

Prediksi Klasifikasi Pembangunan Merek Kosmetik Dengan Metode Enbag K-Logres Berdasarkan Keterlibatan Pengguna Facebook

Fajar Sarasati¹, F. Lia Dwi Cahyanti², Annida Purnamawati³, Riyan Latifahul Hasanah⁴

¹Ilmu Komputer, STMIK Nusa Mandiri

^{2,3}Prodi Sistem Informasi, Universitas Bina Sarana Informatika

⁴Prodi Teknologi Komputer, Universitas Bina Sarana Informatika

¹fajar.fss@nusamandiri.ac.id, ²flia.fdc@bsi.ac.id, ³annida.npm@bsi.ac.id, ⁴riyan.rlt@bsi.ac.id

Abstract: Building a brand new company that starts a business by conducting market research is intended to introduce new products and maintain existing businesses. But the market survey actually requires quite a lot of costs for transportation costs, brochure printing costs, more employee salaries and so forth. Surveys conducted offline also reach a less extensive market, less maximum results and less detail, and require more time. Based on the description above, the researchers conducted a study using Facebook performance metric data that assessed the construction of cosmetics brands using the K-Nearest Neighbor and Logistics Regression (SVM) algorithm by classifying which posts were the most desirable and less desirable by consumers, as well as measuring by the EnBag method K-LoGres of the two algorithms to improve the performance of the two proposed algorithms. Bagging technique was chosen because it has the advantage of being able to improve the measurement results and improve the accuracy of classification measurements by combining two or more algorithms. Based on the measurement results of Facebook metric data which assesses the development of cosmetic brands with the K-NN algorithm it gets an accuracy of 68.67% and a Logistic Regression (SVM) of 72.67% then the two algorithms are processed using the EnBag K-LoGres method getting an accuracy of 73.91%. Based on the results of measurements with the EnBag K-LoGres method the results increased by 1.24%.

Keywords: Brand Development, Cosmetics, K-Nearest-Neighbour, Logistic (SVM), EnBag K-Logres

Abstrak: Membangun merek perusahaan yang baru memulai usaha dengan melakukan riset pasar dimaksudkan untuk memperkenalkan produk baru serta mempertahankan usaha yang sudah ada. Namun survei pasar justru membutuhkan biaya yang cukup banyak untuk biaya transportasi, biaya cetak brosur, gaji karyawan lebih banyak dan lain sebagainya. Survei yang dilakukan secara *offline* juga menjangkau pasar kurang luas, hasil kurang maksimal dan kurang merinci, serta membutuhkan waktu yang lebih lama. Berdasarkan uraian diatas maka peneliti melakukan penelitian dengan memanfaatkan data metrik kinerja facebook yang menilai pembangunan merk kosmetik dengan menggunakan algoritma *K-Nearest Neighbour* dan *Logistic Regression (SVM)* dengan mengklasifikasikan postingan mana yang paling diminati dan kurang diminati oleh konsumen, serta melakukan pengukuran dengan metode *EnBag K-LoGres* dari kedua algoritma untuk meningkatkan kinerja kedua algoritma yang diusulkan. Teknik bagging dipilih karena memiliki kelebihan dapat memperbaiki hasil pengukuran serta meningkatkan akurasi dari pengukuran klasifikasi dengan menggabungkan dua atau lebih algoritma. Berdasarkan hasil pengukuran data metrik facebook yang menilai pembangunan merek kosmetik dengan algoritma K-NN memperoleh akurasi sebesar 68.67% dan *Logistic Regression (SVM)* sebesar 72.67% selanjutnya kedua algoritma diproses dengan metode *EnBag K-LoGres* mendapat akurasi sebesar 73.91%. Berdasarkan hasil pengukuran dengan metode *EnBag K-LoGres* hasilnya mengalami kenaikan sebesar 1.24 %.

Kata kunci: Pembangunan Merek, Kosmetik, K-Nearest Neighbour, Logistic Regression (SVM), *EnBag K-LoGres*



This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2019 by author and IJSE-Indonesian Journal on Software Engineering.

A. PENDAHULUAN

Media sosial menjadi salah satu bukti perkembangan teknologi yang semakin canggih saat ini, jenis konten internet ini memiliki peran penting sebagai alat komunikasi dimana setiap pengguna dapat berbagi informasi, pengetahuan dan saling terhubung. Sosial media merupakan konsep ruang digital dimana setiap pengguna dapat membuat rangkuman profil, mendeskripsikan dirinya untuk berinteraksi dengan orang-orang dari kalangan berbeda, baik dalam lingkungan antar individu maupun dengan perusahaan (Santoso, dkk:2017).

