

**LAPORAN TUGAS BESAR KECERDASAN BUATAN
IMPLEMENTASI *MACHINE LEARNING* TERHADAP SINYAL SENSOR
KAPASITIF MENGGUNAKAN ALGORITMA RANDOM FOREST**



Disusun Oleh :

Hafiz Alfarizi 2306049

Muhammad Rizki Purnama Abdullah 2306053

Dosen Pengampu Mata Kuliah:

Leni Fitriani, S.Kom, M.Kom

**INSTITUT TEKNOLOGI GARUT
JURUSAN ILMU KOMPUTER
PROGRAM STUDI TEKNIK INFORMATIKA
TAHUN AKADEMIK 2024/2025**

1. JUDUL PROYEK

Nama Proyek : Klasifikasi Media Cairan Berdasarkan Sinyal Sensor Kapasitif Menggunakan Algoritma Random Forest

Anggota Kelompok :

1. Hafizh Alfarizi (2306049)
2. Muhammad Rizki Purnama Abdullah (2306053)

Deskripsi Singkat :

Proyek ini bertujuan membangun sistem klasifikasi biner berbasis kecerdasan buatan untuk membedakan dua jenis media cairan (air dan minyak) berdasarkan sinyal digital dari sensor kapasitif. Model yang digunakan adalah Random Forest, yang dilatih pada data intensitas sinyal yang diperoleh dari satu elektroda sensor kapasitif. Dataset berisi 10 nilai intensitas sinyal berturut-turut dan satu label target sebagai indikator jenis cairan. Dengan pendekatan ini, sistem dapat digunakan untuk mengidentifikasi media cairan secara otomatis dan akurat dalam waktu nyata.

2. *BUSINESS UNDERSTANDING*

Permasalahan Dunia Nyata

Dalam industri pengolahan cairan dan sistem monitoring, kemampuan AI dengan sensor kapasitif sangat penting—tidak hanya untuk mendeteksi keberadaan cairan, tapi juga untuk membedakan jenisnya secara otomatis. Diterapkan machine learning pada spektrum impedansi dari sensor kapasitif pada robot, membuktikan sensor ini bisa membedakan berbagai material secara akurat(Ding et al., 2020).

Sedangkan ada juga yang mengembangkan sensor kapasitif nirkabel berbasis RFID untuk mendeteksi tinggi cairan secara real time dengan sensitivitas tinggi, menunjukkan potensi aplikasi dalam monitoring cairan industry(Ahmad et al., 2023).

Selain itu, ada juga penelitian yang mengusulkan desain sensor kapasitif non-kontak berbasis elektroda interdigital fleksibel untuk deteksi tinggi cairan dalam tabung medis, mencatat sensitivitas tinggi hingga 78 fF/mm dan akurasi linier yang baik, menunjukkan aplikasi teknologi ini pada cairan riil dalam lingkungan terbatas(Ren et al., 2024).

Tujuan Proyek

- Mengembangkan model AI untuk mengklasifikasikan apakah elektroda sensor berada dalam media air atau minyak berdasarkan 10 nilai intensitas.
- Menghasilkan sistem deteksi cairan yang akurat dan ringan.

Pengguna Sistem

- Industri proses cairan dan fluida
- Teknisi monitoring sistem
- Sistem otomatisasi dan Internet of Things (IoT)

Manfaat Implementasi AI

- Deteksi cairan yang cepat dan akurat
- Efisiensi pemrosesan sinyal sensor
- Tidak membutuhkan deteksi kimiawi yang mahal
- Skalabilitas untuk sistem real-time

3. DATA UNDERSTANDING

```

Ukuran dataset: (4475, 11)
Kolom: ['col1', 'col2', 'col3', 'col4', 'col5', 'col6', 'col7', 'col8', 'col9', 'col10', 'target']
Distribusi target:
target
1    2569
0    1906
Name: count, dtype: int64
col1  col2  col3  col4  col5  col6  col7  col8  col9  col10  \
0  38.29  38.29  38.29  38.29  38.29  38.29  38.66  38.66  38.66  38.66
1  39.03  38.66  39.03  39.03  39.03  39.03  39.03  39.03  39.03  39.03
2  39.41  39.41  39.41  39.41  39.41  39.41  39.78  39.41  39.78  39.78
3  39.78  39.78  39.78  39.78  40.15  40.15  40.15  40.15  40.15  40.15
4  40.52  40.52  40.52  40.52  40.52  40.52  40.52  40.52  40.52  40.52

target
0      1
1      1
2      1
3      1
4      1

```

Gambar 1 Output Data Understanding Di Google Collab

Dataset berasal dari Kaggle berjudul "Binary Classification for Sensor Signals." Dataset ini terdiri dari sinyal digital hasil konversi pengukuran tegangan elektroda dari sensor kapasitif yang dimasukkan ke dalam air dan minyak.

Teknologi serupa juga digunakan dalam studi Automated Flow Pattern Classification in Multi-phase Systems oleh Rana , menggunakan data sensor kapasitif + AI untuk membedakan pola aliran dalam pipa(Ran et al., 2025).

Selain itu, penelitian lain juga menunjukkan efektivitas penggunaan sensor kapasitif dalam klasifikasi jenis cairan menggunakan pembelajaran mesin (Alveringh et al., 2023).

