

# PORTAL PELANCONGAN DENGAN ANALISIS SENTIMEN

Aljannah binti Nordin

Mohd Ridzwan Yaakub

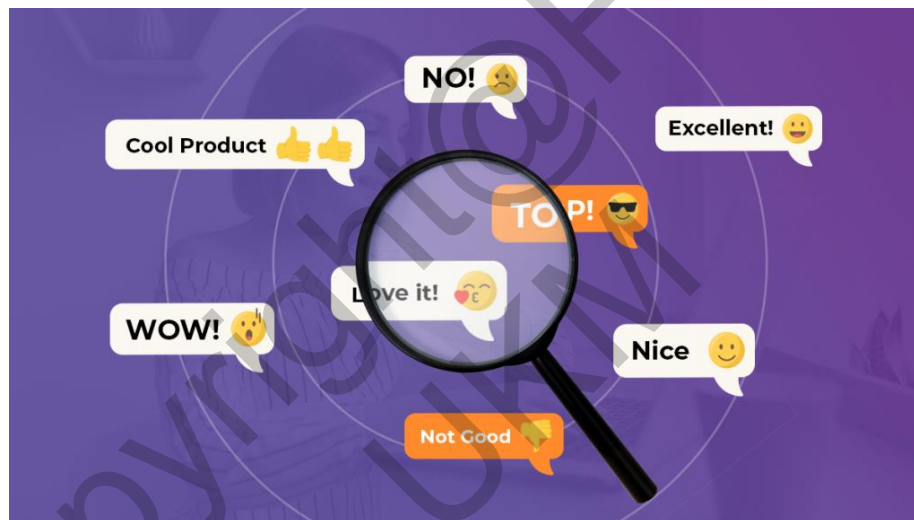
Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia

## ABSTRAK

Sektor pelancongan merupakan sektor yang sentiasa menjadi tumpuan masyarakat. Ini adalah kerana, Malaysia mempunyai pelbagai lokasi pelancongan yang menjadi tarikan pelancong dari dalam negara dan luar negara. Namun begitu, setiap pelancong pasti mempunyai pandangan yang tersendiri terhadap lokasi-lokasi tersebut. Media sosial sering menjadi platform bagi pelancong berkongsi tentang pengalaman mereka semasa bercuti. Namun begitu, pertambahan maklumat tentang pelancongan di media sosial menjadikannya tidak tersusun. Oleh itu, dalam projek ini, Portal Pelancongan Dengan Analisis Sentimen dibangunkan. Objektif portal ini dibangunkan adalah untuk memfokuskan nilai sentimen pelancong terhadap beberapa tempat. Selain itu, portal ini juga dapat memberi panduan kepada pelancong lain untuk melihat nilai sentimen terhadap destinasi pelancongan mereka. Metodologi yang digunakan untuk projek ini adalah *CRISP-DM* kerana metodologi ini sesuai digunakan dengan projek yang berdasarkan perlombongan data. Ini adalah kerana data bagi projek ini diekstrak daripada Twitter. Selain itu, data juga diambil daripada laman web pelancongan seperti *Trivago* atau *Tripadvisor*. Seterusnya, projek ini menggunakan Bahasa pengaturcaraan *Python* dan *HTML* serta beberapa model pengelasan iaitu *Naïve Bayes*, *Support Vector Machine* dan *Logistic Regression*. Hasil akhir bagi projek ini adalah sebuah portal yang mempunyai susun atur yang mudah difahami dan memaparkan nilai sentimen positif, neutral dan negatif sebagai rujukan pengguna.

