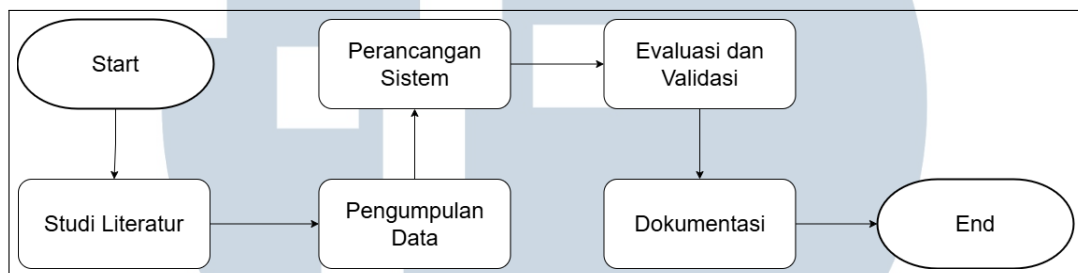


BAB 3 METODOLOGI PENELITIAN

3.1 Metodologi Penelitian

Pada gambar 3.1 merupakan sebuah flowchart atau langkah-langkah dari metodologi penelitian yang dilakukan.



Gambar 3.1. *Flowchart* Metodologi Penelitian

3.1.1 Studi Literatur

Sebelum dilakukannya penelitian, langkah yang harus dilakukan adalah melakukan studi literatur untuk mendapatkan berbagai informasi mengenai subjek yang akan diteliti. Dalam konteks penelitian ini, fokus studi literatur adalah analisis sentimen pengguna Twitter terhadap perdagangan karbon di Indonesia menggunakan algoritma *Naive Bayes*. Sumber informasi yang dikumpulkan melibatkan kajian jurnal ilmiah, karya ilmiah, artikel, dan buku yang berkaitan dengan dua aspek utama penelitian, yaitu perdagangan karbon dan analisis sentimen menggunakan algoritma *Naive Bayes*. Selain itu terdapat juga studi literatur pada jurnal penelitian terdahulu mengenai analisis sentimen menggunakan algoritma *Naive Bayes* pada tabel 3.1.

UNIVERSITAS
MULTIMEDIA
NUSANTARA

Tabel 3.1. Tabel Studi Literatur Analisis Sentimen Menggunakan Algoritma Naive Bayes

Penulis	Judul	Hasil
Adittia Agustian, Tukino, dan Fitria Nurapriani	Penerapan Analisis Sentimen Dan <i>Naive Bayes</i> Terhadap Opini Penggunaan Kendaraan Listrik Di Twitter	Pada penelitian ini <i>Naive Bayes</i> mendapatkan <i>accuracy</i> sebesar 80%, <i>precision</i> 82%, <i>F1-Score</i> 74%, dan <i>recall</i> 80%.
Muhammad Iqbal, Ade Davy Wiranata, Rayhan Suwito, Ridha Faiz Ananda	Perbandingan Algoritma <i>Naive Bayes</i> , <i>KNN</i> , dan <i>Decision Tree</i> terhadap Ulasan Aplikasi Threads dan Twitter	Algoritma <i>Naive Bayes</i> mendapatkan akurasi sebesar 66.41% pada aplikasi Threads dan 85.56% pada aplikasi Twitter, selain itu algoritma <i>Decision Tree</i> mendapatkan akurasi sebesar 72.78% pada aplikasi Twitter dan pada aplikasi Threads sebesar 65.62%, sedangkan algoritma <i>KNN</i> hanya mendapatkan 71.67% pada aplikasi Twitter dan pada aplikasi Threads 62.50%
Sifa Melina Salsabila, Aang Alim Murtopo, Nurul Fadhilah	Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode <i>Naive Bayes Classifier</i>	Algoritma <i>Naive Bayes</i> mendapatkan akurasi sebesar 85.10%

