ISSN 2085-4552

Pembobotan Berdasarkan Tingkat Kesamaan Semantik pada Metode Fuzzy Semi-Supervised Co-Clustering untuk Pengelompokkan Dokumen Teks

Galang Amanda Dwi P., Gregorius Edwadr, Agus Zainal Arifin
Jurusan Teknik Informatika, Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia
galang.amanda13@mhs.if.its.ac.id, gregorius13@mhs.if.its.ac.id, agusza@cs.its.ac.id

Diterima 15 Juni 2014
Disetujui 18 November 2014

***Abstract*—Nowadays, a large number of information can not be reached by the reader because of the misclassification of text-based documents. The misclassified data can also make the readers obtain the wrong information. The method which is proposed by this paper is aiming to classify the documents into the correct group. Each document will have a membership value in several different classes. The method will be used to find the degree of similarity between the two documents is the semantic similarity. In fact, there is no document that doesn't have a relationship with the other but their relationship might be close to 0. This method calculates the similarity between two documents by taking into account the level of similarity of words and their synonyms. After all inter-document similarity values obtained, a matrix will be created. The matrix is then used as a semi-supervised factor. The output of this method is the value of the membership of each document, which must be one of the greatest membership value for each document which indicates where the documents are grouped. Classification result computed by the method shows a good value which is 90 %.**

***Index Terms*—Fuzzy co-clustering, Heuristic, Semantic Similarity, Semi-supervised learning**

I. PENDAHULUAN

Seiring dengan berjalannya waktu, metode *clustering* menjadi salah satu metode yang sangat penting dan sangat banyak diaplikasikan di kehidupan nyata terutama dalam kasus pengelompokan dokumen. Pengelompokan dokumen pada umumnya didefinisikan sebagai proses pengelompokan berbagai dokumen ke dalam berbagai kelas sesuai dengan topik dan kesamaan isi dari masing- masing dokumen tersebut. Terkelompoknya dokumen-dokumen tersebut kedalam kelasnya yang benar akan membantu memberikan manfaat yang baik bagi siapapun yang hendak mencari satu atau lebih dokumen sesuai dengan kelasnya.

Banyak sekali informasi-informasi yang sampai ke pengguna berisi informasi yang salah karena kesalahan dalam mengelompokan dokumen. Akibat dari salahnya informasi-informasi ini, pengguna akhirnya kehilangan pengetahuan yang semestinya bisa dia dapatkan dan dokumen-dokumen yang sudah dikelompokan menjadi tidak ada gunanya karena tidak dapat menampilkan informasi penting saat dibutuhkan.

Banyak algoritma pengelompokan yang telah dikembangkan selama ini, hanya saja dibutuhkan modifikasi khusus untuk algoritma pengelompokan dokumen teks. Algoritma tersebut harus dapat mengelompokan satu dokumen ke dalam banyak kelas sehingga algoritma-algoritma tertentu tidak dapat digunakan. Tentu saja sudah ada beberapa algoritma yang dapat melakukan pengelompokan sesuai dengan kategori diatas namun metode-metode pengelompokan tersebut tidak didasarkan pada kemiripan konten yang dimiliki antar dokumen.

Seperti yang diketahui, dokumen-dokumen yang memiliki banyak kata-kata yang serupa akan cenderung berada di dalam kelas yang sama. Meninjau dari hal tersebut, penelitian kali ini menawarkan cara pengelompokan dokumen dengan diawal proses diasumsikan bahwa dokumen-dokumen yang mempunyai tingkat kemiripan kata yang tinggi akan cenderung berada dalam kelas yang sama.

Dengan menggunakan metode *heuristic semi-supervised fuzzy co-clustering* (SS-HFCR) [1], pengelompokkan dokumen dapat dilakukan secara efektif. Hal ini dikarenakan metode tersebut menggunakan *prior knowledge* untuk menentukan apakah sebuah dokumen memiliki hubungan dengan dokumen yang lain atau tidak, yaitu dengan cara memberikan *constraint* "*must link*" atau "*cannot*

link". Namun demikian, proses penentuan *constraint* dilakukan secara subjektif karena ditentukan secara manual oleh pengguna sehingga dapat memberikan hasil yang kurang akurat.

