

# Analisa Sentimen Pada Twitter Tentang Kekerasan Dalam Rumah Tangga Menggunakan Metode Random Forest dan Extreme Gradient Boosting

Robiatul Asyarah

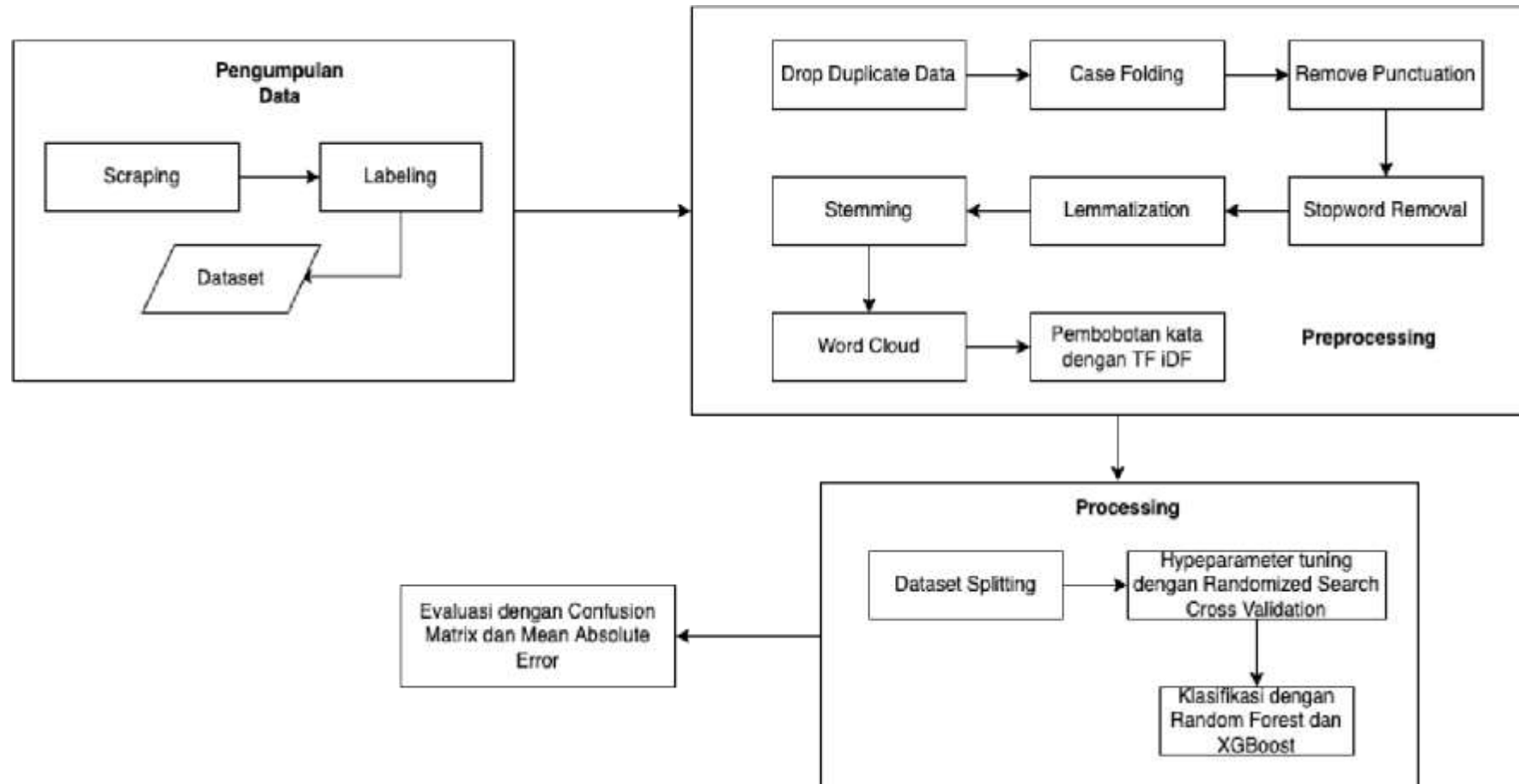
191080200040

# Pendahuluan

- Media sosial Twitter menjadi salah satu pilihan utama masyarakat global dalam berkomunikasi, terbukti dari lonjakan jumlah pengguna Twitter secara global. Pada tahun 2016, Twitter berhasil mencatatkan sekitar 313 juta pengguna yang aktif setiap bulannya [1]. Individu akan berbagi informasi terkini atau pendapat mengenai isu-isu hangat global melalui media sosial Twitter. Isu-isu yang sedang tren dan mendapatkan banyak tanggapan dari pengguna dapat menghasilkan topik yang sedang populer di platform ini, dikenal dengan sebutan "trending topic".
- Dengan meningkatnya jumlah pengguna Twitter, terjadi lonjakan dalam jumlah tweet yang diposting. Tweet-tweet tersebut berisi pandangan dan komentar publik yang beragam, meliputi bidang ekonomi, perilaku sosial, fenomena alam, perdagangan, pendidikan, hiburan, dan berbagai aspek lainnya. Salah satu isu yang menonjol dalam konteks perilaku sosial adalah kasus KDRT yang baru-baru ini mencuat. Pengguna Twitter aktif dalam memberikan komentar dan pandangan mengenai isu tersebut, menghasilkan diskusi yang ramai di platform ini

# Metode Penelitian

- Rangkaian langkah penelitian menggambarkan secara menyeluruh urutan prosedur yang akan dilaksanakan dalam pelaksanaan penelitian ini, dimulai dari awal hingga tahap akhir. Serangkaian langkah penelitian yang akan ditempuh dapat dijelaskan lebih rinci melalui sebuah diagram alir,



# Pengumpulan Data

- Data yang digunakan dalam studi ini diperoleh dari platform media sosial Twitter, dengan melakukan pencarian berdasarkan kata kunci KDRT. Dalam proses perolehan data, para peneliti mengimplementasikan Teknik scraping melalui Bahasa pemrograman Python serta memanfaatkan pustaka (library) Tweepy.

# Preprocessing

- **Case Folding**
- **Remove Punctuation**
- **Tokenization**
- **Stopword Removal**
- **Stemming**
- **Pembobotan Kata**
- **Handling Imbalanced Data**

# Processing

- **Hyperparameter tuning dengan Randomized Cross Validation**
- **Klasifikasi dengan Algoritma Random Forest dan XGBoost**

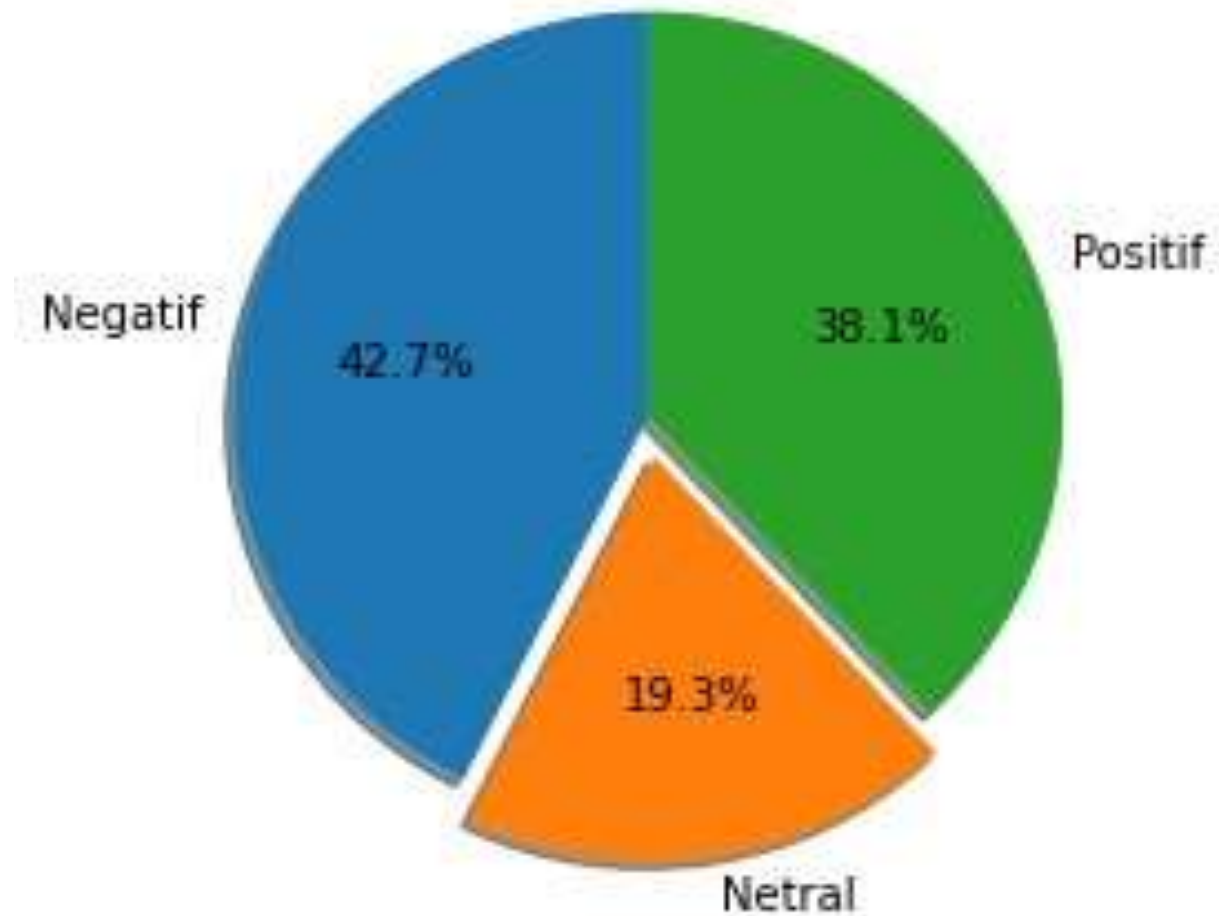
# Evaluasi

Tahap ini memiliki tujuan untuk mengevaluasi kinerja dari model machine learning yang telah dibangun. Dalam tahap evaluasi, peneliti menggunakan confusion matrix untuk menganalisis kinerja model yang telah dihasilkan.



# Hasil dan Pembahasan

- **Dataset**



- **Preprocessing Data**

**Table 3 Hasil Text Preprocessing**

<b>Sebelum Preprocessing</b>	<b>Setelah Preprocessing</b>
Kerja nyatanya adalah mengatasi masalah yang berkaitan dengan perlindungan perempuan dan anak. Saat ini KDRT	kerja nyata adalah atas masalah yang kait dengan lind ung perempuan dan anak

# Processing

**Tabel 6.** Hasil Hyperparameter tuning untuk algoritma Random Forest

Hyperparameter	Hasil
n_estimators	188
max_depth	48
Max_features	0.82
Min_samples_leaf	1

**Tabel 7.** Hasil Hyperparameter tuning untuk algoritma XGBoost

Hyperparameter	Hasil
max_depth	3
learning_rate	0.12
n_estimators	185
subsample	0.78
gamma	1
colsample_bytree	0.69
reg_alpha	0.01
reg_lambda	0.09

# Evaluasi

