

# Upaya Pengoptimalan dalam Model Prediksi Kegagalan Product dari Perusahaan Keep It Dry

Ananda Ravi Kuntadi, Ark, Kuntadi

Departemen Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya,

[anandaravik@student.ub.ac.id](mailto:anandaravik@student.ub.ac.id)

AZRIL JANUAR ATHALLAH, AJA, ATHALLAH

Departemen Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya, [azriljanuar@student.ub.ac.id](mailto:azriljanuar@student.ub.ac.id)

DAVIN DALANA FIDELIO FREDRA, DDFR, FREDRA

Departemen Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya, [davilana04@student.ub.ac.id](mailto:davilana04@student.ub.ac.id)

MUHAMMAD YASIN HAKIM, MYH, HAKIM

Departemen Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya, [yasinh20@student.ub.ac.id](mailto:yasinh20@student.ub.ac.id)

RIDHA ILHAM ADI SETYAWAN

Departemen Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya,

[ridhailham33@student.ub.ac.id](mailto:ridhailham33@student.ub.ac.id)

SALSA ZUFAR RADINKA AKMAL, SZRA, AKMAL

Departemen Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya, [salsazufar@student.ub.ac.id](mailto:salsazufar@student.ub.ac.id)

## ABSTRAK

Proses membangun model pembelajaran mesin sangat diperlukan dalam berbagai industri, tidak terlepas dari manufaktur. Kaggle menyediakan sebuah platform untuk mempelajari dan melatih kemampuan dalam membangun model pembelajaran mesin. Proses melatih tersebut dilakukan dalam proses *trial and error* dalam proses *data preprocessing*, *scaling*, hingga *model selection*. Maka, pada projek ini, kami membandingkan beberapa pemilihan *data preprocessing* dan *model selection* untuk membangun model pembelajaran mesin untuk sebuah kaggle *competition* bernama "PPM-Predict Failures Keep It Dry". Beberapa *data preprocessing* yang akan dibandingkan adalah perbandingan *imputer* nilai kosong dengan *mean* dan *median imputer* serta perbandingan encoding fitur kategorikal dengan *one-hot encoding* dan *label encoding*. Selain itu, model yang dibandingkan adalah K-Nearest Neighbours, Naive Bayes, dan Decision Tree. Hasil yang didapatkan adalah Naive Bayes adalah model yang terbaik diantara ketiga model yang dipilih dan perbandingan *imputer* dan *encoding* tidak memiliki perbedaan yang signifikan.

**Additional Keywords:** *data preprocessing*, *one-hot encoding*, *label encoding*, *Imputer*, *K-Nearest Neighbours*, *Naive Bayes*, *Decision Tree*

## 1 PENDAHULUAN

Pembelajaran mesin memiliki peran penting dalam berbagai industri, termasuk manufaktur. Salah satu aplikasinya adalah memprediksi kegagalan peralatan, yang dapat membantu mencegah downtime dan meningkatkan efisiensi produksi. Pada project ini kami dihadapkan untuk menyelesaikan Kaggle dengan nama "PPM-Predict Failures Keep It Dry", pada kaggle tersebut kami diberi tantangan mengenai bagaimana kami menyelesaikan ini. Kami mencoba berbagai cara untuk menyelesaikan hal tersebut dimulai dari Model dengan KNN, Naive Bayes, dan Decision Tree. Kami juga mencoba melakukan optimalisasi pada model tersebut dengan menggunakan Baseline Model, Data Cleaning, Data Preprocessing, Encoding, Cross Validation.

Dalam konteks ini, kita akan membahas tantangan Kaggle berjudul "PPM-Predict Failures Keep It Dry" [Kaggle Tabular Playground Series]. Tantangan ini menyediakan platform bagi para praktisi dan peminat pembelajaran mesin untuk mengasah kemampuan mereka dalam memprediksi kegagalan pompa industri.

Kompetisi ini menawarkan dataset yang berisi informasi terkait kinerja pompa, termasuk faktor-faktor yang mungkin berkontribusi terhadap kegagalannya. Peserta ditugaskan untuk membangun model pembelajaran mesin yang mampu memprediksi kemungkinan kegagalan pompa berdasarkan data tersebut.

Laporan ini bertujuan untuk merinci proses pengerjaan tantangan Kaggle "PPM-Predict Failures Keep It Dry". Kami akan membahas metodologi yang digunakan untuk membangun model prediksi kegagalan pompa, menganalisis performanya, dan menarik kesimpulan berharga.

### 1.1 Permasalahan

Dalam pembelajaran Kaggle yang berjudul "PPM-Predict Failures Keep It Dry", berisi mengenai perusahaan fiksi Keep It Dry ingin meningkatkan produk utamanya, Super Soaker, yang digunakan di pabrik-pabrik untuk menyerap tumpahan dan kebocoran. Perusahaan ini telah menyelesaikan studi pengujian besar yang melibatkan berbagai prototipe produk dan mengumpulkan data tentang kinerja mereka. Tantangannya adalah menggunakan data ini untuk membangun model prediksi yang dapat mengidentifikasi kegagalan produk. Model yang efektif akan membantu Keep It Dry dalam meningkatkan kualitas dan keandalan Super Soaker, sehingga mengurangi resiko kegagalan di lapangan.

### 1.2 Overview Model

Dalam pengembangan model prediksi untuk meningkatkan kualitas dan keandalan produk Super Soaker dari perusahaan fiksi Keep It Dry. Tiga model yang sering digunakan dalam klasifikasi adalah K-Nearest Neighbors(KNN), Naive Bayes dan Decision Tree.

K-Nearest Neighbors(KNN) adalah algoritma yang menggunakan pendekatan berbasis jarak untuk mengklasifikasikan titik data baru berdasarkan mayoritas label dari K tetangga terdekatnya dalam ruang fitur. Algoritma ini sederhana dan tidak membuat asumsi kuat tentang distribusi data, tetapi dapat menjadi lambat dan kurang efektif dengan dataset besar atau berdimensi tinggi.

Naive Bayes adalah algoritma probabilistik yang memanfaatkan Teorema Bayes dengan asumsi independensi yang kuat antara fitur-fitur. Model ini cepat dan efisien, terutama untuk dataset besar, namun asumsi independensinya sering kali tidak realistis data, situasi praktis.

Decision Tree adalah model berbasis pohon keputusan yang memecah dataset secara rekursif berdasarkan fitur yang memberikan informasi paling baik. Model ini mudah dipahami dan diinterpretasikan, namun rentan terhadap overfitting dan bisa tidak stabil terhadap variasi kecil dalam data.

## 2 TINJAUAN PUSTAKA

### 2.1 DATA PRE-PROCESSING

*Data preprocessing* adalah langkah penting dalam proses *machine learning*, terutama dalam tugas klasifikasi. Data yang belum diolah seringkali mengandung *noise*, *missing values*, dan fitur yang tidak relevan yang dapat mengurangi kinerja model klasifikasi. Oleh karena itu, langkah-langkah *preprocessing* data seperti pembersihan data, transformasi data, dan pengurangan dimensi sangat penting untuk memastikan bahwa data yang digunakan untuk pelatihan model bersih, relevan, dan berkualitas tinggi [1].

Pembersihan data melibatkan identifikasi dan penanganan data yang hilang atau *outliers*. Data yang hilang dapat diatasi dengan berbagai teknik seperti penghapusan data, imputasi dengan *mean/median/mode*, atau menggunakan model prediktif untuk mengisi nilai yang hilang. Transformasi data meliputi normalisasi, standarisasi, dan *encoding* fitur kategorikal. Transformasi data bertujuan untuk menyamakan skala dan distribusi fitur sehingga algoritma *machine learning* dapat bekerja lebih efisien dan akurat.

Pengurangan dimensi adalah proses mengurangi jumlah fitur dalam dataset tanpa kehilangan informasi penting. Teknik seperti Principal Component Analysis (PCA) dan Linear Discriminant Analysis (LDA) sering digunakan untuk mengurangi kompleksitas dataset, menghilangkan fitur yang redundant, dan meningkatkan kinerja model.

