

ANALISIS PERFORMANSI *UNSUPERVISED LEARNING* UNTUK VIDEO INSTRUKSI BERNARASI MENGGUNAKAN KLASTERISASI TEKS

PERFORMANCE ANALYSIS OF *UNSUPERVISED LEARNING* FOR NARRATED INSTRUCTION VIDEOS USING TEXT CLUSTERING

Ricky Hilmi Sudrajad¹, Suryo Adhi Wibowo, S.T., M.T., Ph.D.², Iwan Iwut Tritoasmoro, S.T., M.T.³

^{1,2,3}Prodi S1 Teknik Telekomunikasi, Fakultas Teknik Elektro, Universitas Telkom

¹rickyhilmi@student.telkomuniversity.ac.id, ²suryoadhiwibowo@telkomuniversity.ac.id,

³iwaniwut@telkomuniversity.ac.id

Abstrak

Video instruksi bernarasi merupakan salah satu jenis video yang sering digunakan sebagai media pembelajaran. Video ini mengandung representasi visual yang umumnya diperagakan oleh manusia dan representasi tekstual dengan adanya narasi berupa *subtitle*. Teknologi *machine learning* dapat membangun sistem *unsupervised learning* untuk meringkas video instruksi bernarasi menjadi hanya daftar langkah utama (*key steps*). Sistem tersebut menggunakan klasterisasi teks dengan metode *Multiple Sequence Alignment* (MSA) untuk menyejajarkan langkah-langkah utama yang telah ditemukan. Akan tetapi, sistem menghasilkan performansi yang tidak konsisten di setiap skenario video walaupun memiliki jumlah *Ground Truth* yang sama.

Berdasarkan permasalahan tersebut, pada Tugas Akhir ini telah dilakukan analisis performansi *unsupervised learning* untuk video instruksi bernarasi menggunakan klasterisasi teks berdasarkan parameter. Parameter performansi yang diperhatikan adalah *F1 Score* yang terdiri dari presisi dan *recall*. Sistem dijalankan menggunakan perangkat lunak pengolahan data statistik, sedangkan performansi menggunakan bahasa pemrograman Python.

Penelitian ini menghasilkan performansi *F1 Score* rata-rata adalah 0,721. Skor tersebut didapatkan menggunakan konfigurasi parameter meliputi maksimal langkah ditemukan ($K=10$), konfigurasi *alignment cost* (Cs, Cd)= $(-1, 100)$, *global template* ($L=120$), dan jumlah data video ($N=30$).

Kata Kunci : Video instruksi bernarasi, *Unsupervised learning*, *Multiple sequence alignment*, Klasterisasi teks.

Abstract

Narrated instruction videos is one type of learning video or tutorial video. This type of video contains visual representation that is generally exhibited by person and textual representation in the form of subtitle. Machine learning technology can build an unsupervised learning system to summarize narrated video instructions to only a list of key steps. The system use Multiple Sequence Alignment (MSA) as text clustering method to align key steps. However, the system produces inconsistent performance in each video scenario even though it has the same number of Ground Truth.

Based on these problems, in this final project the author has analyzed the performance of unsupervised learning for narrated video instructions using text clustering based on parameters. Performance parameter is F1 Score which consists of precision and recall. The system is run using statistical data processing software, while performance uses the Python programming language.

This research resulted average performance with F1 Score is 0,721. The Score are obtained using parameter configuration which include maximum discovery steps ($K=10$), alignment cost configuration (Cs, Cd)= $(-1, 100)$, global template ($L=120$), and number of video data ($N=30$).

Keywords : *Narrated instruction videos, Unsupervised learning, Multiple sequence alignment, Text clustering.*

1. Pendahuluan

Video instruksi bernarasi merupakan salah satu jenis video yang sering digunakan sebagai media pembelajaran [1]. Video ini mengandung representasi visual yang umumnya diperagakan oleh manusia dan representasi tekstual dengan adanya narasi berupa *subtitle*. Banyaknya informasi langkah instruksi akan mempengaruhi panjang durasi video tersebut. Dengan teknologi *machine learning* dapat membangun sistem *unsupervised learning* untuk meringkas video instruksi bernarasi menjadi hanya daftar langkah utama (*key steps*).

