

# Analisis Komparatif Berbagai Teori Karakteristik Proses Pendekatan Anak dengan Tujuan Seksual untuk Mendeteksi Percakapan Teks

Livia Ashianti<sup>1</sup>, Kanyadian Idananta<sup>2</sup>

<sup>1,2</sup> Universitas Bina Nusantara, Computer Science

[livia.ashianti@binus.ac.id](mailto:livia.ashianti@binus.ac.id)

[kidananta@binus.edu](mailto:kidananta@binus.edu)

**Abstrak**— Teknologi internet sering disalahgunakan untuk tindakan kriminal, salah satunya adalah proses pendekatan anak untuk tujuan seksual secara online. Dalam melakukan tindakannya, pelaku melakukan pendekatan kepada korban dengan menggunakan teknologi internet. percakapan teks menjadi bukti kejahatannya. Studi sebelumnya menunjukkan bahwa teori karakteristik dapat menentukan tingkat akurasi dalam mendeteksi percakapan teks yang berisi proses pendekatan tersebut. Penelitian ini akan melakukan uji coba menggunakan teori karakteristik yang berbeda untuk mendapatkan teori karakteristik terbaik untuk mendeteksi percakapan teks mengandung proses pendekatan dengan tujuan seksual. Karakteristik tersebut akan digunakan untuk ekstraksi fitur percakapan teks. Kemudian diklasifikasikan menggunakan metode Support Vector Machine (SVM). Hasil penelitian menunjukkan bahwa penggunaan jumlah karakteristik yang berbeda berdampak pada tingkat akurasinya. Namun, menggabungkan 2 teori karakteristik berbeda menghasilkan tingkat akurasi yang lebih baik.

**Kata Kunci**— Proses Pendekatan Seksual, Ekstraksi fitur, Text Mining, Karakteristik, Support Vector Machine

## I. PENDAHULUAN

Cara individu berkomunikasi dan berinteraksi selalu mengalami perkembangan teknologi dan komunikasi di era globalisasi saat ini sangat pesat. Saat ini penggunaan internet di dunia maya sudah menjadi kebutuhan dalam setiap aktivitas. Tidak hanya orang dewasa saja yang menggunakan akses internet tetapi anak-anak juga mengenal dan menggunakan internet. Salah satunya adalah penggunaan media sosial, guna mempermudah berkomunikasi dan berinteraksi tanpa harus bertatap muka [1]. Namun, manfaat dari perkembangan teknologi kerap disalahgunakan untuk perbuatan kriminal, salah satunya yaitu pelecehan seksual secara daring khususnya untuk anak dibawah umur.

Dalam melakukan aksinya, pelaku akan melakukan proses pendekatan terhadap korban. Saat pelaku sudah mendapatkan kepercayaan dari korbannya, pelaku baru melaksanakan aksinya. Menurut UNICEF Indonesia pada tahun 2023, 22% anak-anak secara tidak terduga menemukan konten seksual daring melalui iklan, umpan media sosial, mesin telusur, dan aplikasi perpesanan, dan 9% melaporkan bahwa mereka secara aktif mencari materi semacam itu. 500.000 anak di Indonesia dilaporkan mengalami eksloitasi seksual daring dan perilaku berbahaya dalam satu tahun terakhir, Selain itu 29% anak-anak

mengirimkan informasi personal mereka kepada seseorang yang tidak pernah mereka temui sebelumnya [2].

Dalam sepuluh tahun terakhir terdapat beberapa penelitian terkait analisis dan deteksi menggunakan text mining dengan melakukan klasifikasi percakapan teks. Klasifikasi percakapan untuk melihat percakapan teks mengandung proses pendekatan terhadap anak untuk dengan tujuan melakukan pelecehan seksual. Penelitian yang dilakukan Zhenming Zuo, Jie Li, Philip Anderson, Longzhi Yang dan Nitin Naik menggunakan Fuzzy Rough & klasifikasi teks untuk mengidentifikasi tipe-tipe pelaku berdasarkan pola percakapan [3]. Penelitian dilakukan oleh Lugman H.M. Nasir, Zurina Saaya dan Mohd K. Baharon menggunakan Neural Network dan klasifikasi percakapan teks dalam melakukan pendekatan kepada korban [4].

Penelitian dilakukan oleh Iin Ernawati pada tahun 2019 menggunakan Klasifikasi Naïve Bayes dan Support Vector Machine (SVM) sebagai alternatif solusi untuk *text mining* [5]. Kemudian, penelitian dilakukan oleh Pranoto, Gunawan dan Seowito pada 2015 yang mengidentifikasi dua puluh karakteristik *online child grooming* pada percakapan teks dan membangun model logistic regression untuk mendeteksi proses pendekatan pelaku kepada korban pada percakapan teks [6]. Serta pada tahun 2018, Gunawan, Ashianti, Sekishita, melakukan penelitian terhadap percakapan teks antara pelaku dan korban [7]. Beberapa peneliti telah mengidentifikasi sejumlah karakteristik pendekatan yang dilakukan oleh pelaku kepada korban. Karakteristik yang digunakan oleh masing-masing peneliti berbeda-beda.

Berdasarkan data dari KPAI, semakin tahun kejadian seksual *online* pada anak semakin meningkat dengan berbagai macam varian kejadian [8]. Dengan adanya pandemi COVID-19 yang kita alami saat ini, adanya kejadian seksual ini semakin meningkat, karena hampir semua hal dilakukan secara daring, termasuk dengan adanya proses belajar mengajar secara daring. Wakil ketua Lembaga Perlindungan Saksi dan Korban, Antonius PS Wibowo mengatakan bahwa adanya peningkatan tajam pada kejadian seksual terhadap anak bisa jadi karena Pandemi COVID-19 [9].

