

**DETEKSI ANOMALI SENSOR MESIN INDUSTRI FARMASI  
MENGUNAKAN PSEUDO-LABELING DAN HYBRID  
LEARNING**



**Tugas Akhir**

**Jesse Laurencia**

**00000072698**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS TEKNIK DAN INFORMATIKA  
UNIVERSITAS MULTIMEDIA NUSANTARA  
TANGERANG  
2025**

**DETEKSI ANOMALI SENSOR MESIN INDUSTRI FARMASI  
MENGUNAKAN PSEUDO-LABELING DAN HYBRID  
LEARNING**



Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Komputer

**Jesse Laurencia**

**00000072698**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS TEKNIK DAN INFORMATIKA  
UNIVERSITAS MULTIMEDIA NUSANTARA  
TANGERANG**

**2025**

## HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Jesse Laurencia

Nomor Induk Mahasiswa : 00000072698

Program Studi : Sistem Informasi

Skripsi dengan judul:

DETEKSI ANOMALI SENSOR MESIN INDUSTRI FARMASI  
MENGUNAKAN PSEUDO-LABELING DAN HYBRID LEARNING

Merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 28 November 2025

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA



Jesse Laurencia

## **LEMBAR PERNYATAAN PENGGUNAAN BANTUAN KECERDASAN ARTIFISIAL (AI)**

Saya yang bertanda tangan di bawah ini:

Nama Lengkap : Jesse Laurencia

NIM : 00000072698

Program Studi : Sistem Informasi

Judul Skripsi : DETEKSI ANOMALI SENSOR MESIN  
INDUSTRI FARMASI MENGGUNAKAN PSEUDO-LABELING DAN  
HYBRID LEARNING

Dengan ini saya menyatakan secara jujur menggunakan bantuan Kecerdasan Artifisial (AI) dalam pengerjaan Tugas/Project/Tugas Akhir\*(coret salah satu) sebagai berikut (beri tanda centang yang sesuai):

- ☐ Menggunakan AI sebagaimana diizinkan untuk membantu dalam menghasilkan ide-ide utama saja
- ☐ Menggunakan AI sebagaimana diizinkan untuk membantu menghasilkan teks pertama saja
- ☒ Menggunakan AI untuk menyempurnakan sintaksis dan tata bahasa untuk pengumpulan tugas
- ☐ Karena tidak diizinkan: Tidak menggunakan bantuan AI dengan cara apa pun dalam pembuatan tugas

Saya juga menyatakan bahwa:

- (1) Menyerahkan secara lengkap dan jujur penggunaan perangkat AI yang diperlukan dalam tugas melalui Formulir Penggunaan Perangkat Kecerdasan Artifisial (AI)
- (2) Saya mengakui bahwa saya telah menggunakan bantuan AI dalam tugas saya baik dalam bentuk kata, paraphrase, penyertaan ide atau fakta penting yang disarankan oleh AI dan saya telah menyantumkan dalam sitasi serta referensi

Terlepas dari pernyataan di atas, tugas ini sepenuhnya merupakan karya saya sendiri

Tangerang, 28 November 2025  
Yang Menyatakan,



Jesse Laurencia

## HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama : Jesse Laurencia

NIM : 00000072698

Program Studi : Sistem Informasi

Jenjang : S1

Judul Karya Ilmiah : DETEKSI ANOMALI SENSOR MESIN INDUSTRI FARMASI MENGGUNAKAN PSEUDO-LABELING DAN HYBRID LEARNING

Menyatakan dengan sesungguhnya bahwa saya bersedia (**pilih salah satu**):

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) .
- ☐ Lainnya, pilih salah satu:
  - ☐ Hanya dapat diakses secara internal Universitas Multimedia Nusantara
  - ☐ Embargo publikasi karya ilmiah dalam kurun waktu 3 tahun.

Tangerang, 28 November 2025



Jesse Laurencia

Pilih salah satu

Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

## KATA PENGANTAR

Puji syukur saya panjatkan atas terselesaikannya laporan magang berjudul “Deteksi Anomali Sensor Mesin Industri Farmasi Menggunakan Pseudo-Labeling dan Hybrid Learning”. Laporan ini disusun sebagai bagian dari syarat kelulusan Program Studi Sistem Informasi, Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara. Pencapaian ini tidak terlepas dari dukungan berbagai pihak, makad dari itu saya mengucapkan terimakasih kepada

1. Ir. Andrey Andoko, M.Sc. Ph.D., selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti, S.Kom., selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara.
4. Bapak Ahmad Faza, S.Kom., M.T.I., selaku Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi sehingga terselesainya laporan ini.
5. Kak Tamaria Artpinkkan dan Reyhan Wijaya sebagai Pembimbing Lapangan yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya laporan tugas akhir ini.
6. Kepada Perusahaan PT. Kalbe Farma dan PT. Saka Farma
7. Kepada keluarga, teman – teman seperjuangan sahabat yang telah meluangkan waktu mendengarkan segala curahan hati sepanjang mengerjakan laporan ini.

Semoga Laporan Tugas Akhir ini dapat memberikan manfaat yang berguna bagi pembaca, khususnya dalam memahami dinamika dunia kerja yang professional.

Tangerang, 28 November 2025



Jesse Laurencia

# DETEKSI ANOMALI SENSOR MESIN INDUSTRI FARMASI MENGUNAKAN PSEUDO-LABELING DAN HYBRID LEARNING

Jesse Laurencia

## ABSTRAK

Peningkatan kebutuhan produk farmasi setiap tahun mendorong industri farmasi untuk menjaga keandalan mesin produksi guna meminimalkan risiko kehilangan produksi. Kompleksitas data sensor mesin yang bersifat tidak berlabel, tidak seimbang, dan memiliki pola non-linear menuntut sistem deteksi anomali yang akurat untuk mendukung Predictive Maintenance. Penelitian ini berfokus pada perancangan dan evaluasi model Machine Learning untuk mendeteksi anomali pada data sensor mesin produksi di PT Saka Farma. Dataset yang digunakan merupakan data publik IoT sensor dengan karakteristik yang menyerupai data operasional industri, meliputi fitur suhu, getaran, tekanan, dan konsumsi energi.

Penelitian ini mengevaluasi tiga pendekatan utama, yaitu model supervised, unsupervised, dan hybrid. Supervised digunakan untuk memilih classifier terbaik untuk kombinasi model *Hybrid*. Tahapan penelitian menggunakan framework CRISP-DM meliputi data preprocessing, normalisasi, serta penanganan imbalanced data. Hasil evaluasi model dilakukan menggunakan metrik seperti F1-score, Precision, Recall, PR-AUC dan Confidence Interval.

Hasil eksperimen menunjukkan model supervised XGBoost secara individu memberikan performa stabil dengan F1-score anomali sebesar 0.7932. Model unsupervised mampu mencapai nilai Recall tinggi namun menghasilkan Precision yang sangat rendah sehingga kurang ideal digunakan secara individu. Pendekatan hybrid menunjukkan kinerja terbaik secara keseluruhan. Kombinasi Isolation Forest dan XGBoost berhasil mencapai F1-score tertinggi sebesar 0.8456 dengan Precision 0.8939 dan Recall 0.8022, serta nilai PR AUC sebesar 0.8519 yang mencerminkan keseimbangan, stabilitas, dan kemampuan membedakan yang baik. Hasil ini menunjukkan bahwa penggabungan kemampuan deteksi awal dari model unsupervised dengan klasifikasi non-linear pada model supervised lebih efektif dalam menangani kompleksitas dan heterogenitas data sensor industri.

**Kata kunci:** *Anomaly detection, Manufacture Industry, Machine Learning, Hybrid*

# PSEUDO-LABELING AND HYBRID LEARNING FOR MACHINE SENSOR ANOMALY DETECTION IN THE PHARMACEUTICAL INDUSTRY

Jesse Laurencia

## ABSTRACT (English)

*The increasing demand for pharmaceutical products each year has prompted the pharmaceutical industry to maintain the reliability of production machinery to minimize the risk of production losses. The complexity of machine sensor data, which is unlabeled, unbalanced, and has non-linear patterns, requires an accurate anomaly detection system to support Predictive Maintenance. This study focuses on the design and evaluation of Machine Learning models to detect anomalies in production machine sensor data at PT Saka Farma. The dataset used is public IoT sensor data with characteristics like industrial operational data, including temperature, vibration, pressure, and energy consumption features.*

*This study evaluates three main approaches, namely supervised, unsupervised, and hybrid models. Supervised models are used to select the best classifier for the Hybrid model combination. The research stages use the CRISP-DM framework, including data preprocessing, normalization, and imbalanced data handling. Model evaluation results are performed using metrics such as F1-score, Precision, Recall, PR-AUC, and Confidence Interval.*

*The results of the experiment show that the supervised XGBoost model individually provides stable performance with an anomaly F1-score of 0.7932. The unsupervised model achieve a high Recall value but produces very low Precision, making it not ideal for individual use. The hybrid approach shows the best overall performance. The combination of Isolation Forest and XGBoost achieved the highest F1-score of 0.8456 with Precision of 0.8939 and Recall of 0.8022, as well as a PR AUC value of 0.8519, reflecting good balance, stability, and discrimination capabilities. These results show that combining the early detection capabilities of unsupervised models with non-linear classification in supervised models is more effective in handling the complexity and heterogeneity of industrial sensor data.*

**Keywords:** *Anomaly detection, Manufacture Industry, Machine Learning, Hybrid*

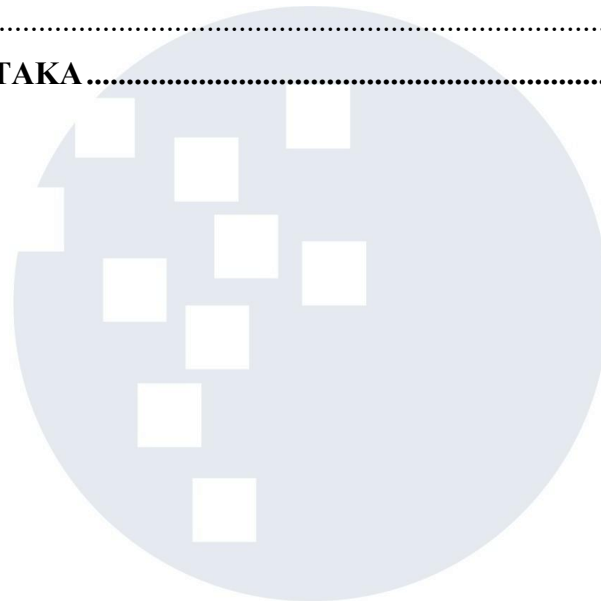


## DAFTAR ISI

|                                                                           |      |
|---------------------------------------------------------------------------|------|
| HALAMAN PERNYATAAN TIDAK PLAGIAT .....                                    | ii   |
| HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH .....                          | iv   |
| KATA PENGANTAR.....                                                       | v    |
| ABSTRAK .....                                                             | vi   |
| ABSTRACT (English) .....                                                  | vii  |
| DAFTAR ISI.....                                                           | viii |
| DAFTAR TABEL .....                                                        | xi   |
| DAFTAR GAMBAR.....                                                        | xii  |
| DAFTAR LAMPIRAN .....                                                     | xiv  |
| BAB I PENDAHULUAN.....                                                    | 1    |
| 1.1    Latar Belakang.....                                                | 1    |
| 1.2    Rumusan Masalah.....                                               | 4    |
| 1.3    Batasan Masalah .....                                              | 4    |
| 1.4    Tujuan dan Manfaat Penelitian .....                                | 5    |
| 1.4.1    Tujuan Penelitian.....                                           | 5    |
| 1.4.2    Manfaat Penelitian.....                                          | 5    |
| 1.5    Sistematika Penulisan .....                                        | 6    |
| BAB II LANDASAN TEORI .....                                               | 8    |
| 2.1    Penelitian Terdahulu .....                                         | 8    |
| 2.2    Teori tentang Topik Skripsi .....                                  | 11   |
| 2.2.1    Machine Learning.....                                            | 12   |
| 2.2.2    Anomaly Detection .....                                          | 13   |
| 2.2.3    Data Mining .....                                                | 14   |
| 2.3    Teori tentang Framework/Algoritma yang digunakan .....             | 15   |
| 2.3.1    CRISP-DM (Cross-Industry Standard Process for Data Mining) ..... | 15   |
| 2.3.2    DBSCAN.....                                                      | 18   |
| 2.3.3    Isolation Forest .....                                           | 19   |
| 2.3.4    AutoEncoder.....                                                 | 20   |
| 2.3.5    SVM .....                                                        | 21   |
| 2.3.6    XGBoost.....                                                     | 23   |
| 2.3.7    Random Forest.....                                               | 25   |

|                |                                                           |           |
|----------------|-----------------------------------------------------------|-----------|
| 2.3.8          | <i>SMOTE (Synthetic Minority Over-sampling Technique)</i> | 26        |
| 2.3.9          | <i>Class Weighting</i>                                    | 27        |
| 2.3.10         | <i>Bootstrap Resampling</i>                               | 28        |
| 2.3.11         | <i>Metrik Evaluasi</i>                                    | 30        |
| 2.4            | <b>Teori tentang Tools / Software yang digunakan</b>      | 34        |
| 2.4.1          | <i>Jupyter Notebook</i>                                   | 34        |
| 2.4.2          | <i>Google Collab</i>                                      | 34        |
| 2.4.3          | <i>Python</i>                                             | 35        |
| <b>BAB III</b> | <b>METODOLOGI PENELITIAN</b>                              | <b>37</b> |
| 3.1            | <b>Gambaran Umum Objek Penelitian</b>                     | 37        |
| 3.2            | <b>Metode Penelitian</b>                                  | 38        |
| 3.2.1          | <i>Metodologi</i>                                         | 38        |
| 3.2.2          | <i>Alur Penelitian</i>                                    | 40        |
| 3.3            | <b>Teknik Pengumpulan Data</b>                            | 47        |
| 3.4            | <b>Variabel Penelitian</b>                                | 47        |
| 3.5            | <b>Teknik Analisis</b>                                    | 48        |
| <b>BAB IV</b>  | <b>ANALISIS DAN HASIL PENELITIAN</b>                      | <b>50</b> |
| 4.1            | <b>Business Understanding</b>                             | 50        |
| 4.2            | <b>Data Understanding</b>                                 | 52        |
| 4.3            | <b>Data Preparation</b>                                   | 65        |
| 4.3.1          | <i>Bootstrapping</i>                                      | 65        |
| 4.3.2          | <i>Split Dataset</i>                                      | 70        |
| 4.3.3          | <i>Scaling</i>                                            | 73        |
| 4.4            | <b>Modelling</b>                                          | 74        |
| 4.4.1          | <i>Unsupervised</i>                                       | 74        |
| 4.4.2          | <i>Supervised</i>                                         | 84        |
| 4.4.3          | <i>Hybrid Approaches</i>                                  | 95        |
| 4.5            | <b>Evaluation</b>                                         | 111       |
| 4.5.1          | <i>Unsupervised</i>                                       | 113       |
| 4.5.2          | <i>Supervised</i>                                         | 116       |
| 4.5.3          | <i>Hybrid</i>                                             | 114       |
| 4.5.4          | <i>Model Individual vs Hybrid</i>                         | 121       |

|                                       |                                                                           |            |
|---------------------------------------|---------------------------------------------------------------------------|------------|
| 4.5.5                                 | <i>Penggunaan SMOTE (Synthetic Minority Over-Sampling Technique).....</i> | <i>123</i> |
| 4.5.6                                 | <i>Rekomendasi model terbaik.....</i>                                     | <i>124</i> |
| 4.6                                   | <b>Deployment .....</b>                                                   | <b>125</b> |
| <b>BAB V SIMPULAN DAN SARAN .....</b> |                                                                           | <b>127</b> |
| 5.1                                   | <b>Simpulan.....</b>                                                      | <b>127</b> |
| 5.2                                   | <b>Saran .....</b>                                                        | <b>128</b> |
| <b>DAFTAR PUSTAKA .....</b>           |                                                                           | <b>130</b> |



UMN  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## DAFTAR TABEL

|                                                                         |     |
|-------------------------------------------------------------------------|-----|
| Tabel 2.1 Perbandingan Penelitian Terdahulu .....                       | 8   |
| Tabel 3.1 Dataset Statistik.....                                        | 37  |
| Tabel 3.2 Deskripsi Atribut Dataset Sensor .....                        | 37  |
| Tabel 3.3 Perbandingan Metodologi .....                                 | 38  |
| Tabel 3.4 Perbandingan <i>Tools</i> .....                               | 48  |
| Tabel 3.5 Overview Dataset Asli Perusahaan.....                         | 53  |
| Tabel 3.6 Overview Dataset Skripsi .....                                | 53  |
| Tabel 3.7 Deskripsi Statistik Dataset Sensor .....                      | 53  |
| Tabel 3.8 Korelasi Fitur.....                                           | 59  |
| Tabel 3.9 Cluster Means .....                                           | 61  |
| Tabel 3.10 Missing Value.....                                           | 66  |
| Tabel 3.11 Perbandingan statistic data asli dan bootstrap .....         | 67  |
| Tabel 3.12 Distribusi Fault .....                                       | 68  |
| Tabel 3.13 Top 3 Parameter DBSCAN .....                                 | 77  |
| Tabel 3.14 Perbandingan param Isolation Forest.....                     | 79  |
| Tabel 3.15 Best Param RF.....                                           | 86  |
| Tabel 3.16 Best Param XGBoost .....                                     | 92  |
| Tabel 3.17 Perbandingan evaluasi metrik .....                           | 112 |
| Tabel 3.18 Perbandingan evaluasi metrik <i>Unsupervised</i> model ..... | 114 |
| Tabel 3.19 Perbandingan evaluasi metrik <i>Supervised</i> Model .....   | 116 |
| Tabel 3.20 Perbandingan evaluasi metrik <i>Hybrid</i> Model .....       | 118 |

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## DAFTAR GAMBAR

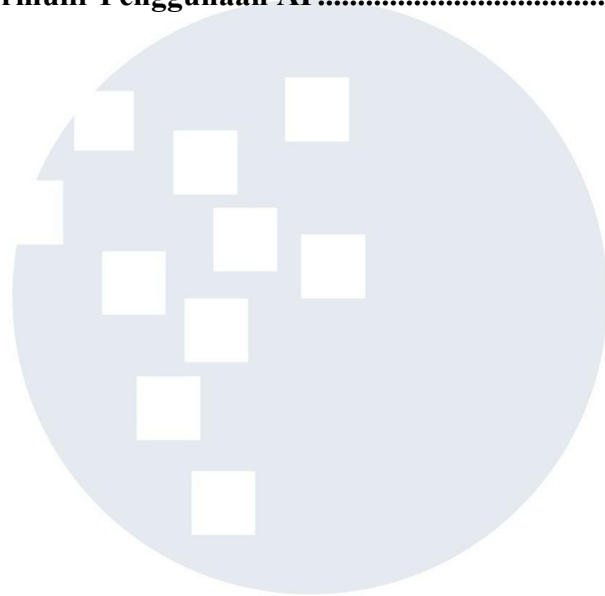
|                                                                                        |    |
|----------------------------------------------------------------------------------------|----|
| Gambar 3.1 Alur Penelitian .....                                                       | 41 |
| Gambar 3.2 Produk Kalbe Consumer Health (Saka Farma) .....                             | 50 |
| Gambar 3.3 Contoh Diagram HVAC Sistem .....                                            | 51 |
| Gambar 3.3 Deskripsi Atribut Dataset Sensor.....                                       | 56 |
| Gambar 3.4 Distribusi fault_flag dataset.....                                          | 57 |
| Gambar 3.5 Visualisasi PCA.....                                                        | 57 |
| Gambar 3.6 Heatmap .....                                                               | 58 |
| Gambar 3.7 Jumlah Cluster Optimal.....                                                 | 59 |
| Gambar 3.8 Clustering Anomali.....                                                     | 60 |
| Gambar 3.9 Visualisasi KDE data normal dan anomali .....                               | 63 |
| Gambar 3.10 Flow Data Preparation .....                                                | 65 |
| Gambar 3.11 Code Bootsrap Resampling.....                                              | 66 |
| Gambar 3.12 Visualisasi distribusi fitur asli dan bootstrap .....                      | 70 |
| Gambar 3.13 Code Split Dataset.....                                                    | 70 |
| Gambar 3.14 Code time-based split data <i>supervised</i> dan <i>unsupervised</i> ..... | 72 |
| Gambar 3.15 Output hasil split .....                                                   | 73 |
| Gambar 3.16 Isi variable X_unsuper_train1 .....                                        | 74 |
| Gambar 3.17 Code validation Split DBSCAN.....                                          | 75 |
| Gambar 3.18 Code param grid DBSCAN.....                                                | 76 |
| Gambar 3.19 Code cari parameter terbaik Isolation Forest .....                         | 78 |
| Gambar 3.20 Code pembuatan arsitektur AutoEncoder .....                                | 81 |
| Gambar 3.21 Code untuk EarlyStopping .....                                             | 81 |
| Gambar 3.22 Kurva Train Loss.....                                                      | 82 |
| Gambar 3.23 Grafik distribusi MSE.....                                                 | 83 |
| Gambar 3.24 Code pencarian optimal threshold .....                                     | 84 |
| Gambar 3.25 Code param grid Random Forest .....                                        | 85 |
| Gambar 3.26 Code GridSearch Random Forest.....                                         | 85 |
| Gambar 3.27 Code SMOTE RF.....                                                         | 87 |
| Gambar 3.28 Param grid SVM.....                                                        | 88 |
| Gambar 3.29 GridSearch SVM .....                                                       | 89 |
| Gambar 3.30 Code best param SVM .....                                                  | 90 |

|                                                                  |            |
|------------------------------------------------------------------|------------|
| <b>Gambar 3.31 Code Param Grid XGBoost.....</b>                  | <b>91</b>  |
| <b>Gambar 3.32 Code SMOTE RF.....</b>                            | <b>92</b>  |
| <b>Gambar 3.33 Code Fit Model XGBoost.....</b>                   | <b>94</b>  |
| <b>Gambar 3.34 Code <i>hybrid</i> Fit DBSCAN .....</b>           | <b>96</b>  |
| <b>Gambar 3.35 Code <i>hybrid</i> fit RF.....</b>                | <b>96</b>  |
| <b>Gambar 3.36 Code SMOTE <i>Hybrid</i> DBSCAN RF.....</b>       | <b>97</b>  |
| <b>Gambar 3.37 Code <i>Hybrid</i> SMOTE SVM .....</b>            | <b>99</b>  |
| <b>Gambar 3.38 Code <i>Hybrid</i> fit Isolation Forest .....</b> | <b>101</b> |
| <b>Gambar 3.39 Code <i>Hybrid</i> fit XGBoost .....</b>          | <b>102</b> |
| <b>Gambar 3.40 Code <i>Hybrid</i> XGBoost Best Param .....</b>   | <b>103</b> |
| <b>Gambar 3.41 Code <i>Hybrid</i> IF- XGB fit.....</b>           | <b>104</b> |
| <b>Gambar 3.42 Code SMOTE <i>Hybrid</i> AE-XGB.....</b>          | <b>107</b> |
| <b>Gambar 3.43 Code <i>Hybrid</i> AE-RF.....</b>                 | <b>109</b> |
| <b>Gambar 3.44 Code SMOTE <i>Hybrid</i> AE-RF .....</b>          | <b>111</b> |



## DAFTAR LAMPIRAN

|                                                     |            |
|-----------------------------------------------------|------------|
| <b>Lampiran A Turnitin Similarity Report.....</b>   | <b>130</b> |
| <b>Lampiran B Formulir Konsultasi Skripsi .....</b> | <b>131</b> |
| <b>Lampiran C Formulir Penggunaan AI .....</b>      | <b>132</b> |



UMN  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA