



PENGELOMPOKAN DOKUMEN PUBLIKASI ILMIAH BERDASARKAN BIDANG KEPAKARAN MENGGUNAKAN ALGORITMA DBSCAN

ARIQ RIZKI FADHILLAH LUBIS



**DEPARTEMEN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

PERNYATAAN MENGENAI LAPORAN AKHIR DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa laporan akhir dengan judul “Pengelompokan Dokumen Publikasi Ilmiah Berdasarkan Bidang Kepakaran Menggunakan Algoritma DBSCAN” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apapun kepada perguruan tinggi manapun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir laporan akhir ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Juni 2024

Ariq Rizki Fadhillah Lubis
G6401201008

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



ABSTRAK

ARIQ RIZKI FADHILLAH LUBIS. Pengelompokan Dokumen Publikasi Ilmiah Berdasarkan Bidang Kepakaran Menggunakan Algoritma DBSCAN. Dibimbing oleh SONY HARTONO WIJAYA dan KARLISA PRIANDANA.

Salah satu cara untuk memvalidasi keahlian para pakar adalah dengan dokumen-dokumen publikasi ilmiah yang dimiliki. Jika keparakan divalidasi secara manual akan membutuhkan waktu yang tidak singkat. Pengelompokan dokumen publikasi ilmiah menjadi salah satu alternatif untuk memverifikasi keparakan seorang dosen. Penelitian ini bertujuan untuk membangun suatu model yang dapat mengelompokkan dokumen berdasarkan bidang keparakan. Penelitian ini menggunakan algoritma *Latent Dirichlet Allocation* (LDA) untuk ekstraksi fitur dan *Density-Based Spatial Application with Noise* (DBSCAN) yang ditingkatkan dengan menambahkan metode *Uniform Manifold Approximation and Projection* (UMAP). Hasil penelitian ini berhasil mengelompokkan dokumen dengan nilai koefisien Silhouette sebesar 0,567. Model yang dibangun dievaluasi dengan mencocokkan klaster yang dihasilkan dengan klaim yang diberikan. Hasil yang diperoleh menunjukkan bahwa sebanyak 63,9% hasil pencocokan sesuai dengan klaim keparakan dosen, 27,6% tidak sesuai dengan klaim keparakan dosen, dan 8,4% tidak diketahui sesuai atau tidak dengan klaim keparakan.

Kata kunci: DBSCAN, dokumen, klasterisasi, pakar, validasi

ABSTRACT

ARIQ RIZKI FADHILLAH LUBIS. Clustering Scientific Publication Documents Based on Field of Expertise Using the DBSCAN Algorithm. Supervised by SONY HARTONO WIJAYA and KARLISA PRIANDANA.

One way to validate experts' expertise is through their scientific publications. If expertise is validated manually, it will take a long time. Grouping scientific publication documents is one alternative to verify the expertise of a lecturer. This research aims to build a model that can classify documents based on their field of expertise. This research uses *Latent Dirichlet Allocation* (LDA) algorithm to extract feature and *Density-Based Spatial Application with Noise* (DBSCAN) which is enhanced by adding *Uniform Manifold Approximation and Projection* (UMAP) method. The results of this study successfully clustered documents with a silhouette coefficient value of 0,567. The built model was evaluated by matching the resulting clusters with the given claims. The results showed that 63,9% of the matching results were in accordance with the lecturer's expertise claim, 27,6% were not in accordance with the lecturer's expertise claim and 8,4% were unknown whether or not it was in accordance with the expertise claim.

Keywords: clustering, DBSCAN, document, expert, validation

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.

2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
- b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2024
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

PENGELOMPOKAN DOKUMEN PUBLIKASI ILMIAH BERDASARKAN BIDANG KEPAKARAN MENGGUNAKAN ALGORITMA DBSCAN

ARIQ RIZKI FADHILLAH LUBIS

Skripsi
Sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Ilmu Komputer

**DEPARTEMEN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



Judul Skripsi : Pengelompokan Dokumen Publikasi Ilmiah Berdasarkan Bidang
Kepakaran Menggunakan Algoritma DBSCAN

Nama : Ariq Rizki Fadhillah Lubis
NIM : G6401201008

@Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:
Dr. Sony Hartono Wijaya, S.Kom., M.Kom.
19810809 200812 1 002



Pembimbing 2:
Dr. Karlisa Priandana, S.T., M.Eng
19851121 201212 2 002



Diketahui oleh

Ketua Departemen Ilmu Komputer:
Dr. Sony Hartono Wijaya, S.Kom., M.Kom.
19810809 200812 1 002



Tanggal Ujian:
28 Juni 2024

Tanggal Lulus:

PRAKATA

Puji dan syukur penulis ucapkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan, dengan judul “Pengelompokan Dokumen Publikasi Ilmiah Berdasarkan Bidang Kepakaran Menggunakan Algoritma DBSCAN”.

