

Pemanfaatan Machine Learning untuk Pengelompokan dan Prediksi Target Tambah Daya Listrik Pelanggan Prabayar (Studi Kasus : PT PLN ULP Watang Sawitto)

Utilization of Machine Learning for Grouping and Predicting Targets for Adding Electricity to Prepaid Customers (Case Study: PT PLN ULP Watang Sawitto)

Rizqa Afthoni¹, Diana Purwitasari², Agus Budi Raharjo³

^{1,2,3}Magister Manajemen Teknologi, Institut Teknologi Sepuluh Nopember
E-mail: ¹rzaftthoni@gmail.com, ²diana@if.its.ac.id, ³agus.budi@its.ac.id

Abstrak

Perkembangan teknologi sistem informasi dan ilmu pengetahuan khususnya dalam bidang pemasaran membuat para pelaku usaha berupaya untuk meningkatkan *competitive advantage* mereka dengan mengerahkan sumber daya yang dimiliki oleh perusahaan. Perusahaan dituntut untuk berinovasi dalam mengelola perusahaannya agar dapat bertahan dalam dunia persaingan. Kemampuan untuk memprediksi pelanggan prabayar yang berpotensi tambah daya listrik merupakan salah satu strategi pendukung untuk keberhasilan program pemasaran tambah daya pelanggan berdasarkan karakteristik konsumsi listriknya. Berdasarkan hal tersebut, penelitian ini mengajukan metode prediksi pelanggan prabayar dengan memanfaatkan algoritma pengelompokan (*Clustering*) dan klasifikasi. Data yang diolah adalah data pelanggan prabayar tarif rumah tangga yang memiliki fitur variabel daya listrik pelanggan (VA), frekuensi beli token listrik, total pemakaian kWh, total rupiah pembelian token, selisih daya VA pelanggan, jam nyala, periode hari pembelian token listrik, dan riwayat tambah daya listrik pelanggan. Pengelompokan dilakukan dengan menerapkan algoritma K-means. Dari hasil tersebut, model prediksi dibangun sesuai target setiap klaster dengan memanfaatkan dua metode, Gradient Boosting dan Artificial Neural Network. Evaluasi prediksi model terbaik dilakukan dengan menerapkan tiga skenario proporsi data latih dan data uji, yang selanjutnya diukur menggunakan matrik akurasi dan Cohen Kappa. Hasil eksperimen menghasilkan empat klaster berdasarkan karakteristik konsumsi listriknya. Gradient Boosting memberikan hasil yang terbaik untuk semua klaster, untuk klaster 1 menghasilkan nilai AUC 0.784, klaster 2 menghasilkan nilai AUC 0.941, klaster 3 menghasilkan nilai 0.884 dan klaster 4 menghasilkan nilai AUC 0.903.

Kata kunci: Tambah daya, K-means, Gradient Boosting dan Artificial Neural Network

Abstract

The development of information system technology and science, especially in the field of marketing, makes business actors seek to increase their competitive advantage by mobilizing the resources owned by the company. Companies are required to innovate in managing their companies in order to survive in the world of competition. The ability to predict prepaid customers who have the potential to add electricity is one of the supporting strategies for the success of the marketing program for adding power to customers based on the characteristics of their electricity consumption. Based on this, this study proposes a prediction method for prepaid customers by utilizing clustering and classification algorithms. The data processed is household tariff prepaid customer data which features variable customer electrical power (VA), frequency of buying electricity tokens per year, total kWh usage per year, total rupiah token purchases per year, customer VA power difference, hours of operation, period days of purchase of electricity tokens, and history of adding electricity to customers. The grouping is done by applying the K-means algorithm. From these results, a prediction model is built according to the target of each cluster by utilizing two methods, Gradient Boosting and Artificial Neural Network. Evaluation of the best model predictions is carried out by applying three scenarios of the proportion of training data and test data, which are then measured using the accuracy matrix and Cohen Kappa. The

experimental results produce four clusters based on the characteristics of their electricity consumption. Gradient Boosting gives the best results for all clusters, for cluster 1 it produces an AUC value of 0.784, cluster 2 produces an AUC value of 0.941, cluster 3 produces an AUC value of 0.884 and cluster 4 produces an AUC value of 0.903.

Keywords: Increase power, K-means, Gradient Boosting and Artificial Neural Network

1. PENDAHULUAN

PT. PLN (Persero) khususnya unit ULP Watang Sawitto masih menggunakan konsep konvensional dalam pemasaran tambah daya listrik. Konsep pemasaran tambah daya listrik, selama ini hanya berdasarkan jenis tarif dan daya VA pelanggan di program-program promonya. Hal ini menyebabkan program pemasaran tambah daya listrik hanya menasar dan tertuju pada jenis tarif dan daya tertentu saja, padahal pelanggan di luar jenis tarif dan daya tersebut ada potensi untuk mengikuti program pemasaran tersebut. Tetapi, karena jenis tarif dan dayanya yang tidak sesuai menyebabkan mereka tidak bisa ikut program promo tambah daya listrik. Hal ini sangat disayangkan karena PT. PLN (Persero) kehilangan potensi pelanggan yang menginginkan untuk mengikuti program promo tambah daya.

Program promo pemasaran tambah daya listrik seharusnya bisa didasarkan pada tren konsumsi listrik pelanggan, sehingga program promo tambah daya tersebut lebih mencerminkan kebutuhan pelanggan. Belum adanya data mining profil pemakaian listrik pelanggan menyebabkan pengambilan keputusan untuk skema program tambah daya kurang sesuai dengan kondisi kebutuhan konsumsi listrik pelanggan. Program promo hanya didasarkan pada tarif dan daya VA semata, tanpa memperhatikan profil konsumsi listrik pelanggan. Banyak data-data konsumsi listrik pelanggan yang tersimpan di database PT PLN (Persero) kurang dimanfaatkan dengan baik, apabila data tersebut diolah maka dapat membangkitkan wawasan baru, yang dapat membantu pengambilan keputusan dalam menjalankan strategi bisnis.

Penelitian Motlagh [1] menggunakan machine learning untuk pengelompokan pelanggan listrik di perumahan New South Wales, Australia. Pengelompokan pelanggan ini berdasarkan karakteristik beban konsumsi listrik di perumahan tersebut. Tantangan pada penelitian ini adalah bagaimana mengekstraksi fitur *time-series* dan memisahkan *noise data* yang terdapat pada data *time-series* beban listrik perumahan agar lebih mudah untuk dikelompokkan. Dari hasil metode *Unsupervised Learning* yang digunakan, terbentuk 12 kluster beban listrik di perumahan. Masing-masing kluster mempunyai karakter konsumsi listrik yang berbeda, sehingga hal tersebut bisa digunakan acuan untuk segmentasi pelanggan. Daniel Robert Thomasa [2] juga melakukan penelitian mengenai karakter konsumsi listrik pelanggan di daerah pedalaman India. Penelitian ini dilakukan dengan menggunakan data kesediaan membayar listrik, tingkat kepuasan pelanggan dan konsumsi *watt* listrik alat rumah tangga. Jumlah *dataset* yang digunakan lebih dari 10.000 pelanggan listrik rumah tangga di lebih dari 200 desa di India. Metode yang digunakan adalah menggunakan *Machine Learning* dan Teknik Regresi, dimana pelanggan listrik terlebih dahulu disegmentasikan berdasarkan kesediaan membayar listrik, tingkat kepuasan pelanggan dan kualitas pencahayaan. Terdapat 3 kluster segmentasi, yaitu segmen pelanggan potensial, segmen pelanggan penggunaan listrik rendah dan segmen pelanggan penggunaan listrik tinggi. Dari hasil segmentasi tersebut, dapat dipetakan prediksi penggunaan listrik di tiap segmennya.

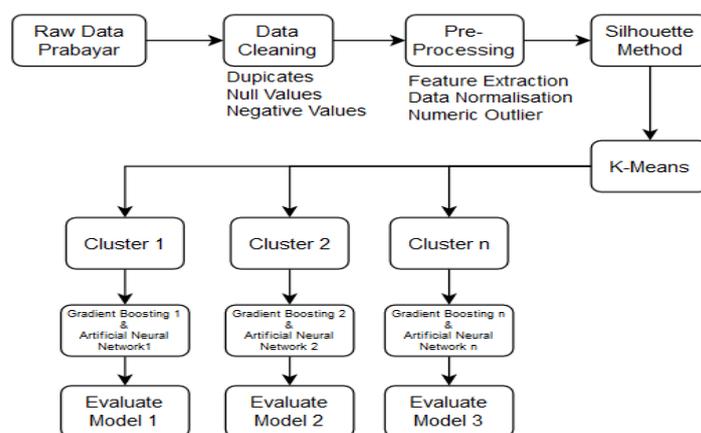
Pemanfaatan Machine Learning tidak hanya pada pelanggan listrik, seperti yang dilakukan oleh Halidu Abu-Bakar [3], yang menggunakan *unsupervised learning dan supervised learning* untuk membuat profil pelanggan berdasarkan karakter konsumsi air. Dataset yang digunakan adalah data *smart meter* dari 10.000 pelanggan air rumah tangga di tahun 2019. Dari hasil pengolahan *Machine Learning*, mengungkapkan bahwa terdapat empat kluster rumah tangga, yaitu Evening Peaks (EP), Late Morning (LM), Early Morning (EM) dan Multiple Peaks (MP). Penelitian ini menyimpulkan bahwa mengidentifikasi ciri utama pola konsumsi perkluster nya akan berpengaruh pada fungsi peramalan permintaan dan prediksi konsumsi per-klaternya sehingga memungkinkan langkah-langkah intervensi yang dapat diambil sesuai dengan

karakteristiknya.

Berdasarkan permasalahan dan hasil penelitian – penelitian diatas, maka dalam penelitian ini, akan dilakukan prediksi target tambah daya listrik pelanggan prabayar berbasis pembelajaran mesin. Pada penelitian ini akan diterapkan 2 tahapan kombinasi pengolahan pembelajaran mesin yaitu *Unsupervised Learning* [4] dan *Supervised Learning* [5]. Dataset input berupa data histori pelanggan prabayar tambah daya, frekuensi pembelian token selama 1 tahun, periode hari lama pembelian token listrik, Daya VA terpasang di pelanggan, jam nyala rata-rata pelanggan, pemakaian kWh pelanggan, total rupiah tenaga listrik yang didapat dan selisih daya. Pertama, penelitian dilakukan dengan menganalisa kluster yang terbentuk dari hasil K-means clustering, hal ini dilakukan agar pelanggan prabayar terkluster sesuai dengan karakteristik komsumsi listriknya. Hasil dari *clustering* tersebut dilanjutkan pada tahapan kedua, yaitu *supervised learning*. Metode yang digunakan pada tahapan kedua ini adalah *Artificial Neural Netwok* [6] dan *Gradient Boosted*. Metode tersebut bertujuan untuk menentukan target pelanggan prabayar yang berpotensi tambah daya listrik. Evaluasi prediksi model terbaik dilakukan dengan menerapkan tiga skenario proporsi data latih dan data uji, yang selanjutnya diukur menggunakan evaluasi AUC [7]. Tujuan yang ingin dicapai adalah menemukan model prediksi target tambah daya listrik pelanggan prabayar pada masing masing kluster-nya. Model yang memberikan hasil yang baik, akan digunakan sebagai data pendukung pengambilan keputusan bisnis PLN, sehingga PLN mampu memanfaatkan hasil model tersebut sebagai acuan dalam penentuan program tambah daya yang sesuai dengan karakteristik komsumsi listrik pelanggan [8] dan mampu menentukan target market *door to door* tambah daya pelanggan.

2.METODE PENELITIAN

Sistematika penelitian ini dibagi menjadi tiga tahapan utama, yaitu tahapan pembersihan dan *preprocessing* data, tahapan pengelompokan data serta tahapan pemilihan model terbaik. Tahapan pertama penelitian ini menerapkan pembersihan data dan *preprocessing* data dimana data yang masih terdapat *duplicate*, *null values*, *negative flows* dibuang selanjutnya dilakukan *preprocessing* data untuk *feature extraction*, *data normalization* dan *numeric outliers* [9].



Gambar 1 Sistematika Penelitian

Tahapan kedua yaitu pengelompokan data, pada tahapan ini meliputi penentuan jumlah kluster optimal dengan menggunakan silhouette score dan pengelompokan dengan menerapkan algoritma K-Means. Data mentah prabayar yang sudah melalui-proses pembersihan data, akan dilakukan proses klustering pelanggan terlebih dahulu, hal ini bertujuan agar pelanggan prabayar terkelompok sesuai dengan karakteristik komsumsi listriknya. Setelah didapatkan jumlah kluster pelanggan prabayar, langkah selanjutnya adalah pengolahan data secara klasifikasi *supervised learning* dengan menggunakan algoritma *Gradient Boosting* dan *Artificial Neural Netwok*. Pada tahapan ini bertujuan untuk menentukan target pelanggan prabayar yang berpotensi tambah daya

listrik sesuai dengan histori tambah daya dan karakter konsumsi listriknya. Setelah model memberikan hasil prediksi, tahapan berikutnya adalah pemilihan model terbaik dengan menerapkan tiga skenario proporsi data latih dan data uji untuk diukur performanya menggunakan AUC.

2.1 Sumber Data

Data pada penelitian ini menggunakan data kuantitatif, diantaranya adalah data daya VA listrik pelanggan prabayar, frekuensi pembelian token pelanggan, total pemakaian kWh, total rupiah pembelian token, jam nyala, selisih daya, periode pembelian token dan data histori tambah daya. Pada tahapan pengumpulan data dan informasi yang diperoleh dalam penelitian ini, penulis mengunduh data tersebut dari aplikasi internal PLN. Jumlah data awal yang digunakan dalam penelitian ini adalah 17.290 pelanggan prabayar tarif rumah tangga.

Tabel 1 Pemilihan Daya VA

Daya	Non Tambah Daya_0	Tambah daya_1	Turun Daya_2	Jumlah
450 VA	103	1	0	104
900 VA	7152	179	240	7571
1300 VA	6346	443	3	6792
2200 VA	1004	683	2	1689
3500 VA	179	365	0	544
4400 VA	46	198	0	244
5500 VA	61	167	0	228
7700 VA	11	41	0	52
11000 VA	5	61	0	66
TOTAL	14907	2138	245	17290

Daya	Non Tambah Daya_0	Tambah daya_1	Jumlah
900 VA	7152	179	7331
1300 VA	6346	443	6789
2200 VA	1004	683	1687
3500 VA	179	365	544
TOTAL	14681	1670	16351



Pada tahapan selanjutnya, dilakukan pemilihan daya VA pelanggan yang sering melakukan permintaan tambah daya. Dari hasil rekap Tabel 1 diatas, menunjukkan daya 900VA, 1300VA, 2200VA dan 3500VA merupakan daya VA dominan melakukan tambah daya. Sedangkan pelanggan yang melakukan turun daya, tidak dimasukkan sebagai sumber data pada penelitian ini, karena mempunyai jumlah data turun daya yang sedikit. Jumlah proporsi pelanggan daya 900VA, 1300VA, 2200VA dan 3500VA yang dipilih menjadi sumber data utama sebanyak pada penelitian ini adalah 94% dari keseluruhan data awal. Daya 450VA, 4400VA, 5500VA, 7700VA dan 11000 VA tidak dimasukkan menjadi sumber data utama, karena jumlah pelanggan di daya tersebut sedikit dan minim melakukan permintaan tambah daya. Daya VA pelanggan yang menjadi acuan pada penelitian ini adalah daya 900VA, 1300VA, 2200VA dan 3500VA.

1.2 Definisi Variabel Penelitian

Penelitian ini menggunakan beberapa variabel input dalam proses pengelompokan dan prediksi target tambah daya listrik, berikut adalah struktur data pada penelitian ini :

Tabel 2 Struktur Data

No	Name Field	Data Type	Keterangan
1.	IDPEL	Char	No ID Pelanggan
2.	Daya VA	Numeric	Daya VA saat ini
3.	Frekuensi beli token	Numeric	Frekuensi Pembelian
4.	Total pemakaian kWh	Numeric	Pemakaian kWh
5.	Total Rupiah Token	Numeric	Rupiah pendapatan
6.	Jam nyala	Numeric	Jam pakai listrik
7.	Selisih daya	Numeric	Selisih Daya mutasi
8.	Periode pembelian token	Numeric	Jarak hari pembelian
9.	Data histori tambah daya	Numeric	Keterangan Tambah Daya

a. Daya listrik Pelanggan (VA)

Daya listrik Pelanggan (VA) merupakan batas kontrak daya listrik yang diberikan PLN kepada pelanggan. Terdapat beberapa daya kontrak PLN ke pelanggan, contohnya adalah daya 900VA, 1300VA, 2200VA dst. Daya yang digunakan dalam penelitian ini adalah daya tarif rumah tangga pelanggan prabayar, yaitu dari daya 900VA, 1300VA, 2200VA dan

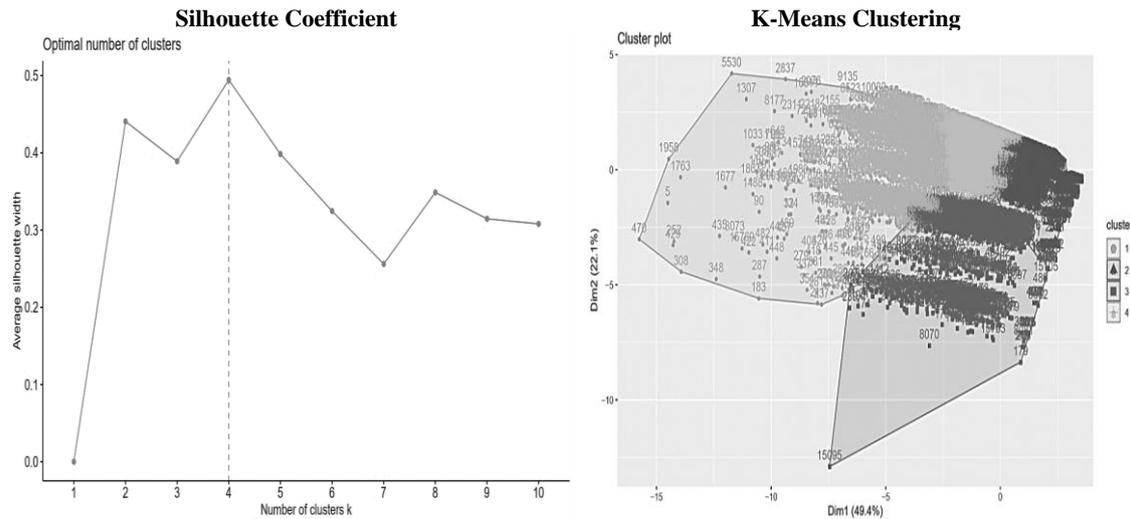
- 3500VA.
- b. Frekuensi pembelian token listrik
Frekuensi beli token listrik merupakan jumlah berapa kali pelanggan membeli token listrik. Pada penelitian ini diambil frekuensi pembelian token dalam satu tahun. Variabel ini menunjukkan karakter dari tiap pelanggan prabayar dalam membeli token setiap tahunnya.
 - c. Total pemakaian kWh
Total pemakaian kWh adalah jumlah total kWh yang pelanggan beli selama satu tahun. Variabel ini merupakan jumlah kWh kumulatif token listrik yang pelanggan beli. Token listrik yang pelanggan beli memiliki kuota kWh berbeda-beda, sesuai dengan daya dan tarifnya masing-masing.
 - d. Total rupiah pembelian token
Variabel ini menunjukkan jumlah total rupiah token yang pelanggan beli selama satu tahun. Variabel ini menunjukkan *gain* pendapatan rupiah yang PLN dapatkan dari pembelian token pelanggan prabayar per tahunnya.
 - e. Selisih daya
Selisih daya merupakan selisih daya awal pelanggan sebelum melakukan permintaan tambah daya dikurangi dengan daya sekarang. Sebagai contoh pelanggan 900 VA melakukan permintaan tambah daya ke 1300 VA, maka selisih daya nya ialah 400 VA.
 - f. Jam nyala listrik
Jam nyala listrik merupakan konversi pemakaian kWh pelanggan pada setiap dayanya ke dalam jumlah jam selama satu bulan (skala 0 sampai dengan 720 jam nyala). Ini memberikan informasi berapa lama pemakaian listrik pelanggan secara jam dalam satu bulan.
 - g. Periode hari lama pembelian token listrik
Variabel ini memberikan informasi periode hari rata-rata antar pembelian token pelanggan selama satu tahun.
 - h. Data histori tambah daya
Data histori tambah daya merupakan variabel target dalam penelitian ini. Data pelanggan yang tidak pernah melakukan tambah daya diberi label 0, sedangkan pelanggan yang pernah melakukan tambah daya diberikan label 1

3. HASIL DAN PEMBAHASAN

3.1 Pengolahan Silhouette Coefficient dan K-Means

Pada tahapan ini, pengujian dilakukan untuk mengetahui relasi antara objek di dalam klaster dan seberapa jauh jarak antar klaster. Metode silhouette coefficient merupakan kombinasi dari dua metode yaitu metode kohesi yang berfungsi untuk mengukur jarak relasi antara objek di dalam klaster, dan metode separasi yang berfungsi untuk mengukur jarak sebuah klaster terpisah dengan klaster lain. Sebanyak 16351 data pelanggan dilakukan proses normalisasi terlebih dahulu yang bertujuan untuk membuat data pada beberapa variabel memiliki rentang nilai yang sama sehingga proses pengolahan data menjadi lebih ringan. Pada penelitian ini, klustering dilakukan agar pelanggan prabayar terkelompok sesuai dengan kemiripan karakter konsumsi listriknya terlebih dahulu. Penentuan jumlah klaster, ditentukan melalui metode silhouette coefficient.

Pada tahapan ini, sebanyak 16351 data pelanggan dilakukan proses penentuan jumlah klaster dengan metode silhouette coefficient. Hasil proses silhouette coefficient menunjukkan nilai optimal koefisien ada pada jumlah 4 klaster dengan nilai silhouette rata-rata 0.5, berdasarkan hasil nilai koefisien silhouette tersebut, maka dapat disimpulkan bahwa terdapat 4 klaster karakteristik konsumsi listrik.



Gambar 2 Penentuan Jumlah Kluster dan Klastering Pelanggan

Setelah mendapatkan jumlah kluster yang optimal melalui metode *silhouette coefficient* diatas, maka langkah selanjutnya adalah memproses 16.351 data pelanggan prabayar agar terbentuk 4 kluster melalui K-means clustering. Metode ini menghitung jarak setiap data dengan titik-titik pada klaternya, data tersebut akan berada pada kelompok kluster yang memiliki jarak terkecil terhadap titik *centroid*.

Dari hasil pengolahan K-means clustering dengan jumlah 4 kluster, maka didapatkan sebaran data kluster 2 dan kluster 4, terlihat lebih dekat dan solid, maka dapat diambil kesimpulan bahwa pelanggan pada kluster ini mempunyai identitas kemiripan yang kuat. Untuk kluster 1 dan kluster 3 dari persebaran datanya, terlihat lebih bervariasi, hal ini menunjukkan bahwa pelanggan pada kluster ini mempunyai variansi konsumsi listrik yang lebih luas, namun masih tetap pada masing-masing klaternya.

3.2 Profil konsumsi listrik kluster

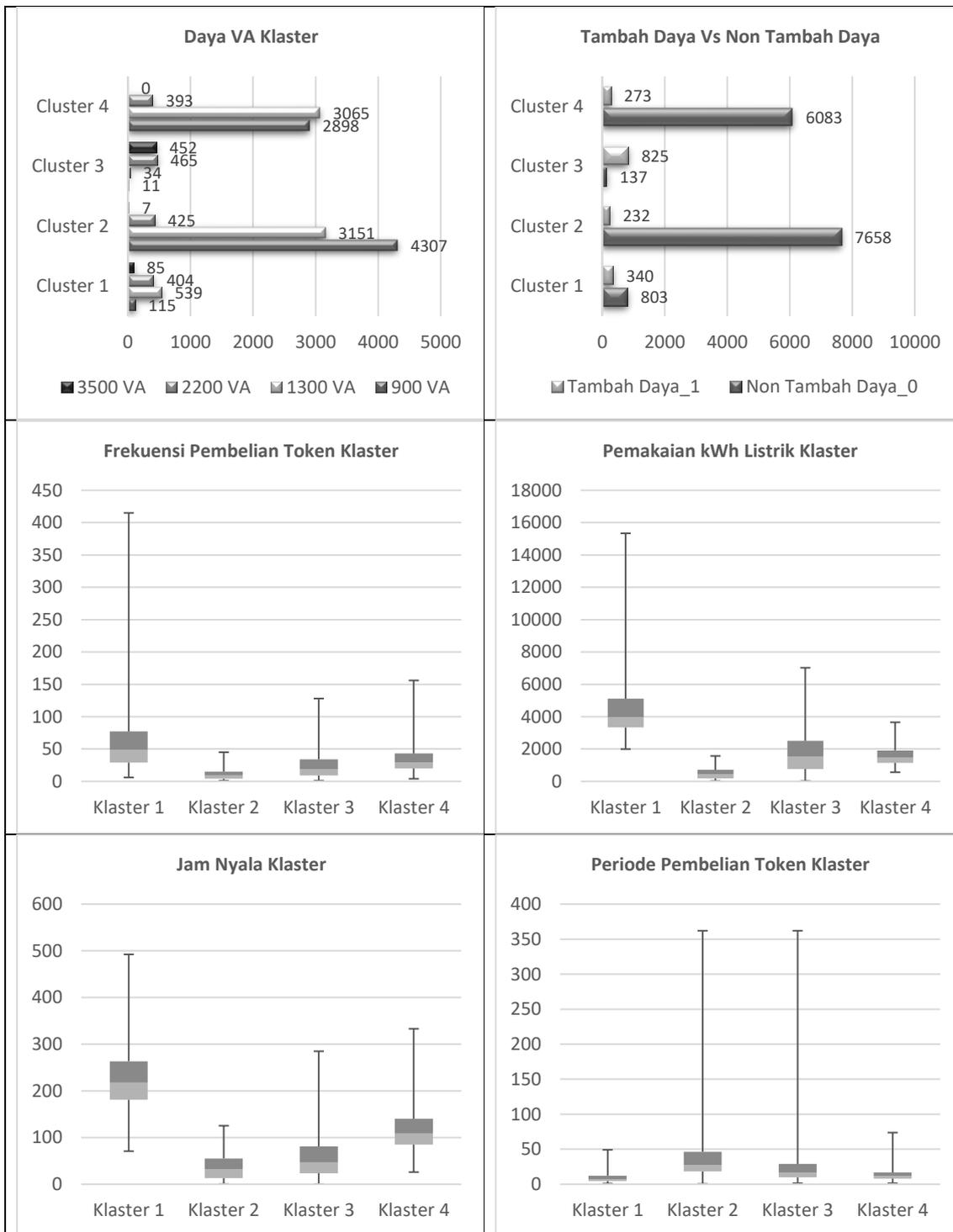
Berdasarkan daya VA listrik pelanggan pada tiap kluster nya, menunjukkan bahwa kluster 2 merupakan kluster dengan jumlah pelanggan terbanyak, yaitu 7890 pelanggan. Kluster 2 mempunyai jumlah pelanggan daya 900 VA dan 1300 VA terbanyak dibandingkan dengan kluster lain. Kluster 4, merupakan satu-satunya kluster yang tidak mempunyai pelanggan daya 3500 VA, tetapi untuk daya 900 VA dan 1300 VA di kluster 4 ini, jumlah pelanggannya terbanyak kedua setelah kluster 2. Kluster 3 jumlah pelanggannya paling sedikit dibandingkan dengan kluster lain.

