Perak's Tourism through the Lens of Social Media: A Computer-based Sentiment Analysis Approach

Raudatul Jannah Rostam
Azilawati Azizan*
Nurkhairizan Khairudin
College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Perak Branch, Tapah Campus, Perak Malaysia
azila899@uitm.edu.my*

Proposed citation:

Abstract

Perak, a state in Malaysia, has a lot of exciting and captivating places for tourists to visit. There are many different natural wonders, cultural landmarks, historical sites, and delicious foods to enjoy. This makes Perak special and can bring in a lot of tourists. In today's digital age, social media plays a big role in sharing information, including tourism. With the presence of technology that can discover feelings and emotions from social media texts, such as sentiment analysis, we can make this even better. Sentiment analysis is the process of analyzing and identifying the feelings conveyed in a text such as positivity or negativity by utilizing natural language processing (NLP) and machine learning approach. This project aims to discover the sentiments of tourist attraction in Perak by analyzing Twitter data. The project has three objectives: first to collect and prepare a suitable and reliable dataset, then classify the data into positive, negative, or neutral sentiments using NLP techniques, and finally develop a web-based application to visualize those sentiments. To accomplish these objectives, a collection of tweets pertaining to Perak's tourist attractions has been gathered and prepared for analysis. TextBlob library in the Python programming language is used to extract sentiment of tweets from Twitter data and classify them into positive, negative, or neutral categories. Then a machine learning approach, Support Vector Machine (SVM) is used to create, train and test the sentiment model. And as a result, utilizing the SVM classifier with a linear kernel and a split of 70:30 between training and testing data yields an increased accuracy rate of 75.50%. This project is important because it provides valuable insight to the tourism sector in Perak.

Keywords:

Perak, sentiment analysis, tourism attractions, Twitter data

1   Introduction

Perak, a pleasant state located in the center of Malaysia, attracts visitors with a variety of thrilling and alluring locations just waiting to be discovered. Within its boundaries, a patchwork of natural splendors, cultural hubs, historical gems, and mouthwatering culinary delights weave together to create an exceptional attraction (Sulaiman, Jaini, Jamaluddin, & Hashim, 2018). The sceneries include breath-taking natural vistas and architectural wonders that reflect a long and colorful past (Isa et al., 2022). Perak stands itself as a genuinely outstanding location due to its intricate combination of amenities, which attracts tourists from all over the world.

Furthermore, the tourism industry is crucial for economic growth, and it's important for the government and authorities to understand tourists' preferences. Perak's economy has played an important role for the development of Malaysia, becoming the 7th largest state in terms of its economic size and contributing 5.5% of the country's GDP in 2020 (Zulkeffeli, Zafri,. Morhalim A.M, 2023). Therefore, the growth of Perak's tourism industry holds importance, and it is crucial to make efforts to bring more tourists to Perak. This effort needs to be coordinated from several main parties in this sector such as the government and tourism authorities of the state, the tourism board, accommodation service providers, food, infrastructure, transport, local businesses, media & marketing, and the local community.

Meanwhile, from a tourist's perspective, searching for truly captivating destinations can also be challenging nowadays due to the overwhelming amount of information (Abbasi-Moud, Vahdat-Nejad, & Sadri, 2021) and fancy reviews on the Internet that may not always match the actual experience (Goodwin, Joseff, & Woolley, 2020). This may lead to difficulties in making decisions and getting bad experiences during trips. However, comments on social media platforms like Facebook, Twitter, and Instagram can be perceived as more genuine than reviews crafted by websites or social media influencers (Kim & Kim, 2022). This is because the comments are often spontaneous and come directly from users who may have experienced a product, service or destination without any specific agenda (Jaitly & Gautam, 2021). Twitter, for instance, has been used as a platform to give feedback, opinions, and reviews by posting tweets. A tweet may contain details or updates on the tourist sites they will be visiting or have already been to, including their experiences during the tour, their impressions of certain destinations, and other tourist attractions (Hermanto, Ziaurrahman, Bianto, & Setyanto, 2018).

Therefore, this project chose to analyze Twitter data to discover sentiments about tourist attractions in Perak. To achieve this, the comments undergo preprocessing, semantically grouped, and sentimentally assessed (Abbasi-Moud et al., 2021). The outcome of the sentiment analysis is then visualized online on a dashboard of a web application.

This project has multiple advantages to both tourists and tourism management bodies. The analysis of this project provides a convenient and quick way to gauge the overall sentiment towards Perak's attractions through the analysis of Twitter data. Moreover, the user-friendly online dashboard allows easy access to sentiment insights

and enhances tourists' engagement with the attractions. On the side of tourism management, the project offers a valuable tool for understanding tourist perception. This empowers them to customize strategies and services to address visitors' preferences and concerns properly. The sentiment analysis outcomes aid in identifying popular attractions, areas that require improvement, guiding decision-making and resource allocation. In the end, this project helps to improve the tourism services and encourages tourists to have a positive impression of Perak's attractions.

2   Literature Review

2.1   Tourism Attraction in Perak

Tourism attraction in Perak includes all the attractions that attract tourists to come and visit Perak during their holiday. There are many types of tourism attractions such as food tourism attraction, nature attraction, and purpose-built attraction.

a) Food Tourism Attraction

Food tourism is essential to destination marketing because tourists constantly make plans for where to go, what to do, and where to locate the tastiest meals and beverages. Food tourism serves as a means of luring both domestic and international visitors. Iconic food also influences the tourists to visit a certain place. Iconic food is unique typical dishes that represent and identify a destination and motivate tourists to visit it (Rousta & Jamshidi, 2020). In addition to enhancing already-existing tourism assets and fostering innovation, food tourism may significantly contribute to the preservation of regional heritage.

For example, Tambun Pomelo is one of the food attractions in Perak. It is located at Tambun, Ipoh and the main season is during Chinese New Year. The citrus fruit is seasonal and there are two varieties of pomelos which are red (bitter or sour) and white (bitter/sweet). It is famous especially in the Tambun area because water from the surrounding limestone hills gives the soil plenty of nutrients that pomelos prefer. Frequent rain also makes the pomelos juicy. Tourists can have a plantation tour and buy pomelos in Tambun. Figure 1 and Figure 2 shows the entrance of Tambun Pomelo Street and the pomelos there.

Figure 1: Tambun Pomelo Street

Figure 2: Pomelos from Tambun

b) Nature Attractions

A natural attraction is an attraction that has been created by nature. Many of these areas have been given a status to protect their environment and provide facilities so that people who visit are able to enjoy the sites. Nature tourism attractions include national parks, waterfalls, caves, mountains, islands and forests. In Perak, there are several nature attractions like Lata Kinjang Waterfall, Gua Tempurung and Pulau Pangkor. A well-known tourist destination in Perak, Pulau Pangkor is one of nature attractions currently developing into a top tourism centre (Sulaiman et al., 2018). These are the nature attractions that are often visited by the visitors when they are in Perak. Figure 3 shows the scenery of Pulau Pangkor (Pangkor Island).

Figure 3: Pulau Pangkor

c) Purpose-built Attractions

Purpose built attractions are attractions that have been built purposely to attract tourists into that area (Liu, 2023). Tourist attractions make people automatically think about purpose built attractions as they are fun, enjoyable and designed for many different reasons. This type of attraction includes zoos, water parks, museums and art galleries. Zoo Taiping, Sunway Lost World of Tambun and Galeri Sultan Azlan Shah are examples of purpose-built attractions located in Perak. Figure 4 shows Zoo Taiping, one of the purpose-built attractions in Perak.

Figure 4: Zoo Taiping

2.2 Sentiment Analysis

Sentiment analysis or also known as opinion mining is a natural language processing (NLP) technique that allocates positive or negative polarity to an entity or items by using different natural language processing tools and predicted high and low performance of various sentiment classifiers (Al Amrani, Lazaar, & El Kadirp, 2018). Sentiment analysis is a technique for computationally assessing text to find opinions and assessments (Jain et al., 2023). The most common text categorization tool determines if an incoming message is good, negative, or neutral by analysing the underlying sentiment. Sentiment analysis analyses people's utterances, including opinions, feelings, evaluations, attitudes, emotions, and appraisals of products, services, organisations, individuals, subjects, events, and their attributes (Flores-Ruiz, Elizondo-Salto, & Barroso-González, 2021). Sentiment analysis methods emphasise sentiments and emotions, urgency, and even intentions in addition to polarity (Sudhir & Suresh, 2021). Sentiment analysis, which extracts and analyses public mood and opinions, is becoming more and more popular in the research community, in academia, in government, and in the service industries (Cui, Wang, Ho, & Cambria, 2023).

Today, people commonly express their mood and feelings on social networking sites like Facebook, Twitter, and Instagram. Twitter has become a valuable platform, providing an ever-growing collection of information on a wide range of topics from worldwide breaking news to opinions on products and services (Rodríguez-Ibánez, Casánez-Ventura, Castejón-Mateos, & Cuenca-Jiménez, 2023). Sentiment analysis has been widely used to help businesses, services and many more fields in improving based on customer's feedback. Numerous company processes, including brand monitoring, product analytics, customer service, and market research, might benefit from sentiment analysis.

3   Methodology

This project utilised the Machine Learning Life Cycle (MLLC) methodology. It is a cyclical approach for developing a successful machine learning project (Rob Ashmore, Radu Calinescu, 2021). Figure 5 depicts the MLLC, which consists of seven major phases.
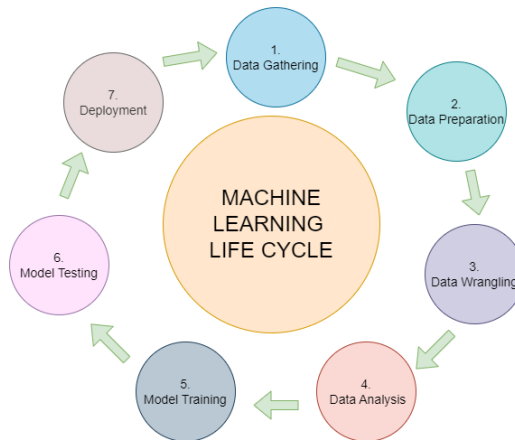
Figure 5: Machine Learning Life Cycle

The process starts with data gathering, particularly from Twitter. Following this, the data preparation phase takes place, involving the cleaning and organisation of the collected data. Subsequently, data wrangling ensures appropriate formatting and structuring of the data for further analysis. Then, during the data analysis stage, the data is classified and closely examined to uncover trends and insights of people's sentiments about Perak's attractions. Later is the model training, in which the computer is taught to understand and predict sentiments based on the prepared data. After that, the model testing, where accuracy of the sentiment prediction is evaluated to ensure dependability (Garcia, Sreekanti, Yadwadkar, & others, 2018). Finally, the outcomes of the tweet analysis are showcased on a dashboard within a web application.

3.1 Data Gathering

In this phase, the data was collected or scraped from Twitter using snscrape library (Santosa, 2022) which was available in Python. The Google Colab platform facilitated this process. The dataset of this project focuses on tourism attractions within Perak. Since Twitter is linked with hashtags or keywords, the data scraping process used keywords and hashtags like '#lostworldoftambun', '#zootaiping', #telukbatik, #latakinjang, #perak, Pulau Pangkor and any related tweets that are necessary. The tweets data was scraped starting from January 2019 until 31st January 2023, yielding a total of 11562 raw tweets. All this data was subsequently saved into an Excel comma-separated values (CSV) format file. The dataset comprises several variables such as date, username and tweets which are important for this project. Figure 6 shows a sample of the collected tweets data.

2022-12-	alan_the	Awat #ipoh trending ni woiii .kome nak serang ipoh ke minggu ni ðŸ¤£ðŸ¤£
2022-12-	walydna0	Our Moon - Color and B&W #Moon #HeritageCity #ipoh #perak #Malaysia #photography #moonphotography https://t.co/7j7yqMC38u
2022-12-	callmeyin	#ipoh jangan nak trending sangat. Kita postpone cuti tau nak balik kampungðŸ¥²
2022-12-	MALAYSI	Taman Rekreasi Gunung Lang, Ipoh, Perak, Malaysia A. Sanusi Husain 14.4.2022 #taman #rekreasi #tasik #airterjun #Ipoh #Perak #Malaysia #Malaysiaindah #Malaysiacantik #Malaysiaboleh #MasyaAllah #SubhanAllah https://t.co/nUOwQgR6to
2022-12-	Meg_Sup	Sedap!! #ipoh https://t.co/JDFqBn3NGd
2022-12-	AhmadS:	Taman Rekreasi Gunung Lang, Ipoh, Perak, Malaysia #taman #rekreasi #tasik #airterjun #Ipoh #Perak #Malaysia #Malaysiaindah #Malaysiacantik #Malaysiaboleh https://t.co/bXEJ9DOYvA
2022-12-	bagirunc	Kenapa ipoh trending wei ? Kalau nk lari pusing pekan ipoh roger. @MohKelpoh #ipoh https://t.co/79YeTltR5e
2022-12-	DrAmarN	Long-tailed Shrike (Lanius schach) seen today â€" closeup image shows the tomial tooth in the beak #BirdsSeenIn2022 #Ipoh #Perak #Malaysia @Avibase @orientbirdclub @IndiAves https://t.co/eG0720pIID
2022-12-	DrAmarN	Zebra (Peaceful) Dove (Geopelia striata) Birds that are considered â€˜commonâ€™™ can be uncommonly beautiful when seen at close range. #BirdsSeenIn2022 #Ipoh #Perak #Malaysia @Avibase @orientbirdclub @IndiAves https://t.co/fiJqZsVpZw
2022-12-	MaslizaM	More trailing daisy/Wedilia (Sphagneticola trilobata) Spotted by the river at Lubuk Timah #Flowers #flowerphotography #NaturePhotography #amateurphotographer #wildflower #itsthelittlethingsinlife #ipoh #Perak https://t.co/8bG05Rb35m

Figure 6: Sample of Raw Tweets

3.2 Data Preparation

During data preparation, the collected dataset has undergone data pre-processing which includes data cleaning, filtering and transformation. During the cleaning process, all the duplicate tweets, retweets, mentions, hashtags and non-related tweets were eliminated from the dataset. Next was filtering the data in which all hyperlinks, emoticons and punctuations were removed from the tweets. Then was the data transformation process; since the collected tweets data contains both Malay and English content, and some with a mix of both languages (Kong et al., 2023), all the tweets were translated to English using Google Cloud Translation API. It is to ensure data consistency. This action was taken because of two concerns. Firstly, is to ensure that all tweet data can be effectively processed and analysed by NLP algorithms and machine learning techniques. Secondly, is to prevent the loss of essential information during text analysis, which could otherwise result in incomplete, inaccurate, or unreliable outcome (Abu Bakar, Idris, Shuib, & Khamis, 2020). These were all performed by creating an automatic function by importing related libraries in the Python programming language. Figure 7 and 8 show the example of data before and after the transformation process being conducted. At the end of this data preparation phase, a cleaned dataset of 501 tweets was successfully produced.

