

Pemodelan Loyalitas Konsumen Susu Pertumbuhan dalam Mengikuti Program *Rewards* Menggunakan Metode *Random Forest* dan *Neural Network*

Ayunda Pratiwi*, Khairil Anwar Notodiputro*, Hari Wijayanto*

*Departemen Statistika Institut Pertanian Bogor

Abstrak—Persaingan bisnis di Indonesia yang semakin kompetitif menjadi alasan perusahaan untuk membuat strategi dalam mempertahankan konsumennya, salah satunya dengan bantuan program loyalitas. Fakta menunjukkan bahwa pada salah satu program *rewards* di perusahaan nutrisi, 37.62% *member* yang terdaftar telah meninggalkan program atau biasa disebut dengan istilah *churn*. Untuk mengantisipasi hal tersebut, dibuat suatu model prediksi (klasifikasi) loyalitas pelanggan berdasarkan profil dan aktivitasnya dalam program selama masa prediksi. Model klasifikasi dibangun pada dua *brand* terbesar dengan metode *random forest* dan *neural network*. Kedua metode tersebut dievaluasi dan dibandingkan berdasarkan kurva ROC (*Relative Operating Characteristics*) dengan melihat nilai *area under curve* (AUC). Kinerja dari kedua metode tersebut tidak berbeda nyata, namun metode *neural network* menghasilkan nilai AUC yang lebih besar, baik pada pemodelan *brand A* maupun *brand B*.

Kata kunci-*churn*; klasifikasi; loyalitas; *neural network*; *random forest*

I. PENDAHULUAN

A. Latar Belakang

Persaingan bisnis yang semakin kompetitif menjadi alasan perusahaan untuk membuat strategi dalam mempertahankan konsumen dengan membentuk suatu program loyalitas. Akan tetapi berdasarkan kasus yang diperoleh dari salah satu program *rewards* perusahaan nutrisi, 36.72% *member* yang terdaftar meninggalkan program tersebut atau biasa disebut dengan istilah *churn*. Menurut Ngai et al. (2009), menganalisis dan memahami perilaku dan karakteristik pelanggan merupakan dasar pengembangan strategi yang kompetitif. Oleh karena itu, salah satu cara perusahaan untuk mempertahankan konsumennya ialah

mengantisipasi adanya *churned member* dengan melakukan prediksi dan memberikan strategi yang tepat untuk mempertahankannya.

Penelitian sebelumnya mengenai prediksi loyalitas dan *churned customer* dilakukan oleh Yudhistira (2017) pada salah satu perusahaan *Internet Service Provider* dengan menggunakan metode *neural network* algoritma *backpropagation* dan menghasilkan akurasi sebesar 99%. Sementara itu, Dror et al. (2012) melakukan prediksi *churn* pada pengguna baru Yahoo Answer dengan membandingkan beberapa pendekatan dan hasilnya menunjukkan bahwa *random forest* menghasilkan nilai *area under curve* (AUC) yang paling tinggi dibandingkan metode lainnya. Pada penelitian ini dibangun model klasifikasi loyalitas *member* dalam mengikuti program *rewards* pada salah satu perusahaan nutrisi dengan membandingkan metode *random forest* dengan *neural network*.

Sesuai dengan uraian di atas, untuk menjamin efektivitas dan efisiensi penelitian, maka pertanyaan-pertanyaan berikut akan dijawab.

- 1) Seperti apa karakteristik konsumen susu formula?
- 2) Karakteristik konsumen seperti apa yang merupakan konsumen loyal atau tidak loyal?
- 3) Peubah-peubah apa yang menentukan loyalitas dari konsumen?
- 4) Sejauh mana metode *random forest* mampu memberikan prediksi yang akurat?
- 5) Sejauh mana metode *neural network* mampu memberikan prediksi yang akurat?
- 6) Bagaimana perbandingan kinerja antara metode *random forest* dan *neural network*?

B. Tujuan

Tujuan penelitian ini untuk membangun model klasifikasi loyalitas konsumen dengan membandingkan pendekatan *random forest* dengan *neural network* berdasarkan data profil dan aktivitas *member* dalam program selama masa prediksi.

II. TINJAUAN PUSTAKA

A. Loyalitas dan Program Loyalitas

Dick and Basu (1994) menggambarkan loyalitas sebagai kekuatan hubungan antara sikap individu dan pengulangan pembelian dengan melibatkan emosi, suasana hati, pengaruh uselera, serta kepuasan dari pelanggan. Adapun upaya peningkatan loyalitas konsumen dilakukan dengan membentuk suatu program loyalitas. Shoemaker and Lewis (1999) mendefinisikan program loyalitas pelanggan (*loyalty programs*) sebagai program yang ditawarkan kepada pelanggan yang bertujuan untuk membangun ikatan emosional terhadap perusahaan.

