

## Implementasi Web Mining dengan Metode Clustering pada Dokumen Akreditasi Program Studi

Renintha Trianjani Susilo<sup>1\*</sup>, Safitri Jaya<sup>1</sup>

<sup>1</sup>Informatika, Universitas Pembangunan Jaya, Tangerang Selatan

Jl. Cendrawasih Raya B7/P No. 65 Bintaro Jaya, Kel. Sawah Baru Ciputat, Tangerang Selatan, Banten 15413

\*Corresponding Author : [renintha.trianjanisusilo@student.upj.ac.id](mailto:renintha.trianjanisusilo@student.upj.ac.id)

### Abstrak

Akreditasi merupakan salah satu bagian dari Sistem Penjaminan Mutu Eksternal Pendidikan Tinggi yang merupakan upaya pemerintah dalam melakukan standarisasi dengan tujuan untuk menentukan kualitas Program Studi dalam melaksanakan proses pendidikan dan menjamin kualitas lulusan agar sesuai dengan kebutuhan kerja. Salah satu tolak ukur dari sistem penjaminan mutu Program Studi adalah kualitas dari sistem tata kelola dokumen mutu yang dilakukan untuk memenuhi kebutuhan instrumen akreditasi. Selain penataan, Program Studi harus mengetahui informasi yang terdapat pada susunan dokumen mutu guna melakukan proses monitoring dan evaluasi agar pengelolaan dokumen mutu dapat dilakukan lebih baik lagi. Penggalan pola menggunakan sistem web mining dapat digunakan sebagai upaya penyediaan informasi dokumen mutu. Informasi didapat dengan melakukan ekstraksi nama dokumen mutu menggunakan metode *text mining* dan mengelompokkan dokumen mutu sesuai dengan kriteria yang telah ditentukan oleh BAN – PT. Dari 65 dokumen mutu yang digunakan sebagai uji coba terdapat 6 dokumen yang tidak dapat dikelompokkan sesuai dengan kriteria. Hasil pengelompokkan dokumen direpresentasikan kedalam grafik sehingga civitas program studi mengetahui perkembangan dokumen mutu.

**Kata kunci:** Dokumen Akreditasi, Web Mining, Clustering

### Abstract

*Accreditation is one part of the Higher Education External Quality Assurance System which is an attempt by the government to standardize with the aim of determining the quality of the Study Program in carrying out the education process and ensuring the quality of graduates to match the work needs. One of the benchmarks of the Study Program quality assurance system is the quality of the quality document management system which is carried out to meet the needs of accreditation instruments. In addition to structuring, the Study Program must know the information contained in the arrangement of quality documents in order to carry out the process of monitoring and evaluation so that the management of quality documents can be done better. Extracting patterns using a web mining system can be used as an effort to provide quality document information. Information is obtained by extracting the name of the quality document using the text mining method and grouping quality documents in accordance with the criteria determined by BAN - PT. Of the 65 quality documents used as trials there are 6 documents that cannot be grouped according to criteria. The results of grouping documents are represented in the graph so that the study program community knows the development of quality documents.*

**Keywords :** Accreditation Document, Web Mining, Clustering

### PENDAHULUAN

Salah satu tolak ukur dari sistem penjaminan mutu Program Studi adalah kualitas dari sistem tata kelola dokumen mutu yang dilakukan oleh Program Studi. Penilaian

terhadap kualitas dokumen mutu Program Studi tersebut dapat secara internal dilakukan maupun eksternal. Program Studi akan mengalami proses akreditasi yang merupakan salah satu bagian dari Sistem Penjaminan Mutu Eksternal Pendidikan

Tinggi yang bertujuan untuk menentukan kualitas Program Studi dalam melaksanakan proses pendidikan dan juga merupakan suatu upaya pemerintah dalam melakukan standarisasi dan menjamin kualitas lulusan agar sesuai dengan kebutuhan kerja. Dalam melaksanakan akreditasi, dokumen mutu Program Studi umumnya disusun dan disimpan sesuai dengan kriteria instrumen sebagai bentuk pelaporan dan akan dievaluasi oleh Lembaga Akreditasi Mandiri (LAM) yaitu lembaga yang bertugas dalam pelaksanaan akreditasi Program Studi. Kualitas mutu Program Studi dapat dinilai berdasarkan kelengkapan dokumen pendukung yang dapat digunakan untuk memenuhi kebutuhan instrumen akreditasi, sehingga Program Studi dituntut untuk dapat mengelola dokumen mutu yang berkaitan dengan bentuk kegiatan berlangsung sebagai bentuk pembuktian mutu dari Program Studi. Upaya penanganan masalah penataan dokumen mutu yang sejauh ini dilakukan Program Studi sudah sampai ke dalam tahap pengembangan website data warehouse dokumen mutu dan sudah cukup membantu dalam upaya pengelolaan dokumen mutu.

Terlepas dari masalah penataan dokumen, dokumen mutu tetap harus dimonitoring serta di evaluasi secara berkala oleh program studi agar kondisi serta status ketersediaan dokumen dapat terus diperbaharui (Jaya, 2018). Ketersediaan informasi mengenai dokumen mutu milik Program Studi masih sangat minim didapat baik dari mekanisme pengelolaan dokumen secara manual maupun yang berbasis web seperti yang dimiliki oleh Program Studi Informatika saat ini. Pencarian pola dengan menggunakan prinsip Web Content Mining dengan sumber data yang diperoleh dari dokumen mutu milik Program Studi Informatika Universitas Pembangunan Jaya yang terdapat dalam data warehouse dapat dijadikan untuk memperoleh informasi dokumen mutu guna memberikan gambaran kepada Program Studi dalam melakukan perencanaan kedepannya sebagai upaya Program Studi untuk terus meningkatkan kualitas mutunya.

Penelitian ini dilakukan dengan tujuan diantaranya :

1. Untuk menggali pola informasi dari data dokumen yang dimiliki oleh program studi berdasarkan kriteria standar instrumen yang ditentukan oleh BAN – PT.

2. Memperoleh gambaran hubungan dari setiap dokumen yang tersedia di program studi
3. Memperoleh gambaran mengenai kualitas dokumen mutu yang dimiliki program studi.
4. Memanfaatkan teknik penambangan data untuk memperoleh informasi dari dokumen program studi.

### Akreditasi

Akreditasi adalah kegiatan penilaian untuk menentukan kelayakan program studi Akreditasi merupakan Sistem Penjaminan Mutu Eksternal sebagai bagian dari Sistem Penjaminan Mutu Pendidikan Tinggi yang memiliki prinsip yaitu *independent*, akurat, objektif, transparan, dan akuntabel dengan tujuan untuk menentukan kelayakan Program Studi berdasarkan kriteria yang mengacu pada Standar Nasional Pendidikan Tinggi serta menjamin mutu Program Studi secara eksternal baik bidang akademik maupun non akademik untuk melindungi kepentingan mahasiswa dan masyarakat.

Berdasarkan Peraturan Menteri RISTEKDIKTI Nomor 32 Tahun 2016, Badan Akreditasi Nasional Perguruan Tinggi (BAN – PT) dibentuk oleh pemerintah memiliki tugas untuk mengembangkan sistem akreditasi termasuk dalam hal melakukan penyusunan instrumen akreditasi. Terhitung 1 Januari 2019, BAN – PT akan efektif memberlakukan ketetapan instrumen akreditasi baru yaitu IAPS 4.0 yang terbagi menjadi 9 kriteria yaitu

1. Visi, Misi, Tujuan dan Strategi
2. Tata Pamong, Tata Kelola dan Kerjasama
3. Mahasiswa
4. Sumber Daya Manusia
5. Keuangan, Sarana dan Prasarana
6. Pendidikan
7. Penelitian
8. Pengabdian Kepada Masyarakat
9. Luaran dan Capaian Tridharma

