PEMANFAATAN ALGORITMA *ID3* UNTUK KLASIFIKASI IPK MAHASISWA PROGRAM STUDI PENDIDIKAN MATEMATIKA UN PGRI KEDIRI

SKRIPSI
Diajukan Untuk Memenuhi Sebagian Syarat Guna
Memperoleh Gelar Sarjana Pendidikan (S.Pd)
Pada Program Studi Pendidikan Matematika

OLEH :

MASLULUKHIDAYAH
NPM: 11.1.01.05.0123

FAKULTAS KEGURUAN DAN ILMU PENDIDIKAN (FKIP)
UNIVERSITAS NUSANTARA PERSATUAN GURU REPUBLIK INDONESIA
KEDIRI
2016

HALAMAN PERSETUJUAN

Skripsi oleh:

MASLULUK HIDAYAH

NPM: 11.1.01.05.0123

judul:

PEMANFAATAN ALGORITMA *ID3* UNTUK KLASIFIKASI IPK MAHASISWA PROGRAM STUDI PENDIDIKAN MATEMATIKA UN PGRI KEDIRI

Telah disetujui untuk diajukan Kepada
Panitia Ujian/sidang Skripsi Program Studi Pendidikan Matematika
Fakultas Keguruan Dan Ilmu Pendidikan Universitas Nusantara PGRI Kediri

Tanggal: 22 Desember 2015

Pembimbing I	Pembimbing II
Drs. Darsono, M. Kom.	**Khomsatun Ni'mah, M.Pd.**
NIDN. 0710016401	NIDN. 0703018502

HALAMAN PENGESAHAN

Skripsi oleh

MASLULUK HIDAYAH

NPM: 11.1.01.05.0123

judul:

PEMANFAATAN ALGORITMA *ID3* UNTUK KLASIFIKASI IPK MAHASISWA PROGRAM STUDI PENDIDIKAN MATEMATIKA UN PGRI KEDIRI

Telah dipertahankan di depan Panitia Ujian/sidang Skripsi
Program Studi Pendidikan Matematika
Fakultas Keguruan Dan Ilmu Pendidikan Universitas Nusantara PGRI Kediri
Tanggal: 5 Januari 2016

Dan Dinyatakan Telah Memenuhi Persyaratan

Panitia Penguji :

1. Ketua : **Drs. Darsono, M.Kom.**
2. Penguji I : **Yuni Katminingsih, S.Pd., M.Pd.**
3. Penguji II : **Khomsatun Ni'mah, M.Pd.**

Mengetahui,
Dekan FKIP

Dr. Hj. Sri Panca Setyawati, M.Pd.
NIDN. 0716046202

iii

PEMANFAATAN ALGORITMA *ID3* UNTUK KLASIFIKASI IPK MAHASISWA PROGRAM STUDI PENDIDIKAN MATEMATIKA UN PGRI KEDIRI

MASLULUK HIDAYAH
NPM: 11.1.01.05.0123
FKIP PENDIDIKAN MATEMATIKA
mas_luluk@ymail.com
Drs. Darsono, M. Kom.dan Khomsatun Ni'mah, M.Pd.
UNIVERSITAS NUSANTARA PGRI KEDIRI

ABSTRAK

Masluluk Hidayah: Pemanfaatan Algoritma *ID3* Untuk Klasifikasi IPK Mahasiswa Progam Studi Pendidikan Matematika UN PGRI Kediri, Skripsi, Pendidikan Matematika, FKIP UN PGRI Kediri, 2015.

Penelitian ini dilatar belakangi karena peneliti ingin memanfaatkan data-data yang melimpah sebagai sumber informasi bagi program studi pendidikan Matematika untuk memprediksi IPK mahasiswa dengan menggunakan teknik-teknik *data mining*. Hal ini juga dapat meningkatkan upaya untuk mendorong meningkatkan IPK mahasiswa. Sehingga selain dapat bermanfaat bagi mahasiswa sendiri, juga dapat meningkatkan nilai akreditasi bagi program studi.

Permasalahan penelitian ini adalah (1) Bagaimana penerapan algoritma *ID3* untuk mengklasifikasi IPK Mahasiswa program studi Pendidikan Matematika UN PGRI Kediri ?