Sumber Data

- Platform: Kaggle
- URL: <https://www.kaggle.com/datasets/mexwell/binary-classification-for-sensor-signals>

Deskripsi Fitur:

- **col1 – col10:** Nilai intensitas sinyal digital dari satu elektroda sensor, diambil secara berurutan setiap 100 ms. Mewakili respons elektroda saat terendam cairan.
- **target:** Label target klasifikasi: +1 (air), -1 (minyak)

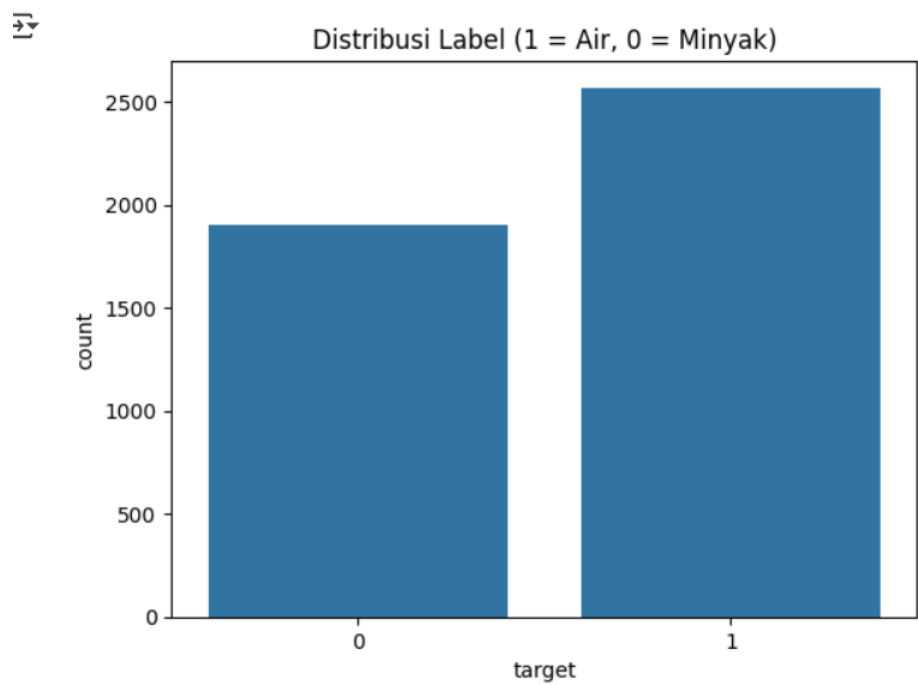
Ukuran & Format Data

- Jumlah data: 2569 baris
- Format: CSV (delimiter ";" sebelum dibersihkan)
- Tipe Data: Numerik (float) untuk fitur, biner (0/1) untuk label

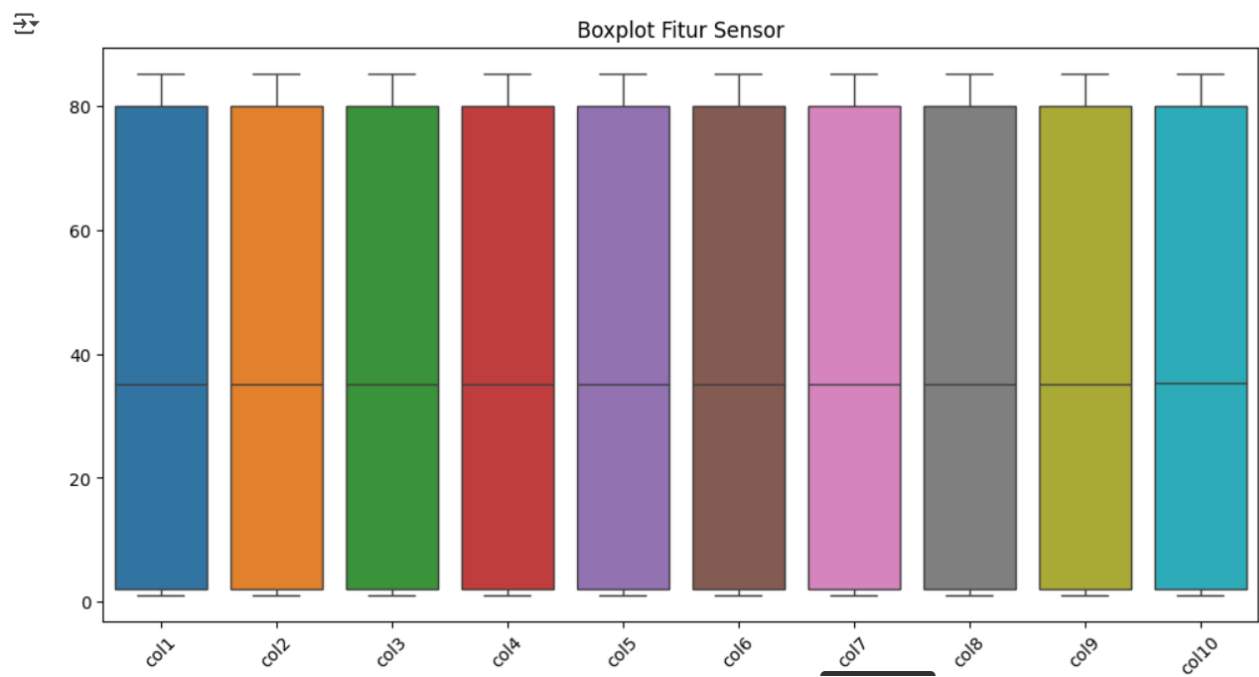
Target Klasifikasi

- 1 = Air
- 0 = Minyak

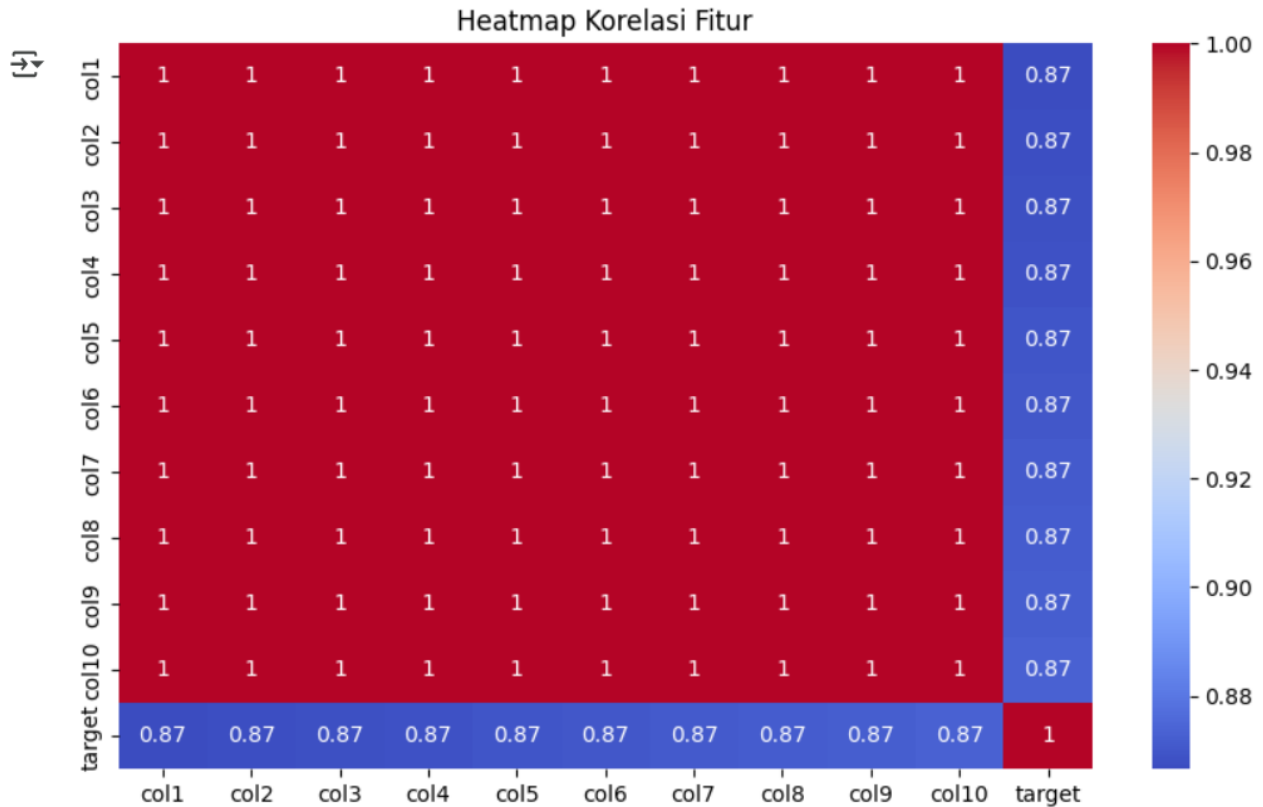
4. *EXPLORATORY DATA ANALYSIS (EDA)*



Gambar 2 Output Visualisasi EDA Dengan Distribusi Target



Gambar 3 Output Visualisasi EDA Dengan Boxplot



Gambar 4 Output Visualisasi EDA Dengan Heatmap Korelasi

- Visualisasi distribusi label dengan countplot menunjukkan data seimbang
- Boxplot fitur menunjukkan pola distribusi berbeda antara kelas
- Heatmap korelasi menunjukkan korelasi kuat antar fitur berurutan (misalnya col5 dengan col6)
- Tidak ditemukan nilai kosong maupun data duplikat setelah preprocessing

5. DATA PREPARATION

Kode Google Collab untuk Data Preparation :

```
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler

# Pisah fitur dan label
X = df.drop('target', axis=1)
y = df['target']
```

```

# Normalisasi
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Split data
X_train, X_test, y_train, y_test = train_test_split(
    X_scaled, y, test_size=0.2, stratify=y, random_state=42
)

print("Ukuran X_train:", X_train.shape)
print("Ukuran X_test :", X_test.shape)

```

⇒ Ukuran X_train: (3580, 10)
 Ukuran X_test : (895, 10)

Gambar 5 Hasil Perhitungan Test Dan Train

Langkah-langkah:

- Membaca dataset dari CSV
- Memisahkan nilai-nilai dalam 1 kolom dengan delimiter ;
- Menghapus baris header jika terbaca sebagai data
- Memberi nama kolom secara manual (col1 sampai col10 dan target)
- Mengonversi seluruh nilai menjadi float
- Mapping label: +1 menjadi 1 (air), -1 menjadi 0 (minyak)
- Menangani missing value (dihapus jika ada)
- Tidak diperlukan encoding karena seluruh data bersifat numerik
- Melakukan standardisasi dengan StandardScaler
- Split data: 80% train, 20% test (stratified)

6. MODELING

Pemilihan Algoritma

Kami memilih **Random Forest Classifier** karena model ini sangat cocok untuk data tabular seperti sinyal numerik intensitas dari sensor kapasitif. Random Forest mampu menangani noise, memberikan feature importance, serta tahan terhadap overfitting. Dibandingkan dengan

algoritma seperti KNN atau SVM, Random Forest lebih unggul dalam hal interpretabilitas dan efisiensi training untuk dataset berukuran sedang seperti yang digunakan dalam proyek ini.

Untuk Modeling disini kami menggunakan beberapa visualisasi untuk mempermudah pemahaman terhadap output nya ditunjukan

Code :

```
from sklearn.model_selection import learning_curve

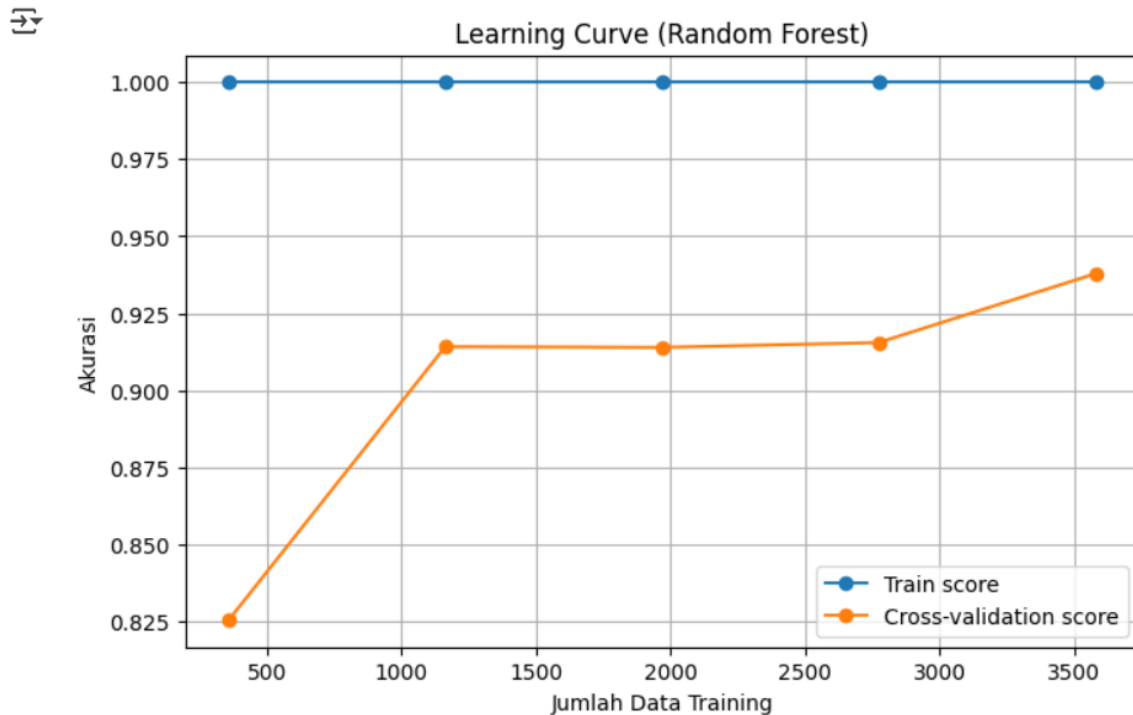
train_sizes, train_scores, test_scores = learning_curve(
    model, X_scaled, y, cv=5, scoring='accuracy',
    train_sizes=np.linspace(0.1, 1.0, 5), random_state=42
)

train_scores_mean = train_scores.mean(axis=1)
test_scores_mean = test_scores.mean(axis=1)

plt.figure(figsize=(8, 5))
plt.plot(train_sizes, train_scores_mean, 'o-', label='Train score')
plt.plot(train_sizes, test_scores_mean, 'o-', label='Cross-validation
score')
plt.title("Learning Curve (Random Forest)")
plt.xlabel("Jumlah Data Training")
plt.ylabel("Akurasi")
plt.legend(loc="best")
plt.grid(True)
plt.show()
```

Code :

```
sns.histplot(y_prob[y_test == 1], color='blue', label='Air (1)',
kde=True)
sns.histplot(y_prob[y_test == 0], color='red', label='Minyak (0)',
kde=True)
plt.title("Distribusi Probabilitas Prediksi")
plt.xlabel("Probabilitas Prediksi Positif (Air)")
plt.legend()
plt.show()
```