### Web Content Mining

Penambangan data dengan sumber data yang berasal dari halaman web yang diekstraksi untuk menemukan pola yang berbeda yang memberikan wawasan yang signifikan. Web *Content Mining* menganalisis konteks web seperti text, data multimedia, dan data terstruktur yang bertujuan untuk memahami isi konten,

melakukan pengindeksan, relevansi halaman web dan mencari informasi yang lainnya yang berhubungan dengan pencarian dan analisis web. Data yang diproses pada kategori web *mining* ini diperoleh dari database yang mendasari web tersebut. Kategori web *mining* ini dapat digunakan untuk mengklasifikasikan dan mengelompokkan halaman web secara otomatis yang sesuai dengan jenisnya. Ada dua strategi yang umum digunakan pada Web Content Mining, pertama langsung melakukan mining terhadap data dan kedua melakukan pencarian serta meningkatkan hasil pencarian seperti search engine atau information retrieval (Baidowi, 2015). Tahapan yang dilakukan oleh web mining dibagi menjadi 3 (tiga) tahapan yaitu

- Tahap *Preprocessing* yang meliputi data *cleaning, Transaction Identification, Integration, Transformation*.
- Tahap *Process* yaitu tahapan penerapan formula statistic untuk mengurangi jumlah atribut dengan cara membuang atribut yang tidak diperlukan. Metode yang dapat dilakukan pada tahapan ini adalah *clustering, association, classification*.
- Tahapan terakhir adalah tahap *Post Processing* yang merupakan analisis lanjut terhadap data yang diproses untuk mengolah hasil penambangan yang sudah didapatkan tahap ini dilakukan apabila hasil data yang sudah diolah tidak bukan merupakan hasil yang dapat divisualisasikan.

## METODE

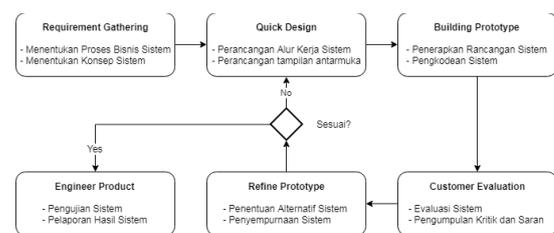
### Pengumpulan Data

Penelitian memiliki tujuan akhir yaitu membuat sebuah sistem web mining untuk dokumen mutu Program Studi. Untuk dapat mencapai tujuan akhir penelitian perlu melakukan pengumpulan data untuk mencari solusi sebagai bentuk penyelesaian masalah. Pengumpulan data diawali dengan melakukan observasi awal mengenai proses pengelolaan dokumen mutu yang sedang diterapkan saat ini. Pengumpulan data juga bertujuan untuk mengetahui permasalahan yang terjadi dan kelemahan yang terdapat pada mekanisme saat ini. Selain melakukan observasi, penelitian juga akan melakukan studi literatur untuk mencari teori yang berkaitan dengan topik permasalahan untuk mendukung jalannya penelitian. Data yang dikumpulkan dari kemudian dianalisis untuk

memberikan solusi yang akan untuk menyelesaikan masalah yang ditemui adalah merekayasa kembali sistem pengelolaan dokumen mutu sesuai dengan kriteria yang ditentukan BAN – PT.

### Pengembangan Sistem

Agar pengembangan dilakukan secara terarah akan diterapkan metode *Software Development Life Cycle* (SDLC) dengan model prototipe hal ini dirasa cukup tepat karena dengan adanya evaluasi dan *feedback* yang interaktif dapat memunculkan ide serta membantu pemilihan alternatif yang mungkin terjadi pada sistem.



Gambar 1. Metode Prototipe

Penerapan metode prototipe pada penelitian yang akan dilakukan akan dijelaskan kedalam tahapan berikut :

#### - Requirement Gathering

Aktivitas yang dilakukan pada tahapan ini yaitu melakukan identifikasi awal untuk mengumpulkan data dengan metode observasi dan dokumentasi yang mengamati secara langsung terhadap mekanisme yang diterapkan saat ini. Pengumpulan data yang dilakukan pada tahap ini akan menjelaskan permasalahan yang umumnya ditemui dan kekurangan yang dimiliki pada mekanisme yang sedang diterapkan saat ini sehingga data yang dikumpulkan dapat dianalisis untuk dijadikan sebagai dasar dari proses bisnis dan konseptual sistem sesuai dengan kebutuhan pengguna untuk menyempurnakan mekanisme yang diterapkan saat ini.

#### - Quick Design

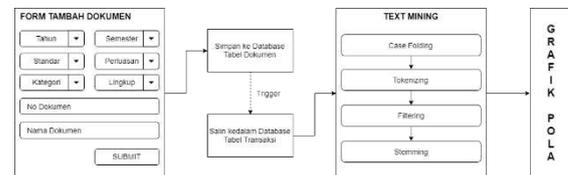
Setelah melakukan analisis, dilanjut dengan melakukan perancangan terhadap konseptual sistem baru meliputi alur kerja sistem dan desain tampilan antarmuka yang akan disajikan kepada pengguna.

- *Building Prototype*  
Hasil rancangan yang telah dibuat akan diimplementasikan kedalam bahasa pemrograman. Pada penelitian ini, sistem akan dibuat berbasis website sehingga bahasa pemrograman yang digunakan yaitu HTML, CSS, PHP, JavaScript dan MySQL.
- *Customer Evaluation*  
Setelah pemrograman selesai sistem akan direview dan dievaluasi oleh pengguna sehingga sistem akan mendapatkan *feedback* berupa kritik dan saran untuk penyempurnaan sistem dan agar sistem yang dibuat lebih sesuai dengan kebutuhan penggunanya
- *Refine Prototype*  
Hasil kritik dan saran yang terima pada proses evaluasi akan dianalisis untuk melihat kesesuaian konsep dengan sistem yang telah dibuat. Apabila sistem yang dibuat masih belum sesuai dan diperlukan adanya perubahan maka penelitian akan kembali ke tahapan *quick design* untuk melakukan perancangan alur sesuai dengan hasil evaluasi
- *Engineer Product*  
Tahapan terakhir dari metode prototipe yaitu penyerahan sistem kepada pengguna. Sebelum diserahkan kepada pengguna, sistem akan dilakukan pengujian terlebih dahulu untuk mengetahui kualitas sistem yang akan dibuat. Pengujian yang akan dilakukan berupa pengujian *black box* yang berfokus kepada fungsionalitas atribut sistem dan juga pengujian *white box* yang akan menguji implementasi struktur internal sistem. Sebelum melakukan pengujian, akan ditentukan skenario poin yang akan diuji.

## HASIL DAN PEMBAHASAN

Gambaran Objek yang diangkat pada penelitian ini adalah dokumen mutu Program Studi Informatika. Penilaian kualitas dokumen mutu dapat didasari dari sistem pengelolaan yang baik, sebaiknya sistem pengelolaan dokumen mutu diharapkan dapat menampilkan ragam informasi yang terdapat pada tumpukan dokumen mutu dalam data warehouse. Sedangkan gambaran subjek dari penelitian ini adalah penerapan sistem baru yang bernama sistem web mining akreditasi. Sistem web mining diterapkan untuk menggali pola – pola

informasi dokumen mutu Program Studi. Untuk dapat menggali pola informasi sistem akan menggunakan nama dokumen mutu sebagai data input, sistem menerapkan tahapan metode *text mining* untuk melakukan pembersihan data inputan yang digunakan pada sistem dan mengelompokkan data untuk mencari pola informasi.



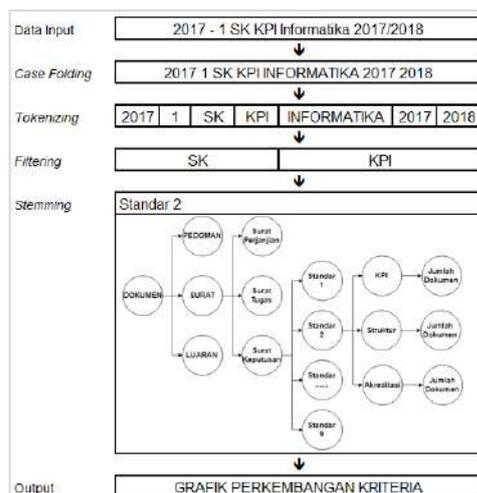
Gambar 2. Blok Diagram Konsep

Desain alur konsep yang digambarkan pada Gambar 2 menjelaskan kebutuhan data input yang digunakan oleh sistem berupa kumpulan nama dokumen mutu yang telah diinputkan pengguna kedalam sistem melalui form input. Namun, untuk melakukan proses *text mining* sistem membutuhkan data yang telah disesuaikan dengan kebutuhan. Untuk memenuhi kebutuhan data tersebut sistem akan membuat sebuah *trigger* pada database untuk menyimpan data input sesuai dengan kebutuhan.