3.1.2 Pengumpulan Data

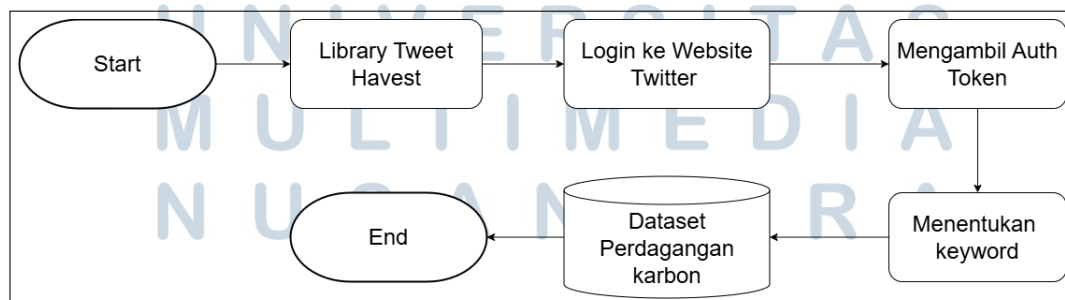
Pada tahap pengumpulan data, penelitian ini menggunakan Google Colab dan *library tweet harvest*. Proses penggunaan *tweet harvest* memerlukan *auth*

token, yang bisa diperoleh dari *cookies* pada browser. Untuk mendapatkan *auth token*, pertama-tama perlu membuka Twitter dan masuk ke akun menggunakan browser pada tampilan desktop seperti komputer atau laptop, karena fitur *inspect* tidak tersedia pada tampilan *mobile*. Setelah masuk ke akun Twitter, buka alat pengembang dengan mengklik kanan pada halaman Twitter dan memilih opsi "*Inspect*" atau dengan menekan *Ctrl+Shift+I* untuk *Windows* atau *Cmd+Opt+I* untuk *Mac*.

Setelah alat pengembang terbuka, navigasikan ke tab "*Application*". Di tab ini, berbagai informasi tentang aplikasi web, termasuk *cookies*, dapat dilihat. Di bawah bagian "*Storage*" di sisi kiri, klik "*Cookies*" dan pilih <https://x.com>. Di sini, cari item *auth token* di daftar *cookies* yang muncul dan salin nilai *auth token* tersebut.

Setelah mendapatkan *auth token*, kembali ke *Google Colab* dan masukkan *auth token* yang telah disalin ke dalam skrip. Langkah selanjutnya adalah menginstal *NodeJS*, *Pandas*, dan *library tweet harvest*. Setelah penginstalan selesai, gunakan *tweet harvest* untuk mengautentikasi permintaan pengumpulan data. Dengan *auth token* yang telah diperoleh, tweet yang relevan dengan topik perdagangan karbon di Indonesia dapat dikumpulkan.

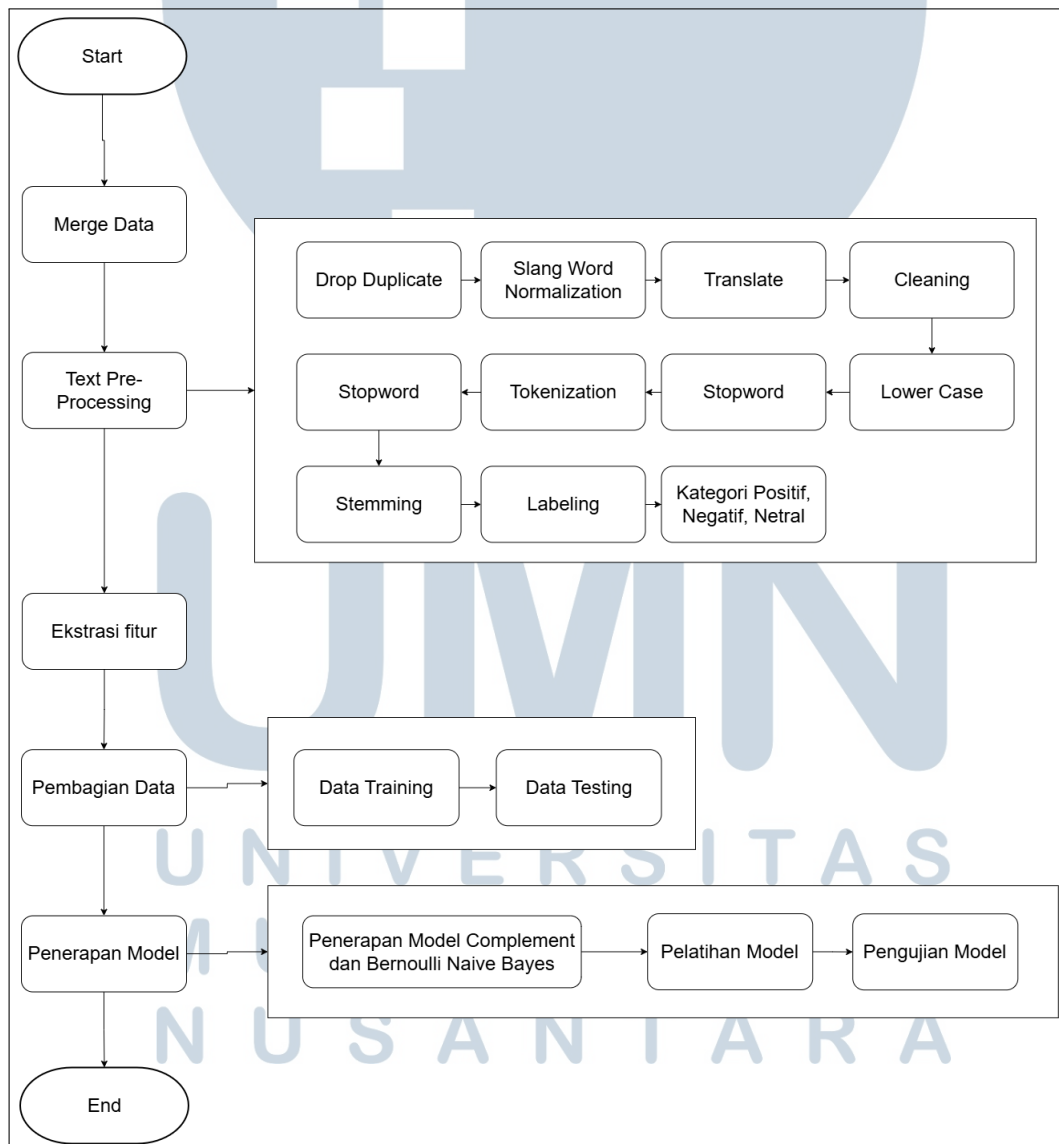
Kata kunci yang digunakan untuk pengumpulan data meliputi "Perdagangan Karbon", "*Carbon Trading*", "penjualan karbon", "Jual Beli Karbon", "*Exchange Carbon*", "Kredit Karbon", "*Credit Carbon*", dan "Bursa Karbon". Data dikumpulkan dalam rentang waktu dari 1 Januari 2020 hingga 15 April 2024. Alasan digunakan nya rentan waktu tersebut dikarenakan di tahun 2007 sudah didirikan sebuah proyek *Katingan Mentaya*. Proyek ini guna dalam melakukan konservasi hutan atau melakukan perawatan hutan dengan menghasilkan kredit karbon yang dapat digunakan untuk perdagangan karbon di indonesia dalam mengurangi emisi karbon. Untuk proses pengambilan data terdapat pada *flowchart*.



