
Jurnal Penelitian Pendidikan

<https://journal.unnes.ac.id/nju/index.php/JPP>

Klasifikasi Persuratan dengan Menggunakan Algoritma Decision Tree di Lembaga Penelitian dan Pengabdian Kepada Masyarakat

Suminar, Yoris Adi Mareta, Muhammad Kurniawan, Dwi Gansar Santi Wijayanti

Universitas Negeri Semarang, Indonesia

Abstract

Penggunaan dokumentasi perkantoran yang meliputi dokumen kertas yang bersifat cetak dan file yang bersifat soft copy tentu membutuhkan ruang dan kapasitas guna penyimpanannya. Pemanfaatan system informasi surat dinas siradi di Universitas Negeri Semarang dan Arsip di UPT Kearsipan UNNES sangat membantu dalam mengadministrasikan dan penyimpanan. Setiap tahun ada ratusan surat dan dokumen arsip yang disimpan dalam bentuk cetak dan soft copy dan ini tentu memerlukan ruang penyimpanan dan kapasitas. Lembaga Penelitian dan Pengabdian Masyarakat UNNES sesuai dengan tugas pokoknya mengadministrasi kegiatan diantaranya kegiatan Penelitian, Pengabdian, Kuliah Kerja Nyata, Jurnal, Buku, Kekayaan Intelektual dan dokumen Keuangan. Setiap tahun rata-rata mendokumentasi 6.000 laporan Kuliah Kerja Nyata dari mahasiswa, 450 laporan Proposal dan Laporan akhir Penelitian dan pengabdian Kepada Masyarakat dan dokumen lainnya. Pengelolaan arsip yang sudah digital menarik untuk dikaji dan diolah sebagai mesin learning untuk menganalisis bagaimana model dan proses yang ada di arsip. Pengolahan data arsip ini dapat diolah dengan algoritma komputasi yang dapat dijadikan sebagai model penelitian ini. Model arsip yang dikelola oleh Lembaga Penelitian dan pengabdian masyarakat yang sesuai dengan kategori surat dan jenis surat dapat diolah dalam mesin learning dan di analisa.

Keywords: klasifikasi, algoritma decision tree, analisa

PENDAHULUAN

Sejak tahun 2015 UPT TIK melalui Kepala bagian Umum Universitas Negeri Semarang membuat Sistem Informasi Surat Dinas disingkat SIRADI. Sistem ini membuat alur persuratan sudah mulai ditata yang semula menggunakan penomoran surat manual berubah menjadi persuratan yang otomatis dan terdokumentasi secara baik. Sistem informasi surat dinas ini digunakan oleh setiap unit di Rektorat, Fakultas, Pascasarjana Lembaga dan Unit di Universitas Negeri Semarang.

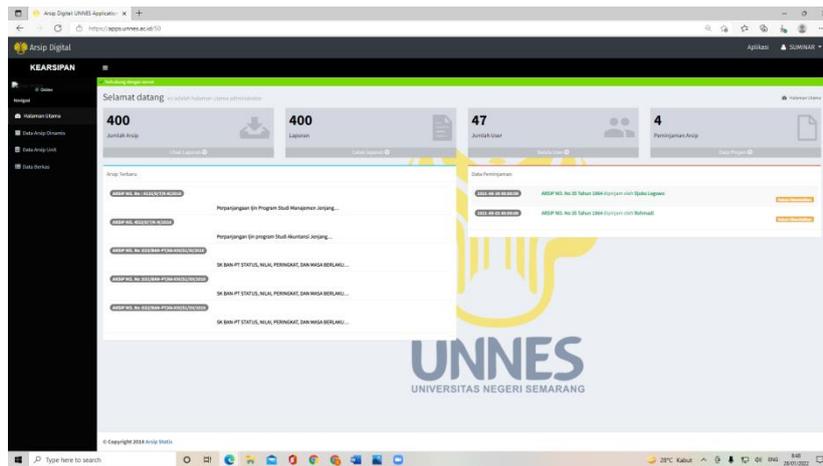
Sejak dibentuknya UPT Kearsipan Universitas Negeri Semarang penataan arsip semakin baik dan semakin dapat dipertanggungjawabkan. Sebelum adanya UPT kearsipan semua arsip dan dokumentasi masih dikelola oleh setiap unit. Arsip yang diciptakan oleh tiap unit sebelumnya masih dikelola oleh unit, sehingga untuk mencari arsip yang dibutuhkan secara cepat tetap menghubungi unit yang mengelola arsip tersebut. (Kebudayaan, 2015)

Lembaga Penelitian pengabdian Masyarakat memiliki 5 Pusat dan 3 Gugus antara lain, 1) Pusat Inovasi dan Komersialisasi, 2) Pusat Pengembang Kuliah Kerja Nyata, 3) Pusat Pengembang Jurnal dan Publikasi, 4) Pusat Pemberdayaan Masyarakat Desa 5) Pusat Desiminasi dan Teknologi dan Kekayaan Intelektual, 6) Gugus Kerjasama, 7) Gugus Inkubator Unit Bisnis 8) Gugus Penjaminan Mutu. Dari setiap Pusat dan Gugus tersebut merupakan unit terkecil dari LPPM yang melaksanakan kegiatan salah satunya membuat surat dan dokumen yang merupakan arsip.

Lembaga Penelitian dan pengabdian kepada Masyarakat dalam menciptakan arsip telah mendokumentasikan arsip dalam aplikasi SIRADI. Dalam aplikasi tersebut terekam semua surat masuk,

surat keluar yang tersimpan dalam server aplikasi. Dengan data arsip yang tersimpan dalam database maka akan sangat mudah mengolah dan manajemen arsip LPPM. Dalam penelitian ini lewat aplikasi peneliti mengambil data yang terdiri surat keluar yang diolah dalam algoritma.

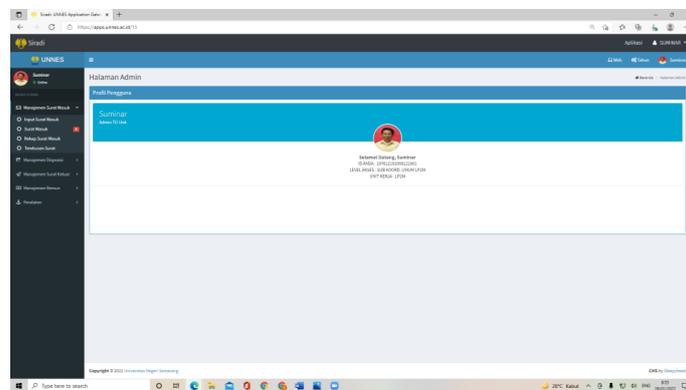
Arsip yang tercipta oleh Lembaga Penelitian dan pengabdian kepada masyarakat dari aplikasi SIRADI dicetak secara fisik dan dilegalisasi dalam setahun tersimpan oleh pencipta arsip. Dokumen tersebut di scan dan disimpan dalam aplikasi arsip Universitas Negeri Semarang dengan alamat arsip.unnes.ac.id. Dalam aplikasi Arsip menyimpan dokumen arsip Dinamis dan arsip stastis.. Dokumen arsip telah di digitalkan melalui soft file dan dimasukkan dalam aplikasi arsip.unnes.ac.id (UNNES, 2013). Aplikasi ini telah dibangun sejak tahun 2019 dan mulai digunakan disetiap unit di Universitas.



Gambar 1 Sistem Informasi arsip unnes

Dengan menerapkan Sistem informasi Surat dinas dan Sistem informasi Arsip sangat memudahkan Lembaga Penelitian dan pengabdian kepada masyarakat dalam menciptakan, mengolah dan melacak. Dengan penggunaan Aplikasi tersebut sewaktu-waktu apabila ada permintaan untuk mencari arsip sangat mudah ditemukan secara digital dan dapat mencetak kembali tanpa harus mencari dokumen di depo arsip. Data Arsip aplikasi ini tersimpan dalam database aplikasi yang pengelolaan aplikasi oleh Unit Pelaksana Teknik Komputer. Saat ini Data base aplikasi tersebut tersimpan dalam data center kedalam satu buah storage server.

Unit Pelaksana Teknis Komputer (UPT Komputer) tugas pokok fungsinya adalah mengelola data center di Universitas Negeri Semarang. UPT Komputer mengelola semua server di lingkungan universitas, salah satunya adalah server yang dijadikan aplikasi SIRADI dan ARSIP. Tahun 2022 UPT Komputer merencanakan akan membuat data center backup yang sering disebut dengan DRC (discovery recovery System). Dengan adanya DRC ini akan menjamin keamanan dan back up data jika terjadi hal hal yang tidak diinginkan.



Gambar 2 Sistem Informasi Surat Dinas

Pengelolaan arsip di LPPM UNNES dalam menciptakan arsip dan mengelola arsip tentu mengalami beberapa permasalahan. Jenis permasalahan penanganan arsip di LPPM UNNES antara lain :

1. Jumlah arsip tiap tahun bertambah banyak jumlah dan dimensinya
2. Media penyimpanan data semakin besar(storage hardisk) dan perlu adanya backup data yang menyeluruh.
3. Media Penyimpan dalam depo arsip statis, dan dinamis di lingkungan lppm tiap tahun diperlukan penanganan yang lebih rapi.
4. Klasifikasi Arsip di LPPM selama ini tidak pernah dianalisis dengan metode algoritma desecion tree.

Latar Belakang Masalah

Klasifikasi arsip persuratan merupakan gambaran tahapan pelaksanaan kegiatan dari tahap perencanaan, persiapan, pelaksanaan dan evaluasi. Skema klasifikasi disusun dalam bentuk berjenjang, yang dijabarkan dari pokok masalah ke sub masalah sampai ke sub-sub masalah. klasifikasi meliputi nama atau judul fungsi, nama kegiatan dan nama transaksi-transaksi kegiatan. Dalam sistem informasi SIRADI klasifikasinya persuratan di Lembaga Penelitian dan pengabdian kepada masyarakat sudah di tentukan melalui peraturan Rektor.

Setiap penciptaan arsip persuratan yang telah dibuat sesuai dengan klasifikasi arsip tersimpan dalam aplikasi SIRADI. Data surat tersebut arsip persuratan tersebut merupakan data besar yang merupakan data mentah dan tidak pernah di evaluasi kebermanfaatannya. Pencipta arsip hanya membuat dokumen surat berdasarkan klasifikasi dan dinput dalam aplikasi.

Arsip persuratan tersimpan dalam aplikasi siradi yang disimpan dalam server. Data ini sangat mudah untuk diambil dan direkap per tahun dan per unit di system informasi. Pengelompokan data ini sangat memudahkan untuk dikelompokkan sesuai kebutuhan.

Rumusan Masalah

Berdasarkan latar belakang tersebut agar permasalahan tidak terlalu meluas dikarenakan keterbatasan maka pada penelitian ini membatasi lingkup permasalahan sebagai berikut:

1. Bagaimana klasifikasi persuratan di Lembaga Penelitian dan pengabdian kepada Masyarakat UNNES menggunakan aplikasi SIRADI.
2. Analisa klasifikasi persuratan di Lembaga Penelitian dan pengabdian kepada masyarakat menggunakan algoritma *decision tree*

Tujuan Penelitian

Penelitian ini menganalisa klasifikasi persuratan di LPPM Universitas Negeri Semarang. Secara khusus, penelitian ini diharapkan dapat:

- a. Mengetahui klasifikasi persuratan menggunakan sistem persuratan di Lembaga Penelitian dan Pengabdian kepada Masyarakat Universitas Negeri Semarang.
- b. Mengetahui klasifikasi persuratan dengan menggunakan metode klasifikasi *decision tree*.

Metode

Penelitian ini akan mengambil data dari system informasi SIRADI dan dilakukan normalisasi. Populasi atau sampel data tersebut diukur dengan instrumen penelitian yang berupa tools klasifikasi data mining algoritma *decision tree*. Analisa data bersifat kuantitatif atau statistik dengan tujuan menguji hipotesis yang telah dibuat. Metode penelitian yang digunakan menggunakan data mining aplikasi *decision tree* dengan tool rapid miner.

Sesuai dengan pedoman penyusunan arsip dinamis sangat diperlukan demi terciptanya arsip yang teratur. Klasifikasi arsip sangat diperlukan sebagai instrument dalam penyusunan arsip dinamis yang dikelola oleh Lembaga. Pimpinan sebuah Lembaga mempunyai tanggung jawab dalam menetapkan klasifikasi arsip. Dalam menetapkan klasifikasi arsip dalam Lembaga harus memiliki ketentuan teknis yang

berupa logis, faktual, perbaikan berkelanjutan, sistimatis, akomodatif dan kronologis.

Universitas Negeri Semarang yang merupakan institusi dibawah kementerian Pendidikan kebudayaan riset dan Teknologi telah menetapkan klasifikasi arsip dinamis. Penyusunan klasifikasi arsip sangat diperlukan sebagai instrument dalam pengelolaan asrip dinamis di lingkungan universitas negeri semarang untuk memfasilitasi penciptaan, akses dan penggunaan, serta penyusutan arsip. Dalam penciptaan arsip dynamin surat keluar merupakan salah satu arsip yang diciptakan oleh perguruan tinggi. Persuratan Universitas jenis klasifikasinya telah ditetapkan dalam peraturan Rektor Nomor 27 tahun 2019.

Prediksi merupakan tujuan yang dihasilkan dalam data mining dalam memproses sebuah data dalam pemrograman. Prediksi merupakan proses menentukan keputusan atas variabel dalam kelas yang sudah ditentukan dalam pengolahan data. Untuk pengolahan Data mining yang besar, setiap data terdiri dari kelas bersama dengan variable dan faktor faktor penentu kelas variabel dalam pengolahan tersebut. Dengan data mining, peneliti dapat menentukan suatu kelas dari variabel data yang dimiliki. (Quinlan, 1993)

Dalam penelitian ini menggunakan pemrograman algoritma C 4.5 untuk output klasifikasi jenis arsip.

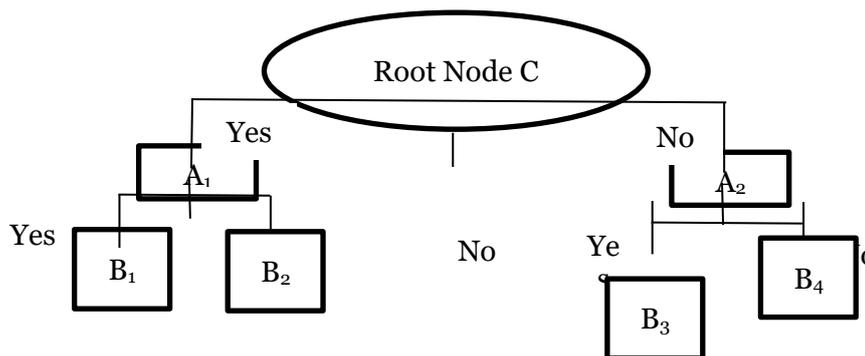
Algoritma ini sering disebut juga pohon keputusan, dikarekan ada beberapa percabangan dari hasil pengolahan datanya. Decision Tree (Pohon Keputusan) merupakan metode klasifikasi (taksonomi) yang menggunakan representasi struktur pohon (tree) di mana setiap node merepresentasikan atribut, cabangnya merepresentasikan nilai dari atribut, dan daun merepresentasikan kelas. Node yang paling atas dari decision tree disebut sebagai root. (Harrington, 2012):

Dalam algoritma ini Root Node dijadikan tatanan yg paling atas, node ini merupakan ujung sehingga dijadikan rujukan, dalam aplikasi rapidminer dinamai *weight*.

Dalam terminologi algoritma ini setelah note root (ujung) membelah menjadi dua yang brarti sesuai atau tidak sesuai, (ya atau tidak) yang merupakan output node root. Untuk Langkah selanjutnya jika dalam node percabangan kemudian memiliki cabang dua lagi sampai seterusnya sampai pada percabangan akhir yang tidak memiliki output atau luaran lainnya.