Terima kasih penulis ucapkan kepada para pembimbing, Dr. Sony Hartono Wijaya, S.Kom., M.Kom., dan Dr. Karlisa Priandana, S.T., M.Eng. yang telah membimbing dan banyak memberi saran. Ucapan terima kasih juga disampaikan kepada pembimbing akademik, moderator seminar, dan penguji luar komisi pembimbing. Ungkapan terima kasih juga disampaikan kepada ayah, ibu, serta seluruh keluarga yang telah memberikan dukungan, doa, dan kasih sayangnya. Ungkapan terima kasih juga disampaikan kepada teman-teman H7, MIT, Etherone 57, beserta orang-orang baik yang sudah menemani dan kebersamai selama kuliah di IPB.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmiah pengetahuan.

Bogor, Juni 2024

Ariq Rizki Fadhillah Lubis



DAFTAR ISI

DAFTAR TABEL	viii
DAFTAR GAMBAR	viii
DAFTAR LAMPIRAN	viii
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	2
1.4 Manfaat	2
1.5 Ruang Lingkup	2
II TINJAUAN PUSTAKA	4
2.1 <i>Latent Dirichlet Allocation</i>	4
2.2 <i>Density-Based Spatial Clustering Application with Noise</i>	4
2.3 <i>Uniform Manifold Approximation and Projection</i>	6
III METODE	7
3.1 Persiapan Data	7
3.1.1 Pengumpulan Data	8
3.1.2 <i>Text Preprocessing</i>	8
3.2 Ekstraksi Fitur	8
3.2.1 Vektorisasi Teks	8
3.2.2 <i>Latent Dirichlet Allocation</i>	9
3.3 Reduksi Dimensi	9
3.4 Klasterisasi	10
3.4.1 <i>Density-Based Spatial Clustering Application with Noise</i>	10
3.4.2 Ekstraksi Istilah	10
3.4.3 Pelabelan Klaster	10
3.4.4 Evaluasi	11
IV HASIL DAN PEMBAHASAN	12
4.1 Persiapan Data	12
4.2 Ekstraksi Fitur	14
4.3 Reduksi Dimensi dan Klasterisasi	15
V SIMPULAN DAN SARAN	21
5.1 Simpulan	21
5.2 Saran	21
DAFTAR PUSTAKA	22
LAMPIRAN	25

Hak Cipta milik IPB University

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

DAFTAR TABEL

1	Gambaran hasil dari vektorisasi teks dengan <i>bag-of-words</i>	8
2	Gambaran hasil keluaran dari LDA	9
3	Gambaran hasil keluaran dari UMAP	10
4	Contoh data karya ilmiah dosen IPB	12
5	Hasil salah satu penggalan kalimat setelah <i>text preprocessing</i>	13
6	Hasil <i>hyperparameter tuning</i> LDA	14
7	Hasil <i>hyperparameter tuning</i> UMAP dan DBSCAN	15
8	Hasil keluaran dari UMAP untuk 10 dokumen pertama	15
9	Jumlah dokumen per kluster	17
10	Hasil dari tahap pertama dan kedua ekstraksi istilah	17
11	Penomoran UDC untuk kata istilah kluster 0	18
12	Hasil pelabelan nama kluster berdasarkan pakar	19
13	Contoh hasil kluster yang sesuai dengan klaim dosen	19
14	Contoh hasil kluster yang tidak sesuai dengan klaim dosen	20
15	Contoh hasil kluster yang tidak diketahui, sesuai atau tidak	20
16	Hasil Pencocokan kluster dokumen dengan klaim dosen	20

DAFTAR GAMBAR

1	Proses <i>Latent Dirichlet Allocation</i>	4
2	Titik <i>border</i> , titik <i>core</i> , dan titik <i>outlier</i>	5
3	Tahapan penelitian	7
4	Persentase karakter yang dihapus	13
5	Tujuh <i>stopwords</i> terbanyak yang dihapus	14
6	Hasil keluaran proses LDA untuk dokumen pertama	15
7	Visualisasi hasil keluaran UMAP	16

DAFTAR LAMPIRAN

1	Contoh pencocokan hasil kluster dengan klaim kepakaran	26
2	Evaluasi klaim kepakaran dosen secara keseluruhan	27

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.