Dalam penelitian kali ini, akan dirancang sebuah metode yang lebih objektif dalam hal pengelompokkan dokumen. Metode yang diusulkan diusulkan akan menggunaan metrik *semantic similarity* [2] dalam menentukan tingkat hubungan yang dimiliki antar dokumen untuk mengganti proses penentuan *constraint* "*cannot link*" dan "*must link*". Tujuannya adalah untuk mengeliminasi segala kemungkinan subjektivitas dalam pengelompokkan dokumen.

Secara umum, artikel ditulis sebagai berikut. Pekerjaan terkait pengelompokkan dokumen dengan algoritma SS-HFCR dan *semantic similarity* akan diulas pada subbab Pekerjaan Terkait. Metode yang diusulkan akan dijelaskan pada subbab Metodologi. Pada subbab Implementasi, akan diilustrasikan kerja algoritma yang diusulkan pada contoh studi kasus.

II. PEKERJAAN TERKAIT

Akan dikelompokkan penelitian-penelitian terkait yang relevan dari metode yang diusulkan ke dalam beberapa kategori, yaitu: *fuzzy co-clustering*, *semi-supervised fuzzy clustering*, *semi-supervised co-clustering*, dan penggunaan metode semantic similarity untuk menghitung nilai kemiripan antar label. Pada bagian ini, penelitian yang sudah ada akan diulas satu persatu. Penelitian terbaru yang menggunakan metode pendekatan *dual-partitioning based fuzzy co-clustering* (HFCR) [3] telah berhasil dirumuskan. Tercatat beberapa pendekatan metode pengelompokan telah mengekslorasi berbagai model *prior knowledge* yang diolah menjadi model *fuzzy clustering*.

Selanjutnya, *dual-partitioning based fuzzy co-clustering* (HFCR) menjelaskan bahwa sebuah mekanisme seleksi aktif untuk memilih syarat yang sesuai dengan tujuan untuk mengurangi akibat kepada performa yang ditimbulkan dari proses seleksi yang terjadi dalam *heuristic semi-supervised fuzzy co-clustering* (SS-HFCR) [4].

Yang Yan [1] menggunakan metode metode *heuristic semi-supervised fuzzy co-clustering* (SS-HFCR). Dalam percobaanya, pengelompokkan dokumen dapat dilakukan secara lebih efektif. Penggunaan *prior knowledge* untuk menentukan apakah sebuah dokumen memiliki hubungan "must link" atau "*cannot link*". Namun begitu, penentuan hubungan "*must link*" atau "*cannot link*" dilakukan secara subjektif karena ditentukan secara manual.

Dalam penelitian yang dilakukan oleh Remco Dijkman [2], digunakan tiga macam metode yang berbeda dalam menentukan tingkat similaritas antar label/kalimat untuk mengelompokkan proses bisnis dalam sebuah repositori. Pengelompokkan berdasarkan tingkat kemiripan bertujuan untuk dokumentasi dan kemudian pencarian informasi pada repositori proses bisnis. Tiga macam metode yang digunakan adalah *syntactic similarity*, *semantic similarity*, dan *contextual similarity*. Dalam penelitian tersebut, metode perhitungan kemiripan dengan metrik *semantic similarity* memiliki nilai presisi yang relatif lebih baik dari dua yang lain.

Fokus dari penelitian ini adalah perancangan sebuah metode yang efektif dan juga objektif dalam hal pengelompokkan dokumen berdasarkan metode yang sudah ada yaitu SS-HFRC dan metrik *semantic similarity*.

III. METODOLOGI

Pada bagian berikut, akan dijelaskan metode yang disulkan yaitu algoritma pembobotan dengan menggunakan metode *semantic similarity* dan SS-HFRC.