## 1 PENGENALAN

Analisis sentimen yang juga dikenali sebagai perlombongan pendapat, adalah merupakan bidang dibawah Pemprosesan Bahasa Tabii (*NLP*) dimana analisis ini mencuba untuk mengenal pasti dan mengekstrak pendapat daripada teks yang diperolehi dari blogs media sosial, forum dan sebagainya (Ozgur Genc 2019). Selain itu, analisis sentimen juga boleh didefinisikan sebagai proses penggunaan *text analytics* untuk mendapatkan pelbagai sumber data dari internet dan platform media sosial (Tsalis Annisa 2020). Seterusnya, analisis sentimen juga merupakan perlombongan teks maklumat yang mengidentifikasikan dan mengekstrak informasi subjektif dan membantu para peniaga memahami sentimen pelanggan dari segi produk mahupun layanan mereka (Rahmat Burhanudin 2018).



Rajah 1 Contoh sentimen

Analisis sentimen sering digunakan bagi tujuan menyusun semua teks yang tidak berstruktur ke bentuk yang lebih tersusun menggunakan *NLP* dan *Open source tools*. Sebagai manusia, sudah tentu kita mudah mengetahui perasaan atau emosi seseorang tapi tidak untuk mesin. Namun begitu, kemajuan yang kita kecapai pada masa kini membenarkan mesin atau komputer untuk memahami sikap atau emosi orang ramai melalui penulisan. Bagi menjalankan proses analisis sentimen ini data diambil daripada media sosial dan juga laman web pelancongan (*travel website*). Hal ini adalah kerana, orang ramai sering menggunakan media sosial untuk berkongsi sesuatu. Segala maklumat akan tersebar dengan cepat di media sosial dengan bantuan kemajuan internet. Analisis dijalankan dengan mengkaji sentimen dalam kandungan media sosial seperti Twitter dan laman web seperti *Trivago* dan *Tripadvisor* mahupun *Google Reviews*.

## IMPORTANCE OF SOCIAL MEDIA SENTIMENT ANALYSIS

Quantzig



Rajah 2 Kepentingan analisis sentimen menggunakan media sosial

Dalam projek yang diberi nama Portal Pelancongan Dengan Analisis Sentimen ini, analisis sentimen digunakan bagi mendapat pandangan orang ramai tentang destinasi pelancongan yang menarik di Malaysia. Data bagi portal ini diambil daripada laman web pelancongan atau diekstrak daripada Twitter. Kemudian, peratusan sentimen positif dan sentimen negatif dipaparkan didalam portal untuk diakses oleh orang ramai dan membuat penilaian tentang destinasi pelancongan mereka. Tujuan portal ini dibina adalah untuk mengasingkan sentimen positif dan negatif supaya menjadi lebih teratur. Selain itu, portal ini juga dapat memudahkan masyarakat yang ingin melancong untuk mencari maklumat tentang destinasi pelancongan mereka tanpa perlu membaca komen yang tidak teratur di media sosial.

## 2 PERNYATAAN MASALAH

Berdasarkan pemerhatian yang dijalankan, laman sosial telah menjadi antara platform utama bagi orang ramai untuk berkongsi maklumat. Perkara ini menjadikan media sosial seperti Twitter atau Facebook sebagai sumber bagi pengumpulan data. Selain daripada media sosial, laman web pelancongan (*travel website*) juga menjadi sumber untuk mendapatkan data kerana kebanyakan pelancong akan meninggalkan komen pada laman web tempahan destinasi pelancongan seperti *Tripadvisor*. Antara maklumat yang sering dikongsi oleh pengguna ialah perasaan atau sentimen mereka terhadap beberapa destinasi pelancongan di Malaysia iaitu sentimen positif, neutral dan negatif. Namun begitu, terdapat beberapa masalah yang dikenal pasti. Antaranya, jumlah sentimen yang sering bertambah di media sosial atau laman web pelancongan menyebabkan percampuran antara sentimen positif dan negatif. Perkara ini

menyukarkan pembahagian jumlah nilai setiap sentimen untuk dipaparkan didalam portal yang dibangunkan. Selain itu, bukan semua orang yang mahir untuk menggunakan media sosial atau pelayar web untuk mendapat maklumat tentang destinasi pelancongan mereka. Seseengah masyarakat mungkin memerlukan platform yang berbeza bagi melihat pandangan orang lain tentang sesuatu tempat.

