Klasifikasi Penerbitan Surat Keputusan Tunjangan Profesi Guru Menggunakan *Naive Bayes* Berbasis *Information Gain*

Rani Pratikaningtyas[1], Purwanto[2], M. Arief Soeleman[3]
[123]Pasca Sarjana Teknik Informatika Universitas Dian Nuswantoro

ABSTRACT

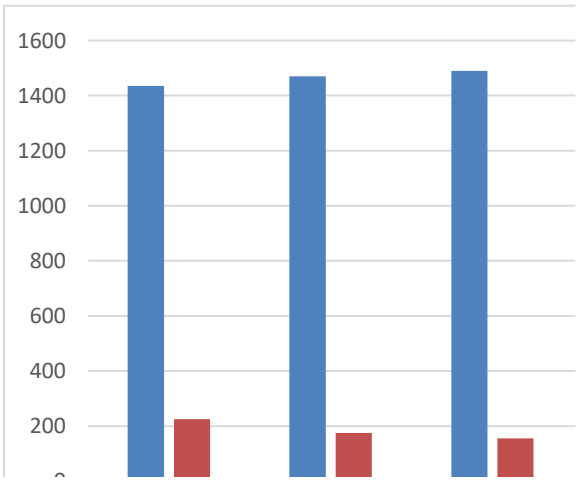
Sertifikasi guru merupakan salah satu upaya pemerintah untuk meningkatkan mutu pendidikan disertai dengan peningkatan kesejahteraan guru. Namun banyaknya penerima sertifikasi yang ternyata tidak cair berpengaruh kepada laporan anggaran belanja negara dan daerah. Penelitian ini bertujuan untuk melakukan seleksi fitur dengan cara memberi bobot pada setiap atribut dari data Penerbitan Surat Keputusan Tunjagan Profesi Guru di Kota Surakarta tahun 2015, menggunakan metode information gain untuk meningkatkan akurasi pada algoritma Naïve Bayes, sehingga dapat mengklasifikaasi penerbitan surat keputusan tunjangan profesi guru dengan baik. Information gain digunakan untuk memilih atribut khususnya dalam menangani data dengan dimensi tinggi. Sedangkan untuk proses klasifikasinya menggunakan algoritma Naïve Bayes yang merupakan teknik prediksi berbasis probabilistic sederhana. Adapun atribut yang digunakan dalam eksperiman ini adalah, NUPTK, Format Bayar, Jenis PTK, Jenis Kelamin, NIP, Status Kepegawaian, Kode Sertifikasi, Area Tugas, Jenjang, JJM Mengajar, Tugas Tambahan, Tugas Mengajar, Golongan, Nama Bank, Keputusan. Hasil Eksperimen dari metode Naïve Bayes didapatkan hasil akurasi sebesar 93,31% sedangkan setelah menggunakan seleksi fitur dengan information gain didapatkan hasil akurasi sebesar 96,11%. Sehingga mengalami peningkatan akurasi sebesar 2,80%.

Keyword: *Sertifikasi guru, Information Gain, Naïve Bayes*

1. PENDAHULUAN

Sertifikasi guru adalah sebuah upaya peningkatan mutu guru dibarengi dengan peningkatan kesejahteraan guru, sehingga diharapkan dapat meningkatkan mutu pembelajaran dan mutu pendidikan di Indonesia secara berkelanjutan. Bentuk peningkatan kesejahteran guru berupa tunjangan profesi sebesar satu kali gaji pokok bagi guru yang telah memiliki sertifikat pendidik. Pascadisahkannya UU No. 14 Tahun 2005 tentang guru dan dosen, profesi guru dan dosen kembali menjadi bahan pertimbangan oleh banyak pihak khususnya bagi mereka yang berkecimpung dalam dunia pendidikan. Mengapa tidak kehadiran undang-undang tersebut manambah wacana baru akan dimantapkannya hak- hak dan kewajiban bagi guru dan dosen. Diantara hak yang paling ditunggu selama ini adalah adanya upaya perbaikan kesejahteraan bagi guru dan dosen, salah satu upaya yang sementara dilaksanakan saat ini dalam rangka implementasi UUGD adalah pelaksanaan sertifikasi guru dalam jabatan sebagaimana telah diatur dalam peraturan Menteri Pendidikan Nasional RI Nomor 18 Tahun 2007.

Setiap triwulan (tiga bulan) pemerintah melakukan pencairan tunjangan profesi guru. Pencairan tunjangan dilakukan setelah pemenuhan beberapa syarat antara lain kualifikasi akademik, jumlah jam mengajar, mengampu mata pelajaran yang linier. Berikut adalah data jumlah penerima tunjangan sertifikasi pada tahun 2015 di ahun 2015 di Kota Surakarta.



(Sumber : UPTD Dikpora Kota Surakarta Bidang PTK)

Gambar 1. Data Pencairan Sertifikasi Guru Kota Surakarta 2015

Dari gambar 1 diketahui adanya perbedaan yang cukup signifikan antar pencairan tunjangan sertifikasi yang tepat waktu dan yang mengalami penundaan. Pada triwulan I di tahun 2015, sebanyak 1660 guru yang menerima sertifikasi, hanya sebanyak 1435 yang menerima pencarian, sedang sebanyak 225 guru tidak menerima pencairan. Pada Triwulan II, sebanyak 1645 guru yang tersertifikasi hanya 1470 guru yang menerima pencairan sedangkan 175 guru tidak menerima pencairan. Sedang pada triwulan III sebanyak 1645 guru yang sertifikasi hanya 1490 yang menerima pencairan sedangkan 155 tidak cair. Adanya keterlambatan dalam pencairan ini tentu akan sangat merugikan baik dari pihak guru maupun pemerintah. Karena dana sertifikasi telah ada namun tidak bisa langsung disalurkan sehingga tentu akan berpengaruh pada laporan keungan kas pemerintah, sedangkan bagi guru tentu saja keterlambatan pencairan akan mengurangi hak yang sudah seharusnya mereka terima. Dimana adanya keterlambatan pencairan akan membuat guru harus kembali mengurus administrasi. Keterlambatan dalam pencairan ini terjadi karena dalam pemasukan data ada berbagai data yang harus dipenuhi oleh guru yang kemudian diolah agar memenuhi syarat untuk mendapatkan Surat Keputusan Tunjangan Profesi (SKTP). Banyak data yang harus diisi sering membuat tertundanya Surat Keputusan Tunjangan Profesi yang tidak turun sehingga pencairan Tunjangan Profesi guru terlambat.

Untuk mengelola data tersebut, dibutuhkan metode yang bisa digunakan untuk menggali informasi – informasi dari data tersebut. Metode tersebut dikenal dengan data mining. Dengan bantuan perangkat lunak, data mining melakukan proses analisa data untuk menemukan pola atau aturan tersembunyi dalam lingkup himpunan data pencairan tunjangan profesi guru tersebut.

Klasifikasi merupakan salah satu teknik dari data mining. Terdapat banyak teknik klasifikasi data mining seperti yang tercantum dalam [1] [2]. Klasifikasi membutuhkan data training untuk mengenali pola tertentu dari data dengan label atau hasil akhir. Kemudian pola tersebut dipakai untuk menentukan label yang belum diketahui dari data baru. Beberapa teknik klasifikasi yang terbaik menurut Wu et al [3] antara lain algoritma C-4.5, Support Vector Machine, k-Nearest Neighbour, serta *Naïve Bayes*.Perbandingan empiris dari kinerja beberapa algoritma klasifikasi tersebut telah menunjukkan bahwa masing-masing baik dalam beberapa masalah domainnya. Dalam kasus ini, tidak ada algoritma klasifikasi terbaik di semua kemungkinan domain [4]. Alasannya, pertama bahwa setiap algoritma pasti mengandung bias untuk memilih generalisasi tertentu, dan akan berhasil jika bias tersebut sesuai dengan karakteristik dari domain aplikasi yang dipengaruhi oleh kompleksitas model. Ketika model tersebut terlalu sederhana, rata-rata prediksi akan menghasilkan bias besar, sebaliknya bila model terlalu kompleks, distribusi prediksi akan

menghasilkan rata-rata prediksi dengan bias kecil. Kedua, keragaman model yang berhubungan dengan tingkat varian yang dihasilkan, artinya tingkat kompleksitas model yang rendah akan menimbulkan nilai varian yang kecil, sedangkan model yang terlalu kompleks menimbulkan nilai varian yang besar [4].

Dari beberapa algoritma klasifikasi yang disebutkan, algoritma *Naïve Bayes* merupakan teknik prediksi berbasis probabilistik sederhana yang modelnya mengunakan fitur bebas. bebas yang dimaksud adalah bahwa pada setiap fitur tidak berkaitan dengan fitur lain dalam data yang sama [5]. Klasifikasi Naive Bayes terbukti memiliki kecepatan yang tinggi saat diaplikasikan ke dalam data yang besar [6] [7].

Peneliti lain juga menyatakan bahwa *Naïve Bayes* dikenal optimal jika pada atribut prediktif bebas, namun percobaan pada data menunjukkan bahwa naive bayes dapat bersaing dengan banyak algoritma klasifikasi lain [8].

Dalam Penelitian Peng-Mian [9] naive bayes adalah algoritma klasifikasi statistik yang efektif dan sukses digunakan dalam bioinformatik, tetapi naive bayes mengasumsikan atribut yang mandiri. Kerangka naive bayes dalam masalah klasifikasi dapat dilihat sebagai masalah menemukan maximum probabilitas yang diberikan satu set variabel yang diamati. Dalam penelitian Peng-Mian, menggunakan data obat dengan 420 fitur, yaitu 20 komposisi asam amino dan 400 komposisi dipeptida. Metode yang diusulkan dalam penelitian itu adalah klasifikasi Naive Bayes, BayesNet, RBFnetwork, Random Forest, J48, SVM dan LogitBoots. Sehingga dapat dihasilkan akurasi dari Naive Bayes sekitar 3%, 4%, 5% dan 7% lebih tinggi dibandingkan dengan metode lain, ini dilihat dari Sn(%) dan auRoc. Hasil ini menunjukan bahwa model Naive Bayes dapat secara efektif digunakan untuk mengklasifikasi data Phage Virion dan Nonvirion Protein [9].

Proses klasifikasi dengan atribut yang terlalu banyak jelas akan memerlukan biaya komputasi yang mahal. Terlebih lagi jika terdapat beberapa atribut redundan atau tidak relevan yang dapat membuat performa algoritma klasifikasi menurun. Seleksi fitur adalah proses pengambilan attribut yang tidak ada hubunganya dengan model yang akan kita bangun. Tujuan seleksi fitur untuk meningkatkan perhitungan yang lebih baik. Proses seleksi fitur ini dilakukan karena ada data yang redundan dan atribut kotor dalam proses membangun pemodelan perhitungan. Salah satu seleksi fitur yang ada adalah information gain pada tahap pre-processing. Information Gain merupakan salah satu algoritma seleksi fitur yang banyak dipakai dan populer [10]. Salah satu kelebihan dari information gain adalah baik digunakan dalam memilih atribut khususnya dalam menangani data dengan dimensi tinggi [11]. Algoritma *Naïve Bayes* (NB) adalah tekhnik machine learning yang populer untuk klasifikasi teks, karena sangat sederhana, efisien dan memiliki performa yang baik pada banyak domain. Namun, *Naïve Bayes* memiliki kekurangan yaitu sangat sensitif pada fitur yang terlalu banyak, sehingga membuat akurasi menjadi rendah. Oleh karena itu, penggunaan Information Gain (IG) untuk seleksi fitur dan metode adaboost untuk mengurangi bias agar dapat meningkatkan akurasi algoritma *Naïve Bayes* [12].

2. PENELITIAN TERKAIT

Sebelum melakukan klasifikasi dengan algoritma *Naïve Bayes* dengan selection fiture Information gain dilakukan beberapa review dari penelitian terkait serta tema yang sama. Beberapa penelitian yang terkait antara lain :

Hotmari Ginting dalam penelitiannya berusaha membuat system pendukung keputusan penentuan prioritas usulan sertifikasi guru dengan menggunakan metode *Simple Additive Weighting* (SAW). Pada hakekatnya metode *Simple Additive Weighting* (SAW) sering juga dikenal dengan istilah metode penjumlahan terbobot. Konsep dasar metode *Simple Additive Weighting* (SAW) adalah mencari penjumlahan terbobot dari rating kinerja pada setiap alternatif pada semua atribut. Metode *Simple Additive Weighting* (SAW) membutuhkan proses normalisasi matriks keputusan, menggunakan MYSQL [13]. Dalam penelitian ini Ari Kurniawan dan Hariadi mencoba menerapkan metode K-Mean clustering untuk mengelompokan kompetensi guru berdasarkan penilaian portofolio dalam sertifikasi guru. Data yang digunakan adalah basis data ASG dan kemudian diolah berdasarkan proses – proses yang ada dalam data

mining. Adapun perangkat lunak yang digunakan adalah MATLAB hasil klasterisasi nilai kompetensi kualifikasi akademeik dan perencanaan dan Pelaksanaan Pembelajaran. Terbangun 2 kelompok dengan kesimpulan : Kompetensi guru dalam membuat RPP tergolong bagus dan merata dari berbagai kualifikasi akademik. Tanda (*) adalah kelompok guru – guru yang perlu mendapat perhatian berkaitan dengan kualifikasi akademik yang dibawah rata – rata [14].

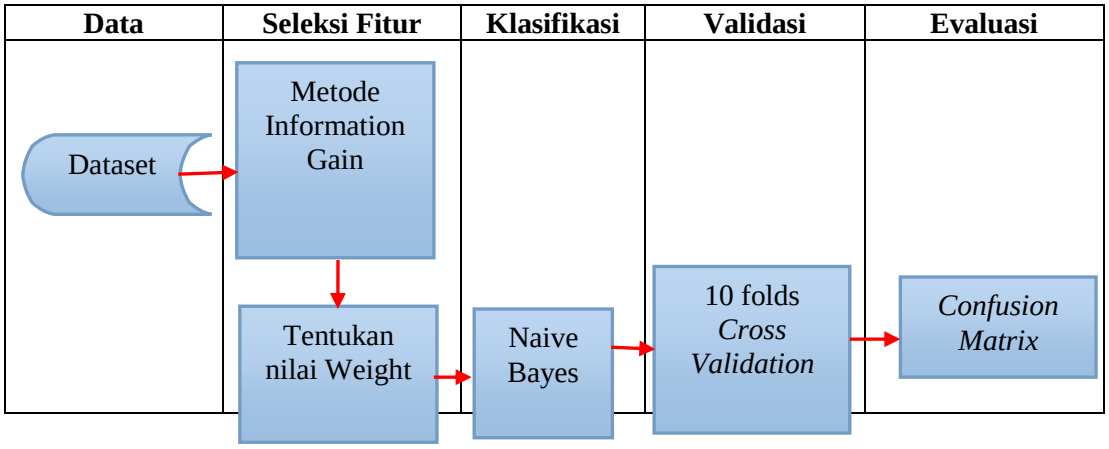
Eddy Purnama dan Mauridhi menggunakan Association Rule dengan metode apriori untuk menemukan pola sertifikasi guru berdasarkan NUPTK. Banyaknya rules yang dihasilkan memberikan banyak kemungkinan untuk melihat pola-pola yang muncul dalam database NUPTK. Sehingga memberikan berbagai kemungkinan yang dapat dijadikan sebagai dasar untuk membuat keputusan. Tidak semua rules yang ditemukan dalam penelitian ini diinterpretasi. Yang diinterpretasi adalah rule-rule yang memiliki nilai Lift yang tinggi (alasan obyektif) dan rule yang memiliki relevansi dengan kebutuhan (alasan subyektif). Dari hasil analisa dan interpretasi Association rule data mining NUPTK, maka dapat disimpulkan bahwa. Association rule mining menggunakan metode apriori berhasil diimplementasikan menemukan 184 rule penting yang tersembunyi dalam database NUPTK [15].

Yuda Septian Nugroho memprediksi kelulusan mahasiswa dengan algoritma naive Bayes. Adapun data yang digunakan adalah data mahasiswa Universitas Dian Nuswantoro Fakultas Ilmu Komputer angkatan 2009 berjenjang DIII dan S1 dengan atribut NIM, Nama, Jenjang, Progdi, Provinsi Asal, Jenis Kelamin, SKS, IPK, dan Tahun Lulus, sebanyak 759 record. Adapun tool yang digunakan dalam proses perhitungan yaitu Rapidminer versi 5.3. Dari hasil perhitungan data mining dan proses pengujian tingkat akurasi dengan menggunakan rapid miner, dapat disimpulkan bahwa angkatan 2009 kelas kelulusan "tidak tepat waktu" lebih besar dari kelas kelulusan "tepat waktu". Sedangkan analisa yang dilakukan terhadap tingkat akurasi menggunakan *Naïve Bayes* menunjukan bahwa nilaiyang dihasilkan oleh algoritma *Naïve Bayes* memiliki tingkat akurasi yang cukup tinggi yaitu 82.08% [16]

Lila Dini Utami beserta Romy Satria Wahono, melakukan penelitian terhadap sentiment review restoran. Klasifikasi sentimen bertujuan untuk mengatasi masalah ini dengan cara mengklasifikasikan ulasan pengguna ke pendapat positif atau negatif. Algoritma *Naïve Bayes* (NB) adalah tekhnik machine learning yang populer untuk klasifikasi teks, karena sangat sederhana, efisien dan memiliki performa yang baik pada banyak domain. Namun, *Naïve Bayes* memiliki kekurangan yaitu sangat sensitif pada fitur yang terlalu banyak, sehingga membuat akurasi menjadi rendah. Oleh karena itu, dalam penelitian ini menggunakan Information Gain (IG) untuk seleksi fitur dan metode adaboost untuk mengurangi bias agar dapat meningkatkan akurasi algoritma *Naïve Bayes*. Penelitian tersebut menghasilkan klasifikasi teks dalam bentuk positif dan negatif dari review restoran. Pengukuran naïve baye berdasarkan akurasi sebelum dan sesudah penambahan metode seleksi fitur. Validasi dilakukan dengan menggunakan 10 fold cross validation. Sedangkan pengukuran akurasi diukur dengan confusion matrix dan kurva ROC. Hasil penelitian menunjukkan peningkatan akurasi *Naïve Bayes* dari 73.00% jadi 81.50% dan nilai AUC dari 0.500 jadi 0.887. Sehingga dapat disimpulkan bahwa integrasi metode information gain dan adaboost pada analisis sentimen review restoran ini mampu meningkatkan akurasi algoritma *Naïve Bayes* [12].

3. METODE PENELITIAN

Metode penelitian pada penelitian ini adalah penelitian eksperimen. Tahapan penelitian dalam penelitian ini antara lain pengumpulan data, pengolahan data awal, eksperimen dan tahap pengujian algoritma dengan menggunakan tools Rapid Miner, kemudian tahap terakhir yang dilakukan adalah evaluasi hasil.



Gambar 2. Model yang Diusulkan

Dalam tahap ini tidak dilakukan koding ataupun membuat sintak tertentu untuk menghitung algoritma. Dikarenakan dalam *tools* Rapid Miner sudah tersedia banyak algoritma data mining. Algoritma seperti *Naïve Bayes* serta algoritma seleksi fitur *information gain* sudah langsung dapat dipakai dengan cara *drag and drop* atau hanya tinggal menempelkan pada posisi data yang sebelumnya telah di *import* ke dalam program.

4. HASIL EKSPERIMEN

Eksperimen dilakukan pada komputer dengan spesifikasi processor intel intel core I3, 2 GB RAM, dan Sistem Operasi Windows 8 dan aplikasi Rapid Miner 5.

Adapun data yang digunakan adalah data sertifikasi tunjangan profesi guru tahun 2015, yang diperoleh dari UPTD Dikpora kota Surakarta bidang PTK. Data yang diperoleh sejumlah 1645 record. Dengan rincian data guru NON PNS 285 record, data guru PNS ada 1360 record. Dalam data penerima tunjangan sertifikasi guru di Kota Surakarta terdapat 14 atribut dan 1 label yatu "Sudah SK" dan "Belum SK". Meta data dari data Penerima Tunjangan Sertifikasi di Kota Solo tahun 2015 triwulan 3 dapat dilihat pada tabel 1.

Dalam eksperimen ini dilakukan perhitungan manual terhapap nilai *Information Gain* dari setiap atribut. Dalam data ini atribut label ada 2 dimana (C1=Sudah SK, C2=Belum SK). Jumlah *record* dalam data ini adalah 1645 record, sehingga nilai s=1076. Dengan rincian 1470 sudah SK sedangkan 175 belum SK. Dengan demikian nilai s1=1470 sampel dari S didalam kelas C1 serta s2=175 sampel dari S di kelas C2. Dengan demikian informasi untuk mengklasifikasikan kelas tersebut adalah :

$$I(s1,s2) = -\frac{1470}{1645} log2 \frac{1470}{1645} - \frac{175}{1645} log2 \frac{175}{1645} = 0{,}489$$

Sehingga S = S11+S12+S21+S22 = 1645 sampel. Jika Sj merupakan jumlah sampel pada masing-masing subset sj maka informasi harapan dari subset NUPTK Valid sudah sk dan NUPTK valid belum SK adalah sebagai berikut :

$$I(S11,S21) = \frac{s11}{s1}\log 2\frac{s11}{s1} - \frac{s21}{s1}\log 2\frac{s21}{s1}$$

$$= \frac{1460}{1578}\log 2\frac{1460}{1578} - \frac{118}{1578}\log 2\frac{118}{1578} = 0{,}18$$

Tabel 1. Meta Data Penerima Tunjangan Sertifikasi Guru di Kota Surakarta Tahun 2015

No	Attribut	Type	Range
1	NUPTK	Binominal	Valid, Tidak Valid
2	Format Bayar	Binominal	Dana Pusat, Transfer Daerah
3	Jenis PTK	Binominal	Guru Kelas, Guru Mapel
4	Jenis Kelamin	Binominal	Laki-laki, Perempuan
5	NIP	Binominal	PNS, Non PNS
6	Status Kepegawaian	Binominal	Valid, Tidak Valid
7	Kode Sertifikasi	Polynominal	020-Guru Kelas, 097-IPA, 100-IPS, 154-PKN, 156-B. Indonesia, 157-B. Inggris, 180-Matematika, 217-Seni Budaya, 220-PJOK, 746-B. Jawa, 800-Pendidikan Luar Biasa
8	Area Tugas	Polynominal	Kecamatan Banjarsari, Kecamatan Jebres, Kecamatan Laweyan, Kecamatan Pasar Kliwon, Kecamatan Serengan
9	Jenjang	Polynominal	SD, SDLB, SLB, SMP, SMPLB, UPTD
10	Jumlah Jam Mengajar	Binominal	Memenuhi, Tidak Memenuhi
11	Tugas Tambahan	Polynominal	Kepsek, Lab, perpustakaan, Wakasek, Tidak Ada
12	Tugas Mengajar	Binominal	Lnier, Tidak Linier
	Golongan	Polynominal	2A Pengatur Muda, 2B Pengatur Muda Tingkat I, 2C Pengatur, 2D Pengatur TK I, 3A penata Muda, 3B Penata Muda Tk I, 3C Penata, 3D Penata Tk I, 4A Pembina, 4B Pembina TK I, 4C Pembina Utama Muda, GTY
13	Nama Bank	Polynominal	BNI, BRI Simpedes, Mandiri
	Keputusan	Binominal	Sudah SK, Belum SK

(Sumber data : Dikpora bidang PTK)

Tabel 2. Pemisahan Atribut NUPTK

S	Jenjang	Sudah SK (s1)	Belum SK (s2)
S1	Valid	1460	118
S2	Tidak Valid	10	57

Sedangkan informasi harapan dari NUPTK tidak valid sudah SK dan NUPTK tidak valid belum SK adalah sebagai berikut :

$$I(S12,S22) = \frac{s12}{s2}\log 2\frac{s12}{s2} - \frac{s22}{s2}\log 2\frac{s22}{s2}$$

$$= \frac{10}{67}\log 2\frac{10}{67} - \frac{57}{67}\log 2\frac{57}{67} = -0{,}21$$

Kemudian nilai *entropy* dari atribut jenjang dapat dihitung dari informasi harapan berdasarkan pemisahan ke dalam subset berikut :

$$E(A) = \frac{s11+s21}{S} I(s11,S21) + \frac{s12+s22}{S} I(s12,s22)$$

$$= \frac{1460+118}{1645} 0{,}18 + \frac{10+57}{1645} - 0{,}21 = 0{,}16$$

Dengan demikian nilai *information gain* dari atribut NUPTK adalah :

$$Gain\ (NUPTK) = |I(S1,S2) - -E(NUPTK)|$$

= 0,488 – 0,16 = **0,328**

Hasil perhitungan setiap atribut bisa dilihat di tabel 3.

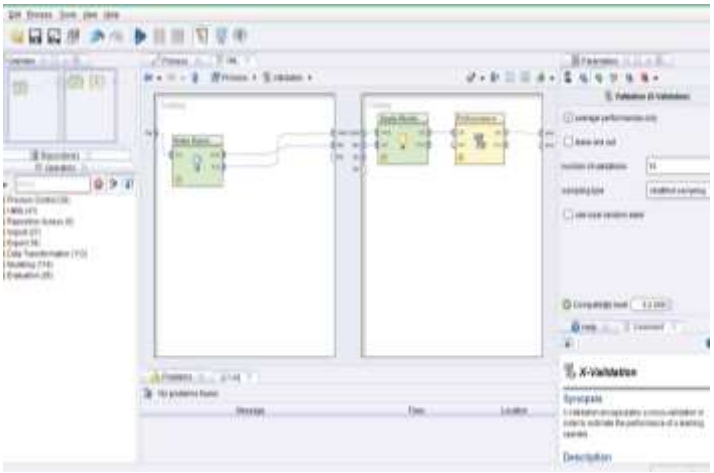
Tabel 3. Nilai Bobot Atribut

ATRIBUT	NILAI BOBOT
NUPTK	0.328
Format Bayar	0.312
Jenis PTK	0.291
Jenis Kelamin	0.2902
NIP	0.306
Status kepegawaian	0.293
Kode Sertifikasi	0.310
Area Tugas	0.355
Jenjang	0.3188
Jumlah Jam Mengajar	0.396
Tugas Tambahan	0.2909
Tugas Mengajar	0.332
Golongan	0.305
Nama bank	0.291

Untuk klasifikasi kelulusan mahasiswa dengan menggunakan algoritma *Naïve Bayes* menggunakan *tools* Rapid Miner. Berikut pengolahan data kelulusan mahasiswa

Gambar 3. Pemodelan Naive Bayes pada Rapid Miner

Gambar 3 adalah gambar percobaan klasifikasi menggunakan *Naïve Bayes* dalam rapid miner. Dalam hal ini menggunakan X Validation untuk membantu menghasilkan tingkat keakurasian berdasarkan dataset kelulusan mahasiswa. Selanjutnya melakukan *testing* dan *training* data. Di kolom *training* terdapat algoritma klasifikasi yang diterapkan yaitu *Naïve Bayes*. Sedangkan di dalam kolom testing terdapat *Apply Model* untuk menjalankan model *Naïve Bayes* dan *Performance* untuk mengukur performa dari model *Naïve Bayes*. Pada gambar 4 menerangkan proses *training* dan *testing* yang terjadi pada dataset. Proses *training* menggambarkan pembuatan model sedangkan proses *testing* untuk menguji hasil dari model yang ada.



Gambar 4. Proses *Training* dan *Testing*

Dari hasil percobaan dengan algoritma *Naïve Bayes* seperti di gambar 4 dengan menggunakan *tools* Rapid Miner diperoleh waktu komputasi yang baik. Ini artinya komputasi menggunakan *Naïve Bayes* berjalan cukup cepat, sesuai dengan kelebihan dari algoritma *Naïve Bayes*.

Hasil akurasi model *Naïve Bayes* dengan tingkat akurasi 93.31% yang dievaluasi menggunakan

confusion matrix. Hasil yang ada memberikan gambaran bahwa hasil akurasi 93.31%, untuk penerbitan surat keputusan tunjangan profesi guru menggunakan *Naïve Bayes* terbukti menunjukan hasil yang baik dengan menggunakan dataset yang sedikit.

Tabel 4. Hasil Penelitian Klasifikasi Penerbitan Surat Keputusan Tunjangan Profesi Guru

Dataset Original Dengan Algoritma *Naïve Bayes*	93.31%	
Atribut	**Nilai *Weight***	**Dataset seleksi fitur *Information Gain***
NUPTK	0.328	95,32%
Format Bayar	0.312	95,32%
Jenis PTK	0.291	95,32%
Jenis Kelamin	0.2902	95,32%
NIP	0.306	95,32%
Status kepegawaian	0.293	95,32%
Kode Sertifikasi	0.310	95,32%
Area Tugas	0.355	95,32%
Jenjang	0.3188	95,32%
Jumlah Jam Mengajar	0.396	96,11%
Tugas Tambahan	0.2909	95,32%
Tugas Mengajar	0.332	95,32%
Golongan	0.305	95,32%
Nama bank	0.291	95,32%

Dari hasil penelitian yang telah dilakukan didapatkan hasil bahwa algoritma seleksi fitur *information gain* dapat meningkatkan performa *Naïve Bayes* untuk klasifikasi pencairan tunjangan sertifikasi.. Dari tabel 4 dapat diketahui bahwa hasil yang diperoleh menggunakan algoritma *Naïve Bayes* hasil akurasinya 93.31%. Setelah dilakukan seleksi fitur tingkat akurasinya menjadi 95.32% Sehingga mengalami peningkatan akurasi sebesar 2.01%. dan atribut dengan bobot terbesar yaitu 0.396 mampu menaikan akurasi sebesr 2.80% yaitu menjadi 96.11%

5. KESIMPULAN

Dari hasil penelitian yang telah dilakukan dapat disimpulkan bahwa penerapan algoritma seleksi fitur *information gain* dapat meningkatkan performa algoritma *Naïve Bayes* untuk pencairan tunjangan profesi guru. Peningkatan akurasi sebesar 2.80%. Akurasi yang diperoleh sebelum menggunakan *information gain* adalah sebesar 93.31% dan setelah menggunakan *information gain* dengan memasukan bobot terbesar akurasinya naik sebesar 96.11%

PERNYATAAN ORISINALITAS

"Saya menyatakan dan bertanggung jawab dengan sebenarnya bahwa artikel ini adalah hasil karya saya sendiri kecuali cuplikan dan ringkasan yang masing-masing telah saya jelaskan sumbernya"

[Rani Pratikaningtyas].

DAFTAR PUSTAKA

[1] I. H. Witten, E. Frank, and M. A. Hall, 2011. Data Mining : Practical Machine Learning Tools and Techniques 3rd Edition. Elsevier.

[2] Daniel T. Larose, 2005. Discovery Knowledge in Data : an Introduction to Data Mining. John Wiley & Sons.

[3] X. Wu, V. Kumar, 2007. Top 10 Algorithms in Data Mining. Vol. 14, no. 1. 2007, pp. 1-37.

[4] Oded Maimon and Lior Rokach, 2010. Data Mining and Knowledge Discovery Handbook. 2nd ed. Springer New York Dordrecht Heidelberg London.

[5] Eko Prasetyo, 2012. Data Mining Konsep dan Aplikasi Menggunakan Matlab. Yogyakarta: Andi Offset

[6] Kusrini and L. E. Taufiq, 2009. Algoritma Data Mining. Yogyakarta: Andi Offset.

[7] Md. Faisal Kabir, 2011. Enhanced Classification Accuracy on Naive Bayes Data Mining Models. Journal of Computer Application (0975 - 8887) Vol. 28 - No. 3.

[8] Wang, Li-Min, 2006. Combining decision tree and Naive Bayes for classification. Knowledge-Based Systems.

[9] Peng-Mian, 2013. *Naïve Bayes* Classifier with Feature Selection to Identify Phage Virion Proteins.

[10] E. Alpaydin, 2010. Introduction to Machine Learning Second Edition.

[11] B. Azhagusundari and A. S. Thanamani, 2013. Feature Selection based on Information Gain. no. 2, pp. 18–21.

[12] Dini Utami, Satria Wahono Romi, 2015. Integrasi Metode Information Gain Untuk Seleksi Fitur dan Adaboost Untuk Mengurangi Bias Pada Analisis Sentimen Review Restoran Menggunakan Algoritma *Naïve Bayes*

[13] Ginting Hotmaria, 2013. Sistem Pendukung Keputusan Penentuan Prioritas Usulan Sertifikasi guru dengan Metode *Simple Additive Weighting*.

[14] Hariadi Moch, et all., 2014. Klasterisasi Kompetensi Guru Menggunakan Hasil Penilaian Portofolio Sertifikasi Guru dengan Metode Data Mining.

[15] Purnama Eddy, 2014. Penerapan Association Rule Mining Pada Data Nomor Unik Pendidik dan Tenaga Kependidikan Untuk Menemukan Pola Sertifikasi Guru.

[16] Yuda Septian Nugroho, n.d. Data Mining Menggunakan Algoritma *Naïve Bayes* untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro.