

## KOMBINASI K-NN DAN *GRADIENT BOOSTED TREES* UNTUK KLASIFIKASI PENERIMA PROGRAM BANTUAN SOSIAL

Elly Firasari<sup>1</sup>, Umi Khultsum<sup>2</sup>, Monikka Nur Winnarto<sup>3</sup>, Risnandar<sup>4</sup>

<sup>1,2,3</sup>STMIK Nusa Mandiri Kramat 18, <sup>4</sup>Pusat Penelitian Informatika-LIPI  
Email: <sup>1</sup>14002379@nusamandiri.ac.id, <sup>2</sup>umu.khultsum@gmail.com, <sup>3</sup>monikkanw@gmail.com,  
<sup>4</sup>risnandar01@gmail.com  
\*Penulis Korespondensi

(Naskah masuk: 11 Januari 2020, diterima untuk diterbitkan: 26 November 2020)

### Abstrak

Kemiskinan bagi pemerintah Indonesia termasuk masalah yang sulit untuk diselesaikan. Upaya yang dilakukan pemerintah dalam mengatasi kemiskinan di Indonesia yaitudengan program bantuan sosial meliputiBLT (Bantuan Langsung Tunai), PKH (Program Keluarga Harapan), Raskin (Beras Miskin), dan lain lain. Dalam Pelaksanaan program bantuan sosial saat masih sangat terbatas sehingga dalam penerimaan program bantuan tidak tepat sasaran. Data mining membantu untuk menentukan keputusan dalam memprediksi data di masa yang akan datang. *Gradient Boosted Trees* dan K-NN merupakan salah satu metode data mining untuk klasifikasi data. Masing-masing metode tersebut memiliki kelemahan. *Gradient Boosted Trees* menghasilkan nilai persentase akurasi lebih rendah dibanding metode K-NN. Dari permasalahan tersebut maka diusulkan metode kombinasi K-NN dan *Gradient Boosted Trees* untuk meningkatkan akurasi pada pelaksanaan program bantuan sosial agar tepat sasaran. Metode K-NN, *Gradient Boosted Trees*, K-NN-*Gradient Boosted Trees*dilakukan pengujian pada data yang sama untuk mendapatkan hasil perbandingan nilai akurasi. Hasil pengujian membuktikan bahwa kombinasi tersebut menghasilkan nilai persentase yang tinggi dibanding metode K-NN atau *Gradient Boosted Trees* yaitu 98.17%.

**Kata Kunci:** K-NN, *Gradient Boosted Trees*, Metode Kombinasi

## COMBINATIONS OF K-NN AND GRADIENT BOOSTED TREES FOR CLASSIFICATION OF RECIPIENTS IN SOCIAL ASSISTANCE PROGRAMS

### Abstract

Poverty for the Indonesian government is a problem that is difficult to solve. The efforts made by the government in overcoming poverty in Indonesia are through social assistance programs including BLT (Bantuan Langsung Tunai), PKH (Program Keluarga Harapan), Raskin (Beras Miskin), and others. In the implementation of the social assistance program when it was still very limited, the acceptance of the aid program was not on target. Data mining helps to determine decisions in predicting data in the future. *Gradient Boosted Trees* and K-NN are data mining methods for data classification. Each of these methods has weaknesses. *Gradient Boosted Trees* produce lower accuracy percentage values than the K-NN method. From these problems, a proposed method of combination of K-NN and *Gradient Boosted Trees* is used to improve the accuracy of the implementation of social assistance programs so that it is right on target. The K-NN, *Gradient Boosted Trees*, and K-NN-*Gradient Boosted Trees* methods are tested on the same data to get a comparison of the accuracy values. The test results prove that the combination produced a high percentage value compared to the K-NN or *Gradient Boosted Trees* method that is 98.17%.