### **3 OBJEKTIF KAJIAN**

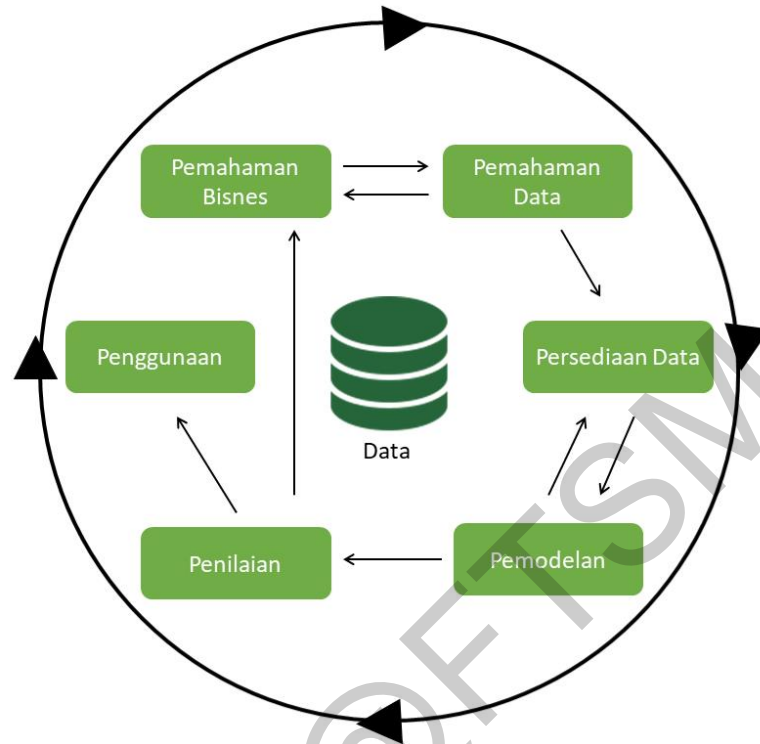
Objektif utama bagi Portal Pelancongan Dengan Analisis Sentimen ini adalah :

1. Membangunkan satu portal pelancongan yang memfokuskan nilai sentimen para pelancong terhadap beberapa tempat. Portal ini dibina menggunakan analisis sentimen.
2. Mengenalpasti perpustakaan yang sesuai untuk domain pelancongan berdasarkan perbandingan perpustakaan analisis sentimen sedia ada.

### **4 METOD KAJIAN**

Metodologi sangat penting dalam penghasilan satu projek kerana kaedah yang sesuai dapat ditentukan untuk penyelesaian masalah yang berkesan. Selain itu, metodologi juga memastikan setiap proses berjalan dengan lancar dan sistematik berdasarkan beberapa fasa. Bagi Portal Pelancongan Dengan Analisis Sentimen ini, metodologi yang digunakan adalah CRISP-DM. CRISP-DM yang merupakan singkatan bagi *Cross-Industry Process for Data Mining* merupakan metodologi yang menyediakan rangka tindakan yang lengkap untuk menjalankan projek perlombongan data (Mayank Aggarwal, 2018). Metodologi ini mempunyai 6 fasa iaitu :

- 1) Pemahaman Bisnes (*Business Understanding*)
- 2) Pemahaman Data (*Data Understanding*)
- 3) Persediaan Data (*Data Preparation*)
- 4) Pemodelan (*Modelling*)
- 5) Penilaian (*Evaluation*)
- 6) Penggunaan (*Deployment*)



Rajah 3 Fasa metodologi CRISP-DM

#### 4.1 Fasa Pemahaman Bisnes

Fasa Pemahaman Bisnes ini merupakan proses memahami keperluan pengguna dan perniagaan kerana memahami keperluan utama projek adalah aspek yang sangat penting. Dalam fasa ini, setiap objektif dikemukakan dengan tepat dan jelas. Perkara ini dapat menjimatkan masa penghasilan projek kerana telah mempunyai matlamat yang perlu diberi perhatian. Selain itu, dalam fasa ini juga, setiap langkah yang diambil juga dijelaskan. Berdasarkan pembinaan Portal Pelancongan ini, keperluan utama adalah memaparkan nilai sentimen positif dan negatif bagi rujukan pelancong.