Penelitian ini menggunakan pendekatan kualitatif dan jenis penelitian ini adalah studi kasus. Variable terikatnya IPK Mahasiswa dan variable bebasnya asal sekolah, jarak tempat tinggal ke kampus UN PGRI Kediri, pendidikan orang tua, pekerjaan orang tua, penghasilan orang tua, prestasi mahasiswa. Penelitian ini menggunakan *Algoritma ID3* yang menggunakan persamaan *Entropy* dan *Gain*. Tempat penelitian di Kampus UN PGRI Kediri. Populasi penelitian ini semua mahasiswa program studi pendidikan matematika yaitu tingkat 1 sampai dengan tingkat 4. Sedangkan sampel dalam penelitian ini menggunakan metode *Slovin.* Pengambilan data berasal dari database Sistem Informasi Akademik dari BAAK UN PGRI Kediri, Program studi pendidikan matematika dan subjek diminta untuk mengisi lembar isian.

Kesimpulan dari penelitian ini adalah variabel yang paling dominan disini adalah penghasilan orang tua. Dan diperoleh variabel yang mempengaruhi IPK Rendah dan Tinggi. Variabel yang membedakan keduanya adalah asal sekolah Paket C dan Pekerjaan orang tua TNI maka IPK pasti akan tinggi.

Implikasi adalah sebagai Sebagai tambahan informasi yang dapat dijadikan bahan pertimbangan dalam mengambil keputusan pada saat mempertimbangkan penerimaan mahasiswa baru dan sebagai acuan atau pertimbangan bagi peneliti berikutnya yang tertarik pada masalah ini dan ingin menambah serta memperdalam penelitian ini. Hasil klasifikasi yang digunakan untuk memprediksi IPK mahasiswa menunjukkan berbagai klasifikasi serta daftar variabel-variabel penting yang dapat menentukan variabel IPK, sehingga disarankan bagi lembaga agar dapat memperhatikan hal tersebut untuk membantu meminimalisir potensi perolehan IPK dengan kategori rendah

Kata kunci:*Data Mining,* Algoritma *Id3, Entropy, Gain, Rule*

I. LATAR BELAKANG

Dalam dunia pendidikan, data yang berlimpah dan berkesinambungan mengenai siswa yang dibina dan alumni terus dihasilkan.Menurut Jing (2004: 6) dan Merceron (2005) dalam Ayub (2007), data yang berlimpah membuka peluang diterapkannya *data mining* untuk pengelolaan pendidikan yang lebih baik dan *data mining* dalam pelaksanaan pembelajaran berbantuan komputer yang lebih efektif.

> Sementara itu, menurut Luan (2002: 7) menunjukkan bahwa data mining dapat digunakan untuk menyelesaikan siswa yang bermasalah dan membantu institusi menjadi lebih proaktif dalam mengidentifikasi dan merespon siswa tersebut. Luan menerapkan data mining sebagai cara untuk memprediksi ciri-ciri siswa yang akan dikeluarkan oleh sekolah dan kemudian kembali ke sekolah tersebut pada tahun berikutnya.

Dari pendapat diatas dapat disimpulkan bahwa, *data mining* dapat membantu sebuah organisasi yang memiliki data melimpah untuk memberikan informasi yang dapat mendukung pengambilan keputusan.

Salah satu lembaga pendidikan yang cukup besar di Kediri saat ini adalah Universitas Nusantara PGRI (UN PGRI) Kediri yang memiliki 5 fakultas dengan jumlah mahasiswa 13.500. Salah satunya adalah Fakultas Keguruan dan Ilmu Pendidikan. Telah memiliki 11 program studi salah satunya adalah program studi pendidikan matematika dengan jumlah mahasiswa 777. Dengan demikian data-data akademik mahasiswa yang ada juga cukup banyak. Jika data yang melimpah ini hanya dibiarkan menumpuk, maka hanya akan menjadi beban database yang dimiliki.