Terdapat beberapa istilah yang berbeda terkait program loyalitas meskipun pada dasarnya memiliki prinsip yang sama seperti *bonus program*, *customer club*, *membership card* dan lain sebagainya. Jenis-jenis dari program loyalitas juga bermacam-macam, misalnya program poin, bonus, kemitraan dan beberapa perusahaan juga menggunakan kombinasinya.

B. Random Forest

Metode *random forest* adalah pengembangan dari metode CART (*Classification and Regression Tree*) dengan menerapkan metode *bootstrap aggregating (bagging)* dan *random feature selection* Breiman (2001). *Random Forest* memprediksi respon berdasarkan kelas yang paling sering muncul sebagai hasil prediksi dari k pohon klasifikasi.

Pada gugus data yang terdiri atas n amatan dan p peubah penjelas, prosedur untuk melakukan *random forest* adalah (Breiman 2001):

- 1) Melakukan penarikan contoh acak berukuran n dengan pemulihan pada gugus data. Tahap ini adalah tahapan *bootstrap*.

- 2) Dengan menggunakan contoh *bootstrap*, pohon dibangun sampai mencapai ukuran maksimum (tanpa pemangkasan). Pembangunan pohon dilakukan dengan menerapkan *random feature selection* pada setiap proses pemilihan pemilah, yaitu m peubah penjelas dipilih secara acak dimana $m < p$, lalu pemilah terbaik dipilih berdasarkan m peubah penjelas tersebut.
- 3) Ulangi langkah 1 dan 2 sebanyak k kali sehingga terbentuk sebanyak k pohon.

C. Neural Network

Neural network atau jaringan syaraf tiruan (JST) adalah salah satu teknik klasifikasi yang dapat memodelkan sistem yang bersifat tidak linear dan kompleks. Karakteristik model ini ditentukan oleh arsitektur jaringan, fungsi aktivasi, dan metode penentuan bobot atau proses pelatihan. Arsitektur jaringan menempatkan neuron pada lapisan yang terdiri dari lapisan masukan, lapisan keluaran, dan jika ada lapisan tersembunyi yang jumlahnya satu atau lebih.

Algoritma pelatihan *backpropagation* atau propagasi balik termasuk metode pelatihan terbimbing (*supervised*) yang secara umum memiliki dua tahapan, yaitu tahapan perambatan maju dan perambatan mundur. Mekanisme pelatihan algoritma *backpropagation* dimulai dengan menginisiasi bobot awal dan membandingkan keluaran yang dihasilkan dengan target aslinya untuk memperoleh sisaan. Sisaan tersebut dimanfaatkan dalam memperbaiki bobot pada iterasi selanjutnya sampai sisaan yang dihasilkan kurang dari batas ambang sisaan yang ditentukan.

Dalam algoritma *backpropagation*, fungsi aktivasi yang dipakai harus memenuhi syarat kontinu dan dapat dideferensialkan, salah satunya adalah fungsi sigmoid biner Kusumadewi (2004). Fungsi sigmoid biner memiliki nilai pada interval 0 sampai 1. Oleh karena itu, fungsi ini digunakan untuk jaringan syaraf yang membutuhkan nilai *output* yang terletak pada interval 0 sampai 1. Fungsi sigmoid biner dirumuskan sebagai berikut:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (1)$$

dengan turunannya,

$$f'(x) = f(x)(1 - f(x)) \quad (2)$$

Resilient Backpropagation (RPROP) adalah salah satu pengembangan dari *backpropagation neural network* dalam perbaikan bobot. RPROP memperkecil besarnya efek turunan parsial dengan cara hanya memanfaatkan tanda turunan dan mengabaikan besarnya nilai turunan. Besarnya perubahan setiap bobot ditentukan oleh suatu faktor yang disebut faktor naik dan faktor turun.

D. Ukuran Kinerja Model Klasifikasi

Kurva ROC (*Receiver Operating Characteristic*) merupakan salah satu cara untuk melihat ukuran kinerja model klasifikasi dari hasil plot antara sensitivitas dan (1 - spesifisitas) untuk semua kemungkinan nilai π_0 . Ketepatan model dalam memprediksi kejadian gagal ($y=0$) dinyatakan sebagai spesifisitas sedangkan kemampuan model memprediksi kejadian berhasil ($y=1$) disebut sensitivitas Hosmer et al. (2013).

Menurut Agresti (2007), kurva ROC lebih informatif daripada tabel klasifikasi karena merangkum kekuatan prediktif untuk semua kemungkinan π_0 . Nilai luas daerah di bawah kurva menunjukkan keakuratan model dalam mengklasifikasikan data. Semakin besar nilai tersebut menunjukkan kinerja yang semakin baik.

III. METODOLOGI

A. Data

Data yang digunakan pada penelitian ini adalah data profil dan aktivitas *member* program loyalitas suatu produk nutrisi yang dihimpun sampai dengan tanggal 31 Desember 2017. Objek dari data tersebut ialah seluruh *member* di Indonesia yang terdaftar dalam program minimal satu tahun. *Member* diasumsikan adalah orang tua dari anak yang mengonsumsi susu pertumbuhan dan keanggotannya dalam program dalam program bersifat sukarela (*voluntary*). Pemodelan klasifikasi dilakukan pada dua *brand* terbesar, yaitu *brand A* yang terdiri dari 14351 observasi dan *brand B* yang terdiri dari 7453 observasi dengan 7 peubah penjelas dan 1 peubah respon atau kelas. Peubah

bebas memanfaatkan informasi profil dan aktivitas *member* di bulan pertama.

Tabel I
DAFTAR PEUBAH PENJELAS

Peubah	Keterangan	Skala
Respon	Kategori loyalitas <i>member</i> , bernilai 1 untuk loyal dan 0 untuk <i>churn</i>	Kategorik (Biner)
Jumlah Anak	Jumlah anak dari <i>member</i>	Numerik
Usia Anak	Usia anak yang mengonsumsi susu pertumbuhan	Numerik
Gram	Total gram yang dibeli dalam satu bulan pertama (masa prediksi)	Numerik
Web	Saluran yang digunakan saat pertama kali mendaftar program loyalitas. Bernilai 1 untuk web dan 0 untuk selainya.	Kategorik (Biner)
Total <i>member</i>	Banyaknya <i>member</i> yang terdaftar pada masing-masing provinsi. <i>Member</i> yang berasal dari provinsi yang sama memiliki nilai peubah yang sama.	Numerik
Persentase <i>member</i> loyal	Persentase <i>member</i> loyal pada masing-masing provinsi. <i>Member</i> yang berasal dari provinsi yang sama memiliki nilai peubah yang sama.	Numerik
<i>Regtrans</i>	Bernilai 1 untuk <i>member</i> yang melakukan transaksi setelah registrasi dan bernilai 0 untuk <i>member</i> yang tidak melakukan transaksi setelah registrasi di hari yang sama	Kategorik (Biner)

B. Prosedur Analisis Data

Tahapan analisis dalam penelitian ini adalah sebagai berikut:

- 1) Melakukan praproses data dengan cara standarisasi data seperti pada persamaan (3) untuk pemodelan *neural network*

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_i}{s(x_i)} \quad (3)$$

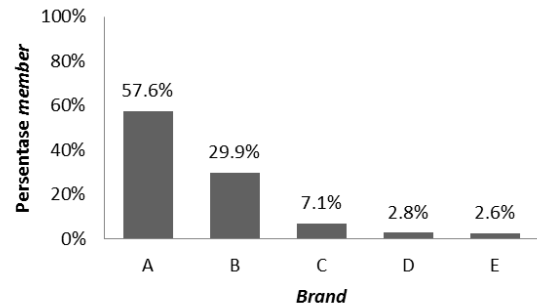
dengan i adalah 7 peubah bebas yang digunakan dan j adalah observasi pada masing-masing model *brand A* dan *B*.

- 2) Melakukan eksplorasi data dengan statistika deskriptif pada masing-masing peubah untuk mengetahui gambaran konsumen susu pertumbuhan dalam program *rewards* tersebut berdasarkan kriteria loyal dan *churn* pada masing-masing *brand*.
- 3) Membagi data menjadi data latih (70%) dan data uji (30%)
- 4) Melakukan proses klasifikasi dengan metode *random forest* dengan menentukan parameter banyak pohon yang paling optimal.
- 5) Melakukan proses klasifikasi dengan metode *neural network*:
 - a) Menentukan node lapisan masukan dan keluaran. Jumlah unit lapisan masukan sama dengan jumlah peubah bebas yang digunakan, yakni sebanyak 7 unit dan satu unit pada lapisan keluaran.
 - b) Menyusun jaringan dengan satu lapisan tersembunyi yang dicobakan pada kondisi 1 sampai dengan 8 neuron.
 - c) Menyusun jaringan dengan dua lapisan tersembunyi yang dicobakan dengan kombinasi sampai dengan 5 neuron di lapisan pertama dan 2 neuron di lapisan kedua.
 - d) Melakukan proses pelatihan dengan arsitektur jaringan yang sudah disusun.
- 6) Mengevaluasi kinerja model *random forest* dan *neural network* dengan menggunakan kurva ROC untuk menentukan model terbaik.

