

Portofolio Proyek: Prediksi Keselamatan Penumpang Titanic

Deskripsi Proyek

Proyek ini bertujuan untuk memprediksi apakah penumpang Titanic akan selamat atau tidak berdasarkan beberapa fitur seperti jenis kelamin, usia, kelas, dan harga tiket. Model yang digunakan adalah Random Forest Classifier.

Langkah-langkah Proyek

- Load dataset Titanic dari seaborn
- Data cleaning: mengisi nilai kosong, menghapus kolom tidak relevan
- Exploratory Data Analysis: visualisasi distribusi usia, survival rate, korelasi
- Encoding data kategorikal
- Split data training dan testing (80:20)
- Pelatihan model RandomForestClassifier
- Evaluasi menggunakan akurasi, confusion matrix, classification report

Potongan Kode Penting

```
from sklearn.ensemble import RandomForestClassifier
model = RandomForestClassifier(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

y_pred = model.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
print(f"Akurasi: {accuracy:.2%}")

from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred)

from sklearn.metrics import classification_report
print(classification_report(y_test, y_pred))
```

Hasil Evaluasi Model

- Akurasi model: $\pm 81\%$
- Confusion Matrix menunjukkan prediksi benar dan salah secara visual
- Classification Report memberikan metrik precision, recall, f1-score

Kesimpulan

Model Random Forest menunjukkan performa cukup baik untuk memprediksi keselamatan penumpang Titanic. Proyek ini menggambarkan tahapan lengkap dalam proses data science mulai dari preprocessing hingga evaluasi model.