

**PERBANDINGAN KINERJA ALGORITMA SUPPORT  
VECTOR MACHINE (SVM) DAN RANDOM FOREST (RF)  
DENGAN TEKNIK RANDOM UNDERSAMPLING UNTUK  
MEMPREDIKSI RISIKO DIABETES MELITUS  
GESTASIONAL**

**LAPORAN NON-REGULER - SCIENTIST**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



Disusun oleh :  
**ANNISA DAMAYANTI**  
**21.11.4260**

**FAKULTAS ILMU KOMPUTER**  
**UNIVERSITAS AMIKOM YOGYAKARTA**  
**YOGYAKARTA**  
**2025**

**PERBANDINGAN KINERJA ALGORITMA SUPPORT  
VECTOR MACHINE (SVM) DAN RANDOM FOREST (RF)  
DENGAN TEKNIK RANDOM UNDERSAMPLING UNTUK  
MEMPREDIKSI RISIKO DIABETES MELITUS  
GESTASIONAL**

**LAPORAN NON-REGULER - SCIENTIST**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



Disusun oleh :

**ANNISA DAMAYANTI**

**21.11.4260**

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2025**

HALAMAN PERSETUJUAN

JALUR NON-REGULER

**PERBANDINGAN KINERJA ALGORITMA SUPPORT VECTOR MACHINE  
(SVM) DAN RANDOM FOREST (RF) DENGAN TEKNIK RANDOM  
UNDERSAMPLING UNTUK MEMPREDIKSI RISIKO  
DIABETES MELITUS GESTASIONAL**

yang disusun dan diajukan oleh

**Annisa Damayanti**

**21.11.4260**

telah disetujui oleh Dosen Pembimbing  
pada tanggal 11 Februari 2025

**Dosen Pembimbing,**



Anna Baita, S.Kom., M.Kom  
NIK. 190302290

**HALAMAN PENGESAHAN**

**JALUR NON-REGULER**

**PERBANDINGAN KINERJA ALGORITMA SUPPORT VECTOR MACHINE  
(SVM) DAN RANDOM FOREST (RF) DENGAN TEKNIK RANDOM  
UNDERSAMPLING UNTUK MEMPREDIKSI RISIKO  
DIABETES MELITUS GESTASIONAL**

yang disusun dan diajukan oleh

**Annisa Damayanti**  
21.11.4260

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 11 Februari 2025

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

Theophilus Bayu Sasongko, S.Kom., M.Eng  
NIK. 190302375

Ainul Yaqin, S.Kom., M.Kom  
NIK. 190302255

Anna Balta, S.Kom., M.Kom  
NIK. 190302290

Laporan ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 11 Februari 2025

**DEKAN FAKULTAS ILMU KOMPUTER**



Hanif Al Fatta, S.Kom., M.Kom., Ph.D.  
NIK. 190302096

**HALAMAN PERNYATAAN KEASLIAN KARYA**

Yang bertandatangan di bawah ini,

Nama mahasiswa : Annisa Damayanti

NIM : 21.11.4260

Menyatakan bahwa Laporan dengan judul berikut.

## Perbandingan Kinerja Algoritma Support Vector Machine (SVM) dan Random Forest (RF) dengan Teknik Random Undersampling untuk Memprediksi Risiko Diabetes Melitus Gestasional

Dosen Pembimbing : Anna Baiza, S.Kom., M.Kom.

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan **gagasan, rumusan** dan kegiatan **SAYA** sendiri, tanpa bantuan pihak lain kecuali arahan dari **Dosen Pembimbing**.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab **SAYA**, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini **SAYA** buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidak-benaran dalam pernyataan ini, maka **SAYA** bersedia menerima **SANKSI AKADEMIK** dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 11 Februari 2025

**Yang Menyatakan,**



Annisa Damayanti

## HALAMAN PERSEMBAHAN

Laporan ini dipersembahkan dengan penuh rasa hormat dan terima kasih kepada kedua orang tua penulis Bapak Muhamad Amin, S.Sos dan Ibu Masni yang selalu memberikan doa, dukungan, fasilitas, dan motivasi tanpa henti. Ucapan terima kasih juga disampaikan kepada dosen pembimbing dan penguji atas bimbingan serta ilmu yang telah diberikan dalam proses penyusunan laporan ini. Tidak lupa, apresiasi diberikan kepada sahabat dan rekan-rekan yang senantiasa memberikan dukungan dan semangat. Penulis berharap laporan ini dapat memberikan manfaat dan kontribusi yang positif.