Dalam algoritma Decision tree banyak motode yang dapat dipakai dalam pembentukan pohon keputusan, salah satunya ada algoritma C4.5. Dalam algoritma ini bisa disusun dalam bentuk tabel dengan atribut dan record yang djadikan atribut. Atribut yang dibuat dijadika parameter yang dibuat sebagai kriteria dalam pembentukan pohon.

Algoritma ini lebih jelas menggunakan kondisi if-then (jika karena) atau dapat dianalogikan ya atau tidak, tetapi tidak membutuhkan parameter dan metrik atau angka yang di nominalkan. Pengolahan data yang strukturnya yang sangat sederhana sangat memungkinkan algoritma ini digunakan untuk memecahkan masalah atribut yang lebih kompleks dan multi-type. Decision tree dapat mengelola nilai-nilai yang hilang atau data noise kemudian dapat di filter dan di olah menjadi keputusan akhir.



Gambar 3 Struktur decision tree Harrington, 2021

Algoritma C4.5 merupakan bentuk dasar dari algoritma pohon keputusan atau decision tree. Sebelumnya dikembangkan algoritma ID3 kemudian disempurnakan menjadi algoritma C4.5. Dalam menentukan pengolahan algoritma ini J. Rose Quinland melakukan perhitungan dengan paling bawah disebut daun dan ujung perhitungan di sebut akar.

Dalam melakukan proses pengolahannya proses alur algoritma C4.5 dalam menentukan pohon

keputusan data mining membutuhkan 4 Proses (Gorunescu, 2011)

Menurut (Hall 1999), tujuan dari seleksi fitur adalah untuk mengidentifikasi beberapa fitur dalam data set isinya sama pentingnya, dan membuang fitur lainnya yang memberikan informasi tidak relevan dan berlebihan. Dalam data set terdapat beberapa atribut yang lebih berguna daripada yang lain, beberapa mungkin lebih membantu proses belajar daripada yang lain (Blum & Langley, 1997).

Untuk menjalankan algoritma Data Mining, karena waktu berjalan tergantung baik pada jumlah record serta pada jumlah atribut dalam setiap record (dimensi). Fitur seleksi terkadang tidak berpengaruh terhadap akurasi tetapi menghemat waktu dalam proses pengolahan data pada atribut (Maimon & Rokach, 2010).

Gain ratio merupakan pengembangan dari informasi gain. Informasi gain digunakan untuk membentuk induksi pohon keputusan (ID3), sedangkan gain ratio digunakan pada C4.5, yang merupakan pengembangan dari ID3 (Quinlan, 1993).

Informasi gain menghasilkan data yang bias, informasi gain lebih memilih fitur dengan banyak variasi nilai daripada fitur yang memiliki sedikit variasi nilai meskipun lebih informatif (Anggraeny & Widayari, 2013). Contoh, fitur unik pada suatu data seperti id siswa dalam tabel siswa di database.

Information gain (IG) adalah pengukuran nilai pada atribut. (Bramer, 2007). Ada 3 tahapan dalam pemilihan fitur menggunakan Information Gain diantaranya adalah sebagai berikut:

- a. menghitung nilai gain informasi untuk setiap atribut dalam dataset asli.
- b. membuang semua atribut yang tidak memenuhi kriteria yang ditentukan.
- c. Hasil Dataset dapat direvisi.

Pengukuran atribut ini dipelopori oleh Claude Shannon pada teori informasi (Gallager, 2001; Risvik, 1997) dituliskan sebagai:

$$Info(D) = - \sum_{i=1}^m p_i \log_2 p_i$$

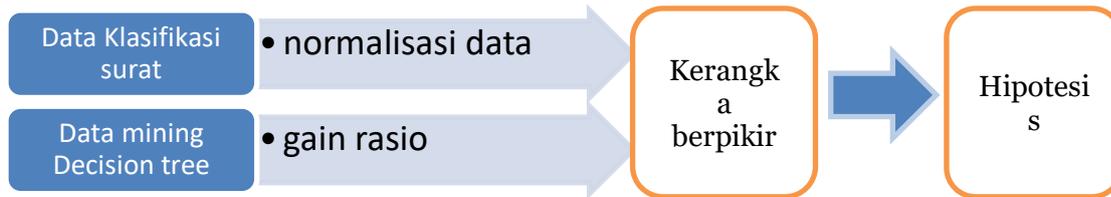
Cross validation merupakan pembuktian dengan membagi data sebagian sebagai data training dan sebagian yang lain sebagai data testing dengan komposisi tertentu. [13].

Dalam penelitian ini dilakukan pengujian dengan 10 folds cross validation yaitu dengan membagi rata data secara acak menjadi 10 bagian kemudian 9 bagian digunakan sebagai data training dan 1 bagian menjadi data testing. Perlakuan seperti ini diulang sampai dengan 10 kali secara berulang dengan mengganti bagian data testing dan data training sampai merata. Lalu hasil akurasi dari 10 kali percobaan tersebut diambil rata-ratanya.

