0.074194** (7.42%)	**0.607835** (60.78%)	**0.317971** (31.79%)
Total number of tweets: 702			

Summarising the results obtained from the analysis and referring to Table 2 for the results, the sentiment analysis raises that, concerning Malaysia Industry 4.0, 7.42% of all the content tweets have negative sentiment, 60.78% of all the content tweets have neutral sentiment, and 31.79% of the content of the tweet have a positive sentiment.

7 Conclusion

The results conclude that most of the tweets and their contents towards Malaysia Industry 4.0 have been neutral to positive. The negative probability of all the tweets is as low as 7.42%, which indicates that most Twitter accounts do not think of this Industry 4.0 in Malaysia as a flawed initiative. The total probability score of Neutral-Positive sentiment is about 92.57%, which is a very optimistic score for the initiative in Malaysia.

The paper's analysis infers that most of the public's sentiment towards the initiative of Industry 4.0 in Malaysia is likely Neutral-positive on the social media platform Twitter. The result infers that on social media, at least twitter in general, users from around the world look at Malaysia's Industry 4.0 attempt with a relatively good optimism and the public opinion might not necessarily create any hindrance in the enactment of the policy.

8 Limitations

The sentiment analysis in this paper employed Machine Learning models and Data Mining to understand how people on the internet perceive their sentiment towards Industry 4.0 in Malaysia. However, it is not a general perception of all the people in the population, let alone all the people on the internet. Moreover, since the analysis in the paper only uses one social media platform, Twitter, as the primary source to collect data, it does not necessarily create a generalisation of the results. Therefore, for a more comprehensive sentiment toward the query, it is recommended for future research to include models which can accommodate more social media platforms. The paper also recommends including non-reachable platforms such as blogging websites to collect a more comprehensive outlook or sentiment on Malaysia Industry 4.0 in future research.

Another limitation faced by the analysis is the consideration and inclusion of languages in the analysis. RoBERTa model is excellent because of its exhaustive training of 58 million tweets; however, it seems that it is designed for English language Sentiment analysis (Hugging Face, n.d.). This feature brings a barrier to figuring out the sentiment of Tweets which are not made in English. Further research could be done with the dataset but with a model which can support multiple local languages from Malaysia and English.

It is also necessary to note that the AI model used in the analysis is not perfect. It has an average accuracy of 70% (Jose Camacho-Collados, 2020/2022). This number implies that there is a 30% chance for the probability score to be not accurate. Further research with better learning models is recommended to estimate the sentiment in the tweets better.

The dataset used for the analyses does not have any region restriction, i.e., the tweets and data collected for this analysis consist of tweets from all over the world with respect to Malaysia Industry 4.0. Thus, if there is to be a need to judge Malaysian public perception precisely, it would be recommended to use the Places function in the Twitter advanced search options (Twitter, Inc., n.d.).

To summarise, this paper aims to introduce how Sentiment analysis can be used to find more profound insights into people's perceptions of Industry 4.0 in Southeast Asia. Further research can inculcate upon the above-raised limitations and bring better and deeper public opinion insights towards industry 4.0 in the region.

9 Bibliography

AI Spectrum (Director). (2022, April 5). *Get UNLIMITED Tweets by Python Without Twitter API*. https://www.youtube.com/watch?v=jtIMnmbnOFo

Arbulu, I., Lath, V., Mancini, M., Patel, A., & Tonby, O. (2018). *Industry 4.0: Reinvigorating ASEAN Manufacturing for the Future*. McKinsey&Company. https://www.mckinsey.com/~/media/mckinsey/business%20functions/operations/our%20insights/industry%204%200%20reinvigorating%20asean%20manufacturing%20for%20the%20future/industry-4-0-reinvigorating-asean-manufacturing-for-the-future.ashx

Association of Southeast Asian Nations (ASEAN). (2021). *Consolidated Strategy on the Fourth Industrial Revolution for ASEAN*. ASEAN: A Community of Opportunities for All. https://asean.org/wp-content/uploads/2021/10/6.-Consolidated-Strategy-on-the-4IR-for-ASEAN.pdf

