

Klasifikasi Penipuan Transaksi Kartu Kredit Menggunakan Metode Random Forest

Tiara Suci Lestari*, Dwi Agustin Nuriani Sirodj

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

*tiarasucilstr@gmail.com, dwi.agustinns@gmail.com

Abstract. In today's technological developments, the use of credit cards is a very easy and practical way for customers to make transactions. However, with the increasing use of credit cards, it will lead to financial fraud, namely fraudulent credit card transactions that can harm customers and the bank or company. One technique that can overcome this problem is data mining techniques, namely the classification used to predict fraudulent actions in credit card transactions. The method used is the random forest method, which is an ensemble method by applying bootstrap aggregating (bagging) and random feature selection, which combines several decision trees to form a forest, then to get the results of the final classification prediction through a voting process. The data used is credit card transaction fraud data for 2019-2020. The purpose of the results of this study is to apply the random forest method to the classification of credit card transaction fraud based on the evaluation of classification accuracy such as confusion matrix, accuracy, sensitivity, precision, f-measure and AUC value. The results of the study showed that the application of the random forest method gave very good classification results in classifying fraudulent credit card transactions.

Keywords: *Random Forest, Classification, Credit Card.*

Abstrak. Pada perkembangan teknologi saat ini, penggunaan kartu kredit merupakan cara yang sangat mudah dan praktis digunakan pelanggan dalam melakukan transaksi. Tetapi dengan meningkatnya penggunaan kartu kredit maka akan menimbulkan kecurangan finansial yaitu penipuan transaksi kartu kredit yang dapat merugikan nasabah maupun pihak bank atau perusahaan. Salah satu teknik yang dapat mengatasi masalah tersebut yaitu teknik data mining yaitu klasifikasi yang digunakan untuk memprediksi tindakan penipuan pada transaksi kartu kredit. Metode yang digunakan yaitu metode random forest yang merupakan metode ensemble dengan menerapkan bootstrap aggregating (bagging) dan random feature selection yaitu menggabungkan beberapa pohon keputusan sehingga membentuk hutan (forest), kemudian untuk mendapatkan hasil dugaan klasifikasi akhir melalui proses voting. Data yang digunakan yaitu data penipuan transaksi kartu kredit tahun 2019-2020. Tujuan hasil dari penelitian ini yaitu menerapkan metode random forest pada klasifikasi penipuan transaksi kartu kredit berdasarkan evaluasi ketepatan klasifikasi seperti seperti confusion matrix, akurasi, sensitivitas, presisi, f-measure dan nilai AUC. Hasil dari penelitian didapatkan bahwa penerapan metode random forest memberikan hasil klasifikasi yang sangat baik dalam mengklasifikasikan penipuan transaksi kartu kredit.

Kata Kunci: *Random Forest, Klasifikasi, Kartu Kredit.*

A. Pendahuluan

Kartu kredit merupakan suatu alat pembayaran yang menggunakan kartu sebagai pengganti pembayaran uang tunai dimana fasilitas yang diberikan yaitu berupa kredit yang mana pada saat jatuh tempo dapat dibayarkan dengan jumlah minimum dan sisanya dijadikan sebagai kredit [7]. Menurut Asosiasi Kartu Kredit Indonesia (AKKI) [1] jumlah penggunaan kartu kredit di Indonesia mengalami peningkatan yang signifikan yaitu dari tahun 2009 sampai 2020, dimana pada tahun 2009 penggunaan kartu kredit di Indonesia sebanyak 12.259.295 dan pada tahun 2020 menjadi 17.469.264. Maka dari itu, dengan meningkatnya penggunaan kartu kredit dan transaksi yang dilakukan maka akan meningkatkan tindakan kriminal berupa kecurangan finansial yaitu salah satunya penipuan kartu kredit. Maka dari itu untuk mengatasi tindakan tersebut harus dilakukan deteksi penipuan untuk mencegah aktivitas penipuan agar tidak membuat kerugian besar bagi pihak bank atau perusahaan dan pemegang kartu kredit, selain itu juga untuk memantau perilaku transaksi pemegang kartu kredit, apakah transaksi yang dilakukan termasuk kedalam kategori penipuan atau tidak.

