

Analisis Pemilihan Algoritma Klasifikasi terbaik dengan studi kasus dalam penentuan Bidang Minat dan bakat santri pada usia 13-15 Di Pondok Pesantren Nurul Jadid

by Jasri Ahyak

Submission date: 12-Dec-2019 01:49PM (UTC+0900)

Submission ID: 1232885840

File name: Faid_Jurnal_JIT_Vol._I,_No._1,_Mei_2017_-_Moh._Jasri_Ahyak.docx (209.31K)

Word count: 2003

Character count: 12033

Analisis Pemilihan Algoritma Klasifikasi terbaik dengan studi kasus dalam penentuan Bidang Minat dan bakat santri pada usia 13-15 Di Pondok Pesantren Nurul Jadid

Moh. Jasri¹ Mochammad faid²

¹STT NURUL JADID PAITON PROBOLINGGO

²STT NURUL JADID PAITON PROBOLINGGO

¹jasri@sttnj.ac.id

²ahmad.faid08@gmail.com

ABSTRAK

Dalam penelitian ini membahas tentang algoritma klasifikasi terbaik dalam hal penentuan minat dan bakat santri, tujuan utama dalam penelitian ini adalah untuk mengetasi masalah klasik yang ada di pesantren nurul jadid, salah satu masalah yang ada di pondok pesantren nurul jadid adalah masalah penentuan minat dan bakat santri karena kendala dilapangan jika minat dan bakat santri tidak sesuai dengan keinginan santri maka akan terjadi adalah santri tidak dapat mengembangkan keilmuannya secara maksimal sesuai dengan minat dan bakatnya, untuk melakukan hal tersebut hal yang paling penting untuk dilakukan adalah mencari algoritma klasifikasi apa yang paling cocok untuk penentuan minat dan bakat santri di pondok pesantren nurul jadid, dan dalam penelitian ini akan menggunakan data santri yang telah menemukan minat dan bakat sesuai dengan keilmuan yang dimiliki santri nurul jadid, data didapat dari pengurus pesantren nurul jadid tepatnya di Biro pesantrenan, data yang digunakan dalam penelitian ini sebesar 1367, dan usia yang dipilih adalah 13-15 tahun karena usia ini dianggap lebih bisa diarahkan untuk dibina sesuai dengan minat bakat santri, sedangkan algoritma yang digunakan dalam penelitian ini adalah algoritma klasifikasi diantaranya C4.5, Naive Bayes, K-NN, dan random forest, adapun hasil akurasi dari C4.5 sebesar 83.9068 %, Naive Bayes 84.3507 %, K-NN 78.9123%, dan random forest 59.7114 %.

Kata kunci: *Klasifikasi data mining, Minat dan bakat, Santri*

ABSTRACT

In this study discusses the classification algorithm best in terms of determining the interests and talents of students, the main objective of this study was to mengetasi classic problem in pesantren nurul jadid, one of the problems that exist in Nurul Jadid is a matter of determining the interest and talent of the students because of the constraints on the field if interests and talents of the students are not in accordance with the wishes of the students that will happen is that students can not develop their knowledge optimally in accordance with their interests and talents, to do the most important thing to do is to find a classification algorithm that is best suited to the interests of the determination and talent of students the pesantren nurul jadid, and in this study will use student data that has found interests and talents in accordance with scientific owned by the student nurul jadid, data obtained from the management of schools Pondok pesantren Nurul Jadid precisely in pesantrenan Bureau, the data used in this study numbered 1367, and the selected age is 13-15 years because this age is considered more geared to be developed in accordance with the interests of students flair, while the algorithms used in this study is the classification algorithms such as C4.5, Naive Bayes, K-NN and random forest, while the results of the accuracy of C4.5 at 83.9068% 84.3507% Naive Bayes, K-NN 78.9123% and 59.7114% random forest.

Keywords: *Data mining classification, Interests and talents, Students*

PENDAHULUAN

Data merupakan sebuah elemen yang tidak bisa terpisahkan dalam suatu objek. Dalam perkembangannya sebuah data tidak hanya menghasilkan sebuah informasi melainkan juga dapat menghasilkan sebuah pengetahuan baru. Sebuah data sangatlah penting bagi sebuah instansi ataupun sebuah organisasi, dalam dunia pendidikan data juga berperan penting, data menjadi sebuah acuan dalam pengambilan sebuah keputusan agar keputusan yang diambil tepat sasaran. Dan sebuah data dalam ilmu teknologi disebut juga dengan database. [1]