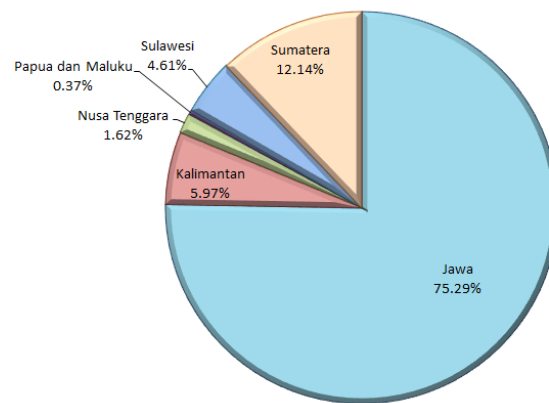
IV. HASIL DAN PEMBAHASAN

A. Karakteristik Konsumen

Berdasarkan data pada program *rewards*, *member* tersebar pada 5 *brand* susu pertumbuhan dengan proporsi seperti pada Gambar 1. *Brand* A dan B memiliki proporsi *member* terbesar dengan total 87%. Hal ini yang menjadi dasar untuk dilakukan eksplorasi lebih lanjut serta pemodelan klasifikasi untuk kedua *brand* tersebut. *Brand* A memiliki 14351 *member* dengan 63.2% *member* loyal sedangkan *brand* B memiliki 7453 *member* dengan 58.5% *member* loyal.



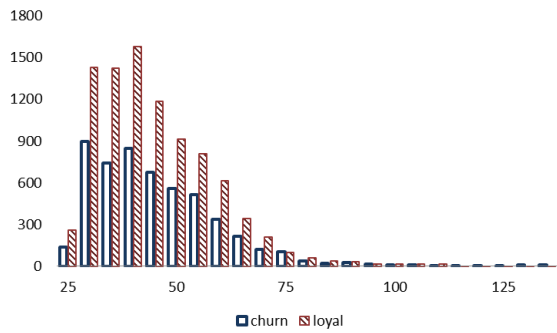
Gambar 1. Persentase *member* pada setiap *brand*



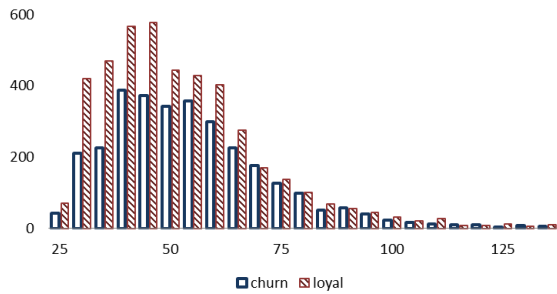
Gambar 2. Diagram persentase *member* berdasarkan pulau

Dari 33 provinsi di Indonesia. Jawa Timur merupakan provinsi dengan jumlah *member* paling banyak, yaitu sebanyak 5646 atau 22.65% dari keseluruhan *member*, disusul oleh Jawa Barat, Jakarta, Jawa Tengah, dan Banten sebagai 5 besar provinsi dengan *member* terbanyak. Sementara itu, provinsi yang memiliki jumlah *member* paling sedikit ialah provinsi Papua Barat. Penyebaran *member* berdasarkan pulau pada Gambar 2 menunjukkan bahwa pulau Jawa merupakan pulau dengan persentase *member* terbesar, yaitu lebih dari 75%. Hal tersebut didukung karena proses pemasaran program *rewards* dan distribusi produk memang lebih aktif dipasarkan di pulau Jawa.