Gambar 3. Trigger data Input

Data yang telah sesuai dengan kebutuhan akan diproses melalui beberapa tahapan *text mining*. *Text Mining* akan digunakan untuk membersihkan data sebelum ditampilkan menjadi grafik pola informasi dokumen mutu.



Gambar 4. Penerapan Konsep Mining

Gambar 4 menjelaskan konsep penerapan *text mining* yang akan dilakukan oleh sistem dalam memproses data yang digunakan pada sistem. Gambaran konsep tersebut lah yang akan diterapkan kedalam sistem dengan menggunakan bahasa pemrograman. Tahapan *text mining* dilakukan dengan mengekstraksi beberapa atribut yang dimiliki oleh dokumen mutu seperti identitas tahun, semester, dan nama dokumen

Tahapan pertama yang dilakukan adalah *case folding* yang bertujuan untuk menghilangkan karakter non huruf yang terdapat pada nama dokumen, saat melakukan case folding semua karakter non huruf (seperti / , . -) akan digantikan dengan spasi. Dalam melakukan pemrosesan data, sistem akan bersifat case sensitive atau membedakan huruf cetak (Uppercase Letter) dan huruf kecil (Lowercase Letter), maka pada proses case folding data input juga akan diubah kedalam bentuk huruf cetak.



Gambar 5. Implementasi Case Folding

Data yang sudah bersih dari simbol akan dirubah menjadi penggalan kata / token. Pemotongan data input memanfaatkan spasi yang dimiliki oleh data input sebagai penanda *delimiter* (pemisah). *Tokenizing* akan disimpan kedalam bentuk array, sehingga hasil proses dapat digambarkan pada gambar 6

Index	0	1	2	3	4	5	6	7
Value	2017	1	SK	KPI	INFORMATIKA	TA	2017	2018

Gambar 6. Implementasi Tokenizing

Tahapan selanjutnya yang dilakukan oleh *text mining* adalah *Filtering* akan mencari kata penting yang terdapat pada kumpulan token dengan membandingkan tiap token dengan *wordlist* atau token yang diinginkan. Pada saat melakukan *filtering* token yang sesuai dengan *wordlist* akan disimpan untuk digunakan pada tahapan selanjutnya dan token yang tidak sesuai akan dengan *wordlist* akan dihapus. Sistem web mining akan menggunakan 2 (dua) *wordlist* yaitu *wordlist* pertama berasal dari kriteria jenis dokumen dan *wordlist* kedua berasal dari kriteria perluasan standar sehingga token terakhir akan menentukan standar sebagai kelompok penyimpanan dokumen. Selain *wordlist* tersebut, terdapat beberapa kriteria perluasan standar yang memiliki sub perluasan tambahan, sehingga akan terdapat beberapa dokumen yang memiliki 3 (tiga) sisa token.

Index	0	1
Value	SK	KPI

Gambar 7. Implementasi Filtering

Sisa token yang dihasilkan pada *filtering* akan menentukan kelompok dokumen dengan cara mengambil id dari *wordlist* perluasan standar yang sesuai dengan token yang dibandingkan kemudian sistem akan menyimpan data kedalam database. token yang tersisa dari proses *filtering* juga akan digunakan untuk menjadi *Counter* yang akan menunjukkan jumlah dokumen mutu berdasarkan kriterianya. Proses ini dilakukan dengan membandingkan kriteria data yang akan disimpan dengan data yang berada pada tabel database, apabila terdapat data dengan kriteria yang sama maka *field* jumlah dokumen akan ditambahkan sedangkan jika tidak ada data pada database yang memiliki kriteria yang sama pada data yang akan disimpan maka data akan disimpan menjadi sebuah data baru pada tabel database. Hasil tabel ini lah yang akan dijadikan sebagai grafik pola informasi

standar	tahun	key1	key2	key3	sum
4	2018	SK	PEMBIMBING	KERJAPROFESI	2
2	2016	SK	STRUKTUR		1
9	2017	SK	BEASISWA	AKADEMIK	1
2	2017	SK	KPI		2
2	2017	SK	AKREDITASI		1