Penelitian *unsupervised learning* untuk skrip suatu skenario pernah dilakukan oleh Michaela Regneri [2] dan Lea Frermann [3]. Kedua penelitian tersebut menggunakan *Multiple Sequence Alignment* (MSA) untuk klasterisasi langkah skenario dan menggunakan data *Event Sequence Description* (EDS), yaitu deskripsi langkah pengerjaan suatu tugas bersumber dari beberapa volunteer yang ditanyakan secara langsung. Seiring berkembangnya teknologi, kumpulan data *unsupervised learning* dapat bersumber dari internet seperti pada penelitian yang dilakukan oleh Jean Alayrac [4]. Pada penelitian tersebut, skrip *subtitle* disederhanakan menjadi frasa *Direct-Object relation* (DOBJ) yang kemudian disejajarkan dengan MSA yang dioptimasi algoritma Frank-Wolfe sehingga menghasilkan ringkasan urutan langkah utamanya saja.

Penelitian [4] menghasilkan nilai evaluasi F1 *Score* yang tidak konsisten di setiap skenario video. Sebagai contoh, pada skenario *changing tire* untuk K=15 memiliki F1 *Score* 0,8 dan skenario *performing CPR* untuk K=15 memiliki F1 *Score* 0,418. Kedua skenario tersebut memiliki jumlah *Ground Truth* yang sama namun menghasilkan nilai F1 *Score* yang cukup jauh. Perbedaan nilai evaluasi juga ditunjukkan pada perbedaan nilai K. Pada skenario *changing tire* untuk K=15 memiliki F1 *Score* 0,66.

Pada Tugas Akhir ini, penulis melakukan analisis performansi *unsupervised learning* untuk video instruksi bernarasi menggunakan klasterisasi teks pada penelitian [4] berdasarkan parameter. Parameter performansi yang diperhatikan adalah F1 *Score* yang terdiri dari presisi dan *recall*. Untuk mendapatkan performansi terbaik, parameter yang akan dioptimalisasi, yaitu maksimum langkah ditemukan (K), *cost similar* (Cs), *cost dissimilar* (Cd), *global template* (L), dan jumlah data video (N). Dengan penelitian ini diharapkan mendapatkan konfigurasi parameter dengan nilai performansi yang lebih baik untuk setiap skenario.

2. Tinjauan Pustaka

2.1 Unsupervised Learning

Unsupervised learning adalah jenis algoritma *machine learning* yang digunakan untuk menarik kesimpulan dari data input yang belum dikategorikan atau diberi label [6]. *Unsupervised learning* mengidentifikasi kesamaan dalam data dan bereaksi berdasarkan ada atau tidaknya kesamaan di setiap bagian data baru. *Unsupervised learning* dapat digunakan pada klasterisasi dan reduksi dimensi. Pada klasterisasi, pembelajaran terjadi secara otomatis dengan mengidentifikasi struktur dalam data. Kelompok atau label hasil analisis dapat diciptakan dengan sendirinya sehingga menjadi wawasan awal yang dapat digunakan untuk menguji hipotesis. Pada reduksi dimensi, pembelajaran digunakan untuk mengompres data dengan mempertahankan struktur dan kegunaannya.

2.2 Video Instruksi Bernarasi

Video instruksi bernarasi atau dikenal juga sebagai video *tutorial* adalah salah satu kategori video yang berisi panduan atau langkah-langkah dalam mengerjakan sesuatu dengan tujuan memberikan pengetahuan dalam melakukan suatu hal atau prosedur. Video instruksi bernarasi biasanya mengandung representasi tekstual berupa *subtitle* dan representasi visual berupa video itu sendiri. Sebuah video instruksi dibawakan oleh narator yang bertugas untuk menyampaikan langkah panduan.

Banyaknya langkah prosedur akan mempengaruhi panjang durasi video instruksi bernarasi. Selain itu, gaya bicara atau penyampaian narator dalam menyampaikan langkah-langkah prosedur juga akan mempengaruhi panjang durasi video. Biasanya video instruksi bernarasi memiliki bagian *intro* yang berisi pembukaan logo dan perkenalan diri bagi narator. Terkadang juga terdapat *credit* atau penutupan pada bagian akhir video.

