

Prediksi Penerimaan Beasiswa Santri Pada Lembaga Pendidikan di Pesantren Nurul Huda Menggunakan Algoritma J48

Faris Muslihul Amin

Universitas Islam Negeri Sunan Ampel Surabaya

Email : farismenk@gmail.com

Terima Naskah : 7 Agustus 2017

Terima Revisi : 25 Agustus 2017

ABSTRAK

Dalam penelitian ini membahas tentang bagaimana memprediksi penerimaan Beasiswa santri pada lembaga pendidikan di Pesantren Nurul Huda Dengan Algoritma J48, salah satu kendala yang ada pada lembaga pendidikan Pesantren Nurul Huda adalah tentang penerima beasiswa santri yang kadang kala tidak tepat saran karena penentuan beasiswanya menggunakan instuisi manusia yang terkadang terjadi human error, penelitian ini mencoba membuat sebuah aplikasi data mining yang bisa memprediksikan penerima beasiswa agar lebih tepat saran dalam menentukan penerima beasiswa santri dengan algoritma J48 yang terintegrasi dengan Java Virtual Mesin (JVM) dan menggunakan Lib Weka, hal yang perlu diperhatikan dalam memprediksi adalah akurasi data semakin tinggi akurasi data maka semakin tepat prediksi penerimaan beasiswanya, data yang akan digunakan dalam penelitian berjumlah 1635 santri, adapun atribut yang akan digunakan dalam penelitian ini meliputi pendapatan orang tua, pekerjaan orang tua, pendidikan orang tua, dan status santri (Yatim, Yatim Piatu, Tidak), dengan akurasi mencapai 87%

Kata kunci: Klasifikasi data mining, Beasiswa Santri, J48

ABSTRACT

In this study discusses about how to predict acceptance of santri scholarship at educational institution at Pesantren Nurul Huda With Algorithm J48, one of the obstacles in Nurul Huda Islamic Boarding School is about scholarship recipients of santri that sometimes inappropriate advice because the determination of its scholarship using human instuisi sometimes human error occurs, this research tries to make a data mining application that can predict scholarship recipients to be more precise suggestion in determining scholarship recipient santri with J48 algorithm integrated with Java Virtual Machine (JVM) and using Lib Weka, things to consider in predict is the accuracy of data the higher the accuracy of the data the more precise predictions of acceptance of scholarship, the data to be used in the study amounted to 1635 students, while attributes to be used in this study include parent income, parent jobs, education g old, and santri status (Orphans, Orphans, No), with an accuracy of 87%.

Keywords: Data mining classification, Scholarship of santri, J48

PENDAHULUAN

Rabithah Makhadi Islamiyah (RMI) mendefinisikan pesantren sebagai lembaga tafaqih fiddin yang mengemban misi meneruskan risalah Muhammad SAW sekaligus melestarikan ajaran islam yang berhaluan Ahlus sunnah wal jama'ah ala Thariqoh al-Madzahib Al-Arba'ah, pondok pesantren nurul huda berdiri pada tahun 1988 di desa paon kecamatan panarukan kabupaten situbondo, pesantren ini memiliki santri sekitar 1635 santri, untuk memotifasi santri dalam

menempuh pendidikan didunia pesantren maka pengurus pesantren memberikan beasiswa kepada santri yang dianggap layak baik itu yang berprestasi ataupun yang kurang mampu, , salah satu kendala yang ada pada lembaga pendidikan Pesantren Nurul Huda adalah tentang penerima beasiswa santri yang kadang kala tidak tepat saran karena penentuan beasiswanya menggunakan instuisi manusia yang terkadang terjadi human error, hal yang terpenting dalam penelitian ini adalah sebuah data yang falid dan lengkap, Data

adalah fakta berupa angka, karakter simbol, gambar, tanda tanda, isyarat, tulisan, suara, bunyi yang mempresentasikan keadaan sebenarnya. 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 huda. Kecenderungan atau pola data penerimaan beasiswa santri ini bisa digunakan untuk memprediksi penerima beasiswa santri yang benar benar layak untuk menerimanya beberapa faktor yang berpengaruh terhadap penerimaan santri diantaranya adalah pendapatan orang tua, pekerjaan orang tua, pendidikan orang tua, dan status santri. Setelah mengetahui faktor faktor tersebut, maka pihak pondok pesantren nurul huda dapat menerapkan sistem penerimaan beasiswa yang tepat untuk santri. 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].

Software data mining yang digunakan dalam penelitian ini adalah java dengan lib weka, adapun algoritma yang akan digunakan dalam penelitian algoritma J48, J48 adalah nama lain dari algoritma C4.5

METODE

Metode yang digunakan pada penelitian ini adalah metode kualitatif, untuk metode klasifikasi data mining yang akan dipakai dalam penelitian ini adalah metode C4.5. Adapun penjelasannya sebagai berikut:

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].

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

Keterangan

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

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

Keterangan

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

HASIL DAN PEMBAHASAN

1. Dataset Santri

Dataset santri yang akan digunakan adalah data penerimaan beasiswa santri dengan attribute sebagai berikut :

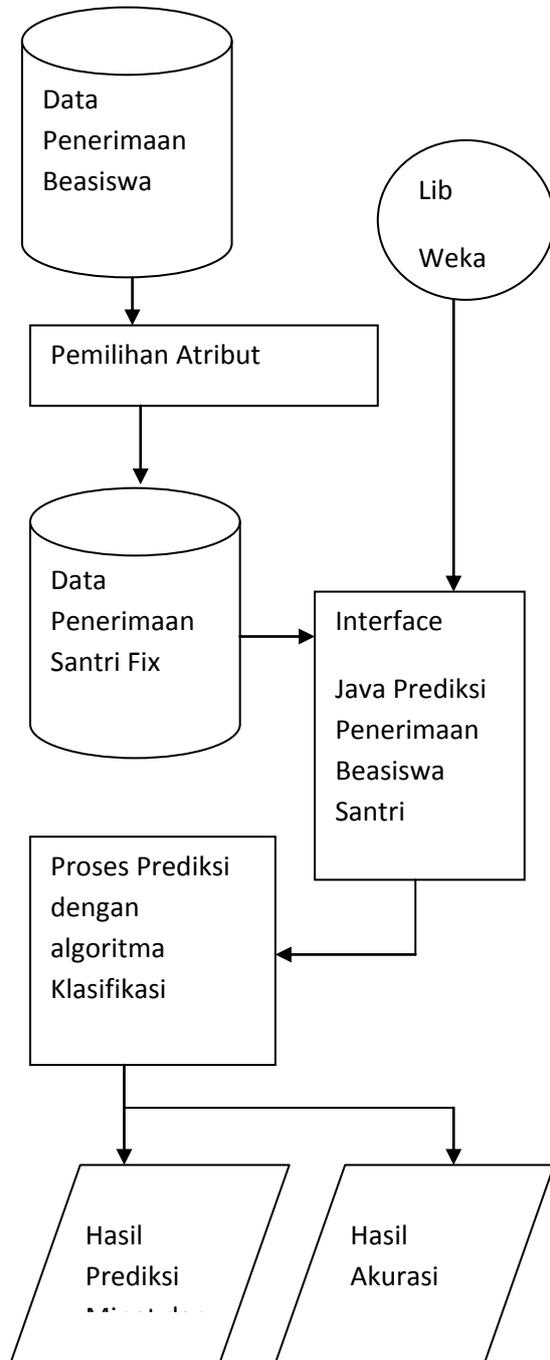
Tabel 1. Attribute yang digunakan dalam penelitian ini

Attribute	Type
Pekerjaan Orang Tua	String
Pendidikan orang tua	String
Penghasilan orang tua	String
Status Santri	String

2. Proses Pengolahan Data

Adapun pengolahan datasetnya sebagai berikut, pemilihan atribut yang memberikan pengaruh terhadap penerimaan beasiswa 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 Status santri (Yatim Piatu, Tidak), adapun alurnya nampak pada gambar 1.



Gambar 1. Block Diagram Tahapan penelitian

3. Hasil pengolahan

Adapun hasil dari akurasi dari perhitungan dari algoritma klasifikasi adalah tampak pada tabel 3.

Tabel 3. Detail Akurasi Masing masing algoritma

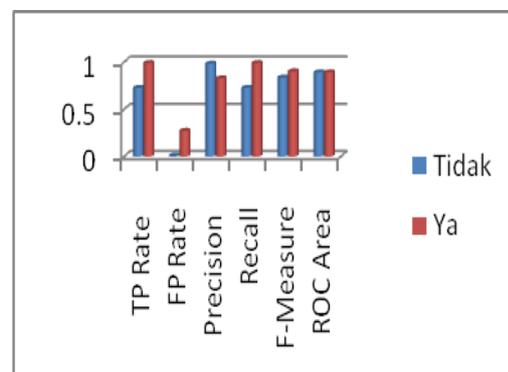
Algoritma	Akurasi
J4.8	87.6989%
	Kappa statistic 0.741
	Mean absolute error 0.1763
	Root mean squared error 0.3029
	Relative absolute error 35.9555%
	Root relative squared error 61.177 %

Dari tabel di atas kita bisa lihat detail akurasi dan errornya dari algoritma J48, Algoritma J48 dipilih karena memiliki tingkat akurasi tertinggi ketimbang akurasi algoritma yang lainnya, adapun perbandingan algoritmanya nampak pada tabel 4

Tabel 4. Cross Validation

Algoritma	Akurasi
J4.8	87.6989%
NBTree	86.4749%
ADTree	87.2705%

Tidak	Ya	Classification
512	192	Tidak
9	921	Ya



Gambar 2. Diagram Batang Detail Akurasi

```

Classifier output
Correctly Classified Instances      1433      87.6989 %
Incorrectly Classified Instances    201      12.3011 %
Kappa statistic                    0.741
Mean absolute error                 0.1763
Root mean squared error             0.3029
Relative absolute error             35.9555 %
Root relative squared error         61.177 %
Total Number of Instances          1634

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area
      0.727   0.01    0.983     0.727   0.836     0.892
      0.99    0.273   0.827     0.99    0.902     0.892
Weighted Avg.  0.877   0.159   0.894     0.877   0.873     0.892

=== Confusion Matrix ===

  a  b  <-- classified as
512 192 |  a = Tidak
  9 921 |  b = Ya

```

Gambar 3. Detail Akurasi dengan bermacam macam metode klasifikasi

SIMPULAN

Dari hasil perhitungan wekanya atribut yang paling berpengaruh adalah atribut Penghasilan orang tua, status santri, tingkat akurasi sebesar 87.6989% dengan algoritma J4.8

DAFTAR PUSTAKA

- [1] SyuhudA.Fatih, .Santri, Pesantren dan tantangan pendidikan Islam. 2008. Pustaka Al-Khoirot.
- [2] S. Santhos Kumar, E. Ramaraj, 2012. Analysis of Sequence Based Classifier Prediction for HIV Subtypes. IJSE.
- [3] Jiawei Han and Micheline Kamber, Data Mining: Concepts and Techniques-Third Edition, Elsevier, 2012
- [4] Ian H. Witten, Frank Eibe, Mark A. Hall, Data mining: Practical Machine Learning Tools and Techniques 3rd Edition, Elsevier, 2011
- [5] Markus Hofmann and Ralf Klinkenberg, RapidMiner: Data Mining Use Cases and Business-Analytics Applications, CRC Press Taylor & Francis Group, 2014
- [6] Daniel T. Larose, Discovering Knowledge in Data: an Introduction to Data Mining, John Wiley & Sons, 2005
- [7] Oded Maimon and Lior Rokach, Data Mining and Knowledge Discovery Handbook Second Edition, Springer, 2010