97

2022-12-30	theoldbikeguy	@denkmit Try Pangkor Island. About 2 hours north of KL. Used to be good.
2022-12-29	timlim12	Ipoh is always about family, friends and food
2022-12-29	azmilsyahmi	Victoria Bridge Kuala Kangsar, Perak - Shot on Insta360 Go #kualakangsar #perak #malaysia
2022-12-26	zairulazham	Back to Nature !!~ #KualaKangsar #Perak
2022-12-25	MohKelpoh	meriah juga padang ipoh waktu malam

Figure 7: Sample Data before Transformation Process

12/30/22	theoldbikeguy	denkmit try pangkor island about 2 hours north kl used be good
12/29/22	timlim12	ipoh is always about family friends food
12/29/22	azmilsyahmi	victoria bridge kuala kangsar perak shot insta360 go kualakangsar perak malaysia
12/26/22	zairulazham	back nature kualakangsar perak
12/25/22	MohKelpoh	merry also ipoh field at night

Figure 8: Sample Data after Transformation Process

3.3 Data Wrangling

Data wrangling is needed to ensure data quality and ready for analysis. The tweets data is saved into CSV format and missing data were also being monitored in this phase. Each tweet was carefully inspected to ensure its validity, reliability and completeness before proceeding to the data analysis phase. This has fulfilled the first objective which is to prepare a suitable and reliable dataset that is pertinent to the project's scope.

3.4 Data Analysis

During this phase, each tweet was processed to classify the sentiment of the text using Textblob library (Lorla, 2020). This library provides an intuitive API (Application Programming Interface) for addressing common NLP tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more (Azmi, Abidin, Mutalib, Zawawi, & Halim, 2022). An API is a way for two or more computer programs to communicate with each other. Figure 9 shows the codes to get polarity of the sentiment using TextBlob library while Figure 10 shows the sentiment of tweets (positive, neutral, negative) that were extracted from TextBlob library.

```
#Create a function to get polarity
def getPolarity(text):
    return TextBlob(text).sentiment.polarity

#Create Polarity column
df['Polarity'] = df['Tweets'].apply(getPolarity)
df
```

Figure 9: Codes to Get Polarity using TextBlob

31-Jan-19	callmeyinyin	concubine lane it s fun come here with boyfriend	0.3	1	Positive
10-Mar-20	athirahsuperb	concubine lane it s hot but if you want get picture bride it s best go when there are no people it s	0.355952	1	Positive
6-Feb-19	na2ngkusuma	concubine lane one places that must be visited ipoh this place is alley china town area that is line	0	0	Neutral
2-Jul-22	noorZaff	concubine lane so crowded so hot street is narrowly compared street art penang i go 1400 its crazy	-0.075	-1	Negative
4-Dec-22	adliana_iknc	concubine lane swears not up my expectation its very narrow but what i like about that place is fc	-0.26	-1	Negative
22-Sep-20	SutthiesRam	concubinelane ipoh gave me that petaling street feel chilllahdei	0	0	Neutral
19-Nov-21	Apkhaw	concubinelane smell coffee is so good	0.7	1	Positive

Figure 10: Sample Tweets with Sentiment Polarity

By utilising this library the second objective is achieved; categorising the data into positive, negative, or neutral sentiments using NLP techniques. The data was then analysed and visualised on the web application dashboard through various forms of charts including clustered bar chart, clustered column chart, pie chart and word cloud. Figure 11 shows the visualised analysis of tweets data through these charts.
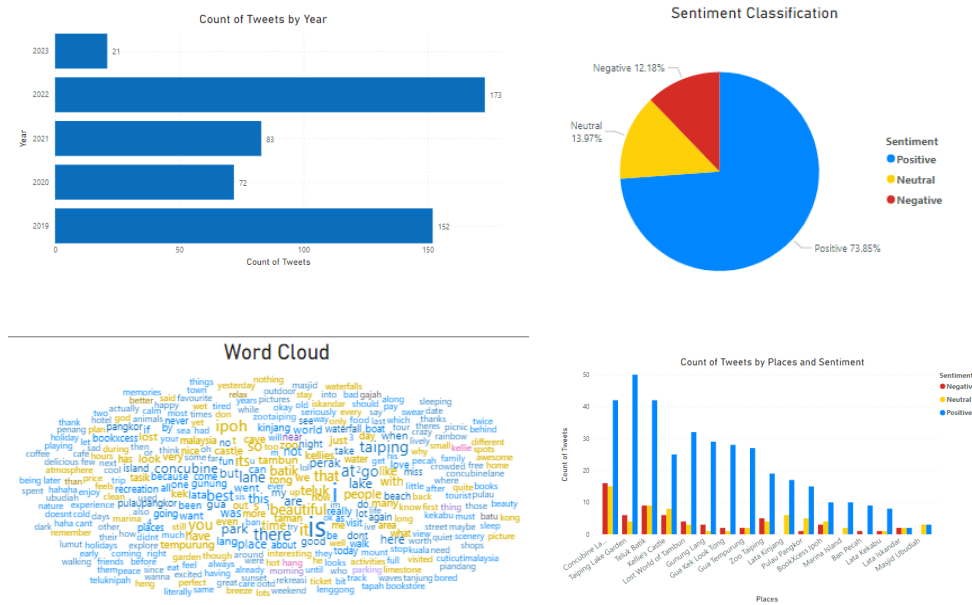

Figure 11: Tweets Analysis

3.5 Model Training

This project employed SVM (Support Vector Machine) classifier to create the sentiment model based on the labelled data produced by TextBlob (Praveen Gujjar & Prasanna Kumar, 2021). SVM is a supervised learning algorithm used to solve classification and regression problems. Performance of the model is evaluated by measuring the classification accuracy variations. The dataset was split into training and testing sets using ratios of 90:10, 80:20, and 70:30 respectively. The results are measured using the confusion matrix, accuracy, recall, sensitivity, precision, and specificity. Figure 12 shows a piece of code that is used for splitting data for training and testing with 42 sets as random states. The random state option sets up the internal random number generator, which determines how the data is divided into train and test