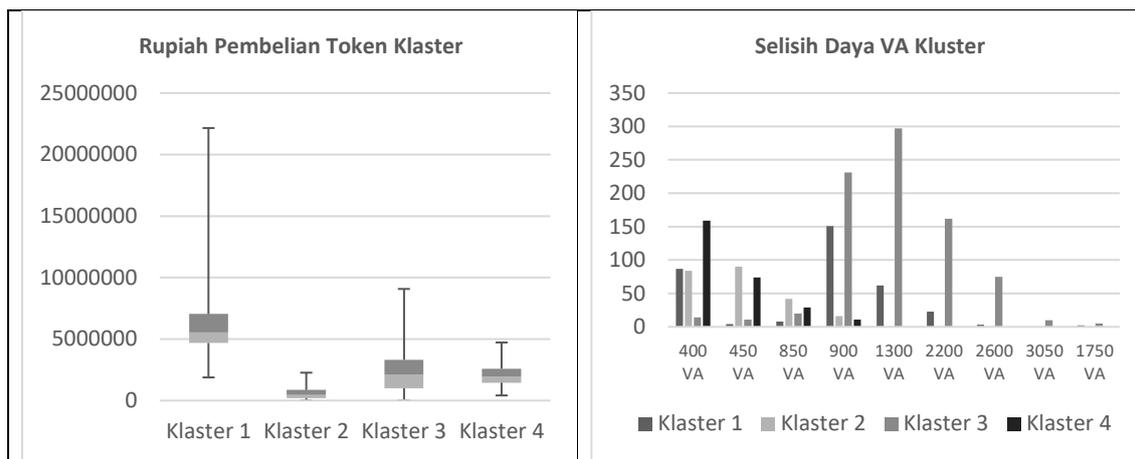
Berdasarkan perbandingan tambah daya dengan non tambah daya menunjukkan bahwa kluster 3 merupakan kluster terbanyak pelanggan melakukan tambah daya, yaitu 825 pelanggan. Kluster 1 mempunyai pelanggan yang melakukan tambah daya paling banyak kedua setelah kluster 3, yaitu sebanyak 340 pelanggan. Pelanggan yang tidak melakukan tambah daya paling banyak ada di kluster 2 dan kluster 3. Pelanggan pada kluster 2 dan kluster 4 masih mempunyai banyak potensi untuk tambah daya, sehingga kluster-kluster tersebut dapat dioptimalkan untuk peningkatan tambah daya.

Berdasarkan frekuensi pembelian token memperlihatkan bahwa kluster 1 merupakan kluster dengan frekuensi paling tinggi pembelian tokennya diantara semua kluster. Ini menunjukkan kluster 1 merupakan kluster yang paling aktif pembelian tokennya dibandingkan dengan kluster lainnya. Kluster 3 dan kluster 4 mempunyai karakteristik frekuensi pembelian token yang hampir sama, tetapi kluster 4 cenderung lebih tinggi frekuensinya dibandingkan dengan kluster 3. Kluster 2 menempati posisi paling rendah untuk frekuensi pembelian tokennya. Dari boxplot frekuensi pembelian token, menunjukkan bahwa kluster 2 masih mempunyai potensi yang kuat untuk ditingkatkan frekuensi pembelian tokennya.

Profil kluster berdasarkan pemakaian kWh listrik memperlihatkan bahwa kluster 1

merupakan klaster dengan pemakaian kWh listrik paling tinggi diantara semua klaster. Klaster 3 menempati posisi kedua dalam komsumsi listriknya, sedangkan untuk klaster 4 menempati posisi ketiga. Klaster 2 merupakan klaster dengan komsumsi listrik paling rendah dibandingkan dengan klaster lainnya. Dapat diambil kesimpulan bahwa klaster 2, klaster 3 dan klaster 4 masih mempunyai banyak potensi untuk tambah daya.





Gambar 3 Boxplot profil konsumsi listrik klaster

Untuk profil klaster berdasarkan jam nyala memperlihatkan bahwa klaster 1 masih mendominasi untuk jam nyala tertinggi diantara semua klaster. Klaster 4 dan klaster 3 mempunyai karakteristik jam nyala yang hampir sama, tetapi klaster 4 cenderung lebih tinggi jam nyalanya dibandingkan dengan klaster 3. Klaster 2 menempati posisi paling rendah jam nyalanya, rentang nilai maksimal frekuensi pembelian token di klaster 2 hanya mencapai 120. Dari rentang nilai jam nyala boxplot tiap klasternya, menunjukkan klaster dengan jumlah pelanggan yang banyak yaitu klaster 2 (7890 pelanggan) dan klaster 4 (6356 pelanggan) masih sangat berpotensi untuk ditingkatkan jam nyalanya dengan meningkatkan daya VA pelanggan atau melalui program-program tertentu agar konsumsi listriknya naik.