2.3 Multiple Sequence Alignment (MSA)

Multiple Sequence Alignment (MSA) adalah metode penjajaran *sequence* yang pada awalnya digunakan untuk menjajarkan urutan informasi biologis seperti DNA, RNA, dan asam nukleat. Metode ini merupakan pengembangan metode *pairwise alignment* yang hanya menjajarkan dua

sequence [9]. MSA membantu mendeteksi hubungan historis dan kekeluargaan antara *sequence* protein atau asam amino dan menentukan struktur atau lokasi tertentu.

MSA dapat digunakan pada *unsupervised learning* sebagai metode untuk klusterisasi teks. MSA akan mengilustrasi ketidaksamaan di antara sekumpulan *sequence* sehingga menghasilkan ringkasan data yang berurutan. Sehingga dimungkinkan untuk membentuk algoritma yang dapat membantu dalam perhitungan berulang MSA. Diketahui m *sequence* $S_i, i = 1, \dots, m$ serupa dengan bentuk persamaan berikut.

$$S := \begin{cases} S_1 = (S_{11}, S_{12}, \dots, S_{1n_1}) \\ S_2 = (S_{21}, S_{22}, \dots, S_{2n_2}) \\ \vdots \\ S_m = (S_{m1}, S_{m2}, \dots, S_{mn_m}) \end{cases}$$

MSA diambil dari rangkaian *sequence* S dengan memasukkan sejumlah *gap* yang diperlukan ke dalam setiap *sequence* S_i dari S hingga *sequence* yang dimodifikasi. Seluruhnya sesuai dengan panjang $L \geq \max\{x_i | i = 1, \dots, r\}$ dan tidak ada nilai dalam *sequence* S pada kolom m yang sama, melainkan hanya terdiri dari *gap*. Bentuk matematika dari MSA untuk *sequence* di atas ditunjukkan pada persamaan berikut.

$$S := \begin{cases} S'_1 = (S'_{11}, S'_{12}, \dots, S'_{1L}) \\ S'_2 = (S'_{21}, S'_{22}, \dots, S'_{2L}) \\ \vdots \\ S'_m = (S'_{r1}, S'_{r2}, \dots, S'_{rL}) \end{cases}$$

2.4 Sum-of-pair Score

Sum-of-pair Score atau *SP-Score* adalah skema penilaian atau matriks akurasi untuk MSA relatif terhadap penjajaran referensi, berdasarkan jumlah pasangan residu yang dijajarkan dengan benar dan dijumlahkan pada semua pasangan *sequence*. Dalam menentukan *global alignment*, dapat menggunakan fungsi harga *sum-of-pairs* pada persamaan berikut.

$$\sum_{(n,m)} \sum_{l=1} c(\varphi(d^n)_l, \varphi(d^m)_l),$$

dimana $c(d_1, d_2)$ menunjukkan harga penjajaran dobj d_1 dan d_2 pada slot l .

2.5 Direct-Object Relation (DOBJ)

Direct-object relation (DOBJ) merupakan relasi gramatikal bahasa Inggris yang tergolong pada *dependency grammar*. Sebuah kalimat disebut dengan *direct object* jika entitas melakukan pekerjaan secara langsung. Berbeda dengan *indirect object* yaitu entitas secara tidak langsung dipengaruhi oleh tindakan. Berdasarkan Treebank Statistics, 12.450 node atau sebesar 5% melekat pada induk kalimat sebagai DOBJ. Mayoritas DOBJ terdiri atas pasangan VERB-NOUN (70%), VERB-PRON (19%), dan VERB-PROPN (5%).