Selain banyaknya *member* pada tiap provinsi, persentase *member* loyal tiap provinsi juga cukup bervariasi. Provinsi yang memiliki persentase *member* loyal yang paling tinggi adalah Kalimantan Utara, yaitu sebanyak 73.33% *member* terma-



Gambar 3. Histogram usia anak dari member pada brand A



Gambar 4. Histogram usia anak dari member pada brand B

suk ke dalam kategori loyal. Dari 33 provinsi, hanya 2 provinsi yang memiliki persentase member loyal dibawah 50%, yaitu Sulawesi Barat sebesar 46.67% dan Papua Barat sebesar 42.86%. Pencapaian program pada wilayah dapat dikatakan baik ketika jumlah member dalam provinsi tersebut banyak dan diiringi dengan proporsi member loyal yang tinggi.

Dari data diketahui mayoritas member memiliki satu anak sedangkan rentang usia anak berada di antara 17 bulan sampai dengan 143 bulan atau hampir berusia 12 tahun. Gambar 3 dan 4 menunjukkan sebaran dari usia anak brand A dan B. Dapat dilihat bahwa sebaran usia anak pada kedua brand menjulur ke kanan serta terdapat persamaan pola sebaran usia anak untuk kategori loyal dan churn. Frekuensi usia anak yang mengonsumsi brand A paling banyak berada di sekitar usia 2 sampai dengan 4 tahun. Sementara frekuensi usia anak yang mengonsumsi brand B paling banyak berusia 3 sampai 4.5 tahun.

Setiap member dapat mengakses program dengan beberapa pilihan saluran, antara lain melalui pesan singkat atau biasa dikenal dengan SMS, call center, atau web. Web merupakan saluran yang paling banyak digunakan oleh member untuk keseluruhan brand hingga mencapai 76%. Selain merupakan saluran yang paling banyak digunakan, persentase member loyal dari pengguna web juga paling tinggi dibandingkan saluran lainnya.

Karakteristik member yang berbeda dimanfaatkan sebagai peubah penjelas untuk memprediksi loyalitas pelanggan susu pertumbuhan. Data aktivitas transaksi pada bulan pertama dimanfaatkan sehingga didapatkan informasi besarnya pembelian susu pertumbuhan dengan satuan gram. Rata-rata pembelian susu pertumbuhan berdasarkan loyalitas dan brand disajikan pada Tabel II. Dari tabel tersebut diketahui bahwa rata-rata pembelian bulan pertama member loyal lebih besar dibandingkan dengan pembelian churned member. Selain itu, rata-rata pembelian pada brand A lebih besar dibandingkan dengan brand B.

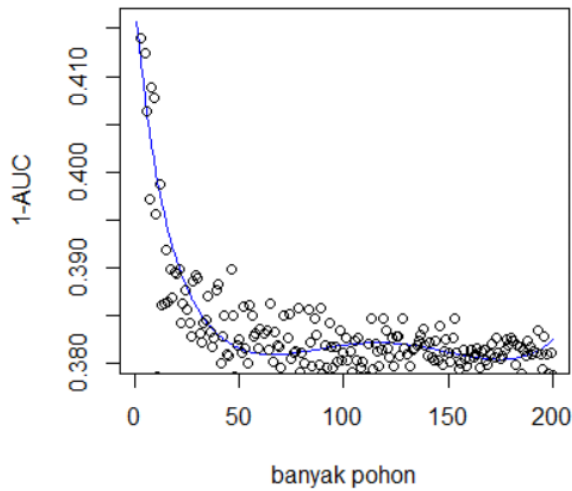
Tabel II
RATA-RATA PEMBELIAN (DALAM GRAM)

Kategori	Brand A	Brand B
Loyal	9071	4357
Churn	5279	3096

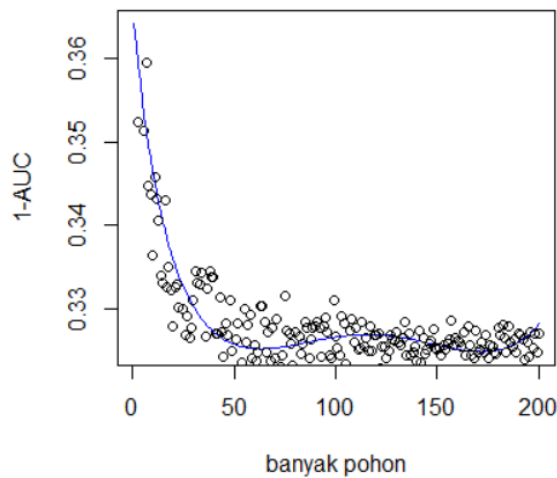
B. Pemodelan Klasifikasi dengan Random Forest

Pemodelan dilakukan pada 70% data dan dievaluasi menggunakan kurva ROC. Evaluasi dilakukan pada data latih yang disebut dengan data internal dan pada data uji yang disebut dengan data eksternal. Pada tahapan pemodelan dengan metode random forest, beberapa kemungkinan banyak pohon dicobakan untuk menentukan model yang optimal pada kedua brand. Dari Gambar 5 dan 6 dapat dilihat bahwa besarnya nilai area di atas kurva ROC mulai stabil ketika pohon yang digunakan sebanyak 50 pohon, baik pada brand A maupun brand B.