## KATA PENGANTAR

Puji syukur kehadiran Allah SWT yang telah memberikan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan laporan skripsi yang berjudul **“Perbandingan Kinerja Algoritma Support Vector Machine (SVM) dan Random Forest (RF) dengan Teknik Random Undersampling untuk Memprediksi Risiko Diabetes Mellitus Gestasional”**. Laporan skripsi ini disusun untuk memenuhi syarat memperoleh gelar Sarjana Komputer Universitas Amikom Yogyakarta. Penulis menyampaikan rasa terima kasih yang tulus serta rasa apresiasi yang mendalam kepada semua pihak yang telah memberikan dukungan dan bantuan yang berharga selama proses penyelesaian skripsi ini. Dengan penuh rasa hormat, ucapan terima kasih yang sebesar-besarnya disampaikan kepada: Prof.

1. Dr. M. Suyanto., M.M selaku Rektor Universitas Amikom Yogyakarta.
2. Bapak Hanif Al-Fatta, M.Kom selaku Dekan Fakultas Ilmu Komputer.
3. Ibu Windha Mega Pradnya Duhita, M.Kom selaku Ketua Program Studi Informatika.
4. Ibu Anna Baita, M.Kom selaku Dosen Pembimbing yang telah memberikan ilmu, bimbingan, nasihat, dukungan serta solusi selama proses penyusunan jurnal hingga laporan skripsi ini.
5. Seluruh bapak ibu dosen serta staff karyawan Universitas Amikom Yogyakarta yang memberikan ilmunya selama masa perkuliahan.
6. Kedua orang tua penulis, Papa tercinta Muhamad Amin, S.Sos, Mama tercinta Masni dan adik tersayang Mery Kurniati yang selalu memberikan cinta, kasih sayang, doa, dukungan, semangat motivasi hingga semua fasilitas keperluan yang di perlukan penulis.
7. Mauhiba Salmaa Ghaisani selaku sahabat penulis yang selalu menemani, membantu, mendengarkan keluh kesah penulis, dan memberikan dukungan kepada penulis dari awal perkuliahan hingga penulisan skripsi ini.

8. Teman-teman seperjuangan, Yanti, Aulia, Titin, Ewi, dan teman-teman penulis lainnya yang tidak dapat disebutkan satu persatu, yang selalu membantu dan memberikan motivasi kepada penulis.
9. Terakhir, terima kasih untuk diri sendiri, yang sudah berada di titik ini dan mampu melewati lika-liku kehidupan selama perkuliahan di dunia rantau ini. Semoga setelah ini akan banyak kabar bahagia yang menanti aamiin.

Penulis menyadari bahwa dalam penyusunan laporan skripsi ini masih sangat jauh dari kata sempurna. Oleh karena itu, penulis memohon maaf atas segala kekurangan dalam penulisan skripsi ini. Semoga laporan skripsi ini dapat memberikan manfaat bagi penulis dan pembaca. Akhir kata, semoga Allah SWT, membalas segala kebaikan semua pihak yang telah berkontribusi dalam penyusunan laporan skripsi ini.