Barbieri, F., Camacho-Collados, J., Neves, L., & Espinosa-Anke, L. (2020). *TweetEval: Unified Benchmark and Comparative Evaluation for Tweet Classification* (arXiv:2010.12421). arXiv. https://doi.org/10.48550/arXiv.2010.12421

Burstein, P. (2003). The Impact of Public Opinion on Public Policy: A Review and an Agenda. *Political Research Quarterly*, *56*(1), 29–40. https://doi.org/10.2307/3219881

DataReportal. (2022, February 15). *Digital 2022: Malaysia*. DataReportal – Global Digital Insights. https://datareportal.com/reports/digital-2022-malaysia

Dima, A. (2021, January 25). Short history of manufacturing: From Industry 1.0 to Industry 4.0. *KFactory*. https://kfactory.eu/short-history-of-manufacturing-from-industry-1-0-to-industry-4-0/

Federal Ministry for Economic Affairs and Climate Action. (2022). *What is Industrie 4.0?* [Government]. Federal Ministry for Economic Affairs and Climate Action - INDUSTRIE 4.0. https://www.plattform-i40.de/IP/Navigation/EN/Industrie40/WhatIsIndustrie40/what-is-industrie40.html

Gil, P. (2021, August 29). *What Is Twitter & How Does It Work?* Lifewire. https://www.lifewire.com/what-exactly-is-twitter-2483331

Government of India. (n.d.). *About I4.0 | SAMARTH Udyog* [Government]. SAMARTH Udyog. https://www.samarthudyog-i40.in/about-i40

Gupta, S. (2018, January 19). *Sentiment Analysis: Concept, Analysis and Applications*. Medium. https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17

Hugging Face. (n.d.). *Cardiffnlp/twitter-roberta-base-sentiment · Hugging Face*. https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment

IBM. (n.d.). *What is Industry 4.0 and how does it work? | IBM* [Technology]. IBM. https://www.ibm.com/topics/industry-4-0

Indonesia Investment Coordinating Board. (n.d.). *Making Indonesia 4.0: Indonesia's Strategy to Enter the 4th Generation of Industry Revolution* [Government]. BKPM. https://www.investindonesia.go.id/en/why-invest/indonesia-economic-update/making-indonesia-4.0-indonesias-strategy-to-enter-the-4th-generation-of-ind

Jose Camacho-Collados. (2022). *TweetEval* [Jupyter Notebook]. Cardiff NLP. https://github.com/cardiffnlp/tweeteval/blob/4fbd22cd78421f05b1ecdb4fc5725bc7a7bd8f66/README.md (Original work published 2020)

Kagermann, H., Anderl, R., Gausemeier, J., Schuh, G., & Wahlster, W. (2016). Industrie 4.0 in a Global Context: Strategies for Cooperating with International Partners (acatech STUDY). *Munich: Herbert Utz Verlag*, 74.

Khan, S. (2021, August 17). Covid just gave Industry 4.0 the push it needed in India. *The Economic Times*. https://economictimes.indiatimes.com/news/company/corporate-trends/covid-just-gave-industry-4-0-the-push-it-needed-in-india/articleshow/85392907.cms

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). *RoBERTa: A Robustly Optimized BERT Pretraining Approach* (arXiv:1907.11692). arXiv. https://doi.org/10.48550/arXiv.1907.11692

Marr, B. (2018, September 2). *What is Industry 4.0? Here's A Super Easy Explanation For Anyone*. Forbes. https://www.forbes.com/sites/bernardmarr/2018/09/02/what-is-industry-4-0-heres-a-super-easy-explanation-for-anyone/

Master Certificações. (2022, January 7). Industry 4.0 in Brazil: What it is, Technologies and Benefits. *Master Certifications*. https://master.org.br/en/news/industry-4-0-brazil/

