ARTIKEL

ANALISA PERBANDINGAN ALGORITMA METODE KLASIFIKASI TERHADAP NASABAH YANG BERPOTENSI MEMBUKA DEPOSITO

Oleh:

YOAN OCTAVIASARI DWI PUTRI LESTARI

14.1.03.03.0085

Dibimbing oleh :

1. **Sucipto, M.Kom**
2. **Teguh Andriyanto, S.T,.M.Cs**

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS TEKNIK

UNIVERSITAS NUSANTARA PGRI KEDIRI

2018

Artikel Skripsi
Universitas Nusantara PGRI Kediri

SURAT PERNYATAAN
ARTIKEL SKRIPSI TAHUN 2018

Yang bertanda tangan di bawah ini:

Nama Lengkap	: Yoan Octaviasari Dwi Putri Lestari
NPM	: 14.1.03.03.0085
Telepun/HP	: 085649028836
Alamat Surel (Email)	: yoanoctaviasaridpl@gmail.com
Judul Artikel	: Analisa Perbandingan Algoritma Metode Klasifikasi Terhadap Nasabah yang Berpotensi Membuka Deposito
Fakultas – Program Studi	: Fakultas Teknik – Sistem Informasi
Nama Perguruan Tinggi	: Universitas Nusantara PGRI Kediri
Alamat Perguruan Tinggi	: Jalan KH Ahmad Dahlan No.76 Mojoroto Kediri Jawa Timur 64112

Dengan ini menyatakan bahwa :

a. artikel yang saya tulis merupakan karya saya pribadi (bersama tim penulis) dan bebas plagiarisme;
b. artikel telah diteliti dan disetujui untuk diterbitkan oleh Dosen Pembimbing I dan II.

Demikian surat pernyataan ini saya buat dengan sesungguhnya. Apabila di kemudian hari ditemukan ketidaksesuaian data dengan pernyataan ini dan atau ada tuntutan dari pihak lain, saya bersedia bertanggungjawab dan diproses sesuai dengan ketentuan yang berlaku.

Mengetahui		Kediri, 18 Juli 2018
Pembimbing I	Pembimbing II	Penulis,
Sucipto, M.Kom	Teguh Andriyanto,S.T,M.Cs	Yoan Octaviasari D.P.L
NIDN.0721029101	NIDN.0701117802	NPM.14.1.03.03.0085

Yoan Octaviasari | 14.1.03.03.0085
Fakultas Teknik – Sistem Informasi
simki.unpkediri.ac.id
|| 1||

ANALISA PERBANDINGAN ALGORITMA METODE KLASIFIKASI TERHADAP NASABAH YANG BERPOTENSI MEMBUKA DEPOSITO

Yoan Octaviasari Dwi Putri Lestari
14.1.03.03.0085
Fakultas Teknik – Sistem Informasi
yoanoctaviasaridpl@gmail.com
Sucipto, M.Kom dan Teguh Andriyanto, S.T, M.Cs
UNIVERSITAS NUSANTARA PGRI KEDIRI

ABSTRAK

Banyaknya algoritma klasifikasi *data mining* yang masing-masing mempunyai tingkat ketepatan akurasi yang berbeda-beda antara satu sama lain. Akurasi digunakan untuk mengukur ketepatan dan juga kemiripan pada hasil dengan waktu yang sama dan membandingkan nilai absolut yang jika semakin mendekati ukuran maka akan semakin tinggi juga tingkat akurasinya. Banyaknya data juga mempengaruhi tingkat akurasi tersebut.

Sehingga peneliti mengidentifikasikan tingkat akurasi beberapa algoritma klasifikasi untuk mengetahui peforma akurasinya masing-masing dan membandingkan akurasi 3 algoritma populer *decision tree, naive bayes*, dan *nearest neighbour*.

Menggunakan dataset Bank Marketing yang dikeluarkan oleh UCI *Machine Learning Repository*. Penelitian dilakukan dengan menggunakan metode kausal komperatif dan dilakukan melalui 5 tahap siklus penelitian.

Untuk algoritma decision tree memliki nilai akurasi tertinggi dibandingkan dengan algoritma klasifikasi naive bayes dan nearest neighbour. Sehingga dapat digunakn oleh pihak bank untuk memprediksi nasabah yang berpotensi membuka deposito.

Saran yang dapat direkomendasikan untuk dikembangkannya penelitian dengan menggunakan algoritma lain untuk melakukan perbandingan sehingga dapat diketahui algoritma yang lebih baik performanya

KATA KUNCI : data mining, algoritma, klasifikasi, perbandingan

I. LATAR BELAKANG

Klasifikasi adalah salah satu teknik dari *data mining* yang termasuk dalam *supervised learning (*pembentukan fungsi menggunakan *training dataset,* digunakan sebagai sebuah pengalaman masa lalu pada sebuah model. Yang memiliki tujuan untuk memprediksi nilai *output* dari sebuah fungsi dengan *input* baru setelah menyelesaikan proses *training*.
(F, 2011)

Dari teknik tersebut memiliki banyak algoritma dengan tingkat ketepatan akurasi yang berbeda-beda antar satu sama lain. Akurasi digunakan sebagai pengatur ketepatan dan kemiripan hasil dengan membandingkan nilai absolut yang semakin mendekati ukuran maka akan semakin tinggi pula tingkat akurasinya. Semakin banyak atribut yang digunakan juga mempengaruhi hasil dari akurasi dan kompleksitas waktu dari masing-masing

algoritma tersebut. Namun tidak hanya pada atribut, *record data* juga mempengaruhi performa algoritma. Penelitian ini akan mengukur kinerja dan membandingkan hasil pengukuran tingkat akurasi algoritma klasifikasi yaitu *naive bayes, nearest neighbour,* dan *decision tree.* Ketiga algoritma ini menghasilkan tingkat akurasi yang berbeda dengan dataset yang sama, beda algoritma beda pula tingkat akurasinya.

II. KAJIAN PENELITIAN TERDAHULU

Penelitian yang diteliti oleh Khafiizh Hastuti dengan judul Analisis Komparasi Logaritma Klasifikasi Data Mining untuk Prediksi Mahasiswa Non Aktif, dengan menggunakan metode *logistic regression, decision tree, naive bayes* , dan *neural network.* Teknik pengujian yang digunakan berdasarkan dataset mahasiswa yang digunakan sebanyak 3.861 mahasiswa dari program studi teknik informatika, sistem informasi, dan desain komunikasi visual tahun akademik 2005-2009. Evaluasi klasifikasi didasarkan pada objek benar atau salah. Dari penelitian tersebut dianalisa bahwa *decision tree* merupakan algpritma yang paling akurat, namun tidak dominan terhadap algoritma lain. *Logistic regression* merupakan algoritma yang paling dominan diantara algoritma lain meskipiun akurasinya rendah. (Hastuti, 2012)

III. METODE

2.1 Metode Penelitian

Dalam penelitian ini menggunakan metode kausal komparatif (penelitian sebab akibat). Penelitian ini lebih berfokus pada hubungan mempengaruhi dan dipengaruhinya variabel yang diteliti yang kemudian memberikan alasan ataupun penjelasan. (Hasibuan, 2007)

Menurut Emzir, alur penelitian dari kausal komparatif adalah yaitu :

1. Permasalahan penelitian
2. Menentukan kelompok data
3. Pemilihan kelompok pembanding
4. Pengumpulan data
5. Analisa data

2.2 Metode Pengumpulan Data

Menggunakan data dokumentasi, yaitu dengan cara pengumpulan data dengan melihat dokumen-dokumen resmi yang telah terjamin keakuratannya. *Dataset* yang digunakan diambil dari repositori publik yang disepakati dalam penelitian oleh para peneliti *data mining (Public Dataset)* yaitu UCI *Machine Learning Repository. Dataset* terkait dengan kampanye pemasaran langsung (telepon seluler) dari insitusi perbankan Portugis.