Kinerja yang dihasilkan dengan 50 pohon tidak berbeda signifikan dibandingkan dengan jumlah



Gambar 5. Plot antara 1-AUC dan banyak pohon pada model brand A

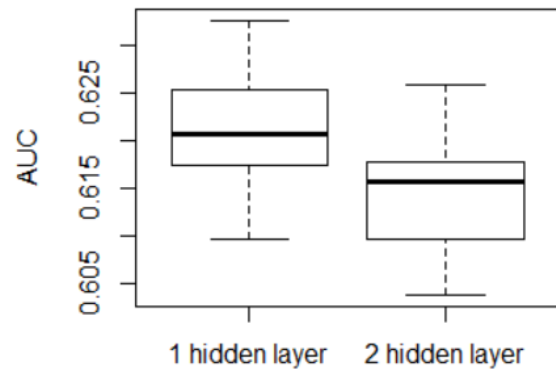


Gambar 6. Plot antara 1-AUC dan banyak pohon pada model brand B

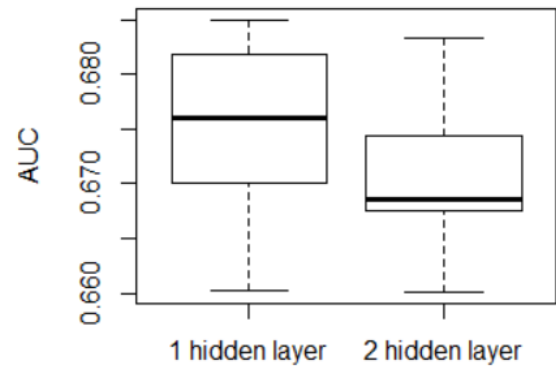
pohon yang lebih banyak. Oleh karena itu, pemodelan klasifikasi *random forest* pada *brand A* dan *B* cukup dilakukan dengan membangun 50 pohon.

C. Pemodelan Klasifikasi dengan Neural Network

Metode *random forest* akan dibandingkan dengan metode *neural network* dengan melihat kinerjanya dalam memprediksi data uji. Pemodelan neural network dilakukan dengan percobaan satu dan dua lapisan tersembunyi. Menurut Tufféry (2011), penambahan lapisan tersembunyi dapat



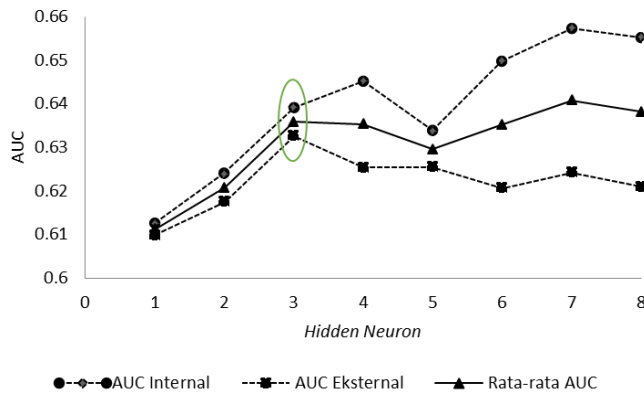
Gambar 7. Perbandingan kinerja berdasarkan lapisan tersembunyi pada model brand A



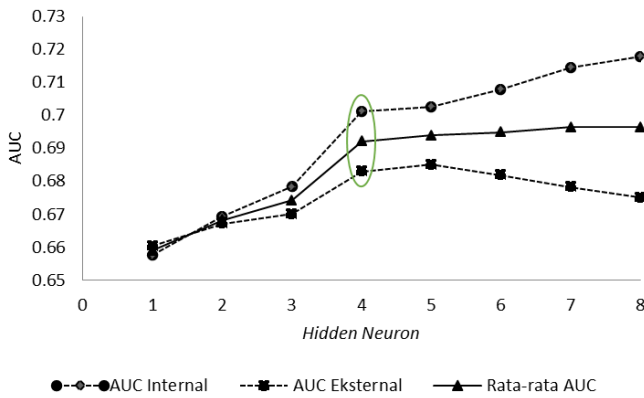
Gambar 8. Perbandingan kinerja berdasarkan lapisan tersembunyi pada model brand B

meningkatkan kinerja prediksi namun banyaknya lapisan harus seminimal mungkin untuk menghindari adanya *overfitting*. Gambar 7 dan 8 menunjukkan bahwa model dengan dua lapisan tersembunyi tidak menunjukkan kinerja lebih baik dibandingkan dengan model satu lapisan tersembunyi. Selanjutnya pemodelan dilakukan dengan mencoba beberapa jumlah neuron dan didapatkan nilai AUC dari model klasifikasi *brand A* seperti pada Gambar 9.