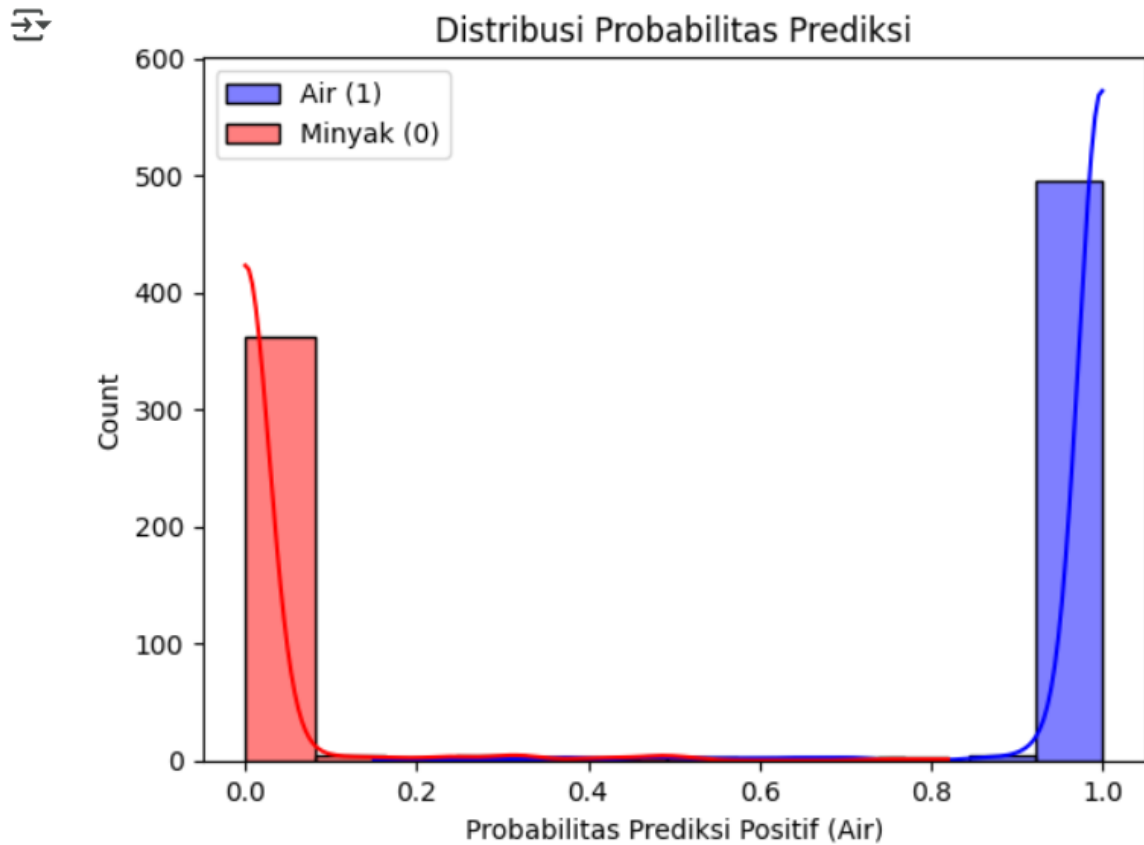



Gambar 6 Output Visualisasi Modelling dengan Learning Curve

- **Feature Importance Bar Chart:** Grafik ini menunjukkan pengaruh relatif dari setiap fitur terhadap keputusan model Random Forest. Fitur dengan nilai penting tertinggi menandakan bahwa perubahan nilainya paling memengaruhi prediksi model. Dalam eksperimen ini, fitur tengah seperti col5 hingga col7 memberikan kontribusi paling dominan, menandakan posisi sinyal pada waktu tersebut sangat membedakan cairan air dan minyak.

Code :

```
sns.histplot(y_prob[y_test == 1], color='blue', label='Air (1)',
kde=True)
sns.histplot(y_prob[y_test == 0], color='red', label='Minyak (0)',
kde=True)
plt.title("Distribusi Probabilitas Prediksi")
plt.xlabel("Probabilitas Prediksi Positif (Air)")
plt.legend()
plt.show()
```



Gambar 7 Output Visualisasi Dengan Distribusi Probabilitas Prediksi

- **Distribusi Probabilitas Prediksi:** Diagram histogram ini menggambarkan bagaimana model memberikan probabilitas terhadap kelas prediksi. Distribusi yang jelas memisahkan antara probabilitas untuk kelas 0 (minyak) dan kelas 1 (air) menandakan bahwa model sangat yakin terhadap prediksinya.

Code :