Gambar 3.2. Pengumpulan Data

3.1.3 Perancangan Sistem

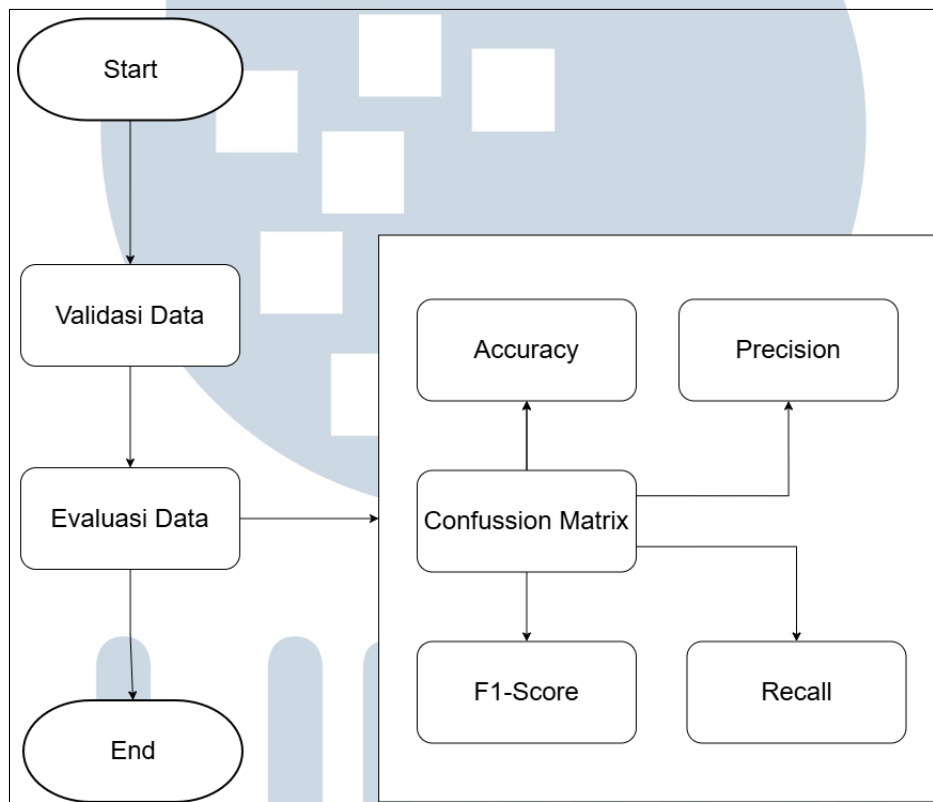
Pada tahap ini, semua data yang telah dikumpulkan akan melalui tahap perancangan sistem, dimana akan dilakukan penggabungan data, kemudian dilakukan *Text Pre-processing*. Proses ini sangat penting untuk memastikan bahwa data siap digunakan dalam analisis atau model *machine learning*. *Text pre-processing* melibatkan serangkaian langkah untuk membersihkan dan menormalkan teks sehingga dapat diproses dengan efisien dan akurat. Selanjutnya akan dilakukan ekstraksi fitur mengubah data ke vektor, dilanjutkan dengan melakukan pembagian data, dan yang terakhir adalah melakukan penerapan model.



Gambar 3.3. Perancangan Sistem

3.1.4 Validasi dan Evaluasi

Pada tahap ini merupakan, hasil evaluasi dan validasi yang dilakukan. Evaluasi yang digunakan adalah menggunakan *Confussion Matrix* yang menghasilkan *accuracy*, *recall*, *f1-score*, dan *precision*. Selain itu terdapat juga hasil validasi yang memberikan seberapa efektif dan akurat suatu model atau sistem dalam pembelajaran mesin.



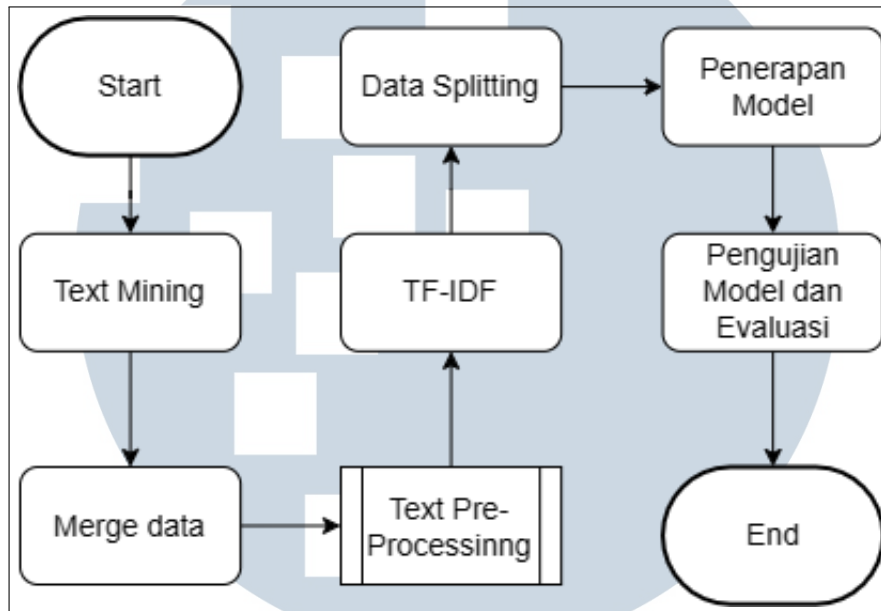
Gambar 3.4. Evaluasi dan Validasi

3.1.5 Dokumentasi

Dalam kegiatan penelitian, dokumentasi menjadi elemen krusial dari keseluruhan proses. Dokumentasi ini melibatkan penyusunan laporan penelitian secara bertahap, mulai dari bagian pendahuluan hingga ke bagian kesimpulan. Selain itu, dokumentasi juga meliputi pencatatan hasil eksperimen, analisis data, serta prosedur yang dilakukan selama penelitian. Dokumentasi yang rapi dan lengkap mendukung transparansi, reproduktibilitas, serta integritas penelitian yang tengah dilakukan.

