

MODEL *CREDIT SCORING* BERDASARKAN DATA DEMOGRAFI DAN JEJARING SOSIAL DI MEDIA SOSIAL(STUDI KASUS: LINKEDIN)

(*CREDIT SCORING MODEL BASED ON DEMOGRAPHIC AND SOCIAL NETWORK DATA IN SOCIAL MEDIA(CASE STUDY: LINKEDIN)*)

Alvin Ramaquita¹, Andry Alamsyah²

Prodi S1 Manajemen Bisnis Telekomunikasi dan Informatika, Fakultas Ekonomi Bisnis, Universitas Telkom
¹alvinrama@students.telkomuniversity.ac.id,² andrya@telkomuniversity.ac.id

Abstrak

Pengguna media sosial terus bertambah di seluruh dunia, khususnya di Indonesia yang memiliki lima *platform* media sosial yang paling aktif digunakan masyarakatnya, salah satunya adalah LinkedIn. Peningkatan pengguna media sosial menghasilkan data yang sangat besar (*big data*), *big data* diolah untuk menghasilkan wawasan dan membuat keputusan-keputusan strategis untuk bisnis, wawasan yang diperoleh dari *big data* dapat diterapkan diberbagai bidang, salah satunya pada lembaga keuangan seperti Bank untuk mendukung keputusan dalam menentukan kelayakan kredit melalui analisis *credit scoring*. Data media sosial memuat informasi dan konten yang dapat digunakan sebagai data alternatif oleh lembaga keuangan seperti Bank dalam melakukan *credit scoring*, meskipun data media sosial memiliki banyak informasi yang tidak relevan terhadap kondisi kredit seseorang. Oleh karena itu, perlu adanya penyesuaian pendekatan teori dan pengolahan data yang tepat. Melalui permasalahan tersebut, penelitian ini bertujuan untuk melakukan analisis *credit scoring* menggunakan data media sosial LinkedIn berupa data demografi dan jejaring sosial. Penelitian ini menggunakan salah satu model dalam metode *data mining* yaitu *prediction analysis* berupa metode *classification* (klasifikasi) dan memanfaatkan algoritma *decision tree* dan *random forest*. Hasil penelitian ini menunjukkan bahwa kombinasi data demografi dan jejaring sosial yang diolah menggunakan algoritma *random forest* terbukti mampu menghasilkan model *credit scoring* terbaik dengan nilai *accuracy* sebesar 77.27 persen.

Kata kunci: *Big Data*, Media Sosial, *Credit Scoring*, *Prediction Analysis*, Klasifikasi

Abstract

Social media users continue to grow throughout the world, especially in Indonesia which has five social media platform that are most actively used by its people, one of which is LinkedIn. Increased social media users produce a huge amount of data (big data), big data is processed to generate insights and make strategic decision for business, and the insights that obtained from big data can be applied in various fields, one of which is on financial institutions such as Bank to support decisions in determining creditworthiness through credit scoring analysis. Social media data contains information and content that can be used as alternative data by financial institutions such as Banks in conducting credit scoring, even though social media data has a lot of irrelevant information to customer's credit conditions. Therefore, it is necessary to adjust the theoretical approach and the proper data processing. Through these problems, this study aims to conduct a credit scoring analysis using LinkedIn social media data in the form of demographic and social network data. This study uses one of the models in the data mining method called prediction analysis with classification method and using decision tree and random forest algorithm. The results of this study indicate that the combination of demographic data and social network data processed using the random forest algorithm is proven to be able to produce the best credit scoring model with an accuracy value of 77.27 percent.

Keywords: *Big Data*, Social Media, *Credit Scoring*, *Prediction Analysis*, Classification

1. Pendahuluan

Fenomena big data kini mulai digunakan dan dikembangkan oleh berbagai perusahaan besar di dunia, salah satunya di bidang perbankan dalam mendukung aspek bisnis perusahaan, seperti: *customer relationship*, segmentasi nasabah, penilaian pinjaman kredit, dan *marketing strategy* [1]. Lembaga keuangan seperti Bank melakukan *credit scoring* dalam pemberian kredit agar dapat membantu lembaga keuangan dalam menentukan siapa nasabah yang berhak menerima atau mendapatkan kredit dan berapa banyak jumlah yang harus diberikan demi meminimalisir risiko kerugian yang timbul dari kesalahan pemberian pinjaman kredit [2]. Namun, dalam melakukan *credit scoring* dengan metode tradisional yang berbasis data keuangan masih terdapat kekurangan dan keterbatasan dalam segi keakuratan datanya dan kesulitan mengimbangi perkembangan pesat industri keuangan [3]. Hambatan-hambatan yang ditemui ini tentu dapat mempengaruhi kegiatan bisnis di bank, karena kenyataannya apabila semakin baik model dan sistem *credit scoring*, maka semakin rendah risiko yang dapat mengakibatkan kerugian serta nilai tambah suatu bank dapat meningkat [4].

Data media sosial yang memuat banyak informasi dan konten data yang kini digunakan oleh lembaga di industri keuangan seperti P2P *lending* sebagai alternatif dalam melakukan *credit scoring* yang ternyata dapat meningkatkan *approval rates* dan mengurangi *credit loss* dengan biaya yang sangat rendah, meski data media sosial memiliki informasi yang tidak relevan dan data yang *noise* terhadap kondisi kredit pengguna [3].

Analisis media sosial berdasarkan data demografi dan jejaring sosial merupakan salah satu hal yang dapat dijadikan sebagai data pendukung dalam melakukan analisis *credit scoring*, termasuk menentukan apakah layak atau tidaknya suatu nasabah untuk mendapatkan kredit di Bank. Metode yang digunakan dalam mengolah data demografi dan jejaring sosial untuk melakukan analisis *credit scoring* pada proses penentuan kelayakan kredit di Bank adalah dengan menggunakan *data mining*, yaitu eksplorasi di dalam basis data, seperti mencari pola atau aturan yang berguna, yang bertujuan mendapatkan suatu pengetahuan yang berguna untuk mendukung pembuatan keputusan strategis [5]. Penelitian ini menggunakan model *prediction analysis* berupa metode klasifikasi dan memanfaatkan algoritma *decision tree* dan *random forest* dalam mendapatkan hasil akhir, di mana dengan teknik ini dapat menemukan suatu hubungan antara variabel independen dan dependen dari data yang dimiliki [6].

2. Dasar Teori dan Metodologi

2.1. Dasar Teori

2.1.1 Kredit

Kredit adalah salah satu produk perbankan yang menyediakan layanan pemberian pinjaman berupa uang atau tagihan berdasarkan kesepakatan pinjam-meminjam antara pihak Bank selaku pihak kreditur dan pihak peminjam selaku pihak debitur [7]. Sedangkan menurut pemaparan Hand & Henley (1997), kredit juga bisa didefinisikan sebagai suatu hal yang merujuk pada sejumlah uang yang dipinjamkan oleh lembaga keuangan kepada konsumen yang sifatnya harus dilunasi, disertai dengan bunga kredit, dan dikembalikan dengan mencicil secara berkala [8].

2.1.2 Credit Scoring

Credit scoring adalah salah satu metode penilaian yang digunakan lembaga keuangan seperti Bank untuk membantu pembuatan keputusan dalam menyetujui pengajuan kredit dari pihak peminjam dengan melihat risiko yang ada, seperti apakah peminjam akan membayar pinjaman sesuai dengan perjanjian yang dibuat atau tidak [5]. Sadatrasoul *et al* (2013) menyebutkan bahwa *credit scoring* merupakan penilaian risiko terkait dengan pemberian pinjaman kepada suatu individu atau organisasi [9].

2.1.3 Media Sosial

Media sosial adalah suatu *online platform* yang digunakan untuk membangun jejaring sosial atau hubungan sosial antar para penggunanya melalui berbagai persamaan yang dimiliki, seperti: minat terhadap suatu hal, berbagai macam kegiatan, latar belakang, atau koneksi kehidupan yang serupa [10]. Sedangkan menurut El-Badawy & Hashem (2015), media sosial diartikan sebagai *platform* yang memberikan kesempatan terhadap suatu individu untuk berinteraksi, komunikasi dua arah, serta membagikan pendapat dengan pengguna media sosial lainnya [11].

2.1.4 Big Data

Menurut O'Reilly (2012), *big data* adalah kumpulan data yang melebihi kapasitas pemrosesan dan strukturnya tidak sesuai dengan arsitektur basis data konvensional dikarenakan datanya terlalu besar, bergerak terlalu cepat, dan harus memilih cara alternatif untuk memprosesnya untuk mendapatkan nilai dari data tersebut [12]. *Big data* dipaparkan oleh Zhan & Fang (2011) sebagai sumber data apa pun yang memiliki setidaknya tiga karakteristik yang unik, atau yang biasa disebut dengan 3 *V's*, yaitu: *volume*, *velocity*, *variety* [13].