99

indices. Setting a random state to a constant value guarantees that the same sequence of random numbers will be created each time the code is run. Typically, people use values like 0, 1, or even 42.

```
# Splitting the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Figure 12: Codes for Splitting Data into Training and Testing

3.6   Model Testing

Model testing is an important stage in determining an SVM model's performance. It determines how well the trained model generalises to new, previously unknown data. It compares the predicted labels from the trained model to the test dataset's ground truth labels. Then the model's performance is evaluated using several evaluation measures such as accuracy; the percentage of properly categorised samples among all samples, precision; the model's ability to identify positive samples, recall; the fraction of positive samples detected correctly by the model, F1 score; a balanced measure of model performance based on the harmonic mean of accuracy and recall, and lastly confusion matrix; a table that displays the total number of true positive, true negative, false positive and false negative predictions (Nurdeni, Budi, & Santoso, 2021).

3.7   Deployment

Before deploying the application to users, the final phase involved creating the web application dashboard. The web application was successfully developed by utilising the combination of HTML (Hypertext Markup Language), CSS (Cascading Style Sheets), and Power BI tool. The web pages were built using HTML, which enabled us to organise and arrange the material. CSS is used because it is crucial in improving the web application's visual appeal since it allows altering the layout, colours, fonts and other design components to provide an aesthetically pleasant user interface. Power BI tool is used to generate charts for the data visualisations of this project (Louis T. Becker, 2019). The dashboard for the web application is shown in Figure 9.

Figure 9: Designated dashboard

This has successfully fulfilled the third objective which is to develop a web application for visualising tourist feedback analysis on attractions in Perak. Figure 10 shows the user interface of the web application.

Figure 10: User Interface of the System

4 Findings and Discussions

4.1 Results of Tweets Analysis

Based on the cleaned dataset produced in the earlier phase, Figure 11 shows a total of 501 tweets related to Perak's tourism attractions from 2019 to 2023. It was found that, the year 2022 witnessed the highest counts of tweets linked to Perak's tourism attractions, reaching 173 tweets, while the second highest was recorded in 2019. In contrast, the year 2020 recorded the lowest tweet count, with only 72 tweets, followed by 83 tweets in 2021. This trend can be linked to the reasons of the Covid-19 pandemic and the subsequent implementation of the Movement Control Order (MCO) executed by the government from 2020 to 2021. Meanwhile, in 2022 the Covid19 pandemic has started to subside and all sectors in the country have re-opened and started operating as usual including the tourism sector. This can be clearly seen when tweets related to tourism increased dramatically in 2022. In fact, the increase in tweets in that year was greater than the growth in tweets in 2019, where the Covid19 pandemic did not exist. This analysis of tweets is visualised interactively by clustered bar graph using Power BI tool. Figure 11 shows the graph trends of count of tweets by year starting from 2019 to 2023.
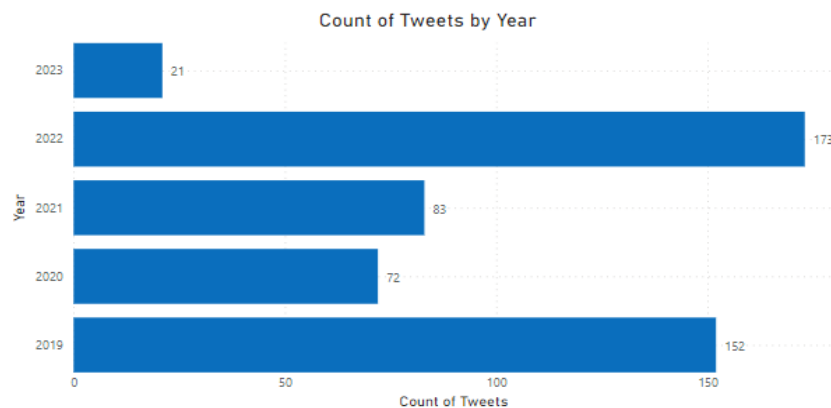

Figure 11: Clustered bar graph for Tweets from 2019 to 2023

The polarity of the data was produced by classifying the tweets into positive, neutral or negative sentiment. This project discovered that a big portion of tweets (73.8%) were positive, which means many people shared positive thoughts about Perak's tourist attractions. This may indicate that most users were giving compliments or recommendations about these attractions.

About 13.97% of tweets were neutral, which is not leaning towards positive or negative. These tweets might include facts or general information. By analysing these neutral tweets, we could find objective information about the attractions without any clear bias.

There were also some tweets (12.18%) that were negative. These represent a group of people who didn't have a positive opinion about the tourist sites. These tweets could point out issues or concerns about safety or bad experiences. To make things better for tourists in Perak, it is important to look at these negative tweets to understand the specific problems and fix them appropriately. A pie chart of the sentiment classification is shown in Figure 12.
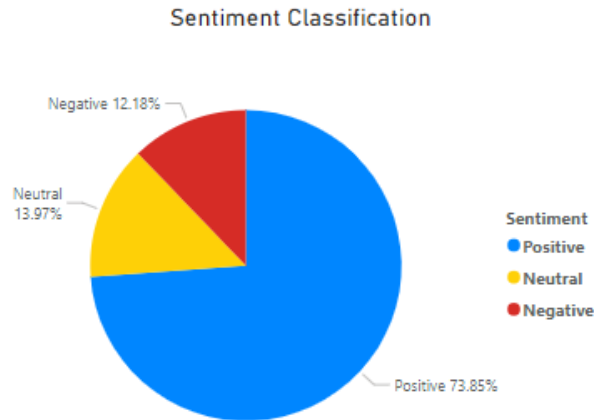

Figure 12: Sentiment Classification of Tweets

Based on the word cloud in Figure 13, it highlights the most frequent discussed or tweeted words. In this word cloud, Taiping, Concubine Lane and Ipoh have been discussed and tweeted by the users a lot seeing that the size of those words is bigger than other words.
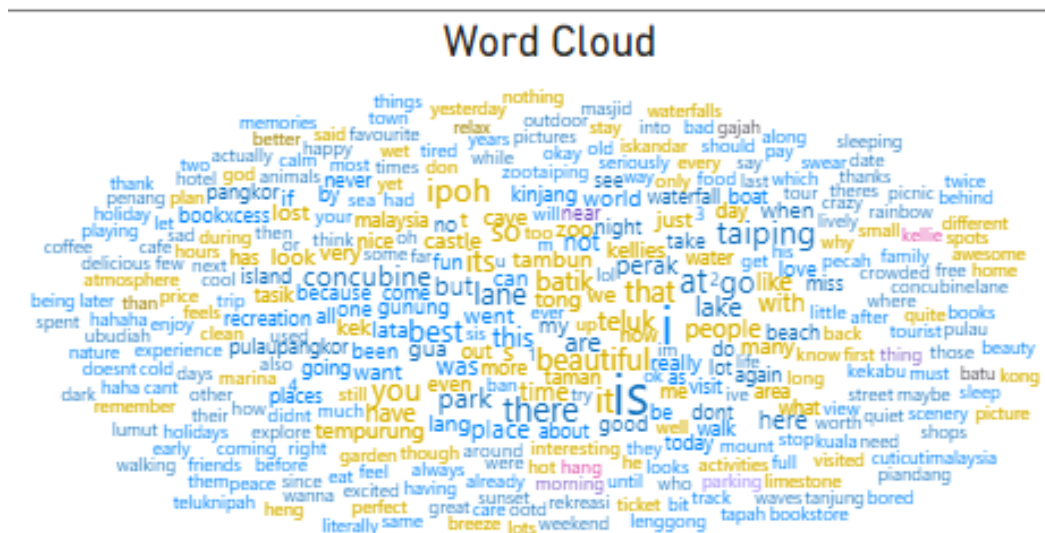

Figure 13: Word cloud

103

Meanwhile, a clustered column chart about count of tweets by sentiment and places were shown in Figure 14. The places that have the greatest number of positive tweets are Taiping Lake Garden with 50 positive tweets while Concubine Lane got the highest number of negative tweets with a total of 16 tweets. Five most tweeted places are Concubine Lane Ipoh, Taiping Lake Garden, Teluk Batik, Kellie's Castle and Lost World of Tambun, Masjid Ubudiah, Kuala Kangsar got the very least tweets from the Twitter users.
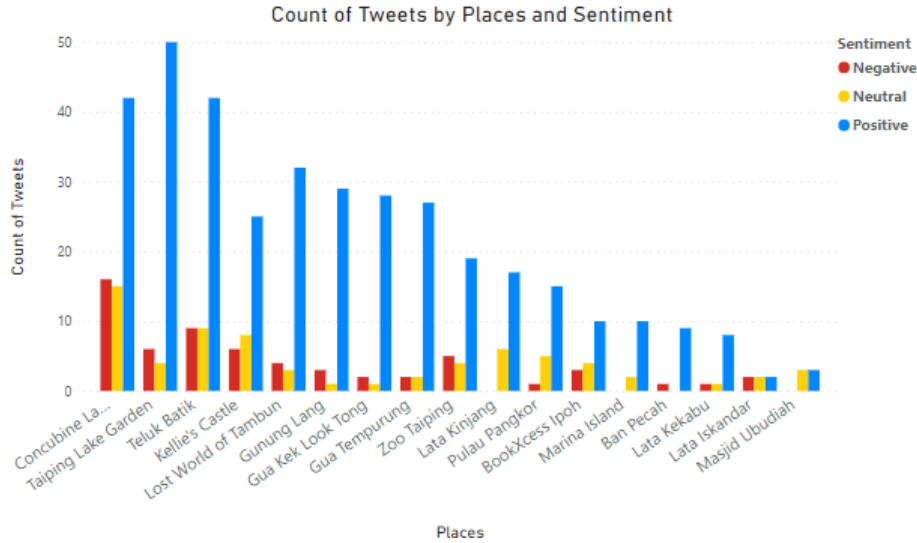

Figure 14: Clustered column chart

4.2   Results of Tweets Analysis

Table 1 lists the accuracy score for each kernel in the SVM model with 70:30 split training and testing. It shows that the accuracy score is 75.50% and it is the highest accuracy score in this project. In reality, an accuracy rate between 70% and 90% is not only acceptable, but also feasible (Brindha, Senthilkumar, Singh, & Sharma, 2022).

Table 1: Accuracy Score for 70:30 Split Training and Testing

Kernel	Accuracy Score (%)
Linear	75.50
Polynomial	75.49
RBF	75.49
Sigmoid	74.83

The results presented in Table 1 highlight the accuracy scores achieved through the utilisation of various kernels within the SVM model. Particularly, the highest accuracy score recorded is an impressive 75.50%, marking a significant achievement within the context of this project. This outcome signifies the effectiveness of the SVM model in accurately categorising sentiments based on the collected data. The attained accuracy

level serves as a promising indicator of the model's reliability and potential utility for sentiment analysis of tourism attractions in Perak.

The robustness of the acquired accuracy is further supported by the 70:30 split between training and testing data. By ensuring that the model's performance is assessed on many datasets, this partitioning technique lowers the possibility of overfitting and improves the model's generalisation skills. The model's capacity to understand the subtleties of sentiment expression in the gathered tweets is highlighted by the consistency of accuracy across the selected kernels.

Additionally, this result establishes the project's sentiment analysis approach as a useful tool for precisely and practically assessing tourist sentiments. The project's effectiveness in successfully classifying attitudes is attested to by the high accuracy score, which can help the Perak state government make wise decisions to improve tourist experiences and promote Perak's attractions.

Though accuracy is a key indicator, it's important to remember that a thorough review might also take other performance metrics like precision, recall, and F1-score into consideration. Regarding imbalanced sentiment classes or different effects of false positives and false negatives, this could offer a broader assessment of the model's performance. Overall, the project's success in creating a trustworthy sentiment analysis model, which opens the way for improved destination management strategies and improved visitor experiences, is highlighted by the achievement of a 75.50% accuracy score.