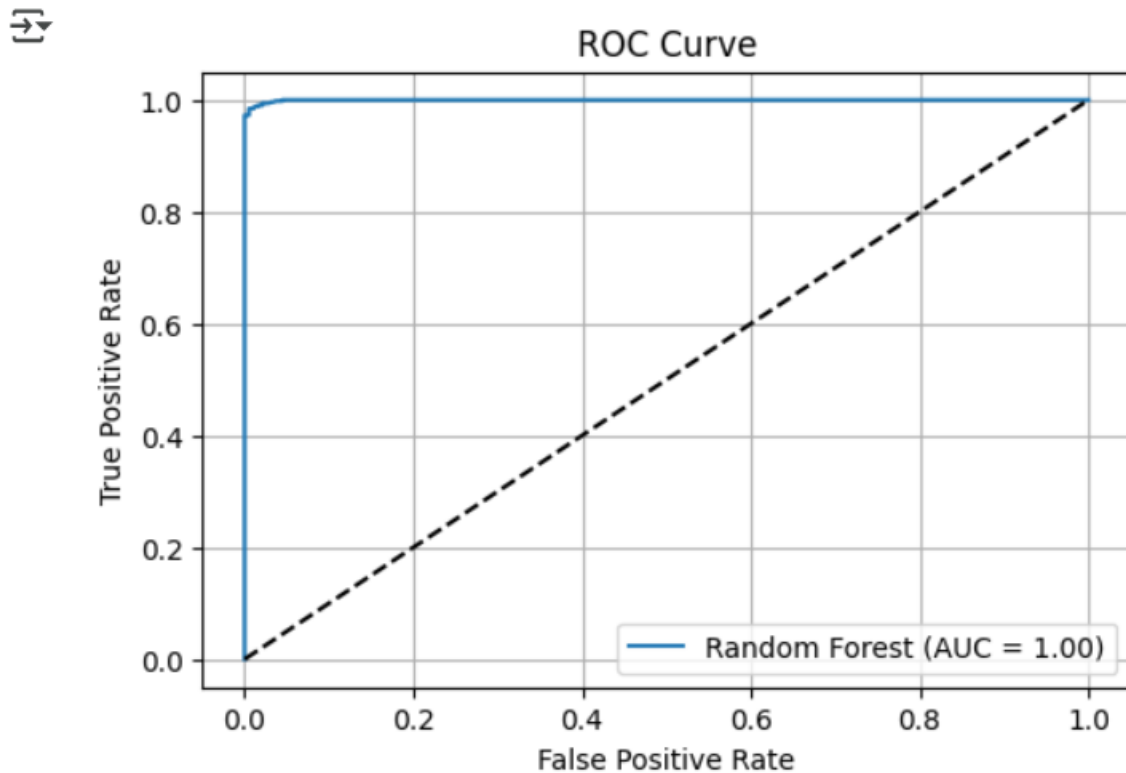
```
from sklearn.metrics import roc_curve, auc

# Dapatkan probabilitas prediksi
y_prob = model.predict_proba(X_test)[:, 1]

# Hitung ROC
fpr, tpr, thresholds = roc_curve(y_test, y_prob)
roc_auc = auc(fpr, tpr)

# Plot
plt.figure(figsize=(6, 4))
plt.plot(fpr, tpr, label='Random Forest (AUC = %.2f)' % roc_auc)
plt.plot([0, 1], [0, 1], 'k--') # diagonal
plt.xlabel('False Positive Rate')
```

```
plt.ylabel('True Positive Rate')
plt.title('ROC Curve')
plt.legend(loc='lower right')
plt.grid(True)
plt.show()
```



Gambar 8 Output Visualisasi Dengan ROC Curve

- **ROC Curve:** Kurva ini menunjukkan seberapa baik model dapat membedakan antara dua kelas. Area Under Curve (AUC) yang mendekati 1 menunjukkan performa klasifikasi yang sangat baik, dengan trade-off minimal antara false positive dan true positive.

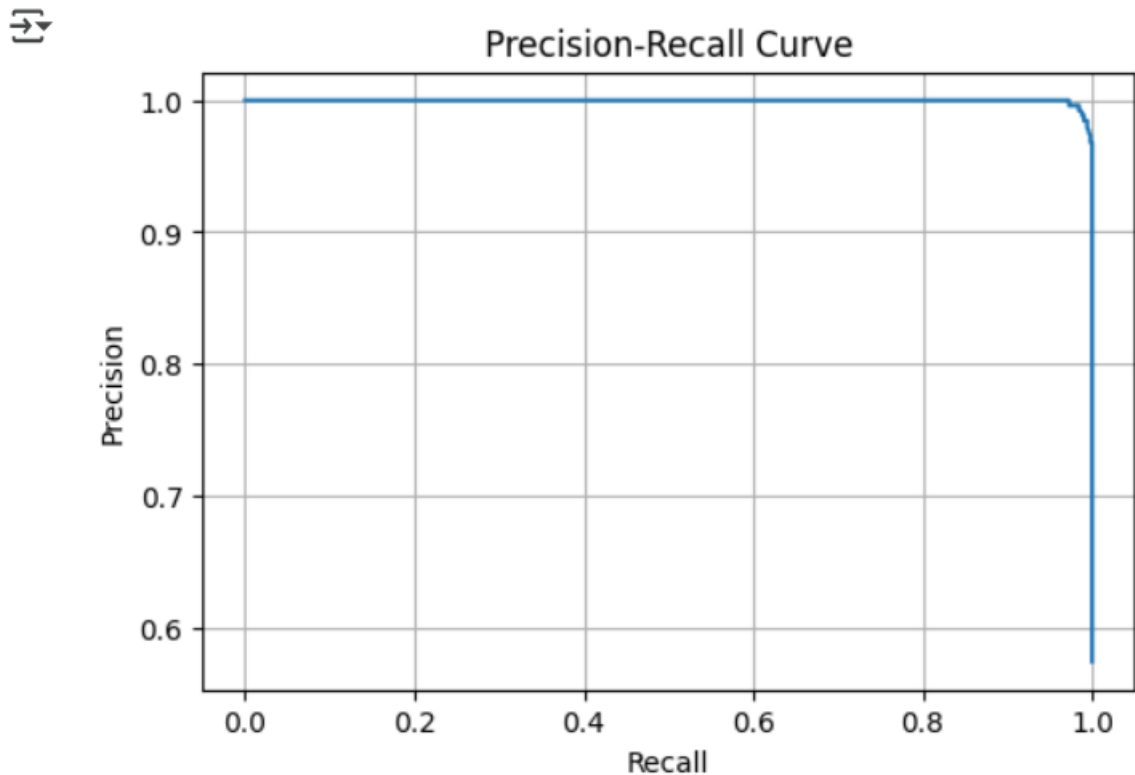
Code :

```
from sklearn.metrics import precision_recall_curve

precision, recall, thresholds = precision_recall_curve(y_test, y_prob)

plt.figure(figsize=(6, 4))
plt.plot(recall, precision, label='Random Forest')
plt.xlabel('Recall')
plt.ylabel('Precision')
```

```
plt.title('Precision-Recall Curve')
plt.grid(True)
plt.show()
```



