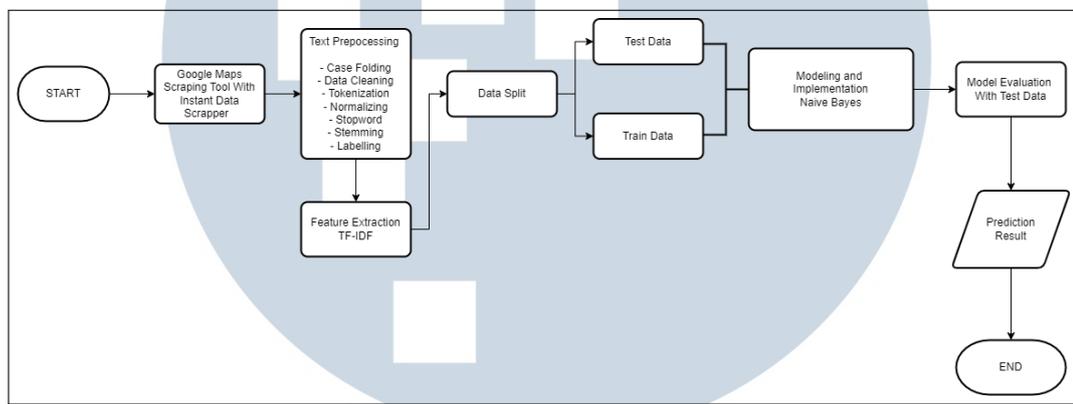


BAB 3 METODOLOGI PENELITIAN

3.1 Gambaran Alur Penelitian

Bab ini berisikan bagaimana alur untuk melakukan penelitian dari awal hingga selesai. Gambar 3.1 merupakan alur untuk melakukan penelitian yang disusun berbentuk diagram.



Gambar 3.1. Gambar Diagram Alur Penelitian

3.2 Metode Penelitian

Dalam penelitian ini dilakukan implementasi *Naive Bayes Classifier* untuk melakukan analisis sentimen terhadap tempat wisata kuliner Pasar Lama Tangerang dengan beberapa Metode Penelitian sebagai berikut:

3.2.1 Studi Literatur

Hal yang digunakan dilakukan untuk mencari informasi yang mendukung penelitian. Dengan melakukan pembelajaran dan mencari teori yang dibutuhkan seperti mendapatkan data melalui ulasan *Google Maps*, *text mining*, *Naive Bayes Classifier*, TF-IDF dan *confusion matrix*. Ulasan yang dikumpulkan berisi tentang komentar mengenai Pasar Lama Tangerang. Pengambilan data pada *Google Maps* menggunakan aplikasi yang bernama *Instan Data Scraper*.

3.2.2 Perancangan Sistem

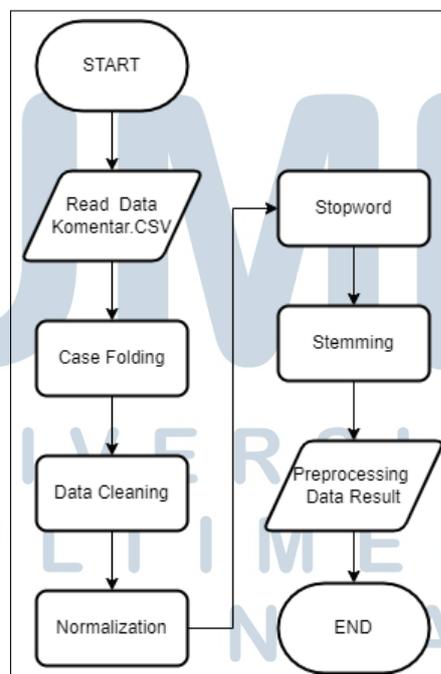
Pada tahap ini, dilakukan proses untuk mengklasifikasikan data yang sudah didapatkan dari *Google Maps*. Berikut beberapa tahapan yang diperlukan untuk melakukan analisis sentimen.

1. Pengumpulan Data

Dataset dikumpulkan dari aplikasi *Google Maps* menggunakan *scraping tool*, yaitu *Instant Data Scraper*. Data yang diambil mengenai informasi tempat wisata kuliner Pasar Lama Tangerang yang berupa komentar orang-orang yang berada pada aplikasi *Google Maps*. Data yang diambil adalah komentar berbahasa Indonesia sebanyak 1.500 data dimana data tersebut diambil dari tanggal 1 November 2022 sampai dengan 30 April 2023.