**Keywords:** K-NN, *Gradient Boosted Trees*, Combination Method

### 1. Pendahuluan

Menurut Badan Pusat Statistik, pada bulan September 2018 angka kemiskinan di Indonesia sebesar 9,66% dimana presentase ini lebih kecil dari

bulan Maret 2018 yang menunjukkan angka sebesar 9,82% dan September 2017 yang mencapai 10,12%. Meskipun data menunjukkan penurunan tingkat kemiskinan di Indonesia tetapi hal tersebut belum memenuhi target yang tertuang dalam RPJM yaitu 7-

8% pada akhir tahun 2019, sedangkan realita di lapangan tingkat kemiskinan di Indonesia tahun 2018 masih di angka 9,66%. Kemiskinan bagi pemerintah Indonesia termasuk masalah yang sulit untuk diselesaikan karena kurangnya pemahaman berbagai pihak tentang penyebab kemiskinan itu sendiri, sehingga program penanggulangan kemiskinan tidak didasarkan pada penyebabnya yang berbeda-beda secara lokal.

Upaya yang dilakukan pemerintah dalam mengatasi kemiskinan di Indonesia yaitudengan program bantuan sosial meliputiBLT (Bantuan Langsung Tunai), PKH (Program Keluarga Harapan), Raskin (Beras Miskin), dan sebagainya. Berdasarkan Data dari Bappenas 2014, masalah kemiskinan saat ini disebabkan oleh beberapa faktor, antara lain: ketidaktepatan sasaran dalam penentuan penerimaan program, mekanisme pendampingan program belum optimal, koordinasi dan pelaksanaan program belum terintegrasi dan prioritas pendanaan untuk program perlindungan sosial yang masih terbatas .

Untuk menentukan klasifikasi tingkat penduduk miskin terdapat banyak metode yang dapat digunakan. Salah satunya yaitu K-NN dan *Gradient Boosted Trees*. Berdasarkan penelitian sebelumnya yang dilakukan oleh Yunus, dkk (2019) tentang Data Mining untuk Memprediksi Hasil Produksi Buah Sawit pada PT Bumi Sawit Sukses menggunakan Metode *K-Nearest Neighbor* (K-NN) menghasilkan *output* dari *Rapidminer* dengan akurasi 85,15% (Yunus, Akbar, & Andri, 2019).

Menurut penelitian sebelumnya yang dilakukan oleh saikin dan kusrini mengenai Karakteristik Data *Traveller* didapatkan hasil kualifikasi metode K-NN dengan pengujian *confusion matrix*didapatkan nilai akurasi sebesar 84% (Saikin & Kusrini, 2019).

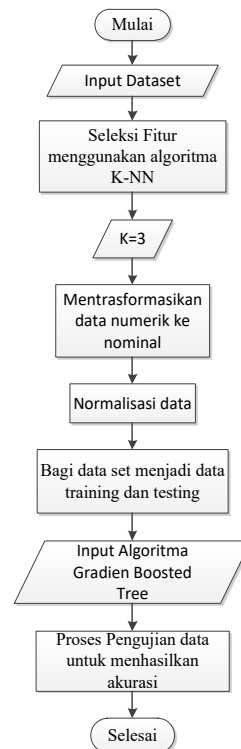
Sedangkan penelitian yang dilakukan oleh Pertiwi, dkk mengenai Komparasi 5 Metode Data Mining yaitu menggunakan metode C4.5, KNN, *Naïve Bayes*, Rule Induction, dan *Gradient Boosted Tree*. Dari hasil perhitungan diperoleh akurasi C4.5 87,50%, K-NN 67,50%, *Naïve Bayes* 25,83%, RI 79,17% dan *Gradient Boosted Tree* 25,83% (Pertiwi, Adiwisatra, & Supriadi, 2019).

Berdasarkan penelitian terdahulu dari masing-masing metode yang digunakan menghasilkan akurasi yang rendah. Kombinasi antara K-NN dan *Gradient Boosted Trees* diperlukan untuk memperoleh nilai akurasi yang lebih tinggi dan telah dibuktikan bahwa hasil akurasi yang diperoleh meningkat.

## 2. METODE PENELITIAN

Penelitian ini menggunakan penggabungan dua metode yaitu metode K-NN dan *Gradient Boosted Tree*. Pada Seleksi Fitur digunakan untuk menyeleksi data yang rusak/tidak lengkap

menggunakan fitur "*Input Missing Value*" dan "*Rename Unused Value*" dengan menggunakan metode K-NN dengan jarak  $k=3$  sehingga didapatkan data set murni. Mentranformasikan data dari numerik ke nominal dan lakukan normaisasi menentukan bentuk data yang paling tepat. Selanjutnya membagi data *training* dan data *testing* dengan perbandingan 80%:20% sehingga dihasilkan data taining 288 dan data testing 73. Pemrosesan data dilakukan dengan menggabungkan metode *Gradient Boosted Tree* dalam proses validasi sehingga diperoleh akurasi yang lebih tinggi dari pengolahan masing-masing metode. Penelitian ini dilakukan untuk pengembangan evaluasi dan pemecahan masalah yaitu meningkatkan akurasi pada pelaksanaan program bantuan sosial agar tepat sasaran. Berikut ini adalah tahapan penelitian yang dilakukan:



Gambar 1. Metode Penelitian

## 3. PENELITIAN TERDAHULU

Kombinasi metode dilakukan pada beberapa penelitian terdahulu menggunakan berbagai algoritma yang berbeda-beda. Kombinasi metode dilakukan karena perhitungan dengan satu metode masih menghasilkan akurasi yang rendah. Berikut merupakan beberapa penelitian terdahulu mengenai perhitungan dengan satu metode dan kombinasi metode:

Prediksi Hasil Produksi Buah Sawit dengan metode K-NN berdasarkan penelitian dari Yunus,dkk (2019) dalam perhitungannya menghasilkan akurasi 85,15% (Yunus et al., 2019).

Komparasi metode C4.5, KNN, *Naïve Bayes*, Rule Induction, dan *Gradient Boosted Tree* dari penelitian yang dilakukan oleh Pertiwi, dkk (2019) mengenai Persentase Wanita Sudah Menikah di Usia 15-49 yang Memakai Alat KB menghasilkan akurasi C4.5 87,50%, K-NN 67,50%, *Naïve Bayes* 25,83%, RI 79,17% dan *Gradient Boosted Tree* 25,83% (Pertiwi et al., 2019).

Kombinasi metode *Naïve Bayes* dan K-NN untuk Klasifikasi Data dari penelitian sebelumnya yang dilakukan oleh Sari, dkk (2015) dalam perhitungannya dapat mengatasi kelemahan metode *naïve Bayes* dan K-NN. Persentase keakuratan menggunakan metode K-NN-*Naïve Bayes* memperoleh hasil yang lebih tinggi dibandingkan pengolahan data dengan masing-masing metode (Sari, Ernawati, & Pranowo, 2015).

Kombinasi metode Logistic Regresion dan *Gradient Boosted Tree* telah dilakukan pada penelitian terdahulu oleh Arfah Anggraina dkk (2019). Kombinasi metode ini digunakan untuk mendeteksi *spam e-mail*. Hasil akurasi dari kombinasi metode ini terbukti lebih tinggi daripada hasil akurasi dari perhitungan masing-masing metode (Anggraina, Primartha, & Wijaya, 2019).

Analisis Komparatif *K-Nearest Neighbor* dan Algoritma *Neighbor K-Nearest* yang Dimodifikasi untuk Klasifikasi Data dari penelitian sebelumnya yang dilakukan oleh Okfalisa, dkk menghasilkan Akurasi tertinggi pada  $k=1$  yaitu 94,95% sedangkan pada metode MKNN menghasilkan akurasi  $k=1$  sebesar 99,51% (Okfalisa, Gazalba, Mustakim, & Reza, 2018)

Penambahan opini Roman Urdu dengan menggunakan klasifikasi K-NN yang telah dilakukan oleh Muhammad Bilal, dkk pada tahun 2015 menghasilkan akurasi 95,00. (Bilal, Israr, Shahid, & Khan, 2015)

Prediksi Kinerja Siswa Menggunakan K-NN, penelitian yang dilakukan Ihsan A. Abu Amra dan Ashraf YA Maghariv menghasilkan akurasi 63,4%. (Amra & Ashraf, 2017)

Penelitian yang dilakukan oleh Xiaohui Li, Siboyang, Rongwei Fan, Xin Yu dan Deying Chen tentang Diskriminasi Jaringan lunak menggunakan kerusakan akibat laser spektroskopi dalam kombinasi dengan  $k$  tetangga terdekat (K-NN) menghasilkan akurasi sebesar 74,75 dengan  $k=3$ . (Li, Yang, Fan, Yu, & Chen, 2018)

Pendekatan lokalisasi pupil mata yang akurat berdasarkan Gradient Boosted tree tahun 2016 Penelitian yang dilakukan oleh Dong Tian, Guanghui He, Jiangxia Wu, Hongtai Chen, dan Yong Jiang menghasilkan akurasi 92,39% (Tian, He, Wu, Chen, & Jiang, 2016)

Identifikasi Sumber Gangguan Kualitas daya Menggunakan *Gradient Boosted Decision Tree* penelitian yang dilakukan oleh Cheng Pan, Jie Tan dan Dandan Feng tahun 2018 menghasilkan akurasi yang cukup tinggi dibanding klasifikasi dengan

metode SVM dan KNN sebesar 98,68% (Pan, Tan, & Feng, 2018)

Kasus mengenai metode KNN telah dilakukan oleh Devika R, Sai Vaishnavi Avilala, dan V. Subramaniaswamy pada tahun 2019 (Devika, Avilala, & Subramaniaswamy, 2019), bahkan pada tahun 2018 untuk penelitian menggunakan metode Gradient Boosted Tree juga pernah dilakukan oleh Peng Sheng, Li Chen, Jing Tian (Sheng, Chen, & Tian, 2018). Dari kedua penelitian tersebut memiliki kelemahan pada bagian hasil yaitu tingkat akurasi yang masih rendah. Penelitian yang dilakukan oleh Devika R, Sai Vaishnavi Avilala, dan V. Subramaniaswamy pada metode KNN menghasilkan akurasi 87.78%, sedangkan penelitian pada metode *Gradient Boosted Tree* yang dilakukan oleh Peng Sheng, Li Chen, Jing Tian menghasilkan akurasi 86.42%. Sehingga kami melakukan perbaikan dan usulan terbaru dengan mengkombinasikan algoritma KNN dan Gradient Boosted Tree untuk menghasilkan tingkat akurasi yang tinggi.

#### 4. HASIL DAN PEMBAHASAN

Pada penelitian ini menggunakan data penduduk desa yang diperoleh dari kantor kelurahan, sebagai berikut:

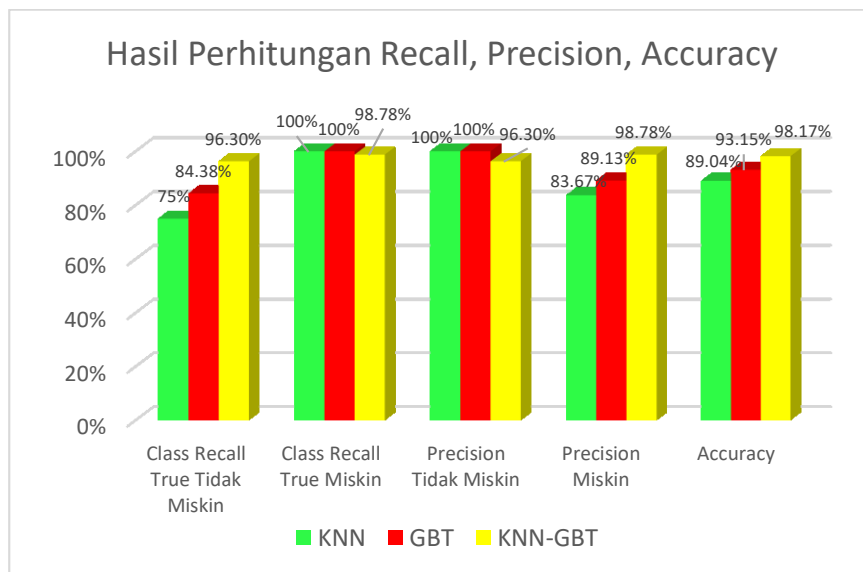
- Pendidikan: Tidak SD, SD, SLTP, SLTA, S1
- Luas Lantai: Kurang dari 13m<sup>2</sup>, Sama 13m<sup>2</sup>, Lebih dari 13 m<sup>2</sup>
- Jenis Lantai: Keramik, Semen, Tanah
- Jenis Dinding: Bambu, Tembok
- Jenis Penerangan: Non PLN, PLN
- Bahan Bakar Masak: Kayu, Gas
- Sumber Air Minum : Sumur, Ledeng
- Jenis Jamban: Umum, Bersama, Sendiri
- Pendapatan: 0 sampai 1,5 juta, 1,5 sampai 3 juta, lebih dari 3 juta
- Pekerjaan: Tidak Bekerja/Buruh, Petani, Pedagang, Wiraswasta, Pegawai
- Kemampuan Membel Obat : Ya, Tidak
- Pakaian: Ya, Tidak
- Konsumsi: Ya, Tidak