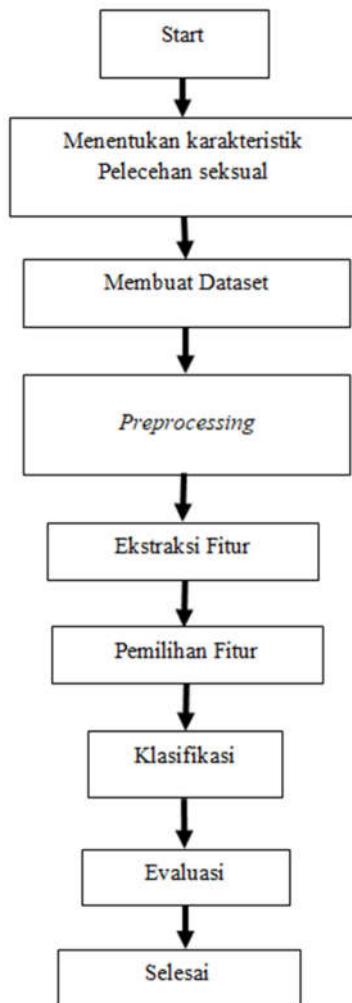
Dari berbagai penelitian yang telah dilakukan sebelumnya, dapat diketahui bahwa karakteristik pelaku kejadian seksual

online pada anak-anak semakin bervariasi dan terdapat karakteristik yang berbeda-beda tergantung sifat pelaku [10]. Namun belum ada yang melakukan analisa perbandingan karakteristik pelaku kejahatan dengan akurasi yang lebih baik.

Atas dasar tersebut akan dilakukan analisis perbandingan hasil identifikasi dengan menggunakan karakteristik berbeda-beda yang telah diteliti oleh peneliti sebelumnya. Analisa ini dimulai dengan memproses karakteristik pada percakapan, yang telah ditentukan oleh peneliti sebelumnya, serta melakukan klasifikasi menggunakan metode SVM dengan berbagai jenis kernel. Selain itu juga akan dilakukan penelitian untuk mengetahui akan hasil yang akan didapatkan jika karakteristik yang digunakan oleh peneliti yang berbeda digabungkan.

## II. METODE PENELITIAN

Berikut adalah metode penelitian yang digunakan dalam penelitian ini:



Gbr 1. Metode Penelitian

### A. Menentukan Karakteristik Pelecehan Seksual

Berdasarkan penelitian-penelitian yang telah dilakukan sebelumnya, terdapat karakteristik-karakteristik yang menggambarkan sebuah percakapan mengandung proses pendekatan pelaku ke korban untuk tujuan pelecehan seksual. Karakteristik tersebut diperlukan untuk mendeteksi proses pendekatan pelaku kepada korban dalam sebuah percakapan teks. Dalam penelitian ini, akan dilakukan dengan menggunakan 2 teori karakteristik yaitu teori pertama terdiri 17 Karakteristik berdasarkan penelitian Gunawan [7], Ashianti and Sekishita in 2018 dan teori kedua terdiri dari 9 Karakteristik berdasarkan penelitian Mcghee in 2011 [10].

TABLE I. Karakteristik proses pendekatan untuk tujuan pelecehan seksual terhadap anak (karakter A)

Variable	Karakteristik
A1	Asking profile
A2	Other way contact
A3	Asking picture
A4	Giving compliment
A5	Talking about activity, favorite, hobby, school
A6	Talking about friend and relationship
A7	Asking questions to know risk of conversation
A8	Acknowledging wrong-doing
A9	Asking if the <i>child</i> is alone or under adult or friend supervision
A10	Trying to build mutual trust
A11	Using falling in love words
A12	Using word to express feeling
A13	Using word about biology, body, intimate parts, and sexual category
A14	Asking hot picture
A15	Introducing sexual stage
A16	Sexual stage
A17	Arranging further contact and meeting

TABLE II. Karakteristik proses pendekatan untuk tujuan pelecehan seksual terhadap anak (Karakter B)

Variable	Karakteristik
B1	<i>Gaining Access</i>
B2	<i>Personal Information</i>
B3	<i>Relationship Information</i>
B4	<i>Activities Information</i>
B5	<i>Compliment</i>
B6	<i>Communicative Desensitization</i>
B7	<i>Reframing</i>
B8	<i>Isolation</i>
B9	<i>Approach</i>

penjelasan dari 17 Karakteristik berdasarkan penelitian Gunawan, Ashianti and Sekishita in 2018 (Karakteristik A) [7]:

1) A1 - *Asking profile*

Pelaku dan korban melakukan pertukaran informasi yang berhubungan dengan profil pribadi masing-masing. Biasa yang dilakukan yaitu bertukar informasi seperti nama, umur, tempat tinggal, tanggal lahir.

2) A2 - *Other way contact*

Pelaku dan korban mendiskusi kontak lainnya yang bisa digunakan untuk berkomunikasi. Pada tahap ini dilakukan untuk mendapatkan alternatif lainnya untuk berkomunikasi. Hal yang biasa ditanyakan seperti nomor telepon, email, dan akun media sosial.

3) A3 - *Asking picture*

Pelaku meminta korban mengirimkan foto diri atau saling bertukar foto.

4) A4 - *Giving compliment*

Pelaku membuat korbannya supaya merasa senang, tersanjung, mulai tergerak hatinya dengan cara memberikan pujiannya kepada korbannya.

5) A5 - *Talk about activity, favorite, hobby, school*

Pelaku dan korban membicarakan hal-hal berkaitan dengan aktivitas sehari-hari, hal-hal yang disukai, hobi maupun aktivitas sekolah.

6) A6 - *Talk about friend and relationship*

Pelaku dan korban membicarakan mengenai pertemanan dan hubungan dengan orang lain. Biasanya pelaku melakukan untuk mengetahui kedekatan korban dengan orang sekitarnya. Semakin korban tidak memiliki teman atau relasi yang kuat dengan keluarga atau orang lain, pelaku akan semakin mudah untuk dekat dengan korbannya.

7) A7 - *Asking questions to know risk of conversation*

Pelaku mencari tahu resiko yang mungkin terjadi yang bisa merugikan pelaku. Pelaku akan bertanya kepada korban terkait dengan kedekatan atau percakapan yang dilakukan oleh mereka berdua apakah ada orang yang mengetahuinya. Jika percakapan tersebut diketahui oleh orang lain, kemungkinan pelaku ketahuan dan dilaporkan akan semakin besar. Tidak hanya itu, biasanya pelaku menanyakan siapa yang sering memeriksa komputer atau *handphone* korbannya, menanyakan letak komputer, dan menanyakan orang tua atau orang lain yang berelasi dengan korban mengetahui *password* dari aplikasi yang mereka gunakan untuk berkomunikasi.

8) A8 - *Acknowledging wrong-doing*

Pelaku memastikan korban mengetahui resiko yang akan dihadapi dan memastikan komunikasi dengan pelaku dilakukan tanpa ada paksaan. Pelaku akan memberitahukan korban bahwa apa yang mereka lakukan salah, dan memiliki risiko hukum bagi pelaku. Dengan mengatakan ini kepada korban, pelaku berharap tidak akan terjerat kasus hukum karena hubungan mereka dibangun karena kemauan korban.

9) A9 - *Asking if the child is alone or under adult or friend supervision*

Pelaku memastikan korban sedang sendiri dan tidak berada dibawah pengawasan orang lain seperti teman atau keluarga.

10) A10 - *Trying build mutual trust*

Pelaku mencoba membangun kepercayaan korban kepada pelaku. Hubungan akan dapat lanjut ke tingkat berikutnya apabila pelaku mendapatkan kepercayaan dari korban.

11) A11 - *Using falling in love words*

Pelaku mulai menggunakan kata-kata yang menggambarkan perasaan jatuh cinta kepada pelaku. Hal ini dilakukan untuk membuat korban jatuh cinta kepada pelaku.

12) A12 - *Using word to express feeling*

Dalam percakapan antara pelaku dan korban menggunakan kata-kata yang mengungkapkan perasaan mereka. Biasanya hal ini dilakukan agar korban semakin terikat secara emosional dengan pelaku.

13) A13 - *Using word about biology, body, intimate parts, and sexual category*

Dalam percakapan antara pelaku dan korban menggunakan kata-kata yang mengandung konteks seksual.

14) A14 - *Asking hot picture*

Pelaku dengan korban akan bertukar foto seksual atau pelaku hanya meminta foto seksual korban. Foto tersebut digunakan untuk berimajinasi. Selain itu terkadang digunakan mengancam korban apabila tidak mau melakukan hal-hal yang diminta oleh pelaku.

15) A15 - *Introducing sexual stage*

Pelaku mengarahkan pembicaraan ke dalam konteks seks seperti pengalaman seks ataupun menanyakan hal-hal yang berhubungan dengan seks untuk mengetahui sejauh mana korban tahu tentang seks.

16) A16 - *Sexual Stage*

Pelaku dan korban melakukan percakapan dimana dalam percakapan tersebut mereka berfantasi melakukan seks.

17) A17 - *Arranging further contact and meeting*

Pelaku mengajak korban untuk bertemu disuatu tempat baik itu di rumah korban, di rumah pelaku atau di tempat lainnya.

Berikut penjelasan dari 9 karakteristik proses pendekatan untuk tujuan pelecehan seksual terhadap anak (Karakter B)

1) B1 - *Gaining Access*

Pelaku mengirimkan sapaan untuk memulai pendekatan terhadap korbannya.

2) B2 - *Personal Information*

Pelaku dan korban bertukar informasi pribadi seperti nama, umur, tempat tinggal, no telepon, foto pribadi, email. Selain itu juga pelaku dan korban berdiskusi mengenai kesukaan atau hal yang tidak disukai.

3) B3 - *Relationship Information*

Pelaku menanyakan mengenai hubungan antara korban dengan keluarga, saudara maupun untuk mencari tahu apakah korban memiliki hubungan spesial dengan orang lain. Pada tahap ini juga pelaku melakukan pendekatan untuk menjalin hubungan romantis dengan korban serta membangun kepercayaan korban kepadanya.

4) B4 - *Activities Information*

Pelaku dan korban membicarakan mengenai aktivitas yang sering dilakukan baik dalam keseharian dalam rumah atau sekolah selain itu juga membicarakan mengenai hobi.

5) B5 - *Compliments*

Dalam percakapan, pelaku memberikan pujiannya kepada korban agar korban semakin dekat dengan pelaku dan bertujuan untuk membuat korban menyukai pelaku.

6) B6 - *Communicative Desensitization*

Pelaku menggunakan kata-kata vulgar atau berhubungan dengan sexualitas. Selain itu juga pelaku akan mendiskusikan kepada korban untuk menganalisa apakah korban bersedia untuk melakukan hubungan sexual dengan pelaku.

7) B7 – *Reframing*

Pelaku membicarakan hal-hal yang berkaitan dengan hubungan sexualitas namun dengan kata-kata non seksual.

8) B8 – *Isolation*

Korban dibuat semakin jauh dengan keluarga dan temannya sehingga korban semakin dekat dengan Pelaku. Selain itu, isolasi ini pelaku meminta korban untuk berbohong kepada keluarga dan temannya.

9) B9 – *Approach*

Pelaku mencoba untuk bertemu dengan korban secara langsung. Dengan cara mengatur pertemuan baik bertemu disuatu tempat atau datang ke rumah korban atau pelaku.

B. *Membuat Set Data*

Untuk melakukan penelitian ini, diperlukan set data yang akan digunakan untuk klasifikasi. Set data yang digunakan untuk penelitian ini berupa percakapan teks. Set data yang akan digunakan terdiri dari 2 yaitu percakapan yang mengandung proses pendekatan untuk tujuan seksual serta percakapan yang tidak mengandung hal tersebut. Untuk Percakapan yang mengandung proses pendekatan diambil dari website *Perverted Justice* ([www.perverted-justice.com](http://www.perverted-justice.com)) sebanyak 100 percakapan (*Positive*). Sedangkan untuk percakapan yang tidak mengandung proses pendekatan diambil dari website literotika ([www.literotika.com](http://www.literotika.com)) yang menyediakan percakapan seksual antara orang dewasa secara legal sebanyak 50 percakapan. Setiap percakapan diambil dan disimpan ke dalam satu dokumen berekstensi “.txt”. Setiap dokumen baik yang positif mengandung proses pendekatan dan negatif dikumpulkan dibagi secara acak ke dalam set data pelatihan dan pengujian sesuai dengan Tabel 3.

TABLE III. Pembagian data untuk klasifikasi

Set Data	Status Dokumen	
	Positif	Negatif
Pelatihan	70	30
Pengujian	35	15
<b>Total</b>	105	45

Setiap dokumen yang telah dibagi akan diberi label sesuai dengan set data dan status dokumen sesuai dengan Tabel 4.

TABLE IV. Label status dokumen untuk klasifikasi

Set Data	Status Dokumen	
	<i>Grooming</i>	Non <i>Grooming</i>
Pelatihan	Positive-Training	Negative-Training
Pengujian	Positive-Testing	Negative-Testing

Pembagian dokumen ke dalam set data dilakukan sebanyak 10 kali sehingga terdapat 10 variasi set data. Dalam

pembagiannya, setiap dokumen minimal 1 kali dimasukan ke dalam set data pelatihan dan pengujian.

C. *Preprocessing Text*

Setiap dokumen akan dilakukan *preprocessing teks*. *Preprocessing text* dilakukan dengan tujuan mempermudah pengolahan percakapan pada tahap selanjutnya. Terdapat 5 tahap yaitu:

1) *Tokenization*

pada tahap ini akan menghilangkan karakter non letter dan setiap dokumen dipecah-pecah menjadi kata-kata.

2) *Transform case*

teks dari dokumen akan ditransformasi menjadi huruf kecil.

3) *Stopwords*

Menghilangkan kata-kata yang tidak diperlukan pada setiap dokumen

4) *Stemming*

Kata-kata pada dokumen ditransformasi menjadi bentuk kata dasar, proses *stemming* ini menggunakan algoritma porter.

5) *Generate Unigram, Bigram dan Trigram*

Kata-kata pada dokumen akan dibentuk menjadi 3 bentuk yaitu

a. *Unigram* yang terdiri dari satu kata

b. *Bigram* yang terdiri dari dua gabungan kata berurutan

c. *Trigram* yang terdiri dari tiga gabungan kata berurutan.

D. *Ekstraksi Fitur*

Dokumen yang telah dilakukan *preprocessing* akan ditransformasi ke dalam model *vector space*. Proses transformasi dengan setiap dokumen dibuat kedalam sebuah vektor yang terdiri fitur-fitur yang diekstrak. Fitur hasil ekstraksi berupa kata atau gabungan kata membentuk daftar kata (*wordlist*). Daftar kata tersebut dinotasikan  $T_1, T_2, \dots, T_t$ . Hasil ekstraksi fitur pada setiap dokumen ( $d_m$ ) seluruh set data dinotasikan ke dalam vektor  $d_m = \{w_{T_1, m}, w_{T_2, m}, \dots, w_{T_t, m}\}$  di mana  $m \in M$ ,  $M$  merupakan jumlah seluruh dokumen dan  $w_{T_i, m}$  bernilai 1 apabila terdapat fitur  $i$  (kata atau *wordlist*) dalam percakapan tersebut dan 0 jika tidak ada dalam percakapan tersebut.

E. *Pemilihan Fitur*

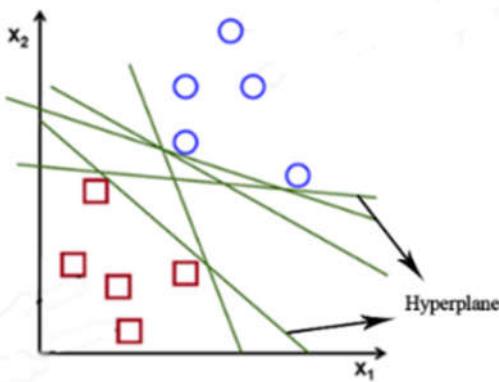
Hasil ekstraksi fitur setiap dokumen dalam model *vector space* kemudian digunakan untuk membuat vektor karakteristik. Vektor karakteristik digunakan untuk klasifikasi pada tahap selanjutnya.. Pada tahap ini akan dibuat dua vektor karakteristik yaitu karakteristik A (17 karakteristik) dan karakteristik B (9 karakteristik). kedua vektor tersebut akan dinyatakan dengan  $c_m = \{c_{m,1}, c_{m,2}, c_{m,3}, \dots, c_{m,17}\}$  di mana  $m \in M$  dan  $c_{m,j}$  merupakan nilai yang menandakan ada tidaknya karakteristik  $j$  pada dokumen  $m$ . Nilai  $c_{m,j}$  akan bernilai 0 apabila dokumen  $m$  tidak mengandung karakteristik  $j$  atau bernilai 1 apabila dokumen  $m$  mengandung karakteristik  $j$ .

Untuk menentukan nilai sebuah karakteristik *grooming* pada dokumen ( $c_{m,j}$ ), fitur berupa kata atau gabungan kata dari hasil ekstraksi akan dipilih sesuai dengan daftar kata atau gabungan kata yang menggambarkan setiap karakteristik *grooming*. Fitur-fitur yang telah dipilih kemudian dijumlahkan nilainya. Apabila hasilnya 0 maka nilai karakteristik  $j$  pada dokumen  $m$  ( $c_{m,j}$ ) bernilai 0. Apabila hasilnya lebih dari 0 maka nilai karakteristik  $j$  pada dokumen  $m$  ( $c_{m,j}$ ) bernilai 1.

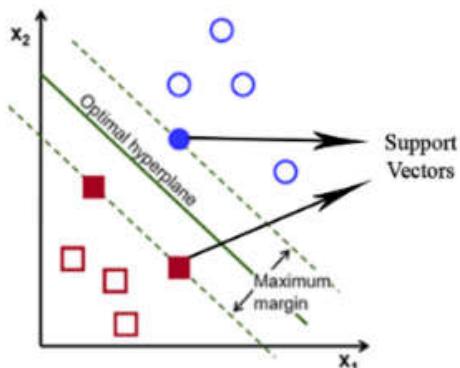
#### F. Klasifikasi

Tahap pertama yaitu membangun model klasifikasi. Model klasifikasi dibangun menggunakan vektor karakteristik setiap dokumen dari set data pelatihan dan status dokumen yang bernilai 0 (positive) atau 1 (negative) untuk label kelas. Dalam melakukan klasifikasi, salah satu metode terbaik untuk melakukan klasifikasi adalah menggunakan *Support Vector Machine* (SVM) [12] [13].

Dalam membangun model klasifikasi, SVM mencari hyperplane terbaik yang memisahkan dua buah kelas yaitu kelas *positive* dan *negative*. SVM merupakan metode numerik untuk menghitung hyperplane yang memisahkan dua kelas set data. SVM dapat bisa digunakan untuk klasifikasi multi kelas [14] [15].



Gbr 2. Pemisah set dengan menggunakan *hyperlane* [16]



Gbr 3. Hyperplane Optimal [16]

Gambar 2 menunjukkan bahwa ada beberapa hyperplane yang dibuat namun tidak semuanya memberikan output yang baik. Algoritma SVM mencari *hyperplane* yang memberikan jarak minimum ke set data pelatihan. Jarak ini dikenal sebagai *margin*. Kendala *hyperplane* optimal memaksimalkan margin data pelatihan seperti yang ditunjukkan pada Gambar 3.

Setelah membangun model, klasifikasi dokumen dilakukan menggunakan vektor karakteristik setiap dokumen dari set data pengujian. Hasil klasifikasi berupa status dokumen yang bernilai 0 yang berarti dokumen diklasifikasi mengandung proses pendekatan untuk tujuan seksual atau 1 yang berarti dokumen diklasifikasi tidak mengandung.

### III. PEMBAHASAN

Hasil dari setiap tahap yang telah dilakukan akan dijelaskan dan dibahas sesuai dengan metode penelitian.

#### A. Ekstraksi Fitur

Setiap dokumen percakapan pada set data pelatihan dan pengujian akan dilakukan *preprocessing*. Contoh tahapan *preprocessing* dari sebuah potongan percakapan dapat dilihat pada tabel 5.

#### B. Ekstraksi Fitur

Pada tahap ini setiap dokumen seluruh set data akan diekstraksi fitur berupa kata dan *wordlist* yang direpresentasikan dalam *model vector space* dengan nilai 0 atau 1. Nilai 1 apabila pada dokumen didapatkan fitur-fitur berupa kata atau *wordlist*. Nilai 0 apabila pada dokumen tidak didapatkan fitur berupa kata atau *wordlist* tersebut. Contoh hasil ekstraksi fitur set data pelatihan ditampilkan pada Tabel 6.

TABLE V. Hasil dari setiap tahap *preprocessing* potongan percakapan dari dokumen berstatus *grooming* (skybluesome sebagai pelaku dan attitude\_ithink\_not sebagai korban)

Dokumen Percakapan	skybluesome 03/23/07 1:09:23 Do you have some pics to share? attitude_ithink_not 03/23/07 1:09:45 sureee hehehe gar skybluesome 03/23/07 1:11:27 OMG so hot attitude_ithink_not 03/23/07 1:11:35 lol ty
Tokenization	skybluesome Do you have some pics to share? attitude_ithink_not sureee hehehe skybluesome OMG so hot attitude_ithink_not lol ty
Transform case	skybluesome do you have some pics to share? attitude_ithink_not sureee hehehe skybluesome omg so hot attitude_ithink_not lol ty
Stopword	skybluesome do you have some pics to share attitude_ithink_not sureee hehehe skybluesome omg so hot attitude_ithink_not lol ty
Stemming	skybluesome do you have some pic to share attitude_ithink_not sureee hehehe skybluesome omg so hot attitude_ithink_not lol ty

**Generate Unigram, Bigram dan Trigram** *skybluesome skybluesome\_do skybluesome\_do\_you do do\_you do\_you\_have you you\_have you\_have\_some have have\_some have\_some\_pic some some\_pic some\_pic\_to\_pic pic\_to\_pic\_to\_share to\_to\_share share share\_attitude attitude attitude\_ithink attitude\_ithink\_not ithink\_ithink\_not ithink\_not\_sureee not\_not\_sureee not\_sureee\_hehehe sureee\_sureee\_hehehe sure\_hehehe\_skybluesome hehehe\_hehehe\_skybluesome\_hehehe\_skybluesome\_omg skybluesome\_skybluesome\_omg skybluesome\_omg\_so\_omg omg\_so\_omg\_so\_hot so\_so\_hot so\_hot\_attitude hot hot\_attitude hot\_attitude\_ithink\_attitude\_attitude\_ithink\_attitude\_ithink\_not ithink\_not\_ithink\_attitude\_ithink\_not ithink\_not\_lol not\_lol not\_lol\_ty lol\_lol\_ty\_ty*

TABLE VI. Contoh hasil ekstraksi

Nama Label	Nama File	Hasil Ekstraksi Fitur Berupa Kata/Gabungan Kata			
		Beauty	Lost	Lost_interest	Love
grooming-training	1.txt	1	1	0	0
grooming-training	2.txt	0	1	1	0
nongrooming-training	200.txt	0	0	0	0
nongrooming-training	201.txt	0	1	0	1

### C. Pemilihan Fitur

Hasil ekstraksi fitur setiap dokumen pada setiap set data akan dipilih untuk membentuk vektor karakteristik. Fitur yang dipilih adalah fitur yang menggambarkan setiap karakteristik A dan karakteristik B. List fitur (kata atau Wordlist) untuk karakteristik A diambil dari Gunawan, Ashianti and Sekisihita in 2018 [7] dan karakteristik B diambil dari Mcghee in 2011 [10]. Pemilihan fitur tersebut akan membentuk 2 buah vektor karakteristik. Pertama vektor karakteristik yang menggunakan karakteristik A. Kedua vektor karakteristik yang menggunakan karakteristik B.

Tahap pertama dalam pemilihan fitur yaitu fitur dari hasil ekstraksi fitur set data akan dipetakan dengan kata atau *wordlist* yang menggambarkan setiap karakteristik. Misalnya pada kata atau *wordlist* karakteristik A1 yaitu "asl", "u\_name" maka apabila fitur hasil ekstraksi adalah "asl" atau "u\_name" akan dipetakan dengan karakteristik A1. Contoh hasil pemetaannya ditampilkan pada Tabel 7.

TABLE VII. Contoh hasil ekstraksi

Nama Label	Nama File	Hasil Ekstraksi Fitur Berupa Kata/Gabungan Kata			
		Beauty	Lost	Lost_interest	Love
grooming-training	1.txt	1	1	0	0
grooming-training	2.txt	0	1	1	0
nongrooming-	200.txt	0	0	0	0

training					
<i>nongrooming</i> - training	201.txt	0	1	0	1

Setelah tahap pertama, tahap kedua dengan membuat vektor karakteristik. Pada setiap dokumen, apabila terdapat fitur pada sebuah karakteristik maka karakteristik tersebut akan diberi nilai 1. Seperti pada dokumen “1.txt”, fitur “asl” dan “u\_name” untuk karakteristik A1 nilainya 1 yang berarti dokumen tersebut mengandung karakteristik A1 “*asking profile*”. Sebaliknya, apabila terdapat tidak ada fitur pada sebuah karakteristik maka karakteristik tersebut akan diberi nilai 0.. Seperti pada dokumen “2.txt”, fitur “asl” dan “u\_name” untuk karakteristik A1 nilainya 0 yang berarti dokumen tersebut tidak mengandung karakteristik A1 “*asking profile*”. Contoh hasil vektor karakteristik A1 ditampilkan pada Tabel 8.

TABLE VIII. Contoh hasil pemetaan fitur dengan karakteristik A pada setiap dokumen pada set data pelatihan

		Kata/Gabungan kata					
		Asl	u_name	Phone_number	Cam	u_pic	cute
Karakteristik		1	1	2	3	3	4
Dokumen	1.txt	1	1	1	1	1	1
	2.txt	0	0	0	1	1	1
	3.txt	1	0	0	1	1	1

TABLE IX. Vektor karakteristik A setiap dokumen pada salah satu set data pelatihan

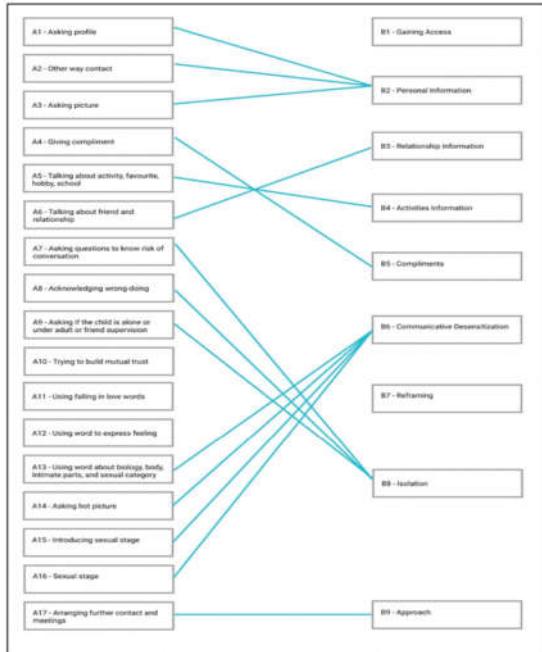
Dokumen	Status	Karakteristik <i>Grooming</i>												
		1	2	3	4	5	6	7	8	...	15	16	17	
1.txt	1	1	1	1	1	1	0	0	0	..	0	0	0	
2.txt	1	0	1	1	1	0	0	0	0	..	0	0	0	
3.txt	1	1	1	1	1	0	0	0	0	..	0	0	0	

Setelah didapatkan hasil vektor karakteristik A dan B maka akan dilakukan klasifikasi pada tahap selanjutnya. Sebelum itu, dilakukan analisis antara teori karakteristik A dan B. Didapatkan terdapat kesamaan antara kedua karakteristik tersebut. Hasil mapping antara kedua karakteristik terdapat pada gambar 4.

Dari hasil analisis diatas dapat terlihat karakteristik no 10,11,12 di karakteristik A tidak ada pada karakteristik B. Sedangkan karakteristik no 1 dan 7 di karakteristik B tidak ada di karakteristik A. Oleh karena itu pada tahap klasifikasi akan

dilakukan klasifikasi dengan menggunakan 4 jenis fitur karakteristik berikut:

- Semua karakteristik A (karakteristik no 1 sampai 17)
- Semua karakteristik B (karakteristik no 1 sampai 9)
- Semua karakteristik A dan karakteristik no 1 dan 7 dari karakteristik B
- Karakteristik no 10, 11, 12 dari karakteristik A dan Semua karakteristik no B



Gbr 4. Pemetaan karakteristik A dan karakteristik B

#### D. Klasifikasi Menggunakan SVM

Klasifikasi menggunakan metode SVM dilakukan dengan beberapa jenis kernel yaitu linear, RBF, polinomial (order 3). Penggunaan beberapa jenis kernel bertujuan untuk mengetahui kernel yang menghasilkan akurasi paling baik. Klasifikasi dilakukan sebanyak 10 kali menggunakan set data pelatihan dan pengujian berbeda.

#### IV. HASIL

Klasifikasi menggunakan metode SVM menggunakan set data pelatihan dan pengujian. Klasifikasi ini dilakukan sebanyak 10 kali percobaan. Setiap percobaan terdiri dari set data pelatihan dan pengujian yang berbeda. Klasifikasi akan menggunakan 4 fitur karakteristik yang berbeda.