2. Text Preprocessing

Dataset yang telah dihimpun akan melalui beberapa tahapan proses untuk dapat digunakan dalam klasifikasi sentimen. Tahapan proses ini disebut dengan *data preprocessing*. Pada Gambar 3.2 merupakan tahapan dari *preprocessing*.



Gambar 3.2. Flowchart *Preprocessing*

(a) *Case Folding*

Case Folding adalah tahap penyeragaman huruf pada dataset menjadi huruf kecil dan menghilangkan karakter selain huruf 'a' sampai 'z', termasuk menghilangkan angka dan tanda baca. Tabel 3.1 ini menunjukkan hasil data teks yang telah dilakukan *case folding*.

Tabel 3.1. *Case Folding*

Sebelum dilakukan <i>case folding</i>	Setelah dilakukan <i>case folding</i>
Aku anak Univeristas Multimedia Nusantara yng lahir di tahun 2003 dan Aku tidak terikat Organisasi apapun di Kampus #AngkatanTua https://t.co/f3wuxtVHec	aku anak universitas multimedia nusantara yng lahir di tahun 2003 dan aku tidak terikat organisasi apapun di kampus #angkatanua https://t.co/f3wuxtVHec

(b) *Cleaning*

Cleaning merupakan tahap untuk menghapus *hashtag*, *mention*, nama pengguna dan *url* pada dataset yang telah diperoleh.

(c) *Tokenizing*

Tokenization adalah tahap untuk memisahkan sebuah kalimat menjadi pecahan yang lebih kecil atau per kata. Tabel 3.2 ini menunjukkan hasil data teks yang telah dilakukan *tokenizing*.

Tabel 3.2. *Tokenizing*

Sebelum dilakukan <i>Tokenizing</i>	Setelah dilakukan <i>Tokenizing</i>
Aku anak Univeristas Multimedia Nusantara yng lahir di tahun 2003 dan Aku tidak terikat Organisasi apapun di Kampus #AngkatanTua https://t.co/f3wuxtVHec	'aku', 'anak', 'universitas', 'multimedia', 'nusantara', 'yng', 'lahir', 'di', 'tahun', '2003', 'dan', 'aku', 'tidak', 'terikat', 'organisasi', 'apapun', 'di', 'kampus'

(d) *Normalization*

Normalization ialah tahap mengubah bentuk kata tidak baku pada dataset menjadi kata baku. Dalam tahap ini akan berpedoman pada Kamus Besar Bahasa Indonesia (KBBI). Tabel 3.3 ini menunjukkan hasil data teks yang telah dilakukan *Normalization*.

Tabel 3.3. *Normalization*

Sebelum dilakukan <i>Normalization</i>	Sesudah dilakukan <i>Normalization</i>
dan gue baru tau tempat makan itu buka siang banget sekitar jam 12-an. normalnya kedai ² gitu buka jam 8-an ga sih...	dan saya baru tahu kedai itu buka siang sekali sekitar lewat dari pukul 12.00 normal kedai begitu buka dari pukul 8.00 tidak sih

(e) *Stopwords*

Stopwords berfungsi untuk menghapuskan kata-kata umum dengan frekuensi kemunculan tinggi. Untuk penelitian ini menggunakan *filter stopwords* bahasa Indonesia dengan *library* Sastrawi. Proses ini bertujuan untuk mengurangi kata-kata yang tidak diperlukan dalam perhitungan akurasi dan model.

(f) *Stemming*

Stemming berfungsi untuk menghapus imbuhan pada tiap kata dan mengubah setiap kata menjadi kata dasar. Contoh kata imbuhan seperti 'menghubungkan' lalu dilakukan proses *stemming*, maka akan menjadi 'hubung' dimana terdapat awalan 'meng-' dan akhiran '-kan'.