Minimnya manfaat yang diperoleh dari database tersebut seolah olah, membuat pekerjaan mengambil data ini. Menjadi non added value atau kurang memberikan nilai tambah terhadap dunia pendidikan salah satu cara memanfaatkan data tersebut adalah dengan mengolahnya sehingga pola atau kecenderungan pada data tersebut dapat ditemukan. Dengan mengetahui pola yang terdapat pada data santri, maka hal tersebut akan dapat digunakan dalam membantu pengambilan keputusan bagi pihak pondok pesantren nurul jadid. Kecenderungan atau pola data santri ini bisa digunakan untuk memprediksi minat dan bakat seseorang serta beberapa faktor yang berpengaruh terhadap minat dan bakat santri tersebut, karena tugas utama suatu pondok pesantren nurul jadid adalah menyediakan pendidikan yang baik dan berkualitas. Mengetahui potensi maupun faktor yang berpengaruh terhadap minat dan bakat santri menjadi hal yang penting. Setelah mengetahui faktor faktor tersebut, maka pihak pondok pesantren nurul jadid dapat menerapkan sistem pembelajaran yang tepat untuk santri selain itu pondok pesantren nurul jadid juga dapat mengantisipasi dan segera menanggapi santri yang memiliki minat

bakat yang rendah. Dalam perkembangannya sebuah data tidak hanya menghasilkan sebuah informasi melainkan juga dapat menghasilkan pengetahuan baru, ilmu yang membahas tentang hal tersebut disebut dengan data mining, data mining atau *knowledge discovery in database* merupakan kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan. Pola hubungan dalam set data yang berukuran besar [2].

Salah satu metode yang digunakan dalam mining adalah klasifikasi, metode klasifikasi dapat digunakan untuk memprediksi atau meramalkan minat dan bakat santri berdasarkan variabel variabel tertentu, tapi pada penelitian ini hanya fokus pada pemilihan algoritma apa yang tepat untuk memprediksi minat dan bakat santri

Penelitian ini dilaksanakan pada santri-santri yang telah menemukan minat dan bakatnya, untuk data santri yang digunakan adalah santri usia 13 sampai 15 tahun, karena pada usia ini adalah masa transisi usia anak anak ke usia remaja, dimana semua karakter dan pengetahuan akan lebih berkembang pada usia ini.

METODE RESEARCH

Metode yang digunakan pada penelitian ini adalah metode kualitatif, untuk metode klasifikasi data mining yang akan dipakai dalam penelitian ini adalah metode C4.5, naive bayes, K-NN, dan random forest. Adapun penjelasannya sebagai berikut:

a) Algoritma C4.5

C4.5 diperkenalkan Quinlan (1996) sebagai versi terbaru dari ID3. Dalam induksi tree hanya bisa dilakukan pada fitur bertipe kategorikal (nominal atau ordinal), sedangkan tipe numerik (interval atau rasio) tidak dapat digunakan. Perbaikan yang membedakan

algoritma C4.5 dan ID3 adalah dapat menangani fitur dengan tipe numerik, melakukan pemotongan decision tree. Algoritma C4.5 juga menggunakan kriteria dalam menentukan fitur yang menjadi pemecah pada pohon yang diinduksi [15].

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \cdot Entropy(S_i)$$

Keterangan

- S : Himpunan kasus
- A : Atribut
- N : jumlah partisi A
- |Si| : jumlah kasus pada partisi ke i
- |S| : jumlah kasus pada S

$$entropy(A) = - \sum_{i=1}^n \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|}$$

2
 Keterangan

- S : Himpunan Kasus
- A : Atribut
- N : Jumlah partisi Atribut A
- |Si| : jumlah kasus pada partisi ke i
- |S| : jumlah kasus pada S

b) Naive Bayes

Naive Bayes mendasarkan pada asumsi penyederhanaan dimana nilai atribut secara kondisional saling bebas apabila diberikan nilai output. Metode ini merupakan sebuah metode yang berakar pada teorema Bayes [10].

c) K- Nearest Neighbor Classifiers

Nearest Neighbor Classifiers Berdasarkan pada proses pembelajaran menggunakan analogi / learning by analogi. Training sampelnya dideskripsikan dalam bentuk atribut numerik n-dimensi. Tiap sampel mewakili sebuah titik pada ruang n-dimensi. Dengan cara ini, semua training sampel disimpan pada pola ruang n-dimensi. Ketika diberikan "unknown"

sampel, K-nearest neighbor classifier mencari pola ruang K training sampel yang paling dekat "unknown" sampel tersebut. K training sampel ini adalah k nearest neighbor dari unknown sampel. "Closeness : dinyatakan dengan euclidean distance, dimana Euclidean distance antara 2 titik,

$$X = (x_1, x_2, \dots, x_n)$$

Dan

$$Y = (y_1, y_2, \dots, y_n)$$

$$\sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2}$$

Untuk melakukan pembobotan variabel dilakukan dengan metode perbandingan pasangan, bobot variabel ditentukan dengan cara normalisasi vektor eigen yang diasosiasikan dengan nilai eigen pada suatu matriks rasio sebelum membandingkan harus ditentukan dulu skala nilai pengaruh atau penting antara variabel [10].

Tabel 1. Skala untuk perbandingan pasangan nilai defenisi

Nilai	Definisi
1	Sama Penting
2	Sama hingga cukup penting
3	Cukup penting
4	Cukup penting hingga tinggi kepentingannya
5	Tinggi kepentingannya
6	Tinggi kepentingannya hingga sangattinggi
7	Sangat tinggi kepentingannya
8	Kepentingannya sangat tinggi hingga amat sangat tinggi
9	Kepentingannya amat sangat tinggi

d) Random Forest

Metode klasifikasi *random forest* merupakan metode pembelajaran ensemble untuk klasifikasi, regresi dan

taks lain, yang beroperasi pada pembangunan pohon keputusan dengan training time dari setiap pohon

HASIL DAN ANALISA

1. Dataset Santri

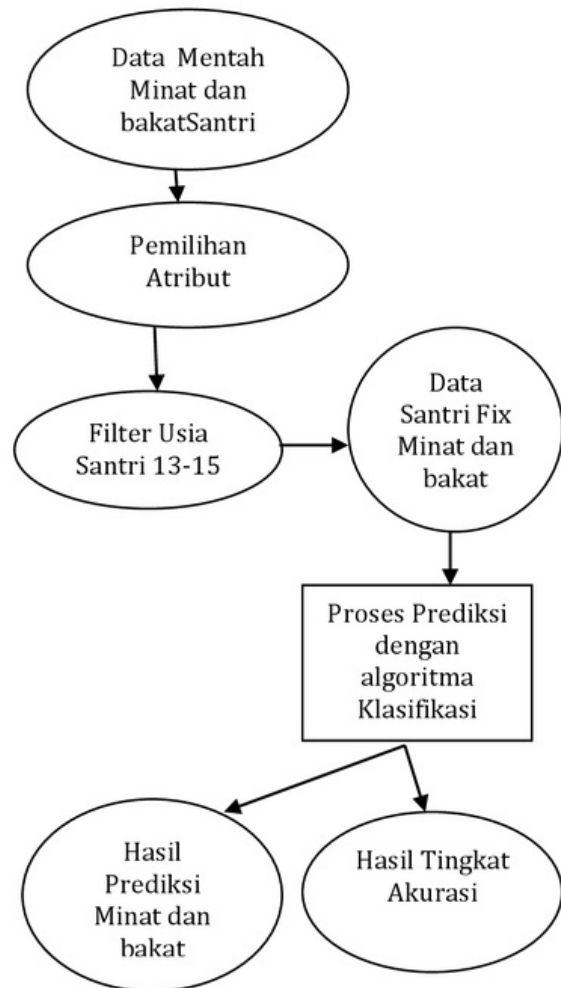
Dataset santri yang akan digunakan adalah data santri yang berusia 13 sampai 15 tahun dengan attribute sebagai berikut :

Tabel 2. Attribute yang digunakan dalam penelitian ini

Attribute	Type
Pekerjaan Orang Tua	String
Pendidikan orang tua	String
Penghasilan orang tua	String
Jenis Kelamin Santri	String
Kelas	String
Minat dan Bakat	Class

2. Proses Pengolahan Data

Adapun pengolahan datasetnya sebagai berikut, pemilihan atribut yang memberikan pengaruh terhadap minat dan bakat dari seorang santri, yang pertama background orang tuanya, karena pendidikan yang pertama kali diterima oleh anak adalah pendidikan dari orang tua baik itu ayah maupun ibu, yang kedua adalah jenis kelamin dari santri tentunya minat dan bakat santri laki-laki dan perempuan pasti akan berbeda, yang ketiga adalah asal daerah santri mengingat setiap daerah akan mempengaruhi cara berfikir santri, sikap ,prilaku santri dan kebudayaan masing masing daerah pasti akan berbeda dan ini pasti akan mempengaruhi terhadap minat dan bakat santri, adapun alurnya nampak pada gambar 1



