



# Sentimen Analysis Ulasan Google Maps Restoran di Makassar



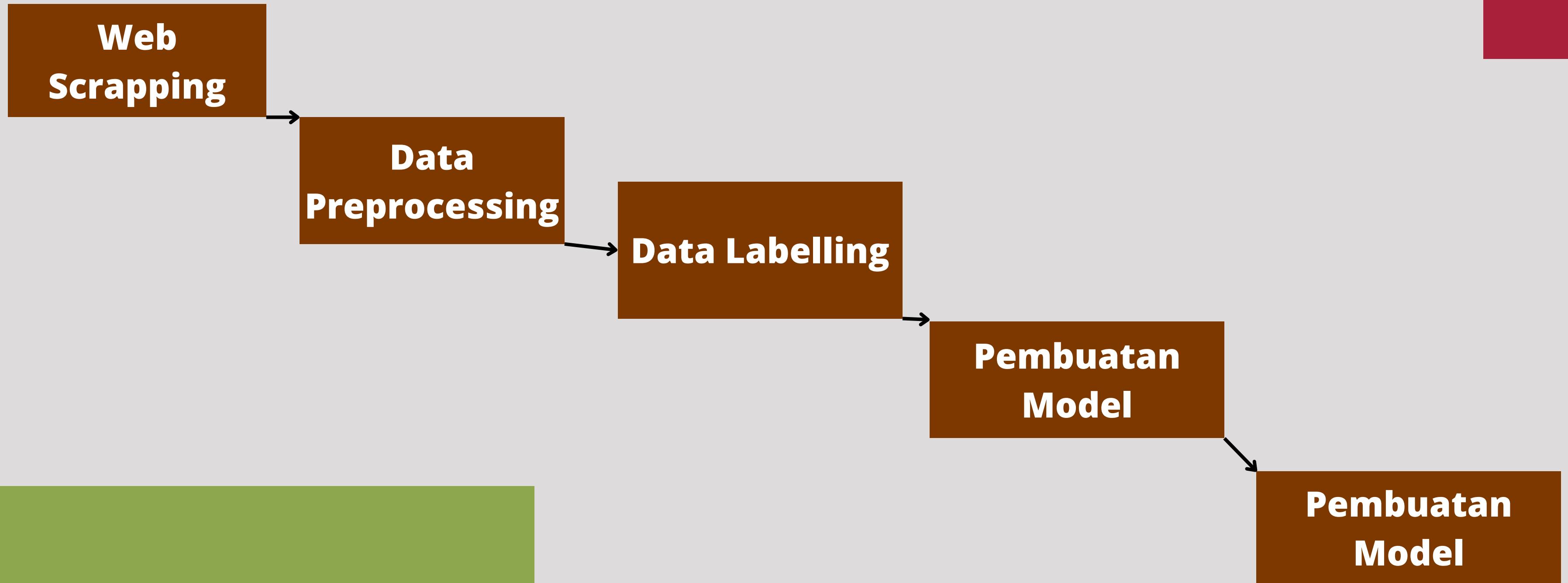


# Tujuan

Proyek Akhir Matakuliah Machine Learning Sentimen Ulasan pada Restoran yang ada di Makassar

Proyek akhir ini bertujuan untuk melakukan analisis sentimen pada ulasan restoran di Makassar menggunakan teknik machine learning. Data yang digunakan dalam proyek ini diambil melalui teknik web scraping pada Google Maps menggunakan library SerpApi. Restoran yang menjadi fokus dalam analisis sentimen adalah Hongkong Restaurant, RM Pallu Kaloa, Rumah Makan Pattene, dan Rich Taste.

# WorkFlow



# Web Scrapping

Metode pengumpulan data yang digunakan adalah scraping data menggunakan library SerpApi. SerpApi adalah sebuah platform yang menyediakan API untuk melakukan scraping data pada berbagai platform seperti Google, Bing, dan lain sebagainya. Pada kernel ini, SerpApi digunakan untuk melakukan scraping data pada ulasan restoran yang ada di Makassar di Google Maps. Dengan menggunakan SerpApi, kita dapat dengan mudah mengambil data ulasan restoran secara otomatis tanpa harus melakukan pengambilan data secara manual satu persatu. Data yang berhasil diambil kemudian akan digunakan sebagai data latih dalam pembuatan model untuk melakukan klasifikasi sentimen ulasan restoran.



# Data Preprocessing

Pada tahapan ini dilakukan preprocessing terhadap data ulasan yang telah diambil. Preprocessing meliputi pembersihan data, tokenisasi, stopword removal, dan stemming.



