

## **ANALISIS SENTIMEN DAN PENGELASAN FILEM DAN DRAMA BERSIRI DI NETFLIX MENGGUNAKAN PEMBELAJARAN MESIN**

Thanusha a/p K.Ravana  
Lailatul Qadri Zakaria

Fakulti Teknologi dan Sains Maklumat, Universiti Kebangsaan Malaysia

### **ABSTRAK**

Medium hiburan atas talian merupakan platform yang semakin membangun pada era terkini dan Netflix merupakan salah satu aplikasi hiburan atas talian yang mendapat tempat di hati rakyat Malaysia. Namun begitu, terdapat terlalu banyak kebebasan pengguna untuk menonton filem dan drama bersiri di Netflix kerana aplikasi ini tidak melalui proses penapisan mengikut garis panduan yang telah ditetapkan oleh FINAS. Oleh itu, kajian ini bertujuan untuk mengkaji masalah kandungan di Netflix yang melibatkan isi-isu yang melanggar norma masyarakat Malaysia mengikut tiga kategori utama iaitu keagamaan, kebudayaan dan nilai moral. Kajian ini menggunakan Teknik analisis pembelajaran mesin bagi mengenalpasti drama atau filem yang tidak mematuhi garis panduan yang telah ditetapkan oleh FINAS dengan menggunakan analisis sentimen dan pengelasan. Kajian ini mengambil ulasan penonton dari Twitter. Ulasan penonton memainkan peranan penting bagi mendapatkan pandangan atau perasaan penonton mengenai drama atau filem tertentu. Analisis sentimen digunakan untuk mengenalpasti drama atau filem yang mendapat ulasan yang negatif dari penonton. Analisis pengelasan dilakukan untuk mengenalpasti kategori ketidaksesuaian kandungan Netflix samaada keagamaan, kebudayaan atau nilai moral. Kajian ini diharap dapat akan dapat membantu pengguna mengenalpastu filem atau drama yang tidak patuh dengan garis panduan yang telah ditetapkan oleh FINAS.

## 1 PENGENALAN

Medium hiburan atas talian atau lebih dikenali sebagai platform Over-The-Top (OTT) semakin berkembang dari masa ke masa. Dalam era modenisasi sekarang ini, jumlah orang yang menggunakan perkhidmatan penstriman media ini meningkat secara drastik terutamanya sewaktu Perintah Kawalan Pergerakan (PKP) yang dilaksanakan berikutan wabak Covid-19 yang melanda seluruh dunia. Menurut Malay Mail, PKP yang dilaksanakan sejak 18 Mac 2020, telah meningkatkan penggunaan OTT kerana orang terpaksa tinggal di rumah dan kebanyakan premis ditutup termasuklah panggung wayang. Jadi, orang ramai beralih ke medium hiburan atas talian seperti Amazon Prime, Astro on the Go, Netflix, iFlix, Viu, Tonton, Dimsum dan HyppTV Everywhere. Platform ini memang mendapat sambutan yang banyak dari orang ramai sejak diperkenalkan beberapa tahun yang lalu kerana penggunaannya yang lebih selesa, mudah dicapai, harga yang berpatutan dan akses kepada pelbagai saluran selaras dengan perkembangan rangkaian Internet di Malaysia.

Menurut Wikipedia, Netflix merupakan satu syarikat penerbitan dan penyedia perkhidmatan penstriman media berdasarkan langganan yang menawarkan penstriman dalam talian dari perpustakaan filem, drama bersiri dan rancangan televisyen. Netflix diperkenalkan pada tahun 1997 di California oleh Marc Randolph dan Reed Hastings. Netflix memulakan perkhidmatan penstriman media atas talian pada tahun 2007 dan mula digunakan dalam skala antarabangsa pada tahun 2010. Netflix mula dicapai di Malaysia pada tahun 2016 dengan beberapa pelan pakej langganan mengikut kriteria yang ditetapkan. Sejurus dengan kemasukan Netflix di Malaysia, servis penstriman ini berjaya mendapat sambutan yang hangat daripada para pengguna. Dengan adanya pelbagai rancangan televisyen, drama bersiri, dokumentari dan filem, pelanggan boleh menonton kandungan yang ditawarkan pada bila-bila masa dan di mana-mana sahaja. Bukan itu sahaja, pelanggan juga boleh menonton Netflix di semua platform atau alat peranti kerana Netflix adalah sangat mesra pengguna. Harga langganannya juga berpatutan. Walau bagaimanapun, Netflix cenderung untuk membuang sesetengah rancangan yang sebelumnya ada dan pelanggan perlu ada akses kepada internet untuk menggunakannya. Netflix juga tiada sebarang jenis penapisan kandungan yang ditayangkan.

Di Malaysia, masyarakat kita terdiri daripada pelbagai kaum, agama, budaya dan etnik. Sebagai rakyat Malaysia, kita mempunyai nilai-nilai sosial, budaya dan norma-norma tersendiri yang menjadi panduan kehidupan orang ramai. Oleh itu, sebilangan kandungan Netflix mungkin tidak sesuai dengan norma masyarakat Malaysia. Penggunaan Netflix di Malaysia sepatutnya mematuhi peraturan kandungan yang dibenarkan oleh Perbadanan Kemajuan Filem Nasional Malaysia (FINAS) yang bersesuaian dengan norma masyarakat Malaysia. Ketua Pegawai Eksekutif FINAS, Ahmad Idham Ahmad Nadzri yang mahukan kandungan Netflix melalui proses tapisan kerajaan, pernah menekankan bahawa isu ini harus dibincangkan agar ia tidak membawa unsur negatif dalam masyarakat dan menjejaskan hak pengguna itu sendiri. Isu-isu yang boleh dikaitkan dengan kandungan Netflix yang tidak sesuai dengan norma masyarakat Malaysia adalah seperti sensitiviti keagamaan, sensitiviti politik Malaysia, unsur tidak senonoh seperti pornografi dan pedofilia, LGBTQ, keganasan yang melampau, unsur yang mengganggu-gugat sosiobudaya dan keharmonian rakyat Malaysia, serta pengamalan ilmu hitam.

