



PENERAPAN TEXT MINING UNTUK PENGELOMPOKAN PENELITIAN DOSEN

Eka Sabna

Teknik Informatika, STMIK Hang Tuah Pekanbaru

Email:

es3jelita@yahoo.com

Abstract

Research is one of the Tri Darma in Higher Education. Every lecturer is obliged to conduct research. Lecturers' research results in the Study Program are increasing and varying. This research was conducted to classify the titles of the lecturers' research using Text Mining. This grouping aims to determine the research topics of lecturers. From these results, it can help the Study Program and other researchers to see the research trends conducted by lecturers. In addition, the Study Program can use the results of this study to determine the supervisor and examiner lecturers of students taking the Thesis course. In addition, the results of this study can also help the lecturer in finding relationships / partners in the Lecturer further research.

Keywords: *Text Mining, Grouping, Lecturer Research.*

Abstrak

Penelitian merupakan salah satu dari Tri Darma Perguruan Tinggi. Setiap Dosen wajib melakukan penelitian. Hasil penelitian Dosen di Program Studi semakin bertambah dan beragam. Penelitian ini dilaksanakan untuk mengelompokkan judul-judul dari penelitian Dosen dengan menggunakan Text Mining. Pengelompokkan ini bertujuan untuk mengetahui Topik-topik penelitian Dosen. Dari hasil ini dapat membantu Program Studi dan Peneliti-peneliti yang lain untuk melihat kecenderungan (trend) penelitian yang dilakukan Dosen. Selain itu Program Studi dapat menggunakan hasil penelitian ini untuk menentukan Dosen Pembimbing dan Penguji dari mahasiswa yang mengambil matakuliah Skripsi. Selain itu hasil penelitian ini juga dapat membantu Dosen dalam mencari relasi/rekanan dalam penelitian Dosen tersebut selanjutnya.

Keywords: Text Mining, Pengelompokan, Penelitian Dosen.

PENDAHULUAN

Setiap Dosen Program Studi berkewajiban untuk melaksanakan penelitian. Dari waktu ke waktu jumlah penelitian Dosen semakin bertambah dan beragam. Penelitian Dosen yang semakin hari semakin banyak ini dapat dilakukan pengelompokan berdasarkan kemiripan dari judul-judul penelitian Dosen dengan menerapkan Text Mining. Dengan pengelompokan judul penelitian ini maka dapat membantu Program Studi untuk menentukan Dosen Pengampu Matakuliah yang sesuai, Dosen Pembimbing Skripsi Mahasiswa dan juga dapat membantu Dosen dalam mencari relasi/rekanan dalam penelitian Dosen tersebut selanjutnya.

Text mining merupakan pengembangan dari metode data mining yang dapat diterapkan untuk mengatasi permasalahan terkait dengan pengklasifikasian tema judul penelitian Dosen [1]. Algoritma-algoritma dalam text mining dibuat untuk dapat mengenali data yang sifatnya semi terstruktur seperti sinopsis, abstrak maupun isi dari dokumen-dokumen. Kategori teks atau klasifikasi teks adalah suatu proses yang mengelompokkan suatu teks kedalam suatu kategori tertentu [2].

Penelitian terkait dengan klasifikasi tema judul penelitian telah dilakukan oleh beberapa peneliti, diantaranya Prilianti & Wijaya (2014) meneliti mengenai pengembangan aplikasi berbasis text mining untuk automasi penentuan trend topik skripsi dengan metode K-Means Clustering [6]. Klasifikasi tugas akhir dan skripsi dengan algoritma K-Means clustering digunakan dalam proses penemuan pola terbukti dapat membantu proses pengelompokan berbagai topik skripsi yang ada sehingga diperoleh informasi yang bermakna dalam menentukan tren penelitian Universitas dari tahun ke tahun.

Dalam Penelitian ini data yang digunakan adalah data dari salah satu Program Studi yang ada di STIKes Hang Tuah Pekanbaru yaitu Program Studi Ilmu Kesehatan Masyarakat (IKM) tahun 2020. Data dari Program Studi ini yang dipilih adalah karena Program Studi (IKM) memiliki jumlah Dosen dan Mahasiswa yang paling banyak sehingga hasil penelitian dapat membantu Program Studi dalam menentukan Dosen Pengampu Matakuliah yang sesuai, Dosen Pembimbing Skripsi Mahasiswa dan juga dapat membantu Dosen dalam mencari relasi/rekanan dalam penelitian Dosen tersebut selanjutnya.