Gambar1 Block Diagram Tahapan penelitian

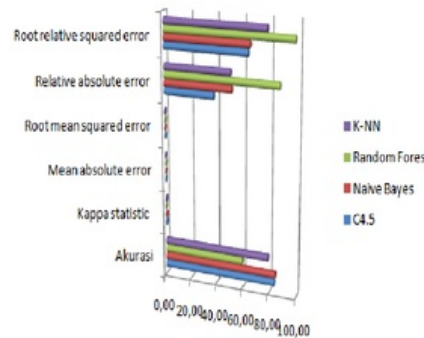
Adapun hasil dari akurasi dari perhitungan dari algoritma klasifikasi adalah tampak pada tabel3 sedangkan grafik perbandingan akurasinya nampak pada gambar2 dan gambar3

Tabel 3. Detail Akurasi Masing masing algoritma

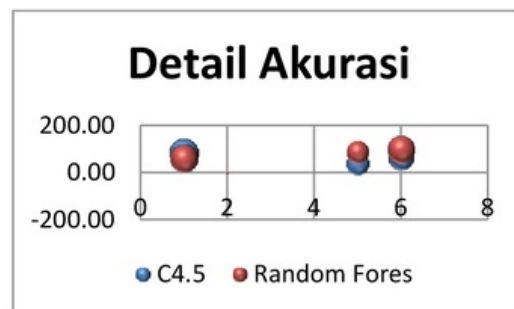
Algoritma	Akurasi
7 C4.5	83.9068 %
	Kappa statistic 0.7142
	Mean absolute error 0.091
	Root mean squared error 0.2205
	Relative absolute error 37.9339 %
	Root relative squared error 63.724 %
8 Naive Bayes	84.3507 %
	Kappa statistic 0.7221
	Mean absolute error 0.1239
	Root mean squared error 0.2253
	Relative absolute error 51.6671 %
	Root relative squared error 65.1093 %
11 Random Fores	59.7114 %
	Kappa statistic 0.1663
	Mean absolute error 0.2081
	Root mean squared error 0.3372
	Relative absolute error 86.7554 %
	Root relative squared error 97.4578 %
10 K-NN	78.9123 %
	Kappa statistic 0.628
	Mean absolute error 0.1215
	Root mean squared error 0.2676
	Relative absolute error 50.6362 %
	Root relative squared error 77.3519 %

Tabel 4. Detail Akurasi Masing masing algoritma

Klasifikasi Minat Santri
FASHION-BATIK
OLAH-RAGA
DAKWAH
HADIST
ALQURAN



Gambar2 Detail Akurasi dengan bermacam metode klasifikasi



Gambar3 Detail Akurasi Algoritma yang Dominan dan Stabil

3. Kesimpulan

Dari hasil perhitungan yang ada tentang algoritma klasifikasi c4.5, Naive bayes, K-NN, dan random fores hasilnya menunjukkan sesuatu yang menarik bahwa algoritma yang memiliki akurasi terbaik diperoleh dari algoritma Naive bayes dengan tingkat akurasi 84.3507 %, namun algoritma yang memiliki keseimbangan akurasi antara Kappa statistic , Mean absolute error, Root

mean squared error, Root relative squared error diperoleh dari C4.5 dan random forest jika aplikasi penentuan minat dan bakat santri ini akan dibuat disarankan untuk menggunakan C4.5 karena selain tingkat akurasi yang cukup tinggi, C4.5 juga memiliki keseimbangan pada tiap akurasi.