Yogyakarta, 11 Februari 2025

Penulis



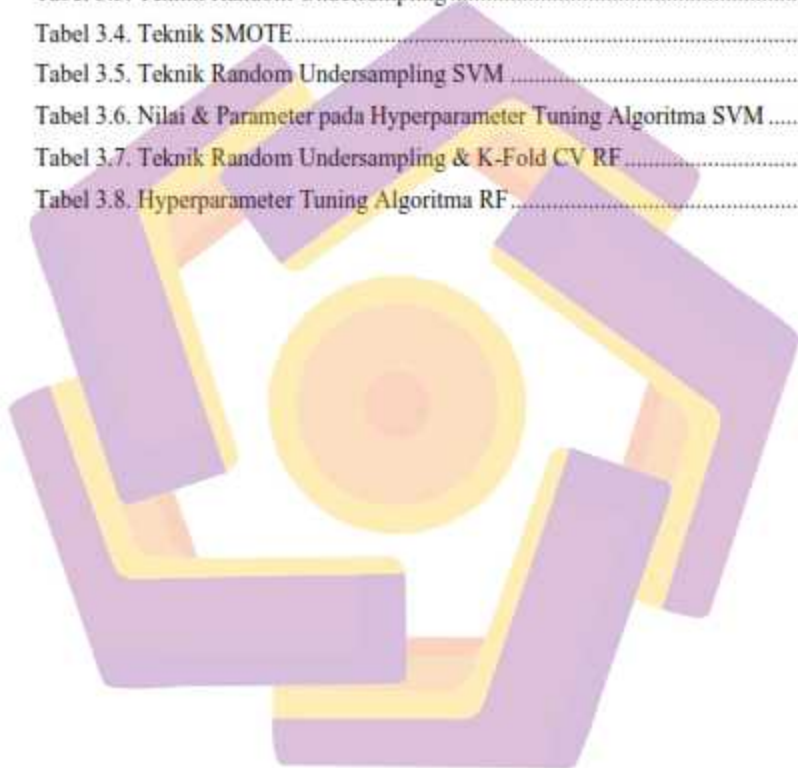
## DAFTAR ISI

|                                          |      |
|------------------------------------------|------|
| HALAMAN JUDUL.....                       | i    |
| HALAMAN PERSETUJUAN.....                 | ii   |
| HALAMAN PENGESAHAN.....                  | iii  |
| sHALAMAN PERNYATAAN KEASLIAN KARYA ..... | iv   |
| HALAMAN PERSEMBAHAN .....                | v    |
| KATA PENGANTAR .....                     | vi   |
| DAFTAR ISI.....                          | viii |
| DAFTAR TABEL.....                        | x    |
| DAFTAR GAMBAR.....                       | xi   |
| DAFTAR LAMPIRAN.....                     | xii  |
| DAFTAR LAMBANG DAN SINGKATAN .....       | xiii |
| DAFTAR ISTILAH .....                     | xiv  |
| INTISARI.....                            | xv   |
| ABSTRACT.....                            | xvi  |
| BAB I PENDAHULUAN.....                   | 1    |
| 1.1. Latar Belakang .....                | 1    |
| 1.2. Rumusan Masalah .....               | 2    |
| 1.3. Batasan Masalah.....                | 3    |
| 1.4. Tujuan .....                        | 3    |
| BAB II TINJAUAN PUSTAKA.....             | 4    |
| 2.1. Studi Literatur .....               | 4    |
| 2.2. Landasan Teori.....                 | 8    |
| 2.2.1. Preprocessing Data.....           | 8    |
| 2.2.2. Random Undersampling & SMOTE..... | 8    |
| 2.2.3. Support Vector Machine (SVM)..... | 8    |
| 2.2.4. Random Forest (RF).....           | 10   |
| 2.2.5. Hyperparameter Tuning .....       | 10   |
| 2.2.6. Cross Validation.....             | 11   |
| 2.2.7. Evaluasi.....                     | 11   |

|                                    |    |
|------------------------------------|----|
| BAB III METODE PENELITIAN.....     | 13 |
| 3.1. Metode.....                   | 13 |
| 3.2. Data Collection.....          | 13 |
| 3.3. Preprocessing Data.....       | 15 |
| 3.4. Handling Imbalance Data.....  | 16 |
| 3.5. Modeling.....                 | 19 |
| BAB IV KESIMPULAN .....            | 26 |
| 4.1. Kesimpulan .....              | 26 |
| 4.2. Saran.....                    | 27 |
| 4.3. Ucapan Terimakasih.....       | 27 |
| REFERENSI .....                    | 28 |
| CURRICULUM VITAE.....              | 32 |
| LAMPIRAN DAN BUKTI PENDUKUNG ..... | 33 |
| a. Letter of Acceptance (LOA)..... | 33 |
| b. Lembar Review .....             | 34 |
| c. Bukti Terbit/Terindex.....      | 35 |
| d. Sertifikat sebagai Penyaji..... | 36 |
| e. Bukti pembayaran .....          | 37 |