Pertama menggunakan semua karakteristik A (karakteristik no 1 sampai 17) didapatkan hasil bahwa SVM linear memberikan paling tinggi dengan akurasi 98.8%, kemudian Polinomial dengan akurasi 98.2% dan RBF dengan akurasi 97.8%.

TABLE X. Hasil klasifikasi percobaan 10 kali dengan menggunakan semua karakteristik A dan karakteristik no 1 dan 7 dari karakteristik B.

Percobaan	RBF	Polinomial	Linear
1	98%	96%	98%
2	98%	100%	100%
3	98%	100%	100%
4	98%	98%	98%
5	96%	98%	100%
6	98%	98%	98%
7	100%	98%	98%
8	94%	98%	98%
9	100%	100%	100%
10	98%	96%	98%
<b>Rata-Rata</b>	<b>97.80%</b>	<b>98.20%</b>	<b>98.80%</b>

Keempat menggunakan karakteristik no 10, 11, 12 dari karakteristik A dan semua karakteristik no B didapatkan hasil bahwa SVM linear memberikan paling tinggi dengan akurasi 97.2%, kemudian Polinomial dengan akurasi 96.4% dan RBF dengan akurasi 95.6%.

TABLE XI. klasifikasi percobaan 10 kali dengan menggunakan semua karakteristik B (karakteristik no 1 sampai 9)

Percobaan	RBF	Polinomial	Linear
1	100%	96%	98%
2	98%	96%	98%
3	96%	100%	100%
4	92%	94%	96%
5	98%	98%	98%
6	98%	96%	98%
7	92%	96%	94%
8	96%	98%	96%
9	88%	92%	92%
10	98%	96%	98%
<b>Rata-Rata</b>	<b>95.60%</b>	<b>96.20%</b>	<b>98.80%</b>

Ketiga menggunakan semua karakteristik A dan karakteristik no 1 dan 7 dari karakteristik B didapatkan hasil bahwa SVM linear memberikan paling tinggi dengan akurasi 99.2%,

kemudian Polinomial dengan akurasi 98.6% dan RBF dengan akurasi 97.6%

TABLE XII. Hasil klasifikasi percobaan 10 kali dengan menggunakan semua karakteristik A dan karakteristik no 1 dan 7 dari karakteristik B

Percobaan	RBF	Polinomial	Linear
1	92%	98%	98%
2	92%	98%	100%
3	100%	100%	100%
4	96%	96%	98%
5	100%	100%	100%
6	100%	98%	98%
7	100%	100%	100%
8	98%	98%	100%
9	98%	98%	98%
10	100%	100%	100%
<b>Rata-Rata</b>	<b>97.60%</b>	<b>98.60%</b>	<b>99.20%</b>

Keempat menggunakan karakteristik no 10, 11, 12 dari karakteristik A dan semua karakteristik no B didapatkan hasil bahwa SVM linear memberikan paling tinggi dengan akurasi 97.2%, kemudian Polinomial dengan akurasi 96.4% dan RBF dengan akurasi 95.6%

Dari 4 jenis fitur karakteristik dapat dilihat bahwa setiap jenis kernel didapatkan rata-rata akurasi klasifikasi menggunakan semua karakteristik B (karakteristik no 1 sampai 9) paling kecil daripada karakteristik lainnya. Ketika menggabungkan karakteristik A dan B yaitu dengan semua karakteristik B dan karakteristik no 10, 11 dan 12 dari karakteristik A ternyata menghasilkan rata-rata akurasi yang lebih tinggi dibandingkan dengan semua karakteristik B saja. Dibandingkan dengan semua karakteristik B, ternyata rata-rata akurasi menggunakan semua karakteristik A lebih tinggi. Bahkan rata-rata akurasi tertinggi didapatkan ketika menggabungkan semua karakteristik A dengan karakteristik no 1 dan 7 dari karakteristik B. Sedangkan berdasarkan jenis kernel yang digunakan untuk klasifikasi dengan 4 jenis fitur karakteristik, kernel linear menghasilkan rata-rata akurasi lebih baik dibandingkan jenis kernel lainnya. Sedangkan kernel RBF menghasilkan rata-rata akurasi lebih rendah dibandingkan jenis kernel lainnya.

TABLE XIII. Hasil Klasifikasi percobaan 10 kali dengan menggunakan karakteristik no 10, 11, 12 dari karakteristik A dan semua karakteristik no B

Percobaan	RBF	Polinomial	Linear
1	100%	100%	100%

2	96%	98%	98%
3	96%	94%	94%
4	100%	100%	100%
5	98%	98%	96%
6	92%	98%	98%
7	94%	92%	92%
8	96%	98%	100%
9	92%	90%	96%
10	92%	96%	98%
<b>Rata-Rata</b>	<b>95.60%</b>	<b>96.40%</b>	<b>97.20%</b>

TABLE XIV. Perbandingan rata-rata akurasi dari 4 jenis fitur karakteristik menggunakan metode SVM dengan 3 jenis kernel (RBF, Polinomial, Linear)

Karakteristik <i>grooming</i> yang digunakan	Jenis Kernel	Rata-Rata Akurasi
Semua karakteristik A (karakteristik no 1 sampai 17)	RBF	97.80%
	Polinomial	98.20%
	Linear	98.80%
Semua karakteristik B (karakteristik no 1 sampai 9)	RBF	95.60%
	Polinomial	96.20%
	Linear	96.80%
Semua karakteristik A (karakteristik no 1 sampai 17) dan karakteristik no 1 dan 7 dari karakteristik B	RBF	97.60%
	Polinomial	98.60%
	Linear	99.20%
Karakteristik no 10, 11, 12 dari karakteristik A dan Semua karakteristik no B (karakteristik no 1 sampai 9)	RBF	95.60%
	Polinomial	96.40%
	Linear	97.20%

## V. KESIMPULAN

Dari berbagai penelitian yang telah dilakukan sebelumnya, dapat diketahui bahwa karakteristik pelaku kejahatan seksual *online* pada anak-anak semakin bervariasi dan terdapat karakteristik yang berbeda-beda tergantung sifat pelaku. Hasil penelitian membuktikan bahwa dengan menggunakan 17 karakteristik tingkat akurasi lebih tinggi daripada menggunakan hanya 9 karakteristik. Selain itu dengan menggabungkan karakteristik yang ada, hasil yang didapatkan menunjukkan bahwa penggunaan karakteristik lebih lengkap akan menghasilkan tingkat akurasi lebih tinggi. Sedangkan berdasarkan jenis kernel yang digunakan untuk klasifikasi, kernel linear menghasilkan rata-rata akurasi lebih baik dibandingkan jenis kernel lainnya. Sedangkan kernel RBF menghasilkan rata-rata akurasi lebih rendah dibandingkan jenis kernel lainnya.