REFERENCES

1. Dr. V. Karthikeyani, 2013. Comparison A Performance of Data Mining Algorithms (CPDMA) In Prediction of Diabetes Disease. IJCSE.
2. S. Santhos Kumar, E. Ramaraj, 2012. Analysis of Sequence Based Classifier Prediction for HIV Subtypes. IJSE.
3. Bendi Venkata Ramana, Prof. M. Surendra Prasad Babu, Prof. N. B. Venkatewarlu, 2011. A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis. IJDMS.
4. Made Sumarwati, Waluyo Sejati, Roisca Dyah Pramitasari, 2008. Eksplorasi Persepsi Penderita Tentang Faktor-Faktor Penyebab dan Dampak Penyakit Diabetes Mellitus di Wilayah Puskesmas Purwokerto Barat, Kecamatan Purwokerto Barat, Kabupaten Banyuwang. Jurnal Keperawatan Soedirman (The Soedirman Journal of Nursing).
5. Handayani, 2012. Modifikasi Gaya Hidup Dan Intervensi Farmakologis Dini Untuk Pencegahan Penyakit Diabetes Mellitus Tipe 2 (DM-2), Media Gizi Masyarakat Indonesia.
6. U.S. Department of Health and Human-Services, 2004. Prevention, Detection, Evaluation, and Treatment of High Blood Pressure
7. Jiawei Han and Micheline Kamber, Data Mining: Concepts and Techniques-Third Edition, Elsevier, 2012
8. Ian H. Witten, Frank Eibe, Mark A. Hall, Data mining: Practical Machine Learning Tools and Techniques 3rd Edition, Elsevier, 2011
9. Markus Hofmann and Ralf Klinkenberg, RapidMiner: Data Mining Use Cases and Business-Analytics Applications, CRC Press Taylor & Francis Group, 2014
10. Daniel T. Larose, Discovering Knowledge in Data: an Introduction to Data Mining, John Wiley & Sons, 2005
11. Ethem Alpaydin, Introduction to Machine Learning, 3rd ed., MIT Press, 2014
12. Florin Gorunescu, Data Mining : Concepts, Models and Techniques, Springer, 2011
13. Oded Maimon and Lior Rokach, Data Mining and Knowledge Discovery Handbook Second Edition, Springer, 2010
14. Warren Liao and Evangelos Triantaphyllou (eds.), Recent Advances in Data Mining of Enterprise Data: Algorithms and Applications, World Scientific, 2007
15. Yosoa Putra Raharja, 2013, rancangan bangun sistem rekomendasi berbasis menggunakan algoritma klasifikasi C4.5, Semarang

Analisis Pemilihan Algoritma Klasifikasi terbaik dengan studi kasus dalam penentuan Bidang Minat dan bakat santri pada usia 13-15 Di Pondok Pesantren Nurul Jadid

ORIGINALITY REPORT

45%

SIMILARITY INDEX

43%

INTERNET SOURCES

9%

PUBLICATIONS

23%

STUDENT PAPERS

PRIMARY SOURCES

1	anzdoc.com Internet Source	10%
2	ejournal.ikado.ac.id Internet Source	7%
3	jurnal.poltekomp.ac.id Internet Source	5%
4	sihite.blogspot.com Internet Source	5%
5	www.scribd.com Internet Source	2%
6	lppm.poltekomp.ac.id Internet Source	2%
7	Submitted to AUT University Student Paper	1%
8	Submitted to Auckland University of Technology Student Paper	1%
9	Larissa Navia Rani. "Klasifikasi Nasabah	

Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit", INOVTEK Polbeng - Seri Informatika, 2016

Publication

1%

10

artemis.library.tuc.gr

Internet Source

1%

11

stevelowe.co.nz

Internet Source

1%

12

www.stikes-bth.ac.id

Internet Source

1%

13

Submitted to Universitas Islam Indonesia

Student Paper

1%

14

Submitted to Universitas Negeri Surabaya
The State University of Surabaya

Student Paper

1%

15

osdir.com

Internet Source

1%

16

dodream.tv

Internet Source

1%

17

sentia.polinema.ac.id

Internet Source

1%

18

Submitted to Universitas Terbuka

Student Paper

<1%

19

www.psiquiatriafmusp.org.br

Internet Source

<1%

20

id.123dok.com

Internet Source

<1%

21

Submitted to Imperial College of Science,
Technology and Medicine

Student Paper

<1%

22

e-archivo.uc3m.es

Internet Source

<1%

23

daniellehamiltonphotography.blogspot.com

Internet Source

<1%

24

id.scribd.com

Internet Source

<1%

25

Submitted to Universitas Negeri Jakarta

Student Paper

<1%

26

P.P.P.A.N.W Fikrul Ilmi R.H Zer, Dedy
Hartama, Sundari Retno Andani. "Analisa
Faktor Dominan Mahasiswa Kesulitan
Memahami Bahasa Pemrograman
Menggunakan Metode C4.5", Prosiding
Seminar Nasional Riset Information Science
(SENARIS), 2019

Publication

<1%

Exclude quotes

Off

Exclude matches

Off

Exclude bibliography

On

Analisis Pemilihan Algoritma Klasifikasi terbaik dengan studi kasus dalam penentuan Bidang Minat dan bakat santri pada usia 13-15 Di Pondok Pesantren Nurul Jadid

GRADEMARK REPORT

FINAL GRADE

/100

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6