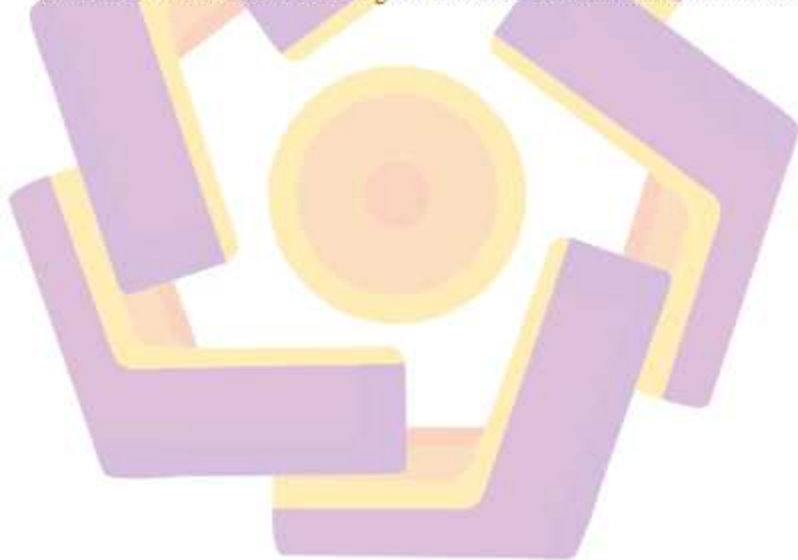
## DAFTAR TABEL

|                                                                             |    |
|-----------------------------------------------------------------------------|----|
| Tabel 2.1 Studi Literatur .....                                             | 4  |
| Tabel 2.2.1 Confusion Matrix .....                                          | 11 |
| Tabel 3.1. Identifikasi Data .....                                          | 14 |
| Tabel 3.2. Hasil Transformasi Data dengan RFE .....                         | 15 |
| Tabel 3.3. Teknik Random Undersampling .....                                | 16 |
| Tabel 3.4. Teknik SMOTE.....                                                | 17 |
| Tabel 3.5. Teknik Random Undersampling SVM .....                            | 19 |
| Tabel 3.6. Nilai & Parameter pada Hyperparameter Tuning Algoritma SVM ..... | 20 |
| Tabel 3.7. Teknik Random Undersampling & K-Fold CV RF .....                 | 21 |
| Tabel 3.8. Hyperparameter Tuning Algoritma RF .....                         | 22 |



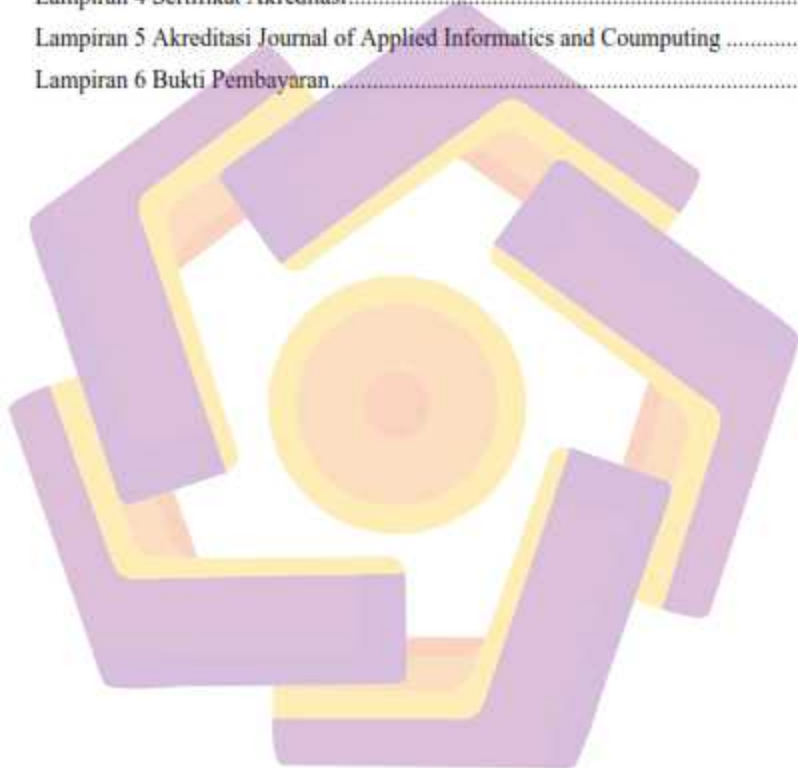
## DAFTAR GAMBAR

|                                                                        |    |
|------------------------------------------------------------------------|----|
| Gambar 3.1. Alur Penelitian.....                                       | 13 |
| Gambar 3.2. Bar Chart Distribusi Class Diabetes_012.....               | 14 |
| Gambar 3.3. Visualisasi jumlah data setelah random undersampling ..... | 17 |
| Gambar 3.4. Visualisasi jumlah data setelah SMOTE .....                | 18 |
| Gambar 3.5. Hasil pemodelan ML Random Undersampling.....               | 18 |
| Gambar 3.6. Hasil pemodelan ML SMOTE .....                             | 18 |
| Gambar 3.7. Pengecekan Overfitting & Underfitting SVM .....            | 21 |
| Gambar 3.8. Pengecekan Overfitting & Underfitting RF.....              | 23 |
| Gambar 3.9. Confusion Matrix Algoritma SVM & RF.....                   | 23 |
| Gambar 3.10. Curva ROC-AUC Algoritma SVM & RF.....                     | 24 |



## DAFTAR LAMPIRAN

|                                                                           |    |
|---------------------------------------------------------------------------|----|
| Lampiran 1 LoA dari Journal of Applied Informatics and Computing .....    | 33 |
| Lampiran 2 Hasil Review dari Reviewer A & B .....                         | 35 |
| Lampiran 3 Accept Submission Artikel.....                                 | 36 |
| Lampiran 4 Sertifikat Akreditasi.....                                     | 36 |
| Lampiran 5 Akreditasi Journal of Applied Informatics and Coumputing ..... | 37 |
| Lampiran 6 Bukti Pembayaran.....                                          | 37 |



## DAFTAR LAMBANG DAN SINGKATAN



|         |                                                                 |
|---------|-----------------------------------------------------------------|
| SVM     | <i>Support Vector Machines</i>                                  |
| RF      | <i>Random Forest</i>                                            |
| DMG     | <i>Diabetes Melitus Gestasional</i>                             |
| RBF     | <i>Radial Basis Function</i>                                    |
| RFE     | <i>Recursive Feature Elimination</i>                            |
| SMOTE   | <i>Synthetic Minority Over-sampling Technique</i>               |
| CV      | <i>Cross Validation</i>                                         |
| TN      | <i>True Negative</i>                                            |
| FP      | <i>False Positive</i>                                           |
| TP      | <i>True Positive</i>                                            |
| FN      | <i>False Negative</i>                                           |
| OvR     | <i>One-vs-Rest</i>                                              |
| ROC-AUC | <i>Receiver Operating Characteristic – Area Under the Curve</i> |
| $\beta$ | <i>Beta</i>                                                     |
| TPR     | <i>True Positive Rate</i>                                       |
| FPR     | <i>False Positive Rate</i>                                      |



## DAFTAR ISTILAH

|                       |                                                                                                                                          |
|-----------------------|------------------------------------------------------------------------------------------------------------------------------------------|
| K-Fold                | Teknik untuk mengevaluasi performa model pembelajaran mesin.                                                                             |
| Hyperparameter Tuning | Proses mencari kombinasi nilai yang optimal untuk parameter-parameter tertentu yang tidak di pelajari oleh model selama proses training. |
| Akurasi               | Nilai hasil untuk menentukan keakuratan model.                                                                                           |
| Confusion Matrix      | Tabel yang menggambarkan kinerja sampel klasifikasi yang digunakan untuk mengukur kinerja pada kumpulan data.                            |
| Precision             | Tingkat probabilitas dengan hasil positif dan ditentukan dengan benar.                                                                   |
| Recall                | Tingkat probabilitas dengan yang hasil positif dan ditentukan dengan benar.                                                              |
| F1-Score              | Nilai hasil untuk mengetahui nilai rata-rata dari perbandingan keluaran <i>precision</i> maupun keluaran <i>recall</i> .                 |
| True Negative (TN)    | Prediksi hasil yang bernilai negatif, data yang sebenarnya negatif.                                                                      |
| False Positive (FP)   | Prediksi hasil yang bernilai positif, data yang sebenarnya negatif.                                                                      |
| False Negative (FN)   | Prediksi hasil yang bernilai negatif, data yang sebenarnya positif.                                                                      |
| True Positive (TP)    | Prediksi hasil yang bernilai positif, data sebenarnya yang positif.                                                                      |
| Grid Search           | Proses pengoptimalan nilai akurasi dengan menemukan parameter terbaik.                                                                   |
| Scoring Metrics       | Alat evaluasi yang digunakan untuk menilai kinerja model.                                                                                |

## INTISARI

Diabetes Melitus Gestasional (DMG) adalah suatu kondisi intoleransi glukosa yang berkembang selama masa kehamilan hingga proses persalinan, yang ditandai dengan peningkatan kadar gula darah yang tidak normal. Diagnosis dini yang akurat sangat penting untuk memberikan informasi yang dapat mempercepat proses pengobatan dan mengurangi komplikasi pada ibu dan bayi. Salah satu metode machine learning yang dapat digunakan untuk memprediksi DMG adalah Support Vector Machine (SVM) dan Random Forest (RF). Penelitian ini bertujuan untuk membandingkan model prediksi DMG menggunakan algoritma SVM dan RF dengan menyeimbangkan data target menggunakan Teknik Random Undersampling. Pendekatan dengan menggunakan teknik ini berhasil meningkatkan akurasi sebesar 18% dari akurasi sebelum menggunakan random undersampling. Model SVM pada penelitian ini menggunakan tuning hyperparameter dengan parameter kernel, C (cost), dan gamma, sedangkan model RF menggunakan Scoring Metrix dan empat parameter lainnya, yaitu N estimator, max\_depth, min\_samples\_split, dan min\_samples\_leaf. Proses pencarian parameter terbaik dilakukan dengan menggunakan GridSearchCV pada kedua model. Hasil penelitian menunjukkan bahwa model klasifikasi SVM dengan teknik random undersampling dan tuning hyperparameter dengan K-Fold mencapai akurasi dengan nilai precision, recall, f1-score rata-rata 100%, dengan parameter terbaik Kernel Linier, nilai  $C=0.1$  dan nilai  $\gamma=0.001$  dengan akurasi tertinggi yaitu 1.0, nilai ROC-AUC 99% yang mengindikasikan performa prediksi yang sangat baik. Sementara model RF menunjukkan hasil akurasi 99%, tuning juga dilakukan dengan menggunakan parameter yang sesuai menghasilkan akurasi yang sama 99%, nilai ROC-AUC 99%. Kedua model menunjukkan algoritma SVM dan RF memiliki performa prediksi sangat baik dalam memprediksi DMG, namun algoritma SVM memprediksi DMG lebih baik dibandingkan RF karena jumlah kesalahan prediksi lebih rendah.

**Kata kunci:** DMG, prediksi, RF, random undersampling, SVM

## ABSTRACT

Gestational diabetes mellitus (GDM) is a condition of glucose intolerance that develops during pregnancy until delivery, characterised by abnormally elevated blood sugar levels. Accurate early diagnosis is essential to provide information that can speed up the treatment process and reduce complications for mother and baby. One of the machine learning methods that can be used to predict DMG is Support Vector Machine (SVM) and Random Forest (RF). This study aims to compare GDM prediction models using SVM and RF algorithms by balancing the target data using the Random Undersampling Technique. The approach using this technique succeeded in increasing accuracy by 18% from accuracy before using random undersampling. The SVM model in this study uses hyperparameter tuning with kernel, C (cost), and gamma parameters, while the RF model uses Scoring Matrix and four other parameters, namely N\_estimator, max\_depth, min\_samples\_split, and min\_samples\_leaf. The process of finding the best parameters was carried out using GridSearchCV on both models. The results showed that the SVM classification model with random undersampling technique and hyperparameter tuning with K-Fold achieved accuracy with an average precision, recall, f1-score value of 100%, with the best parameters Linear Kernel, C value = 0.1 and gamma value = 0.001 with the highest accuracy of 1.0, ROC-AUC value of 99% which indicates excellent prediction performance. While the RF model shows 99% accuracy results, tuning is also done using the appropriate parameters resulting in the same accuracy of 99%, ROC-AUC value of 99%. Both models show that SVM and RF algorithms have excellent prediction performance in predicting GDM, but SVM algorithm predicts GDM better than RF because the number of prediction errors is lower.

**Keyword:** GDM, prediction, RF, random undersampling, SVM