Dari 10 folds cross dibagikan secara acak. Kemudian salah satu bagian dari data (10%) diambil sebagai data testing) dan sisanya menjadi data training. Percobaan tersebut diulang sampai dengan 10 kali sampai keseluruhan bagian data mendapat giliran sebagai data testing. Hasil dari seluruh percobaan diambil akurasi lalu dari keseluruhan tingkat akurasi (a1 sampai dengan a10) diambil rata-rata untuk dijadikan patokan akurasi dari 10 folds cross validation

Penelitian ini merupakan pengolahan data mining, data yang didapat dalam aplikasi SIRADI UNNES diambil dan diolah dalam mesin learning. Mesin learning ini berupa aplikasi rapid miner dengan algoritma decision tree. Dalam teknologi mesin learning ini mampu untuk mengolah data sesuai dengan teori algoritma yang akan dijalankan.

Cara kerja mesin learning berdasarkan algoritma ini meliputi pengumpulan data, eksplorasi data, pemilihan model atau Teknik, memberikan pelatihan terhadap model yang digunakan dan mengevaluasi hasil dari penggunaan rapid miner ini. Untuk lebih lanjut dapat digambarkan pada gambar di bawah ini.



Gambar 4 kerangka berpikir

Penelitian ini menggunakan data mining dengan tahapan penelitian sebagai berikut:

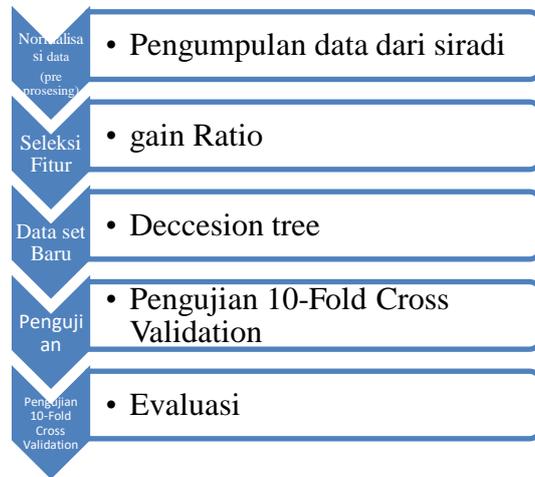
1. Pengumpulan Data
Pengumpulan data di LPPM UNNES ini merupakan langkah awal pada suatu penelitian yang mengambil data dari aplikasi. Data yang digunakan pada penelitian ini adalah data dari persuratan yang tersimpan dalam data siradi UNNES
2. Pengolahan Awal
Pengolahan awal (Pre processing) merupakan tahap untuk mempersiapkan data yang telah diperoleh dari aplikasi dan direkap dalam file format excel, yang akan digunakan pada tahap selanjutnya.
3. Metode yang Digunakan
Tahapan ini akan membahas metode data mining yang meliputi penggunaan algoritma C4.5, kemudian menggunakan fitur seleksi dan di evaluasi menggunakan cross validation.
4. Eksperimen dan Pengujian
Tahapan ini akan membahas tahapan pengolahan penelitian dan teknik pengujian yang akan digunakan. Pengujian yang menggunakan cross validation dan penggunaan table ROC untuk menentukan jenis penelitian ini termasuk baik atau buruk.
5. Evaluasi dan Validasi Penelitian
Tahapan ini akan membahas hasil evaluasi yang telah dimasukkan dalam aplikasi rapid miner dari eksperimen yang telah digunakan. Jika data yang telah di olah dalam aplikasi rapid miner sudah selesai maka disusun validasi data sesuai dengan kebutuhan.
6. Jenis Data
Data yang diambil dari aplikasi penelitian ini menggunakan data dari System informasi surat dinas yang digunakan di Universitas Negeri Semarang terutama pada Lembaga Penelitian dan Pengabdian kepada Masyarakat periode tahun 2018 sampai dengan tahun 2020

Tabel 1 keterangan atribut data set

No	Nama Atribut	Keterangan
1	Nomor Surat	Sesuai klasifikasi penomoran surat di UNNES
2	Jenis Surat	Klasifikasi surat di unnes
3	Hal surat	Perihal yang digunakan
4	Pengirim	Berdasar pengirim LPPM UNNES
5	Tujuan	Tujuan surat yang akan dikirim
6	Penandatanganan	Penanda tanganan Surat
7	sifat	Bersifat segera atau tidak
8	Tembusan	Jika surat terkirim ada tembusannya
9	Jenis surat	Termasuk jenis jenis apa peraturan atau yang lain
10	Kelompok surat	Apakah kelompok penting rutin atau rahasisa
11	Kode surat	Ini klasifikasi kode surat
12	Sifat /keamanan surat	Apakah termasuk terbatas sampai rahasia
13	Derajat surat	Apakah termasuk biasa, segera atau amat segera
14	Isi ringkasan	berisi ringkasan surat
15	Agenda surat	Nomor agenda ini dijadikan kode surat yang telah diterbitkan

7. Pengolahan Data Awal
Data persuratan dalam sistem siradi yang diambil dalam pengolahan ini hanya menggunakan data yang hanya yang akan di uji dan dapat diolah oleh data mining. Data-data tersebut disajikan dalam pengolahan rapid miner untuk mengetahui atribut atribut mana yang memiliki gain yang lebih tinggi
8. Model yang Di usulan

Metode yang diusulkan pada penelitian ini adalah penggunaan pendekatan teknik data mining dan machine learning untuk diklasifikasi jenis jenis surat yang digunakan selama 3 tahun di Lembaga Penelitian Pengabdian kepada Masyarakat menggunakan aplikasi rapidminer 9.10.000 .