3.2 Rancangan Sistem

Pada gambar 3.1 merupakan sebuah *flowchart* yang dilakukan dalam rancangan sistem.



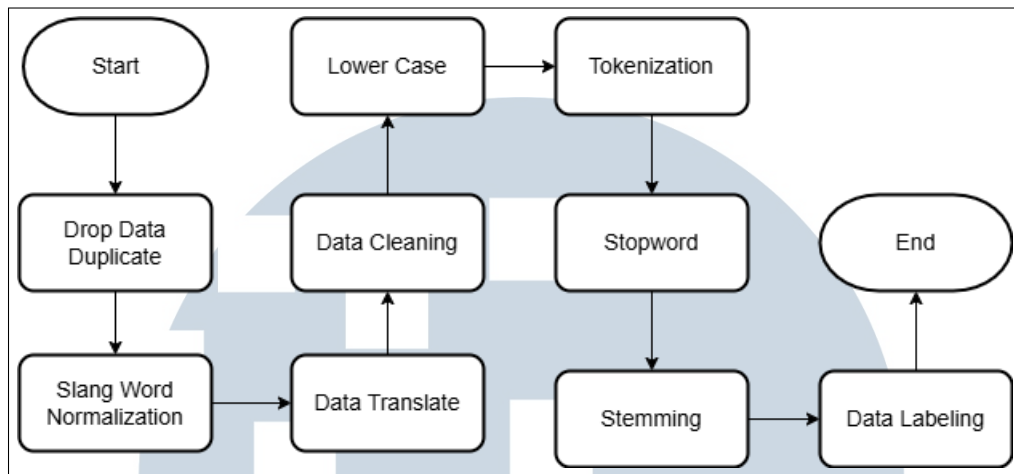
Gambar 3.5. *Flowchart* Rancangan Sistem

3.2.1 *Text Mining*

Pengumpulan data untuk analisis sentimen ini dilakukan dengan menggunakan *library tweet harvest*. Data *tweet* yang dikumpulkan difokuskan pada topik perdagangan karbon di Indonesia. Beberapa kata kunci yang digunakan untuk mengambil *tweet* mencakup "Perdagangan Karbon", "*Carbon Trading*", "penjualan karbon", "Jual Beli Karbon", "*Exchange Carbon*", "Kredit Karbon", "*Credit Carbon*", dan "Bursa Karbon". Data dikumpulkan mulai dari 1 Januari 2020 hingga 15 April 2024.

3.2.2 *Text Pre-processing*

Pada gambar 3.6 merupakan sebuah *flowchart* yang menunjukkan bagaimana dilakukannya *Text Pre-processing* pada penelitian ini.



Gambar 3.6. *Flowchart Text Pre-processing*

Untuk melakukan *Text Pre-processing*, menggunakan beberapa *library* seperti *nlk*, *textblob*, *regex*, *pandas*, dan *stemmer*. Langkah pertama yang dilakukan dalam *Text Pre-processing* adalah menghapus data duplikat atau data yang benar-benar terlihat sama. Penghapusan data duplikat ini dilakukan menggunakan *library pandas* dan akan menyimpan entri pertama pada data tweet.

kemudian, dilakukan *Slang Word Normalization*, yaitu mengubah kata-kata yang menggunakan bahasa gaul atau tidak baku ke dalam bentuk yang lebih formal. Normalisasi kata slang ini dilakukan dengan menggunakan *dictionary* dan *pandas*. Normalisasi dilakukan dengan mengumpulkan kata-kata typo atau kata-kata gaul dan mengubahnya ke kata baku.

Selanjutnya, data tersebut akan diterjemahkan menggunakan *website Deepl*. Setelah terjemahan selesai dilakukan, data tersebut siap untuk dilakukan *Data Cleaning* atau pembersihan data. Pembersihan data dilakukan dengan menghapus angka, simbol, *URL*, spasi berlebihan, dan tanda baca yang tidak diperlukan menggunakan *regex*. Selain itu, dilakukan perubahan pada setiap kata-kata dari huruf besar ke huruf kecil dilakukan menggunakan *pandas*.

