

PREDIKSI TINGKAT KRIMINALITAS MENGUNAKAN METODE XGBOOST DAN KLASIFIKASI MENGUNAKAN RANDOM FOREST

Richard Petrus Haposan Siagian

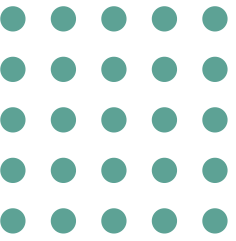
NIM : 21081010250

PROGRAM STUDI INFORMATIKA

2024

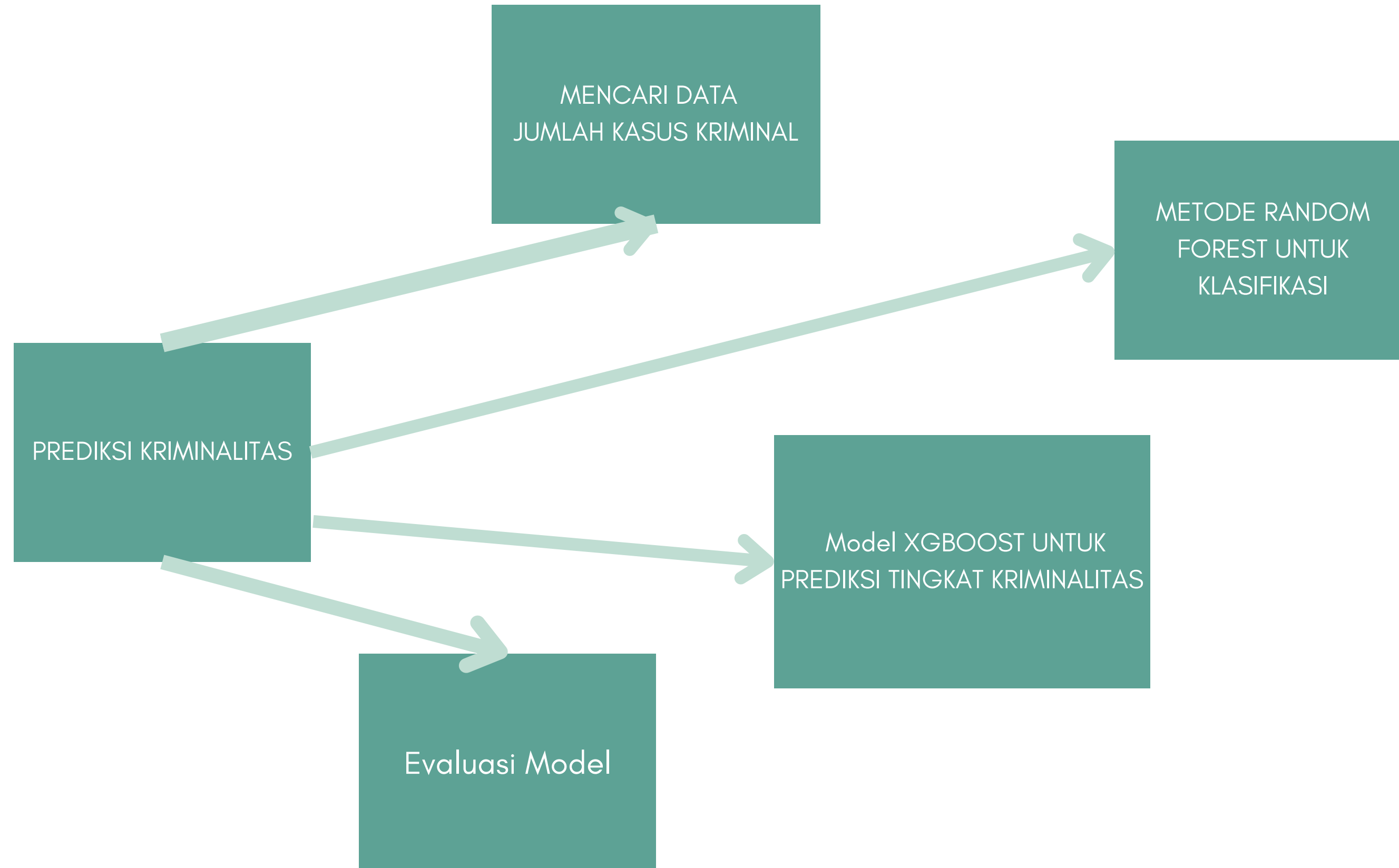
LATAR BELAKANG

Kriminalitas merupakan salah satu permasalahan sosial yang terus meningkat di berbagai negara, termasuk Indonesia. Tingkat kriminalitas tidak hanya berdampak pada keamanan dan ketertiban umum, tetapi juga memengaruhi kualitas hidup masyarakat serta pertumbuhan ekonomi. Dengan meningkatnya populasi, urbanisasi, dan perubahan kondisi sosial ekonomi, pola kriminalitas menjadi semakin kompleks dan sulit diprediksi.