Dari Gambar 9 dapat dilihat bahwa nilai AUC internal pada *brand A* semakin meningkat dengan penambahan neuron pada lapisan tersembunyi sampai dengan 4 neuron, kemudian terjadi penurunan AUC ketika menggunakan sebanyak 5 neu-



Gambar 9. Plot antara AUC dan banyaknya neuron di lapisan tersembunyi pada model brand A



Gambar 10. Plot antara AUC dan banyaknya neuron di lapisan tersembunyi pada model brand B

ron dan kemudian AUC internal terus meningkat. Sementara itu, kinerja dalam memprediksi data eksternal paling baik ketika neuron yang digunakan sebanyak 3. Oleh karena itu, pemodelan pada brand A dilakukan cukup dengan menggunakan node sebanyak 3.

Sedangkan pada Gambar 10, nilai AUC eksternal pada brand B semakin meningkat dengan penambahan neuron pada lapisan tersembunyi sampai dengan 5 neuron, kemudian AUC eksternal turun ketika banyaknya neuron yang digunakan lebih dari 5. Sementara itu, nilai AUC internal terus meningkat dengan penambahan neuron pada lapisan tersembunyi. Namun tidak terdapat perbedaan yang signifikan antara AUC yang dihasilkan dengan dari model dengan 4 neuron dan 5 neuron

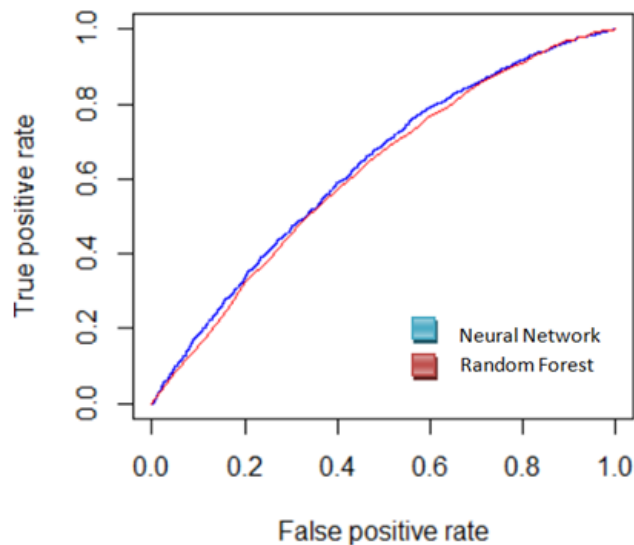
sehingga model dipilih dengan menggunakan satu lapisan tersembunyi dengan neuron sebanyak 4.

D. Perbandingan Random Forest dengan Neural Network

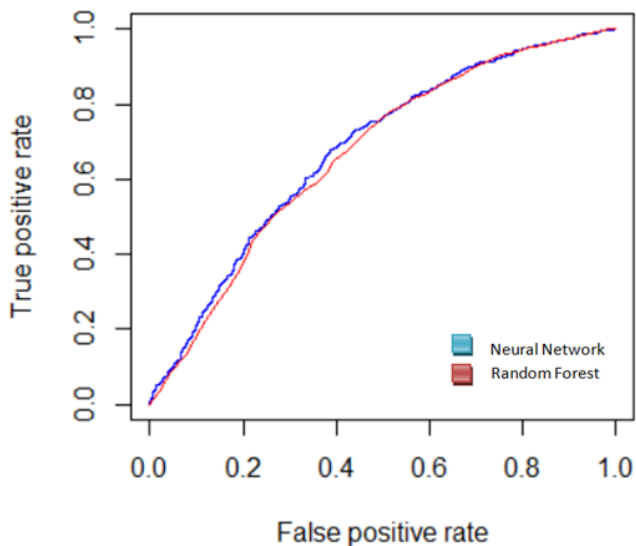
Metode *random forest* dan *neural network* yang keduanya termasuk ke dalam metode non parametrik memiliki kinerja yang berbeda dalam memprediksi loyalitas *member*. Metode *random forest* menghasilkan 50 pohon klasifikasi melalui tahap *bootstrap* dan *random feature selection*, sedangkan model *neural network* menghasilkan bobot pada setiap masukan untuk diteruskan ke lapisan selanjutnya. Arsitektur akhir yang dibentuk ialah jaringan dengan tiga lapisan, yaitu lapisan masukan, satu lapisan tersembunyi, dan lapisan keluaran dengan banyaknya neuron pada lapisan tersembunyi sebanyak 3 untuk brand A dan sebanyak 4 untuk brand B.