Berdasarkan boxplot periode pembelian token tiap klasternya, memperlihatkan bahwa klaster 1 merupakan klaster dengan karakter hari pembelian yang pendek dengan rentang maksimal periode pembelian tokennya 50 hari, hal ini menunjukkan klaster 1 yang berjumlah 1143 pelanggan, merupakan klaster pelanggan aktif membeli token listrik. Klaster 4 dengan jumlah pelanggan 6356 pelanggan, menunjukkan karakteristik periode hari pembelian token yang aktif, yaitu rentang maksimalnya 70 hari pembelian. Klaster 3 dan klaster 2, dari boxplot memperlihatkan kecenderungan karakteristik periode pembelian tokennya yang hampir sama, tetapi secara rata-rata klaster 3 menunjukkan periode hari pembelian cenderung lebih aktif dibandingkan dengan klaster 2. Pergerakan boxplot periode pembelian token menunjukkan karakter periode pembelian token di tiap klasternya, sehingga dapat diambil kesimpulan klaster 2 dan klaster 3 mempunyai potensi besar untuk ditawarkan program-program yang membuat periode hari pembelian tokennya lebih aktif.

Berdasarkan boxplot rupiah pembelian token, klaster 1 masih mendominasi untuk rupiah pembelian token, rata-rata rupiah pembelian tokennya diatas angka 5 juta rupiah. Klaster dengan jumlah pelanggan yang terbesar yaitu klaster 2 dengan jumlah pelanggan 7890, justru menjadi klaster dengan rupiah pembelian terkecil. Klaster 3, menjadi klaster dengan rupiah pembelian token terbesar kedua, tetapi rata-rata nilai rupiah pembelian tokennya masih dibawah 5 juta rupiah. Sedangkan, klaster 4 nilai rupiah pembelian tokennya maksimal mencapai 5 juta, tetapi untuk rata-rata masih cenderung lebih rendah daripada klaster 3.

Berdasarkan boxplot selisih daya VA menunjukkan bawah klaster 1 untuk selisih daya VA dengan pelanggan paling banyak ada di selisih daya 900 VA, yaitu 151 pelanggan. Untuk klaster 2 sendiri, pelanggan paling banyak kenaikan selisih daya VA sebesar 450 VA yaitu 90 pelanggan. Sedangkan di klaster 3, pelanggan paling banyak melakukan tambah daya, ada di selisih daya 1300 VA, sebanyak 297 pelanggan. Klaster 4 sendiri mempunyai karakteristik dengan pelanggan paling banyak peningkatannya di selisih daya 400 VA. Klaster 3 merupakan klaster dengan kenaikan selisih daya tertinggi dibandingkan semua klaster, yaitu 1300 VA, selanjutnya klaster 1 menempati posisi kedua dengan kenaikan selisih daya sebesar 900 VA.

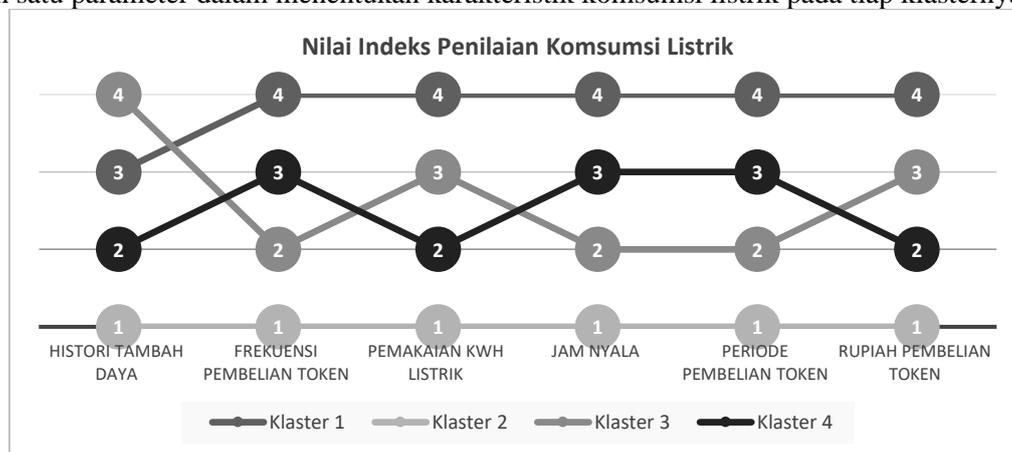
3.3 Evaluasi Profil Komsumsi Listrik

Evaluasi profil komsumsi listrik pada tiap klasternya, menggunakan matriks ranking, dimana klaster pada tiap variabelnya dengan peringkat 1 akan diberi nilai indeks 4, peringkat 2 akan diberi nilai indeks 3, peringkat 3 akan diberi nilai indeks 2 dan peringkat 4 akan diberi nilai indeks 1. Berikut adalah matriks indeks klaster pada tiap variabel yang sudah ditentukan.

Tabel 3 Matriks nilai indeks komsumsi listrik per-klaster

Klaster	Histori Tambah Daya	Frekuensi Pembelian Token	Pemakaian KWh Listrik	Jam Nyala	Periode Pembelian Token	Rupiah Pembelian Token	Modus Nilai Indeks	Karakteristik Komsumsi Listrik
Klaster 1	3	4	4	4	4	4	4	Tertinggi
Klaster 2	1	1	1	1	1	1	1	Terendah
Klaster 3	4	2	3	2	2	3	2	Agak Tinggi
Klaster 4	2	3	2	3	3	2	3	Agak Rendah

Pemberian nilai indeks pada tiap klasternya berdasarkan peringkat klaster di masing-masing variabelnya. Variabel yang digunakan dalam evaluasi profil komsumsi listrik adalah frekuensi beli token listrik, total pemakaian kWh, total rupiah pembelian token, jam nyala, periode pembelian token dan data histori tambah daya. Selanjutnya, modus nilai indeks akan dijadikan salah satu parameter dalam menentukan karakteristik komsumsi listrik pada tiap klasternya.