Gambar 5 alur penelitian Pengelohan Data Primer

9. Pengumpulan data persuratan dari siradi

Proses data yang di ambil dari aplikasi siradi UNNES dengan melakukan dump data base selama 3 tahun terakhir dan di klasifikasikan sesuai kebutuhan data yang akan diolah dalam aplikasi rapidminer. Data yang tidak banyak seragam dari siradi di sederhanakan dalam atribut dan kebutuhan untuk evaluasi. Data tersebut ditampilkan dalam 14 atribut yang akan di evaluasi sesuai kebutuhan aplikasi Rapidminer

10. Pre prosesing

Dalam pengumpulan data dari siradi, didapat data yang masih mentah dengan beragam inputan yang ada. Data tersebut sering terdapat rentang data kotor, misalnya para operator siradi yang menginput surat namun sebetulnya surat itu tidak jadi digunakan atau dibatalkan, hal ini perlu di bersihkan sebelum data-data masuk dalam mesin learning dan dilakukan pemrosesan.

Untuk mengatasi hal ini diperlukan normalisasi data dengan teknik normalisasi fitur data yang diguhkan. Sehingga semua fitur akan diolah berada dalam jangkauan yang sama. Jika tanpa dilakukan normalisasi berakibat beberapa fitur akan lebih dominan fungsi dibanding fitur yang minimal.

Cara melakukan normalisasi adalah dengan normalisasi linier(Prasetyo, 2014)

11. Gain ratio

Dalam data yang disajikan di penelitian ini data nomor surat dalam tabel sangat variatif, karena merupakan data yang dimiliki oleh setiap recordnya. Pemisahan menggunakan nomor surat akan menghasilkan sangat banyak partisi, karena data nomor surat memiliki nilai unik. Nomor surat dijadikan id sehingga tidak mempengaruhi perhitungan.

12. C 4.5

Perhitungan C4.5 dengan menghitung entropi dan gain pada algoritma ini dapat didapatkan gain tertinggi. Perhitungan dilakukan dengan menggunakan manual dan lewat aplikasi rapidminer dan menghitung informasinya.

Secara umum alur proses algoritma C4.5 untuk membangun pohon keputusan dalam data mining ada 3 tahapan (Gorunescu, 2011):

proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama. Perhitungan secara manual dengan menerapkan rumus yang ada dan menungkan dalam tulisan ini. Kemudian membandingkan dengan menggunakan rapidminer. Hasilnya diamati apakah sesuai dengan menggunakan perhitungan manual

PEMBAHASAN

1. Hasil Pengumpulan data

Data yang digunakan dalam penelitian ini adalah menggunakan data sistem yang terekam dalam sistem informasi persuratan unnes, terutama dalam basis data di Lembaga penelitian dan pengabdian kepada masyarakat LPPM pada tahun 2018, 2019, dan 2020.

Dalam pengambilan laporan ini kami membagi menjadi per tahun untuk mengetahui tren persuratan pada tiap tahunnya.

Data diambil tiap tahun dalam menu siradi dengan menu sebagai berikut:

Tabel 2 Data awal yang diambil dari sistem siradi.

Data Primer 2021

	Tahun 2018	Tahun 2019	Tahun 2020
Jumlah Surat	565	572	101

2. Pengolahan awal

Pengolahan data dengan memfilter awal data yang digunakan dalam Siradi. Data kolom yang tidak digunakan dalam proses analisis ini antara lain :

1. Status surat, status ini berisi surat apakah sudah terkirim atau belum
2. Lampiran surat, kolom ini berisi apakah ada lampiran surat yang akan digunakan.
3. Pilihan, kolom ini berisi perlu cetak agenda yang digunakan oleh aplikasi untuk mencetak

Data yang tidak digunakan dalam pengolahan ini dihapus karena tidak mempengaruhi perhitungan prediksi dan klasifikasi.

3. Penerapan metode yang digunakan

Penggunaan aplikasi rapid miner digunakan untuk memudahkan data Mining sebuah data yang dapat mempercepat dan dapat menganalisis penggunaan data yang akan diukur. Dalam penelitian ini , peneliti menggunakan Algoritma decision tree

4. Aplikasi menggunakan rapid miner

Penggunaan rapid miner sebagai salah satu tool untuk melakukan pengolahan data, atau data mining merupakan cara yang lebih efektif untuk menentukan data yang akan di klasifikasi. Rapid Miner adalah merupakan software untuk melakukan data mining algoritma komputer. Dengan software ini, melakukan text mining atau pengolahan data, yang menganalisis teks, mengekstrak pola, dan mengombinasikannya dengan metode statistika, database, dan kecerdasan buatan atau *artificial intelligence (AI)*