Sementara itu, data-data yang melimpah ini sebenarnya bisa dimanfaatkan sebagai sumber informasi bagi program studi pendidikan Matematika untuk memprediksi IPK mahasiswa dengan menggunakan teknik-teknik *data mining*. Hal ini tentunya selain dapat memberikan informasi bagi fakultas dan program studi, juga dapat meningkatkan upaya untuk mendorong meningkatkan IPK mahasiswa. Sehingga selain dapat bermanfaat bagi mahasiswa sendiri, juga dapat meningkatkan nilai akreditasi bagi program studi.

Berdasarkan pada latar belakang tersebut, maka penelitian ini dilakukan untuk menerapkan pendekatan *data mining* menggunakan *algoritma ID3* guna memprediksi dan mengklasifikasi IPK kelulusan mahasiswa Program Strudi Pendidikan Matematika UN PGRI Kediri. Dengan *ID3*kita dapat menemukan aturan atau *rules* baru yang belum pernah ditemui sebelumnya.

II. METODE

Penelitian ini menggunakan pendekatan kualitatif,

Menurut Sugiyono (2013 : 15), pendekatan kualitatif adalah : Metode penelitian yang berlandaskan pada filsafat postpositifsime, digunakan untuk meneliti pada kondisi objek yang alamiah (sebagai lawannya adalah eksperimen) di mana peneliti adalah sebagai instrument kunci, pengambilan sample sumber dan data dilakukan secara purposive dan snowbaal, teknik pengumpulan data dilakukan dengan triangulasi (gabungan) analisis data bersifat induktif / kualitatif, dan hasil penelitian kualitatif lebih menekankan pada makna dari pada generalisasi.

Dari beberapa teori-teori di atas, maka dapat kita simpulkan bahwa yang dimaksud dengan penelitian kualitatif adalah metode penelitian yang berlandaskan pada filsafat *postpositivisme,* digunakan untuk meneliti pada kondisi obyek yang alamiah. Dengan tujuan untuk memahami suatu fenomena dalam konteks sosial secara alamiah dengan mengedepankan proses interaksi komunikasi yang mendalam antara peneliti dengan fenomena yang diteliti.

Jenis penelitian yang digunakan dalam penelitian ini adalah studi kasus.

Menurut Afifuddin dan Saebani (2009: 87) dalam studi kasus yang digali adalah entitas tunggal atau fenomena dari suatu masa tertentu dan aktivitas, serta mengumpulkan detail informasi dengan menggunakan berbagai prosedur pengumpulan data selama kasus ini terjadi.

Studi kasus merupakan metodologi penelitian dengan menggunakan satu kasus atau lebih untuk membuktikan teori yang terjadi pada kehidupan nyata. Studi kasus mampu mempelajari dan membedakan antara fenomena dan konteks sehingga memperdalam pengetahuan. Maka dari itu studi kasus sangat dibutuhkan terutama dalam penelitian ini, karena mampu menjelaskan penggunaan teori secara faktual. Dalam penelitian ini, peneliti beranggapan bahwa studi kasus mampu menciptakan pemahaman mendalam terhadap objek atau fenomena yang diteliti.

Dalam penelitian kualitatif, yang menjadi instrumen atau alat penelitian adalah peneliti itu sendiri.

Menurut Sugiyono (2010: 306) peneliti kualitatif sebagai *human instrument*, berfungsi menetapkan fokus penelitian, memilih informan sebagai sumber data, melakukan pengumpulan data, menilai kualitas data, analisis data, menafsirkan data dan membuat kesimpulan atas temuannya.

Sehingga peneliti merupakan instrument kunci dalam penelitian kualitatif, di mana penggunaan instrumen lain sangat tergantung pada peneliti sendiri. Peneliti kualitatif dituntut untuk memiliki kemampuan dan keterampilan dalam pengumpulan data, oleh karena itu sebagai instrumen yang utama, peneliti juga

hendaknya divalidasi, seberapa jauh peneliti kualitatif siap melakukan penelitian dan selanjutnya terjun ke lapangan. Validasi yang dimaksud menurut Sugiyono (2010: 305) meliputi validasi terhadap pemahaman metode penelitian kualitatif, penguasaan wawasan terhadap bidang yang diteliti serta kesiapan peneliti untuk memasuki obyek penelitian, baik secara akademik maupun logistiknya

Tahap pertama penelitian ini adalah studi literatur atau kepustakaan, Tahap kedua penelitian ini adalah pemilihan obyek penelitian, Tahap ketiga adalah penentuan variabel *data mining*, Tahap keempat adalah penentuan nilai klas variable, Tahap kelima adalah pengumpulan data, Tahap keenam adalah mengolah data, Tahap ketujuh adalah melakukan analisis data.