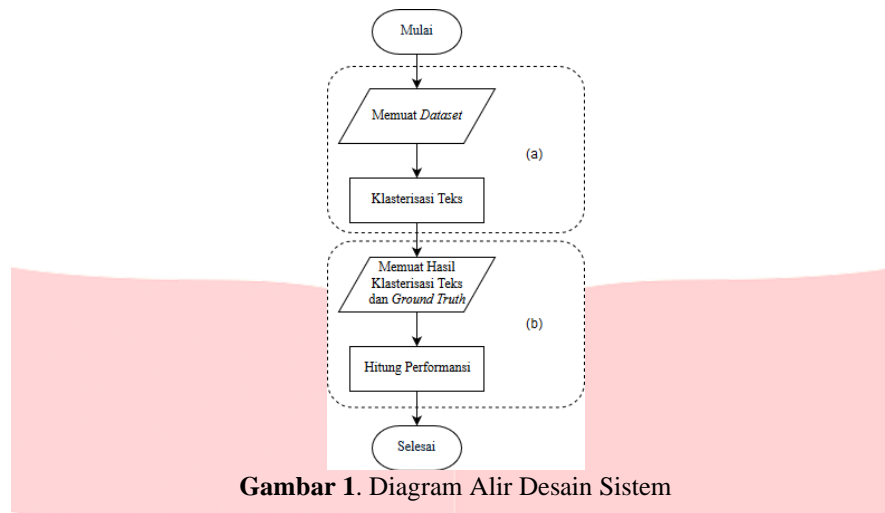
2.6 Matriks Konfusi

Matriks konfusi adalah tabel yang biasa digunakan untuk mendeskripsikan performansi dari model klasifikasi pada sekumpulan data yang telah diketahui nilai sebenarnya (*true values*). Dalam *unsupervised learning*, matriks konfusi dapat digunakan jika diketahui *Ground Truth* sebagai nilai sebenarnya. Matriks konfusi direpresentasikan sebagai tabel yang memiliki dua dimensi, yaitu aktual dan prediksi. Dimensi aktual akan diisi dengan data *Ground Truth*, sedangkan dimensi prediksi akan diisi dengan data hasil *learning*.

3. Perancangan Sistem

3.1 Gambaran Umum Sistem

Desain sistem dalam Tugas Akhir ini akan membahas perancangan *unsupervised learning* untuk meringkas video instruksi bernarasi menjadi langkah-langkah utama hingga menghitung performansi. Terdapat dua tahap, yaitu tahap pertama melalui sistem klusterisasi teks dan tahap kedua melalui proses perhitungan performansi. Kedua tahap tersebut dipisahkan karena perbedaan perangkat lunak yang akan dijalankan. Diagram alir desain sistem yang akan dirancang pada Tugas Akhir ini dijelaskan pada Gambar 1.



Gambar 1. Diagram Alir Desain Sistem

3.2 Spesifikasi Perangkat

Tugas Akhir ini menggunakan perangkat keras berupa laptop dengan spesifikasi sebagai berikut:

- Tipe : ACER Aspire E5-475G
- Processor : Intel® Core™ i5-7200U CPU @2.50GHz
- RAM : 8GB DDR4
- GPU : 2GB NVIDIA® 940MX

3.3 Dataset

Penelitian ini menggunakan dataset yang telah digunakan pada penelitian [4]. Dataset terdiri dari lima skenario video, yaitu *changing tire*, *make coffee*, *repat a plant*, *performing CPR*, dan *jumping cars*. Masing-masing skenario terdapat 1 data kamus DOBJ, 1 data matriks kesamaan, dan 30 data informasi video.

3.3.1 Kamus DOBJ

Kamus DOBJ berisi indeks DOBJ berdasarkan jumlah kemunculannya. DOBJ diperoleh melalui *dependency parsing*, yaitu proses mengekstraksi *dependency parse* dari sebuah kalimat yang merepresentasikan struktur gramatikal dan hubungan kebahasaan antar kata. Proses *dependency parsing* menggunakan Stanford Parser [12]. Masing-masing skenario video memiliki 1 data kamus DOBJ.

3.3.2 Matriks Kesamaan

Matriks kesamaan (*similarity matrix*) merupakan nilai yang menyatakan kesamaan antara dua titik data. Dalam hal ini, matriks kesamaan didapatkan dengan mencari jarak antara dua pasang DOBJ menggunakan Wordnet [13]. Nilai ini akan menjadi kalkulasi untuk proses klasterisasi teks. Masing-masing skenario video memiliki 1 data matriks kesamaan.

3.4 Klasterisasi Teks

Klasterisasi teks merupakan langkah penjajaran dobj dari dataset hingga menjadi langkah-langkah utama saja (key steps). Pada penelitian ini, klasterisasi teks menggunakan metode MSA yang dioptimasi algoritma Frank Wolfe. Langkah awal pada klasterisasi teks adalah menentukan parameter, kemudian dilakukan pemrograman kuadratik. Optimasi menggunakan algoritma Fran-Wolfe juga dilakukan untuk mendapatkan skala yang lebih baik dan fleksibilitas dalam menentukan *alignment cost*. Yang terakhir adalah ekstraksi langkah untuk menghasilkan output *discovery steps* yang nantinya akan dievaluasi.