2.3 Metode Analisis Data

2.3.1 *Data Selection* / Pengumpulan Data

Data yang digunakan sebesar 45.212 *dataset bank marketing* yang terdiri dari 16 atribut prediktor dan 1 atribut label. Sedangkan untuk atribut yang digunakan pada penelitian ini adalah *age, job, marital, education, default, balance, housing, loan, contact, day, month, duration, compaign, pdays, previous, poutcome*, dan atribut label adalah y.

2.3.2 *Pre Processing* / Pemrosesan Awal

Sebelum melakukan proses *data mining*, perlu dilakukan proses *cleaning* pada data yang akan menjadi fokus penelitian. Proses *cleaning* tersebut mencakup antara lain memeriksa data yang inkonsisten, membuang duplikasi data, dan memperbaiki kesalahan pada data. Proses ini dilakukan agar dapat memperoleh hasil akurasi yang benar-benar akurat.

2.4 Proses *Data Mining*

2.4.1 Metode yang digunakan

Untuk memproses data-data yang telah ditentukan tadi dengan *data mining* digunakan metode klasifikasi.

2.4.2 Algoritma yang digunakan

Menggunakan algoritma *nearest neighbour, naive bayes*, dan *decision tree* untuk dibandingkan tingkat keakurasian dari masing-masing algoritma tersebut.

2.5 Evaluasi dan Interprestasi

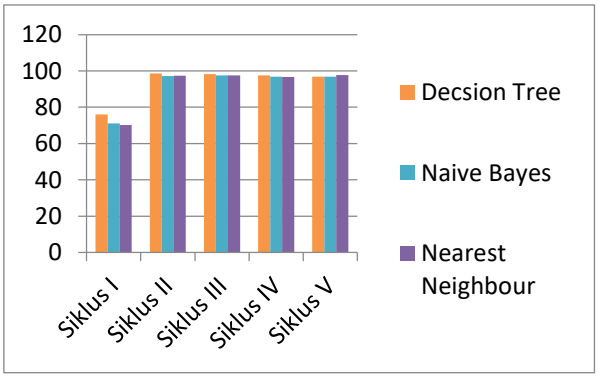
1. *Croos Validation*
2. *T-test*
3. *Confusion Matrix*
4. ROC *Curve*

IV. HASIL

Tabel 4.16 Ringkasan Performa Akurasi 5 Siklus

Siklus	Akurasi		
	Decision Tree	Naive Bayes	Nearest Neighbour
Siklus 1	76.00	71.00	70.25
Siklus 2	98.50	97.12	97.37
Siklus 3	98.29	97.58	97.53
Siklus 4	97.58	96.87	96.62
Siklus 5	96.87	96.87	97.77

Dari kelima penelitian yang telah dilakukan, diambil nilai akurasi tertinggi dari masing-masing algoritma di tiap siklusnya. Dari kelima siklus tersebut, algoritma klasifikasi *decision tree* memiliki nilai akurasi tertinggi dibandingkan dengan algoritma *naive bayes* dan *nearest neighbour*.



Gambar 4.3 Grafik Akhir 5 Siklus

Dari grafik diatas menjelaskan bahwa 4 dari 5 siklus penelitian, algoritma *decision tree* memliki nilai akurasi tertinggi dibandingkan dengan 2 algoritma lainnya *naive bayes* dan *nearest neighbour*.

Tabel 4.17 Ringkasan Performa T-test 5 Siklus

Siklus	T-test		
	Decision Tree	**Naive Bayes**	**Nearest Neighbour**
Siklus 1	0.900	0.960	0.987
Siklus 2	0.960	0.970	0.960
Siklus 3	0.972	0.978	0.989
Siklus 4	0.959	0.961	0.980
Siklus 5	0.958	0.960	0.995

Dari kelima siklus penelitian yang telah dilakukan, diambil nilai AUCnya tertinggi dari masing-masing algoritma di tiap siklusnya. Dari kelima siklus, nilai yang dihasilkan hampir mendekati 1. Yang itu berarti bahwa semua hasil penelitian tersebut baik, karena semakin besar AUC semakin kuat klasifikasi yang digunakan.

V. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Kesimpulan yang dapat diperoleh dari hasil 5 siklus penelitian dari perbandingan algoritma klasifikasi *decision tree, naive bayes*, dan *nearest neighbour* adalah bahwa algoritma klasifikasi *decision tree* memiliki tingkat akurasi tertinggi berdasarkan nilai akurasi *croos validation* dibandingkan dengan algoritma klasifikasi *naive bayes* dan *nearest neighbour*.

5.2 Saran

Dari penelitian yang telah dilakukan, maka penulis dapat memberi saran untuk dikembangkannya penelitian ini yaitu menggunakan algoritma lain untuk melakukan perbandingan sehingga dapat diketahui algoritma yang lebih baik performanya.

VI. DAFTAR PUSTAKA

(2012, Februari 14). Diambil Kembali Dari UCI Machine Learning Repository: Https://Archive.Ics.Uci.Edu

Adi, S. (2015). Perbandingan Kinerja Algoritma Nbc, Svm, C 4.5 Dan Nearest Neighbor : Kasus Prediksi Status Resiko Pembiayaan Di Bank Syariah. Jurnal Ilmiah DASI, Vol. 16, No. 3, 101-106.

B, S. (2007). Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis. Yogyakarta: Graha Ilmu.

Buulolo, E. (2013). Implementasi Algoritma Apriori Pada Sistem Persediaan Obat (Studi Kasus : Apotik Rumah Sakit Estomihi Medan). Pelita Informatika Budi Darma, Volume : IV, Nomor : 1, 71-83.

C, D. A., Baskoro, D. A., Ambarwati, L., & Wicaksana, I. S. (2013). Belajar Data MIning dengan Rapid Miner. Jakarta.

F, G. (2011). Data Mining Concepts, Models and Techniques. Chennai, India: Springer.

Hasibuan, Z. A. (2007). Metodologi Penelitian pada Bidang Ilmu Komputer dan Teknologi Informasi. Depok: Fasilkom Universitas Indonesia.

Hastuti, K. (2012). Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Mahasiswa Non Aktif. Seminar Nasional Teknologi Informasi & Komunikasi Terapan 2012 (Semantik 2012), 241-249.

Jananto, A. (2010). Perbandingan Performansi Algoritma Nearest Neighbor dan SLIQ untuk Prediksi Kinerja Akademik Mahasiswa Baru (Studi Kasus : Data Akademik Mahasiswa Fakultas Teknologi Informasi UNISBANK). Teknologi Informasi Dinamik, Vol. XV, No. 2, 157-169.

Khumaidi, A. (2016). Penerapan Case Based Reasoning dan Algortima Nearest Neighbor untuk Penentuan Lokasi Waralaba. Prosiding Annual Research Seminar, Vol. 2, No. 1, 45-51.

Putranto, R. A., Wuryandari, T., & Sudarno. (2015). Perbandingan Analisis Klasifikasi Antara Decision Tree Dan Support Vector Machine Multiclass Untuk Penentuan Jurusan Pada Siswa SMA. Gaussian, Vol. 4, No. 4, 1007-1016.

Sartika, D., & Sensuse, D. I. (2017). Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian. Jatisi, Vol. 1, No.2, 151-161.

Widiastuti, N. A., Santosa, S., & Supriyanto, C. (2014). Algoritma Klasifikasi Data Mining Naïve Bayes Berbasis Particle Swarm Optimization Untuk Deteksi Penyakit Jantung. Pseudocode, Vol. 1, No. 1, 11-14.

Zaki, M. J., & Jr, W. M. (2014). Data Mining and Analysis. British Library.