METODE PENELITIAN

Metode penelitian terdiri dari 3 tahap yaitu :

1. Metode Pengumpulan Data

Data diperoleh dari Data penelitian Dosen Program Studi IKM Tahun 2020. Data ini bentuk text.

2. Text Mining

Menurut Kurniawan, et al (2012), langkah-langkah yang dilakukan dalam text mining adalah sebagai berikut :

a. Text Preprocessing

Tindakan yang dilakukan pada tahap ini adalah :

1. To lower case, yaitu mengubah semua karakter huruf menjadi huruf kecil.
2. Tokenizing, yaitu proses penguraian deskripsi yang semula berupa kalimat kalimat menjadi kata kata.
3. Remove number, yaitu menghilangkan karakter angka pada kata tersebut.
4. Remove url, yaitu menghilangkan link internet.
5. Remove punctuation, yaitu menghilangkan delimiter delimiter seperti tanda titik (.), koma (,) dan spasi.

b. Feature Selection

Pada tahap ini tindakan yang dilakukan adalah :

1. Stopword adalah kosakata yang bukan merupakan ciri (kata unik) dari suatu dokumen.
2. Stemming adalah proses pemetaan dan penguraian berbagai bentuk (variants) dari suatu kata menjadi bentuk kata dasarnya (stem).

3. Analisa

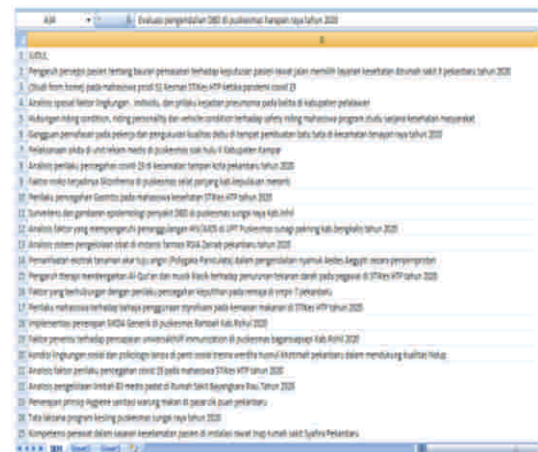
Analisa text mining dengan menggunakan algoritma K-means Clustering untuk mendapatkan pola dan untuk menentukan pengelompokan kemunculan kata di dalam text/dokumen.

4. Proses Text Mining ini menggunakan Aplikasi Rapidminer.

HASIL & PEMBAHASAN

1. Data

Data yang digunakan dalam penelitian ini adalah data penelitian Dosen Program Studi Ilmu Kesehatan Masyarakat (IKM) Tahun 2020.



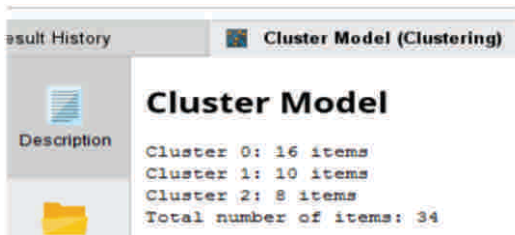
Gambar 1. Data Penelitian Dosen

2. Jalan kan aplikasi Rapidminer. Tahapan proses nya adalah sebagai berikut :

1. Import Data Excel.
2. Memasukan Nominal to Text ke proses Tokenizing
3. Memasukan Tools Process Document from Data (Text Processing)
4. Proses K-means clustering

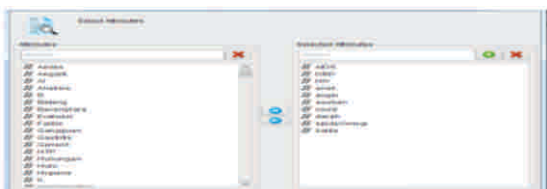


Gambar 2. Proses Text Mining dan Clustering

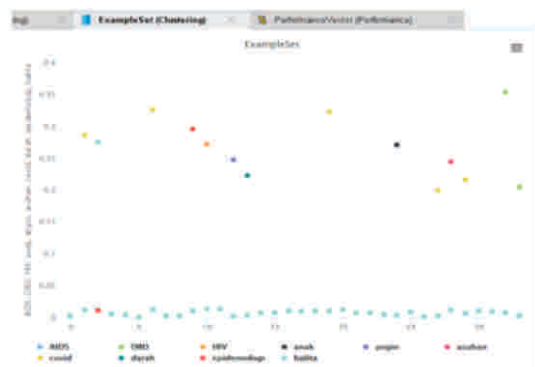


Gambar 3. Hasil dari Cluster Model

Gambar 4. Hasil Clustering setiap kata



Gambar 5. Kata (word) yang dipilih untuk dimunculkan pada Grafik



Gambar 6. Grafik dari kata-kata yang dipilih

Hasil dari proses diatas dapat dilihat pola dari kata-kata yang muncul dalam text/dokumen. Kata yang sering muncul (terbanyak) dari topik data penelitian Dosen adalah Covid dan HIV (2 kata yang tertinggi).