Penelitian ini dilaksanakan di kampus UN PGRI Kediri. Penelitian dilaksanakan pada bulan Mei sampai bulan September 2015. Sumber data yang digunakan dalam penelitian ini adalah data primer dan data sekunder. Pengertian sumber data primer dan sumber data sekunder.

Menurut Sugiyono (2013 : 193) adalah sebagai berikut : Sumber data primer adalah sumber data yang langsung memberikan data kepada pengumpul data, sedangkan sumber data sekunder merupakan sumber yang tidak langsung memberikan data kepada pengumpul data, misalnya lewat orang lain atau dokumen.

Data primer dalam penelitian ini diperoleh langsung dari lapangan. Dalam penelitian ini yang menjadi sumber data penelitian adalah mahasiswa Program Studi Pendidikan Matematika UN PGRI Kediri tingkat 3 dan tingkat 4. Pengambilan data dengan tujuan untuk mengetahui apakah mahasiswa tersebut berprestasi atau tidak berprestasi. Sedangkan data sekunder diperoleh dari Program Studi Pendidikan Matematika yang ada di Fakultas Keguruan dan Ilmu Pendidikan (FKIP) UN PGRI Kediri dan juga berasal dari database Sistem Informasi Akademik dari BAAK UN PGRI Kediri. Data-data tersebut meliputi data kemahasiswaan yang bersifat akademik dan non akademik. Dari informasi tersebut maka peneliti dapat mengolah data sesuai dengan tujuan penelitian.

Analisis data yaitu cara pengolahan data dan menguraikan serta mengevaluasi atas pokok masalah yang diteliti, dari data yang diperoleh akan dibahas dengan menggunakan teori–teori yang mendukung dalam pemecahan masalah.

Menurut Sugiyono (2013: 335) pengertian analisis data adalah proses mencari dan menyusun secara sistematis data yang diperoleh dari hasil wawancara, catatan lapangan dan dokumentasi, dengan cara mengorganisasikan data ke dalam kategori, menjabarkan ke dalam unit-unit, melakukan sintesa, menyusun ke dalam pola, memilih mana yang penting dan

yang akan dipelajari, dan membuat kesimpulan sehingga mudah difahami oleh diri sendiri maupun orang lain.

Teknik analisis data yang digunakan adalah dengan teknik analisis diskriptif, menguraikan dan mendiskripsikan mengenai data–data yang berkaitan dengan data akademik dan non-akademik mahasiswa.

Adapun data yang dideskripsikan antara lain :

1. Mengumpulkan Data

Sumber data adalah data-data yang dikumpulkan dari Program Studi Pendidikan Matematika yang ada di Fakultas Keguruan dan Ilmu Pendidikan (FKIP) di UN PGRI Kediri dan juga berasal dari database Sistem Informasi Akademik dari BAAK UN PGRI Kediri. Data-data tersebut meliputi data kemahasiswaan yang bersifat akademik dan non akademik.

Dan data yang diperoleh langsung dari lapangan. Dalam penelitian ini yang menjadi sumber data penelitian adalah mahasiswa Program Studi Matematika UN PGRI Kediri tingkat 3 dan tingkat 4. Pengambilan data dengan tujuan untuk mengetahui apakah mahasiswa tersebut berprestasi atau tidak berprestasi.

2. Seleksi dan Pembersihan Data

Pada tahapan ini dilakukan pemilihan himpunan data, menciptakan himpunan data target dan atau memfokuskan pada subset variabel (sampel data) yang telah dikumpulkan pada tahap pengumpulan data. Pada tahap ini pula dilakukan pemusnahan data yang tidak konsisten dan mengandung *noise*, duplikasi data dan memperbaiki kesalahan dan ketidaklengkapan data.

