

# **SOLUTION DOCUMENT**

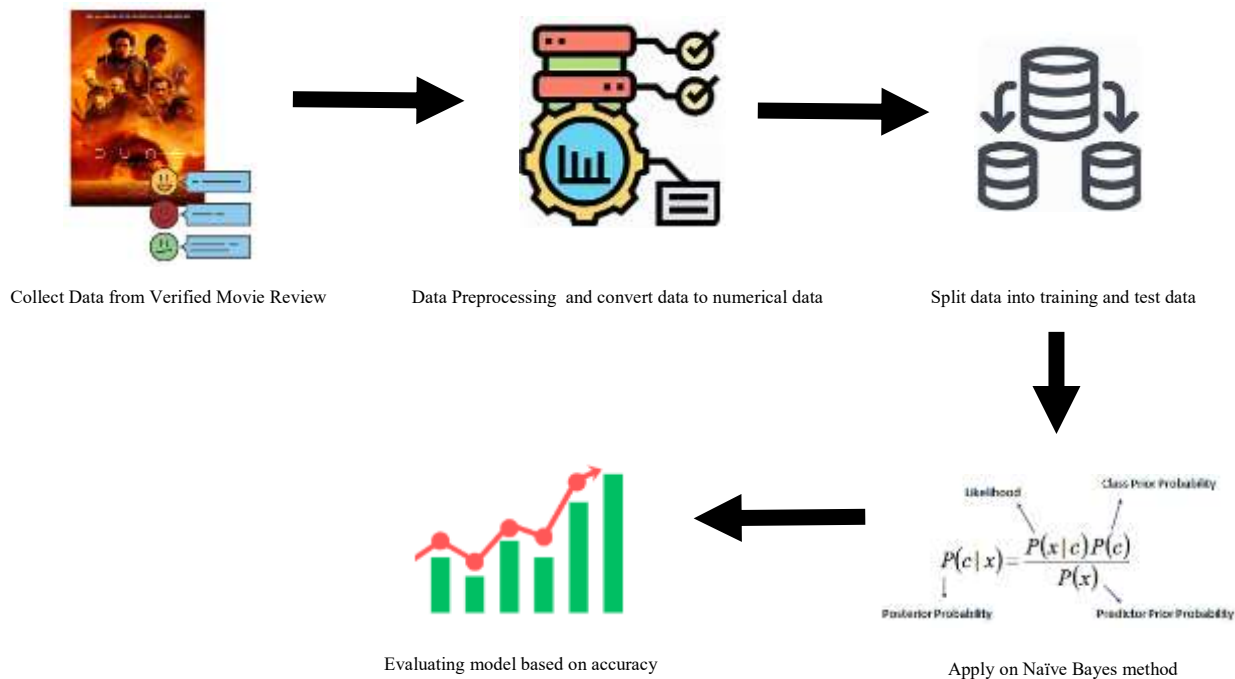
## **Sentiment Analysis on Newly Released Movie Study Case “Dune 2”**



By :

I Nengah Dharma Pradnyandita

# 1. Workflow of Sentiment Analysis Solution



## 1.1 Collect Data from verified movie review

Data diperoleh dari Scrapping website Rotten Tomatoes pada bagian Audience Review. Data yang di scrapping sebanyak 510 baris. Data terdiri dari enam kolom seperti yang terlihat pada gambar 1.

review-data href	audience-reviews__name	audience-reviews__name href	audience-reviews__duration	audience-reviews__review	Sentiment
<a href="https://www.rottentomatoes.com/m/dune-2">https://www.rottentomatoes.com/m/dune-2</a>	Scott B	<a href="https://www.rottentomatoes.com/m/dune-2">https://www.rottentomatoes.com/m/dune-2</a>	Apr 25, 2024	This movie is a stunning, visu	1
<a href="https://www.rottentomatoes.com/m/dune-2">https://www.rottentomatoes.com/m/dune-2</a>	alan t	<a href="https://www.rottentomatoes.com/m/dune-2">https://www.rottentomatoes.com/m/dune-2</a>	Apr 25, 2024	Seen all this before. Star War	0
<a href="https://www.rottentomatoes.com/m/dune-2">https://www.rottentomatoes.com/m/dune-2</a>	Drew F	<a href="https://www.rottentomatoes.com/m/dune-2">https://www.rottentomatoes.com/m/dune-2</a>	Apr 25, 2024	Zpoiler Alert: Wonka helps sa	1
<a href="https://www.rottentomatoes.com/m/dune-2">https://www.rottentomatoes.com/m/dune-2</a>	Campolongo C	<a href="https://www.rottentomatoes.com/m/dune-2">https://www.rottentomatoes.com/m/dune-2</a>	Apr 25, 2024	Question: Is Denis Villeneuve	0
<a href="https://www.rottentomatoes.com/m/dune-2">https://www.rottentomatoes.com/m/dune-2</a>	Nicholas S	<a href="https://www.rottentomatoes.com/m/dune-2">https://www.rottentomatoes.com/m/dune-2</a>	Apr 24, 2024	Honestly I enjoyed the first m	1
<a href="https://www.rottentomatoes.com/m/dune-2">https://www.rottentomatoes.com/m/dune-2</a>	Jamey S	<a href="https://www.rottentomatoes.com/m/dune-2">https://www.rottentomatoes.com/m/dune-2</a>	Apr 24, 2024	If you actually read the books	0

Gambar 1 Table of Data from Dune 2 Movie reviews Rotten Tomatos

Kolom Sentiment berfungsi sebagai label untuk setiap review dengan keterangan “1” berarti positif , “0” berarti negative dan “2” berarti netral.

## 1.2 Data Preprocessing convert data to numerical data



Gambar 2 Data Preprocessing to cleaning the text

Pada praproses data dilakukan beberapa langkah

1. Hilangkan katakter yang tidak diperlukan seperti simbol.
2. Ubah seluruh kata menjadi huruf kecil, hal ini dilakukan agar mengurangi variasi dalam teks.
3. Tokenisasi setiap review, dimana teks dibagi menjadi unit yang lebih kecil yaitu kata.
4. Hilangkan stopwords yang sering muncul dalam teks dan tidak terlalu berpengaruh dalam pemrosesan,
5. Stemming setiap kata agar mengurangi variasi dari kata.
6. Susun kembali setiap kata menjadi sebuah kalimat.

## 1.3 Convert review text into numerical data



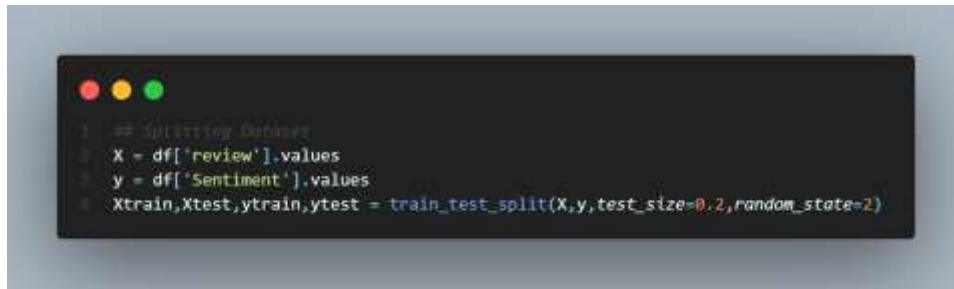
Gambar 3 Convert Review Text into Numerical Data

Pada Proses Vectorizer data setiap review diubah menjadi numerik dengan menggunakan TfidfVectorizer, metode ini digunakan karena data yang akan diimplementasikan adalah kalimat, sehingga digunakan metode ini untuk mengubah kalimat menjadi vector angka yang dapat diproses oleh model. Penggunaan metode ini juga karena setiap kata akan diidentifikasi berdasarkan tingkat kepentingannya, sehingga lebih mudah dalam idenfikasi Sentimen. Hasil dari implementasi pada data seperti pada Gambar 4.

(0, 519)	0.1486252548638906
(0, 1513)	0.17049244489902232
(0, 2291)	0.2601638907180995
(0, 1965)	0.2475024982888377
(0, 274)	0.18660781159600848

Gambar 4 Result of TFidfVectorizer

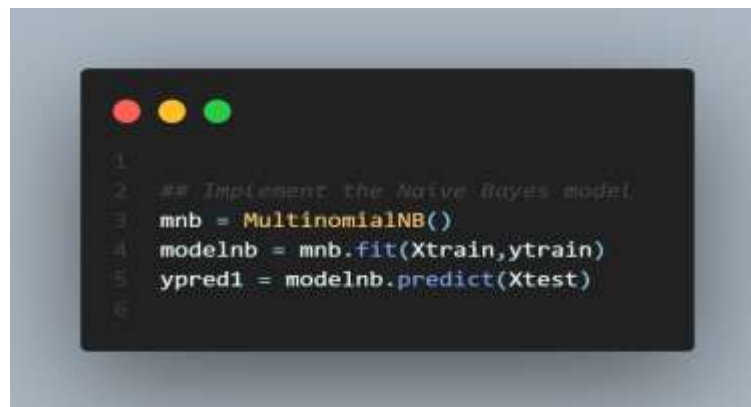
## 1.4 Split data into training and testing data



*Gambar 5 Splitting data into Training and testing*

Pada Proses ini data akan dibagi menjadi dua bagian, yaitu data Training dan data Testing, Data Training digunakan untuk melatih model dan data testing digunakan untuk menguji model. Persentase data Training adalah 80% dari total keseluruhan data, sedangkan data Testing 20% dari total keseluruhan data.

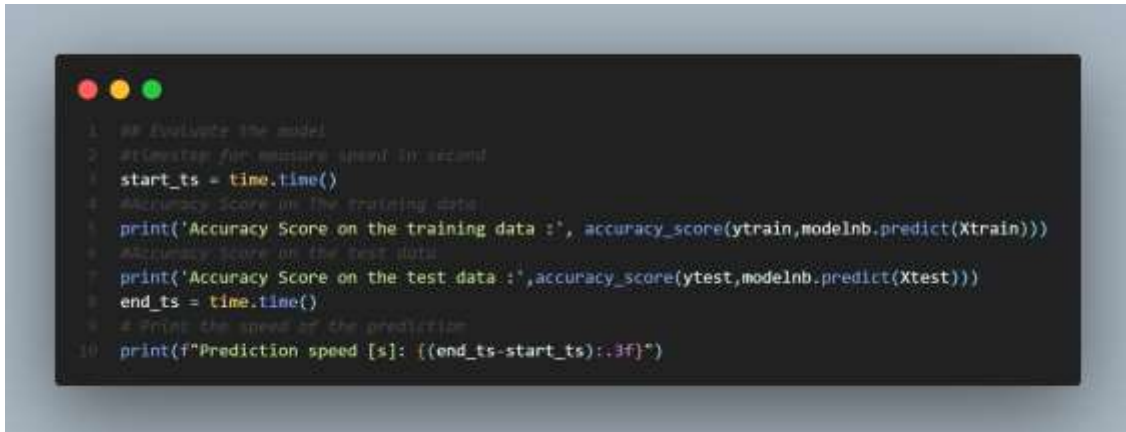
## 1.5 Implement Naïve Bayes Model to predict the data



*Gambar 6 Implement the data into Naïve Bayes Model*

Model Machine Learning yang digunakan untuk mengimplementasikan data sentiment adalah Naïve Bayes. Pemilihan model didasarkan pada akurasi dan tingkat kecocokan data dengan model. Model Naïve Bayes mengklasifikasi data berdasarkan probabilitas kecocokan antara data baru dengan data latih. Dibandingkan model Machine Learning yang lain, Naïve Bayes sangat sesuai dengan jenis data Catagorical, oleh sebab itu model ini dapat mencegah terjadinya overfitting. Pada kasus ini, digunakan Multiominal Naïve Bayes, dikarenakan data yang digunakan berbentuk vector.

## 1.6 Evaluate model based on Accuracy



*Gambar 7 Evaluate Model Based On Accuracy*

Pengukuran Akurasi digunakan untuk mengevaluasi model yang telah dilatih dengan data Sentimen. Hasil Akurasi model yang diperoleh adalah 75%, hal ini membuktikan model cukup baik bekerja pada data review movie. Sedangkan kecepatan dalam pemrosesan data adalah 0.069 detik. Jumlah data yang sedikit berpengaruh pada kecepatan pemrosesan model, semakin besar data semakin lama prosesnya.

```
Accuracy Score on the training data : 0.8088235294117647
Accuracy Score on the test data : 0.7549019607843137
Prediction speed [s]: 0.069
```

*Gambar 8 The result of Model Accuracy*

## 2. Next Steps to Improve the solution

Langkah selanjutnya yang bisa dilakukan untuk meningkatkan solusi adalah pemilihan parameter yang lebih sesuai dengan menggunakan Grid Search. Selain itu data yang digunakan dapat lebih diperbanyak sehingga model memiliki lebih banyak referensi dalam memprediksi data sentimen. Model yang digunakan dapat divariasikan dengan menggunakan Algoritma Optimasi.