Medallion Data Science (Director). (2022, May 5). *Python Sentiment Analysis Project with NLTK and 🤗 Transformers. Classify Amazon Reviews!!* https://www.youtube.com/watch?v=QpzMWQvxXWk

MINISTRY OF INTERNATIONAL TRADE AND INDUSTRY. (2022, August 12). *Industry4WRD* [Government]. MINISTRY OF INTERNATIONAL TRADE AND INDUSTRY. https://www.miti.gov.my/index.php/pages/view/4832

MINISTRY OF INTERNATIONAL TRADE AND INDUSTRY (MITI), Malaysia. (2018). *Industry4WRD: National Policy on Industry 4.0*. MINISTRY OF INTERNATIONAL TRADE AND INDUSTRY. https://www.miti.gov.my/miti/resources/National%20Policy%20on%20Industry%204.0/Industry4WRD_Final.pdf

Page, B. I., & Shapiro, R. Y. (1983). Effects of Public Opinion on Policy. *American Political Science Review*, *77*(1), 175–190. https://doi.org/10.2307/1956018

PR Newswire. (2021, July 30). The Next Digital Behemoth: Thailand's 4.0 Revolution. *Bloomberg.Com*. https://www.bloomberg.com/press-releases/2021-07-30/the-next-digital-behemoth-thailand-s-4-0-revolution

Rotulo, M. (2022, April 10). 😊 😳 Tweets sentiment analysis with RoBERTa. *MLearning.Ai*. https://medium.com/mlearning-ai/tweets-sentiment-analysis-with-roberta-1f30cf4e1035

Singapore Economic Development Board. (2022, August 9). *Advanced Manufacturing Singapore: Key Industry & Activities | Singapore EDB*. Singapore Economic Development Board. https://www.edb.gov.sg/en/our-industries/advanced-manufacturing.html

Tech Target. (2021, October). *What is sentiment analysis (opinion mining)? Definition from SearchBusinessAnalytics*. SearchBusinessAnalytics. https://www.techtarget.com/searchbusinessanalytics/definition/opinion-mining-sentiment-mining

The ASEAN Post Team. (2018, November 1). *Malaysia launches Industry 4.0 policy* [News]. The ASEAN Post. https://theaseanpost.com/article/malaysia-launches-industry-40-policy

Twitter, Inc. (n.d.). *How to use advanced search – find Tweets, hashtags, and more* [Social Media Website]. https://help.twitter.com/en/using-twitter/twitter-advanced-search

Velasquez, N., Estevez, E., & Pesado, P. (2018). Cloud Computing, Big Data and the Industry 4.0 Reference Architectures. *Journal of Computer Science and Technology*, *18*, 258–266. https://doi.org/10.24215/16666038.18.e29

Statutory Declaration

“I herewith declare that I have composed the present the term paper myself and without the use of any other than the cited sources and aids. Sentences or parts of sentences quoted literally are marked as such. Other references are indicated regarding statement and scope and are specified by full details of the publications concerned. The term paper in the same or similar form has not been submitted to any examination body and has not been published. This thesis was not yet, even in part, used in another examination or as a course performance. Moreover, I give my consent to have the present thesis checked by plagiarism detection software.”

Place: Frankfurt am Main Date:28/8/2022

Signature:……

Ehrenwörtliche Erklärung

“Ich versichere hiermit, dass ich die vorliegende Arbeit selbständig und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel verfasst habe. Wörtlich übernommene Sätze oder Satzteile sind als Zitat belegt, andere Anlehnungen, hinsichtlich Aussage und Umfang, unter Quellenangabe kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen und ist nicht veröffentlicht. Sie wurde nicht, auch nicht auszugsweise, für eine andere Prüfungs- oder Studienleistung verwendet. Darüber hinaus stimme ich der Überprüfung der vorliegenden Arbeit durch eine Plagiatssoftware zu.“

Ort, Datum: Frankfurt am Main, 28/8/2022

Unterschrift:……