Untuk mengatasi masalah tersebut dapat dipecahkan menggunakan teknik data mining yaitu klasifikasi. Tujuan dari klasifikasi yaitu untuk memprediksi label kelas dari suatu objek berdasarkan atribut yang ada. Metode yang termasuk kedalam klasifikasi diantaranya yaitu metode *random forest*. *Random forest* mampu mengatasi masalah non-linier karena didasarkan pada teknik pohon keputusan (*decision tree*). Konsep dasar *random forest* yaitu menggunakan lebih dari *classifier* dari metode yang sama kemudian mengkombinasikannya melalui *voting* untuk mendapatkan hasil dugaan klasifikasi akhir [2]. *Random forest* memiliki kelebihan yaitu diantaranya dapat memberikan hasil klasifikasi yang baik disertai dengan hasil residu/*error* yang lebih rendah, dapat mengatasi data training yang berukuran sangat besar secara efisien, selain itu metode *random forest* juga dinilai efektif dalam mengatasi masalah *missing data*. Berdasarkan penelitian yang dilakukan oleh Muttaqin [3] yaitu membandingkan ketiga metode *ensemble* yaitu *random forest*, *Boosting* CART dan *Bagging* CART yang berguna untuk memperbaiki prediksi pohon klasifikasi CART, dimana hasil penelitian menyebutkan bahwa ketiga metode *ensemble* memberikan hasil akurasi yang lebih baik dibandingkan metode pohon klasifikasi CART. Diantara ketiga metode *ensemble* tersebut, metode *random forest* memberikan hasil akurasi prediksi yang lebih baik dibandingkan dengan dua metode lainnya yaitu metode *Boosting* CART dan *Bagging* CART yang diterapkan pada data klasifikasi rumah tangga miskin di Kabupaten Jombang.

Maka dari itu berdasarkan penelitian sebelumnya maka peneliti ingin menerapkan metode *random forest* pada klasifikasi penipuan transaksi kartu kredit berdasarkan evaluasi ketepatan klasifikasi dengan harapan metode *random forest* dapat memberikan hasil klasifikasi *fraud* atau *non-fraud* dengan benar.

B. Metodologi Penelitian

Metode yang digunakan dalam penelitian ini yaitu metode *random forest* yang diterapkan pada data penipuan transaksi kartu kredit yang yang diperoleh dari situs kaggle dan disediakan oleh Kartik Shenoy. Data tersebut berisikan informasi simulasi transaksi kartu kredit yang sah dan penipuan dari 1 Januari 2019 – 31 Desember 2020 didalamnya mencakup 996 pelanggan yang melakukan transaksi ke 693 pedagang dengan jumlah data sebanyak 50.208 transaksi. Variabel yang digunakan yaitu sebanyak 10 variabel dimana 9 variabel merupakan variabel *independent* dan 1 variabel yang merupakan variabel *dependent* seperti yang disajikan pada Tabel 1.

Tabel 1. Variabel Data Transaksi Penipuan Kartu Kredit

Variabel	Skala Pengukuran	Keterangan
Trans_date_trans_time	Nominal	Waktu transaksi
Cc_num	Nominal	Nomor kartu kredit pelanggan
Merchant	Nominal	Nama pedagang

Category	Nominal	Kategori pedagang
Amt	Rasio	Jumlah transaksi (\$)
City	Nominal	Nama kota dari pemegang kartu kredit
State	Nominal	Nama provinsi dari pemegang kartu kredit
Job	Nominal	Pekerjaan dari pemegang kartu kredit
Age	Nominal	Usia dari pemegang kartu kredit
Is_fraud	Nominal	Kelas target penipuan atau tidak

Klasifikasi Random forest

Metode *Random Forest* menerapkan metode *bootstrap aggregating (bagging)* dan *random feature selection* yang merupakan perkembangan dari metode *Classification and Regression Trees (CART)* [4]. Klasifikasi *random forest* dilakukan melalui kombinasi pohon dengan dilakukannya percobaan pada sampel yang telah disediakan. Dalam penentuan kelas klasifikasi pada *Random Forest* yaitu berdasarkan *voting* dari beberapa pohon keputusan yang telah terbentuk dimana pemenang dari pohon tersebut ditentukan dengan *vote* yang paling banyak (*majority vote*).