5 Conclusion

This project aims to identify the sentiment polarity associated with tourist attractions in Perak, primarily using comments from social media platforms, particularly Twitter. The approach involves employing computer-based sentiment analysis techniques that leverage NLP technology. In conclusion, the analysis has shown how people feel about Perak's tourist attraction, based on tweets data collected from January 2019 to January 2023. The rise in tweets in 2022 and the drop in 2020 reflect the impact of Covid-19 and the rules put in place during 2020-2021. The increase in tweets related to tourism in the state of Perak in 2022 indicates that the tourism sector in this state is beginning to regain momentum. The result of sentiments, using a tool called TextBlob, highlighted that most tweets (73.8%) were positive, showing that people are fond of Perak's attractions. Neutral tweets (13.97%) were kind of in the middle, giving facts without a strong feeling. Negative tweets (12.18%) were fewer, and they pointed out areas that need fixing. The charts and pictures help to understand this better. This information can be used by the Perak state government to further improve its tourism sector and will indirectly make tourists happier. The relationship between feelings and tourism is important, showing how data can guide us in making decisions, such as the insight presented in this project, which can help the tourism sector in Perak to be better.

6   References

Abbasi-Moud, Z., Vahdat-Nejad, H., & Sadri, J. (2021). Tourism recommendation system based on semantic clustering and sentiment analysis. *Expert Systems with Applications*, *167*(April). https://doi.org/10.1016/j.eswa.2020.114324

Abu Bakar, M. F. R., Idris, N., Shuib, L., & Khamis, N. (2020). Sentiment Analysis of Noisy Malay Text: State of Art, Challenges and Future Work. *IEEE Access*, *8*, 24687–24696. https://doi.org/10.1109/ACCESS.2020.2968955

Al Amrani, Y., Lazaar, M., & El Kadirp, K. E. (2018). Random forest and support vector machine based hybrid approach to sentiment analysis. *Procedia Computer Science*, *127*(September 2019), 511–520. https://doi.org/10.1016/j.procs.2018.01.150

Azmi, P. A. R., Abidin, A. W. Z., Mutalib, S., Zawawi, I. S. M., & Halim, S. A. (2022). Sentiment Analysis on MySejahtera Application during COVID-19 Pandemic. *2022 3rd International Conference on Artificial Intelligence and Data Sciences: Championing Innovations in Artificial Intelligence and Data Sciences for Sustainable Future, AiDAS 2022 - Proceedings*, 215–220. https://doi.org/10.1109/AiDAS56890.2022.9918748

Brindha, K., Senthilkumar, S., Singh, A. K., & Sharma, P. M. (2022). Sentiment Analysis with NLP on Twitter Data. *2022 International Conference on Smart Generation Computing, Communication and Networking, SMART GENCON 2022*, 1–4. https://doi.org/10.1109/SMARTGENCON56628.2022.10084036

Cui, J., Wang, Z., Ho, S. B., & Cambria, E. (2023). *Survey on sentiment analysis: evolution of research methods and topics*. *Artificial Intelligence Review* (Vol. 56). Springer Netherlands. https://doi.org/10.1007/s10462-022-10386-z

Flores-Ruiz, D., Elizondo-Salto, A., & Barroso-González, M. D. L. O. (2021). Using social media in tourist sentiment analysis: A case study of andalusia during the Covid-19 pandemic. *Sustainability (Switzerland)*, *13*(7), 1–19. https://doi.org/10.3390/su13073836

Garcia, R., Sreekanti, V., Yadwadkar, N., & others. (2018). Context: The missing piece in the machine learning lifecycle. *Acm Cmi*, 1–4. Retrieved from https://web.stanford.edu/~neerajay/Flor_CMI_18_CameraReady.pdf

Goodwin, A. M., Joseff, K., & Woolley, S. C. (2020). Social Media Influencers and the 2020 U.S. Election: Paying "Regular People" for Digital Campaign Communication. *Center for Media Engagement*. Retrieved from https://mediaengagement.org/research/social-media-influencers-and-the-2020-election

Hermanto, D. T., Ziaurrahman, M., Bianto, M. A., & Setyanto, A. (2018). Twitter Social Media Sentiment Analysis in Tourist Destinations Using Algorithms Naive Bayes Classifier. *Journal of Physics: Conference Series*, *1140*(1). https://doi.org/10.1088/1742-6596/1140/1/012037

Isa, M. I., Rahman, N. A., Abdullah, K., Ahmad, A. L., Mohamad, M. R., Rahmat, A., & Omar, F. N. (2022). Development of Rural Tourism in Perak Tengah District Based on Local Authority Perspectives. *Planning Malaysia*, *20*(4), 405–419. https://doi.org/10.21837/pm.v20i23.1175

Jain, T., Verma, V. K., Sharma, A. K., Saini, B., Purohit, N., Bhavika, … Arshad, M. S. (2023). Sentiment Analysis on COVID-19 Vaccine Tweets using Machine Learning and Deep Learning Algorithms. *International Journal of Advanced Computer Science and Applications*, *14*(5), 32–41. https://doi.org/10.14569/IJACSA.2023.0140504

Jaitly, R. C., & Gautam, O. (2021). Impact of social media influencers on customer engagement and brand perception. *International Journal of Internet Marketing and Advertising*, *15*(2), 220–242. https://doi.org/10.1504/IJIMA.2021.114336

Kim, J., & Kim, M. (2022). Rise of Social Media Influencers as a New Marketing Channel: Focusing on the Roles of Psychological Well-Being and Perceived Social Responsibility among Consumers. *International Journal of Environmental Research and Public Health*, *19*(4). https://doi.org/10.3390/ijerph19042362

Kong, J. T. H., Juwono, F. H., Ngu, I. Y., Nugraha, I. G. D., Maraden, Y., & Wong, W. K. (2023). A Mixed Malay–English Language COVID-19 Twitter Dataset: A Sentiment Analysis. *Big Data and Cognitive Computing*, *7*(2), 61. https://doi.org/10.3390/bdcc7020061

Liu, Z. (2023). Sustainable tourism development: A critique. *Journal of Sustainable Tourism*, *11*(6), 459–475. https://doi.org/10.1080/09669580308667216

Lorla, S. (2020). TextBlob Documentation Release 0.16.0. *TextBlob*. Retrieved from https://textblob.readthedocs.io/en/dev/

Louis T. Becker, E. M. G. (2019). 2019-Microsoft Power BI-Extending Excel to Manipulate, Analyze, and Visualize Diverse Data. *Computer Science Serials Review*, *45*, 184-188.

Nurdeni, D. A., Budi, I., & Santoso, A. B. (2021). Sentiment Analysis on Covid19 Vaccines in Indonesia: From the Perspective of Sinovac and Pfizer. *3rd 2021 East Indonesia Conference on Computer and Information Technology, EIConCIT 2021*, 122–127. https://doi.org/10.1109/EIConCIT50028.2021.9431852

Praveen Gujjar, A. P., & Prasanna Kumar, H. R. (2021). Sentiment Analysis:Textblob For Decision Making. *International Journal of Scientific Research & Engineering Trends*, *7*(2), 2395–2566. Retrieved from www.samra.

Rob Ashmore, Radu Calinescu, C. P. (2021). Assuring the Machine Learning Lifecycle: Desiderata, Methods, and Challenges. *ACMComput. Surv*, *0*.

Rodríguez-Ibánez, M., Casánez-Ventura, A., Castejón-Mateos, F., & Cuenca-Jiménez, P. M. (2023). A review on sentiment analysis from social media platforms. *Expert Systems with Applications*, *223*(March). https://doi.org/10.1016/j.eswa.2023.119862

Rousta, A., & Jamshidi, D. (2020). Food tourism value: Investigating the factors that influence tourists to revisit. *Journal of Vacation Marketing*, *26*(1), 73–95. https://doi.org/10.1177/1356766719858649

Santosa, I. M. A. (2022). Sentiment Analysis on the Centralized Isolation Policy for Covid-19 Response in Bali Province. *Compiler*, *11*(2), 45–50. https://doi.org/10.28989/compiler.v11i2.1356

Sudhir, P., & Suresh, V. D. (2021). Comparative study of various approaches, applications and classifiers for sentiment analysis. *Global Transitions Proceedings*, *2*(2), 205–211. https://doi.org/10.1016/j.gltp.2021.08.004

Sulaiman, F. C., Jaini, N., Jamaluddin, E. R., & Hashim, N. I. (2018). The Impact of Marine Tourism Towards the Local Community at Pulau Pangkor, Perak. *Asian Journal of Quality of Life*, *3*(13), 168–176. https://doi.org/10.21834/ajqol.v3i13.174

Zulkeffeli, Zafri,. Morhalim A.M, . (2023). Perak State Economy & Its Potentials, *2020*(June 2022), 1–13.