#### 4.2 Fasa Pemahaman Data

Fasa Pemahaman Data berkait rapat dengan Fasa Pemahaman Bisnes kerana data yang diambil perlulah selari dengan objektif utama projek. Dalam fasa ini, semua data yang berkaitan dikumpulkan. Kemudian data tersebut dikaji seperti menentukan format data, mengenal pasti hubungan setiap data antara satu sama lain dan memastikan data memenuhi keperluan objektif projek. Bagi projek ini, data yang berkaitan tentang sentimen pengguna media sosial atau laman web pelancongan diperlukan dan data tersebut mestilah mempunyai ciri pelancongan. Selain itu, ciri data yang diperlukan juga ialah tarikh, tempat pelancongan dan jenis sentimen (positif dan negatif).

### 4.3 Fasa Persediaan Data

Dalam fasa ini, menentukan set data akhir yang digunakan bagi fasa pemodelan adalah objektif utama. Bagi memastikan set data akhir ini memenuhi semua keperluan projek, data perlu melalui beberapa proses seperti mewujudkan ciri baharu yang berkaitan. Sebagai contoh, menambah ciri jumlah sentimen positif dan negatif kepada data yang sedia ada. Selain itu, data juga mungkin perlu diformat semula. Contohnya, nilai *string* mungkin perlu ditukar kepada nilai *numeric* bagi melakukan operasi matematik. Secara amnya, fasa ini memberi peluang untuk memperbaiki set data yang sedia ada untuk memenuhi semua keperluan utama projek.

### 4.4 Fasa Pemodelan

Fasa Pemodelan merupakan fasa untuk menentukan teknik pemodelan yang akan digunakan. Contoh teknik pemodelan ialah *decision-tree*. Dalam fasa ini, terdapat 4 proses yang perlu dilakukan iaitu:

- i. Menentukan teknik pemodelan
- ii. Mewujudkan reka bentuk ujian (test design)
- iii. Membina model
- iv. Mengakses model

### 4.5 Fasa Penilaian

Model yang dipilih dinilai dan langkah untuk melaksanakan model tersebut dilakukan dalam fasa ini. Pada akhir fasa ini, keputusan daripada perlombongan data (data mining) haruslah diperolehi. Fasa ini mempunyai 3 proses yang perlu dilalui iaitu:

- i. Menilai keputusan dimana hasil akhir perlombongan data haruslah difahami dan menepati keperluan pengguna dan bisnes. Contohnya, adakah jumlah sentimen positif dan negatif mencukupi bagi membuat perbandingan antara kedua-dua sentimen ini?
- ii. Mengkaji semula semua proses yang dilakukan bagi memastikan tiada langkah yang tertinggal
- iii. Menentukan langkah yang iaitu mengenal pasti penggunaan untuk setiap keputusan.

### 4.6 Fasa Penggunaan

Fasa penggunaan merupakan fasa yang melibatkan pengguna untuk menguji hasil projek yang telah dilakukan. Dalam fasa ini plan penggunaan dan penyelenggaraan akan diambil berat. Penyelenggaraan amat penting dalam setiap projek untuk

mengelakkan pelaksanaan perlombongan data yang salah terbiar dalam jangka masa yang lama. Seterusnya, dalam fasa ini juga laporan akhir yang mempunyai ringkasarn projek diwujudkan.