### 2.2 CROSS VALIDATION

*Cross validation* adalah teknik yang digunakan dalam penilaian dari keakuratan suatu model dalam dataset tertentu. Metode K-Fold cross validation merupakan salah satu contoh dari penerapan teknik *cross validation*. Dalam metode tersebut, pertama-tama dataset dibagi dalam K-jumlah partisi. Pembagian tersebut dilakukan secara acak. Pada tiap pembagian tersebut, partisi ke-K berperan sebagai data uji sedangkan partisi yang lain berperan sebagai data latih. Langkah tersebut terus berulang sampai seluruh partisi berperan menjadi data latih dan data uji [3].

Nilai akurasi pada K-Fold cross validation diperoleh menggunakan persamaan berikut :

$$Akurasi = \frac{dataUjiBenar}{jumlahDataUji} \times 100\% \quad (1)$$

Selain akurasi, nilai rata-rata dari setiap pengujian juga dihitung menggunakan persamaan berikut :

$$RataRata = \frac{\sum Akurasi}{\sum UjiCoba} \times 100\% \quad (2)$$

### 2.3 K-NEAREST NEIGHBOR

K-Nearest Neighbor merupakan salah satu teknik dalam klasifikasi data yang memiliki cara kerja menghitung jarak dari data baru dengan beberapa data terdekat [4]. Dalam kata lain, pendekatan ini digunakan untuk mencari solusi dengan mengukur jarak terdekat antara kasus baru dan beberapa kasus lama, serta mencocokkannya berdasarkan bobot dari sejumlah data yang ada. Algoritma ini sering digunakan karena kemudahan dalam perhitungannya.

Berikut merupakan langkah-langkah dalam perhitungan metode K-Nearest Neighbor :

1. Menentukan parameter K
2. Menghitung jarak antara data uji dan data latih

Perhitungan jarak yang paling umum digunakan dalam K-Nearest Neighbor adalah euclidean distance, dengan persamaan sebagai berikut :

$$euclidean = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

dimana :

Pi = data latih

qi = data uji

i = variabel data

n = dimensi data

3. Mengurutkan jarak yang telah dihitung
4. Menentukan jarak terdekat sampai urutan ke-K
5. Melakukan metode voting mayoritas untuk menentukan kelas dari titik data baru. Kelas yang paling sering muncul di antara K tetangga terdekat akan menjadi prediksi untuk titik data baru.

### 2.4 NAIVE BAYES

Naive Bayes adalah salah satu teknik dalam klasifikasi data yang berdasarkan pada teorema Bayes. Metode ini mengasumsikan bahwa setiap fitur dalam data adalah independen satu sama lain. Meskipun asumsi ini sering kali tidak benar dalam kasus nyata (sehingga disebut "naive"), metode ini tetap efektif dan populer dalam banyak aplikasi klasifikasi, terutama dalam pengolahan teks dan klasifikasi spam.

Berikut adalah langkah-langkah dalam perhitungan metode Naive Bayes:

1. Menghitung probabilitas prior: Probabilitas prior adalah probabilitas kelas target sebelum kita melihat data. Ini dihitung dengan menghitung frekuensi relatif kelas target dalam dataset pelatihan.
2. Menghitung probabilitas likelihood: Probabilitas likelihood adalah probabilitas fitur tertentu diberikan kelas target. Ini dihitung dengan menghitung frekuensi relatif fitur dalam setiap kelas target.
3. Menghitung probabilitas posterior: Probabilitas posterior adalah probabilitas kelas target diberikan fitur. Ini dihitung dengan menggunakan teorema Bayes, yang menyatakan bahwa:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

dimana:

- P(A|B) adalah probabilitas posterior (probabilitas kelas target diberikan fitur)

- $P(B|A)$  adalah probabilitas likelihood (probabilitas fitur diberikan kelas target)
  - $P(A)$  adalah probabilitas prior (probabilitas kelas target)
  - $P(B)$  adalah probabilitas fitur
4. Klasifikasi: Untuk setiap titik data baru, kita menghitung probabilitas posterior untuk setiap kelas target dan menetapkan kelas dengan probabilitas tertinggi.

Naive Bayes memiliki beberapa keuntungan, seperti simpel, cepat, dan efisien dalam hal penggunaan memori. Selain itu, metode ini juga dapat bekerja dengan baik bahkan ketika asumsi independensi tidak sepenuhnya benar. Namun, Naive Bayes bisa kurang akurat jika fitur-fitur dalam data sangat bergantung satu sama lain.

## 2.5 DECISION TREE

Decision Tree adalah salah satu teknik dalam klasifikasi dan regresi data yang berdasarkan pada struktur pohon keputusan. Metode ini membagi data menjadi subset yang semakin kecil berdasarkan fitur tertentu, sementara pada saat yang sama, pohon keputusan yang terkait dengan subset tersebut juga semakin berkembang. Proses ini berlanjut sampai pohon keputusan mencapai tingkat maksimum yang ditentukan atau sampai tidak ada lagi peningkatan dalam nilai target.

Berikut adalah langkah-langkah dalam pembuatan Decision Tree:

1. Menentukan fitur yang akan digunakan sebagai node (simpul) pada pohon. Fitur ini biasanya dipilih berdasarkan kriteria tertentu, seperti Information Gain, Gain Ratio, atau Gini Index.
2. Membagi data menjadi subset berdasarkan fitur yang telah ditentukan. Proses ini menghasilkan cabang (branch) pada pohon keputusan.
3. Mengulangi proses di atas untuk setiap subset hingga mencapai tingkat maksimum yang ditentukan atau sampai tidak ada lagi peningkatan dalam nilai target.
4. Setelah pohon keputusan selesai dibuat, dapat digunakan untuk melakukan klasifikasi data baru dengan memulai dari node akar (root) dan bergerak ke bawah melalui pohon berdasarkan nilai fitur dari data baru.

Keuntungan dari Decision Tree adalah mudah untuk dipahami dan diinterpretasikan, dapat menangani data numerik dan kategorikal, dan mampu menangani masalah multi-output. Namun, Decision Tree bisa menjadi sangat kompleks dan cenderung overfit jika kedalaman pohon terlalu besar.

## 2.6 MATRIKS EVALUASI ROC

Matriks Evaluasi ROC (Receiver Operating Characteristic) adalah salah satu teknik dalam evaluasi performa model klasifikasi. Metode ini menggambarkan kemampuan model dalam membedakan antara kelas positif dan negatif di semua ambang klasifikasi.

Berikut adalah langkah-langkah dalam perhitungan Matriks Evaluasi ROC:

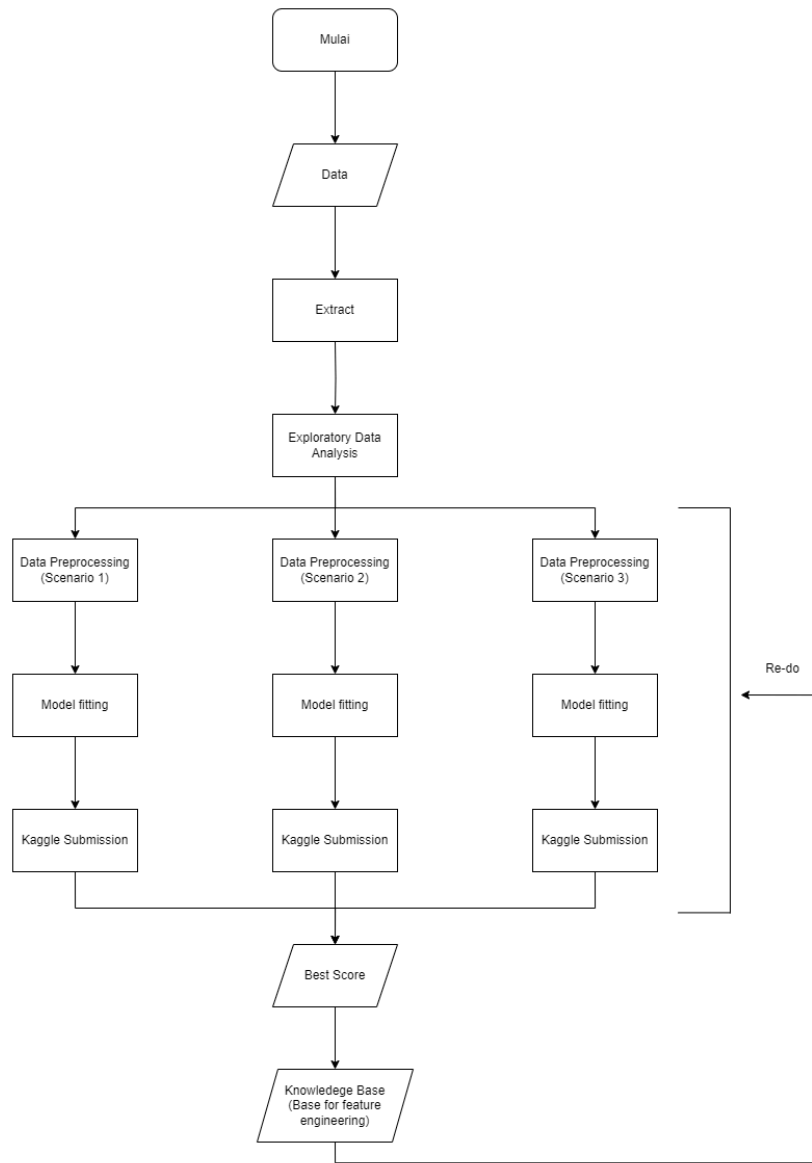
1. Menghitung probabilitas: Probabilitas bahwa sampel positif diprediksi sebagai positif (True Positive Rate, TPR) dan probabilitas bahwa sampel negatif diprediksi sebagai positif (False Positive Rate, FPR) dihitung untuk semua ambang klasifikasi.
2. Membuat plot: Plot dibuat dengan FPR pada sumbu x dan TPR pada sumbu y. Plot ini disebut kurva ROC.

3. Menghitung Area Under Curve (AUC): AUC adalah luas di bawah kurva ROC. Nilai AUC berkisar antara 0 dan 1, dengan nilai yang lebih tinggi menunjukkan model yang lebih baik.

Keuntungan dari Matriks Evaluasi ROC adalah dapat memberikan gambaran yang baik tentang seberapa baik model dapat membedakan antara kelas positif dan negatif. Selain itu, ROC dan AUC tidak dipengaruhi oleh distribusi kelas yang tidak seimbang. Namun, ROC dan AUC mungkin tidak informatif jika ada biaya yang sangat tinggi yang terkait dengan salah satu jenis kesalahan (misalnya, False Positives vs False Negatives).

### **3 METODOLOGI**

### 3.1 DIAGRAM ALUR



### 3.2 PENJELASAN

Proses ini dimulai dengan pengumpulan data dari Kaggle untuk tugas akhir saya yang dibimbing oleh dosen saya. Data tersebut kemudian diekstraksi fiturnya agar nilai-nilai dalam data tersebut dapat digunakan. Langkah selanjutnya adalah melakukan EDA (Exploratory Data Analysis) untuk mendapatkan wawasan dan ide-ide potensial untuk feature engineering.

Setelah itu, akan dilakukan tiga skenario Data Preprocessing dijalankan secara berkala untuk dimasukkan pada model. Setelahnya, model dilatih dan dievaluasi pada masing-masing skenario dengan data split dan cross-validation untuk memastikan keandalan. Prediksi disubmit ke Kaggle, berfokus pada prediksi kegagalan produk sambil menjaga kualitas.

Evaluasi model akan dilakukan dengan membagi dataset menjadi data pelatihan dan data validasi menggunakan data split, serta dengan menggunakan cross-validation untuk memastikan keandalan model. Prediksi akan disubmit ke Kaggle untuk pengujian, namun dalam konteks ini, kita berfokus pada prediksi kegagalan sambil berusaha menjaga agar tetap kering.

Skor terbaik yang diperoleh akan dicatat dan dimasukkan ke dalam basis pengetahuan yang berisi kumpulan ide-ide feature engineering yang telah terbukti efektif, namun dalam hal ini, kita mengevaluasi menggunakan ROC (Receiver Operating Characteristic). Proses ini adalah iteratif dan terus diperbaiki sepanjang waktu untuk terus meningkatkan skor.

Diagram ini menggambarkan proses tersebut dengan jelas, menunjukkan bahwa ini adalah proses yang berkelanjutan dan ditujukan untuk peningkatan berkelanjutan.

## **4 HASIL DAN PEMBAHASAN**

Pada bagian ini, kami akan menjabarkan teknik yang digunakan untuk ketiga model submisi kami.

### **4.1 DATA PREPROCESSING**

Data preprocessing merupakan tahap krusial dalam pengolahan data untuk mempersiapkannya sebelum digunakan dalam model machine learning. Proses pada prediksi kami adalah scaling, encoding, dan imputasi. Scaling dengan menggunakan teknik Power Transformer adalah salah satu teknik yang sering diterapkan untuk menormalkan distribusi data sehingga mendekati distribusi normal, yang dapat meningkatkan performa model terutama yang sensitif terhadap skala data maka dari itu kami menggunakan Power Transformer. Selain itu, encoding dilakukan untuk mengubah data kategorik menjadi bentuk numerik. Metode encoding yang akan kami bandingkan ada dua, yaitu One-Hot Encoding dan Label Encoding. One-Hot Encoding mengubah setiap kategori menjadi kolom biner yang terpisah, sangat efektif untuk data kategorik dengan jumlah kategori terbatas. Sementara itu, Label Encoding mengubah kategori menjadi nilai integer yang unik, cocok untuk data dengan hubungan ordinal antara kategorinya.

Imputasi kita juga lakukan untuk mengisi nilai pada *missing values*. Strategi yang diterapkan adalah imputasi dengan *mean* dan *median*. Imputasi dengan mean menggantikan nilai yang hilang dengan rata-rata dari fitur tersebut, sementara imputasi dengan median menggantikan dengan nilai tengah, yang sering lebih robust terhadap outlier.

Dengan membandingkan kombinasi scaling, encoding, dan imputasi yang tepat, data preprocessing dapat meningkatkan kualitas data dan membantu dalam membangun model machine learning yang lebih akurat dan efisien.



## 4.2 CROSS VALIDATION

Cross Validation kami lakukan untuk melakukan evaluasi terhadap hasil yang didapatkan dengan lebih baik tanpa ada khawatir terjadi overfitting. Cross validation ini dilakukan dengan K fold sebanyak 5. Lalu rata-rata dari lima skor dari setiap fold itu akan dihitung untuk dijadikan sebagai perbandingan.

## 4.3 HASIL

Table 1: Perbandingan Hasil Evaluasi Metode

Metode	ROC Score
<b>KNN</b>	
+ One hot encoding + Mean Imputer	0.5165895996207738
+ One hot encoding + Median Imputer	0.5168354577562194
+ <b>Label encoding + Mean Imputer</b>	<b>0.518051243307824</b>
+ Label encoding + Median Imputer	0.5174566149614785
<b>Naive Bayes</b>	
+ One hot encoding + Mean Imputer	0.5846833825685511
+ <b>One hot encoding + Median Imputer</b>	<b>0.584789261238449</b>
+ Label encoding + Mean Imputer	0.584683378709897
+ Label encoding + Median Imputer	0.5847892166666628
<b>Decision Tree</b>	
+ One hot encoding + Mean Imputer	0.5080368860693827
+ <b>One hot encoding + Median Imputer</b>	<b>0.5102504601750832</b>
+ Label encoding + Mean Imputer	0.5098050274021808
+ Label encoding + Median Imputer	0.5078194268784412

Dari data diatas yang ditampilkan pada Table 1, kami dapat menentukan beberapa kesimpulan seperti berikut:

- Model terbaik dalam kasus ini diantara tiga model yang dipilih adalah Naive Bayes dengan rata-rata ROC Score 0.584
- Kombinasi-kombinasi yang dilakukan tidak terlalu mempengaruhi hasil model
- Di antara kombinasi yang ada, kombinasi One hot encoding dan median imputer adalah kombinasi yang paling banyak menghasilkan nilai tertinggi

## 5 KESIMPULAN

Dalam laporan ini, kami telah membahas proses dan metodologi yang digunakan untuk memprediksi kegagalan produk Super Soaker dari perusahaan Keep It Dry dengan menggunakan tiga model pembelajaran mesin utama: K-Nearest Neighbors (KNN), Naive Bayes, dan Decision Tree. Setiap model memiliki karakteristik unik dengan kelebihan dan kekurangan dalam konteks analisis data dan prediksi.

Dari data yang ditampilkan pada Tabel 1, kami dapat menyimpulkan beberapa hal sebagai berikut:

- Model terbaik di antara tiga model yang dipilih adalah Naive Bayes, dengan rata-rata ROC Score 0.584.
- Kombinasi-kombinasi yang dilakukan tidak terlalu mempengaruhi hasil model.
- Di antara kombinasi yang ada, kombinasi One hot encoding dan median imputer adalah yang paling sering menghasilkan nilai tertinggi.

Langkah-langkah preprocessing data seperti pembersihan, transformasi, dan pengurangan dimensi sangat penting untuk meningkatkan kualitas data sebelum dimasukkan ke dalam model. Optimalisasi model melalui cross-validation dan hyperparameter tuning meningkatkan performa dan keandalan model. Evaluasi menggunakan matriks ROC memberikan gambaran yang baik tentang kemampuan model dalam membedakan antara kelas positif dan negatif.

Secara keseluruhan, pemilihan model terbaik sangat tergantung pada karakteristik dataset dan tujuan prediksi. Melalui iterasi dan evaluasi berkelanjutan, kami berupaya meningkatkan akurasi dan keandalan prediksi kegagalan produk, membantu perusahaan Keep It Dry dalam meningkatkan kualitas dan keandalan produk Super Soaker. Laporan ini memberikan panduan tentang penggunaan model pembelajaran mesin untuk prediksi kegagalan produk, serta pentingnya preprocessing data dan optimalisasi model dalam mencapai hasil yang diinginkan.

## DAFTAR PUSTAKA

- [1] Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques (3rd ed.). Morgan Kaufmann.
- [2] A. Zheng dan A. Casari, Feature Engineering for Machine Learning. Sebastopol: O'Reilly Media, Inc., 2018.
- [3] Hastie, T., Tibshirani, R., & Friedman, J. (2009). Springer Series in Statistics. The Elements of Statistical Learning, 27(2), 83–85.

- [4] Kusriani, Iuthfi Taufiq Emha, (2009), Algoritma Data Mining, Penerbit Andi, Yogyakarta.
- [5] Naive Bayes Classifier From Scratch in Python. Machine Learning Mastery. <https://machinelearningmastery.com/naive-bayes-classifier-scratch-python/> (diakses pada 19 Mei 2024).
- [6] Decision Tree Classifier in Python using Scikit-learn. DataCamp. <https://www.datacamp.com/community/tutorials/decision-tree-classification-python> (diakses pada 19 Mei 2024).
- [7] Support Vector Machines for Machine Learning. Machine Learning Mastery. <https://machinelearningmastery.com/support-vector-machines-for-machine-learning/> (diakses pada 19 Mei 2024).
- [8] Understanding AUC - ROC Curve. Towards Data Science. <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5> (diakses pada 19 Mei 2024).

## LAMPIRAN

Kode Github: <https://github.com/Y716/NoFreeWifi/tree/main>

### Pembagian Tugas:

1. Ananda Ravi Kuntadi:
  - a. Pembuatan laporan bagian tinjauan Pustaka & Metodologi
  - b. Melakukan percobaan menggunakan Hyperparameter optimization di KNN, Naive Bayes dan SVM
  - c. Mencoba menggunakan oversampling dan undersampling
2. Azril Januar Athallah:
  - a. Melakukan percobaan sendiri
  - b. Pembuatan laporan bagian kesimpulan
  - c. Pembuatan powerpoint presentasi
3. Davin Dalan Fidelio Fredra:
  - a. Pembuatan laporan bagian Pendahuluan.
  - b. Melakukan percobaan Iterative Group Classifier.
  - c.
4. Muhammad Yasin Hakim:
  - a. Melakukan percobaan dengan menggunakan data preprocessing yang bermacam
  - b. Melakukan Exploratory Data Analysis
  - c. Pembuatan laporan bagian Hasil dan Pembahasan
5. Ridha Ilham Adi Setyawan:
  - a. Pembuatan laporan bagian Hasil dan Pembahasan
6. Salsa Zufar Radinka Akmal:
  - a. Melakukan Exploratory Data Analysis
  - b. Melakukan prediksi probabilitas kegagalan menggunakan model Gaussian Naive Bayes, KNN, dan Decision Tree dengan beberapa iterasi untuk meningkatkan konsistensi hasil prediksi.
  - c. Pembuatan laporan bagian Tinjauan Pustaka & Metodologi