2.1.5 Data Mining

Data mining adalah sebuah proses menemukan korelasi, pola, dan tren baru dan bermakna dengan menyaring atau mengolah dari data yang jumlahnya besar yang disimpan dalam suatu repositori, menggunakan teknologi pengenalan pola, serta teknik statistik dan matematika [14].

2.1.6 Demografi

Definisi lain berkaitan dengan demografi yaitu suatu pengukuran yang memudahkan untuk mengidentifikasi sekumpulan orang yang dikelompokkan menurut karakteristik berdasarkan variabel jenis kelamin, umur, pendapatan, dan pendidikan [15].

2.1.7 Jejaring Sosial

Jejaring sosial menurut pemaparan Zhan & Fang (2011) adalah struktur sosial yang berasal dari masyarakat atau manusia umum sesuai dengan hubungan yang terbentuk didalam suatu ruang lingkup, di mana hubungan ini menyerupai suatu grafik yang terdiri dari *node* (aktor) dan *edge* (hubungan) [13].

2.1.8 Prediction Analysis

Prediction Analysis adalah model dalam *data mining* yang digunakan dalam memprediksi *output* dari suatu hubungan input data berupa variabel independen dan dependen yang diolah menggunakan kriteria atau formula tertentu [6].

2.1.9 Klasifikasi

Klasifikasi dapat didefinisikan sebagai suatu proses untuk menyatakan suatu objek data sebagai salah satu kategori (kelas) yang telah didefinisikan sebelumnya, serta digunakan dalam berbagai aplikasi seperti klasifikasi profil pelanggan, prediksi penjualan, dan *fraud detection* [16].

2.1.9.1 Decision Tree

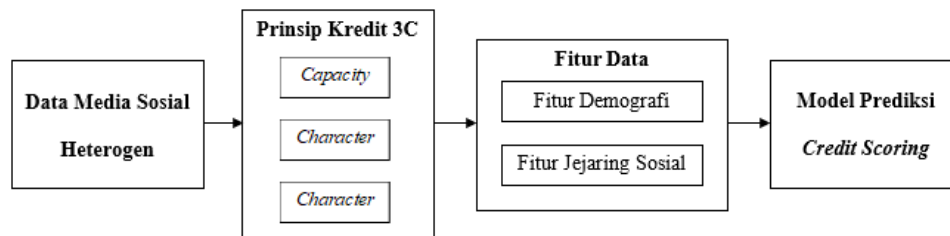
Penjelasan *decision tree* menurut Swamy & Hanumanthappa (2012), adalah model algoritma yang direpresentasikan sebagai pohon biner dua arah yang dapat menunjukkan nilai variabel target (*output*) yang dapat diprediksi dengan seperangkat variabel prediktor (*input*) [17].

2.1.9.2 Random Forest

Menurut penelitian Verma (2018), *random forest* merupakan metode pembelajaran gabungan yang mencoba untuk membangun sejumlah pohon keputusan saat melalui proses *training* dan *output* yang dihasilkan ditentukan dari hasil klasifikasi terbaik dari setiap pohon keputusan tersebut [18].

2.2 Kerangka Pemikiran

Kerangka pemikiran yang digunakan pada penelitian ini diadaptasi dari model penelitian sebelumnya dari Guo *et al* (2016) [3].



Gambar 2.1 Kerangka Pemikiran

Sumber: Guo *et al* (2016) [3]

2.3 Metodologi

Data yang digunakan dalam penelitian ini berupa data media sosial LinkedIn sebanyak 132 akun yang berasal dari koneksi pertemanan salah satu dosen Fakultas Ekonomi Bisnis Telkom University. Dari setiap akun tersebut terdapat informasi untuk diolah menjadi fitur data media sosial, yang terdiri dari data demografi dan jejaring sosial. Analisis kredit didasari oleh prinsip kredit 3C yang diajukan oleh peneliti, yaitu: *character*, *capacity*, dan *conditions* [3]. Kriteria penilaian kredit menggunakan data media sosial ditentukan berdasarkan tabel 2.1.

Tabel 2.1 Credit Scoring Scorecard

No	Atribut	Nilai	Keterangan	No.	Atribut	Nilai	Keterangan
1	<i>Gender</i>	1	Wanita Pria	5	<i>Degree Centrality</i>	1	0 - 5 5 - 10 > 10
		2				2	
		3				3	
2	<i>Age</i>	1	< 19 19 - 34 35 - 54 > 54	6	<i>Betweenness Centrality</i>	1	0 - 1 1 - 10 > 10
		2				2	
		3				3	
		4				3	
3	<i>Job</i>	1	0 - 5.000.000 5.000.0001 - 10.000.000 10.000.001 - 20.000.000 20.000.001 - 40.000.000 > 40.000.000	7	<i>Closeness Centrality</i>	1	0 - 0.51 0.51 - 0.53 > 0.53
		2				2	
		3				3	
		4				3	
		5				3	

			*dalam rupiah				
4	Education	1	Kluster 5	8	Clustering Coefficient	1	0 - 0.5
		2	Kluster 4			2	0.5 - 0.7
		3	Kluster 3			3	> 0.7
		4	Kluster 2				
		5	Kluster 1				

Sumber: Hasil olahan peneliti

3. Pembahasan

3.1 Pembentukan Model Penilaian Kredit

Tabel 3.1 Desain Data Input Model

Model	Fitur Input
Model 1	Fitur Demografi
Model 2	Fitur Jejaring Sosial
Model 3	Fitur Demografi + Jejaring Sosial

Sumber: Hasil olahan peneliti

Penelitian ini mengklasifikasikan fitur data menjadi tiga jenis model seperti pada keterangan tabel 3.1. Setelah data melalui proses *preprocessing*, tahap selanjutnya adalah menentukan *threshold* kelayakan kredit. Suatu akun dapat dikatakan layak mendapatkan kredit apabila memiliki nilai agregasi bobot nilai atribut dalam *credit scoring scorecard* yang melebihi *threshold* yang ditentukan oleh peneliti. Kemudian, data diolah menggunakan algoritma *decision tree* dan *random forest* untuk dievaluasi kinerja *performance metrics*-nya.

3.2 Hasil Model Prediksi

Tabel 3.1 Komparasi Hasil Model Prediksi

Algoritma	Model Percobaan	Performance Matrix			
		Accuracy	Precision	Recall	F-Measure
Decision Tree	Model 1	70.45	65.38	61.82	63.55
	Model 2	61.36	55.00	40.00	46.32
	Model 3	73.48	71.74	60.00	65.35
Random Forest	Model 1	68.18	61.82	61.82	61.82
	Model 2	65.91	59.26	58.18	58.72
	Model 3	77.27	76.60	65.45	70.59

Sumber: Hasil Olahan Peneliti

Berdasarkan hasil komparasi tiga jenis model yang telah dibentuk, model 3 yang diolah menggunakan algoritma *random forest* memiliki nilai evaluasi kinerja *performance metrics* tertinggi apabila dibandingkan dengan model lainnya, dengan rincian nilai *accuracy* sebesar 77.27 persen, nilai *precision* 76.60 persen, nilai *recall* sebesar 65.45 persen, dan nilai *f-measure* sebesar 70.59 persen. Dari hasil tersebut dapat disimpulkan secara keseluruhan bahwa model 3, yang terdiri data *input* berupa kombinasi fitur demografi dan jejaring sosial, yang diolah menggunakan algoritma *random forest* merupakan model terbaik untuk analisis *credit scoring* berbasis data media sosial.

Algoritma *decision tree* memiliki hasil yang baik dalam mengolah model 1 yang terdiri dari data *input* berupa fitur demografi, di mana memiliki nilai *accuracy* sebesar 70.45 persen yang diikuti dengan nilai *precision* sebesar 65.38 persen, nilai *recall* sebesar 61.82 persen, dan nilai *f-measure* sebesar 63.55 persen. Terdapat kemungkinan bahwa model 1 kumpulan data dengan daerah pengambilan keputusan yang cenderung sederhana, sehingga hasil yang diperoleh tersebut dapat terjadi [16]. Disisi lain, algoritma *random forest* memberikan kinerja yang lebih baik daripada algoritma *decision tree* apabila mengolah model 2 yang terdiri dari fitur jejaring sosial yang dapat menghasilkan nilai *accuracy* sebesar 65.91 persen yang diikuti dengan nilai *precision* sebesar 59.26 persen, nilai

recall sebesar 58.18 persen, dan nilai *f-measure* sebesar 58.72 persen. Algoritma *random forest* dapat memberikan hasil yang lebih baik ketika digunakan untuk mengolah tipe data seperti fitur jejaring sosial yang ada pada model 2 dan 3, hal tersebut bisa terjadi karena *random forest* melakukan teknik pembelajaran dengan sistem pemilihan *subset* sampel pelatihan secara acak dan dilanjutkan pemilihan acak dari atribut karakteristik suatu sampel, serta apabila semua pohon keputusan sudah dibangun hasil akhir ditentukan oleh metode pemungutan suara dari hasil bobot yang setara [19].

Secara teori, fenomena *big data* kini menghasilkan sejumlah jejak data digital penggunaan media sosial yang memuat banyak informasi dan konten data dan apabila diaplikasikan lebih lanjut maka dapat diperoleh *insight* dan manfaat yang dapat digunakan untuk kegiatan bisnis dan analisis perilaku sosial [3]. Merujuk pada pernyataan tersebut, hasil penelitian ini dapat menunjukkan bahwa dengan memanfaatkan jejak data digital maupun konten yang dihasilkan oleh setiap pengguna data media sosial LinkedIn dengan memproses data profil akun yang dikonversikan menjadi data demografi dan data pertemanan akun yang dikonversikan menjadi data jejaring sosial ini dapat menghasilkan *insight* dan manfaat yang berguna untuk kegiatan bisnis khususnya lembaga keuangan seperti Bank dalam melakukan analisis *credit scoring* dalam menentukan kelayakan kredit seseorang.

4. Kesimpulan dan Saran

4.1 Kesimpulan

Dari penelitian yang telah dilakukan dapat ditarik kesimpulan sebagai berikut:

1. Proses pembentukan model *credit scoring* berdasarkan data media sosial dapat diawali dengan melakukan pengumpulan dan pemrosesan data profil akun LinkedIn menjadi data demografi dan jejaring sosial yang masing-masing memiliki *insight* yang dapat mewakili tiga prinsip kredit yang diajukan peneliti, yaitu: *character*, *capacity*, dan *conditions*.
 - a) Data yang diperoleh kemudian memasuki tahap *preprocessing data* dilakukan agar didapatkan kualitas data yang baik dan dapat diolah oleh mesin data dengan cara menyaring dan membersihkan data yang tidak relevan dan data kosong.
 - b) Setelah melewati tahap *preprocessing*, pembentukan fitur data dilakukan agar mendapatkan data yang berkualitas dan relevansi yang paling baik sehingga menghasilkan data *input* yang sesuai untuk melakukan proses penentuan kelayakan dalam analisis *credit scoring*.
 - c) Setelah membentuk fitur demografi dan jejaring sosial, selanjutnya adalah membentuk *credit scoring scorecard* yang digunakan sebagai *threshold* yang dapat membantu peneliti dalam proses pelabelan layak tidak suatu data akun dalam mendapatkan kredit.
 - d) Berdasarkan *threshold* serta proses pelabelan data pada tahap sebelumnya, model *credit scoring* dibentuk dengan mengklasifikasikan fitur data menjadi tiga jenis data *input* model: fitur demografi, fitur jejaring sosial, dan fitur demografi + fitur jejaring sosial.
2. Berdasarkan hasil komparasi model prediksi, dapat disimpulkan bahwa model prediksi yang sesuai untuk digunakan dalam melakukan *credit scoring* berdasarkan data demografi dan jejaring sosial adalah model 3 dengan menggunakan algoritma *random forest* menunjukkan nilai evaluasi kinerja *performance metrics* tertinggi. Dari hasil pengoalahan, model 3 menghasilkan nilai *accuracy* sebesar 77.27 persen, nilai *precision* 76.60 persen, nilai *recall* sebesar 65.45 persen, dan nilai *f-measure* sebesar 70.59 persen. Selain itu, dari hasil komparasi model 1 dan 2 juga dapat disimpulkan bahwa fitur jejaring sosial pada model 2 cenderung lebih baik apabila dijadikan data pelengkap untuk meningkatkan hasil nilai evaluasi kinerja mengingat hasil pada model 2 memiliki tingkat *accuracy* data yang tidak terlampau jauh dengan model 1, serta dalam proses pengolahan datanya membutuhkan waktu lebih panjang dan pemahaman lebih lanjut.

4.2 Saran

4.2.1 Saran Teoritis

Agar bisa memberikan manfaat dan wawasan lebih dalam bidang manajemen keuangan yang dipadukan dengan fenomena *big data*, penelitian selanjutnya juga diharapkan dapat membentuk model *credit scoring* berbasis data media sosial yang lebih baik dengan menggunakan sumber data dari media sosial lainnya yang dapat digunakan sebagai pembanding dan mengolah jenis atau ragam bentuk data lainnya selain data demografi dan jejaring sosial, yang dapat memberikan *insight* lebih dan berguna untuk mendukung kegiatan bisnis lembaga keuangan seperti Bank.

4.2.2 Saran Praktis

Saran praktis yang dapat diberikan kepada perusahaan adalah hasil penelitian ini dapat menjadi rekomendasi untuk bagi Bank atau lembaga keuangan lainnya untuk memahami dan memanfaatkan jejak digital berupa data media sosial dalam penerapannya untuk mengelola berbagai kegiatan dan produk perbankan, khususnya dalam melakukan analisis *credit scoring* pada proses penentuan kelayakan kredit. Dengan menggunakan dan memanfaatkan data media sosial, pihak Bank dapat menghemat waktu dan biaya serta memberikan strategi

keputusan yang tepat dalam menentukan penilaian kelayakan kredit seseorang sehingga keunggulan kompetitif perusahaan dapat ditingkatkan.

Daftar Pustaka:

- [1] Davenport, T. H., & Dyché, J. (2013). Big Data in Big Companies. *International Institute for Analytics*, 16.
- [2] Eddy, Y. L., & Nazri, E. M. (2017). Credit Scoring Models: Techniques and Issues. *Journal of Advanced Research in Business and Management Studies*, 29.
- [3] Guo, G., Zhu, F., Chen, E., Liu, Q., & Wu, L. (2016). From Footprint to Evidence: An Exploratory Study of Mining Social Data for Credit Scoring. *Research Collection School of Information Systems Singapore Management University*, 1-40
- [4] Blochlinger, A., & Leippold, M. (2005). Economic Benefit of Powerful Credit Scoring. *Journal of Banking & Finance*, 22.
- [5] Yap, B. W., Ong, S. H., & Mohamed Husain, N. H. (2011). Using Data Mining to Improve Assessment of Credit Worthiness via Credit Scoring Models. *Expert Systems with Applications*, 2.
- [6] Olson, D., & Shi, Y. (2007). *Introduction to Business Data Mining*. New York: McGraw-Hill Irwin.
- [7] Fahmi, I. (2014). *Manajemen Perkreditan*. Bandung: Alfabeta.
- [8] Hand, D. J., & Henley, W. E. (1997). Statistical Classification Methods in Consumer Credit Scoring: a Review. 1.
- [9] Sadatrasoul, S. M., Gholamian, M. R., Siami, M., & Hajimohammadi, Z. (2013). Credit Scoring in Banks and Financial Institutions via Data Mining Techniques: A Literature Review. *Journal of AI and Data Mining*, 2.
- [10] Akram, W., & Kumar, R. (2017). A Study on Positive and Negative Effects of Social Media on Society . *International Journal of Computer Sciences and Engineering*, 1.
- [11] El-Badawy, T. A., & Hashem, Y. (2015). The Impact of Social Media on the Academic Development of School Students. *International Journal of Business Administration*, 1-2.
- [12] O'Reilly. (2012). *Big Data Now: 2012 Edition*. Sebastopol: O'Reilly Media Inc.
- [13] Zhan, J., & Fang, X. (2011). Social Computing: The State of The Art. *International Journal Social Computing and Cyber-Physical Systems*, 1-12.
- [14] Larose, D. T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. Hoboken: John Wiley & Sons.
- [15] Grewal, D., & Levy, M. (2012). *Marketing*. New York: Mc Graw Hill.
- [16] Suyanto. (2017). *Data Mining Untuk Klasifikasi dan Klusterisasi Data*. Bandung: Informatika.
- [17] Swamy, M. N., & Hanumanthappa, M. (2012). Predicting Academic Success from Student Enrolment Data Using Decision Tree Technique. *International Journal of Applied Information Systems (IJ AIS)*, 2.
- [18] Verma, A. (2018). Study and Evaluation of Classification Algorithms in Data Mining. *International Research Journal of Engineering and Technology (IRJET)*, 1297-1308.
- [19] Ren, Q., Cheng, H., & Han, H. (2017). Research on Machine Learning Framework Based on Random Forest Algorithm. *AIP Conference Proceedings*, 1-7.