Random Forest membutuhkan dua parameter, yaitu jumlah *tree* yang akan dipakai dan jumlah variabel *independent* yang digunakan untuk proses pencabangan agar mencapai nilai yang optimal. Untuk mendapatkan nilai *mtry* yaitu disarankan menggunakan persamaan sebagai berikut:

$$mtry_1 = \frac{1}{2} \lfloor \sqrt{p} \rfloor \quad (1)$$

$$mtry_2 = \lfloor \sqrt{p} \rfloor \quad (2)$$

$$mtry_3 = 2 \times \lfloor \sqrt{p} \rfloor \quad (3)$$

Dimana:

mtry : banyaknya variabel independent untuk setiap split

p : banyaknya variabel independent

Algoritma Random Forest untuk klasifikasi yaitu sebagai berikut [4]:

1. Proses *bootstrap sample* yaitu melakukan penarikan sampel acak berukuran *N* pengamatan dari data *training* dengan pengembalian (*replacement*).
2. Menggunakan sampel *bootstrap (BS_j)* sebagai data *training* untuk membangun pohon ke-*j* (*j* = 1, 2, ..., *J*), setiap pohon tumbuh secara maksimal tidak dilakukan pemangkasan. pembangunan pohon klasifikasi dilakukan dengan menerapkan *random future selection* dengan mengulangi langkah-langkah untuk setiap *node* terminal pohon, sampai ukuran *node* minimum tercapai.
 - (i) Pilih variabel *m* secara acak dari variabel prediktor *p* dengan *mtry* < *p*.
 - (ii) Pilih variabel/split terbaik diantara semua kemungkinan *mtry* fitur dari proses (i) sebagai penyekat.
 - (iii) Pisahkan *node* menjadi dua *sub – node* berdasarkan solusi *split* yang diperoleh dari proses (ii).
3. Ulangi langkah 1-3 sebanyak *j* kali sehingga terbentuk sebuah hutan yang terdiri dari *j* pohon. Tiap pohon klasifikasi akan menghasilkan satu suara hasil prediksi sehingga akan didapatkan *j* buah suara. Penentuan hasil prediksi akhir yaitu dengan cara mengaggregasi atau menggabungkan hasil prediksi setiap pohon dengan menggunakan pemungutan suara terbanyak (*majority vote*).

Terdapat dua alasan mengapa *random forest* menggunakan teknik *bagging (bootstrap aggregating)*, dimana yang pertama dapat meningkatkan akurasi dan yang kedua dapat digunakan untuk mengestimasi kesalahan generalisasi dari *ensemble* pohon, dengan kata lain estimasi *strength* dan korelasi atau disebut OOB. Data OOB digunakan dijadikan data untuk memvalidasi pada pohon yang bersesuaian. Perkiraan tingkat kesalahan yang diduga melalui *error OOB (Out-Of-Bag)* berdasarkan data pelatihan yaitu sebagai berikut [5]:

1. Pada setiap iterasi *bootstrap*, lakukan prediksi data yang tidak termuat kedalam sampel

bootstrap atau yang disebut dengan OOB menggunakan pohon yang ditumbuhkan dengan sampel *bootstrap*.

2. Gabungkan prediksi OOB. Secara rata-rata, setiap gugus data asli akan menjadi *out-of-bag* sekitar 36.8 % untuk setiap pohon, Maka dari itu, pada langkah 1, masing-masing amatan gugus data asli mengalami prediksi sekitar sepertiga kali dari banyaknya pohon. Misalkan *a* merupakan sebuah amatan dari data asli, maka hasil prediksi *random forest* terhadap *a* adalah gabungan dari hasil prediksi setiap kali *a* menjadi data *out-of-bag*.
3. Hitung tingkat kesalahan (*error* OOB) dari proporsi misklasifikasi hasil prediksi yang terdapat pada seluruh amatan gugus data asli.

Breiman [5] menyarankan untuk mengamati *error* OOB yang lebih teliti ketika *k* kecil, kemudian memilih *m* yang menghasilkan *error* OOB terkecil. Sehingga kinerja *random forest* dapat dilihat berdasarkan jumlah pohon dan *mtry* yang memiliki nilai *error* OOB yang paling kecil untuk mendapatkan parameter optimal. Ketika metode *random forest* dilakukan untuk menghasilkan *variable importance*, maka disarankan untuk menggunakan *tree* sebanyak 1000 atau lebih, agar *tree* yang dibangun semakin banyak sehingga peubah *respon* yang dianalisis pun semakin banyak, sehingga nilai tersebut dapat lebih besar dan *variable importance* yang dihasilkan semakin stabil. Penggunaan jumlah pohon lebih dari 100 cenderung menghasilkan tingkat misklasifikasi yang rendah.

Evaluasi model

Ketika model klasifikasi telah terbentuk, maka harus dilakukan evaluasi keakuratan pengklasifikasian agar memperoleh ketepatan klasifikasi yang akurat dan baik. Ukuran yang digunakan untuk mengukur kinerja klasifikasi yaitu *confusion matrix*. *Confusion matrix* dapat digunakan untuk mengevaluasi kualitas *classifier* yaitu seberapa baik *classifier* dapat mengenali tupel/fitur dari kelas yang berbeda yang disajikan dalam bentuk tabulasi silang dimana kelas data hasil prediksi ditampilkan sebagai kolom matriks dan kelas data aktual ditampilkan sebagai baris matriks [8].

Tabel 2. Confusion Matrix

		<i>Predicted Class</i>		
		<i>Positive</i>	<i>Negative</i>	Total
<i>Actual Class</i>	<i>Positive</i>	<i>True Positive</i> (TP)	<i>False Negative</i> (FN)	TP+FN
	<i>Negative</i>	<i>False Positive</i> (FP)	<i>True Negative</i> (TN)	FP+TN
	Total	TP+FP	FN+TN	TP+FN+FP+TN

1. *True Positive* (TP) yaitu kelas yang diprediksi positif dan pada kenyataannya positif.
2. *True Negative* (TN) yaitu kelas yang diprediksi negatif dan pada kenyataannya negatif.
3. *False Positive* (FP) yaitu kelas yang diprediksi positif tetapi pada kenyataannya negatif.
4. *False Negative* (FN) yaitu kelas yang diprediksi negatif tetapi pada kenyataannya positif.

Ukuran dalam mengevaluasi kinerja model berdasarkan *confusion matrix* ada berbagai macam diantaranya akurasi, sensitivitas, presisi, *f-measure*, dan nilai AUC. Rumus-rumus yang digunakan yaitu sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

$$Sensitivity = Recall = TP_{rate} = \frac{TP}{TP + FN} \tag{5}$$

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$F\ Measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{7}$$

$$False\ Positive\ Rate = FP_{rate} = \frac{FP}{TN + FP} \tag{8}$$

$$AUC = \frac{1 + TP_{rate} - FP_{rate}}{2} \tag{9}$$

Menurut Gorunescu [6] parameter hasil akurasi, sensitivitas, presisi, dan AUC dapat dikatakan menghasilkan hasil klasifikasi yang baik atau tidak dengan menggunakan pedoman parameter hasil klasifikasi yang ditampilkan pada Tabel 3.

Tabel 3. Parameter Hasil Klasifikasi

Rentang	Kategori
0.90 – 1.00	<i>Excellent classification</i> (klasifikasi sangat baik)
0.80 – 0.90	<i>Good classification</i> (klasifikasi baik)
0.70 – 0.80	<i>Fair classification</i> (klasifikasi cukup)
0.60 – 0.70	<i>Poor classification</i> (klasifikasi buruk)
0.5 – 0.60	<i>Failure</i> (klasifikasi salah/gagal)

Tahapan Analisis Data

1. Persiapan data *credit card fraud* yang didapatkan pada situs Kaggle.
2. Dilakukannya *data preprocessing* seperti memilih variabel yang dibutuhkan, mengatasi dan membersihkan data yang *missing*, dan melakukan pengkodean data *string* ke numerik.
3. Mendeskripsikan data dari variabel-variabel yang informatif.
4. Pembagian data (*splitting*) menjadi dua bagian yaitu *data training* untuk melatih model dan *data testing* untuk evaluasi model. Proporsi pembagian yaitu sebesar 75% untuk *data training* dan sisanya untuk *data testing* sebesar 25%.
5. Melakukan pemodelan pengklasifikasian metode *random forest* dengan menentukan jumlah pohon dan jumlah fitur terbaik berdasarkan nilai *error OOB* terkecil.
6. Menerapkan model klasifikasi sehingga menghasilkan hasil prediksi akhir. Pada tahapan ini akan memperlihatkan hasil ketepatan klasifikasi seperti *confusion matrix*.
7. Interpretasi hasil.

C. Hasil Penelitian dan Pembahasan

Data Preprocessing

Pada bagian ini dilakukan *preprocessing* data yakni pada data klasifikasi penipuan kartu kredit yang terdiri dari 50.208 data dan 10 variabel dimana pada data ini tidak terdapat *missing value*, sehingga data yang digunakan yaitu data asli, kemudian dilanjutkan kedalam deskripsi data yang disajikan pada **Gambar 1**. yaitu persentase transaksi yang melakukan penipuan dan tidak penipuan.



Gambar 1. Presentase Transaksi yang Melakukan Penipuan dan Tidak Penipuan

Berdasarkan **Gambar 1**. terlihat bahwa persentase transaksi yang tidak melakukan penipuan lebih besar dibandingkan dengan tindakan yang melakukan penipuan. Adapun persentase sekitar 84,1% transaksi yang terindikasi tidak melakukan penipuan, sisanya sekitar 15,9% transaksi yang terindikasi melakukan penipuan.

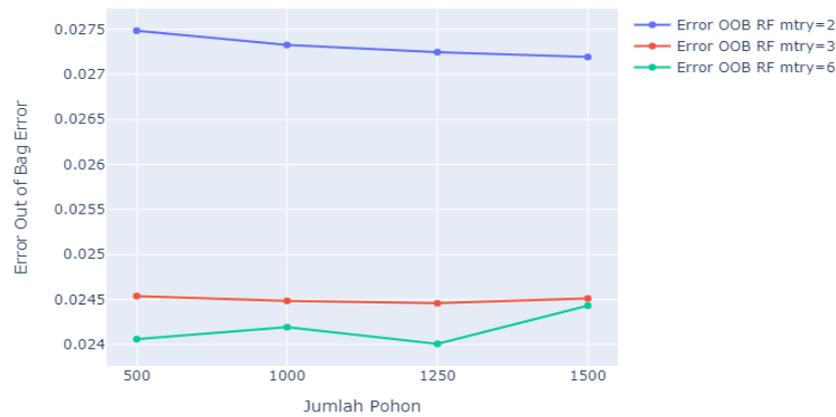
Setelah dilakukan deskripsi data, lalu dilakukan *splitting* data dengan proporsi pembagian data *training* sebesar 75% yaitu sebanyak 37.656 data dan proporsi pembagian data *testing* sebesar 25% yaitu sebanyak 12.552 data, sehingga total data sebanyak 50.208 transaksi.

Analisis Klasifikasi *Random Forest*

Dalam proses klasifikasi dengan menggunakan *random forest* terdapat dua parameter yang harus disesuaikan yaitu dengan menentukan jumlah fitur (*mtry*) dan jumlah pohon. Penentuan nilai *mtry* dilakukan dengan 3 cara yaitu dengan menggunakan Persamaan (1), (2), dan (3) dengan jumlah variabel prediktor sebanyak 9 variabel didapatkan perhitungan sebagai berikut:

1. $mtry_1 = \frac{1}{2} |\sqrt{9}| = 1,5 \approx 2$
2. $mtry_2 = |\sqrt{9}| = 3$
3. $mtry_3 = 2 \times |\sqrt{9}| = 6$

Jumlah fitur yang digunakan yaitu 2, 3, 6 yang selanjutnya akan dicobakan pada tiap-tiap pohon untuk melihat kombinasi mana yang akan menghasilkan nilai misklasifikasi yang paling kecil untuk menentukan parameter optimal yang disajikan dalam **Gambar 2**.



Gambar 2. Hasil Perubahan Error OOB

Berdasarkan **Gambar 2**, terlihat bahwa setiap jumlah fitur (*mtry*) yang digunakan memiliki nilai *error* OOB yang bervariasi. Semakin besar nilai *mtry* yang digunakan, nilai *error* OOB semakin menurun. Hal ini dapat dilihat pada saat $mtry_3 = 6$ fitur mengalami perubahan nilai *error* OOB yang semakin kecil dibandingkan dengan *mtry* lainnya. Pada saat $mtry_3 = 6$ perubahan nilai *error* OOB cenderung naik pada saat jumlah pohon 500 sampai 1000, kemudian turun pada saat jumlah pohon 1.250 dan naik lagi ketika jumlah pohon 1.500. Sedangkan pada saat jumlah fitur 2 dan 3, semakin besar jumlah pohon yang digunakan menghasilkan perubahan nilai *error* OOB yang cenderung menurun, meskipun ketika jumlah fitur 3 mengalami sedikit kenaikan pada saat jumlah pohon 1500.

Evaluasi Model

Dengan menggunakan bantuan bahasa pemrograman *Python* 3.7, **Tabel 4** menyajikan hasil dari klasifikasi *confusion matrix* dari metode *random forest*

Tabel 4. Confusion Matrix Metode Random Forest

Aktual	Prediksi		
	Tidak Penipuan	Penipuan	Total
Tidak Penipuan	10.409	127	10.536
Penipuan	215	1.801	2.016
Total	10.624	1.928	12.552

Berdasarkan **Tabel 4**, dapat dijelaskan terdapat sebanyak 10.536 transaksi yang tidak melakukan penipuan dan 2.016 transaksi yang melakukan penipuan. Pada metode *random forest* tanpa menggunakan *resampling* prediksi kelas yang tidak melakukan penipuan diketahui sebanyak 10.409 transaksi yang diprediksi secara benar dan 127 transaksi yang salah diprediksi. Kemudian prediksi untuk kelas yang melakukan penipuan terdapat sebanyak 1.801 transaksi yang diprediksi secara benar dan 215 transaksi yang salah diprediksi. Dengan menggunakan

persamaan (4) untuk menghitung akurasi, persamaan (5) untuk menghitung *sensitivity*, persamaan (6) untuk menghitung *precision*, persamaan (7) untuk menghitung *F-Measure*, persamaan (8) untuk menghitung *False Positive Rate*, dan persamaan (9) untuk menghitung nilai AUC pada data *testing* maka diperoleh hasil sebagai berikut:

$$Accuracy = \frac{10.409 + 1.801}{10.409 + 215 + 1.801 + 127} \times 100\% = 97,275\%$$

$$Sensitivity = \frac{10.409}{10.409 + 127} \times 100\% = 98,795\%$$

$$Precision = \frac{10.409}{10.409 + 215} \times 100\% = 97,976\%$$

$$F\ Measure = 2 \times \frac{0,97976 \times 0,98795}{0,97976 + 0,98795} \times 100\% = 98,384\%$$

