



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk mengubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB III

METODOLOGI PENELITIAN

3.1 Objek Penelitian

Objek penelitian pada penelitian ini adalah untuk melakukan perbandingan analisis sentimen masyarakat terhadap perusahaan *multiplatform* milik Polytron, yaitu MolaTV. Sebagai perusahaan layanan konten digital yang tergolong baru dan sedang berkembang pesat, MolaTV memiliki tanggung jawab yang besar terhadap kepuasan para pelanggannya.

Penelitian ini memanfaatkan toko aplikasi Google Playstore yang berisikan berbagai ulasan dari para pengguna MolaTV sebagai sumber data dan akan dilakukan analisis sentimen terhadap layanan MolaTV, pengambilan data menggunakan Google Playstore karena MolaTV tersedia di Google Playstore sejak awal dirilis pada tahun 2019 lalu, maka penelitian ini akan berfokus pada hasil dari perbandingan ketiga algoritma untuk menganalisis sentimen masyarakat terhadap MolaTV.

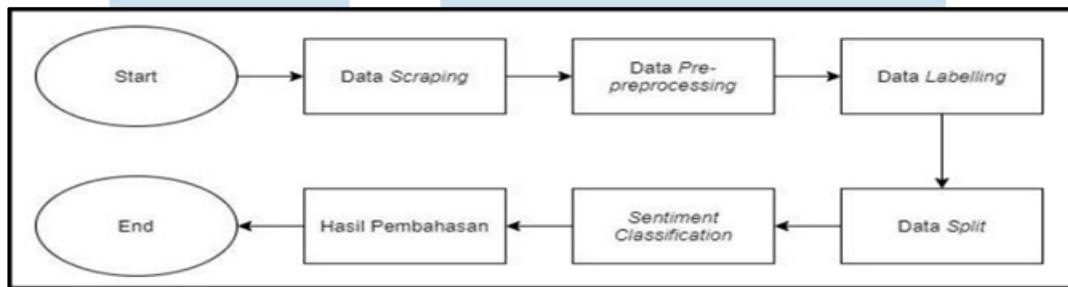
3.2 Teknik Pengumpulan Data

Pengumpulan data dilakukan dengan menggunakan Bahasa pemrograman Python melalui aplikasi Jupyter Notebook dan navigator Anaconda yang dapat mengambil data ulasan yang terdapat pada Google Playstore atau biasa disebut dengan *data scraping*. Dengan fitur Google-play-scraper dan Menggunakan code Python pada JupyterNotebook, maka didapatkan jumlah *review* sebanyak 8480 data, yang kemudian akan diseleksi Kembali sesuai dengan rentang waktu yang

diinginkan yaitu 1 tahun (dari 1 desember 2020 hingga 1 desember 2021), sehingga di dapatkan data sebanyak 4576 data.

3.3 Alur Penelitian

Penelitian terhadap MolaTV didasari oleh diagram alur penelitian yang terdapat pada gambar 3.1. Diagram alur berisikan berbagai proses yang terdapat pada penelitian ini dilihat pada gambar berikut:



Gambar 3. 1 Diagram Alur Penelitian

3.3.1 Data Scraping

Pengumpulan data dilakukan dengan menggunakan Bahasa pemrograman Python melalui aplikasi Jupyter Notebook dan navigator Anaconda yang dapat mengambil data ulasan yang terdapat pada Google Playstore atau biasa disebut dengan data scraping. Fitur pada Python yang digunakan pada pengumpulan data adalah Google-play-scraper.

Data yang didapatkan lalu diseleksi sesuai dengan rentang waktu yang diinginkan dan disimpan menjadi dataset dalam format CSV(Comma Separated Values) dan disimpan kedalam folder yang diinginkan untuk kemudian dipakai pada proses selanjutnya.

3.3.2 Data Pre-Processing

Tahapan *Pre-Processing* merupakan langkah awal yang dilakukan pada melakukan *Text Mining* untuk melakukan perubahan terhadap data yang didapatkan agar menjadi data yang baik dan bersih untuk di analisis. Tahapan ini dilakukan dengan menggunakan R Studio sebagai *tools*. Terdapat beberapa langkah untuk melakukan *Pre-Processing*:

- 1) *Remove Duplicates* = Untuk menghilangkan data yang sama/kembar sehingga dilakukan penghapusan data kembar.
- 2) *Transform Cases* = Mengubah huruf besar menjadi huruf kecil untuk menghilangkan tulisan dari huruf besar.
- 3) *Remove Punctuation* = Membersihkan tulisan dari simbol dan karakter yang bukan termasuk dalam alphabet.
- 4) *Remove URL* = Karena tidak memiliki arti didalam Analisa, maka URL akan dihapus dari tulisan.
- 5) *Remove Emoticon* = emoticon yang terdapat pada data akan dihapus.
- 6) *Tokenize* = text akan dipecah menjadi bagian-bagian kecil berdasarkan kata pada ulasan.
- 7) *Remove Digit* = Berfungsi untuk menghilangkan angka yang terdapat pada data.