Hasil yang diperoleh dari perhitungan masing-masing metode adalah sebagai berikut:

Tabel 1. Perhitungan dengan Metode KNN

Accuracy: 89.04%			
	True Tidak Miskin	True Miskin	Class Precision
Pred. Tidak Miskin	41	8	83.67%
Pred. Miskin	0	24	100%
Class Recall	100%	75%	

Perhitungan dengan metode KNN diperoleh hasil akurasi sebesar 89.04%,



Gambar 2. Hasil Perhitungan K-NN

Tabel 2. Perhitungan Metode *Gradient Boosted Trees*

Accuracy: 93.15%

	True Tidak Miskin	True Miskin	Class Precision
Pred. Tidak Miskin	41	5	89.13%
Pred. Miskin	0	27	100%
Class Recall	100%	84.38%	

Perhitungan dengan metode *Gradient Boosted Trees* diperoleh hasil akurasi sebesar 93.15%,

Tabel 3. Perhitungan Kombinasi Metode KNN dan *Gradient Boosted Trees*

Accuracy: 98.17%

	True Tidak Miskin	True Miskin	Class Precision
Pred. Tidak Miskin	81	1	98.78%
Pred. Miskin	1	26	96.30%
Class Recall	98.78%	96.30%	

Perhitungan dengan kombinasi metode KNN dan *Gradient Boosted Trees* menghasilkan akurasi sebesar 98.17%

Perbandingan persentase keakuratan dari metode K-NN dan *Gradient Boosted Trees* dan metode kombinasi K-NN dan *Boosted Trees* adalah sebagai berikut:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \dots \dots \dots (1)$$

$$\text{KNN} = \frac{41 + 24}{41 + 24 + 0 + 8} \times 100$$

$$\frac{65}{73} \times 100 = 89.04 \%$$

$$\text{GBT} = \frac{41 + 27}{41 + 27 + 0 + 5} \times 100$$

$$\frac{68}{73} \times 100 = 93.15 \%$$

$$\text{KNN} + \text{GBT} = \frac{81 + 26}{81 + 26 + 1 + 1} \times 100$$

$$\frac{107}{109} \times 100 = 98.17 \%$$

Tabel 4. Perbandingan Akurasi

No	Metode	Akurasi
1.	K-NN	89.04%
2.	<i>Gradient Boosted Trees</i>	93.15%
3.	Kombinasi K-NN – <i>Gradient Boosted Trees</i>	98.17%

## 5. KESIMPULAN

Berdasarkan dari hasil penelitian yang telah dilakukan dapat disimpulkan bahwa:

1. Penelitian Kombinasi metode K-NN dan *Gradient Boosted Trees* dapat memperoleh hasil persentase akurasi yang lebih tinggi. Hasil perhitungan kombinasi K-NN dan *Gradient Boosted Trees* menunjukkan persentase akurasi yang lebih tinggi dibandingkan dengan perhitungan masing-masing metode. Kombinasi ini menghasilkan akurasi sebesar 98.17%.
2. Akurasi 98.17% menunjukkan bahwa kombinasi metode ini tepat digunakan dalam penentuan penerima bantuan sosial secara tepat sasaran.