Gambar 8. Database hasil Text Mining

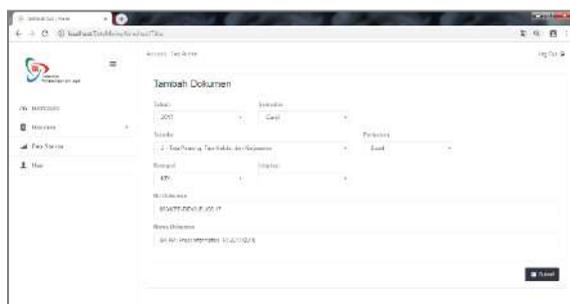
Untuk memastikan semua fungsi dan algoritma yang diimplementasikan ke dalam

sistem telah berjalan dengan baik, Penelitian akan melakukan pengujian terhadap sistem melalui 2 tahapan yaitu pengujian black box yang berfokus kepada eksekusi fungsionalitas dan white box yang berfokus kepada algoritma terapan. Kedua tahapan tersebut memiliki item pengujian yang sama, berikut ini merupakan hasil yang diperoleh dari pengujian black box.

Tabel 1. Hasil Pengujian Black Box

Item Pengujian	Kondisi Normal	Error	Hasil
Case Folding	Menghilangkan semua karakter non huruf yang terdapat pada data input	Karakter non huruf tidak dapat dihilangkan	✓
Tokenizing	Pemotongan data input menjadi beberapa token	Data tidak dapat dijadikan token	✓
Filtering	Pengambilan kata penting dari token, dengan melakukan perbandingan dengan word list	Terdapat token yang sama dengan word list tetapi tidak dapat diambil	✓
Stemming	Melakukan pencarian root dokumen dari token yang tersisa	Root dokumen tidak ditemukan	✓
Display Pattern	Menampilkan grafik pola informasi yang dihasilkan oleh proses web mining	Grafik tidak dapat ditampilkan	✓

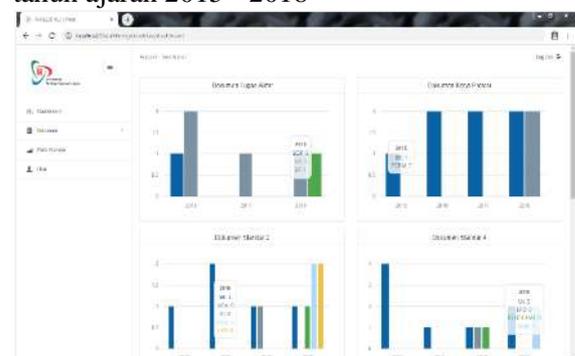
Tampilan interface yang disajikan sebagai bentuk interaksi antara sistem dengan pengguna akan dilampirkan selanjutnya. Form input merupakan interaksi pertama yang dilakukan oleh pengguna terhadap sistem untuk memberikan data yang akan diproses oleh sistem untuk dijadikan sebagai pola informasi.



Gambar 9. Form Input

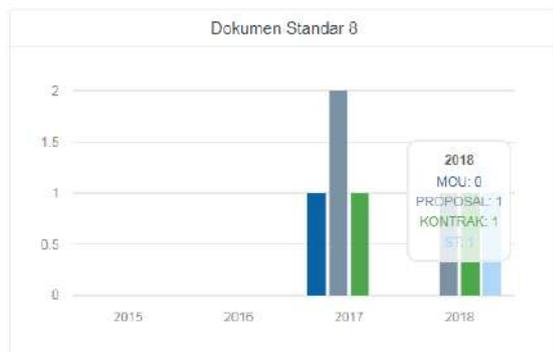
Tampilan dashboard akan berisikan pola – pola informasi yang diperoleh. Untuk mengetahui kualitas dokumen mutu, grafik pola

yang ditampilkan berdasarkan jangka waktu tahun yang telah ditentukan. Pada penelitian ini jangka yang ditentukan adalah dokumen mutu selama 4 tahun ke belakang yaitu dimulai dari tahun ajaran 2015 - 2018