Gambar 4 Nilai indeks penilaian komsumsi listrik per-klaster

Klaster 1 mempunyai karakteristik komsumsi listrik tertinggi dari semua klaster, itu ditunjukkan dengan dominan nilai indeks 4 pada 5 variabel, sehingga dapat diambil kesimpulan bahwa klaster 1 mempunyai karakteristik komsumsi listrik tertinggi. Sedangkan, klaster 2 mempunyai karakteristik komsumsi terendah dari semua klaster, hal ini ditunjukkan dengan dominansi nilai indeks 1 pada 6 variabel, maka dapat diambil kesimpulan bahwa klaster 2 ini karakteristik komsumsi listriknya terendah. Untuk klaster 3 dan klaster 4, terlihat pergerakan nilai indeks nya hampir sama, tetapi apabila dilihat dari nilai indeks variabel histori tambah daya, karena klaster 3 mempunyai nilai indeks tambah daya tertinggi dibandingkan dengan klaster 4, maka dapat diambil kesimpulan klaster 3 karakteristik komsumsi listriknya masuk kategori agak tinggi, sedangkan klaster 4 dikategorikan komsumsi listriknya agak rendah.

3.4 Point Biserial Correlation

Point-Biserial Correlation adalah metode yang digunakan untuk menganalisis keeratan hubungan antara 2 variabel, dimana variabel 1 merupakan data kontinu sedangkan variabel pembanding satunya merupakan variabel nominal atau kategori. Dalam penelitian ini, variabel kontinu yang akan dianalisa korelasi nya adalah daya listrik pelanggan (VA), frekuensi beli token listrik, total pemakaian kWh, total rupiah pembelian token, jam nyala, selisih daya dan periode pembelian token, sedangkan variabel kategori atau target nya adalah data histori tambah daya.

Tabel 4 Nilai Korelasi Point Biserial

Point Biserial Correlation		
Variabel	Nilai Korelasi	
Selisih daya	0.84	84%
Daya listrik pelanggan (VA)	0.491	49.10%
Periode pembelian token	-0.29	-29%
Total pemakaian kWh	0.288	28.80%
Total rupiah pembelian token	0.295	29.50%
Frekuensi beli token	0.095	9.50%
Jam nyala	0.08	8%

Pada Tabel 4 menunjukkan bahwa variabel selisih daya mempunyai korelasi yang sangat kuat, yaitu 84%. Berdasarkan pengamatan data variabel selisih daya yang dilakukan, apabila terdapat tambah daya pada variabel histori tambah daya, maka variabel selisih daya ada nilainya, apabila tidak ada tambah daya pada variabel histori tambah daya maka nilai selisih daya nol, hal ini sangat selaras dengan variabel target histori tambah daya sehingga dalam penelitian ini variabel selisih daya tidak akan digunakan dalam model prediksi. Variabel lainnya, yaitu daya listrik Pelanggan (VA), frekuensi beli token listrik, total pemakaian kWh, total rupiah pembelian token, jam nyala dan periode pembelian token akan digunakan dalam model prediksi Artificial Neural Netwok dan Gradient Boosted.

3.5 Evaluasi Model Prediksi Tambah Daya

Empat klaster yang telah terbentuk, akan diproses target prediksi tambah daya pelanggan prabayar berdasarkan algoritma supervised learning yaitu *Artificial Neural Netwok dan Gradient Boosted*. Pengolahan algoritma tersebut menggunakan tiga skenario proporsi *data training – data testing*, yaitu 75%-25%, 80%-20% dan 60%-40%. Masing-masing klaster akan diuji melalui tiga skenario tersebut untuk mendapatkan nilai AUC terbaik. Berikut adalah hasil dari pengolahan model prediksi dengan tiga skenario proporsi data.

Tabel 5 Nilai AUC model prediksi

Cluster	Training Data – Testing Data	Machine Learning Model	AUC	Accuracy	Cohen Kappa	Error	Correct Classified	Wrong Classified
Klaster 1	75% - 25%	ANN	0.721	72.39%	0.44	27.61%	291	111
		Gradient Boosted	0.78	78.11%	0.56	21.89%	314	88
	80%-20%	ANN	0.702	70.19%	0.40	29.81%	226	96
		Gradient Boosted	0.776	77.64%	0.55	22.36%	250	72
	60%-40%	ANN	0.704	70.45%	0.41	29.55%	453	190
		Gradient Boosted	0.784	78.38%	0.57	21.62%	504	139
Klaster 2	75% - 25%	ANN	0.5	51.29%	0.00	48.71%	1964	1865
		Gradient Boosted	0.932	93.21%	0.86	6.79%	3569	260
	80%-20%	ANN	0.5	49.71%	0.00	50.29%	1523	1541
		Gradient Boosted	0.941	93.96%	0.88	6.04%	2879	185
	60%-40%	ANN	0.5	50.47%	0.00	49.53%	3092	3035
		Gradient Boosted	0.937	93.65%	0.87	6.35%	5738	389
Klaster 3	75% - 25%	ANN	0.839	84.26%	0.68	15.74%	348	65
		Gradient Boosted	0.884	88.14%	0.76	11.86%	364	49
	80%-20%	ANN	0.824	82.73%	0.65	17.27%	273	57
		Gradient Boosted	0.871	88.48%	0.76	11.52%	292	38
	60%-40%	ANN	0.835	83.64%	0.67	16.36%	552	108
		Gradient Boosted	0.884	88.33%	0.77	11.67%	583	77
	75% - 25%	ANN	0.706	70.61%	0.41	23.39%	2148	894

Klaster 4	80%-20%	Gradient Boosted	0.897	89.71%	0.79	10.29%	2729	313	
		ANN	0.7	70.01%	0.40	29.99%	1704	730	
	60%-40%	Gradient Boosted	0.903	90.30%	0.81	9.70%	2198	236	
		ANN	0.697	69.63%	0.39	30.37%	3389	1478	
			Gradient Boosted	0.885	88.49%	0.77	11.51%	4307	560

Dari hasil nilai AUC model prediksi diatas menunjukkan bahwa untuk klaster 1, model prediksi yang menghasilkan nilai AUC paling tinggi ada di model prediksi *Gradient Boosted* dengan proporsi data 60%-40%, nilai AUC dihasilkan adalah 0.784. Sedangkan untuk klaster 2, model prediksi yang menghasilkan akurasi paling tinggi adalah *Gradient Boosted* dengan proporsi data 80% -20%, dengan nilai AUC sebesar 0.941. Untuk klaster 3 sendiri, model prediksi *Gradient Boosted* dengan proporsi data 60%-40% yang menghasilkan nilai AUC sebesar 0.884. Sedangkan klaster 4, model prediksi *Gradient Boosted* dengan proporsi data 80%-20% yang menghasilkan nilai AUC yaitu 0.903.