A. Semantic similarity metric

Dalam membandingkan kedua kalimat, penting untuk melihat tingkat kesetaraan antar kata-katanya, tidak hanya mengasumsikan bahwa kata itu benar-benar sama secara tulisan. Bisa saja kata yang berbeda memiliki kemiripan arti karena kedua kata tersebut merupakan sinonim. Contohnya terdapat dua buah kalimat yaitu: "*Customer inquiry processing*" dan "*Client inquiry query processing*" memiliki arti yang mirip walaupun memiliki perbedaan pada kata-kata penyusunnya.

Oleh karena itu, sebagai dasar untuk menghitung kemiripan antar kedua kalimat, kemiripan antar kedua buah elemen tersebut harus dapat diukur. Akan dipertimbangkan metode *semantic similarity*, dimana tidak hanya diukur kemiripan dua buah elemen dari kata-kata penyusunnya, namun juga akan dipertimbangkan arti dari kata-kata yang terdapat dalam kalimat tersebut.

Diberikan dua buah label (kalimat), nilai *semantic similarity* dari keduanya adalah tingkat kemiripan, berdasarkan kesetaraan dari kata-kata yang terdapat dalam label masing-masing. Diasumsikan sebuah kata yang sama lebih dipilih dari sinonim. Dengan demikian, kata-kata yang identik akan diberi nilai 1, sedangkan kata-kata sinonim diberikan nilai lebih rendah dari 1. Dengan begitu, rumus *semantic similarity metric* dapat didefinisikan sebagai berikut. Ketika menentukan kesetaraan antar kata, simbol-simbol khusus akan diabaikan dan semua karakter akan diubah menjadi huruf kecil. Nilai kemiripan dari kedua buah kalimat yang dibandingkan akan dihitung dengan menggunakan nilai pembobotan

sinonim (0, 0.25, 0.5, 0.75, dan 1). Dalam penelitian ini, digunakan nilai 0.75 sebagai pembobotan kata yang bersinonim karena menghasilkan nilai akurasi paling tinggi yaitu 90% dalam penelitian-penelitian sebelumnya. Rumus *semantic similarity* didefinisikan seperti pada Persamaan 1.

$$sem(f_1, f_2) = \frac{\alpha}{\max(|f_1|,|f_2|)} \quad (1)$$

$$\alpha = |f_1 \cap f_2| + 0.75 \times \sum_{(s,l)\in f_1 \setminus f_2 \times f_2 \setminus f_1} synonym(s,l) \quad (2)$$

f_1, f_2 masing-masing adalah kalimat yang akan dibandingkan, kemudian $f_1 \cap f_2$ adalah jumlah kata yang sama dari kedua kalimat. Synonym (s, l) adalah kata yang memiliki kemiripan arti/sinonim. $\max(|f_1|,|f_2|)$ adalah jumlah kata terbanyak dari kedua kalimat yang dibandingkan.

Tabel 1. Daftar Simbol

Notasi	Penjelasan
D	Asosiasi matriks dalam dokumen(tf-idf)
C/c	Indeks kelompok/jumlah kelompok
M	Jumlah kata
N	Jumlah dokumen
x_i	Dokumen ke i
d_{ij}	Tf-idf dari kata ke-j dalam dokumen ke-i
U,V	Matriks dokumen dan atribut
u_{ci} , v_{cj}	Nilai derajat keanggotaan dokumen dan kata
T_u , T_v	Derajat keanggotaan fuzzy yang ditentukan pengguna
W	Data pembelajaran dimana mengandung semua nilai kesamaan setiap dokumen.
T_d	Bobot faktor dari konstrain