## DAFTAR PUSTAKA

AMRA, IHSAN A ABU & ASHRAF Y.A. MAGHARI. 2017. Student Performance Prediction Using KNN and Naïve Bayesian.

- 2017 7th International Conference on Information Technology (ICIT), 909–913.
- ANGGRAINA, A., PRIMARTHA, R., & WIJAYA, A. 2019. *The Combination of Logistic Regression and Gradient Boost Tree for Email Spam Detection The Combination of Logistic Regression and Gradient Boost Tree for Email Spam Detection*. 1–6. <https://doi.org/10.1088/1742-6596/1196/1/012013>
- BILAL, M., ISRAR, H., SHAHID, M., & KHAN, A. 2015. Sentiment classification of Roman-Urdu opinions using Navie Baysian, Decision Tree and KNN classification techniques. *Journal Of King Saud University-Computer And Information Sciences*. <https://doi.org/10.1016/j.jksuci.2015.11.003>
- BPS. 2019. Jumlah Dan Tingkat Penduduk Miskin Indonesia 1970 – Sep 2018
- DEVIKA, R., AVILALA, S. V., & Subramaniaswamy, V. 2019. Comparative Study of Classifier for Chronic Kidney Disease prediction using Naive Bayes , KNN and Random Forest. *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, (Iccmc), 679–684.
- LI, X., YANG, S., FAN, R., YU, X., & CHEN, D. 2018. Discrimination of soft tissues using laser-induced breakdown spectroscopy in combination with k nearest neighbors (kNN) and support vector machine (SVM) classifiers. *Optics and Laser Technology*, 102, 233–239. <https://doi.org/10.1016/j.optlastec.2018.01.028>
- OKFALISA, GAZALBA, I., MUSTAKIM, & REZA, N. G. I. 2018. Comparative analysis of k-nearest neighbor and modified k-nearest neighbor algorithm for data classification. *Proceedings-2017 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2017, 2018-Janua*, 294–298. <https://doi.org/10.1109/ICITISEE.2017.8285514>
- PAN, C., TAN, J., & FENG, D. (2018). Identification of Power Quality Disturbance Sources Using Gradient Boosting Decision Tree. *2018 Chinese Automation Congress (CAC)*, 2589–2592. <https://doi.org/10.1109/CAC.2018.8623162>
- PERTIWI, M. W., ADIWISASTRA, M. F., & SUPRIADI, D. 2019. Analisa Komparasi Menggunakan 5 Metode Data Mining dalam Klasifikasi Persentase Wanita Sudah menikah di Usia 15-49 yang Memakai Alat KB (Keluarga Berencana). *Jurnal Khatulistiwa Informatika*, VII(1), 37–42.
- SAIKIN, & KUSRINI. 2019. Model Data Mining Untuk Karakteristik Data Traveller Pada Perusahaan Tour And Travel (Studi Kasus : Lombok Ceria Holiday). *Jurnal Manajemen Informatika & Sistem Informasi*, 2(2), 61–68.
- SARI, M. K., ERNAWATI, & PRANOWO. 2015. Kombinasi Metode K-Nearest Neighbor Dan Naïve Bayes Untuk Klasifikasi Data. *Seminar Nasional Teknologi Informasi Dan Multimedia 2015*, 6–8.
- SHENG, P., CHEN, L., & TIAN, J. 2018. Learning-based Road Crack Detection Using Gradient Boost Decision Tree. *2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, 1228–1232.
- TIAN, D., HE, G., WU, J., CHEN, H., & JIANG, Y. 2016. *An Accurate Eye Pupil Localization Approach Based on Adaptive Gradient Boosting Decision Tree*. 27–30. <https://doi.org/978-1-5090-5316-2/2016>
- YUNUS, A., AKBAR, M., & ANDRI. 2019. Data Mining Untuk Memeprediksi Hasil Produksi Nuah Sawit Pada Pt Bumi Sawit Saukses (Bss) Menggunakan Metode K-Nearest Neighbor. *Bina Darma Conference on Computer Science*, 198–207.

*Halaman ini sengaja dikosongkan*