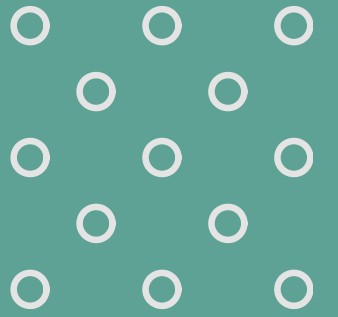
PERUMUSAN MASALAH

1. Bagaimana kinerja algoritma Random Forest dan XGBoost dalam memprediksi tingkat kriminalitas berdasarkan faktor lingkungan?
2. Bagaimana hubungan antara faktor lingkungan (ekonomi, demografi, dan infrastruktur) dengan tingkat kriminalitas di suatu wilayah?
3. Menerapkan model XGBoost dan Random forest yang mengintegrasikan data eksternal dan internal untuk memprediksi dan mengklasifikasi tingkat kriminalitas di Indonesia



Keuntungan Menggunakan XGBoost:

- Efektif untuk pencarian solusi di ruang parameter yang besar dan kompleks.
- Algoritma ini mampu menangkap pola non-linear yang kompleks dalam data.
- Dengan menggunakan fitur Feature Importance, XGBoost dapat mengidentifikasi faktor dominan yang paling memengaruhi tingkat kriminalitas, seperti tingkat pengangguran, kemiskinan, atau kepadatan penduduk.



Keuntungan Menggunakan Random Forest:

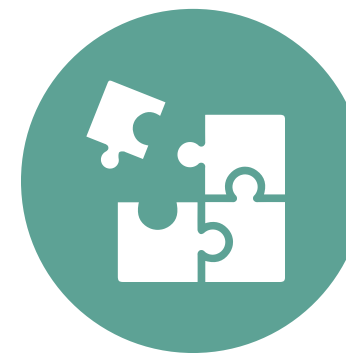
- Membuat banyak Decision Tree dan menggabungkan hasilnya untuk menghasilkan prediksi yang lebih stabil dan akurat.
- random forest tetap bekerja dengan baik bahkan ketika ada lebih banyak fitur daripada data, yang sering terjadi dalam kasus analisis bisnis atau ekonomi.





Mencari Data Kriminalitas

Data data dari BPS
(Badan Pusat Statistik)



Preprocessing

Menginput data yang
sudah dibuat dalam
model csv

Wilayah Jawa Barat	Tingkat Pengangguran Terbuka Kabupaten/Kota (Persen)						
	2018	2019	2020	2021	2022	2023	2024
Provinsi Jawa Barat	8,23	8,04	10,46	9,82	8,31	7,44	6,75
Bogor	9,83	9,11	14,29	12,22	10,64	8,47	7,34
Sukabumi	7,84	8,05	9,6	9,51	7,77	7,32	7,11
Cianjur	10,23	9,81	11,05	9,32	8,41	7,71	5,99
Bandung	5,07	5,51	8,58	8,32	6,98	6,52	6,36
Garut	7,12	7,35	8,95	8,68	7,6	7,33	6,96
Tasikmalaya	6,92	6,31	7,12	6,16	4,17	3,89	3,74
Ciamis	4,64	5,16	5,66	5,06	3,75	3,52	3,37
Kuningan	9,1	9,68	11,22	11,68	9,81	9,49	7,78
Cirebon	10,64	10,35	11,52	10,38	8,11	7,65	6,74
Majalengka	5	4,37	5,84	5,71	4,16	4,12	4,01
Sumedang	7,54	7,7	9,89	9,18	7,72	6,94	6,16
Indramayu	8,46	8,35	9,21	8,3	6,49	6,46	6,25

CO Untitled0.ipynb ☆

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

```
print("Preview Data:")
print(data.head())

# 2. Preprocessing Data

target_column = 'Crime_Rate'
X = data.drop(columns=[target_column])
y = data[target_column]

# Mengisi nilai yang hilang jika ada
X = X.fillna(X.mean())

# Membagi data menjadi Training dan Testing Set (80:20)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# 3. Training Model Random Forest
print("\nTraining Random Forest Model...")
rf_model = RandomForestRegressor(n_estimators=100, random_state=42)
rf_model.fit(X_train, y_train)

# 4. Training Model XGBoost
print("\nTraining XGBoost Model...")
xgb_model = XGBRegressor(n_estimators=100, learning_rate=0.1, random_state=42)
xgb_model.fit(X_train, y_train)

# 5. Evaluasi Model
def evaluate_model(model, X_test, y_test, model_name):
    predictions = model.predict(X_test)
    mse = mean_squared_error(y_test, predictions)
```




Terima Kasih