3.5 Performansi

Performansi dilakukan dengan mengukur F1 Score yang terdiri dari *precision* dan *recall*. Perhitungan performansi menggunakan bahasa pemrograman python. Perhitungan *precision* menggunakan persamaan berikut.

$$Precision = \frac{TP}{DS}$$

Sedangkan perhitungan *recall* menggunakan persamaan berikut.

$$Recall = \frac{TP}{GT}$$

Setelah mendapatkan nilai *precision* dan *recall*, dapat dilakukan perhitungan *F1 Score* dengan persamaan berikut.

$$F1 = 2 \frac{\text{presisi} \cdot \text{recall}}{\text{presisi} + \text{recall}}$$

4. Pengujian dan Analisis Sistem

4.1 Pengujian Sistem

Pengujian sistem dilakukan untuk mengetahui performansi sebuah sistem. Pada penelitian ini, performansi sistem diukur berdasarkan parameter *F1 Score*. Pengujian sistem yang dilakukan dengan mengubah parameter maksimum langkah ditemukan (*K*), *cost similar* (*Cs*), *cost dissimilar* (*Cd*), *global template* (*L*), dan jumlah data video (*N*). Pengujian ini dilakukan terhadap masing-masing skenario video, yaitu *changing tire*, *make coffee*, *repot a plant*, *performing CPR*, dan *jumping cars*.

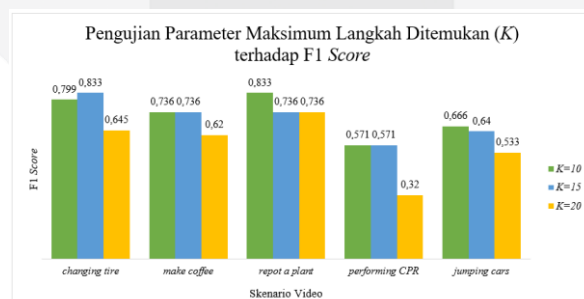
1. Pengujian terhadap parameter maksimum langkah ditemukan (*K*), nilai *K* yang digunakan adalah 10, 15, 20. Digunakan parameter nilai *Cs*=-1, *Cd*=100, *L*=120, dan *N*=30. Nilai *K* yang menghasilkan *F1 Score* terbaik akan digunakan pada pengujian selanjutnya.
2. Pengujian terhadap parameter *cost similar* (*Cs*), nilai *Cs* yang digunakan adalah -1, -10, -100. Digunakan parameter nilai *Cd*=100, *L*=120, *N*=30 dan *K* yang didapatkan dari pengujian pertama.
3. Pengujian terhadap parameter *cost dissimilar* (*Cd*), nilai *Cd* yang digunakan adalah 1, 10, 100. Digunakan parameter nilai *Cs*=-1, *L*=120, *N*=30 dan *K* yang didapatkan dari pengujian pertama.
4. Pengujian terhadap konfigurasi *alignment cost* yang terdiri dari *Cs* dan *Cd* atau ditulis dengan (*Cs,Cd*), nilai *Cs* yang digunakan adalah 1, 10, 100, sedangkan nilai *Cd* yang digunakan adalah 1, 10, dan 100. Digunakan parameter *L*=120, *N*=30, dan *K* yang didapatkan dari pengujian pertama. Konfigurasi *alignment cost* terbaik akan digunakan pada pengujian selanjutnya.
5. Pengujian terhadap parameter *global template* (*L*), nilai *L* yang digunakan adalah 60, 120, 240, dan 360. Digunakan parameter *N*=30, konfigurasi (*Cs,Cd*) dan *K* yang didapatkan dari pengujian sebelumnya.
6. Pengujian terhadap parameter jumlah data video (*N*), nilai *L* yang digunakan adalah 10, 20, 30. Digunakan parameter konfigurasi (*Cs,Cd*), *K*, dan *L* yang didapatkan dari pengujian sebelumnya.