B. Perumusan

Kelompok-kelompok dokumen akan direprentasikan menjadi vektor. Misalkan D *dataset* dari N objek (dokumen) yang diambil dari fitur M-dimensi (kata) ruang, maka tujuan dari *clustering* adalah mengelompokan setiap dokumen pada dokumen secara benar kedalam C kelompok. Pekerjaan [1] sebelumnya telah membuktikan bahwa SS-HFCR memiliki hasil yang baik namun matriks *must link* dan *cannot link* (ml/cnl) harus diinisiasi secara manual. Untuk meninggkatkan objektivitas, maka ditawarkanlah sebuah metode baru dimana dalam metode ini menggunakan matriks W untuk menggantikan matriks ml/cnl. Jika pada matriks ml/cnl, hanya bisa terdapat nilai biner, dimana nilai 1 melambangkan bahwa kedua dokumen terkait harus berada didalam kelas yang sama dan 0 berarti kedua dokumen terkait harus berada dikelas yang berbeda, pada matriks W berisisi sebuah nilai kontinu yang berada dalam rentang 0 sampai 1 yang melambangkan tingkat kemiripan antara dua dokumen. Jika suatu nilai pada matriks W mendekati 1 maka hal itu berarti ada kesamaan yang tinggi diantara kedua dokumen terkait, jika semakin mendekati 0 maka hal ini melambangkan sebaliknya. *Objective function* dari metode ini dirumuskan seperti pada Persamaan 3.

$$F = \sum_{c=1}^{C}\sum_{i=1}^{N}\sum_{j=1}^{K} u_{ci}u_{cj}d_{ij} - T_u\sum_{c=1}^{C}\sum_{i=1}^{N} u_{ci} ln\, u_{ci} - T_v\sum_{c=1}^{C}\sum_{j=1}^{K} v_{cj} ln\, v_{cj} - T_u \times T_v \times \left(\sum_{i=1}^{C}\sum_{j=1}^{C} u_{ci}u_{cj}W_{ij}\right) \quad (3)$$

Tabel 2. Langkah-langkah Pengerjaan

Input : *Dataset* D, Jumlah kelas C
Output : Matrix U and V
Metode : Memberi bobot $Tu, Tv, Td, \tau maks$, dan batas kesalahan (ε) Inisiasi matriks W Ulangi Perbaharui v_{cj} Dengan Persamaan 5 Perbaharui u_{ci} Dengan Persmaan 4 $\tau = \tau + 1$ Sampai $(max_c \lvert u_{ci}^{t+1} - u_{ci}^{t} \rvert \leq \varepsilon)$ atau $\tau > \tau_{maks}$

Dapat dilihat bahwa ada perbedaan pada *objective function* yang ditawarkan pada penelitian ini dengan *objective function* yang berada pada penelitian sebelumnya. Perubahan yang terjadi adalah adanya penggunaan matriks W menggantikan matriks ml/cnl. Hal yang sama juga diterapkan pada fungsi perbaharuan U. Tidak ada perubahan pada fungsi pembaruan matriks V, karena matriks V hanya mereprentasikan kata-kata. Fungsi pembaruan pada matriks U dan V dapat dilihat pada persamaan 4 dan 5.

Daftar simbol yang digunakan telah dirangkum pada Tabel 1. Simbol yang ditulis dengan huruf besar dan ditebalkan melambangkan matriks, sedangkan simbol yang ditulis dengan huruf besar dan miring berarti skalar.

C. Algoritma dan kompleksitas

Alur dari SS-HFCR dimulai dengan matriks U dan V dimana matriks U adalah matriks dokumen dan matriks V adalah matriks atribut. Saat dimulai matriks U tidak berisi angka negatif. Kedua matriks tersebut kemudian diperbarahui dengan

$$u_{ci} = \frac{exp\left\{\frac{\sum_{j=1}^{K} v_{cj} d_{ij}}{\tau_u \sum_{j=1}^{K} v_{cj}} + \tau_d\left[\sum_{i=1}^{c}\sum_{j=1}^{c} u_{cs} W_{ij}\right]\right\}}{\sum_{f=1}^{c} exp\left\{\frac{\sum_{j=1}^{K} v_{fj} d_{ij}}{\tau_u \sum_{j=1}^{K} v_{fj}} + \tau_d\left[\sum_{i=1}^{c}\sum_{j=1}^{c} u_{fs} W_{ij}\right]\right\}} \quad (4)$$

$$v_{cj} = \frac{exp\left\{\frac{\sum_{i=1}^{N} u_{ci} d_{ij}}{\tau_v \sum_{i=1}^{N} u_{ci}}\right\}}{\sum_{q=1}^{K} exp\left\{\frac{\sum_{i=1}^{N} u_{ci} d_{iq}}{\tau_v \sum_{i=1}^{N} u_{ci}}\right\}} \quad (5)$$

Persamaan 4 dan Persamaan 5 sampai dengan jumlah perulangan yang telah ditentukan. Langkah pengerjaan SS- HCFR dapat dilihat pada Tabel 2. Kompleksitas dari SS-HFCR adalah $(CNM\tau)$ dimana τ adalah jumlah iterasi. Kompleksitas dari metode ini sama dengan HFCR.

Tabel 3. Matrix TF-IDF dan Similaritas

Atribut	Similaritas	Hasil
$\begin{pmatrix} 1.0 & 0.8 & 0.3 & 0.5 & 0.6 & 0.0 & 0.4 \\ 1.0 & 0.7 & 0.4 & 0.5 & 0.5 & 0.1 & 0.4 \\ 0.1 & 0.5 & 1.0 & 0.7 & 0.2 & 0.4 & 0.3 \\ 0.1 & 0.6 & 0.9 & 0.6 & 0.1 & 0.6 & 0.2 \\ 0.2 & 0.7 & 0.9 & 0.8 & 0.3 & 0.5 & 0.2 \\ 0.1 & 0.3 & 0.7 & 0.9 & 0.5 & 1.0 & 0.8 \end{pmatrix}$	$\begin{pmatrix} 1.0 & 0.8 & 0.4 & 0.3 & 0.3 & 0.5 \\ 0.8 & 1.0 & 0.3 & 0.2 & 0.2 & 0.4 \\ 0.4 & 0.3 & 1.0 & 0.7 & 0.9 & 0.6 \\ 0.3 & 0.2 & 0.7 & 1.0 & 0.8 & 0.5 \\ 0.3 & 0.2 & 0.9 & 0.8 & 1.0 & 0.5 \\ 0.5 & 0.4 & 0.6 & 0.5 & 0.5 & 1.0 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

IV. MENGGUNAKAN TEMPLATE

Pada bagian ini metode yang diusulkan akan diujicoba dalam dua kali percobaan. Percobaan pertama yaitu pengujian terhadap cara kerja metode. Pengujian dilakukan dengan *dataset* yang sederhana. Percobaan kedua adalah pengujian dengan menggunakan *dataset* yang besar dan kompleks dengan tujuan untuk mendapatkan akurasi dari metode yang diusulkan.

A. Pengujian metode

Pada bagian ini akan digunakan sebuah *dataset* yang digunakan untuk memperlihatkan tingkat kinerja dari metode yang ditawarkan, yaitu Weighted SS-HFCR. Pada Tabel 3 dapat dilihat matriks atribut, matriks similaritas, dan matriks hasil aktualnya. *Dataset* yang digunakan dalam percobaan diatas berisi 6 buah dokumen dengan 7 buah atribut yang dapat diekstraksi. Kolom Atribut dalam Tabel 3 mengambarkan nilai ketujuh buah atribut terhadap masing-masing dokumen. Kolom Similaritas menggambarkan tinggkat kesamaan antara masing-masing dokumen.

Dapat dilihat bahwa dokumen yang berada dalam kelas yang sama cenderung mempunyai tingkat kesamaan yang tinggi. Kolom Hasil berisi hasil pengelompokan dokumen secara benar yaitu dokumen satu dan dua berada pada kelompok satu, dokumen tiga, empat, dan lima berada pada kelompok dua dan dokumen ke enam berada pada kelompok tiga.

Hasil pada percobaan diatas menunjukan bahwa dokumen-dokumen diatas telah terbagi kedalam kelompoknya masing setelah iterasi ke 100. Iterasi proses dapat dilihat pada Tabel 4. Pada iterasi-iterasi awal dokumen satu dan dua telah berhasil dikelompokan dengan baik, dokumen tiga, empat dan lima juga sudah dapat mengelompok dengan agak baik, namun terjadi sedikit masalah pada dokumen keenam.