3. Transformasi Data

Tahapan ini mentransformasikan atau menggabungkan data ke dalam format yang lebih tepat untuk melakukan proses *mining* dengan cara melakukan peringkasan (*agregation*) disertai dengan proses pendeskripsian data. Proses transformasi data dilakukan untuk data yang bersifat kategorikal.

4. Klasifikasi

klasifikasikan sesuai dengan variable independent yang terdiri dari Asal Sekolah, jarak tempat tinggal ke UN PGRI Kediri, pendidikan orang tua, pekerjaan orang tua, penghasilan orang tua dan prestasi.

Pemrosesan Algoritma ID3

Tabel 3.2
klasifikasi

Variabel			Rentang klas	Keterangan
Nama Variabel	**Klas**	**Nama Klas**		
IPK (Y)	1	Tinggi	≥ 3,00	Sesuai dengan syarat IPK minimal CPNS tahun 2014
	2	Rendah	< 3,00	
Asal sekolah (X1)	1	SMA		
	2	SMK		
	3	MA		
	4	Paket C		
Jarak tempat tinggal ke kampus UN PGRI Kediri (X2)	1	Sangat dekat	$<10km$	
	2	Cukup dekat	$10km \leq C < 40km$	
	3	Dekat	$40km \leq D < 70km$	
	4	Jauh	$70km \leq J < 100km$	
	5	Sangat jauh	$>100km$	
Pendidikan orang tua (X3)	1	Tidak sekolah		
	2	Pendidikan dasar dan menengah		
	3	Pendidikan tinggi		
Pekerjaan orang tua (X4)	1	PNS		
	2	Wiraswasta		
	3	TNI/ POLRI		
	4	Petani/ Nelayan		
	5	Swasta		
	6	Tidak bekerja		
Penghasilan orang tua (X5)	1	Sangat Kurang	Dibawah 1 juta	
	2	Kurang	1 – 3 juta	
	3	Besar	3 – 5 juta	
	4	Sangat Besar	Lebih dari 5 juta	
Prestasi (X6)	1	Berprestasi	≥ 42	Sesuai dengan syarat Danum minimal CPNS tahun 2014
	2	Tidak berprestasi	< 42	

Tabel 3.3
Klasifikasi data akademik dan non akademik mahasiswa

No.	Nama Mahasiswa	NPM	Asal sekolah (X1)	Jarak tempat tinggal ke UN PGRI Kediri (X2)	Pendidikan orang tua (X3)	Pekerjaan orang tua (X4)	Penghasilan orang tua/bln (X5)	Prestasi (X6)	IPK (Y)
1	Risky Purbayanti	11.1.01.05.0123	SMA	Dekat	SD	Petani	Kecil	beprestasi	Tinggi
2									
3									
4									
...									

Dari tabel diatas maka algoritma *ID3* :

1. Pilih atribut sebagai akar

Untuk memilih atribut sebagai akar, didasarkan pada nilai *Gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung *Gain* digunakan rumus seperti tertera dalam Rumus 1.

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

Sedangkan perhitungan nilai *Entropy* dapat dilihat pada rumus 2:

$$Entropy(A) = \sum_{i=1}^{n}\left(-pi * \log 2pi\right)$$

2. Buat cabang untuk masing-masing nilai
3. Bagi kasus dalam cabang
4. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.
5. Ekstraksi *Rule* dari *Decision Tree*

Dalam penelitian ini untuk mengecek keabsahan temuan peneliti menggunakan tehnik Triangulasi Sumber.

Menurut Sugiyono (2014: 373) Triangulasi Sumber, Menguji kredibilitas data dilakukan dengan cara mengecek data yang telah diperoleh melalui beberapa sumber. Data yang diperoleh kemudian dideskripsikan dan dikategorisasikan sesuai dengan apa yang diperoleh dari berbagai sumber tersebut. Peneliti akan melakukan pemilahan data yang sama dan data yang berbeda untuk dianalisis lebih lanjut.

Pengujian keabsahan temuan ini dilakukan dengan mengecek data yang diperoleh melalui berbagai sumber. Sumber data yang pertama adalah data-data yang dikumpulkan dari Program Studi Pendidikan Matematika. Sumber data yang kedua berasal dari database Sistem Informasi Akademik dari BAAK UN PGRI Kediri. Sumber data yang ketiga diperoleh langsung dari lapangan.

Adapun langkah-langkah yang digunakan peneliti dalam melakukan teknik triangulasi sumber yaitu, Data yang diperoleh kemudian dideskripsikan dan dikategorisasikan sesuai dengan apa yang diperoleh dari berbagai sumber tersebut. Peneliti akan melakukan pemilahan data yang sama dan data yang berbeda untuk dianalisis lebih lanjut.

III. HASIL DAN KESIMPULAN

Berdasarkan pembahasan pada bab IV maka hasil penelitian dapat disimpulkan sebagai berikut:

- Variabel yang paling dominan mempengaruhi IPK mahasiswa UN PGRI Kediri adalah **Penghasilan orang tua (X5).** Terbukti bahwa penghasilan orangtua menempati posisi *root.*
- variabel yang mempengaruhi *Rule* dengan klasifikasi IPK rendah adalah

(X1) Asal sekolah: SMK, MA, SMA.

(X2) Jarak tempat tinggal ke kampus UN PGRI Kediri: cukup dekat, dekat, sangat dekat, jauh, sangat jauh.

(X4) Pekerjaan orang tua: PNS, petani, swasta, tidak bekerja, wiraswasta,

(X5) Penghasilan orang tua: sangat kurang, kurang, besar, sangat besar.

(X6) Prestasi: tidak berprestasi, berprestasi.

Sehingga *rule* yang dihasilkan ada 26 *rule*.

- variabel yang mempengaruhi *Rule* dengan klasifikasi IPK tinggi adalah

(X1) Asal sekolah: SMA, MA, SMK, Paket C

(X2) Jarak tempat tinggal ke kampus UN PGRI Kediri: sangat dekat, cukup dekat, dekat, sangat jauh, jauh.

(X4) Pekerjaan orang tua: petani, swasta, tidak bekerja, wiraswasta, PNS, TNI.

(X5) Penghasilan orang tua: sangat kurang, kurang, besar, sangat besar.

(X6) Prestasi: tidak berprestasi, berprestasi.

Sehingga *rule* yang dihasilkan ada 45 *rule*.

Dari beberapa variabel yang mempengaruhi IPK, variabel yang membedakan keduanya adalah bahwa jika asal sekolah **Paket C** dan Pekerjaan orang tua **TNI** maka IPK pasti akan **Tinggi**.

IV. DAFTAR PUSTAKA

Afifuddin&Saebani, B.A. *2009.Metodologi PenelitianKualitatif*. Bandung: Pustaka Setia.

Arikunto, S. 2010. *Prosedur Penelitian*. Cetakan ke empat belas. Jakarta: PT Rineka Cipta

Kusrini & Luthfi, E.T. 2009. *Algoritma Data Mining*. Yogyakarta: Andi Offset.

Luan, J. 2002. *Data Mining and Knowledge Management in Higher Education Applications, Paper presented at the Annual Forum for the Association for nstitutional Research*. Canada: Toronto, Ontario. Tersedia : http://eric.ed.gov/ERICWebPortal/detail?accno=ED474143 diunduh 19 Januari 2015

Sugiyono. 2010. *Metode Penelitian Kuantitatif, kualitatif, dan R & D*. Bandung : Alfabeta.

_____. 2013. *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif dan R&D*. Cetakan ke enam belas. Bandung : Alfabeta.

_____. 2014. *Metode Penelitian Kuantitatif Kualitatif dan R & D*. Cetakan ke dua puluh. Bandung: Alfabeta.

Sulianta, F. & Juju, D. 2010. *Data Mining*. Jakarta: PT Gramedia.

Susanto, S. & Suryadi, D. 2010. *Pengantar Data Mining Menggali Pengetahuan dari Bongkahan Data*. Yogyakarta: C. V Andi Offset