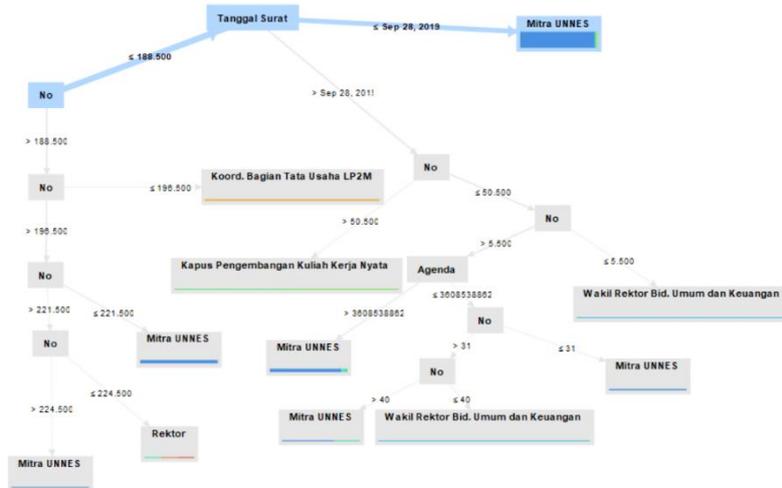
Langkah-langkah yang dilakukan dalam proses penelitian ini adalah sebagai berikut :

1. Pengolahan data dari 3 tahun

Data penelitian ini menggunakan data 3 tahun series yang di klasifikasikan dalam per tahun dan diolah dalam bentuk data yang siap saji dalam pengolahan rapid miner. Pemilahan data yang disesuaikan kebutuhan pengolahan sesuai dengan table yang diolah dengan fitur yang sama.

2. Penerapan gain rasio

Hasil proses data mining dengan algoritma Gain Ratio menggunakan Rapid Miner dapat dilihat pada gambar 6:



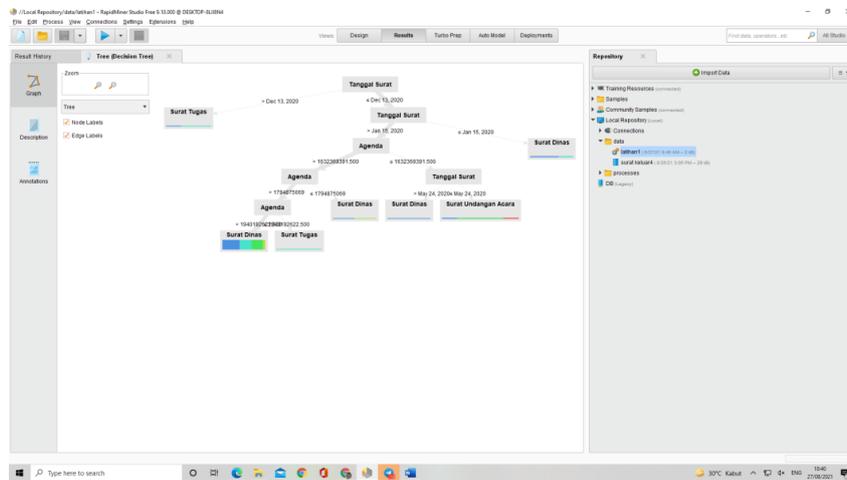
Gambar 6 Hasil Pengolahan Rapidminer

Data Primer 2021

Berdasarkan hasil analisis decision tree di pada gambar di atas dapat dinyatakan bahwa atribut yang paling dominan adalah arsip surat keluar untuk mitra UNNES atau surat keluar yang diberikan diluar instansi dalam universitas. Peringkat kedua arsip surat yang ditujukan kepada Wakil Rektor Bidang umum dan Keuangan.

3. Hasil menggunakan C4.5

Dari hasil penelitian ini menunjukkan algoritma klasifikasi decision tree yang signifikan yang dapat menentukan prediksi penggunaan arsip dari waktu ke waktu sesuai agenda surat.



Gambar 7 Hasil menggunakan rapid miner fitur seleksi

Data Primer 2021

Tingkat akurasi dalam perhitungan algoritma ini menghitung luaran klasifikasi surat berupa tujuan surat sebesar 90.34 %. Hasil ini merupakan data data surat di Lembaga penelitian dan pengabdian kepada masyarakat dapat diprediksi sebesar 90.34 % surat keluar ditujukan kepada mitra UNNES, dengan error margin sebesar +/- 5,35 %. Sedangkan data surat selain itu adalah surat yang ditujukan kepada Unit di sekitar UNNES.

accuracy: 90.34% +/- 5.35% (micro average: 90.31%)

	true Mitra...	true Wakt...	true Reito...	true Reitor	true Deka...	true Kapu...	true Kapu...	true -	true Kaju...	true Keta...	true Koord...	true Deka...	true Wakt...	class prec...
pred. Mitra...	197	3	1	3	1	1	1	1	1	1	1	0	0	93.36%
pred. Wakt...	2	1	0	1	0	0	0	0	0	0	0	0	0	25.00%
pred. Reit...	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. Reitor	0	0	0	0	0	0	0	0	0	0	0	1	1	0.00%
pred. Dek...	2	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. Kap...	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. Kap...	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred -	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. Kaju...	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. Keta...	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
pred. Koord...	0	0	0	0	0	0	0	0	0	0	7	0	0	100.00%
pred. Dek...	0	0	0	1	0	0	0	0	0	0	0	0	0	0.00%
pred. Wakt...	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
class recall	98.01%	25.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	87.50%	0.00%	0.00%	

Gambar 8 akurasi performa

SIMPULAN

Klasifikasi persuratan yang berdasarkan penelitian ini dapat menjawab pertanyaan penelitian :

- a. Bagaimana klasifikasi kearsipan di Lembaga Penelitian dan Pengabdian kepada Masyarakat.

Klasifikasi arsip di LPPM telah diatur melalui peraturan rector Nomor 27 tahun 2019 tentang Pola klasifikasi arsip dinamis dilingkungan universitas negeri semarang. Klasifikasi arsip tersebut merupakan acuan bagi setiap unit kerja dilingkungan UNNES dalam rangka pengelolaan arsip dinamis. Demikian juga dengan LPPM yang merupakan bagian dari UNNES semua pola klasifikasi arsip persuratan mengacu pada peraturan rector tersebut. Kode klasifikasi arsip dijadikan tanda pengenal urusan sesuai fungsi dan tugas unit kerja serta berfungsi sebagai dasar pemberkasan dan penataan arsip persuratan.

Klasifikasi arsip secara substantif berdasarkan alur kegiatan Perguruan tinggi sedangkan fungsi fasilitatif di Perguruan tinggi diurutkan berdasarkan kronologis fungsi fasilitatif. Dalam klasifikasi tersebut antara lain, Penerimaan Mahasiswa, Kurikulum, Tenaga Pendidik, Kemahasiswaan, Perkuliahan, Data dan Informasi dan Pengembangan Akademik, Penunjang akademik, Penelitian, Pengabdian kepada Masyarakat, Publikasi Jurnal Buku, Wisuda dan Alumni, Penjaminan Mutu dan Tata Pamong.

Sedangkan sebagai fungsi fasilitatif Klasifikasi arsip di Universitas Negeri semarang meliputi; Perencanaan, Hukum, Organisasi dan Ketatalaksanaan, Kearsipan, Ketatausahaan, Kerumahtanggan, Perlengkapan, Hubungan Masyarakat, Pendidikan dan Pelatihan, Teknologi Informasi dan Komunikasi, Pengawasan, Kepegawaian, Keuangan, dan Badan Usaha.

- b. Untuk mengetahui klasifikasi persuratan dengan menggunakan metode klasifikasi decision tree dengan memanfaatkan mesin learning dengan menjadikan tujuan arsip persuratan sebagai label. Label ini berupa arsip surat yang dikirim dalam lingkungan unnes dan label arsip surat yang dikirim diluar lingkungan UNNES.

Dari mesin learning didapat hasil akurasi sebesar 90,34 %

Algoritma ini dapat terlihat bahwa penggunaan surat dinas dengan tujuan Mitra UNNES atau surat keluar lebih besar dengan tingkat margin sebesar +/- 5,35 %. Performa algoritma ini menggunakan menggunakan 10 cross validation dengan type sampling secara otomatis.

Untuk lebih dapat mengembangkan penelitian lebih lanjut maka dapat mempergunakan algoritma lain untuk menganalisa hasil penelitian atau menggunakan mesin algoritma lain yang menjalankan analisa data

DAFTAR PUSTAKA

- Anggraeny, F. T., & Widyasri, M. (2013). Klasifikasi Voting ANN PSO Biclass dengan Seleksi Fitur Gain Ratio. *Smntf-Upn Veteran Jatim*, 3–8.
- Blum, A. L., & Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97(1–2), 245–271. [https://doi.org/10.1016/S0004-3702\(97\)00063-5](https://doi.org/10.1016/S0004-3702(97)00063-5)

- Bramer, M. (2007). Principles of Data Mining. In *Springer*. https://doi.org/10.1007/978-1-4471-4884-5_1
- Gallager, R. G. (2001). Claude E. Shannon: A retrospective on his life, work, and impact. *IEEE Transactions on Information Theory*, 47(7), 2681–2695. <https://doi.org/10.1109/18.959253>
- Gorunescu, F. (2011). Data Mining. In *Data mining - Concepts, Models and Technique*. <https://doi.org/10.1007/978-3-642-19721-5>
- Harrington, P. (2012). *Machine learning in action*. http://library02.embl.de/InmagicGenie/DocumentFolder/TableOfContents_H897.pdf
- Kebudayaan, K. P. dan K. (2015). *Permen 23 tahun 2015 Organisasi dan Tata kelola UNNES.pdf* (p. 27).
- Maimon, O., & Rokach, L. (2010). Data Mining and Knowledge Discovery Handbook. In *Vasa*. <https://doi.org/10.1007/978-0-387-09823-4>
- Prasetyo. (2014). DATA MINING MENGOLAH DATA MENJADI INFORMASI MENGGUNAKAN MATLAB. *Penerbit Andi*, 218–241. <https://doi.org/10.1017/CBO9781107415324.004>
- Quinlan, J. R. (1993). C4.5: Programs for Machine Learning. In M. Kaufmann (Ed.), *Morgan Kaufmann San Mateo California* (Vol. 1, Issue 3). Morgan Kaufmann. <http://portal.acm.org/citation.cfm?id=152181>
- Risvik, K. M. (1997). *Discretization of Numerical Attributes Preprocessing for Machine Learning*. Department of Computer and Information Science Norwegian University of Science and Technology.
- UNNES. (2013). *Peraturan Rektor Universitas Negeri Semarang nomor 24 Tahun 2013 tentang Pedoman Pengelolaan Arsip Dinamis Universitas Negeri Semarang* (p. 2). Universitas Negeri Semarang.
- Peraturan Kepala Arsip Nasional Republik Indonesia Nomor 19 tahun 2012 tentang Penyusunan Klasifikasi arsip
- UNNES (2019) Peraturan Rektor Universitas Negeri Semarang Nomor 27 tahun 2019 Tentang Pola Klasifikasi arsip Dinamis dilingkungan Universitas Negeri Semarang