## 5 HASIL KAJIAN

Dalam bahagian ini, hasil projek Portal Pelancongan Dengan Analisis Sentimen ditunjukkan dalam beberapa bahagian iaitu keputusan analisis sentimen, antara muka bagi portal pelancongan dan papan pemuka. Dalam projek ini dua perpustakaan digunakan bagi proses analisis sentimen iaitu *VADER* dan *TextBlob*. Tujuan dua perpustakaan digunakan adalah untuk membuat perbandingan keputusan dan memilih perpustakaan yang mempunyai keputusan yang paling baik. Berdasarkan perbandingan keputusan, *VADER* memberikan yang lebih baik iaitu 1685 sentimen positif, 125 sentimen negatif dan tiada sentimen neutral. Rajah dibawah menunjukkan keputusan analisis sentimen menggunakan *VADER*.

	reviews	name	category	clean_data	scores	compound	sentiment
0	Our stay at King room is very pleasant. The ro...	Holiday Inn Kuala Lumpur Glenmarie	hotel	our stay at king room is very pleasant the roo...	{'neg': 0.0, 'neu': 0.536, 'pos': 0.464, 'comp...	0.9774	positive
1	Very nice and cosy place for staycation, relax...	Holiday Inn Kuala Lumpur Glenmarie	hotel	very nice and cosy place for staycation relaxi...	{'neg': 0.0, 'neu': 0.703, 'pos': 0.297, 'comp...	0.7645	positive
2	I have good lunch today in this hotel. The buf...	Holiday Inn Kuala Lumpur Glenmarie	hotel	I have good lunch today in this hotel the buff...	{'neg': 0.0, 'neu': 0.662, 'pos': 0.338, 'comp...	0.9230	positive
3	Very pleasant hotel to stay although it is an ...	Holiday Inn Kuala Lumpur Glenmaife	hotel	very pleasant hotel to stay although it is an ...	{'neg': 0.0, 'neu': 0.583, 'pos': 0.417, 'comp...	0.9690	positive
4	As a company we've used the Glenmarie Holiday ...	Holiday Inn Kuala Lumpur Glenmarie	hotel	as a company weve used the glenmarie holiday i...	{'neg': 0.0, 'neu': 0.917, 'pos': 0.083, 'comp...	0.4019	positive
...	...	...	...	...	...	...	...
1805	the best place for my kids and grandchildren t...	Zoo Negara	attraction	the best place for my kids and grandchildren t...	{'neg': 0.0, 'neu': 0.724, 'pos': 0.276, 'comp...	0.6369	positive
1806	my kids love to watch animals. we keep visitin...	Zoo Negara	attraction	my kids love to watch animals we keep visitin...	{'neg': 0.0, 'neu': 0.724, 'pos': 0.276, 'comp...	0.6369	positive
1807	Weekday visit is recommended. The animal are h...	Zoo Negara	attraction	weekday visit is recommended the animal are ha...	{'neg': 0.0, 'neu': 0.736, 'pos': 0.264, 'comp...	0.7845	positive
1808	I enjoyed the experience exploring the zoo. Ani...	Zoo Negara	attraction	i enjoyed the experience exploring the zoo ani...	{'neg': 0.049, 'neu': 0.611, 'pos': 0.34, 'com...	0.9231	positive
1809	Terrible zoo management. The are not allowing ...	Zoo Negara	attraction	terrible zoo management the are not allowing m...	{'neg': 0.129, 'neu': 0.871, 'pos': 0.0, 'comp...	-0.7096	negative