Isu mengenai kandungan Netflix bukan sahaja menjadi isu yang membimbangkan masyarakat Malaysia, malah negara lain seperti Singapura dan Saudi Arabia juga banyak mengungkitkan masalah ini. Oleh itu, topik ini banyak dibincangkan terutamanya dari pihak FINAS, Lembaga Penapis Filem (LPF) dan juga ibu bapa yang bimbang akan pendedahan anak-anak mereka terhadap kandungan Netflix yang tidak dikawal. Sebagai cara penyelesaian masalah ini, teknik pembelajaran mesin boleh dilakukan ke atas data-data di media sosial seperti Twitter yang berkaitan dengan kandungan Netflix yang tidak terkawal. Proses analisis sentimen dapat mengenal pasti corak dalam sesuatu data dan data tersebut dapat diproses untuk menjadi bentuk yang lebih mudah untuk dibaca dan difahami. Proses ini secara umumnya mengenal pasti sentimen di sebalik sesuatu teks. Terdapat tiga jenis sentimen utama iaitu positif, negatif dan neutral. Dalam membuat analisis sentimen untuk ulasan filem atau drama bersiri dari Netflix, corak pendapat umum dapat diberi pertimbangan untuk memahami perasaan penonton mengenai tema-tema yang ditunjukkan dalam filem tersebut. Oleh kerana platform media sosial banyak digunakan orang pada masa kini, mereka lebih cenderung untuk mengkritik atau meluahkan pendapat mengenai sesuatu filem atau drama bersiri di Twitter atau Facebook. Jadi dalam kajian ini, klasifikasi sentimen teks media sosial berkenaan filem atau

drama bersiri Netflix akan dilaksanakan mengikut kandungan yang bersesuaian dengan penonton yang terdiri daripada masyarakat Malaysia.

## **2 PENYATAAN MASALAH**

Medium hiburan atas talian merupakan platform yang sangat hangat pada masa sekarang. Walaupun dunia hiburan ini sangat besar, penggunaan servis Over-The-Top (OTT) seperti Netflix dan Amazon Prime sangat popular terutamanya pada masa negara sedang dilanda wabak Covid-19 ini. Kajian yang sedia ada banyak memberikan kesimpulan keseluruhan mengenai masalah kandungan di Netflix yang melibatkan isi-isu yang melanggar norma kehidupan umum dan tidak hanya khusus kepada norma masyarakat Malaysia. Cara kehidupan di negara Barat seperti Amerika Syarikat dan cara kehidupan di Malaysia mempunyai sedikit sebanyak perbezaan yang boleh dianggap ketara. Model-model atau proses sentimen ulasan filem yang diambil kira dari web seperti IMDB dan Rotten Tomatoes juga kesemuanya adalah sangat umum. Perkara yang seharusnya dipertimbangkan adalah siaran kandungan Netflix yang sepatutnya mematuhi peraturan kandungan yang dibenarkan oleh Perbadanan Kemajuan Filem Nasional Malaysia (FINAS) yang bersesuaian dengan norma masyarakat Malaysia. Medium yang sangat rapat dengan masyarakat pada masa kini adalah media sosial seperti Twitter. Maka, semua jenis luahan pendapat dan kritikan sering dibuat dalam bentuk cicipan. Oleh itu, sentimen perlu ditetapkan menerusi analisis pandangan pengguna terhadap kandungan Netflix yang ditonton melalui cicipan pengguna tersebut.

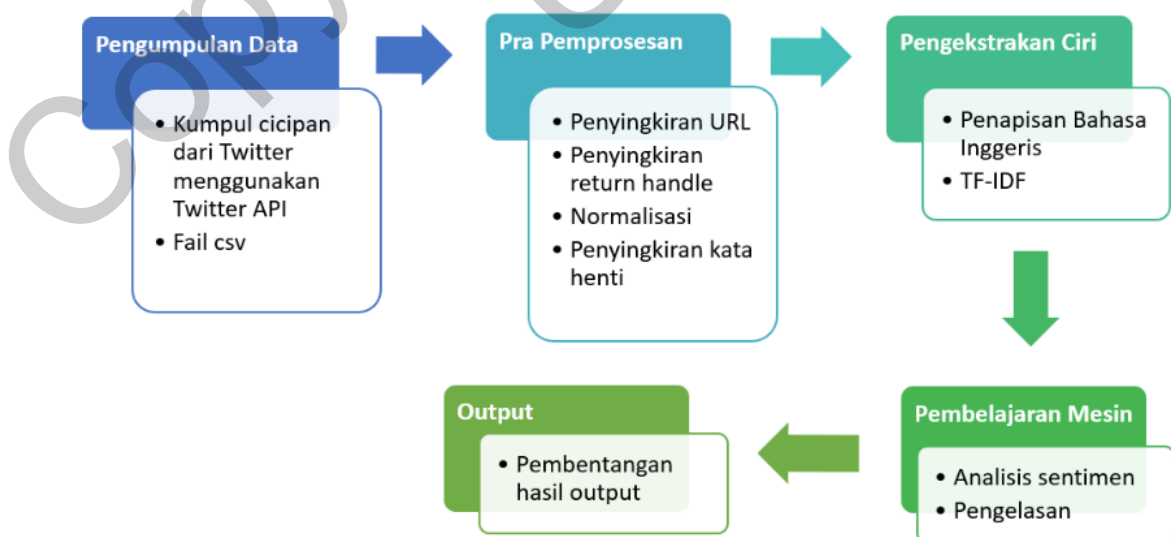
## **3 OBJEKTIF KAJIAN**

Terdapat banyak model yang dapat menentukan sentimen teks kepada nada positif, negatif atau neutral. Dalam cicipan yang dimaksudkan untuk mengkritik filem atau drama bersiri dari Netflix, banyak perkataan spesifik yang harus dipertimbangkan untuk memahami konteks teks secara keseluruhan. Oleh itu tujuan utama kajian ini adalah untuk mengklasifikasi kandungan cicipan berkenaan filem atau drama bersiri Netflix dengan sewajarnya mengikut kandungan yang bersesuaian dengan penonton yang terdiri daripada masyarakat Malaysia. Untuk mencapai matlamat ini, objektif berikut digariskan:

- a. Mengenal pasti sentimen filem atau drama bersiri Netflix yang dianalisis dengan menggunakan analisis sentimen.
- b. Mengenal pasti kategori filem atau drama bersiri Netflix tersebut berdasarkan tiga kategori yang telah ditetapkan oleh Perbadanan Kemajuan Filem Nasional Malaysia (FINAS) menerusi pendekatan Naïve Bayes.

#### 4 METOD KAJIAN

Pembangunan model yang sesuai penting untuk memastikan kajian berjalan dengan baik dan mendapatkan hasil keputusan yang tepat dan berkualiti. Model metodologi yang digunakan untuk membangunkan algoritma ini adalah Model Klasifikasi Sentimen. Model ini mudah untuk difahami dan digunakan dalam proses analisis sentimen. Fasa pembangunan algoritma senang untuk dipantau dan risiko kegagalan semasa pembangunan algoritma adalah kurang. Rajah 1 merujuk kepada model metodologi kajian ini. Beberapa model sistem bagi projek ini dirujuk daripada kajian salah seorang pelajar Universiti Kebangsaan Malaysia (Tan Fui Ken, 2019).



Rajah 1 Model Klasifikasi Sentimen

#### **4.1 Fasa Perancangan**

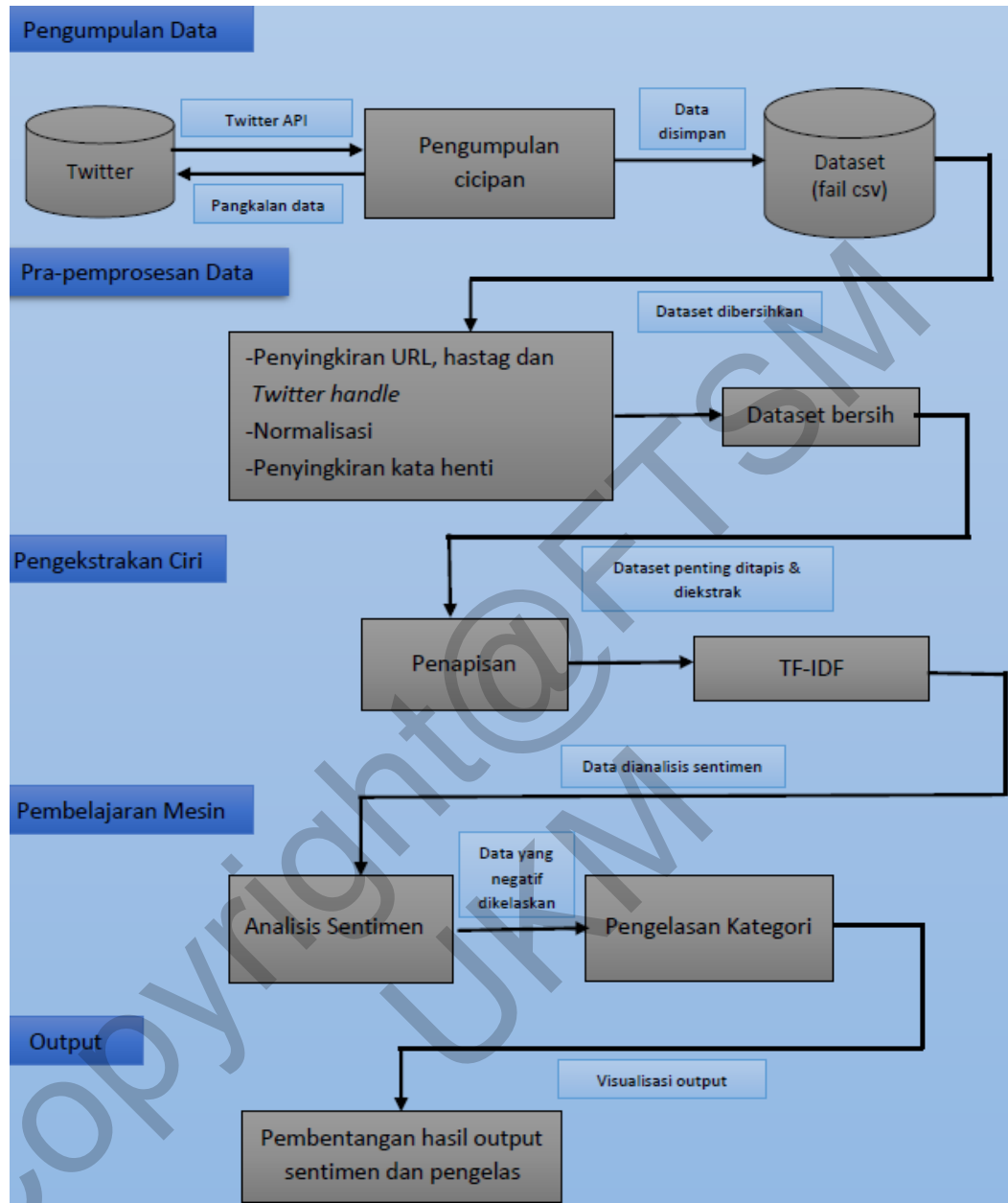
Fasa ini melibatkan proses pengumpulan dan analisis keperluan kajian seperti set data yang berkaitan dengan objektif kajian dan permasalahan kajian serta menganalisis sorotan kajian susastera yang melibatkan pengumpulan maklumat, pencarian jurnal, pembacaan jurnal dan kajian lepas sebagai rujukan. Antara topik yang dikaji adalah berkaitan pengesanan sentimen ulasan filem daripada *IMDB*, *Rotten Tomatoes* dan media sosial menggunakan pembelajaran mesin sedia ada. Set data yang berkaitan dengan kajian daripada Twitter dikumpulkan.

#### **4.2 Fasa Analisis**

Fasa ini melibatkan analisis dan tafsiran terhadap data yang dikumpulkan. Hal ini bertujuan untuk memastikan data yang dikumpulkan adalah sesuai dan relevan dengan topik kajian serta diselaraskan kepada Bahasa Inggeris sahaja. Kemudian, kesemua data yang ditapis dibersihkan lagi ke dataset yang lebih bermakna melalui pra-pemprosesan data dan pengekstrakan ciri. Sebanyak 2,316 set data iaitu cicipan dari Twitter berjaya dikumpul dan dilabelkan secara manual kepada tiga kategori yang menjadi skop kajian ini iaitu nilai moral, kebudayaan dan keagamaan (Wan Amizah, Chang & Jamaluddin 2009). Ini bertujuan untuk memudahkan data untuk dianalisis. Analisis data dijalankan bagi memastikan tiada data yang hilang, tiada data yang berulang, mengetahui jenis data, statistik set data dan visualisasi taburan data bagi mendapatkan set data yang berkualiti.

#### **4.3 Fasa Reka Bentuk**

Fasa ini melibatkan reka bentuk seni bina yang secara umumnya memberi gambaran awal secara keseluruhan bagi pembangunan system kajian ini. Reka bentuk ini juga menunjukkan hubung kait antara sub-sistem secara terperinci. Merujuk kepada rajah 2 di bawah, reka bentuk kajian ini terbahagi kepada lima fasa utama iaitu pengumpulan data, pra-pemprosesan data, pengekstrakan ciri, pembelajaran mesin dan pembentangan output.



Rajah 2 Reka Bentuk Seni Bina Kajian

- i) Pengumpulan data melibatkan pengumpulan cicipan daripada aplikasi Twitter yang mengandungi ulasan filem atau drama bersiri dari Netflix. Perpustakaan Tweepy digunakan bagi mendapatkan set data mentah daripada aplikasi Twitter menggunakan Twitter API. 'Kunci API' dan 'kunci rahsia API' dari akaun Twitter Developer digunakan untuk mencipta akses kepada token. Kemudian, API yang digunakan dalam perisian mempunyai kata kunci yang tertentu seperti 'nama filem' atau 'nama drama bersiri' berserta dengan 'netflix' bagi memperincikan cicipan yang dikumpulkan adalah sesuai dengan projek ini. Filem-filem dan

drama bersiri yang digunakan dalam kajian ini adalah *Cuties*, *Messiah*, *Big Mouth*, *Bonding*, *Enola Holmes*, *One of Us*, *The Devil All The Time*, *Lucifer*, *The Keepers*, *Indian Matchmaking*, *Sex/Life*, dan *Elite*. Sebanyak 3170 cicipan (data mentah) yang terkini berjaya dikumpulkan sepanjang perancangan projek ini iaitu bermula bulan Disember 2020 sehingga Mei 2021. Cicipan yang dikumpulkan disimpan dalam beberapa fail csv mengikut tajuk filem atau drama bersiri Netflix tetapi kemudian disatukan dalam satu fail csv secara keseluruhan selepas ditapis secara manual dan dilabel kategorinya berdasarkan pemerhatian pengkaji. Data yang dikumpulkan juga ditapis dan diselaraskan kepada satu Bahasa sahaja iaitu Bahasa Inggeris. Jumlah data selepas ditapis adalah 2,316.

ii) Pra-pemprosesan teks penting bagi membersihkan teks dari elemen hingar (*noise*). Langkah pra pemprosesan yang dijalankan ialah menyingkir karakter atau simbol yang sinonim dengan data Twitter seperti nama pengguna (@), tanda hastag (#), pautan URL, dan ulangan cicipan (RT), proses normalisasi seperti penukaran cicipan kepada huruf kecil, penghapusan tanda baca serta penyingkiran kata henti. Pakej NLTK diimport untuk proses normalisasi. Pakej ini adalah rangkaian perpustakaan dan program untuk pra pemprosesan bahasa semula jadi simbolik dan statistik (NLP) untuk perisian Python dalam bahasa Inggeris. Pakej NLTK ini secara kebiasaannya digunakan untuk menyokong fungsi klasifikasi, tokenisasi, *stemming*, *tagging*, penghuraian, dan penaakulan semantik. Rajah 3 menunjukkan output data sebelum dan selepas diproses.

	Text	Processed
0	@SodoMe1 @GOJOT1TTIES eh, it's netflix. same people who thought cuties was a great idea	eh netflix people thought cuties great idea
1	@spideysnews @netflix Even Spider-Man won't get me to use Netflix. Not until they remove Cuties and apologize.	even spider man get use netflix remove cuties apologize
2	@IGN Damn! Sounds like Sony supports pedo. Cuties is still on Netflix. I canceled Netflix. I do not support pedo. #RemoveCuties	damn sounds like sony supports pedo cuties still netflix canceled netflix support pedo
3	@spideysnews @netflix Bleep you, @netflix I canceled Netflix due to Cuties! I refuse to return till you remove Cutá❏ <a href="https://t.co/TH7RcsvRRV">https://t.co/TH7RcsvRRV</a>	bleep you canceled netflix due cuties refuse return till remove cut
4	@fridaybobb Does not make it true. Netflix aired Cuties sooo?	make true netflix aired cuties sooo
5	@fridaybobb I don't think they care tbh. Netflix aired "Cuties" what does that tell you about Netflix?	think care tbh netflix aired cuties tell netflix
6	@MiriaAkagi5 @Memelin_sequel I don't know, Miria, you look like the type of person who hates lolicon, but loves Netflix's Cuties.	know miria look like type person hates lolicon loves netflix cuties
7	Because Netflix is the same company made Cuties. <a href="https://t.co/CIAMQIU2hD">https://t.co/CIAMQIU2hD</a>	netflix company made cuties
8	@GameSpot ugh so the only way I will see spiderman is rental or theaters I WILLNOT WATCH ANYTHING on netflix until❏ <a href="https://t.co/nw67W1DhaF">https://t.co/nw67W1DhaF</a>	ugh way see spiderman rental theaters willnot watch anything netflix until
9	if you enjoyed @netflix movie #cuties" it means you're a chomo. #StopChildAbuse #stopsexualtrafficing #ProtectTheKids	enjoyed movie means chomo
10	@TheePill @VitoGesualdi You're arguing with the guy that defended Netflix Cuties for days on end. He does it to be❏ <a href="https://t.co/qcpZdu4NcB">https://t.co/qcpZdu4NcB</a>	arguing guy defended netflix cuties days end be
11	hold on all these baseball tweets I thought we were boycotting MLB like we all gave up our Netflix accounts for cuties?	hold baseball tweets thought boycotting mlb like gave netflix accounts cuties