Hingga pada iterasi ke-100 tingkat kesalahan pada dokumen keenam masih sekitar 7%. Tingkat keimiripan dokumen keenam yang cukup tinggi dengan dokumen-dokumen lainnya mungkin juga ikut berperan dalam menyebabkan susahnya dokumen keenam mencapai tingkat akurasi yang tinggi. Jika dibandingkan dengan dokumen satu dan dua yang tidak mempunyai tingkat kesamaan dengan dokumen dari kelas lainnya, tingkat kesamaan dapat disimpulkan mempunyai peran yang besar.

Tabel 4. Iterasi Proses

Iterasi ke	Hasil
Iterasi ke 5	$\begin{pmatrix} 0.90 & 0.08 & 0.02 \\ 0.92 & 0.07 & 0.01 \\ 0.01 & 0.89 & 0.10 \\ 0.03 & 0.91 & 0.06 \\ 0.01 & 0.85 & 0.14 \\ 0.12 & 0.09 & 0.79 \end{pmatrix}$

Tabel 4. Iterasi Proses (lanjutan)

Iterasi ke 10	$\begin{pmatrix} 0.95 & 0.05 & 0 \\ 0.94 & 0.05 & 0.01 \\ 0.01 & 0.92 & 0.07 \\ 0.02 & 0.94 & 0.04 \\ 0.02 & 0.88 & 0.10 \\ 0.10 & 0.07 & 0.83 \end{pmatrix}$
Iterasi ke 20	$\begin{pmatrix} 0.97 & 0.03 & 0 \\ 0.98 & 0.01 & 0.01 \\ 0 & 0.93 & 0.07 \\ 0.02 & 0.94 & 0.04 \\ 0.01 & 0.90 & 0.09 \\ 0.09 & 0.06 & 0.85 \end{pmatrix}$
Iterasi ke 100	$\begin{pmatrix} 0.99 & 0.01 & 0 \\ 1 & 0 & 0 \\ 0 & 0.98 & 0.02 \\ 0 & 0.99 & 0.02 \\ 0.01 & 0.98 & 0.01 \\ 0.05 & 0.02 & 0.93 \end{pmatrix}$

B. Pengujian dengan dataset Iris, Reuters, dan WebKB

Pada bagian ini, uji coba akan dilaku kan dengan menggunakan tiga buah *dataset* besar. *Dataset* yang pertama adalah data Iris, data yang kedua adalah Reuters-21578 R8, dan data yang ketiga adalah data WebKB. Data iris berjumlah 150. Masing-masing kelas berjumlah 50. Jumlah kelas dalam data ini adalah tiga. Rincian dari masing-masing *dataset* dapat dilihat pada Tabel 5, Tabel 6 dan Tabel 7.

Tabel 5. Dataset Iris

Kelas	Jumlah Data
Iris-Setosa	50
Iris-Versicolor	50
Iris-Virginica	50
Jumlah	**150**

Tabel 6. Dataset Reuters-21578 R8

Kelas	Jumlah Data
Earn	50
Mobey-fx	50
Trade	50
Interest	50
Crude	50
Ship	50
grain	50
acq	50
Jumlah	**400**

Tabel 7. Dataset WebKB

Kelas	Jumlah Data
Project	100
Course	100
Faculty	100
Student	100
Jumlah	**400**

Akurasi terbaik dari hasil percobaan pada tiga buah *dataset* yang telah dipilih dapat dilihat pada Tabel 8.

V. KESIMPULAN

Dalam penelitian ini, diusulkan sebuah metode yang efektif dan juga objektif dalam hal pengelompokkan dokumen berdasarkan metode yang sudah ada yaitu SS-HFRC dan metrik *semantic similarity*. Tujuannya adalah untuk mengeliminasi segala kemungkinan subjektivitas dalam pengelompokkan dokumen.