1810 rows × 7 columns

Rajah 4 Keputusan analisis sentimen menggunakan *VADER*

```
df_vader['sentiment'].value_counts()

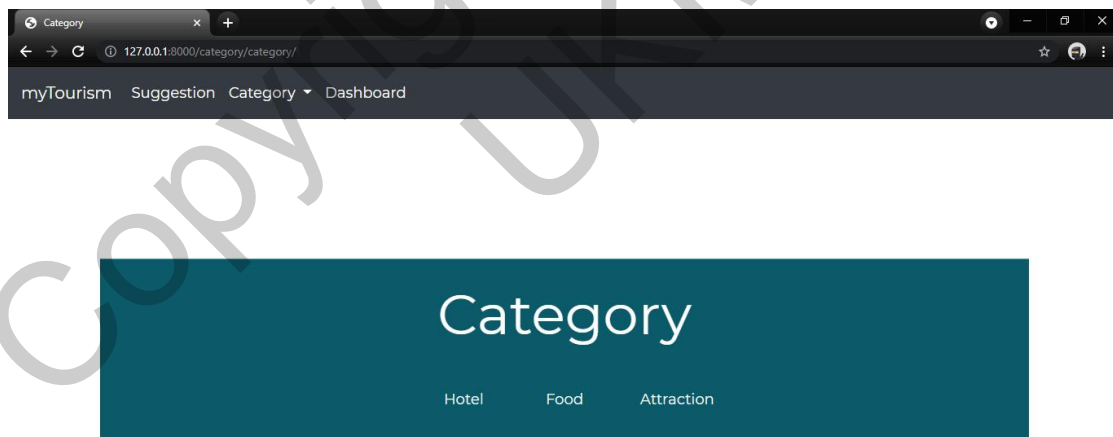
positive    1685
negative    125
Name: sentiment, dtype: int64
```

Rajah 5 Jumlah setiap sentimen menggunakan *VADER*

Seterusnya, portal dihasilkan dan sentimen yang dipaparkan adalah hasil daripada keputusan *VADER*. Portal ini mempunyai 3 kategori iaitu *Hotel*, *Food* dan *Suggestion* serta satu laman *Dashboard*. Rajah dibawah menunjukkan antara muka yang terdapat didalam Portal Pelancongan Dengan Analisis Sentimen ini.