Rajah 3 Data Sebelum dan Selepas Diproses



iii) Analisis sentimen dilakukan dengan menggunakan Textblob. Fasa ini penting bagi mengenalpasti data yang mendapat nilai sentimen yang positif, negatif dan neutral. Textblob akan mengabaikan kata-kata yang tidak diketahui, ia akan mempertimbangkan perkataan dan frasa yang dapat memberikan polariti dan purata untuk memperoleh skor akhir. Pengiraan ciri sentimen cicipan dibuat dalam bentuk polariti dan subjektiviti. Kemudian, perlu mengklasifikasikan polariti sebagai positif, neutral atau negatif. Cicipan dilabel mengikut keputusan sentimen yang diperolehi. Bagi kajian ini, hanya data yang mendapat nilai sentimen yang negatif sahaja akan dianalisis dengan lebih terperinci. Data yang mendapat nilai sentimen yang negatif menunjukkan ketidaksesuaian kandungan filem atau drama bersiri tersebut serta peranan negatif netizen terhadap kandungan Netflix yang dibincangkan. Oleh itu, analisis sentimen ini penting bagi mendapatkan data yang bermakna bagi memenuhi objektif kajian ini.

iv) Fasa pengekstrakan ciri dilakukan bagi mendapatkan ciri penting yang akan digunakan sebagai perwakilan data. Ciri utama yang diekstrak ialah *Term Frequency-Inverse Document Frequency (TF-IDF)*. TF-IDF adalah ukuran statistik yang menilai kadar relevan sesuatu perkataan dalam sesebuah dokumen dengan kumpulan dokumen tersebut. Ini dilakukan dengan mendarabkan dua metrik iaitu, frekuensi istilah dan frekuensi dokumen songsang. Frekuensi Istilah (TF) merujuk kepada berapa kali sesuatu perkataan muncul dalam dokumen dan Frekuensi Dokumen Songsang (IDF) merujuk kepada kekerapan songsang perkataan tersebut dalam sekumpulan dokumen. TF dan IDF tidak perlu dikira secara manual, sebaliknya *sklearn* mempunyai fungsi yang boleh mendapatkannya secara langsung iaitu menggunakan *TfidfVectorizer* dalam bahasa pengaturcaraan Python. Vektor akan mengira berat setiap perkataan dalam korpus dan akan mengembalikan TF-IDF.

iv) Fasa pengelasan data dilakukan dengan menggunakan teknik *Multinomial Naïve Bayes*. Teknik ini dipilih kerana secara kebiasannya digunakan untuk data set media sosial dan sangat mudah dan pantas untuk meramalkan kelas set data ujian. Ia juga menunjukkan prestasi yang baik dalam ramalan *multi-class* sepertimana yang digunakan dalam kajian ini juga. *Naïve Bayes* menggunakan kebarangkalian bersyarat. Algoritma yang digunakan dalam pengelasan *Naïve Bayes* dalam kajian ini menggunakan semua ciri vector TF-IDF dan pengelas menganalisis ciri secara bebas. Data dibahagi kepada data set latihan sebanyak 80%

dan data set pengujian sebanyak 20%. Data akan dikategorikan dalam tiga kelas utama iaitu keagamaan, nilai moral dan kebudayaan.

#### 4.4 Fasa Pengujian

Fasa ini melibatkan pengujian teknik analisis sentimen dan model pengelas yang dipilih bagi kajian ini. Terdapat banyak teknik analisis sentiment yang boleh digunakan bagi menganalisis sentimen di sebalik cicipan yang dikumpulkan. Oleh itu, kajian ini telah menjalankan analisis awal (*preliminary analysis*) bagi membandingkan dan memilih teknik analisis sentimen yang bersesuaian, iaitu sama ada *VADER Sentiment Analysis* atau *TextBlob Sentiment*. Analisis awal ini dijalankan dengan menggunakan 200 cicipan daripada data mentah. 200 cicipan ini dilabel terlebih dahulu mengikut pemerhatian pengkaji sama ada positif, negatif atau neutral. Kemudian, kesemua cicipan tersebut menjalani kaedah pra pemrosesan seperti yang diterangkan di fasa reka bentuk. Seterusnya, cicipan yang sudah diproses dimasukkan pula ke dalam *VADER* dan *TextBlob* bagi mendapatkan output sentiment mengikut teknik masing-masing. Output dari kedua-dua teknik ini dibandingkan dengan label sentimen yang sudah dibuat secara manual oleh pengkaji terlebih dahulu tadi. Hasil analisis awal mendapati *TextBlob Sentiment Analysis* adalah lebih tepat, maka teknik ini yang digunakan bagi pembangunan algoritma yang selanjutnya. Pengujian yang seterusnya bagi projek ini melibatkan pengklasifikasian cicipan yang bersentimen negatif kepada tiga kategori utama yang sudah ditetapkan iaitu nilai moral, keagamaan dan kebudayaan menggunakan vektor TF-IDF. Bagi penilaian pengelas, ketepatan pengelas *Naïve Bayes* bagi latihan dan pengujian adalah seperti dalam rajah 4 di bawah.

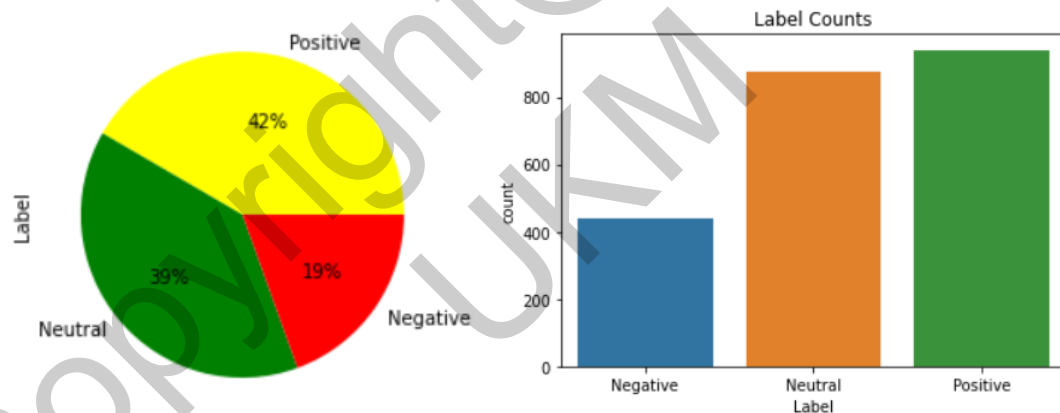
```
nb = BernoulliNB()
nb.fit(X,y)
train_acc = nb.score(X, y)
test_acc = nb.score(X_test, y_preds)
print("Training Accuracy for Naive Bayes:",train_acc)
print("Testing Accuracy for Naive Bayes:",test_acc)
```

```
Training Accuracy for Naive Bayes: 0.9726651480637813
Testing Accuracy for Naive Bayes: 1.0
```

Rajah 4 Ketepatan Pengelas *Naïve Bayes*

## 5 HASIL KAJIAN

Hasil keputusan divisualisasikan dalam bentuk graf bar dan carta pai bagi memudahkan pengguna menilai ulasan pengguna Twitter berkaitan dengan filem dan drama bersiri di Netflix yang mungkin tidak sesuai ditayangkan di Malaysia selaras dengan garis panduan yang ditetapkan oleh FINAS. Keputusan analisis sentimen menggunakan teknik *TextBlob* dikeluarkan dalam bentuk carta pai dalam Rajah 5 di mana carta tersebut menggambarkan peratusan cicipan yang bersentimen positif, negatif dan neutral. Didapati sentimen dataset yang digunakan untuk kajian ini terdiri daripada 42% cicipan positif, 19% cicipan negatif dan 39% cicipan neutral. Rajah 5 juga merujuk kepada graf bar yang menunjukkan jumlah cicipan yang bersentimen positif, negatif dan neutral. Didapati jumlah cicipan yang bersentimen positif adalah 940, jumlah cicipan yang bersentimen negatif adalah 439 dan jumlah cicipan yang bersentimen neutral adalah 877.



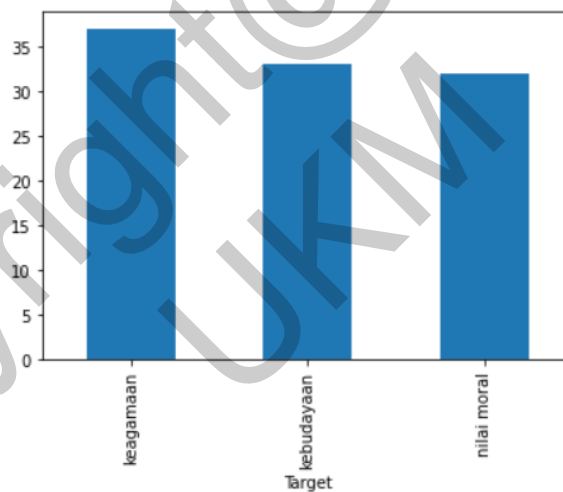
Rajah 5 Output Analisis Sentimen

Antara contoh cicipan positif, negatif dan neutral adalah seperti dalam rajah 6 berikut. Berdasarkan *TextBlob*, polariti adalah apungan yang terletak di antara  $[-1,1]$ , di mana  $-1$  menunjukkan sentimen negatif dan  $+1$  menunjukkan sentimen positif manakala subjektiviti adalah apungan yang berada dalam julat  $[0,1]$ . Ayat atau frasa subjektif umumnya merujuk kepada pendapat, emosi, atau penilaian peribadi. *Textblob* akan mempertimbangkan perkataan dan frasa yang dapat memberikan polariti dan purata untuk memperoleh skor akhir dan kemudian mengklasifikasikan polariti tersebut sebagai positif, neutral atau negatif. Cicipan dilabel mengikut keputusan sentimen yang diperolehi.

Processed	Polarity	Subjectivity	Label
nooooo i love big mouth especially lola she is the reason i watch the show lol	0.325	0.6	Positive
big mouth is one show that always makes my mood better	0.25	0.3	Positive
pedophilia is wrong cuties was a terrible film	-0.75	0.95	Negative
i just started watching the keepers on netflix and i just wanted to say i am disgusted men cant be trusted	-1	1	Negative
our feisty indian aunty watches and believes that dna testing has nothing on indian matchmaking	0	0	Neutral
elite season is out on netflix overrated show	0	0	Neutral

Rajah 6 Hasil Output *TextBlob*

Hasil keputusan pengelas *Naive Bayes* dikeluarkan dalam bentuk graf bar bagi pandangan keseluruhan output model pengelas iaitu mengklasifikasikan cicipan yang bersentimen negatif kepada 3 kategori yang tidak sesuai ditayangkan di Malaysia selaras dengan garis panduan yang ditetapkan oleh FINAS. Merujuk kepada rajah 7, jumlah cicipan yang dikategorikan di bawah nilai moral, kebudayaan dan keagamaan digambarkan berdasarkan 20% data ujian yang dimasukkan ke dalam model pengelas *Naive Bayes*.

Rajah 7 Output Pengelas *Naive Bayes*

Berikut adalah contoh cicipan yang telah dikenalpasti sebagai kategori keagamaan. Merujuk kepada rajah 8, beberapa perkataan dari drama bersiri *Messiah* menggambarkan perkataan unik seperti “religious”, “faith”, “jesus”, “christian”, “devil” dan seumpamanya dapat dikaitkan dengan kategori keagamaan. Antara perkataan unik lain yang menggambarkan kategori keagamaan daripada cicipan lain adalah “jewish”, “god”, “church”, “mosque”, “priest”, “catholic”, “Islam” dan sebagainya.