Meskipun metode yang ditawarkan mempunyai akurasi yang cukup tinggi namun butuh waktu yang sangat lama dalam melakukan komputasi. *Dataset* yang telah dikumpulkan berjumlah sangat besar namun hanya sebagian kecil yang dapat digunakan karena membutuhkan waktu komputasi yang sangat lama. Untuk mencari tingkat kesamaan antar dokumen kompleksitas komputasinya akan sangat tinggi karena harus mencari tingkat kesamaan untuk setiap dokumen ($O(n^2)$) dan setiap iterasi harus membandingkan setiap kata dalam setiap dokumen. Jika jumlah dokumen yang diuji terlalu besar maka percobaan akan memakan waktu yang sangat lama dan mempunyai kemungkinan tidak selesai karena komputer yang digunakan tidak mampu untuk

melakukan komputasi.

Setelah selesai mendapatkan matriks similaritas dan matriks tfidf, perhitungan setiap iterasi juga masih memakan waktu komputasi yang banyak karena setiap iterasi membutuhkan beberapa informasi dari matriks *u* dan *v* seperti pada Persamaan 4 dan Persamaan 5. Informasi dari masing-masing matriks tersebut tidak dapat disimpan karena terus berubah dalam setiap iterasinya sehingga setiap iterasinya membutuhkan waktu yang sangat banyak.

Hasil yang didapat dari metode cenderung stabil karena walaupun parameternya diubah-ubah hasilnya tidak menunjukan perubahan yang terlalu signifikan. Setelah melewati iterasi-iterasi tertentu, umumnya hasil suatu percobaan menjadi hampir sama dengan percobaan lainnya yang memiliki parameter-parameter yang berbeda.

Pada penelitian selanjutnya, fokus penelitian adalah bagaimana untuk mengoptimasi metode yang saat ini telah diusulkan mengingat kompleksitasnya sangat tinggi seperti yang telah dijelaskan pada subbab Pembahasan.

Tabel 8. Hasil Uji Coba

Data	Tu(10^{-4})	Tv(10^{-4})	Td(10^{-5})	Akurasi
Iris	2	1	10	0,9
Reuters-21578 R8	5	2	2	0,8
WebKb	2	10	10	0,8

Ucapan Terima Kasih

Penulis mengucapkan terimakasih kepada dosen pembimbing yaitu Bapak Dr. H. Agus Zainal Arifin, S.Kom., M.Kom. yang telah membantu dan membimbing kami dalam mengerjakan penelitian ini. Penulis juga mengucapkan terimakasih kepada Kemendikbud RI yang telah memberikan beasiswa pada penulis sehingga penulis dapat melanjutkan pendidikan Double Degree Master di Prancis. Selain itu, penulis juga mengucapkan terimakasih kepada keluarga penulis dan teman-teman penulis yang juga sudah memberi kontribusi baik langsung maupun tidak langsung dalam penelitian ini.

Tentunya ada hal-hal yang ingin penulis berikan kepada masyarakat dari hasil penelitian ini. Karena itu penulis berharap semoga penelitian ini dapat menjadi sesuatu yang berguna bagi semua pihak.

Penulis menyadari bahwa dalam menyusun artikel ini masih jauh dari kesempurnaan, untuk itu penulis sangat mengharapkan kritik dan saran yang bersifat membangun guna sempurnanya artikel ini. Penulis berharap semoga penulis ini bisa bermanfaat bagi penulis khususnya dan bagi pembaca pada umumnya.

Daftar Pustaka

[1] L. C. Yang Yan, "Fuzzy semi-supervised co-clustering for text documents," *Fuzzy Sets and System,* vol. 2015, pp. 75-79, 2013.

[2] R. Dijkman, "Similarity of Business Process Models: Metrics and Evaluation".

[3] N.Grira, "Active semi-supervised fuzzy clustering," in *Pattern Recognition*, 2008.

[4] N.Grira, "Semi-supervised fuzzy clustering with pair-wise-constrained competitive agglomeration," in *International Conference on Fuzzy Systems*, 2005.

[5] L.Chen, "A heuristic-based fuzzy co-clustering algorithm for categorization of high-dimensional data," *Fuzzy Sets and System,* vol. 159, 2008.