KESIMPULAN DAN SARAN

Kesimpulan dalam penelitian ini adalah :

1. Pengumpulan data dari Topik Penelitian Dosen di konversi kedalam bentuk teks.
2. Dari file text tersebut dapat menentukan pola dari kata-kata. Pola tersebut di ketahui dari jumlah munculnya kata-kata, dimana kata-kata yang sering muncul (2 urutan tertinggi) dari file

tersebut adalah kata Covid dan HIV. Dari hasil ini maka kita dapat mengetahui bahwa topik penelitian Dosen Program Studi IKM di tahun 2020 lebih banyak mengangkat Topik yang terkait dengan Covid dan HIV.

SARAN

Berdasarkan kesimpulan yang telah disampaikan dapat dikemukakan beberapa saran yang dapat dilakukan oleh penelitian selanjutnya adalah peneliti dapat menggunakan data dengan jumlah yang lebih banyak kemudian dilakukan proses menggunakan text mining untuk menganalisa pola dari hasil penelitian Dosen tersebut.

DAFTAR PUSTAKA

- Fitriastuti, F., Rahmalisa, U., & Girsang, A. S. (2019, March). Multi-Criteria Decision Making On Succesfull Of Online Learning Using Ahp And Regression. In *Journal Of Physics: Conference Series* (Vol. 1175, No. 1, P. 012071). Iop Publishing.
- Irawan, Y. (2018). Sistem Pendukung Keputusan Penentuan Penerimaan Beasiswa Berbasis Web Sma Islam Darul Huda Menggunakan Metode Analytical Hierarchy Process (Ahp). *Jurnal Ilmu Komputer*, 7(1), 1-6.
- Irawan, Y. (2019). Penerapan Data Mining Untuk Evaluasi Data Penjualan Menggunakan Metode Clustering Dan Algoritma Hirarki Divisive Di Perusahaan Media World Pekanbaru. *Jurnal Teknologi Informasi Universitas Lambung Mangkurat (JTIULM)*, 4(1), 13-20.
- J. Han, M.Kamber, And J.Pei. 2006. *Data Mining : Concept And Techniques*, Second Edition, 2nd Ed. Morgan Kaufmann California.
- Kurniawan, B., Effendi, S., And Sitompul, O,S. 2012. *Klasifikasi Konten Berita Dengan Metode Text Mining*. *Dunia Teknologi Informasi Vol.1, No.1*: Hal.14-19.
- Larose,D.T.2005. *Discovering Knowledge In Data : An Introduction To Data Mining*. Hoboken:Interscience, John Wiley & Sons, Inc.
- Mardeni, S. Penerapan Algoritma K-Nearest Neighbor Dalam Klasifikasi Data Hasil Produksi Kelapa Sawit Pada Kud Tirta Kencana.
- Santoso, B. 2007. *Data Mining : Teknik Pemanfaatan Data Untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Susanto,S., And Suryadi, D.2010. *Pengantar Data Mining*. Yogyakarta : C.V Andi Offset.
- Wahyuni, R., & Irawan, Y. (2019). Web-Based Heart Disease Diagnosis System With Forward Chaining Method (Case Study Of Ibnu Sina Islamic Hospital). *Journal Of Applied Engineering And Technological Science (Jaets)*, 1(1), 43-50.

- Wahyuningtyas, N., Zainul, R., Adri, M., Wedi, A., Surahman, E., Aisyah, E. N., ... & Santoso, Y. (2020, July). Development Of Moodle-Based Content Learning System In MkdK Student Development Subjects At Lptk In Indonesia. In *Journal Of Physics: Conference Series* (Vol. 1594, No. 1, P. 012021). Iop Publishing.
- Wahyuni, R., & Irawan, Y. (2019). Web-Based Heart Disease Diagnosis System With Forward Chaining Method (Case Study Of Ibnu Sina Islamic Hospital). *Journal of Applied Engineering and Technological Science (JAETS)*, 1(1), 43-50.