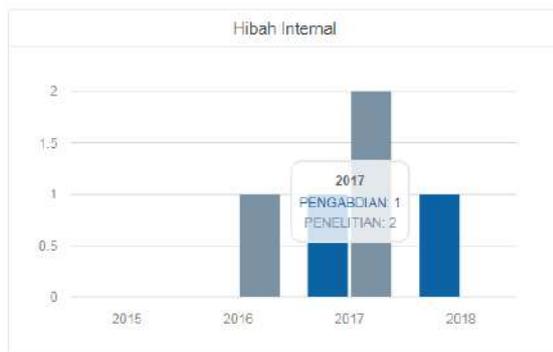


Gambar 10. Tampilan Dashboard

Pola grafik yang ditampilkan pada dashboard dapat berupa jenis dan jumlah dokumen berdasarkan tahun setiap standarnya atau berdasarkan kriteria yang diinginkan.



Gambar 11. Pola Grafik (1)



Gambar 12. Pola Grafik (2)

## SIMPULAN DAN SARAN

Web *content mining* dilakukan dengan menggunakan data yang berasal dari beberapa atribut yang dimiliki oleh dokumen mutu seperti identitas tahun, semester, dan nama dokumen. *Text Mining* sudah dapat berjalan dengan baik dalam proses penarikan *keyword* dari dokumen mutu untuk menentukan kriteria kelompok dokumen. Dari 65 dokumen mutu yang digunakan sebagai dokumen uji coba terdapat 6 dokumen yang tidak dapat dikelompokkan hal tersebut dikarenakan nama dari dokumen mutu kurang memberikan gambaran yang sesuai dengan kriteria kelompok. Grafik perkembangan dokumen mutu yang diperoleh yaitu jumlah jenis dokumen berdasarkan kriteria pertahun, dan jumlah jenis dokumen berdasarkan standar per tahun.

Pengembangan selanjutnya diharapkan dapat membuat kamus data yang mendetail agar penentuan *keyword* dokumen mutu dapat dilakukan dengan baik dan meminimalisir kesalahan pengelompokan dokumen. Menggabungkan hasil yang telah diproses menggunakan *text mining* dengan algoritma lain seperti C4.5 agar pengelompokan dapat dilakukan lebih akurat.

## UCAPAN TERIMAKASIH

Ucapan terimakasih disampaikan kepada semua civitas program studi Informatika UPJ yang telah membantu terutama kepada pembimbing penulis. Serta ucapan terimakasih kepada penyelenggara SEMNASTEK atas kesempatan yang diberikan kepada penulis untuk melakukan publikasi ini.

## DAFTAR PUSTAKA

- Baidowi. A.T., Widiastuti. N.I., 2015 Web Content Mining Menggunakan Partitional Clustering K-Means pada News Aggregator. Makalah sajian dalam Jurnal Sistem Komputer Vol. 5 No.2 ISSN : 2087 - 4685
- Han, Jiawei, Micheline Kamber, Jian Pei 2012, Data Mining Concepts and Techniques, Third Edition. Waltham : Morgan Kaufmann Publisher.
- Jaya, S, Handoko, P., Anwar, C., 2018 *Mekanisme Pengaturan Dokumen Program Studi Berdasarkan Standar BAN – PT Berbasis Online*. Makalah disajikan dalam SEMNASTEK 2018 – Fakultas Teknik Universitas Muhammadiyah Jakarta ISSN : 2407 - 1846
- Peraturan Menteri RISTEKDIKTI No 32, 2016.
- Vahrun, 2018. Press Release BAN-PT tanggal 26 Juli 2018. PELUNCURAN INSTRUMEN AKREDITASI PROGRAM STUDI BERBASIS OUTCOME (IAPS 4.0) DAN EKPOSE PENERAPAN TANDA TANGAN ELEKTRONIK (TT-e). Dari halaman web : <https://lldikti1.ristekdikti.go.id/details/apps/1190>