$$FP_{rate} = \frac{FP}{TN + FP} = \frac{215}{215 + 1801} = 0,10664$$

$$AUC = \frac{1 + 0,98795 - 0,10664}{2} \times 100\% = 94,065\%$$

Dari perhitungan yang dilakukan diperoleh hasil akurasi sebesar 97,275% yang menunjukkan bahwa metode *random forest* untuk melakukan prediksi dari data testing dengan hasil prediksi tepat sebesar 97,275%. Nilai sensitivitas yaitu sebesar 98,795% yang berarti ketepatan klasifikasi pada kelas yang tidak melakukan penipuan sebesar 98,795%. Nilai presisi sebesar 97,976% yang berarti terdapat 97,976% yang benar diklasifikasikan tidak melakukan penipuan dari keseluruhan transaksi yang diprediksi tidak melakukan penipuan. Selain itu didapatkan nilai *F-Measure* sebesar 98,384% dan nilai AUC sebesar 94,065% yang termasuk kedalam klasifikasi yang sangat baik karena dari keseluruhan hasil klasifikasi berada pada rentang 90-100%.

D. Kesimpulan

Berdasarkan hasil dan pembahasan dalam penelitian ini dapat diambil kesimpulan bahwa metode *random forest* dapat diterapkan pada data klasifikasi penipuan transaksi kartu kredit dengan parameter optimal yang digunakan yaitu jumlah fitur sebanyak 6 fitur dan jumlah pohon sebanyak 1.250 pohon, didapatkan hasil akurasi sebesar 97,275%, sensitivitas sebesar 98,795%, presisi sebesar 97,976%, *F-Measure* sebesar 98,384%, dan nilai AUC sebesar 94,065% yang termasuk kedalam klasifikasi yang sangat baik karena dari keseluruhan hasil klasifikasi berada pada rentang 90-100%.

Acknowledge

Terimakasih kepada pihak-pihak yang senantiasa selalu memberi dukungan dan membantu penulis dalam menyelesaikan penelitian ini, terutama kepada Allah SWT, keluarga, dosen pembimbing, teman-teman yang tiada hentinya selalu memberikan masukan dan semangat kepada penulis sehingga bisa terselesaikannya penelitian ini.

Daftar Pustaka

- [1] akki.or.id. (2021, Februari 09). Pertumbuhan Kartu Kredit. Retrieved from Asosiasi Kartu Kredit Indonesia (Indonesian Credit Card Association): <https://www.akki.or.id/index.php/credit-card-growth>.
- [2] Gumilang, N. R., & Wachidah, L. (2020). Identifikasi Faktor-Faktor yang Mempengaruhi Perilaku Seksual pada Remaja Menggunakan Metode Regresi Random Forest. *Prosiding Statistika*, Vol. 6, No. 1.

- [3] Muttaqin , M. J. (2013). Metode Ensemble pada CART untuk Perbaikan Klasifikasi Kemiskinan di Kabupaten Jombang. Master Theses of Statistics Institut Teknologi Sepuluh November.
- [4] Breiman, L. (2001). Random Forests. (R. E. Schapire, Ed.) Machine Learning, 45, 5-32. Retrieved April 22, 2021, from <https://link.springer.com/article/10.1023/A:1010933404324>
- [5] Liaw, A., & Wiener, M. (2002). Classification and Regression by Random Forest. R News, 2, 18-22.
- [6] Pertiwi, D. P., Wiranto, & Anggrainingsih, R. (2019). Evaluation Of Campaign Categories On Kitabisa.Com By Naive Bayes Classifier Method. ITSMART: Jurnal Ilmiah Teknologi dan Informasi, 8(1), 26-33.
- [7] Sumarto, Subroto, A., & Arianto, A. (2011, April). Penggunaan Kartu Kredit dan Perilaku Belanja Kompulsif: Dampaknya pada Risiko Gagal Bayar. Jurnal Manajemen Pemasaran, 6(1), 1-7.
- [8] Han, J., Kamber, M., & Pei, J. (2011). Data Mining Concepts and Techniques (3rd ed.). USA: Morgan Kaufmann.