(g) *Labelling Data*

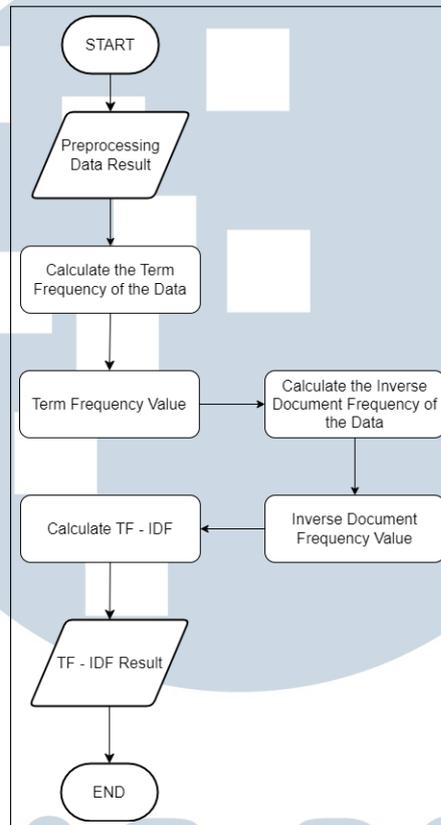
Labelling Data merupakan tahap untuk mengelompokan *tweets* ke dalam kelas positif, negatif. Tabel 3.4 menunjukkan hasil data teks yang telah dilakukan *labelling*.

Tabel 3.4. *Labelling*

<i>clean text</i>	Kelas Sentimen
Tempat wisata yang terlalu ramai dan juga banyak sampah dimana mana	negatif
Tempat wisata yang sangat bagus dan apik	positif

3. Menerapkan Fitur TF-IDF *Term Frequency - Inverse Document Frequency (TF-IDF)* merupakan salah satu metode yang digunakan untuk mengekstraksi ciri dari suatu text selain itu juga memberikan nilai bobot pada setiap kata

yang dihitung berdasarkan banyaknya kata yang bermunculan pada dokumen dari hasil *preprocessing*. Pada Gambar 4.16 merupakan jalannya proses TF-IDF.

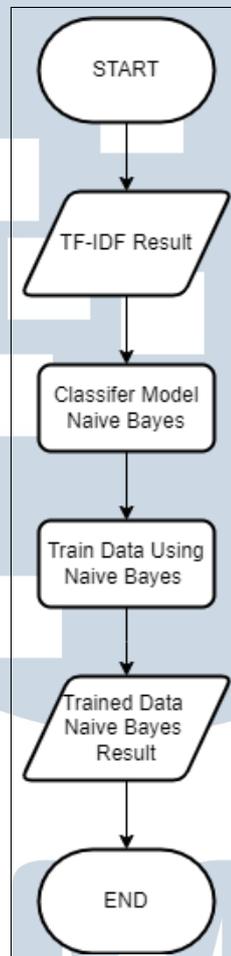


Gambar 3.3. Flowchart menerapkan TF-IDF

Langkah pertama pastikan sudah melakukan langkah *preprocessing*, kemudian menghitung nilai TF terlebih dahulu untuk menentukan frekuensi kemunculan kata. Setelah melakukan perhitungan TF, maka dilakukan perhitungan IDF. Setelah menghitung TF dan IDF dan menggabungkan keduanya, itu akan menampilkan nilai TF-IDF dengan sendirinya.

4. Menerapkan *Naïve Bayes Classifier* Setelah melakukan langkah-langkah TF-IDF, langkah selanjutnya adalah menerapkan model *Naïve Bayes Classifier* menggunakan hasil dari data proses TF-IDF yang sebelumnya telah dibagi menjadi *train* dan *test data*. nilai yang diperoleh dari proses sebelumnya diterapkan ke *training data*, lalu dilatih untuk mengklasifikasikan label pada data komentar. Langkah dimana model *Naïve Bayes Classifier* akan dilatih untuk melakukan klasifikasi menggunakan *training data* sehingga model ini

dapat digunakan untuk memprediksi *test data*. Pada Gambar 3.4 merupakan jalannya proses penerapan *Naïve Bayes Classifier*:



Gambar 3.4. Flowchart menerapkan Naïve Bayes Classifier

5. Evaluasi Model

Hasil yang didapatkan setelah melakukan implementasi algoritma *Naïve Bayes Classifier* hasilnya digunakan untuk melakukan validasi dan mencari parameter terbaik dalam *confusion matrix* pada algoritma *Naïve Bayes Classifier* dalam melakukan klasifikasi data komentar yang tidak terstruktur.