Tweet	Target
i watched messiah on netflix yesterday i not religious but it was not nice	keagamaan
messiah was a clever thought provoking series about contemporary faith deceit mystery very disappoi	keagamaan
jesus chilling out in the rock tomb watching messiah on netflix	keagamaan
hating messiah season hate how jesus is portrayed as problematic but also portrayed inaccurately as u	keagamaan
please let us dispose this christian leftist idea of making castes taboo bollywood and netflix are	keagamaan
i watching lucifer really sad devil guy	keagamaan

Rajah 8 Contoh Cicipan Kategori Keagamaan

Berikut adalah contoh cicipan yang telah dikenalpasti sebagai kategori kebudayaan. Tambahan pula, merujuk kepada rajah 9, perkataan-perkataan yang boleh dikaitkan dengan kategori kebudayaan adalah seperti “bdsbm”, “matchmaking”, “creepy”, “gay”, “divorce”, “lgbt”, “whores”, “sex”, “trashy”, “caste”, “history” dan sebagainya.

Tweet	Target
if you want bad bdsm watch bonding on netflix	kebudayaan
bonding season one on netflix is sadly a flaccid look at the world of bdsm skip it	kebudayaan
randomly started watching indian matchmaking on netflix and it actually way more creepy than what i thought	kebudayaan
netflix needs to stop making all the gays act likes whores its so frustrating	kebudayaan
stopped watching elite when the girl took off her hijab for a white man i hate netflix	kebudayaan
elite its a spanish show on netflix but i watch it illegally a warning the show has a lot of sex so	kebudayaan

Rajah 9 Contoh Cicipan Kategori Kebudayaan

Berikut adalah contoh cicipan yang telah dikenalpasti sebagai kategori nilai moral. Merujuk kepada rajah 10, beberapa perkataan dari filem seperti Cuties menggambarkan perkataan unik seperti “strippers”, “pedo”, “child porn”, “disgusting”, dan seumpamanya boleh dikaitkan dengan kategori nilai moral. Antara perkataan unik lain yang menggambarkan nilai moral daripada cicipan lain adalah “wrong”, “worst”, “shameless”, “horrifying”, “molester”, “disgusted” dan sebagainya.

Tweet	Target
isnt cuties a film about little girl strippers yall are so disgusting	nilai moral
people still watch netflix i canceled mine soon as they put that disgusting cuties	nilai moral
i quit netflix when they started catering to pedo with that cuties series shit	nilai moral
i am being accused for endorsing child porn apparently i think cuties is bad and sh	nilai moral
sorry to burst your misguided bubble my guy but he didn say anyone who still use	nilai moral
and unfortunately one of them shows is the pedo attraction movie cuties	nilai moral

Rajah 10 Contoh Cicipan Kategori Nilai Moral

## 6 KESIMPULAN

Secara ringkasnya, kajian ini membangunkan satu sistem yang dapat mengekstrak data dari Twitter mengenai ulasan filem atau drama bersiri Netflix dan menganalisis sentimen data tersebut di samping mengelaskan kandungan yang didapati negatif kepada 3 kategori yang dikelaskan terlebih dahulu. Data Twitter API dikumpulkan dalam bentuk fail csv dan data tersebut akan dibersihkan dan dilabelkan terlebih dahulu dalam pra-pemrosesan dan pengekstrakan ciri. Teknik yang digunakan bagi kajian ini adalah pembelajaran mesin. Bagi proses analisis sentimen, *TextBlob Sentiment Analysis* digunakan kerana alat tersebut sesuai digunakan untuk dataset daripada media sosial untuk mengenalpasti cicipan yang bernada positif dan negatif. Kemudian, *Naive Bayes Classifier* digunakan untuk mengelaskan cicipan-cicipan yang negatif kepada 3 kelas kategori yang bersesuaian.

Hasil output bagi kajian ini telah dapat memenuhi objektif kajian ini sepenuhnya di samping dapat divisualisasikan dalam bentuk yang tersusun seperti graf dan carta pai. Bagi memenuhi objektif yang pertama, dirujuk rajah 5 untuk output analisis sentimen bagi mengenal pasti cicipan yang bernada positif, negatif dan neutral. Bagi objektif kedua, dirujuk rajah 7 untuk output pengelas bagi mengenal pasti kandungan Netflix yang dianalisis adalah sesuai atau tidak untuk ditonton oleh masyarakat Malaysia mengikut kategori keagamaan, kebudayaan dan nilai moral.

Kajian ini berdasarkan teknik sentimen analisis dan pengelas. Pembangunan sistem ini diharap dapat membantu meningkatkan lagi pengetahuan dalam pembelajaran mesin yang semakin banyak digunakan oleh orang ramai dan industri besar-besaran pada masa kini. Kajian ini juga diharap untuk meningkatkan kemahiran menggunakan bahasa pengaturcaraan Python. Selain itu, mungkin kajian lain pada masa hadapan boleh menambahbaik algoritma kajian ini untuk menjadi sebuah sistem atau aplikasi yang boleh mencapai objektif yang sama. Hal ini kerana, sistem atau aplikasi seperti ini, kalau dibangunkan, akan mendapat sambutan yang tinggi terutamanya dari sektor hiburan. Oleh itu, dapatlah menyumbang tenaga untuk pengembangan analisis teks dalam komuniti sains data serta industri pemasaran dan hiburan.