Tabel 6 Kategori klasifikasi AUC

Klaster	Model Prediksi	Data Proporsi	Nilai AUC	Kategori
Klaster 1	Gradient Boosted	60% - 40%	0.784	Fair Classification
Klaster 2	Gradient Boosted	80% - 20%	0.941	Excellent Classification
Klaster 3	Gradient Boosted	60%-40%	0.884	Good Classification
Klaster 4	Gradient Boosted	80% - 20%	0.903	Excellent Classification

Dari nilai AUC diatas, klaster 1 dengan nilai AUC sebesar 0.784 menunjukkan bahwa kinerja model prediksi pada klaster 1 masuk kategori *fair classification*, untuk klaster 2 dengan nilai AUC sebesar 0.941 masuk kategori *excellent classification*, klaster 3 dengan nilai AUC sebesar 0.884 masuk kategori *good classification*, klaster 4 dengan nilai AUC sebesar 0.903 masuk kategori *excellent classification*. Dari hasil AUC model prediksi diatas membuktikan bahwa model prediksi berhasil dengan sangat baik. Model prediksi tambah daya yang telah dipilih pada tiap klasternya dapat digunakan sebagai model untuk menentukan pelanggan yang berpotensi tambah daya listrik, sehingga pelanggan yang terprediksi tambah daya di tiap klasternya dapat didatangi atau ditawarkan program-program tambah daya listrik.

4. KESIMPULAN DAN SARAN

Berdasarkan hasil klasifikasi dengan beberapa metode klasifikasi, didapatkan kesimpulan bahwa klaster 1 merupakan klaster dengan karakteristik komsumsi listrik tertinggi dari semua klaster, kemudian diikuti oleh klaster 3 dengan kategori konsumsi listrik agak tinggi, klaster 4 menempati posisi ketiga dengan kategori konsumsi listrik agak rendah. Untuk klaster 2 mempunyai karakteristik komsumsi kategori terendah dari semua klaster, hal tersebut ditunjukkan dengan nilai dominansi indeks 1 pada semua variabel.

Berdasarkan evaluasi model prediksi tambah daya, model prediksi pada klaster 1, yang menghasilkan nilai AUC paling tinggi ada di model prediksi *Gradient Boosted (60%-40%)*, nilai AUC dihasilkan adalah 0.784 (*Fair Classification*). Untuk klaster 2, model prediksi yang menghasilkan akurasi paling tinggi adalah *Gradient Boosted (80% -20%)*, dengan nilai AUC sebesar 0.941 (*Excellent Classification*). Untuk klaster 3, model prediksi *Gradient Boosted (60%-40%)* yang menghasilkan nilai AUC sebesar 0.884 (*Good Classification*). Sedangkan klaster 4, model prediksi *Gradient Boosted (80%-20%)* yang menghasilkan nilai AUC yaitu 0.903 (*Excellent Classification*).

Rekomendasi bisnis pada klaster 1 dan klaster 3 yang merupakan klaster dengan karakter komsumsi listrik kategori tertinggi dan agak tinggi akan menjadi klaster prioritas utama dalam implementasi program tambah daya. Sedangkan, klaster 2 dan klaster 4 tidak akan menjadi fokus utama dalam implementasi strategi bisnis tambah daya karena kecenderungan komsumsi listriknya sangat rendah sehingga minat tambah daya kecil. Rekomendasi bisnis untuk klaster 2

dan kluster 4 bersifat pemeriksaan terhadap kondisi alat pengukur dan pembatas serta instalasi di pelanggan. Untuk target prediksi pelanggan yang potensi tinggi tambah daya pada kluster 1 dan kluster 3 menggunakan model *Gradient Boosted (60%-40%)* yang hasilnya nanti akan menjadi target *door to door* pelanggan tambah daya.

Pada penelitian ini masih terbatas pada banyak hal, sehingga saran yang diberikan adalah metode klasifikasi yang digunakan bersifat adaptif yang dilengkapi dengan algoritma optimizer seperti Adaboosting dan Cross Validation. Penelitian dapat dikembangkan lebih lanjut dengan memperpanjang periode dataset yang digunakan dalam pengolahan, misalkan data diambil dalam rentang periode 5 tahun. Untuk metode klustering sendiri, dapat dikembangkan lagi dengan metode klustering lain, seperti DBScan clustering, Hierarchical clustering dan Fuzzy C-means.

DAFTAR PUSTAKA

- [1] O. Motlagh, A. Berry and L. O'Neil, "Clustering of residential electricity customers using load time series," *Elsevier BV*, vol. 237, pp. 11-24, 2019.
- [2] D. R. Thomasa, S. Agrawalb, S. Harishc, A. Mahajand and J. Urpelaine, "Understanding segmentation in rural electricity markets: Evidence from India," *Energy Economics*, vol. 87, p. 104697, 2020.
- [3] H. Abu-Bakar, L. Williams and S. H. Hallett, "An empirical water consumer segmentation and the characterisation of consumption patterns underpinning demand peaks," *Elsevier*, vol. 174, p. 105792, 2021.
- [4] A. Al-Wakeel, J. Wu and N. Jenkins, "k-means based load estimation of domestic smart meter measurements," *Elsevier BV*, vol. 194, pp. 333-342, 2017.
- [5] F. Wang, X. Lu, X. Chang, X. Cao, S. Yan, K. Li, N. Duic, M. Shafie-khah and J. P. Catalao, "Household profile identification for behavioral demand response: A semi-supervised learning approach using smart meter data," *Energy*, vol. 238, p. 121728, 2022.
- [6] X. Kong, X. Zhao, C. Liu, Q. Li, D. Dong and Y. Li, "Electricity theft detection in low-voltage stations based on similarity measure and DT-KSVM," *International Journal of Electrical Power and Energy Systems*, vol. 125, p. 106544, 2021.
- [7] R. Razavia, A. Gharipourb, M. Fleuryc and I. J. Akpan, "A practical feature-engineering framework for electricity theft detection in smart grids," *Applied Energy*, vol. 238, p. 481–494, 2019.
- [8] D. Jaiswal, V. Kaushal, P. K. Singh and A. Biswas, "Green market segmentation and consumer profiling: a cluster approach to an emerging consumer market," *Emerald Group Publishing Ltd*, vol. 28, no. 3, pp. 792-812, 2021.
- [9] S. Hussain, M. W. Mustafa, T. A. Jumani, S. K. Baloch, H. Alotaibi, I. Khan and A. Khan, "A novel feature engineered-CatBoost-based supervised machine learning framework for electricity theft detection," *Energy Reports*, vol. 7, p. 4425–4436, 2021.