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

3.3.3 Data Labeling

Setelah melakukan tahap pembersihan data atau *pre-processing* pada RStudio selanjutnya data sudah bersih akan diberikan label, data labeling diperlukan untuk memberikan label positif atau negatif terhadap tiap-tiap data.

Pada penelitian ini, pelabelan positif negatif data MolaTV dilakukan dengan menggunakan *software* RapidMiner. Dengan menggunakan operator yang terdapat di rapid miner yang bernama *Extract Sentiment* pelabelan jg menggunakan operator sentiwordnet yang berfungsi sebagai basis data yang diciptakan untuk mengklasifikasikan sentimen pada saat melakukan *opinion mining*.

Pemberian label menggunakan sentiwordnet akan menghasilkan angka yang berfungsi sebagai nilai dari sentimen yang terkandung pada kalimat atau tulisan, angka ini nantinya sebagai indikator nilai label positif, negatif maupun netral.

3.3.4 Data Split

Data *split* diperlukan untuk mengevaluasi performa model *machine learning*. Metode ini bekerja dengan membagi *dataset* menjadi data *training* dan data *testing* sesuai dengan proposi yang diinginkan (biasanya 67%:33%). *Random sampling* memilih dan mengambil beberapa *unit sample* yang ditentukan besarnya peluang suatu *sample* untuk terpilih kedalam *sample*.

Teknik *simple random sampling* adalah memilih secara acak *sample* yang akan dipilih. Biasanya data akan dipisah menjadi data *training* dan data *testing*, masing-masing bernilai 67% dan 33%.

3.3.5 Sentiment Classification

Dengan menggunakan ketiga metode *Naïve Bayes*, *Decision Tree*, dan *Support Vector Machine* pada ulasan yang terdapat pada toko aplikasi GooglePlaystore, proses *sentiment classification* dilakukan sebagai perbandingan ketida metode tersebut. Tahapan ini dilakukan untuk mengetahui hasil dari perbandingan ketiga metode dan menarik kesimpulan, metode mana yang memiliki tingkat akurasi paling tinggi untuk melakukan klasifikasi pada penelitian ini.

operator *cross validation* yang terdapat didalam aplikasi Rapid Miner dipakai untuk melakukan proses *sentiment classification*. *Cross Validation* adalah metode *text mining* yang berfungsi untuk membantu mendapatkan hasil akurasi yang lebih baik. *Cross validation* juga adalah sebuah Teknik validasi *model split validation* yang melakukan validasi dengan cara melakukan pengukuran *training error* terhadap data *test* yang diuji. *K-fold cross validation* bekerja dengan menggunakan parameter yang sama, melakukan proses pada data sebanyak k kali.

Biasanya model pengujian *Cross validation* dilakukan dengan menerapkan 10-cross. Hal ini adalah standar karena mampu melakukan validasi yang lebih efisien dan praktis dan dapat meningkatkan nilai performanya. *Operator Cross*

validation juga terdapat di dalam aplikasi RapidMiner, operator ini menggunakan perhitungan jumlah *folds* yang digunakan untuk memberikan nilai K.

Konsepnya, *Cross validation* akan bekerja dengan menilai kinerja/fungsi dari proses sebuah metode algoritma dengan membagi *sample* data secara acak dan dikelompokan sebanyak nilai K.

3.3.6 Hasil Penelitian

Setelah proses klasifikasi sentimen, hasil klasifikasi tiap metode akan dievaluasi menggunakan pengujian *confusion matrix* yang memiliki parameter *accuracy*, *precision* dan *recall*. Dari hasil pengujian *confusion matrix*, maka didapatkan hasil untuk evaluasi, dari ketiga metode ini yang mana yang memiliki akurasi yang paling baik pada penelitian ini.

Dari hasil penelitian, nilai rasio prediksi yang mendekati hasil sebenarnya ditentukan dari nilai *accuracy*, nilai yang memiliki kesamaan dan ketepatan proses klasifikasi ditunjukkan pada nilai *precision* dan nilai yang menunjukkan rasio klasifikasi benar positif dibandingkan dengan keseluruhan dokumen saat klasifikasi adalah *recall*.

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A