**7 RUJUKAN**

- Tay Fui Kien. 2019. *Analisis Sentimen Twitter Mengenai Peristiwa-Persitiwa penting yang Berlaku di Sekitar UKM, Bangi*. Bangi: Universiti Kebangsaan Malaysia.
- Lembaga Penapis Filem. 2010. *Garis Panduan Penapisan Filem*. W.P Putrajaya: Bahagian Kawalan Penapisan Filem dan Penguatkuasaan Kementerian Dalam Negeri.
- Wan Amizah Wan Mahmud, Chang Peng Kee & Jamaluddin Aziz. 2009. *Film Censorship in Malaysia: Sanctions of Religious, Cultural and Moral Values*. Jurnal Komunikasi, Malaysian Journal of Communication Vol, 25: 42-49.
- Lina L. Dhande & Dr. Prof. Girish K. Patnaik. 2014. *Analyzing Sentiment of Movie Review Data using Naive Bayes Neural Classifier*. International Journal of Emerging Trends & Technology in Computer Science, Vol 3, Issue 4.
- Palak Baid, Apoorva Gupta & Neelam Chaplot. 2017. *Sentiment Analysis of Movie Reviews using Machine Learning Techniques*. International Journal of Computer Applications. Vol 179, No.7.
- Kuat Yessenov & Sasa Misailovic. 2009. *Sentiment Analysis of Movie Review Comments*, 6.863 Spring 2009 final project.
- Rucha Jadhavar & Agastya Kumar Komarraju. 2018. *Sentiment Analysis of Netflix and Competitor Tweets to Classify Customer Opinions*. Paper 2708-2018.
- Lum Choi Kian. 2019. *Analisis Sentimen Dalam Bitcoin Tweets*. Bangi: Universiti Kebangsaan Malaysia.
- Ahmad Imam Fahrur Rozi, Elok Nur Hamdana & Muhammad Balya Iqbal Alfahmi. 2018. *Pengembangan Aplikasi analisis Sentimen Twitter Menggunakan Metode Naive Bayes Classifier (Studi Kasus SAMSAT Kota Malang*. Jurnal Informatika Polinema 4(2):149.
- Asma Md. Isa, Wan Amizah Wan Mahmud, Wan Idros Wan Sulaiman & Muhammad Adnan Pitchan. 2019. *Netflix and Dilemma of Content Regulation in Malaysia*. International Journal of Advanced Science and Technology, Vol. 28, No. 16.
- Jiena G. McLellan. 2017. *Sentiment Analysis of Real-Time Twitter Data*. The Beef Cattle Institute, Kansas State University.

- Alexandra Balahur. 2013. *Sentiment Analysis in Social Media Texts*. European Commission Joint Research Centre, Italy.
- Ahmad Faezal. 2020. Filem Cuties: Netflix Cuba Normalisasikan Pedophilia?. *The Patriots*, 14 September. <https://www.thepatriots.asia/filem-cuties-netflix-cuba-normalisasikanpedophili a/>
- Afiezy Azaman. 2019. Penapisan kandungan Netflix: Bincang dalam-dalam agar hak pengguna tidak terjejas. *Astro Awani*, 23 November. <https://www.astroawani.com/beritamalaysia/penapisan-kandungan-netflix-bincang-dalamdalam-agar-hak-pengguna-tidakterjejas-223455>
- Norhafzan Jaafar. 2019. Kawal kandungan program siaran Netflix elak kebimbangan ibu bapa. *Berita Harian Online*, 16 November. <https://www.bharian.com.my/berita/nasional/2019/11/629027/kawal-kandunga nprogram-siaran-netflix-elak-kebimbangan-ibu-bapa>
- Ida Lim. 2020. MCO: Malaysians stuck at home give Netflix, Tonton biggest traffic growth rates against competitors. *Malay Mail*, 1 April. <https://www.malaymail.com/news/malaysia/2020/04/01/mco-malaysians-stuck-at-homegive-netflix-tonton-biggest-traffic-growth-rat/1852432>
- Rahimi Yunus. 2020. Increasing streaming subscribers signals viewing preferences. *The Malaysian Reserve*, 3 August. <https://themalaysianreserve.com/2020/08/03/increasingstreaming-subscriberssignals-viewing-preferences/>
- Nur Haziqah A Malek. 2020. Malaysians turn to social media, streaming services to fill time during MCO. *The Malaysian Reserve*, 6 April. <https://themalaysianreserve.com/2020/04/06/malaysians-turn-to-social-mediastreamingservices-to-fill-time-during-mco/>
- Bernamea. 2019. Finas calls for law to regulate contents on Netflix and VOD. *Free Malaysia Today*, 16 November. <https://www.freemalaysiatoday.com/category/nation/2019/11/16/finas-calls-fo r-law-toregulate-contents-on-netflix-and-vod/>
- Mix. 2020. Here are the 9 titles Netflix purged due to government censorship. *The next Web*, 7 February. <https://thenextweb.com/insights/2020/02/07/netflix-government-censorship/>
- Harris Andrea. 2020. What are the Advantages and Disadvantages of Netflix?. *Tech 21 Century*. <https://www.tech21century.com/pros-and-cons-of-netflix/>



Julius. 2019. Sekatan Dalam Menonton Filem. *Kata Malaysia*, 30 Oktober.

<https://katamalaysia.my/social/sekatan-dalam-menonton-filem/>

Artem Oppermann. 25 February 2019. Sentiment Analysis with Deep Learning of Netflix Reviews. *towardsdatascience*. <https://towardsdatascience.com/sentiment-analysis-with-deep-learning-62d4d0166ef6>

1 Oktober 2019. Pra-Pemrosesan Tweet untuk Analisis Sentimen. *Ichi.Pro*.  
<https://ichi.pro/id/pra-pemrosesan-tweet-untuk-analisis-sentimen-183953141635390>

Shubham Jain. 2018. Ultimate guide to deal with Text Data (using Python) – for Data Scientists and Engineers. *Analytics Vidhya*, 27 Februari.  
<https://www.analyticsvidhya.com/blog/2018/02/the-different-methods-deal-text-datapredictive-python/>

Deepak Das. 2019. Social Media Sentiment Analysis using Machine Learning : Part — I. *towardsdatascience*, 7 September.  
<https://towardsdatascience.com/social-mediasentiment-analysis-49b395771197>

Avinash Navlani. 2018. Naive Bayes Classification using Scikit-learn?. *datacamp*, 4 December.  
<https://www.datacamp.com/community/tutorials/naive-bayes-scikit-learn>

AFAF5. (2021, Jan 6). Analyticsvidhya. Retrieved from Sentiment Analysis: VADER or TextBlob?: <https://www.analyticsvidhya.com/blog/2021/01/sentiment-analysis-vader-or-textblob/>

Baris Hasdemir. (2020, Aug 16). Retrieved from Sentiment analysis on the tweets about distance learning with TextBlob: <https://towardsdatascience.com/sentiment-analysis-on-the-tweets-about-distance-learning-with-textblob-cc73702b48bc>

Netflix. 10 2016. <https://ms.wikipedia.org/wiki/Netflix>