Perkembangan teknologi yang semakin canggih saat ini telah merambah di hampir seluruh lini kehidupan tak terkecuali pada bidang bisnis yang membutuhkan saluran penting untuk menjangkau konsumen dan *client*.

Berdasarkan hasil survei dari APJII (2016:1) mengungkapkan bahwa ada tiga besar konten internet yang dikonsumsi pengguna, yakni media sosial sebanyak 129,2 juta (97,7%), hiburan 128,4 juta (96,8%) dan berita 127,9 juta (96,4%). Sisanya konten pendidikan 124 juta pengguna, komersial 123,5 juta dan layanan publik 121,5 juta. Terkait media sosial, Facebook yang terbesar dengan 71,6 juta pengguna (54%). Disusul Instagram dan Youtube, masing-masing dengan 19,9 juta pengguna dan 14,5 juta. Berdasarkan data tersebut maka dapat menjadi potensi yang besar bagi perusahaan untuk membangun merek serta memasarkan produknya dengan memanfaatkan media sosial.

Melihat hasil survei dari APJII tahun 2016 media sosial Facebook menjadi salah satu media sosial yang memiliki potensi besar dalam bidang perekonomian khususnya bisnis, banyaknya pengguna Facebook memiliki dampak langsung untuk meningkatkan nilai perusahaan dan membangun merek perusahaan sehingga secara tidak langsung akan berdampak pada pendapatan perusahaan.

Bagi wanita, produk kosmetik selalu menjadi bagian dari kehidupan sehari-hari, demi mendapatkan dan mempertahankan kecantikan dari waktu ke waktu. Kondisi ini dimanfaatkan betul oleh produsen kosmetik. Jumlah penduduk sekitar 250 juta jiwa menjadikan Indonesia pasar yang menjanjikan bagi perusahaan kosmetik (Briliani *et al*, 2016:545)

Namun untuk membangun merek produk kosmetik yang banyak diminati sekalipun, perusahaan harus melakukan survei pasar untuk meminimalisir kerugian produksi, karena tidak semua produk diminati oleh konsumen serta banyaknya pesaing sehingga membutuhkan survei pasar mengenai produk mana yang paling diminati dan dibutuhkan oleh pelanggan. Sedangkan untuk melakukan survei pasar perusahaan membutuhkan biaya yang cukup banyak untuk biaya transportasi, biaya cetak brosur, gaji karyawan lebih banyak dan lain sebagainya. Selain itu apabila survei dilakukan secara *offline* jangkauan pasar kurang luas, hasil kurang maksimal dan kurang merinci, serta membutuhkan waktu yang lebih lama. Media sosial Facebook dipilih karena mempunyai potensi yang besar untuk memasarkan produk mengingat jumlah penggunanya yang masih menempati posisi pertama dari sekian banyak media sosial.

Berdasarkan eksperimen yang dilakukan oleh (Moro *et al.*, 2016) dengan data yang sama menghasilkan metrik kinerja berbasis visualisasi lebih sulit untuk dimodelkan untuk mencapai hasil terbaik, dimana “jangkauan seumur hidup oleh orang-orang yang menyukai halaman”. Dengan 37,5 % dan perbedaan lebih dari 10 % berdasarkan interaksi terbaik. Dengan demikian prediksi kedua metrik berbasis visualisasi yang dianggap terbaik hanya dianggap pengguna yang menyukai halaman, sehingga kedua metrik murni berbasis visualisasi diprediksi buruk. Maka berdasarkan *eksperimen* tersebut peneliti memberikan solusi dengan memanfaatkan dataset metrik kinerja Facebook tahun 2016 untuk dapat dimanfaatkan sebagai acuan untuk mengambil keputusan atau melakukan peningkatan prediksi dalam pembangunan merek dengan menganalisa tingkat keterlibatan konsumen produk kosmetik berdasarkan jumlah dari jumlah komentar, suka dan berbagi pos. Berdasarkan uraian diatas maka peneliti menggunakan algoritma *K-Nearest Neighbour* dan *Logistic Regression (SVM)* dengan metode yang telah diadopsi dari metode *Ensemble* teknik *Bagging* yang diberi nama metode *EnBag K-LoGres*. Penelitian ini menghasilkan keputusan berdasarkan tingkat akurasi dari data yang telah diolah dan disajikan dalam bentuk kurva ROC.

B. TINJAUAN PUSTAKA

1. Data Mining

Menurut Afrisawati dalam Maulida (2018:169) menjelaskan bahwa data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. Data mining adalah proses yang menggunakan teknik *statistic*, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar

Menurut ACM dalam Suyanto (2017:1) mengungkapkan bahwa Data mining merupakan gabungan sejumlah disiplin ilmu komputer. Sedangkan menurut Clifton dalam Suyanto (2017:1) mendefinisikan data mining sebagai proses penemuan pola-pola baru dari kumpulan-kumpulan data sangat besar, meliputi metode-metode yang merupakan irisan dari *artificial intelligence*, *machine learning*, *statistics* dan *database systems*.

Menurut Suyanto (2017:3) menjelaskan bahwa secara umum kegunaan data mining dapat dibagi menjadi dua: deskriptif dan prediktif. Deskriptif berarti data mining digunakan untuk mencari pola-pola yang dapat dipahami manusia yang menjelaskan karakteristik data. Sedangkan prediktif berarti data mining digunakan untuk membentuk sebuah model pengetahuan yang akan digunakan untuk melakukan prediksi.

Menurut Fayed dkk dalam Suyanto (2017:3) berdasarkan fungsionalitasnya, tugas-tugas data mining bisa dikelompokkan kedalam enam kelompok berikut ini:

1. Klasifikasi (*classification*): men-generalisasi struktur yang diketahui untuk diaplikasikan pada data-data baru. Misalnya, klasifikasi penyakit ke dalam sejumlah jenis, klasifikasi email kedalam spam atau bukan.
2. Klasterisasi (*clustering*): mengelompokkan data, yang tidak diketahui label kelasnya, kedalam sejumlah kelompok tertentu sesuai dengan ukuran kemiripan.
3. Regresi (*regression*): menemukan suatu fungsi yang memodelkan data dengan alat (kesalahan prediksi) seminimal mungkin.
4. Deteksi anomali (*anomaly detection*): mengidentifikasi data yang tidak umum, bisa berupa *outlier* (pencilan), perubahan atau deviasi yang mungkin sangat penting dan perlu investigasi lebih lanjut.
5. Pembelajaran aturan asosiasi (*asociation rule learning*) atau pemodelan kebergantungan (*dependency modeling*): mencari relasi antar variabel.
6. Perangkuman (*summarization*): menyediakan representasi data yang lebih sederhana, meliputi visualisasi dan pembuatan laporan.

Menurut Mardhiyah & Yuli dalam Joyendri (2017:77) tahapan *data mining* ada 6 yaitu: *data cleaning*, *data integration*, *data selection*, *data transformation*, *proses mining* dan *evaluasi pola*.

2. Knowledge Discovery In Database

Menurut Ndaumanu dkk dalam Putra dan Wadisman (2018:73) menjelaskan bahwa KDD (*Knowledge Discovery In Database*) adalah keseluruhan proses *non-trivial* untuk mencari dan mengidentifikasi pola (*pattern*) dalam data, dimana pola yang ditemukan bersifat sah, baru dapat bermanfaat dan dapat dimengerti

Proses KDD menurut Nasari dan Surya dalam Putra dan Wadisman (2018:73) dapat dijelaskan sebagai berikut:

1. *Data Selection*
Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses *Data mining* disimpan dalam suatu berkas, terpisah dari basis data operasional.
2. *Pre-processing* atau *Cleaning*
Sebelum proses *Data mining* dapat dilaksanakan, perlu dilakukan proses pembersihan pada data yang menjadi fokus KDD. Proses pembersihan mencakup antara lain membuang *duplikasi data*, memeriksa data yang *inkosisten*, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (*tipografi*).
3. *Transformation*
Coding adalah transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *Data mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. Data Mining

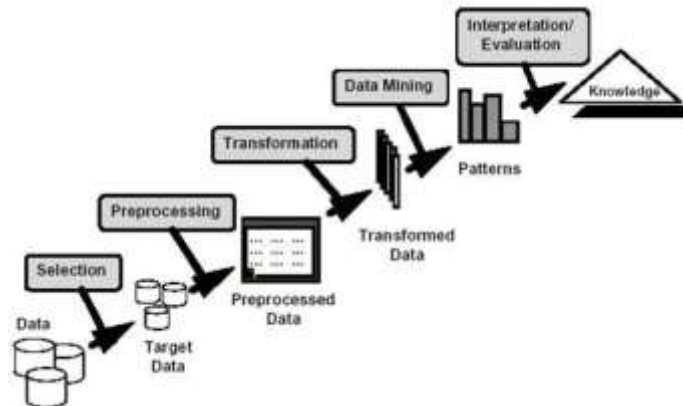
Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *Data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan. Seperti halnya Susanto dan Sudiyatno (2014) meneliti tentang penerapan data mining untuk memprediksi prestasi belajar siswa berdasarkan status sosial ekonomi orang tua, motivasi, kedisiplinan siswa dan prestasi masa lalu dengan algoritma J48.

5. Interpretation atau Evaluation

Pola informasi yang dihasilkan dari proses *Data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya

Menurut Sembiring, Muhammad Ardiansyah (2016:61) menjelaskan bahwa *Data mining* juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan.

Tahapan-tahapan pada data mining dapat digambarkan seperti gambar II.1 berikut:



Sumber: (<https://dosen.perbanas.id/>:2016)

Gambar 1. Tahapan Data Mining

3. Classification

Classification adalah metode yang paling umum pada data mining. Persoalan bisnis seperti *Churn Analysis*, dan *Risk Management* biasanya melibatkan metode *Classification*.

Classification adalah tindakan untuk memberikan kelompok pada setiap keadaan. Setiap keadaan berisi sekelompok atribut, salah satunya adalah *class attribute*. Metode ini butuh untuk menemukan sebuah model yang dapat menjelaskan class attribute itu sebagai fungsi dari *input attribute*.

Yang termasuk dalam *Classification Algorithm* adalah *Decision Trees*, *Neural Network* dan *Naives Bayes*. (<https://www.dosenpendidikan.co.id/metode-data-mining/>:2020)

4. K-Nearest Neighbor

Algoritma *k-nearest neighbor* (k-NN atau KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Ketepatan algoritma k-NN ini sangat dipengaruhi oleh ada atau tidaknya fitur-fitur yang tidak relevan, atau jika bobot fitur tersebut tidak setara dengan relevansinya terhadap klasifikasi. Riset terhadap algoritma ini sebagian besar membahas bagaimana memilih dan memberi bobot terhadap fitur, agar performa klasifikasi lebih baik. KNN juga merupakan contoh teknik *lazy learning*, yaitu teknik yang menunggu sampai pertanyaan (query) datang agar sama dengan data training. (Wu dalam Dewi, 2016)

5. Logistic Regression

Regresi logistik (*Logistic Regression*) adalah bagian dari analisis regresi yang digunakan ketika variabel dependen (respon) merupakan variabel dikotomi. Variabel dikotomi biasanya hanya terdiri atas dua nilai (Santosa dalam Dewi, 2016) yang mewakili kemunculan atau tidak adanya suatu kejadian yang biasanya diberi angka 0 atau 1. Tidak seperti regresi linier biasa, regresi logistik tidak mengasumsikan hubungan antara variabel independen dan dependen secara linier.

Ada beberapa penelitian yang menggunakan komparasi algoritma klasifikasi untuk mengukur akurasi terhadap dataset marketing bank:

1. *Could Decision trees Improve the Classification Accuracy and Interpretability of Loan Granting Decision?* penelitian yang dilakukan (Zurada,2010). Yang melakukan komparasi dari beberapa metode diantaranya adalah regresi logistik(LR), jaringan saraf (NN), dasar fungsi jaringan saraf radial (RBFNN), SVM, CBR, dan pohon keputusan (DTs). Dari semua model ternyata tingkat klasifikasi akurasi yang mengungguli adalah *Decision Trees*, DTs tidak hanya mengklasifikasikan lebih baik dari model-model yang lain tapi juga memiliki pengetahuan dalam membentuk aturan yang mudah ditafsirkan, masuk akal dalam menjelaskan tentang alasan penolakan pinjaman.
2. *Comparing decision trees with logistic regression for credit risk analysis* (Satchidananda & Simha, 2006). Penelitian ini membandingkan dua model algoritma untuk analisa resiko kredit, yaitu Pohon Keputusan dan Regresi Logistik. Data diambil dari dua bank yang berbeda, kemudian untuk mengelompokkan kasus positif dan negatif maka dilakukan klustering data dengan menggunakan k-means. Hasil analisa dari masing-masing model dikomparasi dan kemudian diukur, kemudian didapatkan bahwa algoritma pohon keputusan mempunyai tingkat akurasi yang tinggi dibandingkan algoritma regresi logistik.

C. METODE PENELITIAN

Metode yang diusulkan menggunakan metode *Classification* dengan algoritma *K-Nearest Neighbour* dan *Logistic Regression (SVM)* kemudian kedua algoritma diproses menggunakan metode *Ensemble* dengan teknik *Bagging*. Pada penelitian ini dataset yang digunakan ialah dataset sekunder dan berisi jumlah analisis sentimen pengguna Facebook pada tahun 2016 yang berisi jumlah komentar, *like*, *share* dan atribut lainnya tentang pembelian kosmetik.

Pada penelitian ini dimulai dengan membagi dataset menjadi data *training* dan data *testing* menggunakan operator *split validation*.

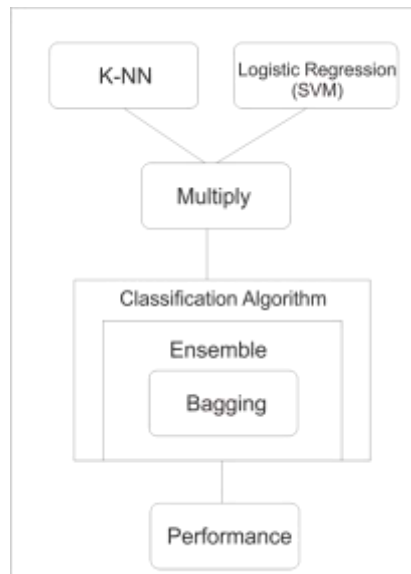
Berikut alur sistem yang dapat digambarkan setelah penelitian dilakukan.

1. Alur Proses Sistem

Metodologi yang diusulkan peneliti yaitu dengan mengukur data dengan masing-masing algoritma, namun hasilnya kurang bagus sehingga diukur kembali dengan menggabungkan kedua algoritma dengan bantuan metode *Ensemble* teknik *Bagging* hingga diperoleh hasil yang lebih bagus dibanding dengan hasil sebelumnya. Pada penelitian ini peneliti menggunakan gabungan algoritma K-NN dan *Logistic Regression (SVM)* sehingga peneliti menggunakan metode baru yang diberi nama *EnBag K-LoGres* dimana nama tersebut hasil adopsi dari kedua algoritma yang dipakai.

Pada prosesnya metodologi ini menggunakan Operator *Multiplies* yang digunakan pada saat kondisi satu data diproses oleh dua algoritma *K-NN* dan *Logistic Regression (SVM)*.

Kemudian setelah dilakukan pengukuran dengan menggunakan masing-masing algoritma dan mendapatkan hasil akurasi kurang bagus, maka dilakukan penanganan dengan metode *EnBag K-LoGres* yang bertujuan untuk meningkatkan akurasi pengklasifikasian dengan menggabungkan pengklasifikasi tunggal. Seperti yang ditampilkan pada gambar dibawah ini:



Sumber: Hasil Analisa (2020)

Gambar 2 Metode *EnBag K-LoGres*

Metode *EnBag K-LoGres* merupakan hasil gabungan dari metode *Classification* dengan algoritma *K-NN* dan *Logistic Regression (SVM)*, keduanya digabungkan menjadi algoritma baru dan diproses dengan operator *Multiply* yang berfungsi untuk membuat salinan objek pada Rapid Miner. Metode *Ensemble* untuk mencari solusi prediksi terbaik dibanding algoritma lain, serta menggunakan teknik *Bagging* yang digunakan meningkatkan stabilitas dan akurasi. Penggabungan kedua algoritma ini bertujuan untuk meningkatkan akurasi pengklasifikasian dengan menggabungkan pengklasifikasi tunggal.

D. HASIL DAN PEMBAHASAN

1. Sumber Data

Data metrik kinerja Facebook pada tahun 2016 yang terdiri dari 18 atribut diantaranya: *Page total likes*, *Category*, *Post Month*, *Post Weekday*, *Post Hour*, *Paid*, *Lifetime Post Total Reach*, *Lifetime Post Total Impressions*, *Lifetime Engaged Users*, *Lifetime Post Consumers*, *Lifetime Post Consumptions*, *Lifetime Post Impressions by people who have liked your Page*, *Lifetime Post reach by people who like your Page*, *Lifetime People who have liked your Page and engaged with your post*, *comment*, *like*, *share*, dan *Total Interactions* yang berhubungan dengan kinerja Facebook mengenai sentimen pembelian kosmetik yang diperoleh dari <https://archive.ics.uci.edu/>

2. Pengolahan Data

Pada tahap *Pre-processing* dilakukan pembersihan data dengan melakukan *replace missing value* yaitu dengan mengisi *record* yang nilainya kosong dengan bantuan aplikasi Rapid Miner kemudian dimasukkan secara manual pada *record* yang kosong di Ms Excel. Kemudian diproses dengan metode dan algoritma yang telah diusulkan

3. Proses

Pada penelitian ini sebelumnya dataset diukur dengan algoritma *K-Nearest Neighbour* dan *Logistic Regression (SVM)* kemudian selanjutnya kedua algoritma dilakukan pengukuran dengan metode *Ensemble* dengan teknik *Bagging* secara bersamaan.

Berdasarkan penerapan model yang diusulkan diperoleh pengukuran seperti ditunjukkan pada Tabel 4.1 berikut:

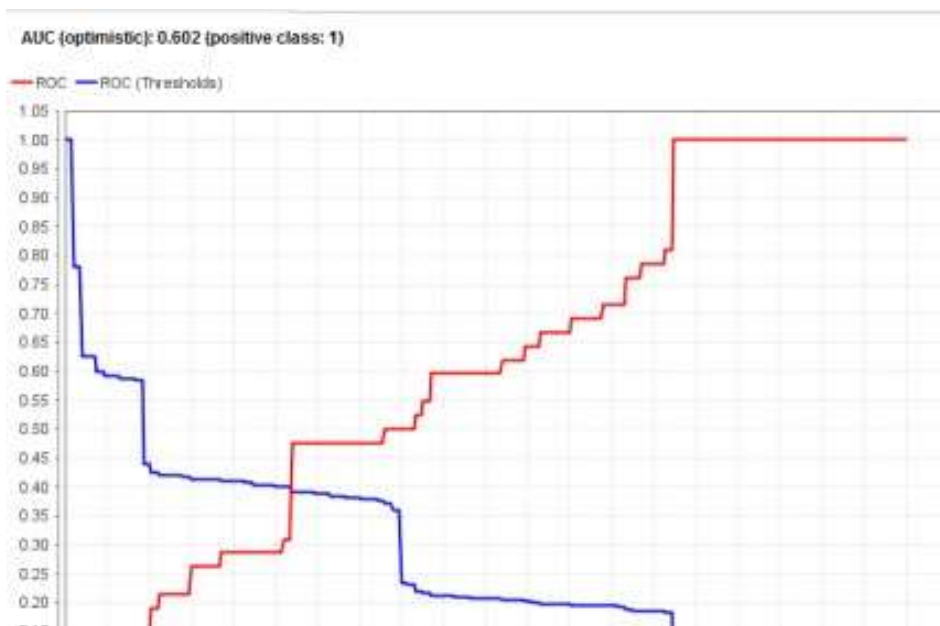
Tabel 1 Hasil Pengukuran Kinerja Model

Model	Accuracy	AUC	Recall	Pression
K-NN	68.67%	0.602	11.90%	33.33%
Logistic Regression (SVM)	72.67%	0.570	2.38%	100.00%
EnBag K-Logres	73.91%	0.584	4.00%	100.00%

Sumber: Hasil Analisa (2020)

a. Hasil AUC Pengukuran Kinerja Model

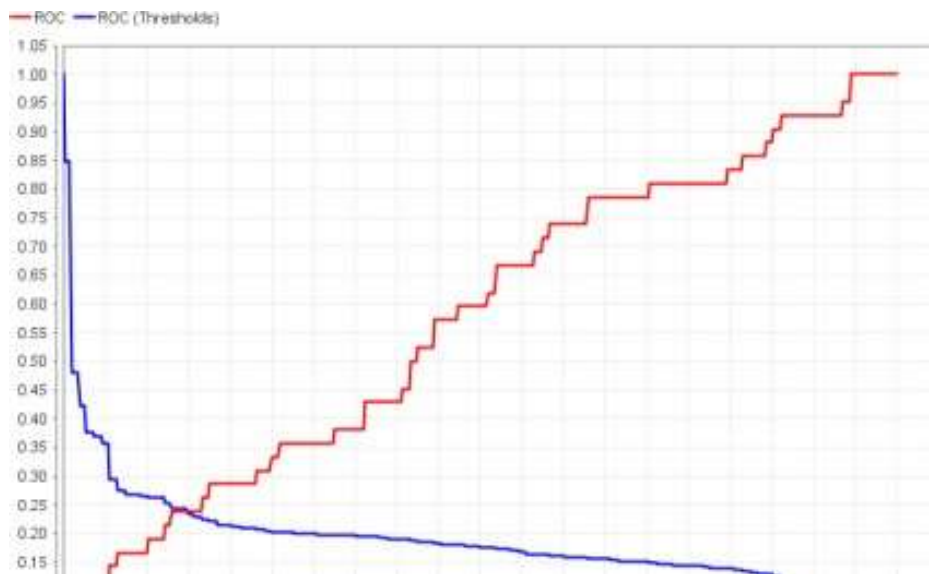
Hasil pengukuran dari masing-masing algoritma beserta hasil penggabungan dari kedua algoritma yang diproses dengan menggunakan metode *EnBag K-LoGres* disajikan dalam bentuk kurva Receiver Operating Characteristic (ROC) dibawah ini:



Sumber: Hasil Analisa (2020)