Setelah data dibersihkan, proses *Tokenization* dilakukan untuk memecah teks menjadi daftar kata-kata terpisah menggunakan *library nltk*. Selanjutnya, dilakukan penghapusan *Stopwords* untuk menghilangkan kata-kata umum yang tidak memiliki makna penting dalam analisis sentimen, juga menggunakan *nltk*.

Kemudian, data tersebut akan dilakukan *Stemming*, yaitu mengubah kata-kata ke bentuk dasar. Proses ini dilakukan menggunakan *library stemmer*. Langkah terakhir dalam *Text Pre-processing* adalah melakukan Labeling untuk memberi label atau anotasi pada data sesuai dengan kebutuhan analisis sentimen atau tugas

lainnya. Proses *labeling* dapat dilakukan menggunakan *library pandas* untuk mengelola dan memberi label pada data.

3.2.3 *TF-IDF*

Setelah data melalui tahap *pre-processing*, langkah berikutnya adalah menerapkan metode *TF-IDF* (*Term Frequency-Inverse Document Frequency*). Proses ini melibatkan pembobotan data teks dengan memberikan nilai penting pada setiap kata, berdasarkan frekuensi kemunculan kata tersebut dalam dokumen (*TF*) dan pentingnya kata tersebut dalam keseluruhan kumpulan dokumen (*IDF*).

Dalam langkah ini, setiap kata dalam dokumen diberi skor *TF-IDF*. Skor ini dihitung dengan mengalikan frekuensi kemunculan kata dalam dokumen (*TF*) dengan logaritma invers dari frekuensi kemunculan kata tersebut dalam semua dokumen (*IDF*). Hasilnya adalah nilai bobot yang menunjukkan seberapa pentingnya kata tersebut dalam konteks dokumen tersebut dan dalam kumpulan dokumen secara keseluruhan.

Skor *TF-IDF* membantu mengidentifikasi dan menekankan kata-kata yang lebih signifikan untuk analisis, sementara mengurangi pengaruh kata-kata yang umum dan kurang informatif. Dengan demikian, penerapan *TF-IDF* memungkinkan model *machine learning* untuk lebih efektif dalam mempelajari pola dari data teks yang telah diberi bobot, sehingga meningkatkan akurasi dalam tugas-tugas seperti klasifikasi sentimen dan penggalian informasi.

3.2.4 *Data Splitting*

Sebelum melakukan pelatihan dan pengujian model, data perlu dibagi terlebih dahulu. Pada pembagian ini, 80% data digunakan untuk pelatihan (*training data*), sedangkan 20% sisanya digunakan untuk pengujian (*testing data*). Selain itu, terdapat juga pembagian data dengan proporsi 70% untuk pelatihan dan 30% untuk pengujian, serta 60% untuk pelatihan dan 40% untuk pengujian [14].

3.2.5 *Naive Bayes*

Setelah data dibagi menjadi set data pelatihan dan pengujian, langkah selanjutnya adalah menerapkan penggunaan dari *Naive Bayes*. Model *Naive Bayes* yang digunakan adalah *Bernoulli Naive Bayes* dan *Complement Naive Bayes* dengan menggunakan pembobotan pada *TF-IDF*. Dalam *Bernoulli Naive Bayes*,

setiap fitur atau kata dalam vektor TF-IDF dianggap sebagai variabel biner (1 atau 0), menunjukkan apakah kata itu muncul atau tidak dalam dokumen. Sementara itu, *Complement Naive Bayes* mencoba menyeimbangkan distribusi kelas dengan memberikan bobot lebih pada kelas minoritas saat membuat prediksi.

3.2.6 Pengujian Model

Setelah model selesai dilatih, kedua model diuji menggunakan data uji yang sama dengan memanfaatkan *confusion matrix*. Evaluasi ini menghasilkan laporan yang mencakup *precision*, *recall*, dan *F1-score* untuk masing-masing kategori sentimen positif dan negatif. Selain itu, keseluruhan akurasi juga dihitung dan disajikan dalam laporan tersebut. Proses validasi juga dilakukan menggunakan *k-fold cross validation* untuk memastikan keandalan model.

