

Analisis Konsumsi Energi Listrik Pelanggan Dan Biaya Pokok Produksi Penyediaan Energi Listrik dengan Machine Learning

Raditya Hari Nugraha¹, Eko Yuwono², Latif Prasetyohadi³, Yanuardhi Arief B⁴, Harry Patria⁵

^{1,2,3,4,5}Institut Teknologi Sepuluh Nopember

Jl. HOS Cokroaminoto No 12A, +62315613922/+62315666172

Abstract

PT PLN (Persero) during the Covid-19 pandemic was one of the companies whose sales growth was affected by the decline in electricity consumption in several sectors. Another condition is that several power plant and substation construction projects have fulfilled the realization commitment to the RUPTL from PT PLN (Persero). This has resulted in PT PLN (Persero) being faced with an over supply condition between power capacity and customer usage load. Realization of sales growth until July 2021 was 4.44% (144,788 TWh). Energy consumption in July 2021 was 20.55 TWh where the growth of kWh sales in July 2021 comparing with July 2020 began to show a recovery of +1.82%. The factor that most affected business and industrial growth was the manufacturing sector in Indonesia experiencing a slowdown/contraction as reflected in the PMI (Purchasing Managers Index) which decreased from 53.5 to 40.1. Growth is strongly influenced by consumer behavior in responding to government regulations, especially related to controlling the spread of Covid-19 in Indonesia in the form of restrictions on social activities (PSBB, PPKM, or Lockdown) which have been effectively implemented since April 2020 until now. Based on the analysis of the customer's electrical energy consumption data per industrial sector, as well as using technical data on the availability of power per electrical sub-system and the cost of producing electrical energy in an area, an evaluation model will be obtained that can be used in selecting the criteria for prospective customers who will be given program offers "SEMAKIN PRODUKTIF". By using "SEMAKIN PRODUKTIF" program data modeling, it is hoped that prospective customers will be given program offers so that they can be an opportunity to increase sales growth of electrical energy which is targeted to grow 6% in December 2021.

Keywords: Growth, Energy sales, Semakin Produktif, Machine Learning.

Abstrak

PT PLN (Persero) di masa pandemic Covid-19 merupakan salah satu perusahaan yang growth penjualannya terdampak akibat dari konsumsi pemakaian listrik di beberapa sektor mengalami penurunan. Kondisi lainnya, beberapa proyek pembangunan pembangkit dan gardu induk sudah memenuhi komitmen realisasi terhadap RUPTL dari PT PLN (Persero). Hal ini mengakibatkan PT PLN (Persero) dihadapkan pada kondisi over supply antara kapasitas daya terhadap beban pemakaian pelanggan. Realisasi pertumbuhan penjualan sd Juli 2021 sebesar 4,44% (144,788 TWh). Konsumsi energi jual pada bulan juli 2021 sebesar 20,55 TWh dimana pertumbuhan kWh penjualan Juli 2021 terhadap Juli 2020 mulai menunjukkan pemulihan yaitu sebesar +1,82%. Faktor yang paling mempengaruhi pertumbuhan bisnis dan industri yaitu karena sektor manufaktur di Indonesia mengalami perlambatan/kontraksi yang tercermin pada PMI (Purchasing Managers Index) mengalami penurunan dari 53,5 ke 40,1. Pertumbuhan sangat dipengaruhi oleh Perilaku Konsumen dalam menyikapi regulasi pemerintah terutama terkait pengendalian penyebaran Covid-19 di Indonesia berupa pembatasan aktivitas sosial (PSBB, PPKM, atau Lockdown) yang secara efektif diimplementasikan sejak April tahun 2020 sampai dengan saat ini. Berdasarkan dari analisa data konsumsi energi listrik pelanggan per sektor industrinya, serta menggunakan data teknis ketersediaan daya per sub-system kelistrikan dan biaya produksi energi listrik di suatu daerah, maka akan didapatkan evaluasi model yang dapat digunakan dalam pemilihan kriteria calon

pelanggan yang akan diberikan penawaran program "SEMAKIN PRODUKTIF". Dengan menggunakan pemodelan data program "SEMAKIN PRODUKTIF" tersebut, maka diharapkan akan diperoleh calon pelanggan yang akan diberikan penawaran program sehingga dapat menjadi peluang meningkatkan Growth penjualan energi listrik yang ditargetkan tumbuh 6% pada Desember tahun 2021.

Kata kunci: Pertumbuhan listrik, Penjualan listrik, Semakin Produktif, Machine Learning.

1. PENDAHULUAN

Pelanggan PT PLN (Persero) terdiri dari pelanggan social, rumah tangga, bisnis, industry dan pemerintah dengan total pelanggan sebesar 80,59 Juta dan total konsumsi kWh jual sampai dengan bulan juli 2021 adalah sebesar 144,78 TWh. Pertumbuhan kWh Jual tertinggi adalah dari golongan tarif Rumah Tangga yaitu sebesar 45,87% (66,41 TWh), hal ini sangat wajar mengingat kontribusi jumlah Pelanggan Rumah Tangga yang sangat signifikan yaitu sebesar 91,83% (74,01 juta Pelanggan) dengan kWh per Pelanggan sebesar 128 kWh. Jumlah Pelanggan dengan komposisi terkecil kedua adalah pelanggan Industri yaitu sebesar 0,18% (145.224 pelanggan) namun menghasilkan kontribusi kWh Jual kedua terbesar setelah pelanggan Rumah Tangga yaitu sebesar 30,41% (44,03 TWh), Rupiah per Pelanggan tertinggi yaitu sebesar Rp 47.284.746/Pelanggan dan kWh per Pelanggan tertinggi yaitu sebesar 43.317 kWh/Pelanggan. Jumlah Pelanggan dengan komposisi terkecil kedua adalah pelanggan Industri yaitu sebesar 0,18% (145.224 pelanggan) namun menghasilkan kontribusi kWh Jual kedua terbesar setelah pelanggan Rumah Tang ga yaitu sebesar 30,41% (44,03 TWh), Rupiah per Pelanggan tertinggi yaitu sebesar Rp 47.284.746/Pelanggan dan kWh per Pelanggan tertinggi yaitu sebesar 43.317 kWh/Pelanggan.

Komposisi pelanggan terkecil atau kecil namun memberikan kontribusi kWh yang terbesar dari golongan tarif lainnya tentunya harus diberikan perhatian khusus. Berdasarkan hal tersebut diperlukan pemetaan pelanggan berdasarkan konsumsi kWh tertinggi atau selanjutnya didefinisikan dengan Pelanggan 10K. Perilaku pelanggan 10K tersebut akan dimanfaatkan sebagai peluang dalam membuat suatu produk layanan serta dapat dijadikan penentuan target *probing* suatu program yang telah diciptakan perusahaan. Saat ini PT PLN (Persero) memiliki program insentif khusus yakni "SEMAKIN PRODUKTIF" dimana pelanggan akan diberikan keringanan berupa tarif khusus sesuai ketentuan yang berlaku dengan tujuan sebagai stimulus bagi peningkatan produksi disisi pelanggan. Produk ini sangat berdampak bagi pelanggan potensial sehubungan dengan pandemic COVID-19 mengakibatkan produktivitas pelanggan menjadi tidak maksimal sehubungan dengan biaya pemakaian energi listrik yang memberatkan pelanggan serta mekanisme perhitungan biaya pemakaian minimum atas kapasitas langganan ke PT PLN (Persero). Dalam penentuan target pelanggan yang akan diberikan penawaran program tersebut digunakan teknologi *machine learning* sehingga dapat lebih tepat sasaran.

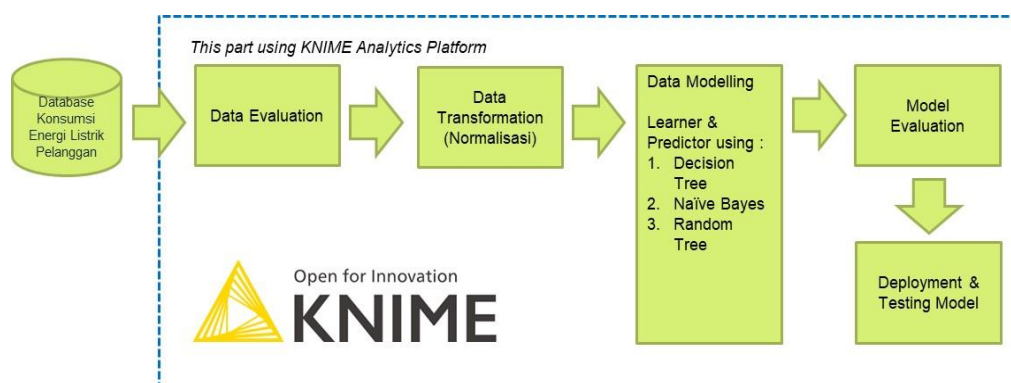
Teknologi *Machine Learning* adalah mesin yang dikembangkan untuk bisa belajar dengan sendirinya tanpa arahan dari penggunanya.

Pembelajaran mesin dikembangkan berdasarkan disiplin ilmu lainnya seperti statistika, matematika dan data mining sehingga mesin dapat belajar dengan menganalisa data tanpa perlu di program ulang atau diperintah. (Dicoding, 2020). Dalam hal ini *Machine Learning* memiliki kemampuan untuk memperoleh data yang ada dengan perintah ia sendiri. *Machine Learning* memiliki dua teknik dasar belajar, yaitu *supervised learning* dan *unsupervised learning*.

Menurut (Wikipedia, 2020), dalam *supervised learning*, seseorang dapat melatih mesin dengan menggunakan data yang diberi label. Artinya, beberapa data sudah diberi label dengan jawaban yang benar. Ini dapat dibandingkan dengan pembelajaran yang berlangsung di hadapan pengawas atau guru. Algoritma pembelajaran yang terarah dapat mempelajari pola tersembunyi dari data pelatihan yang telah berlabel, hal ini akan membantu kita memprediksi hasil untuk data yang belum pernah dipelajari sebelumnya. Untuk dapat berhasil membangun, mengatur, dan menerapkan model mesin pembelajar yang terarah dengan akurasi tinggi, dibutuhkan waktu dan keahlian teknis dari tim *data scientist* yang sangat terampil. Selain itu, para *data scientist* sebaiknya harus mampu membangun kembali model untuk memastikan prediksi yang dihasilkan tetap benar walaupun datanya berganti. *Supervised Learning* memungkinkan kita untuk mengumpulkan data atau menghasilkan keluaran data berdasarkan dari pengalaman sebelumnya. Hal ini dapat membantu mengoptimalkan kriteria kinerja berdasarkan pengalaman mesin. Pembelajaran terarah juga dapat membantu memecahkan berbagai jenis masalah komputasi dunia nyata. *Supervised Learning* memiliki beberapa metode, diantaranya *Decision Tree*, *Naïve-Bayes*, dan *Random Forest*.

2. METODOLOGI PENELITIAN

Metode prediksi model menggunakan 3 fasa, antara lain data *pre-processing*, pemodelan data, serta pengukuran performa dari model yang dapat ditunjukkan dari gambar dibawah ini.



Gambar 1. Metodologi Penelitian

2.1. Data *Pre-processing*

Data *pre-processing* merupakan sekumpulan teknik yang diterapkan pada database untuk menghapus *noise*, *missing value*, dan data yang tidak konsisten. Data preprocessing dibagi menjadi beberapa langkah, yaitu *cleaning data*, *data transformation*, dan *data reduction*. Data *preprocessing* ini digunakan karena dalam data *realtime database* seringkali tidak lengkap dan tidak konsisten sehingga mengakibatkan hasil data mining tidak tepat dan kurang akurat. Oleh karena itu, untuk meningkatkan kualitas data yang akan dianalisis, perlu dilakukan langkah-langkah *preprocessing data*. Langkah-langkah tersebut tidak harus semuanya dilakukan. Pada *data mining*, data *preprocessing* merupakan salah satu langkah penting agar data dapat diolah sesuai dengan metode dan tools yang digunakan. Data yang baru saja dikumpulkan pasti mengandung kesalahan, *missing value*, *error*, dan data-data yang tidak penting lainnya khususnya data primer yang diambil dari internet, seperti data review produk, data dari sosial media, dan lain sebagainya. Hal ini menyebabkan seorang data scientist harus memahami langkah-langkah data *preprocessing* agar data yang ia gunakan tepat sasaran dan menghasilkan hasil analisis yang akurat.

Pada Jurnal ini data yang digunakan adalah data 10.000 pelanggan dengan konsumsi energi listrik tertinggi di PT PLN (Persero) pada bulan Agustus 2021. Dimana data tersebut dapat digunakan untuk melakukan prediksi model tertentu berdasarkan perilaku konsumsi energi listrik pelanggan.

Tabel 1. Data sampel konsumsi energi listrik pelanggan PT PLN (Persero) bulan Agustus 2021

NO	NAMA PLG	TARIF	DAYA	SEGMENT	PEMKWH_202108	RPTAG_202108	JAM NYALA	RP/KWH
1	XXXXXXXXXX	I4	220000000	TT	67,315,965	69,109,410,403	306	997
2	XXXXXXXXXX	LI4	180000000	TT	45,324,000	49,099,135,673	252	1,664
3	XXXXXXXXXX	I4	95000000	TT	42,981,854	44,916,403,231	452	997
4	XXXXXXXXXX	I4	97000000	TT	41,475,840	42,580,847,625	428	997
6	XXXXXXXXXX	I4	280000000	TT	38,112,000	43,836,397,361	136	997
7	XXXXXXXXXX	LI4	70000000	TT	28,147,500	31,941,563,450	402	1,664
8	XXXXXXXXXX	I4	89000000	TT	37,699,200	38,703,589,626	424	997
9	XXXXXXXXXX	I4	105000000	TT	28,740,000	28,904,124,368	274	997
.....
9998	XXXXXXXXXX	LI3	345000	TM	149,436	215,911,306	433	1,664

Dari data table diatas variable yang digunakan, antara lain:

- a) Nama PLG
Merupakan nama pelanggan yang terdaftar di Data Induk Langgan di PT PLN (Persero).
- b) Tarif
Pengelompokkan pelanggan berdasarkan peruntukannya sesuai aturan berlangganan di PT PLN (Persero).
- c) Daya
Kapasitas penggunaan energi listrik yang dapat digunakan oleh pelanggan di PT PLN (Persero) dengan satuan VoltAmpere (VA)
- d) Segmen
Pengelompokkan pelanggan berdasarkan tegangan pelayanan, antara lain Tegangan Tinggi (TT) dengan tegangan 150 KV, Tegangan Menengah (TM) dengan tegangan 20 KV, Tegangan Rendah (TR) dengan tegangan 220 V.
- e) PemKwh_202108
Berupa konsumsi energi listrik yang digunakan pelanggan pada bulan agustus tahun 2021 dalam satuan KWh.
- f) Jam_Nyala
Merupakan lama penggunaan energi listrik pelanggan kurun waktu 1 bulan dalam satuan Jam.
- g) Rp/Kwh
Nilai harga satuan per Kwh yang akan dibebankan kepada pelanggan yang merupakan ketentuan dari kementerian ESDM.
- h) Program
Merupakan kriteria pelanggan yang akan diberikan penawaran produk layanan PLN.

2.2. Data Processing

Model diolah menggunakan aplikasi KNIME dan hasil *pre-processing* selanjutnya akan dilakukan *data training* dan *data testing*. Terdapat 3 metode yang digunakan untuk menghasilkan model prediksi antara lain *decision tree*, *naïve bayes*, dan *random forest*. Ketiga metode tersebut akan dilakukan sebanyak 70% *training* dan 30% *testing* menggunakan node partitioning.

a) Decision Tree

Decision tree merupakan metode klasifikasi dan prediksi yang sangat kuat dan banyak di minati (Wu, 2009). Dalam *decision tree* ini data yang berupa fakta dirubah menjadi sebuah pohon keputusan yang berisi aturan dan tentunya dapat lebih mudah dipahami. Model pohon keputusan banyak digunakan pada kasus data dengan output yang bernilai diskrit. Walaupun tidak menutup kemungkinan dapat juga digunakan untuk kasus data dengan atribut numerik.

b) Naïve Bayes

Naïve Bayes merupakan metode klasifikasi yang berakar pada teori Bayes. Metode pengklasifikasian ini menggunakan metode probabilitas dan statistik yg disampaikan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang masa depan didasarkan

dari pengalaman masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Ciri utama dari *Naïve Bayes Classifier* ini adalah asumsi yg sangat kuat (naïf) akan independensi dari masing-masing kondisi / kejadian.

c) *Random Forest*

Klasifikasi *random forest* dilakukan melalui penggabungan pohon (*tree*) dengan melakukan training pada sampel data yang dimiliki. Penggunaan pohon (*tree*) yang semakin banyak akan mempengaruhi akurasi yang akan didapatkan menjadi lebih baik. Penentuan klasifikasi dengan *random forest* diambil berdasarkan hasil voting dari *tree* yang terbentuk. Pemenang dari *tree* yang terbentuk ditentukan dengan vote terbanyak. Pembangunan pohon (*tree*) pada *random forest* sampai dengan mencapai ukuran maksimum dari pohon data. Akan tetapi, pembangunan pohon *random forest* tidak dilakukan pemangkasan (*pruning*) yang merupakan sebuah metode untuk mengurangi kompleksitas ruang. Pembangunan dilakukan dengan penerapan metode *random feature selection* untuk meminimalisir kesalahan. Pembentukan pohon (*tree*) dengan sample data menggunakan variabel yang diambil secara acak dan menjalankan klasifikasi pada semua *tree* yang terbentuk. *Random forest* menggunakan *Decision Tree* untuk melakukan proses seleksi. Pohon yang dibangun dibagi secara rekursif dari data pada kelas yang sama.

2.3. Pengukuran Kinerja

Evaluasi dan validasi hasil model prediksi antara lain:

a) *Cross Validation*

Cross Validation merupakan salah satu teknik untuk menilai/memvalidasi keakuratan sebuah model yang dibangun berdasarkan dataset tertentu. *Cross Validation* juga merupakan pengujian standar yang dilakukan untuk memprediksi *error rate*.

b) *Confusion Matrix*

Confusion matrix adalah suatu metode yang digunakan untuk melakukan perhitungan akurasi pada konsep *data mining*. Rumus ini melakukan perhitungan dengan 4 keluaran, yaitu: *recall*, *precision*, *acuraccy* dan *error rate*. Evaluasi model klasifikasi didasarkan pada pengujian untuk memperkirakan obyek yang benar dan salah (Wu, 2009).

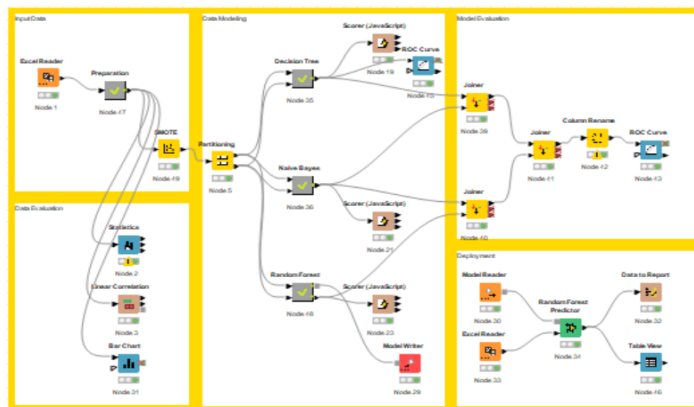
c) *ROC Curve*

ROC curves merupakan salah satu cara melakukan analisa terhadap model *classifier* yang telah dibuat. Penggunaan *ROC curves* untuk menentukan parameter model yang diinginkan sesuai dengan karakteristik dari model *classifier* yang Metode klasifikasi bisa dievaluasi berdasarkan kriteria seperti tingkat akurasi, kecepatan, kehandalan, skalabilitas dan interpretabilitas (Vecellis, 2009).

d) Validasi Menurut (Gurenescu,2011) Diperlukan cara yang sistematis untuk mengevaluasi kinerja suatu metoda. Evaluasi klasifikasi didasarkan pada pengujian pada obyek benar dan salah, menurut (Ian.H,2011) Validasi data digunakan untuk menentukan jenis terbaik dari skema belajar yang digunakan, berdasarkan data pelatihan untuk melatih skema pembelajaran untuk memaksimalkan penggunaan data.

3. HASIL DAN PEMBAHASAN

Model prediksi yang dijalankan pada aplikasi KNIME memiliki diagram alir seperti ditunjukkan pada gambar dibawah ini dimana dibagi menjadi beberapa bagian antara lain input data, data evaluasi, pemodelan data, evaluasi pemodelan, dan *deployment model*. Proses simulasi dilakukan pengukuran tingkat akurasi dan persentase error dari tiga metode *supervised learning* antara lain *decision tree*, *naïve bayes*, dan *random forest* sehingga hasil terbaik akan dijadikan model prediksi dalam penentuan rencana penawaran produk layanan ke pelanggan.



Gambar 2. Simulasi pemodelan pada aplikasi KNIME

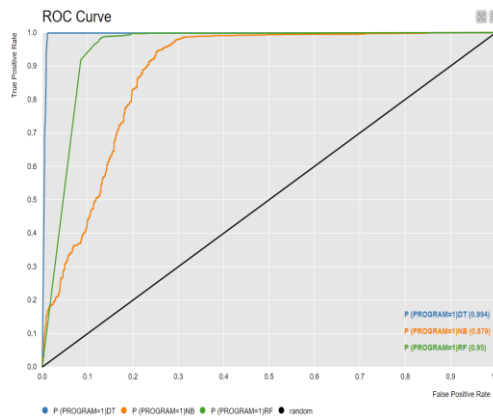
Dari *data-preprocessing* maka dapat dihasilkan data seperti yang ditunjukkan pada tabel dibawah dimana karakteristik konsumsi energi listrik pelanggan berupa pemakaian kwh, jam nyala, serta Rp/kWh maka dapat di klasifikasikan pelanggan yang berhak menerima penawaran program semakin produktif.

Tabel 2. Karakteristik konsumsi listrik pelanggan

NO	NAMA PLG	TARIF	DAYA	SEGMENT	PEMKWH_2021 08	RPTAG_202 108	JAM NYALA	RP/KWH	PROGRAM
1	XXXXXXXXXX	14	220000000	TT	67,315,965	69,109,410, 403	306	997	SEMAKIN PRODUKTIF
2	XXXXXXXXXX	LI4	180000000	TT	45,324,000	49,099,135, 673	252	1,664	SEMAKIN PRODUKTIF
3	XXXXXXXXXX	14	95000000	TT	42,981,854	44,916,403, 231	452	997	PROGRAM LAIN
4	XXXXXXXXXX	14	97000000	TT	41,475,840	42,580,847, 625	428	997	PROGRAM LAIN
6	XXXXXXXXXX	14	280000000	TT					SEMAKIN

NO	NAMA PLG	TARIF	DAYA	SEGMENT	PEMKWH_2021	RPTAG_202	JAM NYALA	RP/KWH	PROGRAM
					38,112,000	43,836,397,361	136	997	PRODUKTIF
7	XXXXXXXXXX	LI4	70000000	TT	28,147,500	31,941,563,450	402	1,664	PROGRAM LAIN
8	XXXXXXXXXX	I4	89000000	TT	37,699,200	38,703,589,626	424	997	PROGRAM LAIN
9	XXXXXXXXXX	I4	105000000	TT	28,740,000	28,904,124,368	274	997	PROGRAM LAIN
.....
9998	XXXXXXXXXX	LI3	345000	TM	149,436	215,911,306	433	1,664	PROGRAM LAIN

Selanjutnya data konsumsi energi listrik dari pelanggan 10K yang sudah dimiliki akan dilakukan proses simulasi kesesuaian parameter model terhadap skema klasifikasi program semakin produktif. Melalui karakteristik pelanggan 10K dengan karakter konsumsi energi listrik yang berbeda-beda tersebut diharapkan dapat menghasilkan model prediksi yang akurat dan dapat digunakan untuk pelanggan lain. Karakteristik pelanggan yang dapat menerima program semakin produktif yakni pelanggan dengan jam nyala <300 jam dan Rp/kWh >1000 selain itu hal lain yang mempengaruhi antara lain pelanggan dengan jenis tarif produktif seperti tarif Industri (I), tarif Bisnis (B), dan tarif layanan premium (L). Ketersediaan pasokan energi listrik di suatu daerah juga menjadi salah satu hal yang menjadi pertimbangan dalam penentuan kriteria model prediksi program dimana untuk beberapa wilayah di Indonesia Timur masih belum dapat menerima program semakin produktif.



Gambar 3. Kurva ROC hasil simulasi pada aplikasi KNIME

Untuk menentukan model prediksi yang akan digunakan maka terlebih dahulu dapat dengan menjalankan kurva ROC dengan hasil yang ditunjukkan seperti gambar diatas. Dengan hasil simulasi kurva ROC maka dapat ditunjukkan metode mana yang memiliki tingkat kesesuaian parameter model dengan skema klasifikasi yang diharapkan. Dari gambar kurva ROC diatas dapat terlihat bahwa metode *decision tree* memiliki tingkat kesesuaian (*True Positive Rate*) yang paling baik dibandingkan dengan metode *naïve bayes* dan *random forest*.

Tabel 3. Hasil simulasi

Metode	Accuracy (Learning/Testing)			% Error (Learning/Testing)		
	80/20	75/25	70/30	80/20	75/25	70/30
Decision Tree	99,29	99,23	99,44	0,71	0,77	0,56
Naïve Bayes	89,30	88,43	89,08	10,70	11,57	10,92
Random Forest	91,56	92,77	70,16	8,44	7,23	29,84

Dari tabel didapatkan informasi bahwa dengan komposisi *data training* dan *data testing* sebesar 80%/20%, 75%/25%, dan 70%/30% dihasilkan kestabilan akurasi prediksi pada metode *Decision Tree* dengan tingkat akurasi diatas 99%.

4. SIMPULAN

Pada penelitian ini memberikan hasil analisa dari ketiga metode supervised learning dan dapat disimpulkan bahwa metode *Decision Tree* memiliki tingkat akurasi yang paling baik dibandingkan dengan metode *Naïve Bayes* dan *Random Forest* dengan tingkat akurasi >99%. Dari table diatas juga kita dapat memilih skema training dan testing dari model yang terbaik adalah dengan 70% *data learning* dan 30% *data testing*.

Metode *Naïve Bayes* memiliki tingkat akurasi yang paling rendah dibandingkan dengan 2 metode lainnya dengan akurasi 80-90% namun masih dalam tingkat akurasi yang cukup baik. Sedangkan untuk metode *Random Forest* hasil simulasi menunjukkan tingkat akurasi dengan skema learning 80/20 dan 75/25 antara *data learning* dan *data testing* sudah >90% namun saat menggunakan skema 70/30 hasil akurasi menjadi kurang baik yakni hanya di angka 70,16%. Berdasarkan kesimpulan diatas maka metode prediksi yang akan digunakan sebagai acuan rencana pemberian program insentif tarif khusus "SEMAKIN PRODUKTIF" yakni menggunakan *model decision tree*. Selanjutnya model prediksi dengan metode *decision tree* tersebut akan di *deploy* dan selanjutnya.

DAFTAR PUSTAKA

- [1]. Gurenescu, 2011, Data mining : Concept and Techniques. Verlag berlin Heidelberg: Springer.
- [2]. Ian H. Witten, frank Eibe, and Mark A. Hall, Data Mining: Practical Machine Learning Tools and Techniques, 3rd ed., Asma Stephan and Burlington, Eds. United States
- [3]. Liao. 2007. Recent Advances in Data Mining of Enterprise Data: Algorithms and Application . Singapore: World Scientific Publishing
- [4]. Wu, Xindong & Kumar, Vipin. 2009. The Top Ten Algorithms in Data Mining. Boca Raton: CRC Press
- [5]. Zhang, Guazhen, Zhou, faming, etl., 2008, Knowledge creation in marketing based on data mining, Intelligent Computation Technology and Automation (ICICTA), 2008 International Conference on Page(s): 782 – 786

- [6]. Vercellis, C. 2009. *Business-Intelligent: Data Mining and Optimization for Decision Making*. Southern Gate: John Wiley & Sons Inc.
- [7]. Shukla, A. Tiwari, R., & Kala, R. 2010. *Real Life Application of Soft Computing*. Taylor and Francis Groups, LLC.
- [8]. Diyah, Puspitaningrum. 2006. *Pengantar Jaringan Syaraf Tiruan*, Penerbit Andi, Yogyakarta.
- [9]. M. DASH and H. LIU, "Feature selection for classification," *Intell. Data Anal.*, vol. 1, no. 1-4, pp. 131-156, 1997.
- [10]. M. Wibowo, F. Noviyanto, S. Sulaiman, and S. M. Shamsuddin, "Machine Learning Technique For Enhancing Classification Performance In Data Summarization Using Rough Set And Genetic Algorithm," *Int. J. Sci. Technol. Res.*, vol. 8, no. 10, pp. 1108-1119, 2019.
- [11]. A. Luque, A. Carrasco, A. Martín, and A. de las Heras, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognit.*, vol. 91, pp. 216-231, Jul. 2019.
- [12]. D. Kurniawan, A. Anggrawan, and H. Hairani, "Graduation Prediction System On Students Using C4.5 Algorithm," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 19, no. 2, pp. 358-365, 2020.
- [13]. S. Helal et al., "Predicting academic performance by considering student heterogeneity," *Knowledge-Based Syst.*, vol. 161, pp. 134-146, Dec. 2018.
- [14]. A. Khan and S. K. Ghosh, "Student performance analysis and prediction in classroom learning: A review of educational data mining studies," *Educ. Inf. Technol.*, vol. 26, no. 1, pp. 205-240, Jan. 2021.
- [15]. A. Namoun and A. Alshantqi, "Predicting Student Performance Using Data Mining and Learning Analytics Techniques: A Systematic Literature Review," *Appl. Sci.*, vol. 11, no. 1, p. 237, Dec. 2020.
- [16]. D. H. Kamagi and S. Hansun, "Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa," *J. Ultim.*, vol. 6, no. 1, pp. 15-20, 2014.
- [17]. X. Xu, J. Wang, H. Peng, and R. Wu, "Prediction of academic performance associated with internet usage behaviors using machine learning algorithms," *Comput. Human Behav.*, vol. 98, pp. 166-173, Sep. 2019.