Gambar 3 Hasil Pengukuran AUC Algoritma K-NN

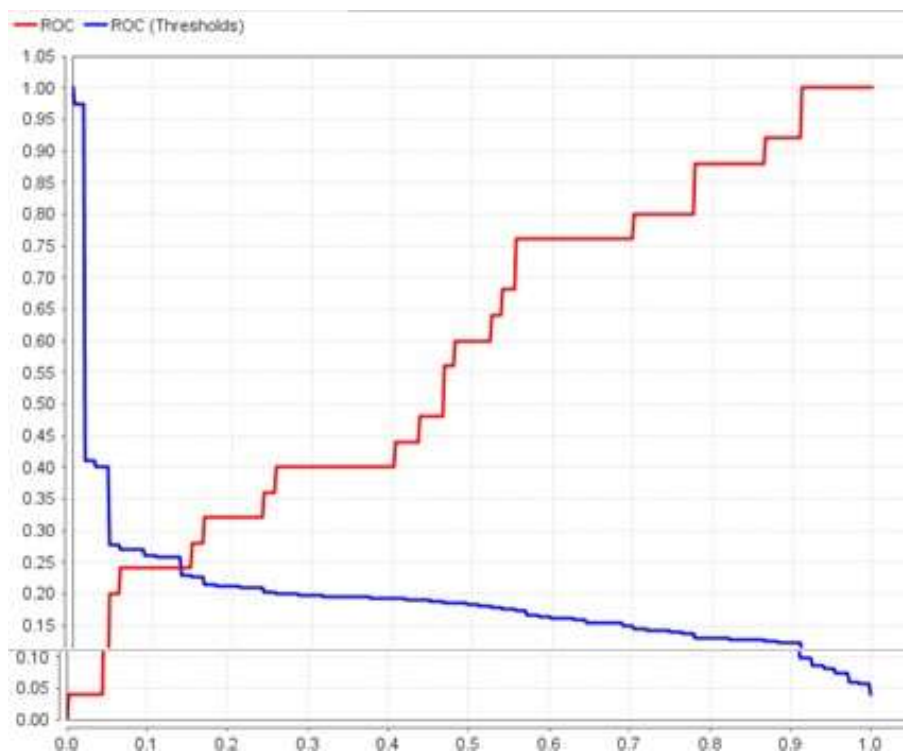
Pada kurva ROC hasil pengukuran algoritma K-NN menunjukkan bahwa jarak antara *True Positif Rate* dan *False Positive Rate* berdekatan sehigga Tresholnya kurang bagus.



Sumber: Hasil Analisa (2020)

Gambar 4 Hasil Pengukuran AUC dengan Algoritma *Logistic Regression (SVM)*

Pada kurva ROC hasil pengukuran algoritma *Logistic Regression (SVM)* menunjukkan bahwa jarak antara *True Positif Rate* dan *False Positive Rate* lebih jauh dibanding dengan algoritma K-NN sehingga tresholdnya lebih bagus dari pada Treshold K-NN .



Sumber: Hasil Analisa (2020)

Gambar 5 Hasil Pengukuran AUC dengan Metode *EnBag K-LoGres*

Pada kurva ROC hasil pengukuran dengan metode *EnBag K-LoGres* menunjukkan bahwa jarak antara *True Positif Rate* dan *False Positive Rate* lebih lebar dari algoritma *Logistic*

Regression (SVM) sehingga tresholdnya lebih bagus dibanding kedua algoritma sebelumnya. Berdasarkan keterangan kurva ROC dapat dijelaskan sebagai berikut, dimana Garis warna biru menunjukkan ROC (*True Positive Rate*) dan Garis warna merah menunjukkan ROC (*False Positive Rate*)

Maka semakin tinggi *True Positive Rate* maka semakin kecil *False Positive Rate* maka Tresholdnya semakin bagus.

E. KESIMPULAN

Berdasarkan hasil pengukuran dari 2 algoritma yang diusulkan tersebut maka hasilnya mengalami kenaikan sebesar 4.00 % dan setelah diproses dengan teknik *Bagging* kedua algoritma tersebut yaitu *K-NN* dan *Logistic Regression (SVM)*, maka hasilnya mengalami kenaikan sebesar 1.24 %. Sehingga berdasarkan hasil pengukuran yang telah diketahui, hasil penerapan metode *Ensemble* dengan teknik *Bagging* pada dataset Metrik Facebook pengguna kosmetik dapat ditarik kesimpulan bahwa metode *Ensemble* dengan metode *EnBag K-LoGres* dapat meningkatkan nilai *Accuracy* dan *Pression* yang dihasilkan.