4.2 Hasil dan Analisis Pengujian Sistem

Pengujian sistem dilakukan berdasarkan perubahan parameter *K*, *cost similar* (*Cs*), *cost dissimilar* (*Cd*), *global template* (*L*), jumlah data video (*N*). Pengaruh parameter akan diukur berdasarkan performansi dengan parameter *F1 Score*.

4.2.1 Pengujian Parameter Maksimum Langkah Ditemukan (*K*)

Pengujian ini dilakukan untuk mengetahui pengaruh parameter nilai maksimum langkah ditemukan (*K*) terhadap nilai *F1 Score*. Nilai *K* yang digunakan adalah 10, 15, dan 20. Grafik pengujian nilai *K* terhadap *F1 Score* ditunjukkan pada Gambar 2.



Gambar 2. Pengujian Parameter Maksimum Langkah Ditemukan (*K*) terhadap *F1 Score*

Berdasarkan hasil pengujian, nilai *F1 Score* cenderung rendah ketika nilai *K* besar. Hal ini terjadi karena memperbesar nilai *K* berarti memperbesar jumlah langkah ditemukan atau *Discovery*

Steps (DS). Jumlah DS yang besar akan mengecilkan nilai *precision* menurut persamaan. Selain itu, jumlah DS yang besar akan memunculkan banyak DOBJ yang bahkan dapat muncul lebih dari sekali. Kemudian dicari nilai K yang menghasilkan F1 *Score* rata-rata terbaik untuk seluruh skenario video. Hasil perhitungan F1 *Score* rata-rata berdasarkan nilai K dijelaskan pada Tabel 1.

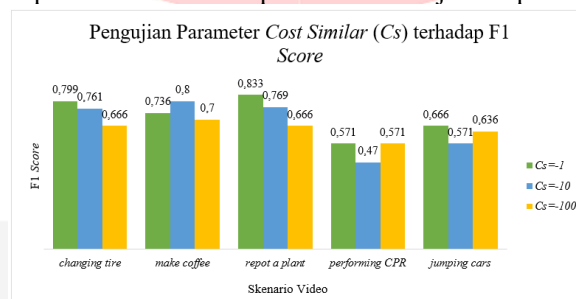
Tabel 1. Hasil Pengujian Parameter Maksimum Langkah Ditemukan (K)

Maksimum Langkah Ditemukan (K)	F1 <i>Score</i> Rata-rata
10	0,721
15	0,703
20	0,570

Nilai F1 *Score* rata-rata terbaik yang dihasilkan dari seluruh skenario video adalah 0,721 dengan nilai K adalah 10. Maka dari itu, pengujian berikutnya akan menggunakan nilai K tersebut.

4.2.2 Pengujian Parameter *Cost similar* (Cs)

Pengujian ini dilakukan untuk mengetahui pengaruh parameter *cost similar* (Cs) terhadap nilai F1 *Score*. Nilai Cs yang digunakan adalah -1, -10, dan -100, nilai K=10, Cd=100, L=120, dan N=30. Grafik pengujian parameter Cs terhadap F1 *Score* ditunjukkan pada Gambar 3.

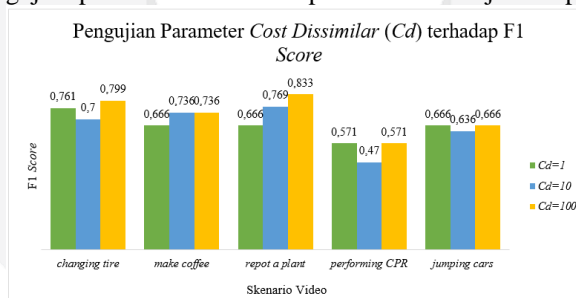


Gambar 3. Pengujian Parameter *Cost Similar* (Cs) terhadap F1 *Score*

Berdasarkan hasil pengujian, terdapat dua skenario video yang menunjukkan bahwa semakin kecil nilai Cs akan semakin kecil nilai F1 *Score*. Akan tetapi terdapat dua skenario video yang menunjukkan bahwa nilai Cs=-100 menghasilkan F1 *Score* lebih baik daripada Cs=-10. Maka dari itu, belum dapat diketahui pengaruh parameter *cost similar* secara langsung.

4.2.3 Pengujian Parameter *Cost similar* (Cd)

Pengujian ini dilakukan untuk mengetahui pengaruh parameter *cost dissimilar* (Cd) terhadap nilai F1 *Score*. Nilai Cd yang digunakan adalah 1, 10, dan 100, nilai K=10, Cs=-1, L=120, dan N=30. Grafik pengujian parameter Cd terhadap F1 *Score* ditunjukkan pada Gambar 4.



Gambar 4. Pengujian Parameter *Cost Dissimilar* (Cd) terhadap F1 *Score*

Berdasarkan hasil pengujian, terdapat tiga skenario video yang menunjukkan bahwa Cd=10 menghasilkan F1 *Score* yang lebih rendah daripada Cd=1 dan Cd=100. Hanya satu skenario video yang menunjukkan semakin besar nilai Cd maka semakin besar nilai F1 *Score*. Oleh karena itu, belum dapat diketahui pengaruh parameter *cost dissimilar* secara langsung.

4.2.4 Pengujian Konfigurasi *Alignment cost*

Pada dua pengujian sebelumnya, tidak dapat diketahui pengaruh parameter *cost similar* dan *cost dissimilar* secara langsung. Pengujian ini menggunakan konfigurasi kedua parameter tersebut menjadi konfigurasi *alignment cost* (Cs,Cd). Nilai Cs yang digunakan adalah -1, -10, dan -100,

sedangkan nilai Cd yang digunakan adalah 1, 10, dan 100. Digunakan nilai K=10, L=120, dan N=30. Hasil perhitungan F1 Score berdasarkan konfigurasi *alignment cost* dijelaskan pada Tabel 1.

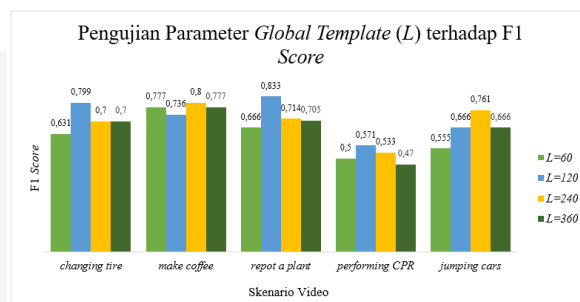
Tabel 2. Hasil Pengujian Konfigurasi *Alignment cost*

(Cs, Cd)	Changing Tire	Make Coffee	Repot a Plant	Performing CPR	Jumping Cars
(-1,1)	0,761	0,666	0,666	0,571	0,666
(-1,10)	0,700	0,736	0,769	0,470	0,636
(-1,100)	0,799	0,736	0,833	0,571	0,666
(-10,1)	0,761	0,777	0,769	0,533	0,666
(-10,10)	0,761	0,631	0,769	0,533	0,600
(-10,100)	0,761	0,800	0,769	0,470	0,571
(-100,1)	0,761	0,777	0,769	0,571	0,666
(-100,10)	0,761	0,777	0,769	0,533	0,666
(-100,100)	0,666	0,700	0,666	0,571	0,636

Berdasarkan hasil pengujian, didapatkan F1 Score pada setiap skenario video dengan skor tertinggi yang tercetak **tebal**. Konfigurasi (Cs,Cd) yang paling sering menghasilkan F1 Score tertinggi pada setiap skenario adalah (-1,100).

4.2.5 Pengaruh Nilai *Global template* (L)

Pengujian ini dilakukan untuk mengetahui pengaruh parameter *global template* (L) terhadap nilai F1 Score. Nilai L yang digunakan adalah 60, 120, 240, dan 360 dengan konfigurasi (Cs,Cd)=(-1,100), K=10, dan N=30. Grafik pengujian nilai *global template* terhadap F1 Score ditunjukkan pada Gambar 5.



Gambar 5. Pengaruh Nilai *Global template* (L) terhadap F1 Score

Berdasarkan hasil pengujian, diperoleh F1 Score tertinggi pada L=120 untuk tiga skenario video. Namun, terdapat dua skenario video memperoleh F1 Score tertinggi pada L=240. Oleh karena itu, dihitung F1 Score rata-rata untuk seluruh skenario video yang ditunjukkan pada Tabel 3.

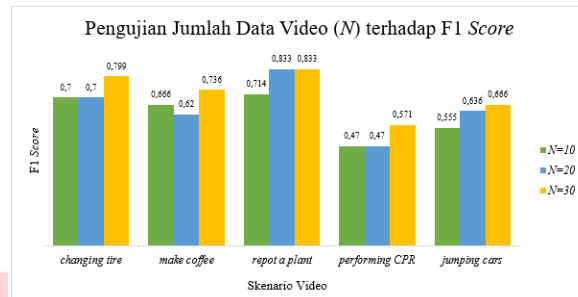
Tabel 3. Hasil Pengujian *Global Template*

<i>Global Template</i> (L)	F1 Score Rata-rata
60	0,6258
120	0,7210
240	0,7016
360	0,6636

F1 Score rata-rata yang terbaik adalah 0,7210 pada L=120. Sedangkan pada L=60 menjadi parameter menghasilkan F1 Score terendah. Terjadi penurunan F1 Score ketika nilai L diperbesar dari 120 ke 360. Akan tetapi, nilai L juga perlu memiliki batas minimal agar dapat menghasilkan F1 Score yang baik. skenario adalah (-1,100).

4.2.6 Pengaruh Jumlah Data Video (N)

Nilai N yang digunakan adalah 10, 20, dan 30 dengan konfigurasi (Cs,Cd)=(-1,100), K=10, dan L=120. Grafik pengujian jumlah data video terhadap F1 Score ditunjukkan pada Gambar 6.



Gambar 6. Pengaruh Jumlah Data Video (L) terhadap F1 Score

Berdasarkan hasil pengujian, diperoleh bahwa F1 Score cenderung tinggi pada N=30. F1 Score rata-rata yang dihasilkan pada seluruh skenario video dijelaskan pada Tabel 4.

Tabel 4. Hasil Pengujian Jumlah Data Video

Jumlah Data Video (N)	F1 Score Rata-rata
10	0,621
20	0,652
30	0,721

F1 Score rata-rata yang terbaik adalah 0,721 pada N=30. Hal ini dikarenakan semakin banyak jumlah data maka semakin besar potensi kesesuaian dengan *ground truth* sehingga memperbaiki nilai F1 Score. Jumlah data video yang sedikit berarti membatasi data *unsupervised learning* sehingga dalam *labelling* klusterisasi teks tidak berjalan dengan baik.

5. Kesimpulan

Berdasarkan hasil pengujian dan analisis yang dilakukan pada sistem *unsupervised learning* untuk video instruksi bernarasi menggunakan klusterisasi teks, dapat disimpulkan bahwa sistem menghasilkan performansi F1 Score rata-rata adalah 0,721 dengan menggunakan konfigurasi parameter meliputi maksimal langkah ditemukan ($K=10$), konfigurasi *alignment cost* (C_s, C_d)= $(-1, 100)$, *global template* ($L=120$), dan jumlah data video ($N=30$). Jadi, konfigurasi parameter yang telah dilakukan pada penelitian sebelumnya sudah menjadi konfigurasi parameter yang menghasilkan F1 Score terbaik untuk seluruh skenario video.

Daftar Pustaka:

- [1] W. J. Hsin and J. Cigas, "Short Videos Improve Student Learning in Online Education," in *Journal of Computing Sciences in Colleges*, no. 28, pp. 253-259, 2013.
- [2] M. Regneri, A. Koller, and M. Pinkal, "Learning Script Knowledge with Web Experiments," in *Association for Computational Linguistics*, pp. 979-988, 2010.
- [3] L. Frermann, I. Titov, and M. Pinkal, "A hierarchical Bayesian model for Unsupervised Induction of Script Knowledge," in *European Association for Computational Linguistics*, pp. 49-57, 2014.
- [4] J. Alayrac, P. Bojanowski, N. Agrawal, J. Sivic, I. Laptev and S. Lacoste-Julien, "Learning from Narrated Instruction Videos," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 9, pp. 2194-2208, 2018.
- [5] Russel, Stuart J., *Artificial Intelligence: A Modern Approach*, Third Ed. NJ: Prentice Hall Press, 2009.
- [6] S. Devin, "Supervised vs Unsupervised Learning," *Towards Data Science*, 2018. [Online]. Available: <https://www.datascience.com/blog/supervisedand-unsupervised-machine-learning-algorithms>
- [7] Mount D. M., *Bioinformatics: Sequence and Genome Analysis*, 2nd ed., NY: Cold Spring Harbor Laboratory Press, 2004.