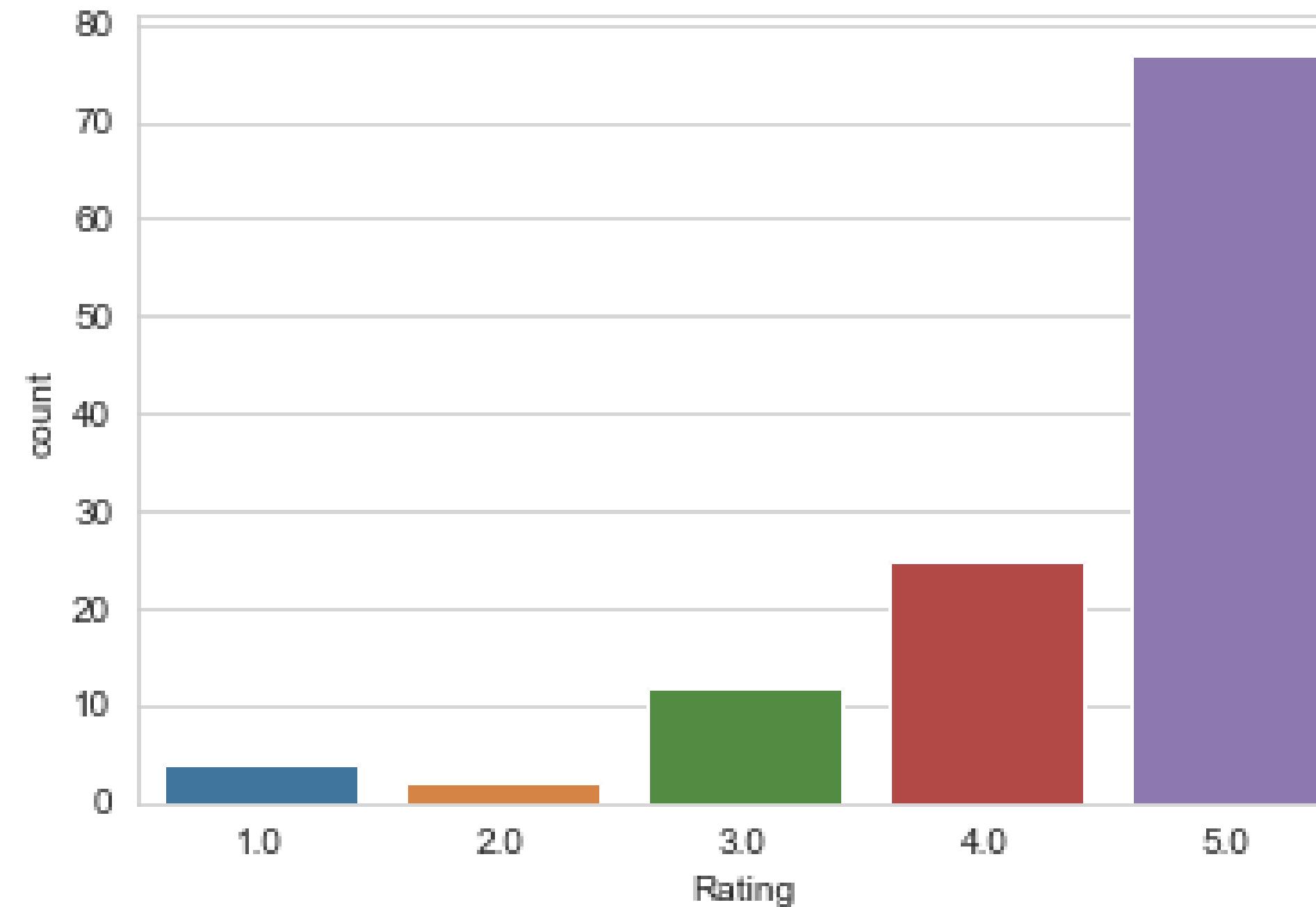
# Data Labelling

**1 = Positif**

**0 = Negatif**

Pada tahapan ini dilakukan labelisasi pada data ulasan. Labelisasi dilakukan dengan mengkategorikan ulasan menjadi positif, negatif, atau netral berdasarkan rating yang diberikan oleh pengguna. Ulasan dengan rating 1-2 dikategorikan sebagai negatif, rating 3 dikategorikan sebagai netral, dan rating 4-5 dikategorikan sebagai positif.

# Hasil Labelling



Dari hasil Labelling kita melakukan proses EDA (Exploratory Data Analysis) dari proses ini didapatkan visualisasi seperti disamping, dari visualisasi disamping kita dapat simpulkan bahwa sentimen yang di berikan oleh reviewer dominan positif atau rating 5

# Pembuatan Model

Proses pembuatan model menggunakan beberapa algoritma machine learning untuk melakukan prediksi

Multinomial Naive Bayes

Support Vector Machine (SVM)

Random Forest Classifier

Logistic Regression

K-Nearest Neighbor (KNN)

# Evaluasi Model

beberapa algoritma machine learning telah diuji coba untuk melakukan klasifikasi sentimen pada ulasan pelanggan. Algoritma yang digunakan antara lain adalah Naive Bayes, Logistic Regression, Random Forest, SVM, KNN, dan Extra Trees.

Ketika dilakukan uji coba dengan dataset yang telah disiapkan, didapatkan hasil akurasi yang berbeda-beda pada setiap algoritma. Hasil akurasi terbaik diperoleh dari Naive Bayes, Random Forest, dan SVM dengan nilai 0.8361 atau 83.61%. Sedangkan akurasi terendah diperoleh dari KNN dengan nilai 0.8014 atau 80.14%.

Namun, selain akurasi, perlu juga diperhatikan nilai precision, recall, dan f1-score dari masing-masing algoritma. Precision dan recall adalah metrik evaluasi yang digunakan untuk mengukur kinerja suatu algoritma pada kelas yang berbeda. Sedangkan f1-score adalah nilai rata-rata harmonic dari precision dan recall.

Berdasarkan hasil evaluasi precision, recall, dan f1-score, didapatkan bahwa algoritma Naive Bayes memiliki nilai f1-score terbaik yaitu 0.91, diikuti oleh Random Forest dan SVM yang sama-sama memiliki nilai f1-score sebesar 0.91 juga. Sedangkan nilai f1-score terendah diperoleh dari KNN dengan nilai 0.8863.

Berdasarkan hasil evaluasi tersebut, dapat disimpulkan bahwa algoritma Naive Bayes, Random Forest, dan SVM dapat digunakan untuk melakukan klasifikasi sentimen pada ulasan pelanggan restoran di Makassar.

**Naive Bayes**  
**accuracy - 0.83611111111111**  
**precision - 0.83611111111111**  
**recall - 1.0**  
**f1 - 0.910000000000001**

**Logistic Regression**  
**accuracy - 0.82361111111111**  
**precision - 0.8343253968253969**  
**recall - 0.9857142857142858**  
**f1 - 0.9023809523809524**

**Random Forest**  
**accuracy - 0.83611111111111**  
**precision - 0.83611111111111**  
**recall - 1.0**  
**f1 - 0.910000000000001**

**SVM**  
**accuracy - 0.83611111111111**  
**precision - 0.83611111111111**  
**recall - 1.0**  
**f1 - 0.910000000000001**

**KNN**  
**accuracy - 0.801388888888889**  
**precision - 0.8279761904761905**  
**recall - 0.9571428571428571**  
**f1 - 0.8863095238095239**

**Extra Trees**  
**accuracy - 0.825**  
**precision - 0.8440476190476189**  
**recall - 0.9714285714285715**  
**f1 - 0.901547619047619**

# Hasil

Hasil kedua menunjukkan bahwa penggunaan teknik pembobotan kata TF-IDF (Term Frequency-Inverse Document Frequency) dengan algoritma Extra Trees Classifier memberikan akurasi prediksi yang lebih tinggi, yaitu sebesar 0.8888888888888888.

Pada dasarnya, TF-IDF adalah teknik pembobotan kata yang digunakan untuk mengevaluasi seberapa penting suatu kata dalam dokumen atau kumpulan dokumen. Sedangkan algoritma Extra Trees Classifier adalah algoritma yang termasuk ke dalam kelompok algoritma ensemble learning, yang melakukan prediksi dengan cara menggabungkan beberapa pohon keputusan yang dibuat secara acak.

Dengan menggunakan teknik pembobotan kata TF-IDF, kata-kata yang muncul pada dokumen yang sama namun tidak relevan dengan topik dapat diidentifikasi dan diberi bobot yang lebih rendah, sehingga meningkatkan akurasi prediksi. Sementara itu, penggunaan algoritma Extra Trees Classifier mampu meningkatkan performa prediksi dengan cara mengurangi overfitting dan meningkatkan generalisasi model.

Dalam konteks kasus ini, hasil yang lebih tinggi pada penggunaan TF-IDF dan Extra Trees Classifier menunjukkan bahwa kombinasi teknik pembobotan kata dan algoritma yang digunakan lebih tepat untuk mengekstrak sentimen dari ulasan restoran di Makassar. Namun, tetap perlu diingat bahwa hasil prediksi juga dipengaruhi oleh kualitas data dan proses preprocessing yang dilakukan sebelumnya.