F. REFERENSI

- Anbe, S., & Kobayashi, I. (2014). An approach to category classification of cosmetics reviews based on brand names. *2014 Joint 7th International Conference on Soft Computing and Intelligent Systems, SCIS 2014 and 15th International Symposium on Advanced Intelligent Systems, ISIS 2014*, 971–975. <https://doi.org/10.1109/SCIS-ISIS.2014.7044788>
- Arfi Joyendri. 2017. Strategi Customer Relationship Management Untuk Meningkatkan Loyalitas Pelanggan Dan Volume Penjualan Menggunakan Teknik Clustering K-Means. *Telemat. - J. Inform. Telekomun. Komputasi Elektron. dan Ind.* 14: 75–82.
- BRILIANI, RIZKA ASRI *et al.* (2016). Analisis Kecenderungan Pemilihan Kosmetik Wanita Di Kalangan Mahasiswi Jurusan Statistika Universitas Diponegoro Menggunakan Biplot Komponen Utama. *None*, 5(3), 545–551.
- Dewi S. 2016. Pada Prediksi Keberhasilan Pemasaran Produk Layanan Perbankan. *Techno Nusa Mandiri XIII*: 60–66.
- Dosen Pendidikan 3.2020. Metode Data Mining. <https://www.dosenpendidikan.co.id/metode-data-mining/> (Diakses pada 05 Mei 2020)
- Hapsari, Y., Fikri Hidayattullah, M., Dairoh, D., & Khambali, M. (2018). Opinion Mining Terhadap Toko Online Di Media Sosial Menggunakan Algoritma Naïve Bayes (Studi Kasus: Akun Facebook Dugal Delivery). *Jurnal Informatika: Jurnal Pengembangan IT*, 3(2), 233–236. <https://doi.org/10.30591/jpit.v3i2.919>
- Hermawanti, L., & Safriandono, A. (2016). Penggabungan Algoritma Forward Selection Dan K-Nearest Neighbor Untuk Mendiagnosis Penyakit Diabetes Di Kota Semarang. *Jurnal Momentum UNWAHAS*, 12(2), 138462. <https://doi.org/10.36499/jim.v12i2.1631>
- Indonesia Internet Service Provider Association. 2016. Edisi 05 November 2016. Jakarta: APJII
- Kristiyanti, D. A., & Wahyudi, M. (2017). Feature selection based on Genetic algorithm, particle swarm optimization and principal component analysis for opinion mining cosmetic product review. *2017 5th International Conference on Cyber and IT Service Management, CITSM 2017*. <https://doi.org/10.1109/CITSM.2017.8089278>
- Luqyana, W. A., Cholissodin, I., & Perdana, R. S. (2018). Analisis Sentimen Cyberbullying Pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 2(11), 4704–4713.
- Maulida L. 2018. Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov . Dki Jakarta Dengan K-Means. *J. Inform. Sunan Kalijaga* 2: 167–174.
- Moro S, Rita P, Vala B. 2016. Predicting social media performance metrics and evaluation of the impact on brand building: A data mining approach. *J. Bus. Res.* 69: 3341–3351.
- Muhammad Ardiansyah Sembiring. 2016. PENERAPAN METODE DECISION TREE ALGORITMA C45 UNTUK MEMPREDIKSI HASIL BELAJAR MAHASISWA BERDASARKAN RIWAYAT

- AKADEMIK. JURTEKSI (JURNAL Teknol. DAN Sist. INFORMASI).
- Oktanisa, I., & Supianto, A. A. (2018). Perbandingan Teknik Klasifikasi Dalam Data Mining Untuk Bank a Comparison of Classification Techniques in Data Mining for. *Teknologi Informasi Dan Ilmu Komputer*, 5(5), 567–576. <https://doi.org/10.25126/jtiik20185958>
- Perbanas Institut. 2016. Sarana Tukar Menukar Informasi dan Pemikiran Dosen. <https://dosen.perbanas.id/apa-itu-data-mining/> (Diakses tanggal 03 Juni 2020).
- Randi Rian Putra CW. 2018. IMPLEMENTASI DATA MINING PEMILIHAN PELANGGAN POTENSIAL MENGGUNAKAN ALGORITMA K-MEANS IMPLEMENTATION OF DATA MINING FOR POTENTIAL CUSTOMER SELECTION USING K-MEANS ALGORITHM. *J. Inf. Technol. Comput. Sci.* 1: 72–77.
- Satvika, G. A. J., Nasution, S. M., & Nugrahaeni, R. A. (2018). Determination of the Best Vehicle Pathway with Classification of Data Mining Twitter using K-Nearest Neighbor. *2018 International Conference on Information Technology Systems and Innovation, ICITSI 2018 - Proceedings*, 72–76. <https://doi.org/10.1109/ICITSI.2018.8695947>
- Suyanto. (2017). Data Mining Untuk Klasifikasi dan Klasterisasi Data. Bandung: Informatika
- UC Irvine Machine Learning Repository. 2016. *Facebook Metrics*, [online] Tersedia di: <https://archive.ics.uci.edu/> [Diakses 2 Oktober 2019]