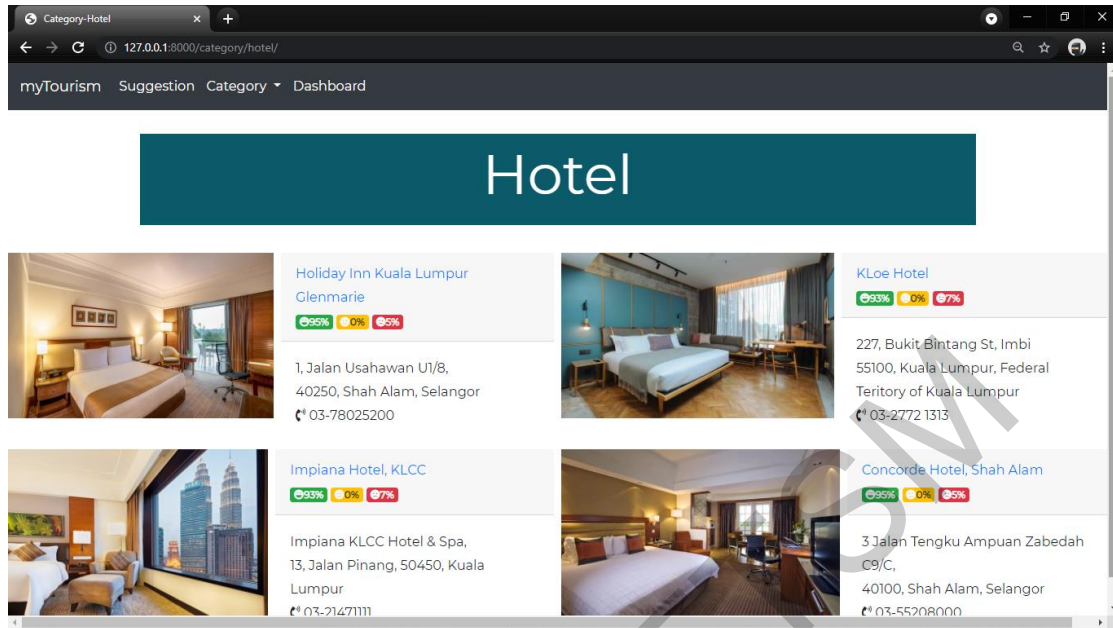


Rajah 6 Antara muka bagi laman utama

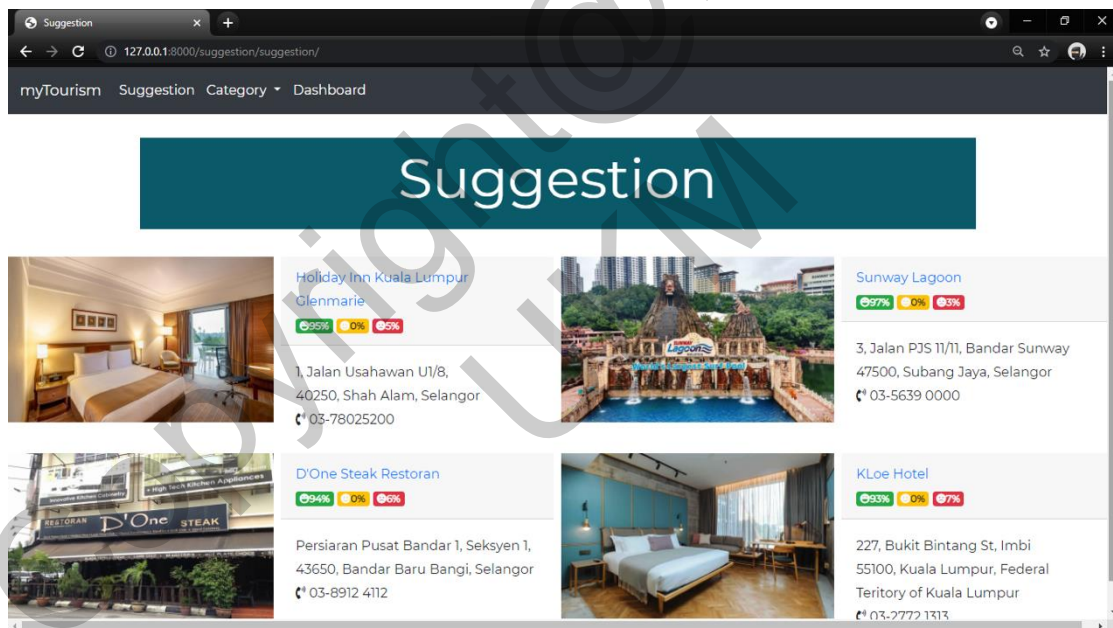


Rajah 7 Antara muka bagi lama “Category”

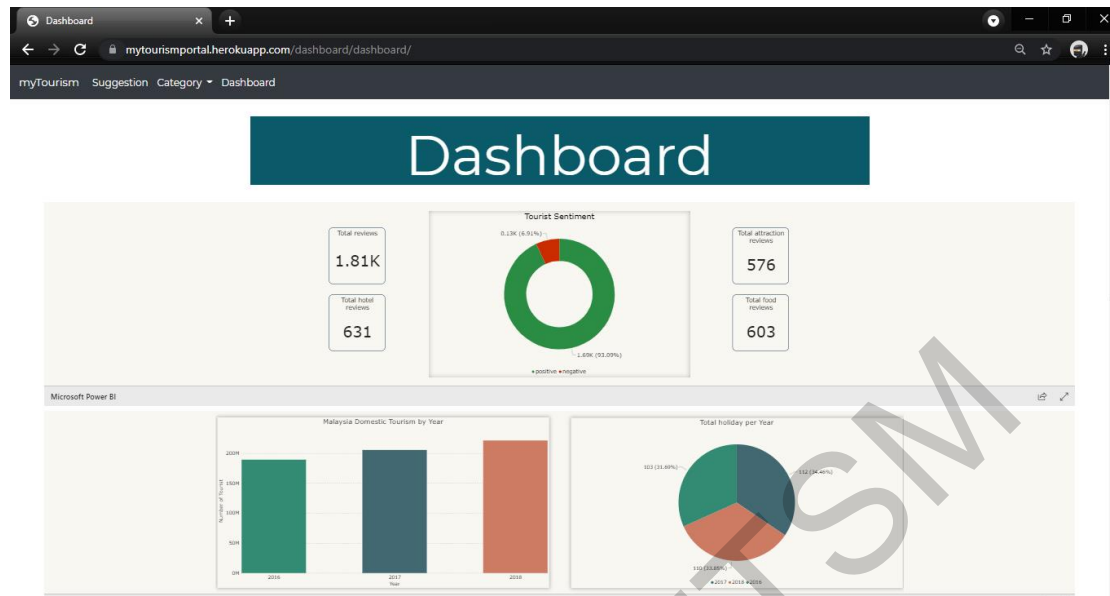




Rajah 8 Antara muka bagi laman “Category-Hotel”



Rajah 9 Antara muka bagi laman “Suggestion”



Rajah 10 Antara muka bagi laman “Dashboard”

Laman Dashboard memaparkan keputusan sentimen daripada pelancong dan beberapa statistik pelancongan antaranya jumlah pelancong domestik mengikut tahun dan cuti tahunan Malaysia mengikut tahun. Mengikut data, tahun 2016 mempunyai pelancong domestik yang sedikit berbanding tahun 2017 dan 2018 dikaitkan dengan jumlah cuti tahunan pada 2016 adalah lebih sedikit berbanding tahun 2017 dan 2018. Rakyat Malaysia sudah tentu merebut peluang semasa tarikh cuti untuk melancong bersama keluarga atau rakan.

## 6 KESIMPULAN

Kesimpulannya, pada akhir proses, projek ini dapat mencapai objektif iaitu membangunkan satu portal pelancongan yang memfokuskan nilai sentimen para pelancong terhadap beberapa tempat dan mengenalpasti perpustakaan yang sesuai untuk domain pelancongan berdasarkan perbandingan perpustakaan analisis sentimen sedia ada. Portal berjaya dibina dan memaparkan sentimen daripada pelancong hasil daripada keputusan analisis sentimen menggunakan perpustakaan VADER.

Walaupun bagaimanapun, projek ini tetap mempunyai kekangan antaranya adalah data yang diperolehi adalah terhad. Ini adalah kerana data ataupun *reviews* yang dikumpulkan hanya daripada media sosial dan laman web pelancongan. Seterusnya, maklumat yang terbaharu juga sukar untuk didapati akibat pandemik COVID-19 dimana rakyat Malaysia kurang keluar melancong bagi mengelakkan virus terus merebak. Namun begitu, portal ini diharapkan dapat memberi manfaat dengan memaparkan sentimen bagi destinasi menarik mengikut kategori.

## 7 RUJUKAN

Kechit Goyal. March 2020. Top 4 Types of Sentiment Analysis & Where to Use.

Sumit Shukla. June 2018. What is Machine Learning and Why it Matters

Matthew Hudson. June 2020. What is Social Media?

Shona McCombes. February 2019. How to write a literature review.

MonkeyLearn. 2020. Sentiment Analysis: A Definitive Guide.

Nurul Huda Husain, Sinar Harian. Julai 2020. Sektor pelancongan negara rugi RM45 bilion.

Mark Kraeling, Lindsley Tania. 2019. Software Engineering for Embedded Systems (Second Edition).

Ulf Eriksson. April 2012. Why is the difference between functional and non-functional requirement important ?

Altexsoft. November 2019. Non-functional Requirements: Examples, Types, How to Approach.

Ralph Hughes MA, PMP, CSM. 2016. Agile Data Warehousing for the Enterprise.

Techopedia. Ogos 2015. System Requirements.

Margaret Rouse. July, 2020. What is web server and how does it work ?

Nikhil Pavanam. November 2020. Understanding Naïve Bayes and Support Vector Machine and their implementation in Python.

Sunil Ray. September 2017. 6 Easy Steps to Learn Naïve Bayes Algorithm with codes in Python.

Sunil Ray. September 2017. Understanding Support Vector Machine (SVM).

Copyright@FTSM  
UKM