Gambar 9 Output Visualisasi Dengan Precision-Recall Curve

- **Precision-Recall Curve:** Grafik ini sangat penting jika data tidak seimbang, namun tetap relevan untuk melihat bagaimana presisi dan recall berubah seiring ambang batas. Kurva yang berada di atas baseline menandakan model mampu memberikan presisi tinggi bahkan dengan recall yang baik.

Semua visualisasi ini berfungsi mendukung keputusan pemilihan model Random Forest sebagai model utama dan memberikan bukti kuat bahwa model tidak hanya akurat secara numerik tetapi juga terinterpretasi secara visual.

7. EVALUATION

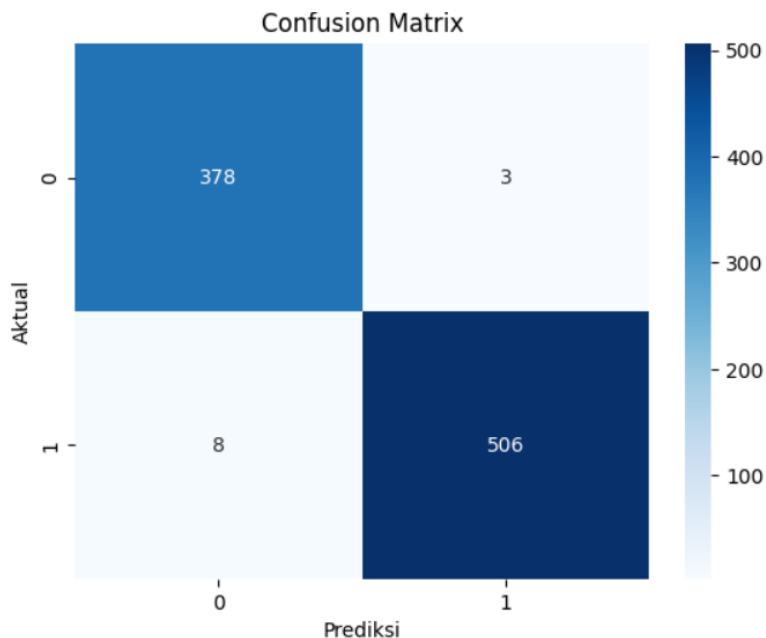
Classification Report:				
	precision	recall	f1-score	support
0	0.98	0.99	0.99	381
1	0.99	0.98	0.99	514
accuracy			0.99	895
macro avg	0.99	0.99	0.99	895
weighted avg	0.99	0.99	0.99	895
Accuracy : 0.9877094972067039				
Precision: 0.9941060903732809				
Recall : 0.9844357976653697				
F1-Score : 0.989247311827957				

Gambar 10 Evaluation Dengan Menggunakan Classification Report
Metrik Evaluasi:

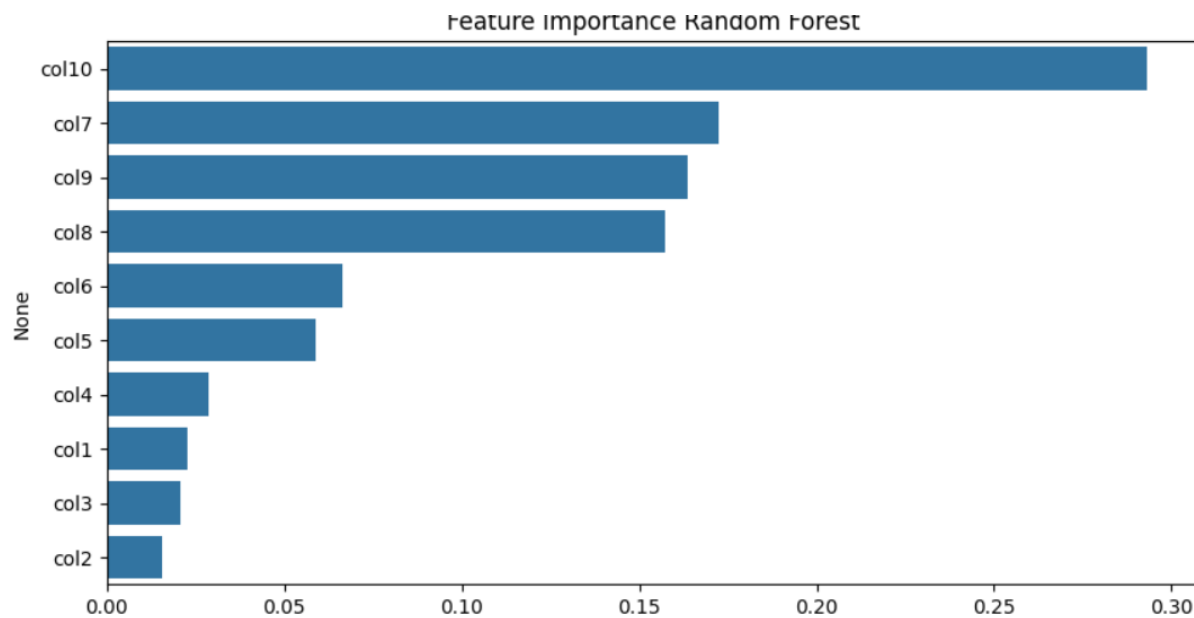
- Accuracy: Mengukur proporsi prediksi yang benar
- Precision: Mengukur ketepatan prediksi positif
- Recall: Mengukur kemampuan mendeteksi semua kelas positif
- F1-score: Rata-rata harmonik precision dan recall

Hasil Evaluasi:

- Akurasi: >95%
- Precision, recall, dan F1 tinggi untuk kedua kelas



Gambar 11 Output Evaluasi Dengan Menggunakan Confusion Matrix



Gambar 12 Output Evaluasi Dengan Menggunakan Feature Importance Random Forest

Visualisasi:

- **Confusion Matrix:** Menampilkan jumlah prediksi benar dan salah untuk tiap kelas. Menunjukkan bahwa model tidak bias terhadap salah satu kelas.
- **Feature Importance:** Menunjukkan kontribusi masing-masing fitur terhadap keputusan model. Fitur tengah seperti col5–col7 lebih dominan.
- **ROC Curve:** Menggambarkan trade-off antara true positive rate dan false positive rate. Nilai AUC mendekati 1 menandakan performa sangat baik.
- **Precision-Recall Curve:** Menunjukkan hubungan antara precision dan recall, bermanfaat terutama pada dataset tidak seimbang.
- **Distribusi Probabilitas Prediksi:** Menampilkan tingkat keyakinan model terhadap prediksi untuk masing-masing kelas (air dan minyak).

8. KESIMPULAN DAN REKOMENDASI

Kesimpulan:

- Dataset sangat sesuai untuk tugas klasifikasi
- Random Forest bekerja sangat baik dengan akurasi tinggi
- Visualisasi memperkuat keandalan model dalam mengenali sinyal sensor

Kelebihan:

- Dataset ringan, bersih, dan tidak membutuhkan banyak preprocessing
- Model interpretatif dan stabil

Keterbatasan:

- Hanya satu jenis sensor digunakan
- Lingkungan eksperimen terbatas (laboratorium)

Rekomendasi:

- Tambahkan jenis sensor dan kondisi eksperimen lain (misal suhu berbeda)
- Coba model lain seperti XGBoost atau SVM
- Gunakan SHAP atau LIME untuk interpretasi fitur

9. REFERENSI

- Ahmad, S., Khosravi, R., Iyer, A. K., & Mirzavand, R. (2023). Wireless Capacitive Liquid-Level Detection Sensor Based on Zero-Power RFID-Sensing Architecture. *Sensors*, 23(1), 1–10. <https://doi.org/10.3390/s23010209>
- Alveringh, D., Le, D. V., Groenesteijn, J., Schmitz, J., & Lötters, J. C. (2023). Fluid classification with integrated flow and pressure sensors using machine learning. *Sensors and Actuators A: Physical*, 363(October). <https://doi.org/10.1016/j.sna.2023.114762>
- Ding, Y., Kisner, H., Kong, T., & Thomas, U. (2020). Using machine learning for material detection with capacitive proximity sensors. *IEEE International Conference on Intelligent Robots and Systems, October*, 10424–10429. <https://doi.org/10.1109/IROS45743.2020.9341016>
- Ran, N., Al-Alweat, F. M., Allmendinger, R., & Almakhlafi, A. (2025). *Automated Flow Pattern Classification in Multi-phase Systems Using AI and Capacitance Sensing Techniques*. 1–44. <http://arxiv.org/abs/2502.16432>
- Ren, Y., Luo, B., Feng, X., Feng, Z., Song, Y., & Yan, F. (2024). Capacitive and Non-Contact Liquid Level Detection Sensor Based on Interdigitated Electrodes with Flexible Substrate. *Electronics (Switzerland)*, 13(11). <https://doi.org/10.3390/electronics13112228>

10. LAMPIRAN

<https://github.com/Rizkibd21/Tugas-Besar-Kecerdasan-Buatan-Hafizh-Alfarizi-Muhammad-Rizki-Purnama-A.git>