Kelebihan metode *random forest* dibandingkan dengan metode *neural network* ialah dari segi waktu komputasi. Metode *neural network* relatif membutuhkan waktu komputasi yang lebih lama. Selain itu, terdapat kesulitan dalam interpretasi baik pada metode *random forest* maupun *neural network*. Pada pohon klasifikasi tunggal, interpretasi dapat dilakukan secara umum bagaimana peubah-peubah bebas menjelaskan kategorinya sedangkan pada metode *random forest*, interpretasi sulit dilakukan karena model dibangun dari banyak model di belakangnya. Kompleksnya arsitektur jaringan juga membuat interpretasi hasil dari *neural network* semakin sulit untuk dilakukan.

Metode *random forest* dan *neural network* dibandingkan dengan melihat dari nilai *area under curve* (AUC). Metode yang memiliki AUC lebih besar merupakan metode yang lebih baik. Dari Gambar 11 dan 12 terlihat bahwa kinerja metode *neural network* tidak jauh berbeda dengan *random forest*. Namun metode *neural network* memiliki nilai AUC yang lebih besar untuk model brand A maupun brand B. Pada model klasifikasi loyalitas brand A, metode *neural network* memiliki nilai AUC sebesar 63.02% sementara metode *random forest* memiliki nilai AUC sebesar 61.69%. Hal yang sama juga terjadi untuk model brand B, nilai



Gambar 11. Kurva ROC pada model klasifikasi loyalitas brand A



Gambar 12. Kurva ROC pada model klasifikasi loyalitas brand B

AUC dari metode *neural network* adalah sebesar 67.91%, lebih besar dibandingkan nilai AUC dari metode *random forest*, yaitu 66.77%.

V. SIMPULAN

Metode *random forest* dan *neural network* memiliki kinerja memprediksi yang tidak jauh berbeda baik pada pemodelan *brand A* maupun

brand B. Namun, metode *neural network* memiliki nilai AUC yang lebih besar dengan selisih 1.33% pada pemodelan *brand A* dan 1.14% pada pemodelan *brand B*. Akan tetapi, metode *neural network* membutuhkan waktu komputasi yang relatif lebih lama dibandingkan metode *random forest*. Kelebihan dari kedua metode tersebut adalah tidak adanya asumsi, namun kelemahannya dalam segi interpretasi juga sulit dilakukan.

DAFTAR PUSTAKA

- Agresti, A. (2007). Logistic regression. *An Introduction to Categorical Data Analysis, Second Edition*, 99–136.
- Breiman, L. (2001). Random forests. *Machine learning* 45(1), 5–32.
- Dick, A. S. and K. Basu (1994). Customer loyalty: toward an integrated conceptual framework. *Journal of the academy of marketing science* 22(2), 99–113.
- Dror, G., D. Pelleg, O. Rokhlenko, and I. Szpektor (2012). Churn prediction in new users of yahoo! answers. In *Proceedings of the 21st International Conference on World Wide Web*, pp. 829–834. ACM.
- Hosmer, J. D. W., S. Lemeshow, and R. X. Sturdivant (2013). *Applied logistic regression*, Volume 398. John Wiley & Sons.
- Kusumadewi, S. (2004). Membangun jaringan syaraf tiruan menggunakan matlab dan excel link. *Yogyakarta: Graha Ilmu*.
- Ngai, E. W. T., L. Xiu, and D. C. K. Chau (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert systems with applications* 36(2), 2592–2602.
- Shoemaker, S. and R. C. Lewis (1999). Customer loyalty: the future of hospitality marketing. *International journal of hospitality management* 18(4), 345–370.
- Tufféry, S. (2011). *Data mining and statistics for decision making*, Volume 2. Wiley Chichester.
- Yudhistira, A. R. (2017). Analisa customer churn pada perusahaan internet service provider xyz menggunakan backpropagation neural network.