## REFERENSI

- [1] A. Holivia and T. Suratman, "Child Cyber Grooming sebagai Bentuk Modus Baru Cyber Space Crimes," *Bhirawa Law*, vol. 2, no. 1, pp. 1-13, May 2021.
- [2] "UNICEF Indonesia," UNICEF, 2023. [Online]. Available: <https://www.unicef.org/indonesia/child-protection/jagabareng>.
- [3] Z. Zuo, J. Li, P. Andreson, L. Yang and N. Naik, "Grooming Detection using Fuzzy-Rough Feature Selection and Text Classification," *IEEE World Congress on Computational*, July 2018.
- [4] L. H. M. Nasir, Z. Saaya and M. K. Baharun, "Identifying Online Sexual Grooming Context in Social Media Using Neural Networks Classification Technique," *Journal of Advanced Computing Technology and Application*, vol. 4, no. 2, pp. 33-42, 2 November 2022.
- [5] I. Ernawati, "NAÏVE BAYES CLASSIFIER DANSUPPORT VECTOR MACHINE SEBAGAI ALTERNATIF SOLUSI UNTUK TEXT MINING," *Jurnal Teknologi Informasi dan Pendidikan*, vol. 12, no. 2, 12 December 2019.
- [6] H. Pranoto, F. E. Gunawan and B. Soewito, "Logistic Models for Classifying Online Grooming Conversation," *Procedia Computer Science*, vol. 59, pp. 357-365, 2015.
- [7] F. E. Gunawan, L. Ashianti and N. Sekishita, "A Simple Classifier for Detecting Online Child Grooming Conversation," *TELKOMNIKA*, vol. 16, no. 3, pp. 1239-1248, June 2018.
- [8] A. Ikhsanudin, "Detik News," 26 July 2019. [Online]. Available: <https://news.detik.com/berita/d-4640789/kpai-sebut-kasus-kekerasan-seksual-anak-meningkat-akibat-pengaruh-digital>
- [9] "Kompas.com," Kompas.com, 29 November 2021. [Online]. Available: <https://megapolitan.kompas.com/read/2021/11/29/19563311/lpsk-kasus-pelecehan-seksual-terhadap-anak-naik-tajam-diduga-efek-pandemi?page=all>.
- [10] V. Kozlovskiy and L. Pankratova, "Age and Sexuality Online: Trends, Risks and Perspectives," *Proceedings of the 2018 2nd International Conference on Education Science and Economic Management (ICESEM 2018)*, August 2018.
- [11] I. Mcghee, J. Bayzick, A. Kontostathis, L. Edwards, A. McBride and E. Jakubowski, "Learning to Identify Internet Sexual Predation," *International Journal of Electronic Commerce*, vol. 15, pp. 103-122, April 2011.
- [12] D. M. Abdullah and A. M. Abdulazeez, "Machine Learning Applications based on SVM Classification A Review," *Qubahan Academic Journal*, vol. 1, no. 2, pp. 81-90, 28 April 2021.
- [13] A. Kurani, P. Doshi, A. Vakharia and M. Shah, "A Comprehensive Comparative Study of Artificial Neural Network (ANN) and Support Vector Machines (SVM) on Stock Forecasting," *Springer Link*, pp. 183-208, 2 June 2021.
- [14] D. A. Otchere, T. O. A. Ganat, R. Gholami and S. Ridha, "Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models," *Journal of Petroleum Science and Engineering*, vol. 200, pp. 1-20, May 2021.
- [15] R. Hu, X. Zhu, Y. Zhu and J. Gan, "Robust SVM with adaptive graph learning," pp. 1945-1968, 27 December 2019.
- [16] Q. Ye, P. Huang, Z. Zhang, Y. Zheng, L. Fu and W. Yang, "Multiview Learning With Robust Double-Sided Twin SVM," *IEEE Transactions on Cybernetics*, vol. 52, no. 12, pp. 12745-12758, December 2022.

## NOMENKLATUR

Nomenklatur disertai arti dari semua persamaan matematika ataupun nomenklatur lain di dalam artikel, dituliskan pada bagian ini.

- $T_r$  Daftar kata atau *wordlist*
- $d_m$  Hasil ekstraksi fitur pada setiap dokumen
- $M$  Jumlah seluruh dokumen
- $w_{Ti,m}$  Bernilai 1 apabila terdapat fitur  $i$  (kata atau *wordlist*) dalam percakapan tersebut dan 0 jika tidak ada dalam percakapan tersebut.
- $C_m$  Vektor Karakteristik setiap dokumen
- $c_{m,1}$  Nilai yang menandakan ada tidaknya karakteristik  $j$  pada dokumen  $m$ .
- $j$  Karakteristik proses pendekatan terhadap anak dibawah umur untuk tujuan seksual.

## BIODATA PENULIS

### Livia Ashianti

Seorang tenaga pengajar dari Universitas Bina Nusantara untuk jurusan Teknik Informatika. Sudah mengajar sejak tahun 2015.

### Kanyadian Idananta

Seorang tenaga pengajar dari Universitas Bina Nusantara untuk jurusan Teknik Informatika. Sudah mengajar sejak tahun 2015.