BAB IV
HASIL DAN PEMBAHASAN

4.1 *Preprocessing Data*

Setelah dilakukan perancangan maka dilakukan implementasi dari tiap Langkah-langkah tersebut yang dimulai dari *data collection*, *data cleaning* dan *data labeling*.

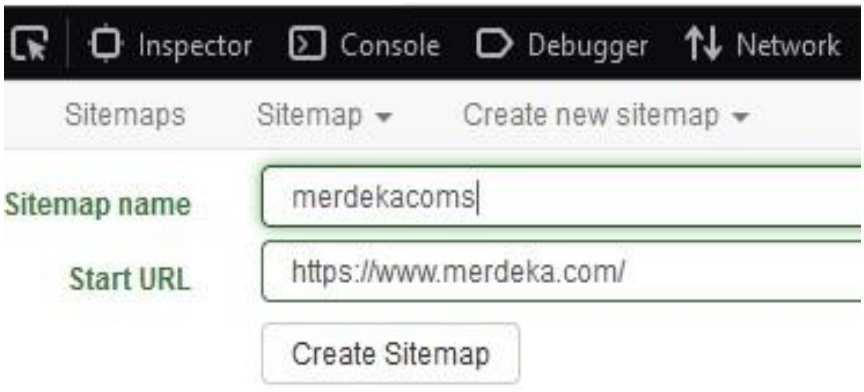
4.1.1 Hasil *Data Collection*

Berdasarkan penjelasan yang diberikan pada sub bab sebelumnya, pengumpulan *data collection* atau *scraping* data dilakukan melalui beberapa tahap berikut:

1. Menentukan *Website* dan Membuat Proyek Baru

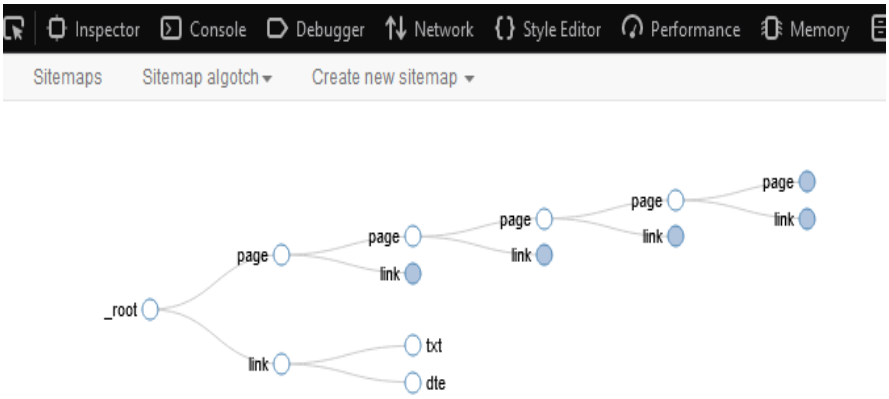
   Menentukan *website* atau *blog* yang akan di *scraping* kemudian membuat proyek baru yang berisikan *link website*, nama projek, dan *pattern* dari *scraping data*.

   Fungsi dari *link website* digunakan untuk mengambil data dari *website* yang dituju dan nama projek diberikan yang bertujuan untuk meminimalisir terjadinya duplikat pengambilan data *website* dari *website* yang di *scraping*. Gambar 4.1 berikut merupakan hasil dari membuat proyek baru yang diberi nama dan *link*.



Gambar 4.1 Membuat Proyek Baru

Kemudian *pattern* dari proses *scraping* ini adalah dengan mengambil *link*, *text* (konten dari *website*), dan tanggal *publish website* yang digunakan sebagai syarat dalam penelitian ini dimana data yang akan digunakan merupakan *website* yang di *posting* atau diterbitkan secara *online* dari tahun 2019-2022. Gambar 4.2 berikut merupakan *pattern* dari pengambilan data dari *website*



Gambar 4.2 *Pattern* Pengambilan Data

2. Memulai *Scraping Data*

   Untuk melihat apakah proses *scraping* data sudah dimulai dapat dilihat pada *window* baru *web browser* yang secara otomatis terbuka. Hasil dari memulai *scraping* data dapat dilihat pada gambar 4.3 berikut:

Gambar 4.3 Mulai Proses *Scraping*

3. *Monitoring*

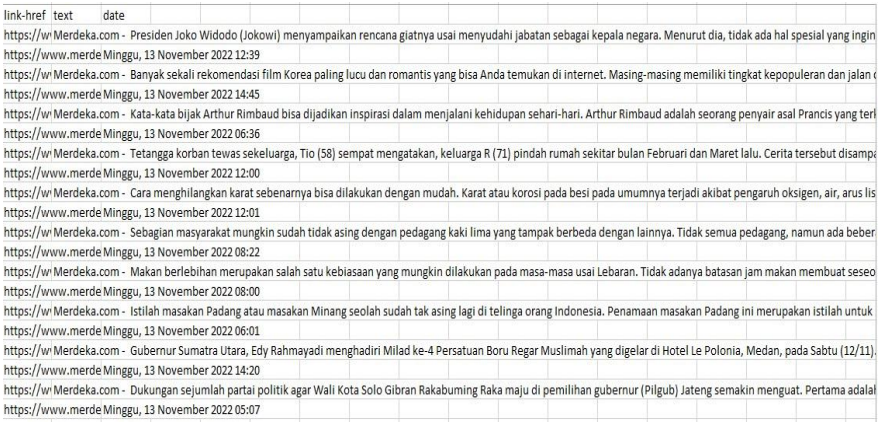
   Monitoring data selama proses *scraping* berlangsung dilakukan untuk memastikan bahwa *pattern* atau format data yang diambil sudah sesuai yaitu *link*, *text*, dan tanggal *posting*, hasil *monitoring* dapat dilihat pada gambar 4.4 berikut:

https://www.merdeka.com/uang/masyarakat-diminta-waspada-minyakita-palsu-begini-cara-membedakannya.html	"Ini buat pembelajaran bersama, kami temukan ini di Sragen," kata Direktur Jenderal Perlindungan Konsumen dan Tertib Niaga Kemendag Veri Anggrijono, saat pengawasan distribusi Minyakita di Pasar Gayamsari, Semarang dikutip dari Antara, Jumat (17/2).	
https://www.merdeka.com/uang/masyarakat-diminta-waspada-minyakita-palsu-begini-cara-membedakannya.html		Jumat, 17 Februari 2023 20:00

Gambar 4.4 *Monitoring* Proses *Scraping*

4. Simpan Hasil *Scraping*

   Setelah prosess *scraping* selesai maka hasil *scraping* disimpan ke dalam format *file csv*. Hasil *scraping* data dapat dilihat pada gambar 4.5 berikut:



link-href	text	date
https://w	Merdeka.com -	Presiden Joko Widodo (Jokowi) menyampaikan rencana giatnya usai menyudahi jabatan sebagai kepala negara. Menurut dia, tidak ada hal spesial yang ingin
https://www.merde		Minggu, 13 November 2022 12:39
https://w	Merdeka.com -	Banyak sekali rekomendasi film Korea paling lucu dan romantis yang bisa Anda temukan di internet. Masing-masing memiliki tingkat kepopuleran dan jalan
https://www.merde		Minggu, 13 November 2022 14:45
https://w	Merdeka.com -	Kata-kata bijak Arthur Rimbaud bisa dijadikan inspirasi dalam menjalani kehidupan sehari-hari. Arthur Rimbaud adalah seorang penyair asal Prancis yang terl
https://www.merde		Minggu, 13 November 2022 06:36
https://w	Merdeka.com -	Tetangga korban tewas sekeluarga, Tio (58) sempat mengatakan, keluarga R (71) pindah rumah sekitar bulan Februari dan Maret lalu. Cerita tersebut disampa
https://www.merde		Minggu, 13 November 2022 12:00
https://w	Merdeka.com -	Cara menghilangkan karat sebenarnya bisa dilakukan dengan mudah. Karat atau korosi pada besi pada umumnya terjadi akibat pengaruh oksigen, air, arus lis
https://www.merde		Minggu, 13 November 2022 12:01
https://w	Merdeka.com -	Sebagian masyarakat mungkin sudah tidak asing dengan pedagang kaki lima yang tampak berbeda dengan lainnya. Tidak semua pedagang, namun ada beber
https://www.merde		Minggu, 13 November 2022 08:22
https://w	Merdeka.com -	Makan berlebihan merupakan salah satu kebiasaan yang mungkin dilakukan pada masa-masa usai Lebaran. Tidak adanya batasan jam makan membuat seseo
https://www.merde		Minggu, 13 November 2022 08:00
https://w	Merdeka.com -	Istilah masakan Padang atau masakan Minang seolah sudah tak asing lagi di telinga orang Indonesia. Penamaan masakan Padang ini merupakan istilah untuk
https://www.merde		Minggu, 13 November 2022 06:01
https://w	Merdeka.com -	Gubernur Sumatra Utara, Edy Rahmayadi menghadiri Milad ke-4 Persatuan Boru Regar Muslimah yang digelar di Hotel Le Polonia, Medan, pada Sabtu (12/11).
https://www.merde		Minggu, 13 November 2022 14:20
https://w	Merdeka.com -	Dukungan sejumlah partai politik agar Wali Kota Solo Gibran Rakabuming Raka maju di pemilihan gubernur (Pilgub) Jateng semakin menguat. Pertama adalal
https://www.merde		Minggu, 13 November 2022 05:07

Gambar 4.5 Hasil *Scraping*

Berdasarkan gambar 4.5 di atas hasil *scraping* dari *domain website* https://www.merdeka.com/ diperoleh jumlah data sebanyak 98 data *website* dengan setiap data yang diambil dari *website* tersebut memiliki *link* masing-masing yang berbeda untuk setiap konten atau postingan yang ada dalam *domain link website* tersebut.

Sebagai contoh dalam *website* https://www.merdeka.com/ ketika proses *scraping* telah selesai dilakukan maka *link website* yang akan digunakan atau disimpan ke dalam *dataset* menjadi seperti berikut:

https://www.merdeka.com/peristiwa/cita-cita-jokowi-usai-purnatugas-presiden-jadi-rakyat-biasa-aktif-lingkungan-hidup.html

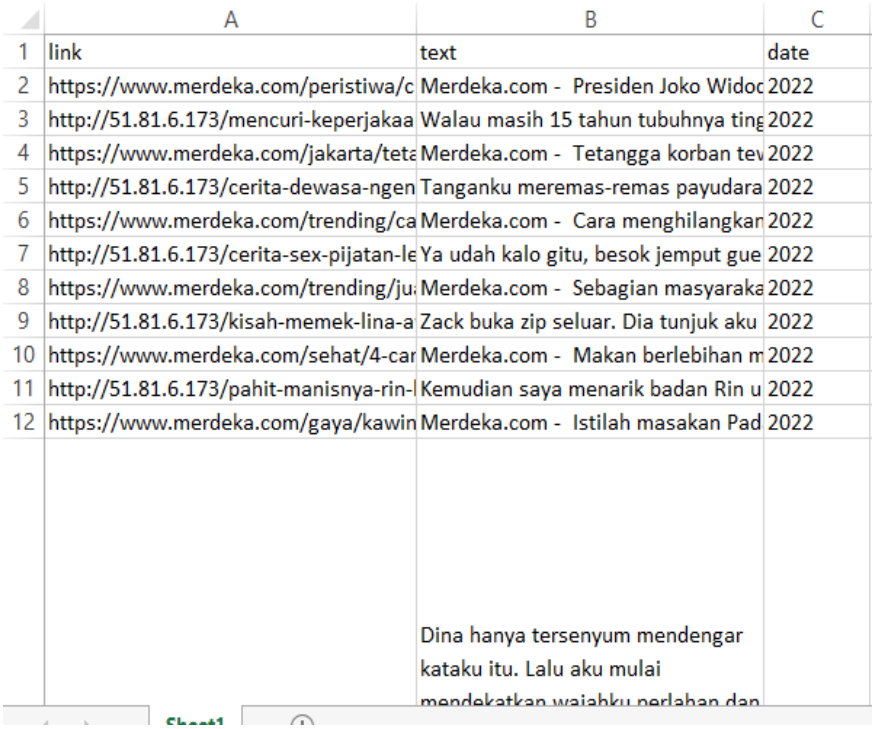
https://www.merdeka.com/jatim/20-film-korea-paling-lucu-dan-romantis-tonton-bersama-orang-terdekat-kln.html

https://www.merdeka.com/jateng/25-kata-kata-bijak-arthur rimbaud-inspiratif-dan-penuh-makna-kln.html dan sterusnya sampai dengan konten data yang terakhir, secara singkat cara kerja dari proses *scraping* ini adalah dengan menggunakan *domain website* seperti https://www.merdeka.com/ sebagai *link* utama yang digunakan untuk memulai pengambilan data atau konten-konten yang ada didalam *domain website* tersebut.

Dapat dilihat pada gambar di atas bahwa data yang didapatkan dari hasil *scraping* masih belum bisa digunakan karena masih terdapat kolom kosong pada hasil *scraping* tersebut serta harus di *filter* kembali tahun *posting* atau *publish* dari *website* tersebut untuk disesuaikan dengan syarat penggunaan *dataset* pada penelitian ini, oleh karena itu data yang dikumpulkan dari hasil *scraping* tersebut diolah kembali sehingga dapat digunakan untuk proses pengklasifikasian dan pembuatan model *Machine Learning*.

Dalam proses finalisasi ini kumpulan-kumpulan data yang disimpan sebagai hasil *scraping* apabila diproses kembali sebagai finalisasi *dataset* maka setiap jumlah *file* data berdasarkan satu *domain website* akan mengalami perubahan dimana jumlah data dari satu domain *website* tersebut akan berkurang, seperti contohnya pada *domain website* https://www.merdeka.com/ hasil dari *scraping data* sebanyak 98 data namun setelah diproses kembali sebagai finalisasi *dataset* maka data yang dapat digunakan dari *website* tersebut hanya sebanyak 45 data, hal ini disebabkan karena selama proses *filtering* data peneliti tidak hanya berfokus pada *link*, teks (konten) dan tahun melainkan juga memperhatikan isi konten teks dari *website-website* tersebut, apabila isi teks dari *website* memiliki ukuran yang kecil atau sedikit maka data tersebut tidak akan digunakan sebagai *dataset*. Dalam hal ini perlu diketahui bahwa dalam satu *domain website* yang telah di *scraping* peneliti dapat memperoleh ratusan ataupun ribuan data dari satu *domain website* tersebut.

Gambar dari hasil pengolahan *dataset* tersebut dapat dilihat pada gambar berikut 4.6



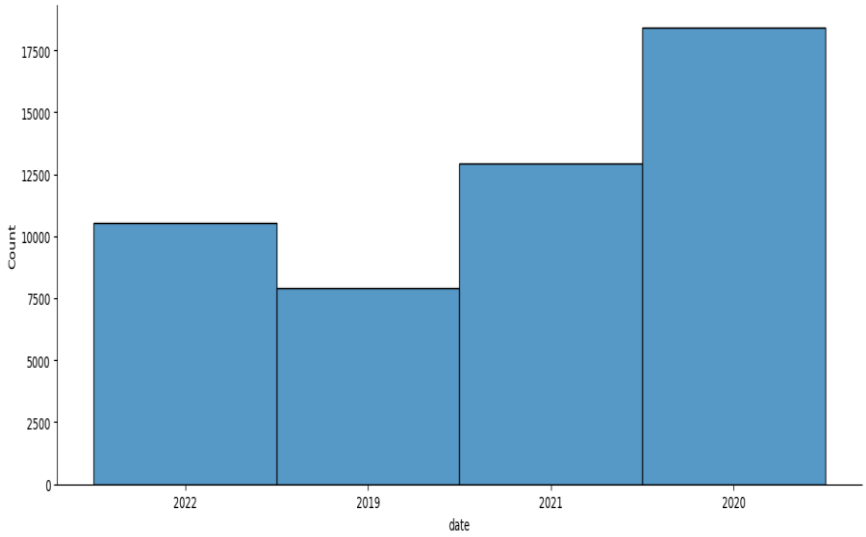
	A	B	C
1	link	text	date
2	https://www.merdeka.com/peristiwa/c	Merdeka.com - Presiden Joko Widoc	2022
3	http://51.81.6.173/mencuri-keperjakaa	Walau masih 15 tahun tubuhnya ting	2022
4	https://www.merdeka.com/jakarta/teta	Merdeka.com - Tetangga korban tev	2022
5	http://51.81.6.173/cerita-dewasa-ngen	Tanganku meremas-remas payudara	2022
6	https://www.merdeka.com/trending/ca	Merdeka.com - Cara menghilangkan	2022
7	http://51.81.6.173/cerita-sex-pijatan-le	Ya udah kalo gitu, besok jemput gue	2022
8	https://www.merdeka.com/trending/ju	Merdeka.com - Sebagian masyaraka	2022
9	http://51.81.6.173/kisah-memek-lina-a	Zack buka zip seluar. Dia tunjuk aku	2022
10	https://www.merdeka.com/sehat/4-car	Merdeka.com - Makan berlebihan m	2022
11	http://51.81.6.173/pahit-manisnya-rin-	Kemudian saya menarik badan Rin u	2022
12	https://www.merdeka.com/gaya/kawin	Merdeka.com - Istilah masakan Pad	2022
		Dina hanya tersenyum mendengar kataku itu. Lalu aku mulai mendekatkan wajahku perlahan dan	

Gambar 4.6 Finalisasi Pengolahan *Dataset*

Hasil finalisasi dari pengolahan *dataset* tersebut dapat dilihat pada gambar 4.6 di atas dimana terdapat 3 variabel kelas dalam *dataset* yaitu *link*, *text* dan *date*. Pengolahan atau finalisasi *dataset* ini bertujuan untuk memudahkan proses selanjutnya yaitu *data cleaning*.

Dari hasil finalisasi *dataset* kemudian ditampilkan frekuensi atau jumlah data yang diambil berdasarkan tahun *publish website* yaitu selama empat tahun dimulai dari tahun 2019 – 2022.

Gambar jumlah data berdasarkan tahun *publish website* dapat dilihat pada gambar 4.7 berikut:



Gambar 4.7 Frekuensi Jumlah *Dataset* Hasil *Scraping*

Berdasarkan gambar 4.7 di atas dapat disimpulkan bahwa terdapat 49.743 ribu jumlah data dari *dataset* hasil *scraping data* dengan jumlah data pada tahun 2022 sebanyak 10.536 ribu data *website*, tahun 2019 terdapat 7.901 data *website*, tahun 2021 sebanyak 12.920 ribu data *website* dan tahun 2020 sebanyak 18.392 ribu data *website*, sehingga secara keseluruhan jumlah *dataset* yang digunakan dalam penelitian ini berjumlah 49.743 ribu data *website*.

4.1.2 Hasil *Data Cleaning*

Data Cleaning dilakukan dengan tujuan membersihkan dan mempersiapkan data teks untuk analisis atau pemrosesan selanjutnya.

Sebelum melanjutkan proses *cleaning* data, tabel 4.1 berikut adalah tabel sampel data yang digunakan pada proses *cleaning*.

Tabel 4.1 Sampel Data

ID	Text
D1	"Bagikan

ID	Text
	tweet ########## Cara Membuat Best Nine Instagram Tanpa Aplikasi 2021 – Tentunya sekarang untuk membuat kolase foto tidak begitu sulit dan juga tidak membutuhkan aplikasi loh. Karena di artikel kali ini, kita akan bagikan tips cara membuat best nine instagram tanpa aplikasi terbaru di tahun 2021. Penasaran caranya seperti apa, silakan baca terus artikel kami sampai selesai agar teman-teman paham proses pembuatan kolase foto menampilkan 9 kota foto. A B C D DAFTAR ISI

ID	Text
	Apa Itu Best Nine?Best Nine & Top Nine Instagram 2020Best Nine Instagram Tanpa Aplikasi 2021Akhir Kata Apa Itu Best Nine? Best Nine adalah kolase foto dengan menampilkan foto foto terdiri dari 9 kotak dalam susunan 3 kali 3 (gride). Sembilan foto tersebut di pilih berdasarkan foto terbaik yang di unggah oleh penggunannya sepanjang tahun. Proses pembuatan best nine cukup mudah dan tentunya ini tidak menggunakan aplikasi tambahan. Baca Juga : Rekomendasi Hp Oppo Dengan Kamera Terbaik Dan Harga Super Terjangkau Cuma 2 jutaanBest Nine & Top Nine Instagram 2020 Untuk best nine instagram di tahun 2020 bisa kalian buat di 2 situs ternama yang dapat kalian kunjungi link nya di bawah ini! Situs tersebut bernama Top Nine dan Best Nine. Top Nine: https://creatorkit.com/top-nine-best-of-2020/ @ Best Nine: https://bestnine.net/en Kedua situs di atas untuk membuat kolase foto yang sudah kamu unggah sepanjang tahun 2020. Cara membuatnya cukup mudah, karena disini kamu hanya perlu memasukan foto foto kamu ke dalam situs tersebut agar dapat membuat kolase foto yang terdiri dari 9 kotak dan 3×3 (grid). Baca Juga : Tips: How to Develop Android APP Menggunakan Android StudioUntuk kedua situs di atas berbeda cara penggunaanya, jika kamu menggunakan top nine maka akan di minta memasukan user

ID	Text
	IG kamu dan alamat email. Namun jika kamu menggunakan situs Best Nine, kamu akan di minta mengisi username ig saja tanpa meminta email instagram kamu. Untuk keamanan privasi di dua situs ini sangat aman, jadi gak perlu takut tentang akun kamu. Best Nine Instagram Tanpa Aplikasi 2021 Cara pembuatan Best Nine instagram tanpa aplikasi 2021 akan segera hadir, karena ini masih merupakan awal tahun 2021 jadi di tunggu saja update *website* untuk membuat best nine kolase foto terbaik anda di tahun ini. Baca Juga : 3 Cara Melihat Instagram di Private Tanpa FollowJika kamu ingin membuat kolase foto di tahun sebelumnya dapat mengunjungi ke dua situs di atas. Karena sangat mudah di gunakan dan tentunya ini gratis tanpa menggunakan aplikasi tambahan. Akhir Kata Jadi itulah sedikit tutorial dari kami tentang Cara Membuat Best Nine instagram tanpa aplikasi 2021, semoga artikel yang kami bagikan kali ini bermanfaat bagi teman-teman yang membacanya. Jangan lupa share artikel ini sebagai dukungan teman-teman untuk kemajuan web topgloabl1.com"
D2	Aku beristirahat sebentar. Lalu aku bertanya pada Fani, “Fan.. Aku boleh nggak, masukin penisku ke vagina kamu?” Fani tampak berpikir sejenak. Namun dalam kondisi horny seperti itu, pikirannya tentu agak kacau. Fani akhirnya mengangguk lemas. Perlahan-lahan, aku merebahkan badannya di ranjang dan menaruh kedua kakinya di bahuku. Aku mulai menyentuhkan kepala penisku ke bibir liang senggamanya. Lalu aku perlahan-lahan memasukkan batang kejantananku ke dalam liang senggamanya. Aduh.. Sulitnya batang

ID	Text
	kejantananku masuk ke liang senggamanya. Walaupun liang senggamanya sudah basah, namun liang senggamanya masih sangat rapat. Ketika batang kejantananku perlahan-lahan masuk, Fani mulai mengerang kesakitan. Aku mencoba menenangkannya. Aku pun merasakan sakit, karena batang kejantananku seperti ditekan oleh liang senggamanya yang sempit.
D3	Hampir 2 jam PakSu memantat aku.. Habis sengal-sengal seluruh badan aku.. Cipap aku tak yah cerita la.. Rasa kembang semacam aje.. Jalan pun rasa lain aje lepas tu.. Rasa menyesal ada juga pasal aku dah tak virgin lagi. Dah kena robek kat PakSu aku. Tapi rasa menyesal tu rasa berbaloi juga dengan nikmat yang aku kecapi. Aku tak sangka sex begitu sedap. Patutla orang Nak kahwin sangat. Bagi aku at least aku dah rasa, Nak tunggu kahwin lambat lagi, paling awal pun ayah aku bagi kahwin umur 21. Tak sanggup rasanya Nak tunggu 7 tahun lagi. Cipap aku ni asyik terkemut-kemut aje bila tengok balak hensem. Last sekali sebagai upacara penutup PakSu suruh aku kulum batang dia.. Aku mengikut aje walaupun tak pandai aku cuba juga. Separuh aje batang PakSu dapat aku kulum. Kira okay jugaklah untuk yang tak ada pengalaman macam aku ni. Itulah first time aku tengok batang lelaki dewasa, selalunya aku tengok konek adik aku yang sebesar ibu jari je. Terkejut juga aku bila tengok batang PakSu yang hampir sebesar lengan aku. Panjangnya lebih kurang 6 inci saja tapi agak besar. Kepala batangnya pun besar macam cendawan.. Suka betul aku bila tengok kepala batang PakSu mengembang dan berkilat bila kena kulum. Masa aku kulum batang PakSu ramas-ramas buah dada aku.. Sekali-sekali jari jahat dia korek lubang dubur aku. Pengotor betul PakSu aku ni. Ada ka dia kata lubang dubur aku cute. Kalau kata cipap aku cute logik juga.. Pasal cipap aku belum ada banyak bulu.. Ada

ID	Text
	bulu pahat saja.. Nipis dan halus. Nampak bersih dan cute.. Lebih kurang pukul 2 barulah MakSu dan Atie balik.. Sempatlah aku aku mandi dan berehat lepas kena kongkek kat PakSu.
D4	nan
D5	PADANG, KOMPAS.com - Senator asal Sumatera Barat, Emma Yohanna, yang mengungguli suara Jokowi-Ma'ruf Amin pada Pemilu 2019 lalu berniat maju kembali sebagai calon anggota Dewan Perwakilan Daerah (DPD). Incumbent tiga periode itu menyerahkan syarat dukungan 2.000 dukungan ke Komisi Pemilihan Umum (KPU) Sumbar, Kamis (22/12/2022). Pada Pemilu 2019 lalu, Emma berhasil mendapatkan 531.104 suara. Jumlah itu menjadi yang tertinggi dibanding 20 calon anggota DPD daerah Pemilihan Sumbar. Baca juga: Tangisan Iringi Pemakaman Aiptu Ruslan, Polisi yang Tewas Ditikam Bawahannya di SPN Polda Riau Sementara, pasangan Capres Joko Widodo-Ma'ruf Amin di Sumbar mendapatkan 407.761 suara. "Hari ini saya bersama tim menyerahkan syarat dukungan suara untuk maju jadi calon DPD RI," kata Emma usai menyerahkan syarat dukungan. Emma menyebut, ada 4.900 dukungan yang tersebar di 19 kabupaten dan kota di Sumbar. Jumlah itu sudah melampaui syarat minimal 2.000 dukungan yang tersebar di 10 kabupaten dan kota di Sumbar. Baca juga: Kasus Penipuan Rp 1,1 Miliar Berkedok Investasi Pariwisata Sumbar, Polisi Periksa 4 Orang Saksi Emma mengaku sengaja datang ke KPU Sumbar bertepatan dengan peringatan Hari Ibu 2022. "Kita membawa tim perempuan sebanyak 22 orang dan kemudian membawa bunga yang dibagi-bagikan kepada masyarakat," kata Emma. Untuk target, Emma berharap bisa terpilih kembali jadi anggota DPD RI. "Kalau dulu suara sampai 500.000 lebih dan mengalahkan suara Pak Jokowi-Ma'ruf Amin. Sekarang ya mudah-

ID	Text
	mudahan bisa kembali terpilih lagi," kata Emma. Sementara itu, Komisioner KPU Sumbar, Gebril Daulay yang menerima syarat dukungan dari Emma Yohanna menyebutkan, hingga sekarang baru satu orang yang datang menyerahkan persyaratan tersebut. "Buk Emma yang pertama. Sementara peminat ada 30 orang yang mendaftar di akun Silon," kata Gebril. Gebril mengatakan, KPU memberi kesempatan hingga 29 Desember 2022 untuk pihak yang ingin menyerahkan persyaratan dukungan. "Persyaratannya yaitu minimal 2.000 dukungan yang tersebar di 10 kabupaten dan kota," kata Gebril. Dapatkan update berita pilihan dan breaking news setiap hari dari Kompas.com. Mari bergabung di Grup Telegram "Kompas.com News Update", caranya klik link https://t.me/kompascomupdate, kemudian join. Anda harus install aplikasi Telegram terlebih dulu di ponsel.

Terdapat beberapa hal yang harus perhatikan yaitu, *ID* merupakan penomoran untuk setiap dokumen yang ada dalam *dataset* untuk memudahkan penulis dalam menulis naskah penelitian ini, terdapat lima dokumen data dalam *dataset* sehingga diberi nomor *ID* sebagai berikut D1 untuk dokumen satu, D2 untuk dokumen dua, D3 untuk dokumen tiga, D4 untuk dokumen empat dan D5 untuk dokumen lima, selain pemberian nomor *ID* dapat dilihat pada tabel 4.1 di atas dimana pada D4 terdapat satu baris teks yang memiliki *values* sebagai "*nan*" hal ini diasumsikan bahwa D4 merupakan dokumen yang tidak memiliki nilai atau kosong dalam kumpulan *dataset* yang nantinya akan dihapus atau dihilangkan dalam tahap *cleaning.* Berikut merupakan beberapa proses yang dilakukan dalam tahap *cleaning* data:

1. *Case Folding*

   Implementasi *case folding* dapat dilihat pada tabel 4.2 berikut:

Tabel 4.2 Hasil *Case Folding*

ID	Text
D1	"bagikan tweet ########## cara membuat best nine instagram tanpa aplikasi 2021 – tentunya sekarang untuk membuat kolase foto tidak begitu sulit dan juga tidak membutuhkan aplikasi loh. karena di artikel kali ini, kita akan bagikan tips cara membuat best nine instagram tanpa aplikasi terbaru di tahun 2021. penasaran caranya seperti apa, silakan baca terus artikel kami sampai selesai agar teman-teman paham proses pembuatan kolase foto menampilkan 9 kota foto. a b c d

ID	Text
	daftar isi apa itu best nine?best nine & top nine instagram 2020best nine instagram tanpa aplikasi 2021akhir kata apa itu best nine? best nine adalah kolase foto dengan menampilkan foto foto terdiri dari 9 kotak dalam susunan 3 kali 3 (gride). sembilan foto tersebut di pilih berdasarkan foto terbaik yang di unggah oleh penggunannya sepanjang tahun. proses pembuatan best nine cukup mudah dan tentunya ini tidak menggunakan aplikasi tambahan. baca juga : rekomendasi hp oppo dengan kamera terbaik dan harga super terjangkau cuma 2 jutaanbest nine & top nine instagram 2020 untuk best nine instagram di tahun 2020 bisa kalian buat di 2 situs ternama yang dapat kalian kunjungi link nya di bawah ini! situs tersebut bernama top nine dan best nine. top nine: https://creatorkit.com/top-nine-best-of-2020/ @ best nine: https://bestnine.net/en kedua situs di atas untuk membuat kolase foto yang sudah kamu unggah sepanjang tahun 2020. cara membuatnya cukup mudah, karena disini kamu

ID	Text
	hanya perlu memasukan foto foto kamu ke dalam situs tersebut agar dapat membuat kolase foto yang terdiri dari 9 kotak dan 3×3 (grid). baca juga : tips: how to develop android app menggunakan android studiountuk kedua situs di atas berbeda cara penggunaanya, jika kamu menggunakan top nine maka akan di minta memasukan user ig kamu dan alamat email. namun jika kamu menggunakan situs best nine, kamu akan di minta mengisi username ig saja tanpa meminta email instagram kamu. untuk keamanan privasi di dua situs ini sangat aman, jadi gak perlu takut tentang akun kamu. best nine instagram tanpa aplikasi 2021 cara pembuatan best nine instagram tanpa aplikasi 2021 akan segera hadir, karena ini masih merupakan awal tahun 2021 jadi di tunggu saja update *website* untuk membuat best nine kolase foto terbaik anda di tahun ini. baca juga : 3 cara melihat instagram di private tanpa followjika kamu ingin membuat kolase foto di tahun sebelumnya dapat mengunjungi ke dua situs di atas. karena sangat mudah di gunakan dan tentunya ini gratis tanpa menggunakan aplikasi tambahan. akhir kata jadi itulah sedikit tutorial dari kami tentang cara membuat best nine instagram tanpa aplikasi 2021, semoga artikel yang kami bagikan kali ini bermanfaat bagi teman-teman yang membacanya. jangan lupa share artikel ini sebagai

ID	Text
	dukungan teman-teman untuk kemajuan web topgloabl1.com"
D2	aku beristirahat sebentar. lalu aku bertanya pada fani, “fan.. aku boleh nggak, masukin penisku ke vagina kamu?” fani tampak berpikir sejenak. namun dalam kondisi horny seperti itu, pikirannya tentu agak kacau. fani akhirnya mengangguk lemas. perlahan-lahan, aku merebahkan badannya di ranjang dan menaruh kedua kakinya di bahuku. aku mulai menyentuhkan kepala penisku ke bibir liang senggamanya. lalu aku perlahan-lahan memasukkan batang kejantananku ke dalam liang senggamanya. aduh.. sulitnya batang kejantananku masuk ke liang senggamanya. walaupun liang senggamanya sudah basah, namun liang senggamanya masih sangat rapat. ketika batang kejantananku perlahan-lahan masuk, fani mulai mengerang kesakitan. aku mencoba menenangkannya. aku pun merasakan sakit, karena batang kejantananku seperti ditekan oleh liang senggamanya yang sempit.
D3	hampir 2 jam paksu memantat aku.. habis sengal-sengal seluruh badan aku.. cipap aku tak yah cerita la.. rasa kembang semacam aje.. jalan pun rasa lain aje lepas tu.. rasa menyesal ada juga pasal aku dah tak virgin lagi. dah kena robek kat paksu aku. tapi rasa menyesal tu rasa berbaloi juga dengan nikmat yang aku kecapi. aku tak sangka sex begitu sedap. patutla orang nak kahwin sangat. bagi aku at least aku dah rasa, nak tunggu kahwin lambat lagi, paling awal pun ayah aku bagi kahwin umur 21. tak sanggup rasanya nak tunggu 7 tahun lagi. cipap aku ni asyik

ID	Text
	terkemut-kemut aje bila tengok balak hensem. last sekali sebagai upacara penutup paksu suruh aku kulum batang dia.. aku mengikut aje walaupun tak pandai aku cuba juga. separuh aje batang paksu dapat aku kulum. kira okay jugaklah untuk yang tak ada pengalaman macam aku ni. itulah first time aku tengok batang lelaki dewasa, selalunya aku tengok konek adik aku yang sebesar ibu jari je. terkejut juga aku bila tengok batang paksu yang hampir sebesar lengan aku. panjangnya lebih kurang 6 inci saja tapi agak besar. kepala batangnya pun besar macam cendawan.. suka betul aku bila tengok kepala batang paksu mengembang dan berkilat bila kena kulum. masa aku kulum batang paksu ramas-ramas buah dada aku.. sekali-sekali jari jahat dia korek lubang dubur aku. pengotor betul paksu aku ni. ada ka dia kata lubang dubur aku cute. kalau kata cipap aku cute logik juga.. pasal cipap aku belum ada banyak bulu.. ada bulu pahat saja.. nipis dan halus. nampak bersih dan cute.. lebih kurang pukul 2 barulah maksu dan atie balik.. sempatlah aku aku mandi dan berehat lepas kena kongkek kat paksu.
D4	nan
D5	padang, kompas.com - senator asal sumatera barat, emma yohanna, yang mengungguli suara jokowi-ma'ruf amin pada pemilu 2019 lalu berniat maju kembali sebagai calon anggota dewan perwakilan daerah (dpd). incumbent tiga periode itu menyerahkan syarat dukungan 2.000 dukungan ke komisi pemilihan umum (kpu) sumbar, kamis (22/12/2022). pada pemilu 2019 lalu, emma berhasil

ID	Text
	mendapatkan 531.104 suara. jumlah itu menjadi yang tertinggi dibanding 20 calon anggota dpd daerah pemilihan sumbar. baca juga: tangisan iringi pemakaman aiptu ruslan, polisi yang tewas ditikam bawahannya di spn polda riau sementara, pasangan capres joko widodo-ma'ruf amin di sumbar mendapatkan 407.761 suara. "hari ini saya bersama tim menyerahkan syarat dukungan suara untuk maju jadi calon dpd ri," kata emma usai menyerahkan syarat dukungan. emma menyebut, ada 4.900 dukungan yang tersebar di 19 kabupaten dan kota di sumbar. jumlah itu sudah melampaui syarat minimal 2.000 dukungan yang tersebar di 10 kabupaten dan kota di sumbar. baca juga: kasus penipuan rp 1,1 miliar berkedok investasi pariwisata sumbar, polisi periksa 4 orang saksi emma mengaku sengaja datang ke kpu sumbar bertepatan dengan peringatan hari ibu 2022. "kita membawa tim perempuan sebanyak 22 orang dan kemudian membawa bunga yang dibagi-bagikan kepada masyarakat," kata emma. untuk target, emma berharap bisa terpilih kembali jadi anggota dpd ri. "kalau dulu suara sampai 500.000 lebih dan mengalahkan suara pak jokowi-ma'ruf amin. sekarang ya mudah-mudahan bisa kembali terpilih lagi," kata emma. sementara itu, komisioner kpu sumbar, gebril daulay yang menerima syarat dukungan dari emma yohanna menyebutkan, hingga sekarang baru satu orang yang datang menyerahkan persyaratan tersebut. "buk emma yang pertama. sementara peminat ada 30 orang yang mendaftar di akun silon," kata gebril. gebril mengatakan, kpu memberi kesempatan hingga 29 desember 2022 untuk pihak yang ingin menyerahkan

ID	Text
	persyaratan dukungan. "persyaratannya yaitu minimal 2.000 dukungan yang tersebar di 10 kabupaten dan kota," kata gebril. dapatkan update berita pilihan dan breaking news setiap hari dari kompas.com. mari bergabung di grup telegram "kompas.com news update", caranya klik link https://t.me/kompascomupdate, kemudian join. anda harus install aplikasi telegram terlebih dulu di ponsel.

Berikut ini adalah hasil dari proses *case folding*:

Contoh Teks Awal: "Bagikan Cara Membuat Instagram Best Nine Tanpa Aplikasi"

Setelah *Case Folding* menjadi: "bagikan cara membuat instagram best nine tanpa aplikasi"

Proses *case folding* atau mengubah kalimat dalam *dataset* menjadi huruf kecil atau biasa disebut juga dengan *lower casing*, dapat dilihat pada tabel 4.2 di atas bahwa semua kalimat dalam *dataset* yang sebelumnya memiliki kombinasi huruf kapital dan huruf kecil diubah dan disimpan menjadi kalimat yang tersusun atas huruf kecil secara keseluruhan dalam dokumen *dataset*.

2. *Remove Special character*

   Hasil *remove special character* dapat dilihat pada tabel 4.3 berikut:

Tabel 4.3 Hasil *Remove Special Character*

ID	Text
D1	bagikan tweet ########## cara membuat best nine instagram tanpa aplikasi 2021 ? tentunya sekarang untuk membuat kolase foto tidak begitu sulit dan juga tidak membutuhkan aplikasi loh. karena di artikel kali ini, kita akan bagikan tips cara membuat best nine instagram tanpa

ID	Text
	aplikasi terbaru di tahun 2021. penasaran caranya seperti apa, silakan baca terus artikel kami sampai selesai agar teman-teman paham proses pembuatan kolase foto menampilkan 9 kota foto. a b c d daftar isi apa itu best nine?best nine & top nine instagram 2020best nine instagram tanpa aplikasi 2021akhir kata apa itu best nine? best nine adalah kolase foto dengan menampilkan foto foto terdiri dari 9 kotak dalam susunan 3 kali 3 (gride). sembilan foto tersebut di pilih berdasarkan foto terbaik yang di unggah oleh penggunannya sepanjang tahun. proses pembuatan best nine cukup mudah dan tentunya ini tidak menggunakan aplikasi tambahan. baca juga :? rekomendasi hp oppo dengan kamera terbaik dan harga super terjangkau cuma 2 jutaanbest nine & top nine instagram 2020 untuk best nine instagram di tahun 2020 bisa kalian buat di 2 situs ternama yang dapat kalian kunjungi link nya di bawah ini! situs tersebut bernama top nine dan best nine. top nine: @ best nine: kedua situs di atas untuk membuat kolase foto yang sudah kamu unggah sepanjang tahun 2020. cara membuatnya cukup mudah, karena disini kamu hanya perlu memasukan foto foto kamu ke dalam situs tersebut agar dapat membuat kolase foto yang terdiri dari 9 kotak dan 3?3 (grid). baca juga :? tips: how to develop android app menggunakan android studiountuk kedua situs di atas berbeda cara penggunaanya, jika kamu menggunakan top nine maka akan di minta memasukan user ig kamu dan alamat email. namun jika kamu menggunakan situs best nine, kamu akan di minta mengisi username ig saja tanpa meminta email instagram kamu. untuk keamanan privasi di

ID	Text
	dua situs ini sangat aman, jadi gak perlu takut tentang akun kamu. best nine instagram tanpa aplikasi 2021 cara pembuatan best nine instagram tanpa aplikasi 2021 akan segera hadir, karena ini masih merupakan awal tahun 2021 jadi di tunggu saja update *website* untuk membuat best nine kolase foto terbaik anda di tahun ini. baca juga :? 3 cara melihat instagram di private tanpa followjika kamu ingin membuat kolase foto di tahun sebelumnya dapat mengunjungi ke dua situs di atas. karena sangat mudah di gunakan dan tentunya ini gratis tanpa menggunakan aplikasi tambahan. akhir kata jadi itulah sedikit tutorial dari kami tentang cara membuat best nine instagram tanpa aplikasi 2021, semoga artikel yang kami bagikan kali ini bermanfaat bagi teman-teman yang membacanya. jangan lupa share artikel ini sebagai dukungan teman-teman untuk kemajuan web topgloabl1.com
D2	aku beristirahat sebentar. lalu aku bertanya pada fani, ?fan.. aku boleh nggak, masukin penisku ke vagina kamu?? fani tampak berpikir sejenak. namun dalam kondisi horny seperti itu, pikirannya tentu agak kacau. fani akhirnya mengangguk lemas. perlahan-lahan, aku merebahkan badannya di ranjang dan menaruh kedua kakinya di bahuku. aku mulai menyentuhkan kepala penisku ke bibir liang senggamanya. lalu aku perlahan-lahan memasukkan batang kejantananku ke dalam liang senggamanya. aduh.. sulitnya batang kejantananku masuk ke liang senggamanya. walaupun liang senggamanya sudah basah, namun liang senggamanya masih sangat rapat. ketika batang kejantananku perlahan-lahan masuk, fani mulai

ID	Text
	mengerang kesakitan. aku mencoba menenangkannya. aku pun merasakan sakit, karena batang kejantananku seperti ditekan oleh liang senggamanya yang sempit.
D3	hampir 2 jam paksu memantat aku.. habis sengal-sengal seluruh badan aku.. cipap aku tak yah cerita la.. rasa kembang semacam aje.. jalan pun rasa lain aje lepas tu.. rasa menyesal ada juga pasal aku dah tak virgin lagi. dah kena robek kat paksu aku. tapi rasa menyesal tu rasa berbaloi juga dengan nikmat yang aku kecapi. aku tak sangka sex begitu sedap. patutla orang nak kahwin sangat. bagi aku at least aku dah rasa, nak tunggu kahwin lambat lagi, paling awal pun ayah aku bagi kahwin umur 21. tak sanggup rasanya nak tunggu 7 tahun lagi. cipap aku ni asyik terkemut-kemut aje bila tengok balak hensem. last sekali sebagai upacara penutup paksu suruh aku kulum batang dia.. aku mengikut aje walaupun tak pandai aku cuba juga. separuh aje batang paksu dapat aku kulum. kira okay jugaklah untuk yang tak ada pengalaman macam aku ni. itulah first time aku tengok batang lelaki dewasa, selalunya aku tengok konek adik aku yang sebesar ibu jari je. terkejut juga aku bila tengok batang paksu yang hampir sebesar lengan aku. panjangnya lebih kurang 6 inci saja tapi agak besar. kepala batangnya pun besar macam cendawan.. suka betul aku bila tengok kepala batang paksu mengembang dan berkilat bila kena kulum. masa aku kulum batang paksu ramas-ramas buah dada aku.. sekali-sekali jari jahat dia korek lubang dubur aku. pengotor betul paksu aku ni. ada ka dia kata lubang dubur aku cute. kalau kata cipap aku cute logik juga.. pasal cipap aku belum ada banyak bulu.. ada

ID	Text
	bulu pahat saja.. nipis dan halus. nampak bersih dan cute.. lebih kurang pukul 2 barulah maksu dan atie balik.. sempatlah aku aku mandi dan berehat lepas kena kongkek kat paksu.
D4	nan
D5	padang, kompas.com - senator asal sumatera barat, emma yohanna, yang mengungguli suara jokowi-ma'ruf amin pada pemilu 2019 lalu berniat maju kembali sebagai calon anggota dewan perwakilan daerah (dpd). incumbent tiga periode itu menyerahkan syarat dukungan 2.000 dukungan ke komisi pemilihan umum (kpu) sumbar, kamis (22/12/2022). pada pemilu 2019 lalu, emma berhasil mendapatkan 531.104 suara. jumlah itu menjadi yang tertinggi dibanding 20 calon anggota dpd daerah pemilihan sumbar. baca juga: tangisan iringi pemakaman aiptu ruslan, polisi yang tewas ditikam bawahannya di spn polda riau sementara, pasangan capres joko widodo-ma'ruf amin di sumbar mendapatkan 407.761 suara. "hari ini saya bersama tim menyerahkan syarat dukungan suara untuk maju jadi calon dpd ri," kata emma usai menyerahkan syarat dukungan. emma menyebut, ada 4.900 dukungan yang tersebar di 19 kabupaten dan kota di sumbar. jumlah itu sudah melampaui syarat minimal 2.000 dukungan yang tersebar di 10 kabupaten dan kota di sumbar. baca juga: kasus penipuan rp 1,1 miliar berkedok investasi pariwisata sumbar, polisi periksa 4 orang saksi emma mengaku sengaja datang ke kpu sumbar bertepatan dengan peringatan hari ibu 2022. "kita membawa tim perempuan

ID	Text
	sebanyak 22 orang dan kemudian membawa bunga yang dibagi-bagikan kepada masyarakat," kata emma. untuk target, emma berharap bisa terpilih kembali jadi anggota dpd ri. "kalau dulu suara sampai 500.000 lebih dan mengalahkan suara pak jokowi-ma'ruf amin. sekarang ya mudah-mudahan bisa kembali terpilih lagi," kata emma. sementara itu, komisioner kpu sumbar, gebril daulay yang menerima syarat dukungan dari emma yohanna menyebutkan, hingga sekarang baru satu orang yang datang menyerahkan persyaratan tersebut. "buk emma yang pertama. sementara peminat ada 30 orang yang mendaftar di akun silon," kata gebril. gebril mengatakan, kpu memberi kesempatan hingga 29 desember 2022 untuk pihak yang ingin menyerahkan persyaratan dukungan. "persyaratannya yaitu minimal 2.000 dukungan yang tersebar di 10 kabupaten dan kota," kata gebril. dapatkan update berita pilihan dan breaking news setiap hari dari kompas.com. mari bergabung di grup telegram "kompas.com news update", caranya klik link kemudian join. anda harus install aplikasi telegram terlebih dulu di ponsel.

Berikut ini adalah hasil dari proses *remove special character*:

Contoh Teks Awal: "bagikan \n\n\n\n\\t\t\t\t\t\t\n\t\n\t #tweet @ best @nine: https://bestnine.net/en "

Setelah Remove *Special Character* menjadi: “bagikan tweet @ best nine:”

Fungsi *remove special Character* digunakan untuk menghapus karakter-karakter khusus tertentu seperti '\t', '\n', dan '\u', serta mengganti karakter ", kemudian mengubah semua karakter menjadi

karakter *ASCII*. Selanjutnya, fungsi tersebut menghapus semua tanda '#' dan '@' diikuti oleh karakter alfanumerik, serta menghapus semua *URL* yang terdapat pada teks seperti pada tabel 4.3 di atas.

3. *Remove Number*

   Hasil dari *remove* number dapat dilihat pada tabel 4.4 berikut:

Tabel 4.4 Hasil *Remove Number*

ID	Text
D1	bagikan tweet ########## cara membuat best nine instagram tanpa aplikasi ? tentunya sekarang untuk membuat kolase foto tidak begitu sulit dan juga tidak membutuhkan aplikasi loh. karena di artikel kali ini, kita akan bagikan tips cara membuat best nine instagram tanpa aplikasi terbaru di tahun . penasaran caranya seperti apa, silakan baca terus artikel kami sampai selesai agar teman-teman paham proses pembuatan kolase foto menampilkan kota foto. a b c d daftar isi apa itu best nine?best nine & top nine instagram best nine instagram tanpa aplikasi akhir kata apa itu best nine? best nine adalah kolase foto dengan menampilkan foto foto terdiri dari kotak dalam susunan kali (gride). sembilan foto tersebut di pilih berdasarkan foto terbaik yang di unggah oleh penggunannya sepanjang tahun. proses pembuatan best nine cukup mudah dan tentunya ini tidak menggunakan aplikasi tambahan. baca juga :? rekomendasi hp oppo dengan kamera terbaik dan harga super terjangkau cuma jutaanbest nine & top nine instagram untuk best nine instagram di tahun bisa kalian buat di situs ternama yang dapat kalian kunjungi link nya di bawah ini! situs tersebut bernama top nine dan best nine. top nine: @ best nine: kedua situs di atas untuk membuat

ID	Text
	kolase foto yang sudah kamu unggah sepanjang tahun . cara membuatnya cukup mudah, karena disini kamu hanya perlu memasukan foto foto kamu ke dalam situs tersebut agar dapat membuat kolase foto yang terdiri dari kotak dan ? (grid). baca juga :? tips: how to develop android app menggunakan android studiountuk kedua situs di atas berbeda cara penggunaanya, jika kamu menggunakan top nine maka akan di minta memasukan user ig kamu dan alamat email. namun jika kamu menggunakan situs best nine, kamu akan di minta mengisi username ig saja tanpa meminta email instagram kamu. untuk keamanan privasi di dua situs ini sangat aman, jadi gak perlu takut tentang akun kamu. best nine instagram tanpa aplikasi cara pembuatan best nine instagram tanpa aplikasi akan segera hadir, karena ini masih merupakan awal tahun jadi di tunggu saja update *website* untuk membuat best nine kolase foto terbaik anda di tahun ini. baca juga :? cara melihat instagram di private tanpa followjika kamu ingin membuat kolase foto di tahun sebelumnya dapat mengunjungi ke dua situs di atas. karena sangat mudah di gunakan dan tentunya ini gratis tanpa menggunakan aplikasi tambahan. akhir kata jadi itulah sedikit tutorial dari kami tentang cara membuat best nine instagram tanpa aplikasi , semoga artikel yang kami bagikan kali ini bermanfaat bagi teman-teman yang membacanya. jangan lupa share artikel ini sebagai dukungan teman-teman untuk kemajuan web topgloabl.com
D2	aku beristirahat sebentar. lalu aku bertanya pada fani, ?fan.. aku boleh nggak, masukin penisku ke vagina kamu?? fani

ID	Text
	tampak berpikir sejenak. namun dalam kondisi horny seperti itu, pikirannya tentu agak kacau. fani akhirnya mengangguk lemas. perlahan-lahan, aku merebahkan badannya di ranjang dan menaruh kedua kakinya di bahuku. aku mulai menyentuhkan kepala penisku ke bibir liang senggamanya. lalu aku perlahan-lahan memasukkan batang kejantananku ke dalam liang senggamanya. aduh.. sulitnya batang kejantananku masuk ke liang senggamanya. walaupun liang senggamanya sudah basah, namun liang senggamanya masih sangat rapat. ketika batang kejantananku perlahan-lahan masuk, fani mulai mengerang kesakitan. aku mencoba menenangkannya. aku pun merasakan sakit, karena batang kejantananku seperti ditekan oleh liang senggamanya yang sempit.
D3	hampir jam paksu memantat aku.. habis sengal-sengal seluruh badan aku.. cipap aku tak yah cerita la.. rasa kembang semacam aje.. jalan pun rasa lain aje lepas tu.. rasa menyesal ada juga pasal aku dah tak virgin lagi. dah kena robek kat paksu aku. tapi rasa menyesal tu rasa berbaloi juga dengan nikmat yang aku kecapi. aku tak sangka sex begitu sedap. patutla orang nak kahwin sangat. bagi aku at least aku dah rasa, nak tunggu kahwin lambat lagi, paling awal pun ayah aku bagi kahwin umur . tak sanggup rasanya nak tunggu tahun lagi. cipap aku ni asyik terkemut-kemut aje bila tengok balak hensem. last sekali sebagai upacara penutup paksu suruh aku kulum batang dia.. aku mengikut aje walaupun tak pandai aku cuba juga. separuh aje batang paksu dapat aku kulum. kira okay jugaklah untuk yang tak ada pengalaman macam aku ni. itulah first time aku tengok

ID	Text
	batang lelaki dewasa, selalunya aku tengok konek adik aku yang sebesar ibu jari je. terkejut juga aku bila tengok batang paksu yang hampir sebesar lengan aku. panjangnya lebih kurang inci saja tapi agak besar. kepala batangnya pun besar macam cendawan.. suka betul aku bila tengok kepala batang paksu mengembang dan berkilat bila kena kulum. masa aku kulum batang paksu ramas-ramas buah dada aku.. sekali-sekali jari jahat dia korek lubang dubur aku. pengotor betul paksu aku ni. ada ka dia kata lubang dubur aku cute. kalau kata cipap aku cute logik juga.. pasal cipap aku belum ada banyak bulu.. ada bulu pahat saja.. nipis dan halus. nampak bersih dan cute.. lebih kurang pukul barulah maksu dan atie balik.. sempatlah aku aku mandi dan berehat lepas kena kongkek kat paksu.
D4	nan
D5	padang, kompas.com - senator asal sumatera barat, emma yohanna, yang mengungguli suara jokowi-ma'ruf amin pada pemilu lalu berniat maju kembali sebagai calon anggota dewan perwakilan daerah (dpd). incumbent tiga periode itu menyerahkan syarat dukungan . dukungan ke komisi pemilihan umum (kpu) sumbar, kamis (//). pada pemilu lalu, emma berhasil mendapatkan . suara. jumlah itu menjadi yang tertinggi dibanding calon anggota dpd daerah pemilihan sumbar. baca juga: tangisan iringi pemakaman aiptu ruslan, polisi yang tewas ditikam bawahannya di spn polda riau sementara, pasangan capres joko widodo-ma'ruf amin di sumbar mendapatkan . suara. "hari ini saya bersama tim menyerahkan syarat dukungan

ID	Text
	suara untuk maju jadi calon dpd ri," kata emma usai menyerahkan syarat dukungan. emma menyebut, ada . dukungan yang tersebar di kabupaten dan kota di sumbar. jumlah itu sudah melampaui syarat minimal . dukungan yang tersebar di kabupaten dan kota di sumbar. baca juga: kasus penipuan rp , miliar berkedok investasi pariwisata sumbar, polisi periksa orang saksi emma mengaku sengaja datang ke kpu sumbar bertepatan dengan peringatan hari ibu . "kita membawa tim perempuan sebanyak orang dan kemudian membawa bunga yang dibagi-bagikan kepada masyarakat," kata emma. untuk target, emma berharap bisa terpilih kembali jadi anggota dpd ri. "kalau dulu suara sampai . lebih dan mengalahkan suara pak jokowi-ma'ruf amin. sekarang ya mudah-mudahan bisa kembali terpilih lagi," kata emma. sementara itu, komisioner kpu sumbar, gebril daulay yang menerima syarat dukungan dari emma yohanna menyebutkan, hingga sekarang baru satu orang yang datang menyerahkan persyaratan tersebut. "buk emma yang pertama. sementara peminat ada orang yang mendaftar di akun silon," kata gebril. gebril mengatakan, kpu memberi kesempatan hingga desember untuk pihak yang ingin menyerahkan persyaratan dukungan. "persyaratannya yaitu minimal . dukungan yang tersebar di kabupaten dan kota," kata gebril. dapatkan update berita pilihan dan breaking news setiap hari dari kompas.com. mari bergabung di grup telegram "kompas.com news update", caranya klik link kemudian join. anda harus install aplikasi telegram terlebih dulu di ponsel.

Contoh Teks Awal: “bagikan *tweet* cara membuat best nine instagram tanpa aplikasi 2021”

Setelah *Remove Number* menjadi: “bagikan *tweet* cara membuat best nine instagram tanpa aplikasi”

Dapat dilihat pada tabel 4.4 di atas bahwa semua angka yang terdapat dalam dokumen *dataset* telah dihapus.

4. *Remove Punctuation*

   Hasil dari implementasi *remove* punctuation dapat dilihat pada tabel 4.5 berikut:

Tabel 4.5 Hasil *Remove Punctuation*

ID	Text
D1	bagikan tweet cara membuat best nine instagram tanpa aplikasi tentunya sekarang untuk membuat kolase foto tidak begitu sulit dan juga tidak membutuhkan aplikasi loh karena di artikel kali ini kita akan bagikan tips cara membuat best nine instagram tanpa aplikasi terbaru di tahun penasaran caranya seperti apa silakan baca terus artikel kami sampai selesai agar temanteman paham proses pembuatan kolase foto menampilkan kota foto a b c d daftar isi apa itu best ninebest nine top nine instagram best nine instagram tanpa aplikasi akhir kata apa itu best nine best nine adalah kolase foto dengan menampilkan foto foto terdiri dari kotak dalam susunan kali gride sembilan foto tersebut di pilih berdasarkan foto terbaik yang di unggah oleh penggunannya sepanjang tahun proses pembuatan best nine cukup mudah dan tentunya ini tidak menggunakan aplikasi tambahan baca juga rekomendasi hp oppo dengan kamera terbaik dan harga super terjangkau cuma jutaanbest nine top nine instagram untuk best nine instagram di tahun

ID	Text
	bisa kalian buat di situs ternama yang dapat kalian kunjungi link nya di bawah ini situs tersebut bernama top nine dan best nine top nine best nine kedua situs di atas untuk membuat kolase foto yang sudah kamu unggah sepanjang tahun cara membuatnya cukup mudah karena disini kamu hanya perlu memasukan foto foto kamu ke dalam situs tersebut agar dapat membuat kolase foto yang terdiri dari kotak dan grid baca juga tips how to develop android app menggunakan android studiountuk kedua situs di atas berbeda cara penggunaanya jika kamu menggunakan top nine maka akan di minta memasukan user ig kamu dan alamat email namun jika kamu menggunakan situs best nine kamu akan di minta mengisi username ig saja tanpa meminta email instagram kamu untuk keamanan privasi di dua situs ini sangat aman jadi gak perlu takut tentang akun kamu best nine instagram tanpa aplikasi cara pembuatan best nine instagram tanpa aplikasi akan segera hadir karena ini masih merupakan awal tahun jadi di tunggu saja update *website* untuk membuat best nine kolase foto terbaik anda di tahun ini baca juga cara melihat instagram di private tanpa followjika kamu ingin membuat kolase foto di tahun sebelumnya dapat mengunjungi ke dua situs di atas karena sangat mudah di gunakan dan tentunya ini gratis tanpa menggunakan aplikasi tambahan akhir kata jadi itulah sedikit tutorial dari kami tentang cara membuat best nine instagram tanpa aplikasi semoga artikel yang kami bagikan kali ini bermanfaat bagi temanteman yang membacanya jangan lupa share artikel ini sebagai dukungan temanteman untuk kemajuan web topgloablcom

ID	Text
D2	aku beristirahat sebentar lalu aku bertanya pada fani fan aku boleh nggak masukin penisku ke vagina kamu fani tampak berpikir sejenak namun dalam kondisi horny seperti itu pikirannya tentu agak kacau fani akhirnya mengangguk lemas perlahanlahan aku merebahkan badannya di ranjang dan menaruh kedua kakinya di bahuku aku mulai menyentuhkan kepala penisku ke bibir liang senggamanya lalu aku perlahanlahan memasukkan batang kejantananku ke dalam liang senggamanya aduh sulitnya batang kejantananku masuk ke liang senggamanya walaupun liang senggamanya sudah basah namun liang senggamanya masih sangat rapat ketika batang kejantananku perlahanlahan masuk fani mulai mengerang kesakitan aku mencoba menenangkannya aku pun merasakan sakit karena batang kejantananku seperti ditekan oleh liang senggamanya yang sempit
D3	hampir jam paksu memantat aku habis sengalsengal seluruh badan aku cipap aku tak yah cerita la rasa kembang semacam aje jalan pun rasa lain aje lepas tu rasa menyesal ada juga pasal aku dah tak virgin lagi dah kena robek kat paksu aku tapi rasa menyesal tu rasa berbaloi juga dengan nikmat yang aku kecapi aku tak sangka sex begitu sedap patutla orang nak kahwin sangat bagi aku at least aku dah rasa nak tunggu kahwin lambat lagi paling awal pun ayah aku bagi kahwin umur tak sanggup rasanya nak tunggu tahun lagi cipap aku ni asyik terkemutkemut aje bila tengok balak hensem last sekali sebagai upacara penutup paksu suruh aku kulum batang dia aku mengikut aje walaupun tak pandai aku cuba juga separuh aje batang paksu dapat aku

ID	Text
	kulum kira okay jugaklah untuk yang tak ada pengalaman macam aku ni itulah first time aku tengok batang lelaki dewasa selalunya aku tengok konek adik aku yang sebesar ibu jari je terkejut juga aku bila tengok batang paksu yang hampir sebesar lengan aku panjangnya lebih kurang inci saja tapi agak besar kepala batangnya pun besar macam cendawan suka betul aku bila tengok kepala batang paksu mengembang dan berkilat bila kena kulum masa aku kulum batang paksu ramasramas buah dada aku sekalisekali jari jahat dia korek lubang dubur aku pengotor betul paksu aku ni ada ka dia kata lubang dubur aku cute kalau kata cipap aku cute logik juga pasal cipap aku belum ada banyak bulu ada bulu pahat saja nipis dan halus nampak bersih dan cute lebih kurang pukul barulah maksu dan atie balik sempatlah aku aku mandi dan berehat lepas kena kongkek kat paksu
D4	nan
D5	padang kompascom senator asal sumatera barat emma yohanna yang mengungguli suara jokowimaruf amin pada pemilu lalu berniat maju kembali sebagai calon anggota dewan perwakilan daerah dpd incumbent tiga periode itu menyerahkan syarat dukungan dukungan ke komisi pemilihan umum kpu sumbar kamis pada pemilu lalu emma berhasil mendapatkan suara jumlah itu menjadi yang tertinggi dibanding calon anggota dpd daerah pemilihan sumbar baca juga tangisan iringi pemakaman aiptu ruslan polisi yang tewas ditikam bawahannya di spn polda riau sementara pasangan capres joko widodomaruf amin di sumbar mendapatkan suara hari ini saya bersama tim

ID	Text
	menyerahkan syarat dukungan suara untuk maju jadi calon dpd ri kata emma usai menyerahkan syarat dukungan emma menyebut ada dukungan yang tersebar di kabupaten dan kota di sumbar jumlah itu sudah melampaui syarat minimal dukungan yang tersebar di kabupaten dan kota di sumbar baca juga kasus penipuan rp miliar berkedok investasi pariwisata sumbar polisi periksa orang saksi emma mengaku sengaja datang ke kpu sumbar bertepatan dengan peringatan hari ibu kita membawa tim perempuan sebanyak orang dan kemudian membawa bunga yang dibagibagikan kepada masyarakat kata emma untuk target emma berharap bisa terpilih kembali jadi anggota dpd ri kalau dulu suara sampai lebih dan mengalahkan suara pak jokowimaruf amin sekarang ya mudahmudahan bisa kembali terpilih lagi kata emma sementara itu komisioner kpu sumbar gebril daulay yang menerima syarat dukungan dari emma yohanna menyebutkan hingga sekarang baru satu orang yang datang menyerahkan persyaratan tersebut buk emma yang pertama sementara peminat ada orang yang mendaftar di akun silon kata gebril gebril mengatakan kpu memberi kesempatan hingga desember untuk pihak yang ingin menyerahkan persyaratan dukungan persyaratannya yaitu minimal dukungan yang tersebar di kabupaten dan kota kata gebril dapatkan update berita pilihan dan breaking news setiap hari dari kompascom mari bergabung di grup telegram kompascom news update caranya klik link kemudian join anda harus install aplikasi telegram terlebih dulu di ponsel

Contoh Teks Awal: “bagikan *tweet* ########## cara membuat best nine instagram tanpa aplikasi ? @ best nine!”

Setelah Remove Punctuation menjadi: "bagikan *tweet* cara membuat best nine instagram tanpa aplikasi best nine"

Pada tabel 4.5 di atas menunjukan bahwa semua tanda baca seperti titik (.), koma (,), tanda seru (!), tanda tanya (?) dll telah dihapus.

5. *Remove White Space dan Multiple White Space*

   Hasil implementasi *remove white space* dan *multiple white space* dapat dilihat pada tabel 4.6 berikut:

Tabel 4.6 Hasil *Remove White Space & Multiple White Space*

ID	Text
D1	bagikan tweet cara membuat best nine instagram tanpa aplikasi tentunya sekarang untuk membuat kolase foto tidak begitu sulit dan juga tidak membutuhkan aplikasi loh karena di artikel kali ini kita akan bagikan tips cara membuat best nine instagram tanpa aplikasi terbaru di tahun penasaran caranya seperti apa silakan baca terus artikel kami sampai selesai agar temanteman paham proses pembuatan kolase foto menampilkan kota foto a b c d daftar isi apa itu best ninebest nine top nine instagram best nine instagram tanpa aplikasi akhir kata apa itu best nine best nine adalah kolase foto dengan menampilkan foto foto terdiri dari kotak dalam susunan kali gride sembilan foto tersebut di pilih berdasarkan foto terbaik yang di unggah oleh penggunannya sepanjang tahun proses pembuatan best nine cukup mudah dan tentunya ini tidak menggunakan aplikasi tambahan baca juga rekomendasi hp oppo dengan kamera terbaik dan harga super terjangkau cuma jutaanbest nine top nine instagram untuk best nine instagram di tahun

ID	Text
	bisa kalian buat di situs ternama yang dapat kalian kunjungi link nya di bawah ini situs tersebut bernama top nine dan best nine top nine best nine kedua situs di atas untuk membuat kolase foto yang sudah kamu unggah sepanjang tahun cara membuatnya cukup mudah karena disini kamu hanya perlu memasukan foto foto kamu ke dalam situs tersebut agar dapat membuat kolase foto yang terdiri dari kotak dan grid baca juga tips how to develop android app menggunakan android studiountuk kedua situs di atas berbeda cara penggunaanya jika kamu menggunakan top nine maka akan di minta memasukan user ig kamu dan alamat email namun jika kamu menggunakan situs best nine kamu akan di minta mengisi username ig saja tanpa meminta email instagram kamu untuk keamanan privasi di dua situs ini sangat aman jadi gak perlu takut tentang akun kamu best nine instagram tanpa aplikasi cara pembuatan best nine instagram tanpa aplikasi akan segera hadir karena ini masih merupakan awal tahun jadi di tunggu saja update *website* untuk membuat best nine kolase foto terbaik anda di tahun ini baca juga cara melihat instagram di private tanpa followjika kamu ingin membuat kolase foto di tahun sebelumnya dapat mengunjungi ke dua situs di atas karena sangat mudah di gunakan dan tentunya ini gratis tanpa menggunakan aplikasi tambahan akhir kata jadi itulah sedikit tutorial dari kami tentang cara membuat best nine instagram tanpa aplikasi semoga artikel yang kami bagikan kali ini bermanfaat bagi temanteman yang membacanya jangan lupa share artikel ini sebagai dukungan temanteman untuk kemajuan web topgloablcom

ID	Text
D2	aku beristirahat sebentar lalu aku bertanya pada fani fan aku boleh nggak masukin penisku ke vagina kamu fani tampak berpikir sejenak namun dalam kondisi horny seperti itu pikirannya tentu agak kacau fani akhirnya mengangguk lemas perlahanlahan aku merebahkan badannya di ranjang dan menaruh kedua kakinya di bahuku aku mulai menyentuhkan kepala penisku ke bibir liang senggamanya lalu aku perlahanlahan memasukkan batang kejantananku ke dalam liang senggamanya aduh sulitnya batang kejantananku masuk ke liang senggamanya walaupun liang senggamanya sudah basah namun liang senggamanya masih sangat rapat ketika batang kejantananku perlahanlahan masuk fani mulai mengerang kesakitan aku mencoba menenangkannya aku pun merasakan sakit karena batang kejantananku seperti ditekan oleh liang senggamanya yang sempit
D3	hampir jam paksu memantat aku habis sengalsengal seluruh badan aku cipap aku tak yah cerita la rasa kembang semacam aje jalan pun rasa lain aje lepas tu rasa menyesal ada juga pasal aku dah tak virgin lagi dah kena robek kat paksu aku tapi rasa menyesal tu rasa berbaloi juga dengan nikmat yang aku kecapi aku tak sangka sex begitu sedap patutla orang nak kahwin sangat bagi aku at least aku dah rasa nak tunggu kahwin lambat lagi paling awal pun ayah aku bagi kahwin umur tak sanggup rasanya nak tunggu tahun lagi cipap aku ni asyik terkemutkemut aje bila tengok balak hensem last sekali sebagai upacara penutup paksu suruh aku kulum batang dia aku mengikut aje walaupun tak pandai aku cuba juga separuh aje batang paksu dapat aku

ID	Text
	kulum kira okay jugaklah untuk yang tak ada pengalaman macam aku ni itulah first time aku tengok batang lelaki dewasa selalunya aku tengok konek adik aku yang sebesar ibu jari je terkejut juga aku bila tengok batang paksu yang hampir sebesar lengan aku panjangnya lebih kurang inci saja tapi agak besar kepala batangnya pun besar macam cendawan suka betul aku bila tengok kepala batang paksu mengembang dan berkilat bila kena kulum masa aku kulum batang paksu ramasramas buah dada aku sekalisekali jari jahat dia korek lubang dubur aku pengotor betul paksu aku ni ada ka dia kata lubang dubur aku cute kalau kata cipap aku cute logik juga pasal cipap aku belum ada banyak bulu ada bulu pahat saja nipis dan halus nampak bersih dan cute lebih kurang pukul barulah maksu dan atie balik sempatlah aku aku mandi dan berehat lepas kena kongkek kat paksu
D4	nan
D5	padang kompascom senator asal sumatera barat emma yohanna yang mengungguli suara jokowimaruf amin pada pemilu lalu berniat maju kembali sebagai calon anggota dewan perwakilan daerah dpd incumbent tiga periode itu menyerahkan syarat dukungan dukungan ke komisi pemilihan umum kpu sumbar kamis pada pemilu lalu emma berhasil mendapatkan suara jumlah itu menjadi yang tertinggi dibanding calon anggota dpd daerah pemilihan sumbar baca juga tangisan iringi pemakaman aiptu ruslan polisi yang tewas ditikam bawahannya di spn polda riau sementara pasangan capres joko widodomaruf amin di sumbar mendapatkan suara hari ini saya bersama tim

ID	Text
	menyerahkan syarat dukungan suara untuk maju jadi calon dpd ri kata emma usai menyerahkan syarat dukungan emma menyebut ada dukungan yang tersebar di kabupaten dan kota di sumbar jumlah itu sudah melampaui syarat minimal dukungan yang tersebar di kabupaten dan kota di sumbar baca juga kasus penipuan rp miliar berkedok investasi pariwisata sumbar polisi periksa orang saksi emma mengaku sengaja datang ke kpu sumbar bertepatan dengan peringatan hari ibu kita membawa tim perempuan sebanyak orang dan kemudian membawa bunga yang dibagibagikan kepada masyarakat kata emma untuk target emma berharap bisa terpilih kembali jadi anggota dpd ri kalau dulu suara sampai lebih dan mengalahkan suara pak jokowimaruf amin sekarang ya mudahmudahan bisa kembali terpilih lagi kata emma sementara itu komisioner kpu sumbar gebril daulay yang menerima syarat dukungan dari emma yohanna menyebutkan hingga sekarang baru satu orang yang datang menyerahkan persyaratan tersebut buk emma yang pertama sementara peminat ada orang yang mendaftar di akun silon kata gebril gebril mengatakan kpu memberi kesempatan hingga desember untuk pihak yang ingin menyerahkan persyaratan dukungan persyaratannya yaitu minimal dukungan yang tersebar di kabupaten dan kota kata gebril dapatkan update berita pilihan dan breaking news setiap hari dari kompascom mari bergabung di grup telegram kompascom news update caranya klik link kemudian join anda harus install aplikasi telegram terlebih dulu di ponsel

Contoh Teks Awal: “ bagikan *tweet* cara membuat best nine instagram tanpa aplikasi best nine”

Setelah *Remove White Space* menjadi: "bagikan *tweet* cara membuat
best nine instagram tanpa aplikasi best nine "
Setelah *Multiple Remove White Space* menjadi: “bagikan tweet cara membuat best nine instagram tanpa aplikasi best nine”
Dapat dilihat pada tabel 4.6 bahwa spasi atau karakter kosong dalam teks dokumen telah dihilangkan, fungsi *remove white space* dan *multiple white space* ini digunakan untuk membersihkan data yang tidak terstruktur.

6. *Remove Single Character*
   Hasil dari implementasi *remove single character* dapat dilihat pada tabel 4.7 berikut:

Tabel 4.7 Hasil *Remove Single Character*

ID	Text
D1	bagikan tweet cara membuat best nine instagram tanpa aplikasi tentunya sekarang untuk membuat kolase foto tidak begitu sulit dan juga tidak membutuhkan aplikasi loh karena di artikel kali ini kita akan bagikan tips cara membuat best nine instagram tanpa aplikasi terbaru di tahun penasaran caranya seperti apa silakan baca terus artikel kami sampai selesai agar temanteman paham proses pembuatan kolase foto menampilkan kota foto daftar isi apa itu best ninebest nine top nine instagram best nine instagram tanpa aplikasi akhir kata apa itu best nine best nine adalah kolase foto dengan menampilkan foto foto terdiri dari kotak dalam susunan kali gride sembilan foto tersebut di pilih berdasarkan foto terbaik yang di unggah oleh penggunannya sepanjang tahun proses pembuatan best nine cukup mudah dan tentunya ini tidak menggunakan aplikasi

ID	Text
	tambahan baca juga rekomendasi hp oppo dengan kamera terbaik dan harga super terjangkau cuma jutaanbest nine top nine instagram untuk best nine instagram di tahun bisa kalian buat di situs ternama yang dapat kalian kunjungi link nya di bawah ini situs tersebut bernama top nine dan best nine top nine best nine kedua situs di atas untuk membuat kolase foto yang sudah kamu unggah sepanjang tahun cara membuatnya cukup mudah karena disini kamu hanya perlu memasukan foto foto kamu ke dalam situs tersebut agar dapat membuat kolase foto yang terdiri dari kotak dan grid baca juga tips how to develop android app menggunakan android studiountuk kedua situs di atas berbeda cara penggunaanya jika kamu menggunakan top nine maka akan di minta memasukan user ig kamu dan alamat email namun jika kamu menggunakan situs best nine kamu akan di minta mengisi username ig saja tanpa meminta email instagram kamu untuk keamanan privasi di dua situs ini sangat aman jadi gak perlu takut tentang akun kamu best nine instagram tanpa aplikasi cara pembuatan best nine instagram tanpa aplikasi akan segera hadir karena ini masih merupakan awal tahun jadi di tunggu saja update *website* untuk membuat best nine kolase foto terbaik anda di tahun ini baca juga cara melihat instagram di private tanpa followjika kamu ingin membuat kolase foto di tahun sebelumnya dapat mengunjungi ke dua situs di atas karena sangat mudah di gunakan dan tentunya ini gratis tanpa menggunakan aplikasi tambahan akhir kata jadi itulah sedikit tutorial dari kami tentang cara membuat best nine instagram tanpa aplikasi semoga artikel yang kami bagikan kali ini

ID	Text
	bermanfaat bagi temanteman yang membacanya jangan lupa share artikel ini sebagai dukungan temanteman untuk kemajuan web topgloablcom
D2	aku beristirahat sebentar lalu aku bertanya pada fani fan aku boleh nggak masukin penisku ke vagina kamu fani tampak berpikir sejenak namun dalam kondisi horny seperti itu pikirannya tentu agak kacau fani akhirnya mengangguk lemas perlahanlahan aku merebahkan badannya di ranjang dan menaruh kedua kakinya di bahuku aku mulai menyentuhkan kepala penisku ke bibir liang senggamanya lalu aku perlahanlahan memasukkan batang kejantananku ke dalam liang senggamanya aduh sulitnya batang kejantananku masuk ke liang senggamanya walaupun liang senggamanya sudah basah namun liang senggamanya masih sangat rapat ketika batang kejantananku perlahanlahan masuk fani mulai mengerang kesakitan aku mencoba menenangkannya aku pun merasakan sakit karena batang kejantananku seperti ditekan oleh liang senggamanya yang sempit
D3	hampir jam paksu memantat aku habis sengalsengal seluruh badan aku cipap aku tak yah cerita la rasa kembang semacam aje jalan pun rasa lain aje lepas tu rasa menyesal ada juga pasal aku dah tak virgin lagi dah kena robek kat paksu aku tapi rasa menyesal tu rasa berbaloi juga dengan nikmat yang aku kecapi aku tak sangka sex begitu sedap patutla orang nak kahwin sangat bagi aku at least aku dah rasa nak tunggu kahwin lambat lagi paling awal pun ayah aku bagi kahwin umur tak sanggup rasanya nak tunggu

ID	Text
	tahun lagi cipap aku ni asyik terkemutkemut aje bila tengok balak hensem last sekali sebagai upacara penutup paksu suruh aku kulum batang dia aku mengikut aje walaupun tak pandai aku cuba juga separuh aje batang paksu dapat aku kulum kira okay jugaklah untuk yang tak ada pengalaman macam aku ni itulah first time aku tengok batang lelaki dewasa selalunya aku tengok konek adik aku yang sebesar ibu jari je terkejut juga aku bila tengok batang paksu yang hampir sebesar lengan aku panjangnya lebih kurang inci saja tapi agak besar kepala batangnya pun besar macam cendawan suka betul aku bila tengok kepala batang paksu mengembang dan berkilat bila kena kulum masa aku kulum batang paksu ramasramas buah dada aku sekalisekali jari jahat dia korek lubang dubur aku pengotor betul paksu aku ni ada ka dia kata lubang dubur aku cute kalau kata cipap aku cute logik juga pasal cipap aku belum ada banyak bulu ada bulu pahat saja nipis dan halus nampak bersih dan cute lebih kurang pukul barulah maksu dan atie balik sempatlah aku aku mandi dan berehat lepas kena kongkek kat paksu
D4	nan
D5	padang kompascom senator asal sumatera barat emma yohanna yang mengungguli suara jokowimaruf amin pada pemilu lalu berniat maju kembali sebagai calon anggota dewan perwakilan daerah dpd incumbent tiga periode itu menyerahkan syarat dukungan dukungan ke komisi pemilihan umum kpu sumbar kamis pada pemilu lalu emma berhasil mendapatkan suara jumlah itu menjadi yang tertinggi dibanding calon anggota dpd daerah pemilihan

ID	Text
	sumbar baca juga tangisan iringi pemakaman aiptu ruslan polisi yang tewas ditikam bawahannya di spn polda riau sementara pasangan capres joko widodomaruf amin di sumbar mendapatkan suara hari ini saya bersama tim menyerahkan syarat dukungan suara untuk maju jadi calon dpd ri kata emma usai menyerahkan syarat dukungan emma menyebut ada dukungan yang tersebar di kabupaten dan kota di sumbar jumlah itu sudah melampaui syarat minimal dukungan yang tersebar di kabupaten dan kota di sumbar baca juga kasus penipuan rp miliar berkedok investasi pariwisata sumbar polisi periksa orang saksi emma mengaku sengaja datang ke kpu sumbar bertepatan dengan peringatan hari ibu kita membawa tim perempuan sebanyak orang dan kemudian membawa bunga yang dibagibagikan kepada masyarakat kata emma untuk target emma berharap bisa terpilih kembali jadi anggota dpd ri kalau dulu suara sampai lebih dan mengalahkan suara pak jokowimaruf amin sekarang ya mudahmudahan bisa kembali terpilih lagi kata emma sementara itu komisioner kpu sumbar gebril daulay yang menerima syarat dukungan dari emma yohanna menyebutkan hingga sekarang baru satu orang yang datang menyerahkan persyaratan tersebut buk emma yang pertama sementara peminat ada orang yang mendaftar di akun silon kata gebril gebril mengatakan kpu memberi kesempatan hingga desember untuk pihak yang ingin menyerahkan persyaratan dukungan persyaratannya yaitu minimal dukungan yang tersebar di kabupaten dan kota kata gebril dapatkan update berita pilihan dan breaking news setiap hari dari kompascom mari bergabung di grup telegram

ID	Text
	kompascom news update caranya klik link kemudian join anda harus install aplikasi telegram terlebih dulu di ponsel

Contoh Teks Awal: “bagikan *tweet* cara membuat best nine instagram tanpa aplikasi best nine a b c d e”

Setelah *Remove Single Character* menjadi: “bagikan *tweet* cara membuat best nine instagram tanpa aplikasi best nine “

Dapat dilihat pada tabel 4.7 di atas bahwa semua karakter tunggal yang terdapat dalam dokumen telah dihilangkan.

7. *Tokenizing*

Hasil dari *tokenizing* dapat dilihat pada tabel 4.8 berikut:

Tabel 4.8 Hasil *Tokenizing*

ID	Text
D1	['bagikan', 'tweet', 'cara', 'membuat', 'best', 'nine', 'instagram', 'tanpa', 'aplikasi', 'tentunya', 'sekarang', 'untuk', 'membuat', 'kolase', 'foto', 'tidak', 'begitu', 'sulit', 'dan', 'juga', 'tidak', 'membutuhkan', 'aplikasi', 'loh', 'karena', 'di', 'artikel', 'kali', 'ini', 'kita', 'akan', 'bagikan', 'tips', 'cara', 'membuat', 'best', 'nine', 'instagram', 'tanpa', 'aplikasi', 'terbaru', 'di', 'tahun', 'penasaran', 'caranya', 'seperti', 'apa', 'silakan', 'baca', 'terus', 'artikel', 'kami', 'sampai', 'selesai', 'agar', 'temanteman', 'paham', 'proses', 'pembuatan', 'kolase', 'foto', 'menampilkan', 'kota', 'foto', 'daftar', 'isi', 'apa', 'itu', 'best', 'ninebest', 'nine', 'top', 'nine', 'instagram', 'best', 'nine', 'instagram', 'tanpa', 'aplikasi', 'akhir', 'kata', 'apa', 'itu', 'best', 'nine', 'best', 'nine', 'adalah', 'kolase', 'foto', 'dengan', 'menampilkan', 'foto', 'foto', 'terdiri', 'dari', 'kotak', 'dalam', 'susunan', 'kali', 'gride', 'sembilan', 'foto', 'tersebut', 'di',

ID	Text
	'pilih', 'berdasarkan', 'foto', 'terbaik', 'yang', 'di', 'unggah', 'oleh', 'penggunannya', 'sepanjang', 'tahun', 'proses', 'pembuatan', 'best', 'nine', 'cukup', 'mudah', 'dan', 'tentunya', 'ini', 'tidak', 'menggunakan', 'aplikasi', 'tambahan', 'baca', 'juga', 'rekomendasi', 'hp', 'oppo', 'dengan', 'kamera', 'terbaik', 'dan', 'harga', 'super', 'terjangkau', 'cuma', 'jutaanbest', 'nine', 'top', 'nine', 'instagram', 'untuk', 'best', 'nine', 'instagram', 'di', 'tahun', 'bisa', 'kalian', 'buat', 'di', 'situs', 'ternama', 'yang', 'dapat', 'kalian', 'kunjungi', 'link', 'nya', 'di', 'bawah', 'ini', 'situs', 'tersebut', 'bernama', 'top', 'nine', 'dan', 'best', 'nine', 'top', 'nine', 'best', 'nine', 'kedua', 'situs', 'di', 'atas', 'untuk', 'membuat', 'kolase', 'foto', 'yang', 'sudah', 'kamu', 'unggah', 'sepanjang', 'tahun', 'cara', 'membuatnya', 'cukup', 'mudah', 'karena', 'disini', 'kamu', 'hanya', 'perlu', 'memasukan', 'foto', 'foto', 'kamu', 'ke', 'dalam', 'situs', 'tersebut', 'agar', 'dapat', 'membuat', 'kolase', 'foto', 'yang', 'terdiri', 'dari', 'kotak', 'dan', 'grid', 'baca', 'juga', 'tips', 'how', 'to', 'develop', 'android', 'app', 'menggunakan', 'android', 'studiountuk', 'kedua', 'situs', 'di', 'atas', 'berbeda', 'cara', 'penggunaanya', 'jika', 'kamu', 'menggunakan', 'top', 'nine', 'maka', 'akan', 'di', 'minta', 'memasukan', 'user', 'ig', 'kamu', 'dan', 'alamat', 'email', 'namun', 'jika', 'kamu', 'menggunakan', 'situs', 'best', 'nine', 'kamu', 'akan', 'di', 'minta', 'mengisi', 'username', 'ig', 'saja', 'tanpa', 'meminta', 'email', 'instagram', 'kamu', 'untuk', 'keamanan', 'privasi', 'di', 'dua', 'situs', 'ini', 'sangat', 'aman', 'jadi', 'gak', 'perlu', 'takut', 'tentang', 'akun', 'kamu', 'best', 'nine', 'instagram', 'tanpa', 'aplikasi', 'cara', 'pembuatan', 'best', 'nine', 'instagram', 'tanpa', 'aplikasi', 'akan', 'segera', 'hadir', 'karena', 'ini',

ID	Text
	'masih', 'merupakan', 'awal', 'tahun', 'jadi', 'di', 'tunggu', 'saja', 'update', '*website*', 'untuk', 'membuat', 'best', 'nine', 'kolase', 'foto', 'terbaik', 'anda', 'di', 'tahun', 'ini', 'baca', 'juga', 'cara', 'melihat', 'instagram', 'di', 'private', 'tanpa', 'followjika', 'kamu', 'ingin', 'membuat', 'kolase', 'foto', 'di', 'tahun', 'sebelumnya', 'dapat', 'mengunjungi', 'ke', 'dua', 'situs', 'di', 'atas', 'karena', 'sangat', 'mudah', 'di', 'gunakan', 'dan', 'tentunya', 'ini', 'gratis', 'tanpa', 'menggunakan', 'aplikasi', 'tambahan', 'akhir', 'kata', 'jadi', 'itulah', 'sedikit', 'tutorial', 'dari', 'kami', 'tentang', 'cara', 'membuat', 'best', 'nine', 'instagram', 'tanpa', 'aplikasi', 'semoga', 'artikel', 'yang', 'kami', 'bagikan', 'kali', 'ini', 'bermanfaat', 'bagi', 'temanteman', 'yang', 'membacanya', 'jangan', 'lupa', 'share', 'artikel', 'ini', 'sebagai', 'dukungan', 'temanteman', 'untuk', 'kemajuan', 'web', 'topgloablcom']
D2	['aku', 'beristirahat', 'sebentar', 'lalu', 'aku', 'bertanya', 'pada', 'fani', 'fan', 'aku', 'boleh', 'nggak', 'masukin', 'penisku', 'ke', 'vagina', 'kamu', 'fani', 'tampak', 'berpikir', 'sejenak', 'namun', 'dalam', 'kondisi', 'horny', 'seperti', 'itu', 'pikirannya', 'tentu', 'agak', 'kacau', 'fani', 'akhirnya', 'mengangguk', 'lemas', 'perlahanlahan', 'aku', 'merebahkan', 'badannya', 'di', 'ranjang', 'dan', 'menaruh', 'kedua', 'kakinya', 'di', 'bahuku', 'aku', 'mulai', 'menyentuhkan', 'kepala', 'penisku', 'ke', 'bibir', 'liang', 'senggamanya', 'lalu', 'aku', 'perlahanlahan', 'memasukkan', 'batang', 'kejantananku', 'ke', 'dalam', 'liang', 'senggamanya', 'aduh', 'sulitnya', 'batang', 'kejantananku', 'masuk', 'ke', 'liang', 'senggamanya', 'walaupun', 'liang', 'senggamanya', 'sudah', 'basah', 'namun', 'liang', 'senggamanya', 'masih', 'sangat', 'rapat', 'ketika',

ID	Text
	'batang', 'kejantananku', 'perlahanlahan', 'masuk', 'fani', 'mulai', 'mengerang', 'kesakitan', 'aku', 'mencoba', 'menenangkannya', 'aku', 'pun', 'merasakan', 'sakit', 'karena', 'batang', 'kejantananku', 'seperti', 'ditekan', 'oleh', 'liang', 'senggamanya', 'yang', 'sempit']
D3	['hampir', 'jam', 'paksu', 'memantat', 'aku', 'habis', 'sengalsengal', 'seluruh', 'badan', 'aku', 'cipap', 'aku', 'tak', 'yah', 'cerita', 'la', 'rasa', 'kembang', 'semacam', 'aje', 'jalan', 'pun', 'rasa', 'lain', 'aje', 'lepas', 'tu', 'rasa', 'menyesal', 'ada', 'juga', 'pasal', 'aku', 'dah', 'tak', 'virgin', 'lagi', 'dah', 'kena', 'robek', 'kat', 'paksu', 'aku', 'tapi', 'rasa', 'menyesal', 'tu', 'rasa', 'berbaloi', 'juga', 'dengan', 'nikmat', 'yang', 'aku', 'kecapi', 'aku', 'tak', 'sangka', 'sex', 'begitu', 'sedap', 'patutla', 'orang', 'nak', 'kahwin', 'sangat', 'bagi', 'aku', 'at', 'least', 'aku', 'dah', 'rasa', 'nak', 'tunggu', 'kahwin', 'lambat', 'lagi', 'paling', 'awal', 'pun', 'ayah', 'aku', 'bagi', 'kahwin', 'umur', 'tak', 'sanggup', 'rasanya', 'nak', 'tunggu', 'tahun', 'lagi', 'cipap', 'aku', 'ni', 'asyik', 'terkemutkemut', 'aje', 'bila', 'tengok', 'balak', 'hensem', 'last', 'sekali', 'sebagai', 'upacara', 'penutup', 'paksu', 'suruh', 'aku', 'kulum', 'batang', 'dia', 'aku', 'mengikut', 'aje', 'walaupun', 'tak', 'pandai', 'aku', 'cuba', 'juga', 'separuh', 'aje', 'batang', 'paksu', 'dapat', 'aku', 'kulum', 'kira', 'okay', 'jugaklah', 'untuk', 'yang', 'tak', 'ada', 'pengalaman', 'macam', 'aku', 'ni', 'itulah', 'first', 'time', 'aku', 'tengok', 'batang', 'lelaki', 'dewasa', 'selalunya', 'aku', 'tengok', 'konek', 'adik', 'aku', 'yang', 'sebesar', 'ibu', 'jari', 'je', 'terkejut', 'juga', 'aku', 'bila', 'tengok', 'batang', 'paksu', 'yang', 'hampir', 'sebesar', 'lengan', 'aku', 'panjangnya', 'lebih', 'kurang', 'inci', 'saja', 'tapi', 'agak', 'besar', 'kepala',

ID	Text
	'batangnya', 'pun', 'besar', 'macam', 'cendawan', 'suka', 'betul', 'aku', 'bila', 'tengok', 'kepala', 'batang', 'paksu', 'mengembang', 'dan', 'berkilat', 'bila', 'kena', 'kulum', 'masa', 'aku', 'kulum', 'batang', 'paksu', 'ramasramas', 'buah', 'dada', 'aku', 'sekalisekali', 'jari', 'jahat', 'dia', 'korek', 'lubang', 'dubur', 'aku', 'pengotor', 'betul', 'paksu', 'aku', 'ni', 'ada', 'ka', 'dia', 'kata', 'lubang', 'dubur', 'aku', 'cute', 'kalau', 'kata', 'cipap', 'aku', 'cute', 'logik', 'juga', 'pasal', 'cipap', 'aku', 'belum', 'ada', 'banyak', 'bulu', 'ada', 'bulu', 'pahat', 'saja', 'nipis', 'dan', 'halus', 'nampak', 'bersih', 'dan', 'cute', 'lebih', 'kurang', 'pukul', 'barulah', 'maksu', 'dan', 'atie', 'balik', 'sempatlah', 'aku', 'aku', 'mandi', 'dan', 'berehat', 'lepas', 'kena', 'kongkek', 'kat', 'paksu']
D4	nan
D5	['padang', 'kompascom', 'senator', 'asal', 'sumatera', 'barat', 'emma', 'yohanna', 'yang', 'mengungguli', 'suara', 'jokowimaruf', 'amin', 'pada', 'pemilu', 'lalu', 'berniat', 'maju', 'kembali', 'sebagai', 'calon', 'anggota', 'dewan', 'perwakilan', 'daerah', 'dpd', 'incumbent', 'tiga', 'periode', 'itu', 'menyerahkan', 'syarat', 'dukungan', 'dukungan', 'ke', 'komisi', 'pemilihan', 'umum', 'kpu', 'sumbar', 'kamis', 'pada', 'pemilu', 'lalu', 'emma', 'berhasil', 'mendapatkan', 'suara', 'jumlah', 'itu', 'menjadi', 'yang', 'tertinggi', 'dibanding', 'calon', 'anggota', 'dpd', 'daerah', 'pemilihan', 'sumbar', 'baca', 'juga', 'tangisan', 'iringi', 'pemakaman', 'aiptu', 'ruslan', 'polisi', 'yang', 'tewas', 'ditikam', 'bawahannya', 'di', 'spn', 'polda', 'riau', 'sementara', 'pasangan', 'capres', 'joko', 'widodomaruf', 'amin', 'di', 'sumbar', 'mendapatkan', 'suara',

ID	Text
	'hari', 'ini', 'saya', 'bersama', 'tim', 'menyerahkan', 'syarat', 'dukungan', 'suara', 'untuk', 'maju', 'jadi', 'calon', 'dpd', 'ri', 'kata', 'emma', 'usai', 'menyerahkan', 'syarat', 'dukungan', 'emma', 'menyebut', 'ada', 'dukungan', 'yang', 'tersebar', 'di', 'kabupaten', 'dan', 'kota', 'di', 'sumbar', 'jumlah', 'itu', 'sudah', 'melampaui', 'syarat', 'minimal', 'dukungan', 'yang', 'tersebar', 'di', 'kabupaten', 'dan', 'kota', 'di', 'sumbar', 'baca', 'juga', 'kasus', 'penipuan', 'rp', 'miliar', 'berkedok', 'investasi', 'pariwisata', 'sumbar', 'polisi', 'periksa', 'orang', 'saksi', 'emma', 'mengaku', 'sengaja', 'datang', 'ke', 'kpu', 'sumbar', 'bertepatan', 'dengan', 'peringatan', 'hari', 'ibu', 'kita', 'membawa', 'tim', 'perempuan', 'sebanyak', 'orang', 'dan', 'kemudian', 'membawa', 'bunga', 'yang', 'dibagibagikan', 'kepada', 'masyarakat', 'kata', 'emma', 'untuk', 'target', 'emma', 'berharap', 'bisa', 'terpilih', 'kembali', 'jadi', 'anggota', 'dpd', 'ri', 'kalau', 'dulu', 'suara', 'sampai', 'lebih', 'dan', 'mengalahkan', 'suara', 'pak', 'jokowimaruf', 'amin', 'sekarang', 'ya', 'mudahmudahan', 'bisa', 'kembali', 'terpilih', 'lagi', 'kata', 'emma', 'sementara', 'itu', 'komisioner', 'kpu', 'sumbar', 'gebril', 'daulay', 'yang', 'menerima', 'syarat', 'dukungan', 'dari', 'emma', 'yohanna', 'menyebutkan', 'hingga', 'sekarang', 'baru', 'satu', 'orang', 'yang', 'datang', 'menyerahkan', 'persyaratan', 'tersebut', 'buk', 'emma', 'yang', 'pertama', 'sementara', 'peminat', 'ada', 'orang', 'yang', 'mendaftar', 'di', 'akun', 'silon', 'kata', 'gebril', 'gebril', 'mengatakan', 'kpu', 'memberi', 'kesempatan', 'hingga', 'desember', 'untuk', 'pihak', 'yang', 'ingin', 'menyerahkan', 'persyaratan', 'dukungan', 'persyaratannya', 'yaitu', 'minimal', 'dukungan', 'yang', 'tersebar', 'di', 'kabupaten',

ID	Text
	'dan', 'kota', 'kata', 'gebril', 'dapatkan', 'update', 'berita', 'pilihan', 'dan', 'breaking', 'news', 'setiap', 'hari', 'dari', 'kompascom', 'mari', 'bergabung', 'di', 'grup', 'telegram', 'kompascom', 'news', 'update', 'caranya', 'klik', 'link', 'kemudian', 'join', 'anda', 'harus', 'install', 'aplikasi', 'telegram', 'terlebih', 'dulu', 'di', 'ponsel']

Contoh Teks Awal: “bagikan *tweet* cara membuat best nine instagram tanpa aplikasi best nine”

Setelah *Tokenizing* menjadi: [“bagikan”, “*tweet”,* “cara”, “membuat”, “best”, “nine”, “instagram”, “tanpa”, “aplikasi”, “best” “nine”]

Dapat dilihat pada tabel 4.8 di atas bahwa semua kata yang terdapat dalam dokumen telah diubah menujadi sekumpulan token dimana semua kata yang telah menjadi token dipisahkan oleh tanda petik (‘’) dan tanda koma (,).

8. *Stopword Removal dan Filtering*

   Hasil *stopword removal* dan *filtering* dapat dilihat pada tabel 4.9 berikut:

Tabel 4.9 Hasil *Stopword Removal & Filtering*

ID	Text
D1	bagikan tweet cara membuat best nine instagram aplikasi tentunya sekarang membuat kolase foto sulit membutuhkan aplikasi loh artikel kali bagikan tips cara membuat best nine instagram aplikasi terbaru tahun penasaran caranya apa silakan baca terus artikel selesai temanteman paham proses pembuatan kolase foto menampilkan kota foto daftar isi apa best ninebest nine top nine instagram best nine instagram

ID	Text
	aplikasi akhir kata apa best nine best nine kolase foto menampilkan foto foto terdiri kotak susunan kali gride sembilan foto tersebut pilih berdasarkan foto terbaik unggah penggunannya sepanjang tahun proses pembuatan best nine cukup mudah tentunya menggunakan aplikasi tambahan baca rekomendasi hp oppo kamera terbaik harga super terjangkau cuma jutaanbest nine top nine instagram best nine instagram tahun kalian buat situs ternama kalian kunjungi link bawah situs tersebut bernama top nine best nine top nine best nine kedua situs atas membuat kolase foto kamu unggah sepanjang tahun cara membuatnya cukup mudah disini kamu perlu memasukan foto foto kamu situs tersebut membuat kolase foto terdiri kotak grid baca tips how to develop android app menggunakan android studiountuk kedua situs atas berbeda cara penggunaanya kamu menggunakan top nine minta memasukan user ig kamu alamat email kamu menggunakan situs best nine kamu minta mengisi username ig meminta email instagram kamu keamanan privasi situs sangat aman gak perlu takut akun kamu best nine instagram aplikasi cara pembuatan best nine instagram aplikasi segera hadir merupakan awal tahun tunggu update *website* membuat best nine kolase foto terbaik tahun baca cara melihat instagram private followjika kamu membuat kolase foto tahun sebelumnya mengunjungi situs atas sangat mudah gunakan tentunya gratis menggunakan aplikasi tambahan akhir kata sedikit tutorial cara membuat best nine instagram aplikasi semoga artikel bagikan kali bermanfaat temanteman membacanya jangan

ID	Text
	lupa share artikel dukungan temanteman kemajuan web topgloablcom
D2	aku beristirahat sebentar lalu aku bertanya fani fan aku masukin penisku vagina kamu fani tampak berpikir sejenak kondisi horny pikirannya kacau fani akhirnya mengangguk lemas perlahanlahan aku merebahkan badannya ranjang menaruh kedua kakinya bahuku aku mulai menyentuhkan kepala penisku bibir liang senggamanya lalu aku perlahanlahan memasukkan batang kejantananku liang senggamanya aduh sulitnya batang kejantananku masuk liang senggamanya walaupun liang senggamanya basah liang senggamanya sangat rapat batang kejantananku perlahanlahan masuk fani mulai mengerang kesakitan aku mencoba menenangkannya aku merasakan sakit batang kejantananku ditekan liang senggamanya sempit
D3	hampir jam paksu memantat aku habis sengalsengal seluruh badan aku cipap aku tak yah cerita la rasa kembang semacam aje jalan rasa aje lepas tu rasa menyesal pasal aku dah tak virgin dah kena robek kat paksu aku rasa menyesal tu rasa berbaloi nikmat aku kecapi aku tak sangka sex sedap patutla orang nak kahwin sangat aku at least aku dah rasa nak tunggu kahwin lambat paling awal ayah aku kahwin umur tak sanggup rasanya nak tunggu tahun cipap aku ni asyik terkemutkemut aje bila tengok balak hensem last sekali upacara penutup paksu suruh aku kulum batang aku mengikut aje walaupun tak pandai aku cuba separuh aje batang paksu aku kulum kira okay jugaklah tak pengalaman macam aku ni first time aku tengok batang lelaki dewasa

ID	Text
	selalunya aku tengok konek adik aku sebesar ibu jari je terkejut aku bila tengok batang paksu hampir sebesar lengan aku panjangnya lebih kurang inci besar kepala batangnya besar macam cendawan suka betul aku bila tengok kepala batang paksu mengembang berkilat bila kena kulum masa aku kulum batang paksu ramasramas buah dada aku sekalisekali jari jahat korek lubang dubur aku pengotor betul paksu aku ni ka kata lubang dubur aku cute kalau kata cipap aku cute logik pasal cipap aku banyak bulu bulu pahat nipis halus nampak bersih cute lebih kurang pukul barulah maksu atie balik sempatlah aku aku mandi berehat lepas kena kongkek kat paksu
D4	nan
D5	padang kompascom senator asal sumatera barat emma yohanna mengungguli suara jokowimaruf amin pemilu lalu berniat maju calon anggota dewan perwakilan daerah dpd incumbent tiga periode menyerahkan syarat dukungan dukungan komisi pemilihan umum kpu sumbar kamis pemilu lalu emma berhasil mendapatkan suara jumlah menjadi tertinggi dibanding calon anggota dpd daerah pemilihan sumbar baca tangisan iringi pemakaman aiptu ruslan polisi tewas ditikam bawahannya spn polda riau pasangan capres joko widodomaruf amin sumbar mendapatkan suara hari bersama tim menyerahkan syarat dukungan suara maju calon dpd ri kata emma usai menyerahkan syarat dukungan emma menyebut dukungan tersebar kabupaten kota sumbar jumlah melampaui syarat minimal dukungan tersebar kabupaten kota sumbar baca

ID	Text
	kasus penipuan rp miliar berkedok investasi pariwisata sumbar polisi periksa orang saksi emma mengaku sengaja datang kpu sumbar bertepatan peringatan hari ibu membawa tim perempuan sebanyak orang kemudian membawa bunga dibagibagikan masyarakat kata emma target emma berharap terpilih anggota dpd ri kalau dulu suara lebih mengalahkan suara pak jokowimaruf amin sekarang mudahmudahan terpilih kata emma komisioner kpu sumbar gebril daulay menerima syarat dukungan emma yohanna menyebutkan hingga sekarang baru satu orang datang menyerahkan persyaratan tersebut buk emma pertama peminat orang mendaftar akun silon kata gebril gebril mengatakan kpu memberi kesempatan hingga desember pihak menyerahkan persyaratan dukungan persyaratannya minimal dukungan tersebar kabupaten kota kata gebril dapatkan update berita pilihan breaking news hari kompascom bergabung grup telegram kompascom news update caranya klik link kemudian join install aplikasi telegram terlebih dulu ponsel

Contoh Teks Awal: “"bagikan tweet cara membuat best nine instagram tanpa aplikasi tentunya sekarang untuk membuat kolase foto tidak begitu sulit dan juga tidak membutuhkan aplikasi”

Setelah *Stopword removal dan filtering* menjadi: “bagikan tweet cara membuat best nine instagram aplikasi tentunya sekarang membuat kolase foto sulit membutuhkan aplikasi”

Dapat dilihat pada tabel 4.9 di atas bahwa semua kata yang termasuk didalam *stoplist* dan *filtering* telah dihapus

9. *Stemming*

   Hasil stemming dapat dilihat pada tabel 4.10 berikut:

Tabel 4.10 Hasil *Stemming*

No	Text
D1	bagi tweet cara buat best nine instagram aplikasi tentu sekarang buat kolase foto sulit butuh aplikasi loh artikel kali bagi tips cara buat best nine instagram aplikasi baru tahun penasaran cara apa sila baca terus artikel selesai temanteman paham proses buat kolase foto tampil kota foto daftar isi apa best ninebest nine top nine instagram best nine instagram aplikasi akhir kata apa best nine best nine kolase foto tampil foto foto diri kotak susun kali gride sembilan foto sebut pilih dasar foto baik unggah gun panjang tahun proses buat best nine cukup mudah tentu guna aplikasi tambah baca rekomendasi hp oppo kamera baik harga super jangkau cuma jutaanbest nine top nine instagram best nine instagram tahun kalian buat situs nama kalian kunjung link bawah situs sebut nama top nine best nine top nine best nine dua situs atas buat kolase foto kamu unggah panjang tahun cara buat cukup mudah sini kamu perlu pasu foto foto kamu situs sebut buat kolase foto diri kotak grid baca tips how to develop android app guna android studiountuk dua situs atas beda cara penggunaanya kamu guna top nine minta pasu user ig kamu alamat email kamu guna situs best nine kamu minta isi username ig minta email instagram kamu aman privasi situs sangat aman gak perlu takut akun kamu best nine instagram aplikasi cara buat best nine instagram aplikasi segera hadir rupa awal tahun tunggu update *website* buat best nine kolase foto baik tahun baca cara lihat instagram private followjika kamu buat kolase foto tahun belum unjung situs atas sangat mudah guna tentu gratis guna aplikasi tambah akhir kata sedikit tutorial cara buat

No	Text
	best nine instagram aplikasi moga artikel bagi kali manfaat temanteman baca jangan lupa share artikel dukung temanteman maju web topgloablcom
D2	aku istirahat sebentar lalu aku tanya fani fan aku masukin penis vagina kamu fani tampak pikir jenak kondisi horny pikir kacau fani akhir angguk lemas perlahanlahan aku rebah badan ranjang taruh dua kaki bahuku aku mulai sentuh kepala penis bibir liang senggamanya lalu aku perlahanlahan masuk batang jantan liang senggamanya aduh sulit batang jantan masuk liang senggamanya walaupun liang senggamanya basah liang senggamanya sangat rapat batang jantan perlahanlahan masuk fani mulai erang sakit aku coba tenang aku rasa sakit batang jantan tekan liang senggamanya sempit
D3	hampir jam paksu pantat aku habis sengalsengal seluruh badan aku cipap aku tak yah cerita la rasa kembang macam aje jalan rasa aje lepas tu rasa sesal pasal aku dah tak virgin dah kena robek kat paksu aku rasa sesal tu rasa berbaloi nikmat aku kecapi aku tak sangka sex sedap patutla orang nak kahwin sangat aku at least aku dah rasa nak tunggu kahwin lambat paling awal ayah aku kahwin umur tak sanggup rasa nak tunggu tahun cipap aku ni asyik terkemutkemut aje bila tengok balak hensem last sekali upacara tutup paksu suruh aku kulum batang aku ikut aje walaupun tak pandai aku cuba paruh aje batang paksu aku kulum kira okay jugaklah tak alam macam aku ni first time aku tengok batang lelaki dewasa selalu aku tengok konek adik aku besar ibu jari je kejut aku bila tengok batang paksu

No	Text
	hampir besar lengan aku panjang lebih kurang inci besar kepala batang besar macam cendawan suka betul aku bila tengok kepala batang paksu kembang kilat bila kena kulum masa aku kulum batang paksu ramasramas buah dada aku sekalisekali jari jahat korek lubang dubur aku kotor betul paksu aku ni ka kata lubang dubur aku cute kalau kata cipap aku cute logik pasal cipap aku banyak bulu bulu pahat nipis halus nampak bersih cute lebih kurang pukul baru maksu atie balik sempat aku aku mandi rehat lepas kena kongkek kat paksu
D4	nan
D5	padang kompascom senator asal sumatera barat emma yohanna unggul suara jokowimaruf amin milu lalu niat maju calon anggota dewan wakil daerah dpd incumbent tiga periode serah syarat dukung dukung komisi pilih umum kpu sumbar kamis milu lalu emma hasil dapat suara jumlah jadi tinggi banding calon anggota dpd daerah pilih sumbar baca tangis iring makam aiptu ruslan polisi tewas tikam bawah spn polda riau pasang capres joko widodomaruf amin sumbar dapat suara hari sama tim serah syarat dukung suara maju calon dpd ri kata emma usai serah syarat dukung emma sebut dukung sebar kabupaten kota sumbar jumlah lampau syarat minimal dukung sebar kabupaten kota sumbar baca kasus tipu rp miliar kedok investasi pariwisata sumbar polisi periksa orang saksi emma aku sengaja datang kpu sumbar tepat ingat hari ibu bawa tim perempuan banyak orang kemudian bawa bunga dibagibagikan masyarakat kata emma target emma harap pilih anggota dpd

No	Text
	ri kalau dulu suara lebih kalah suara pak jokowimaruf amin sekarang mudahmudahan pilih kata emma komisioner kpu sumbar gebril daulay terima syarat dukung emma yohanna sebut hingga sekarang baru satu orang datang serah syarat sebut buk emma pertama minat orang daftar akun silon kata gebril gebril kata kpu beri sempat hingga desember pihak serah syarat dukung syarat minimal dukung sebar kabupaten kota kata gebril dapat update berita pilih breaking news hari kompascom gabung grup telegram kompascom news update cara klik link kemudian join install aplikasi telegram lebih dulu ponsel

Contoh Teks Awal: "bagikan tweet cara membuat best nine instagram aplikasi tentunya sekarang membuat kolase foto sulit membutuhkan aplikasi".

Setelah *Stemming* menjadi: "bagi tweet cara buat best nine instagram aplikasi tentu sekarang buat kolase foto sulit butuh aplikasi".

Dapat dilihat pada tabel 4.10 di atas bahwa seluruh teks yang ada dalam dokumen telah diubah menjadi bentuk yang lebih sederhana dan mudah untuk diproses tanpa kehilangan makna atau konteks dari teks awal.

10. *Remove Null Values*

    Gambar 4.8 berikut merupakan dokumen *dataset* sebelum diterapkanya *remove null values*

0	text
0	bagi tweet cara buat best nine instagram aplik...
1	aku istirahat sebentar lalu aku tanya fani fan...
2	hampir jam paksu pantat aku habis sengalsengal...
3	NaN
4	padang kompascom senator asal sumatera barat e...

Gambar 4.8 Sebelum Implementasi *Remove Null Values*

Gambar 4.9 berikut adalah dokumen *dataset* yang sudah diimplementasikan *remove null values*.

0	text
0	bagi tweet cara buat best nine instagram aplik...
1	aku istirahat sebentar lalu aku tanya fani fan...
2	hampir jam paksu pantat aku habis sengalsengal...
4	padang kompascom senator asal sumatera barat e...

Gambar 4.9 Setelah *Remove Null Values*

Seperti yang sudah dijelaskan sebelumnya bahwa keseluruhan dokumen dalam *dataset* berjumlah lima dokumen yang diberi ID D1, D2, D3, D4 dan D5, kemudian pada D4 yang diberikan nilai “nan” atau nilai kosong dimana pada kolom *text* tidak mengandung kata atau kalimat apapun maka dokumen dengan ID D4 akan dihapus dari kumpulan *dataset* yang ada agar tidak mempengaruhi proses-proses yang akan dilakukan selanjutnya

Analisa hasil dari proses *remove null values* berdasarkan gambar 4.8 dan gambar 4.9 adalah terdapat dokumen kosong di dalam *dataset* seperti pada gambar 4.8 pada dokumen 4 atau D4 kemudian diterapkannya *remove null values* dimana dokumen kosong tersebut telah dihapus seperti pada gambar 4.9 di atas. Sehingga jumlah dokumen yang tersisa berjumlah empat dokumen *dataset*, sehingga susunan ID dalam kumpulan *dataset* berubah menjadi D1 untuk dokumen satu, D2 untuk dokumen dua, D3 untuk dokumen tiga dan D4 untuk dokumen empat. Kemudian ID dari sampel data ini akan digunakan untuk tahap selanjutnya yaitu *data labeling* sehingga tidak perlu menuliskan kembali teks dari setiap dokumen yang ada dalam *dataset.*

4.1.3 Hasil *Data Labeling*

Tahap *data labeling* ini merupakan tahap meberikan label pada *dataset* apakah termasuk dalam label kelas pornografi atau non-pornografi, ada beberapa tahap yang dilakukan dalam *data labeling* ini diantaranya adalah:

1. Kamus Kata Pornografi dan Non-pornografi
   Hasil dari pembuatan kamus kata pornografi dan non-pornografi ini adalah dua kamus kata yang mengandung informasi tentang kelas label yang terkait dengan setiap kata yang ada dalam dokumen. Kamus tersebut dapat dilihat pada tabel 4.11 berikut:

Tabel 4.11 Kamus Kata Pornografi dan Non-Pornografi

Kamus Kata Pornografi	Kamus Kata Non-Pornografi
aaaaggggggghh	abadi
basah	benteng
ccccrooot	cari
desah	data

Kamus Kata Pornografi	Kamus Kata Non-Pornografi
embat	elemen
fuck	fajar
gendong	gudang
hhmmmm	haji
isap	iklan
julur	juara
kelamin	komoditas
lemas	lomba
mulus	masak
nafsu	nelayan
onani	olahraga
pentil	pemerintah

Tabel 4.11 di atas merupakan list kata dari kamus pornografi dan non-pornografi yang kemudian dimasukan ke dalam *notepad* dan disimpan dalam bentuk format *file txt* dengan nama *'porn_dict'* untuk kamus kata pornografi dan *'nonporn_dict'* untuk kamus kata non-pornografi.

2. Total Kata Dalam Dokumen

   Untuk menghitung jumlah kata dalam dokumen dalam penelitian ini diasumsikan bahwa sampel data yang digunakan sama dengan sampel data yang digunakan pada proses *data cleaning* dimana dokumen teks divisualisasikan dengan ID sama seperti pada proses

data cleaning sebelumnya yaitu sebagai "D1 untuk dokumen satu, D2 untuk dokumen dua, D3 untuk dokumen tiga dan D4 untuk dokumen empat", kemudian total kata divisualisasikan sebagai "TK".

Hasil dari proses ini adalah angka yang menunjukkan jumlah kata dalam dokumen. Hasil dari penjumlahan total kata pada dokumen dapat dilihat pada tabel 4.12 berikut:

Tabel 4.12 Total Kata Dalam Dokumen

ID	TK
D1	289
D2	85
D3	231
D4	235

Berdasarkan tabel 4.12 di atas didapatkan hasil sebagai berikut:

D1 = 289 kata

D2 = 85 kata

D3 = 231 kata

D4 = 235 kata

3. Jumlah Kata Pornografi dan Non Pornografi Dalam Dokumen

   Hasil dari proses ini adalah angka yang menunjukkan jumlah kata pornografi dan non-pornografi yang terdapat dalam dokumen berdasarkan kemunculan setiap kata pada kamus kata, jumlah kata pornografi divisualisasikan sebagai "JKP" dan jumlah kata non-pornografi divisualisasikan sebagai "JKN".

   Hasil dari penjumlahan kata pornografi dan non-pornografi dapat dilihat pada tabel 4.13 berikut:

Tabel 4.13 Jumlah Kata Pornografi dan Non-Pornografi

ID	TK	JKP	JKN
D1	289	243	283
D2	85	85	66
D3	231	206	175
D4	235	193	222

Berdasarkan tabel 4.13 diperoleh hasil dimana:

D1 = Kata pornografi (243 kata) dan kata non-pornografi (283 kata)

D2 = Kata pornografi (85 kata) dan kata non-pornografi (66 kata)

D3 = Kata pornografi (206 kata) dan kata non-pornografi (175 kata)

D4 = Kata pornografi (193 kata) dan kata non-pornografi (222 kata)

4. Persentase Pornografi dan Non Pornografi Dalam Dokumen

   Hasil dari proses ini adalah Persentase kata pornografi dan non-pornografi yang terdapat dalam setiap dokumen yang diperoleh dengan cara membagi jumlah kata pada label kelas dengan jumlah kata pada dokumen, Persentase pornografi dapat divisualisasikan sebagai “PP” dan untuk Persentase non-pornografi divisualisasikan sebagai “PN”.

   Hasil dari perhitungan Persentase pornografi dan non pornografi dapat dilihat pada gambar 4.14 berikut:

Tabel 4.14 Hitung Persentase Pornografi Dan Non-Pornogarfi

ID	TK	JKP	JKN	PP	PN
D1	289	243	283	0.84083045	0.979238754
D2	85	85	66	1	0.776470588

ID	TK	JKP	JKN	PP	PN
D3	231	206	175	0.891774892	0.757575758
D4	235	193	222	0.821276596	0.944680851

Hasil dari tabel 4.13 di atas dapat didapatakan dengan menggunakan persamaan berikut

$$Presentase = \frac{jumlah\ kata\ setiap\ kelas}{jumlah\ total\ kata} \qquad 4.1$$

Sehingga didapatkan hasil sebagai berikut:

D1 = $\frac{243}{289} = 0.84083045\ (persentase\ porn)$

$\frac{283}{289} = 0.979238754\ (persentase\ non)$

D2 = $\frac{85}{85} = 1\ (persentase\ porn)$

$\frac{66}{85} = 0.776470588\ (persentase\ non)$

D3 = $\frac{206}{231} = 0.891774892\ (persentase\ porn)$

$\frac{175}{231} = 0.757575758\ (persentase\ non)$

D4 = $\frac{193}{235} = 0.821276596\ (presentase\ porn)$

$\frac{222}{235} = 0.944680851\ (persentase\ non)$

Berdasarkan hasil perhitungan menggunakan persamaan 4.1 di atas maka disimpulkan bahwa D1 memiliki Persentase kata non-pornografi sebesar 97.08% yang mana Persentase kata non-pornografi pada D1 lebih besar daripada Persentase kata pornografi yaitu 84.08%, sedangkan D2 memiliki Persentase kata pornografi sebesar 1% dan untuk Persentase kata non-pornografi sebesar 77.64%, sama halnya dengan D2 dimana Persentase kata pornografi pada D3 lebih besar dari Persentase kata non-pornografi dimana

persentase kata pornografi untuk D3 sebesar 89.17% sedangkan untuk Persentase kata non-pornografi pada D3 sebesar 75.75% kemudian untuk D4 Persentase dari kata pornografi lebih kecil dibandingkan dengan Persentase kata non-pornografi yang mana Persentase untuk kata pornografi pada D4 sebesar 82.12% sedangkan Persentase kata non-pornografi sebesar 94.46% .

5. *Labeling Data*

Setelah menghitung Persentase maka dokumen akan diberi label berdasarkan Persentase kata yang terdapat dalam dokumen, apabila Persentase kata pornografi lebih besar dari Persentase kata non-pornografi maka, dokumen akan diberi label sebagai pornografi (YA) dan apabila Persentase kata non-pornografi lebih besar dari Persentase kata pornografi maka dokumen diberi label non-pornografi (TIDAK).

Hasil dari pemberian label pornografi dan non-pornografi data dapat lihat pada tabel 4.15 berikut:

Tabel 4.15 Hasil *Labeling Data*

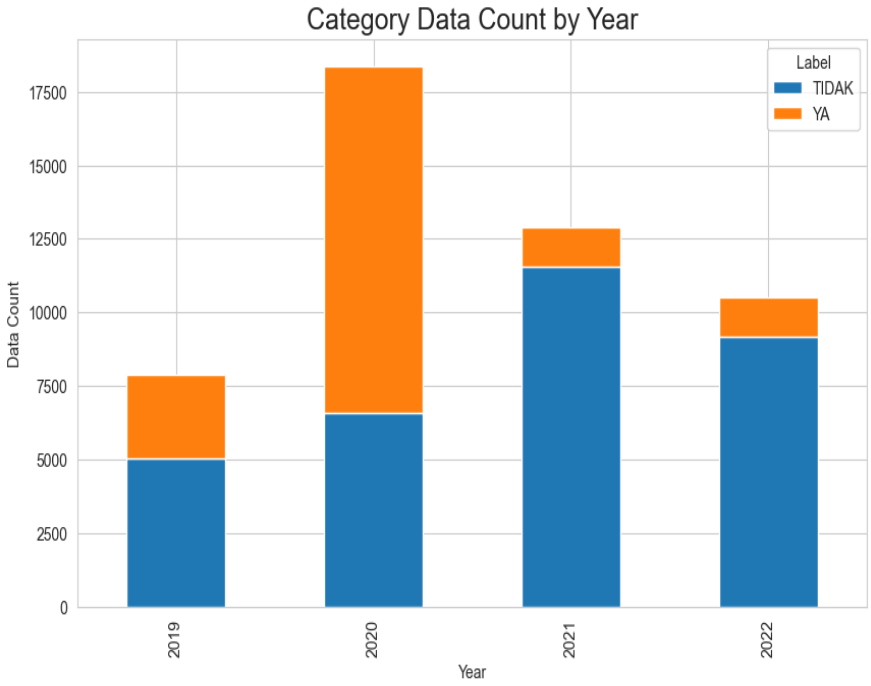
ID	TK	JKP	JKN	PP	PN	label
D1	289	243	283	0.84083045	0.979238754	TIDAK
D2	85	85	66	1	0.776470588	YA
D3	231	206	175	0.891774892	0.757575758	YA
D4	235	193	222	0.821276596	0.944680851	TIDAK

Berdasarkan tabel 4.15 di atas hasil yang diperoleh adalah:

D2 dan D3 diberi label “pornografi” karena Persentase kata pornografi pada D2 dan D3 lebih besar dibandingkan dengan Persentase kata non-pornografi, dimana Persentase kata pornografi sebesar 1% untuk D1 serta 89.17% untuk D3 dan sebaliknya yaitu

pada D1 dan D4 diberi label “non pornografi” karena Persentase kata non-pornografi lebih besar dibandingkan dengan Persentase kata pornografi dimana Persentase kata non-pornogrrafi yaitu sebesar 97.92% untuk D1 dan 94.46% untuk D4.

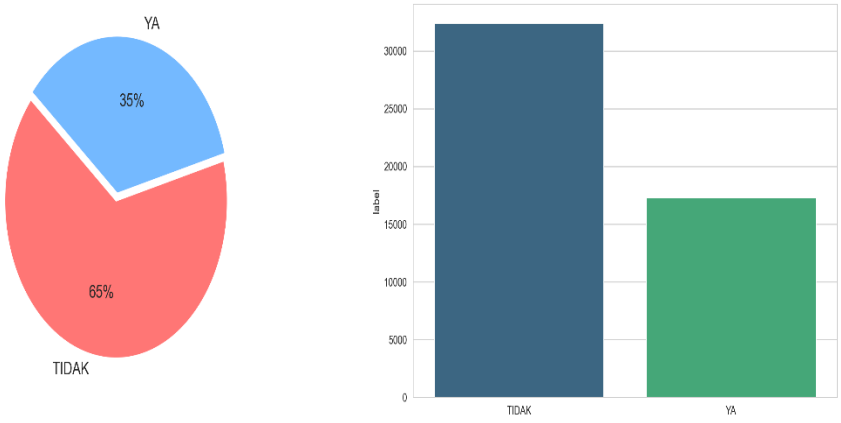
Setelah memberikan label pada *dataset* maka secara keseluruhan *dataset* dapat divisualisasikan berdasarkan label dokumen per tahun sebagai berikut:



Gambar 4.10 Label Data Berdasarkan Tahun

Analisa berdasarkan gambar 4.10 di atas adalah pada tahun 2019 label non-pronografi (TIDAK) sebanyak 5.058 ribu data, untuk label pornografi (YA) sebanyak 2.834 ribu data, pada tahun 2020 label non-pornografi (TIDAK) sebanyak 6.593 ribu data dan label pornografi (YA) sebanyak 11.781 ribu data, kemudian untuk tahun 2021 label non-pornogarfi (TIDAK) sebanyak 11.576 ribu data serta label pornografi (YA) sebanyak 1.332 ribu data dan yang terakhir yaitu tahun 2022 label non-pornogarfi sebanyak 9.177 ribu data sedangkan untuk label pornografi sebanyak 1.351 ribu data.

Maka secara keseluruhan jumlah label data dapat dilihat pada gambar 4. 11 berikut:



Gambar 4.11 Keseluruhan Label

Berdasarkan gambar 4.11 di atas didapatkannya Persentase keseluruhan jumlah label dalam *dataset* dimana Persentase untuk label kelas non-pornografi lebih besar dibandingkan dengan label kelas pornografi, dimana label kelas non-pornogarfi memiliki Persentase sebesar 65% dengan jumlah label data dalam *dataset* sebanyak 32.405 ribu data sedangkan Persentase dari label kelas pornografi adalah sebesar 35% dengan jumlah label dalam *dataset* sebanyak 17.297 ribu data.

Setelah memberikan label untuk kelas pornografi dan non-pornografi pada *dataset* kemudian dilakukan perubahan atau transformasi label menjadi angka dengan nilai 0 sebagai label non-pornografi (TIDAK) dan 1 sebagai label pornografi (YA).

Gambar 4.12 berikut merupakan hasil dari transformasi label menjadi angka:

	text	label
0	bagi tweet cara buat best nine instagram aplik...	0
1	aku istirahat sebentar lalu aku tanya fani fan...	1
2	hampir jam paksu pantat aku habis sengalsengal...	1
3	padang kompascom senator asal sumatera barat e...	0

Gambar 4.12 Hasil *Transform Label*

Pada gambar 4.12 di atas dapat dilihat bahwa label telah diubah menjadi 0 dan 1 yang sebelumnya masih dalam bentuk tipe data *string* atau teks yaitu YA dan TIDAK. Tujuan dari mengubah label menjadi *integer* atau *numeric* adalah agar label data menjadi lebih mudah diproses oleh model *machine learning*

4.2 Hasil *Feature Engineering*

Tahap selanjutnya yang dilakukan adalah *feature engineering* yang bertujuan untuk memeriksa kerja fitur dengan model yang dibuat dan meningkatkan fitur untuk direpresentasikan dalam bentuk *numeric* data atau data *matrix*.

Sebelum diimplementasikan nya *feature engineering*, hal pertama yang dilakukan terlebih dahulu adalah dengan membagi *dataset* menjadi data *training* dan *data testing* dengan rasio pembagian *data training* dan *data testing* adalah 70:30. Dimana hasil dari pembagian *data training* dan *data testing* berjumlah 34.791 ribu *data training* dan 14.911 ribu data untuk *data testing.*

Setelah pembagian *data training* dan *data testing* telah diterapkan maka *feature engineering* siap diimplementasikan, terdapat 2 proses yang dilakukan dalam *feature engineering* ini, proses tersebut dapat dilihat sebagai berikut:

4.2.1 *Term Frequency-Inverse Document Frequency* (*TF-IDF*)

Hasil perhitungan *TFIDF* ini akan dilampirkan pada bagian lampiran "Tabel Hasil Perhitungan *TFIDF*"

1. *Term frequency*

   Hasil dari perhitungan *term frequency* menggunakan persamaan 3.1 dapat dilihat pada gambar 4.13 berikut:

Term	TF			
	D1	D2	D3	D4
adik	0	0	1	0
aduh	0	1	0	0
aiptu	0	0	0	1
aje	0	0	5	0
akun	1	0	0	1
alam	0	0	1	0
alamat	1	0	0	0
aman	2	0	0	0
amin	0	0	0	3
android	2	0	0	0

Gambar 4.13 Hasil *Term Frequency*

Hasil *"term frequency"* pada gambar 4.13 di atas menunjukkan frekuensi kemunculan term atau kata-kata pada empat dokumen yang berbeda (D1, D2, D3, dan D4) dalam bentuk tabel. *"TF"* merupakan singkatan dari *term frequency*, yaitu jumlah kemunculan kata-kata tersebut dalam setiap dokumen.

Sebagai contoh, kata "adik" muncul satu kali dalam dokumen D3, sedangkan kata "aduh" muncul satu kali dalam dokumen D2. Kata "aiptu" muncul satu kali dalam dokumen D4, dan kata "aje" muncul lima kali dalam dokumen D3. Kata "akun" muncul satu kali dalam dokumen D1 dan satu kali dalam dokumen D4.

Setelah mendapatkan nilai dari *term frequency* kemudian dilakukan perhitungan untuk mencari bobot nilai dari setiap *term* dengan persamaan 3.2

Dimana *"term" frekuensi dalam dokumen j"* adalah jumlah kemunculan *"term" dalam dokumen j* dan *"total kata dalam dokumen j"* adalah *jumlah total kata dalam dokumen j*. hasil dari perhitungan bobot *term frequency* dapat dilihat pada gambar 4.14 berikut:

Term	TF				F	Bobot Nilai TF			
	D1	D2	D3	D4		D1	D2	D3	D4
adik	0	0	1	0	1	0	0	0.25	0
aduh	0	1	0	0	1	0	0.25	0	0
aiptu	0	0	0	1	1	0	0	0	0.25
aje	0	0	5	0	1	0	0	1.25	0
akun	1	0	0	1	2	0.25	0	0	0.25
alam	0	0	1	0	1	0	0	0.25	0
alamat	1	0	0	0	1	0.25	0	0	0
aman	2	0	0	0	1	0.5	0	0	0
amin	0	0	0	3	1	0	0	0	0.75

Gambar 4.14 Hasil Hitung Bobot *Term Frequency*

Gambar 4.14 di atas menunjukkan perhitungan TF dan bobot nilai TF dari suatu dokumen (D1, D2, D3, D4) terhadap beberapa term (adik, aduh, aiptu, aje, akun, alam, alamat, aman, amin, android).

Pada kolom TF, angka-angka menunjukkan frekuensi kemunculan term tersebut dalam setiap dokumen. Misalnya, pada dokumen D3, term "adik" muncul sebanyak satu kali, sehingga nilai TF-nya adalah 1.

Setelah nilai *TF* dihitung, kemudian dihitung bobot nilai *TF*-nya pada kolom Bobot Nilai *TF*. Bobot ini menggambarkan tingkat pentingnya suatu term dalam dokumen tersebut dibandingkan dengan dokumen lainnya. Bobot ditentukan dengan mengalikan

nilai *TF* dengan logaritma dari jumlah dokumen (dalam hal ini 4) dibagi dengan jumlah dokumen yang mengandung term tersebut. Misalnya, bobot nilai TF term "adik" pada dokumen D3 adalah 0.25.

Dengan demikian, tabel tersebut memberikan informasi tentang frekuensi kemunculan suatu *term* dalam dokumen dan tingkat pentingnya term tersebut dalam dokumen tersebut dibandingkan dengan dokumen lainnya.

Cara menghitung nilai *term frequency* berdasarkan persamaan 3.2 adalah sebagai berikut:

Terdapat empat dokumen dalam *dataset* yang divisualisasikan sebagai D1, D2, D3 dan D4. Maka untuk menghitung bobot *term frequency* dari kata adik dalam D1, D2, D3 dan D4 adalah:

$$TF(\text{adik,D1}) = \frac{0}{4} = 0$$

$$TF(\text{adik,D2}) = \frac{0}{4} = 0$$

$$TF(\text{adik,D3}) = \frac{1}{4} = 0.25$$

$$TF(\text{adik,D4}) = \frac{0}{1} = 0$$

Dengan demikian dapat disimpulkan bahwa bobot *term frequency* dari *term "adik"* pada D1 sama dengan 0, D2 sama dengan 0, D3 sama dengan 0.25 dan D4 sama dengan 0.

2. *Inverse Document Frequency (IDF)*

   Persamaan yang digunakan untuk menghitung *IDF* adalah dengan menggunakan persamaan 3.3. Dimana "total dokumen + 1" adalah jumlah total dokumen dalam *dataset* ditambah satu dan "dokumen yang mengandung term + 1" adalah jumlah total dokumen yang mengandung term dalam *dataset* ditambah satu.

Untuk menghitung *IDF* dapat menggunakan persamaan 3.3 pada bab sebelumnya, kemudaian gambar 4.15 berikut adalah hasil perhitungan *IDF*

Term	IDF
adik	1.916290732
aduh	1.916290732
aiptu	1.916290732
aje	1.916290732
akun	1.510825624
alam	1.916290732
alamat	1.916290732
aman	1.916290732
amin	1.916290732
android	1.916290732

Gambar 4.15 Hasil Hitung *IDF*

Bobot nilai *IDF (Inverse Document Frequency)* pada tabel tersebut menunjukkan besarnya bobot relatif dari setiap term dalam seluruh dokumen yang dianalisis. Nilai *IDF* dihitung dengan rumus, di mana jumlah dokumen ditambah satu dibagi dengan jumlah dokumen yang mengandung *term* ditambah satu dan ditambah satu.

Sebagai contoh, pada term "adik" terdapat di dokumen D3 saja dengan frekuensi TF 1. Nilai IDF untuk *term* "adik" dihitung dengan LOG((4+1)/(1+1))+1 = 1.916290732, di mana 4 adalah jumlah seluruh dokumen dan 1 adalah jumlah dokumen yang mengandung term "adik". Dalam hal ini, semakin tinggi nilai *IDF* suatu term, semakin tinggi bobot relatifnya dalam seluruh dokumen.

Dari tabel tersebut, term dengan bobot *IDF* tertinggi adalah "adik", "aduh", "aiptu", "aje", "alam", "alamat", "aman", "amin", dan "android", dengan nilai IDF yang sama yaitu 1.916290732. Ini menunjukkan bahwa term-term tersebut memiliki bobot relatif yang sama tingginya dalam seluruh dokumen. Sedangkan *term* dengan bobot IDF terendah adalah "akun", dengan nilai IDF sebesar 1.510825624.

Maka untuk menghitung *IDF* dari term *"adik"* menggunakan persamaan 3.3 adalah sebagai berikut:

$$IDF(adik) = LOG\,\frac{(4+1)}{1+1} + 1 = \ 1.916290732$$

Berdasarkan hasil perhitungan di atas maka *IDF* dari *term "adik"* adalah 1.916290732

3. TFIDF

Setelah mendapatkan nilai dari *term frequency* dan nilai *IDF* maka untuk menghitung bobot *TFIDF* dapat dilakukan menggunakan persamaan 3.4 pada bab sebelumnya maka hasil dari perhitungan *TF-IDF* dan cara untuk menghitung *TF-IDF* dari term adik adalah sebagai berikut:

	TFIDF			
	D1	**D2**	**D3**	**D4**
adik	**0**	**0**	**1.916290732**	**0**
aduh	**0**	**1.916290732**	**0**	**0**
aiptu	**0**	**0**	**0**	**1.916290732**
aje	**0**	**0**	**9.581453659**	**0**
akun	**1.510825624**	**0**	**0**	**1.510825624**
alam	**0**	**0**	**1.916290732**	**0**
alamat	**1.916290732**	**0**	**0**	**0**
aman	**3.832581464**	**0**	**0**	**0**
amin	**0**	**0**	**0**	**5.748872196**
android	**3.832581464**	**0**	**0**	**0**

Gambar 4.16 Hasil Hitung *TF-IDF*

TFIDF (Term Frequency - Inverse Document Frequency) adalah suatu metode penghitungan bobot kata yang biasanya digunakan dalam *text mining* atau *information retrieval* untuk mengevaluasi seberapa penting suatu kata dalam suatu dokumen atau kumpulan dokumen. Pada gambar 4.13, *TF (Term Frequency)* merupakan jumlah kemunculan suatu kata dalam suatu dokumen, *IDF (Inverse Document Frequency)* pada gambar 4.15 adalah kebalikan dari jumlah dokumen yang mengandung kata tersebut, sedangkan *TFIDF* adalah perkalian antara *TF* dan *IDF*. Semakin besar nilai *TFIDF* suatu kata dalam suatu dokumen, maka semakin penting kata tersebut dalam dokumen tersebut.

Pada gambar 4.16 di atas, setiap term (kata) memiliki frekuensi *TF* seperti pada gambar 4.13 dan bobot nilai *IDF* seperti pada gambar 4.15 maka Sebagai contoh untuk menghitung bobot nilai *TFIDF* adallah, kata "adik" muncul hanya pada dokumen D3, sehingga nilai TF untuk D3 adalah 1, sedangkan untuk dokumen lainnya nilai TF-nya adalah 0. Nilai IDF untuk "adik" adalah 1.916290732, karena kata "adik" hanya muncul pada satu dokumen. Nilai *TFIDF* untuk "adik" pada D3 adalah 1.916290732 *(TF * IDF).*

Demikian pula, kata "android" muncul pada dokumen D1 dengan jumlah kemunculan sebanyak dua kali. Nilai *IDF* untuk "android" adalah 3.832581464, karena kata "android" muncul pada D2 sebanyak dua kali. Nilai *TFIDF* untuk "android" pada D1 adalah 3.832581464 *(TF x IDF).*

Perhitungan bobot *TFIDF* dari *term* “adik” dapat dihitung menggunakan persamaan 3.4 sebagai berikut:

$$TFIDF(adik) = (0, D1) * (1.916290732) = 0$$

$$TFIDF(adik) = (0, D2) * (1.916290732) = 0$$

$$TFIDF(adik) = (1, D3) * (1.916290732) = 1.916290732$$
$$TFIDF(adik) = (0, D4) * (1.916290732) = 0$$

Perhitungan *TFIDF* yang telah dihitung di atas merupakan perhitungan *TFIDF* dengan menggunakan persamaan yang digunakan ketika menghitung *TFIDF* pada *python*, berbeda halnya apabila menghitung atau mencari bobot *TFIDF* secara manual. Untuk menghitung bobot *TFIDF* secara manual tanpa menggunakan *python* maka persamaan yang digunakan adalah persamaan yang terdapat pada bab dua yaitu persamaan 2.5, 2.6 dan 2.7, perhitungan *TFIDF* menggunakan ketiga persamaan tersebut akan dilampirkan pada halaman lampiran dengan nama "Tabel Perhitungan *TFIDF* Tanpa Menggunakan *Python*".

4.2.2 *Synthetic Minority Over-Sampling* (*SMOTE*)

Hasil *SMOTE* dapat dilihat pada gambar 4.17 berikut:

```
from imblearn.over_sampling import SMOTE
print("Before OverSampling, counts of label '1': {}".format(sum(y_train==1)))
print("Before OverSampling, counts of label '0': {} \n".format(sum(y_train==0)))

sm = SMOTE(random_state=2)
X_train_res, y_train_res = sm.fit_resample(X_train, y_train.ravel())

print('After OverSampling, the shape of train_X: {}'.format(X_train_res.shape))
print('After OverSampling, the shape of train_y: {} \n'.format(y_train_res.shape))

print("After OverSampling, counts of label '1': {}".format(sum(y_train_res==1)))
print("After OverSampling, counts of label '0': {}\n".format(sum(y_train_res==0)))
```

```
Before OverSampling, counts of label '1': 12108
Before OverSampling, counts of label '0': 22683

After OverSampling, the shape of train_X: (45366, 5000)
After OverSampling, the shape of train_y: (45366,)

After OverSampling, counts of label '1': 22683
After OverSampling, counts of label '0': 22683
```

Gambar 4.17 Hasil Implementasi Teknik *SMOTE*

Pada gambar 4.17 di atas merupakan penerapan teknik *SMOTE* (*Synthetic Minority Over-sampling Technique*) untuk menangani ketidakseimbangan kelas (*imbalanced class*) pada data *training*.

Pada awalnya, dihitung jumlah sampel untuk label '1' dan label '0' pada data *training* yang belum di-*oversampling* dengan jumlah kelas pada label 1 adalah 12.108 ribu kelas dan jumlah kelas pada label 0 berjumlah 22.683 ribu kelas.

Sebelum diterapkannya teknik *SMOTE* ukuran *set data training* untuk *X_train* adalah 34.791 ribu dengan jumlah fitur sebanyak 5.000 ribu fitur serta untuk *y_train sebanyak* 34.791 ribu, kemudian diimplementasikannya teknik *SMOTE* dimana, objek *SMOTE* dibuat dengan parameter *random_state=2*. *SMOTE* kemudian diterapkan pada *X_train* dan *y_train* menggunakan metode *fit_resample(),* sehingga *X_train_res* dan *y_train_res* adalah hasil dari *data training* yang sudah di-*oversampling*. Hasil dari *oversampling training set X_train* adalah 45.366 ribu data dengan 5.000 fitur dan untuk *y_train* sebesar 45.366 ribu data

Setelah *oversampling*, dihitung kembali jumlah sampel untuk label '1' dan label '0' pada y_train_res untuk memastikan bahwa jumlah sampel untuk kedua label sudah seimbang. Sebelum *oversampling* label 1 berjumlah 12.108 ribu label data 1 dan untuk label 0 berjumlah 22.683 ribu label data 0.

Dapat dilihat pada gambar 4.17 bahwa setelah diterapkannya teknik *SMOTE* kedua jumlah kelas pada label 0 dan 1 telah seimbang dengan jumlah kelas pada kedua kelas label tersebut sebanyak 22.683 ribu kelas label data.

4.3 Hasil *Model Training*

Model training merupakan tahap yang dilakukan untuk melatih model *Machine Learning* yang dibuat serta penerapan algoritma *Naïve Bayes* untuk klasifikasi teks atau dokumen data, proses *model training* yang dilakukan adalah sebagai berikut:

4.3.1 *Model Selection*

Terdapat tiga model klasifikasi dalam algoritma *Naïve Bayes* yaitu *Gaussian Naïve Bayes*, *Bernoulli Naïve Bayes* dan *Multinomial Naïve Bayes*, ketiga model klasifikasi tersebut di *training* untuk mendapatkan model klasifikasi *Machine Learning* yang terbaik dimana model dengan nilai akurasi yang tinggi merupakan model yang terbaik. Hasil dari *model selection* dapat dilihat pada tabel 4.16 berikut:

Tabel 4.16 Hasil *Model Selection*

Model	**Akurasi**	***Precision***	***Recall***	***F1-Score***
MultinomialNB	93.38072564	0.857312925	0.971478127	0.910832053
BernoulliNB	93.37401918	0.898803873	0.912314511	0.905508799
GaussianNB	92.60277647	0.836462451	0.97880131	0.902051328

Hasil yang diberikan pada tabel 4.16 di atas merupakan hasil evaluasi performa dari tiga jenis model klasifikasi Naive Bayes yang berbeda, yaitu *MultinomialNB, BernoulliNB,* dan *GaussianNB*. Evaluasi performa tersebut dilakukan dengan membandingkan hasil prediksi model dengan nilai sebenarnya dari data yang digunakan sebagai *dataset*.

Akurasi: persentase klasifikasi benar dari seluruh data yang diprediksi. Semakin tinggi nilai akurasi, semakin baik performa model dalam melakukan klasifikasi.

Precision: rasio antara jumlah *true positive (TP)* dengan jumlah seluruh data yang diprediksi sebagai positif *(TP+FP*). Precision mengukur sejauh mana model memberikan hasil positif yang benar (tidak salah dinyatakan positif).

Recall: rasio antara jumlah *true positive (TP)* dengan jumlah seluruh data yang seharusnya diprediksi sebagai positif *(TP+FN). Recall* mengukur sejauh mana model mampu menemukan data yang seharusnya diprediksi positif (tidak ada data positif yang terlewat).

F1-Score: rata-rata harmonik antara *precision* dan *recall*. *F1-Score* menggabungkan nilai *precision* dan *recall* untuk memberikan gambaran yang lebih baik tentang performa model secara keseluruhan. Semakin tinggi nilai *F1-Score*, semakin baik performa model dalam melakukan klasifikasi.

Berdasarkan hasil evaluasi performa tersebut, dapat dilihat bahwa model *MultinomialNB* dan *BernoulliNB* memiliki akurasi yang lebih baik dibandingkan dengan model *GaussianNB*, dimana akurasi *MultinomialNB* adalah 93.38% dan *BernoulliNB* sebesar 93.37% sedangkan *GaussianNB* sebesar 92.60%. Model *MultinomialNB* juga memiliki *precision* yang lebih rendah dan *recall* yang lebih tinggi dibandingkan dengan model *BernoulliNB* dimana *MultinomialNB* memiliki nilai *precision* sebesar 85.73% dan *recall* sebesar 97.14% sedangkan *BernoulliNB* memiliki nilai *precision* sebesar 89.88% dan *recall* 91.23% , meskipun *F1-Score* keduanya hampir sama yaitu *F1-score* untuk *MultinomialNB* sebesar 91.08% dan *F1-score* untuk *BernoulliNB* sebesar 90.55%. Hal ini menunjukkan bahwa model *MultinomialNB* lebih unggul dalam menemukan data yang seharusnya diprediksi sebagai positif (*recall*), namun model *BernoulliNB* lebih akurat dalam memprediksi data sebagai positif (*precision*). Sementara itu, model *GaussianNB* memiliki nilai *recall* yang sangat tinggi sebesar 97.88%, namun *precision*-nya lebih rendah sebesar 89.88% dibandingkan dengan model lainnya, sehingga menyebabkan nilai *F1-Score* yang lebih rendah yaitu sebesar 90.20%.

Berdasarkan hasil penjelasan di atas maka sebagai hasil akhir dari *model selection* adalah dengan menggunakan model *Multinomial Naïve Bayes* yang merupakan model dengan akurasi atau performa kinerja model yang terbaik

dibandingkan dengan model *Gaussian Naïve Bayes* dan *Bernoulli Naïve Bayes*.

4.3.2 *Grid Search*

Grid search dilakukan untuk mencari *hyperparameter* yang terbaik untuk model *Machine Learning* dengan tujuan untuk meningkatkan performa dan akurasi model klasifikasi. Hasilnya dapat dilihat pada gambar 4.18 dibawah ini:

```
for i,j in grid_models:
    grid = GridSearchCV(estimator=i,param_grid = j, scoring = 'accuracy',cv = 5)
    grid.fit(X_train_res, y_train_res)
    best_accuracy = grid.best_score_
    best_param = grid.best_params_
    print('{}:\nBest Accuracy : {:.2f}%'.format(i,best_accuracy*100))
    print('Best Parameters : ',best_param)
    print('')
    print('----------------')
    print('')
```

```
MultinomialNB():
Best Accuracy : 94.61%
Best Parameters :  {'alpha': 0.1}
```

Gambar 4.18 Hasil *Grid Search*

Gambar 4.18 merupakan implementasi dari *grid search* pada model klasifikasi *Multinomial Naïve Bayes*. *Grid search* dilakukan dengan menggunakan parameter alpha pada model dan menghitung nilai akurasi dengan matrik *'accuracy'* menggunakan *cross-validation* dengan *5 fold*.

Pada setiap iterasi, parameter *grid* yang diberikan diiterasi dan diuji menggunakan *GridSearchCV*. Untuk menemukan parameter terbaik dan nilai akurasi tertinggi. Pada output yang dihasilkan seperti pada gambar 4.18 di atas, dapat dilihat bahwa model *Multinomial Naïve Bayes* memiliki akurasi terbaik sebesar 94.61% dengan parameter alpha sebesar 0.1. Ini menunjukkan bahwa parameter *alpha* yang lebih rendah menghasilkan akurasi yang lebih baik pada model *Multinomial Naïve Bayes*, dan model ini dapat digunakan untuk melakukan klasifikasi pada data yang belum dilihat sebelumnya.

4.3.3 *Model Tuning Hyperparameter*

Finalisasi model klasifikasi *Machine Learning* berdasarkan akurasi model terbaik pada proses *model selection* dan penggunaan parameter yang terbaik untuk model klasifikasi *Machine Learning* pada proses *grid search*. Hasilnya dapat dilihat pada gambar 4.19 berikut:

```
#Fitting MNB Model
classifier = MultinomialNB(alpha = 0.1)
classifier.fit(X_train_res, y_train_res)

y_pred = classifier.predict(X_test)
y_prob = classifier.predict_proba(X_test)[:,1]
cm = confusion_matrix(y_test, y_pred, labels=[1,0])

print(classification_report(y_test, y_pred))
print(f'ROC AUC score: {roc_auc_score(y_test, y_prob)}')
print('Accuracy Score: ',accuracy_score(y_test, y_pred))
```

Gambar 4.19 *Model Tuning Hyperparameter*

Pemilihan nilai *alpha=0.1* sebagai *hyperparameter* dalam model *MultinomialNB* dapat meningkatkan akurasi model karena *alpha* adalah parameter *smoothing* yang digunakan untuk menghindari nilai probabilitas nol dalam model.

Probabilitas nol dalam model referensi pada situasi ketika probabilitas suatu kejadian dalam data pelatihan adalah nol. Misalnya, dalam klasifikasi teks, jika ada kata yang tidak muncul dalam data pelatihan, probabilitasnya akan menjadi nol dan ini akan menyebabkan kesalahan dalam perhitungan probabilitas pada saat pengujian. Ini dapat terjadi pada model probabilitas yang tidak dihaluskan atau tanpa *smoothing*. Oleh karena itu, *smoothing* seperti yang dilakukan dalam model *MultinomialNB* digunakan untuk menghindari probabilitas nol dan memberikan probabilitas yang lebih realistis pada kata-kata yang mungkin tidak muncul dalam data pelatihan. Penggunaan parameter *smoothing* (atau sering disebut dengan *smoothing*) adalah teknik yang digunakan untuk mengurangi *overfitting* pada model statistik, terutama pada model yang kompleks atau dengan jumlah parameter yang besar.

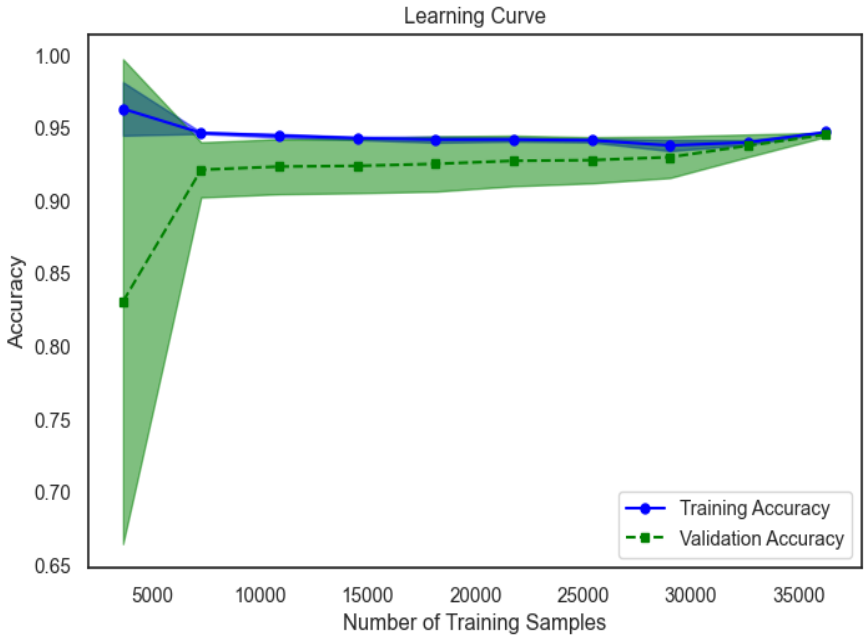
Sehingga *output* yang dihasilkan seperti pada gambar 4.19 dalam *tuning hyperparameter* ini adalah terjadinya peningkatan performa model *MultinomialNB* dari 93.38% menjadi 93.50% setelah menggunakan *hyperparameter* menunjukkan bahwa *hyperparameter* yang dipilih berhasil meningkatkan kinerja model walaupun peningkatan performa model *Multinomial Naïve Bayes* dengan menggunakan *hyperparameter* ini tidak terlalu signifikan.

4.3.4 *Report Model Classification*

Report model classification dapat dilihat pada poin-poin berikut:

1. *Overfitting* dan *Underfitting*

   Hasil pengecekan *overfitting* dan *underfitting* pada model *Machine Learning* yang telah dibuat dapat dilihat pada gambar 4.20 *learning curve* dan gambar 4.21 *validation curve*:



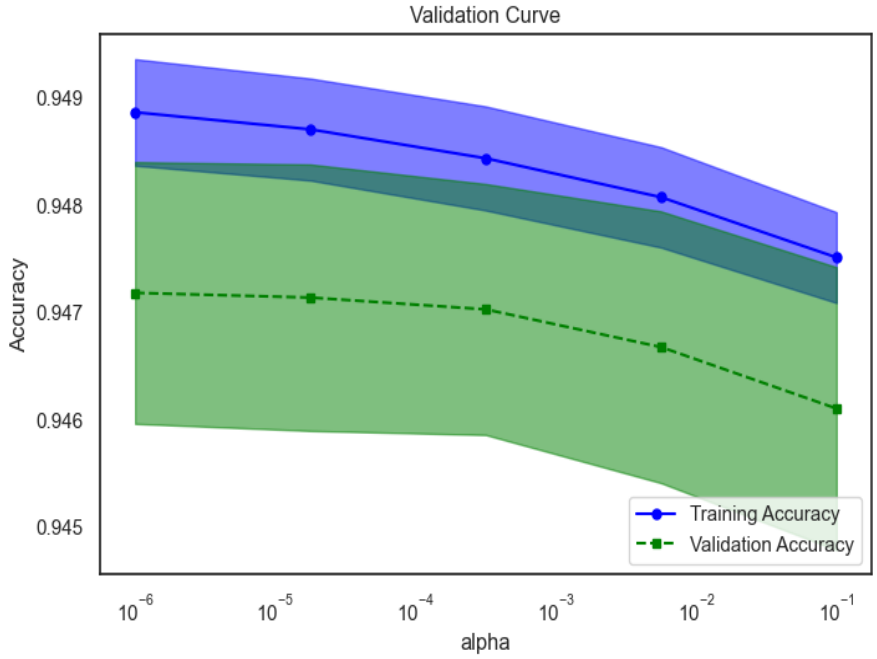
Gambar 4.20 *Learning Curve*

Learning curve menunjukkan kinerja model dalam hubungannya dengan ukuran *dataset* pelatihan yang digunakan.

Pada sumbu x, *learning curve* menunjukkan jumlah data pelatihan yang digunakan untuk melatih model.

Pada sumbu y, *learning curve* menunjukkan kinerja model, yaitu akurasi.

Berdasarkan *learning curve* di atas menunjukan bahwa seiring ditambahnya data *training* maka akurasi dari dari model atau performa model semakin stabil dan berhenti pada maksimum data *training* yaitu 35.000 data *training* dengan performa atau kinerja model berdasarkan *training* dan *validation* data sebesar 94%. Hal ini menunjukkan bahwa model sudah *overfitting* dan tidak akan meningkatkan kinerjanya meskipun data *training* diperbesar.



Gambar 4.21 *Validation Curve*

Validation curve menunjukkan kinerja model dalam hubungannya dengan *hyperparameter* yang digunakan dalam model

Pada sumbu x, *validation curve* menunjukkan nilai *hyperparameter* yang berbeda yang diuji.

Pada sumbu y, *validation curve* menunjukkan kinerja model, yaitu nilai akurasi.

Berdasarkan *validation curve* di atas menunjukan bahwa seiring menurunnya nilai *hyperparameter*, performa model semakin menurun Hal ini menunjukkan bahwa *hyperparameter* perlu disetel ke nilai yang lebih rendah untuk meningkatkan kinerja model.

Berdasarkan gambar di atas serta analisa yang diberikan sebelumnya di mana nilai *hyperparameter* harus diturunkan tidak menjadi suatu masalah atau mempengaruhi kinerja model yang diberikan dalam penelitian ini karena perbedaan nilai antara akurasi dan validasi nilai *hyperparameter* tidak terlampau jauh dan masih dapat dibandingkan dimana pada gambar di atas nilai akurasi dari *training set* 0.948% dan nilai akurasi dari validation adalah 0.947%.

Selain kedua cara di atas secara sederhana untuk mengetahui apakah model yang telah dibuat ini mengalami *overfitting* atau *underfitting* dengan melihat nilai dari *training set* dan nilai dari *test set* seperti pada gambar 4.22 berikut:

```
# print the scores on training and test set

print('Training set score: {:.4f}'.format(classifier.score(X_train_res, y_train_res)))

print('Test set score: {:.4f}'.format(classifier.score(X_test, y_test)))
```

```
Training set score: 0.9473
Test set score: 0.9351
```

Gambar 4.22 *Simple Overfitting* & *Underfitting*

Dapat dilihat bahwa model yang dibangun menggunakan *MultinomialNB* dengan *hyperparameter alpha=0.1* memiliki akurasi yang tinggi pada data pelatihan dan data pengujian.

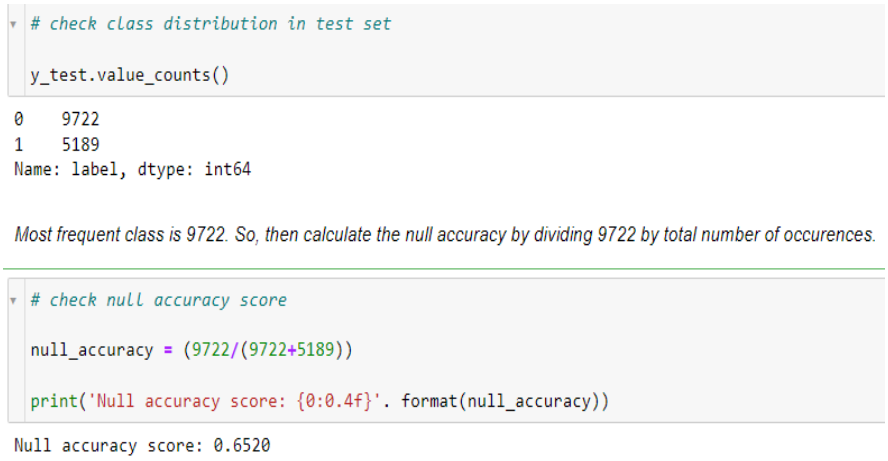
Training set score menunjukkan akurasi model pada data pelatihan, di mana model mencapai akurasi sebesar 0.9473 atau sekitar 94.73%. Artinya, model mampu memprediksi dengan benar sekitar 94.73% dari data pelatihan yang digunakan untuk melatih model.

Test set score menunjukkan akurasi model pada data pengujian yang belum pernah dilihat sebelumnya, di mana model mencapai akurasi sebesar 0.9351 atau sekitar 93.51%. Artinya, model mampu memprediksi dengan benar sekitar 93.51% dari data pengujian yang belum pernah dilihat sebelumnya.

Hasil ini menunjukkan bahwa model *MultinomialNB* dengan *hyperparameter alpha=0.1* memiliki kinerja yang baik dalam melakukan klasifikasi teks pada data yang belum pernah dilihat sebelumnya. Cara mendapatkan nilai akurasi ini adalah dengan menggunakan metode *score* dari kelas *MultinomialNB*. Pada data pelatihan, metode *score* dipanggil dengan memasukkan X_train_*res* dan *y_train_res* sebagai parameter untuk menghitung akurasi pada data pelatihan. Pada data pengujian, metode *score* dipanggil dengan memasukkan *X_test* dan *y_test* sebagai parameter untuk menghitung akurasi pada data pengujian.

2. *Null Accuracy*

   Menghitung *null accuracy* model klasifikasi yang dapat dilihat pada gambar 4.23 berikut:



Gambar 4.23 Hasil Hitung *Null Accuracy*

Dari output yang dihasilkan pada gambar 4.23 di atas, dapat dilihat bahwa data pengujian memiliki jumlah sampel sebanyak 14.911, di

mana sebanyak 9.722 (sekitar 65.20%) di antaranya termasuk ke dalam kelas 0 (Non-pornografi / TIDAK) dan sebanyak 5.189 (sekitar 34.80%) termasuk ke dalam kelas 1 (Pornografi / YA).

Null accuracy score menunjukkan akurasi yang akan dihasilkan oleh model jika hanya melakukan prediksi pada kelas mayoritas (yaitu kelas 0 atau "Non-Pornografi / TIDAK") tanpa memperhatikan fitur atau atribut dari data. Dalam hal ini, *null accuracy score* adalah sebesar 0.6520 atau sekitar 65.20%, yang artinya jika model hanya memprediksi kelas mayoritas (yaitu Non-Pornografi / TIDAK), maka model akan mencapai akurasi sebesar 65.20%.

Hasil ini dapat dijadikan sebagai baseline atau acuan untuk mengevaluasi kinerja model, sehingga model yang dibangun diharapkan mampu menghasilkan akurasi yang lebih baik daripada *null accuracy score*.

Cara mendapatkan nilai *null accuracy score* ini adalah dengan menghitung jumlah sampel pada kelas mayoritas (yaitu kelas 0) dan membaginya dengan jumlah total sampel pada data pengujian.

$$null\ accuracy = \frac{9722}{(9722 + 5189)} = 0.652001878$$

Namun penhitungan *null accuracy* ini tidak memberi tahu apa pun tentang jenis kesalahan atau *type error* yang dibuat oleh model klasifikasi yang dibuat, teknik lainya yang digunakan untuk mencari spesifik kesalahan yang dibuat oleh model klasifikasi disebut dengan *confusion matrix*.

3. *Confusion Matrix*

   Hasil dari *confusion matrix* ini berupa ringkasan prediksi berupa nilai benar dan salah yang dikelompokan berdasarkan masing-

masing kategori label dan disajikan dalam bentuk tabel dengan keterangan dapat dilihat pada pemaparan berikut:

a. *True Positive*s (*TP*)

   Jumlah prediksi yang benar dan sesuai dengan label kelas positif pada data yang diklasifikasikan.

b. *True Negative*s (*TN*)

   Jumlah prediksi yang benar dan sesuai dengan label kelas negatif pada data yang diklasifikasikan.

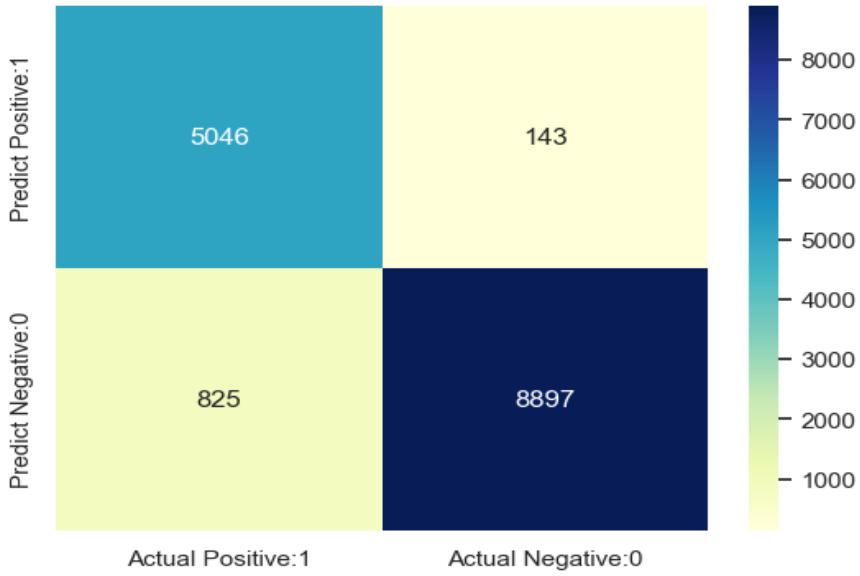
c. *Values Positive* (*FP*)

   Jumlah prediksi yang salah dan tidak sesuai dengan label kelas positif pada data yang diklasifikasikan, *values* positive ini merupakan tipe kesalahan yang disebut *Type I error*.

d. *Values Negative*s (*FN*)

   Jumlah prediksi yang salah dan tidak sesuai dengan label kelas negatif pada data yang diklasifikasikan, merupakan kesalahan yang sangat serius dan disebut dengan *Type II error*.

Berikut adalah hasil *confusion matrix* dari hasil prediksi kelas pada model klasifikasi *Machine Learning* yang dibuat:



Gambar 4.24 *Confusion Matrix*

$True\ Positives\ (TP) = 5046$

$True\ Negatives\ (TN) = 8897$

$False\ Positives\ (FP) = 143$

$False\ Negatives\ (FN) = 825$

Confusion Matrix menunjukan bahwa $5046 + 8897 = 13.943$ prediksi benar dan $143 + 825 = 968$ prediksi salah. Dalam hal ini dapat disimpulkan bahwa hasil dari *confusion matrix* adalah sebagai berikut:

TP (Actual *Positive*: 1 and Predict *Positive*: 1) - 5046

TN (Actual *Negative*: 0 and Predict *Negative*: 0) - 8897

FP (Actual *Negative*: 0 but Predict *Positive*: 1) - 143 (*Type I error*)

FN (Actual *Positive*: 1 but Predict *Negative*: 0) - 825 (*Type II error*)

Setelah nilai *confusion matrix* telah didapatkan berikutnya adalah menghitung hasil dari poin-poin di bawah ini:

- *Accuracy*

  Menghitung akurasi model *MultinomialNB* dapat dilihat pada gambar 4.25 dibawah ini:

```
# print classification accuracy

classification_accuracy = (TP + TN) / float(TP + TN + FP + FN)

print('Classification accuracy : {0:0.4f}'.format(classification_accuracy))
```

```
Classification accuracy : 0.9351
```

Gambar 4.25 Nilai *Accuracy Model MultinomialNB*

Hasil dari akurasi model klasifikasi *MultinomialNB* dapat dihitung dengan persamaan 2.8 yang kemudian memperoleh hasil sebagai berikut:

Di mana:

$$TP = 5046$$
$$TN = 8897$$
$$FP = 143$$
$$FN = 825$$

Hasil perhitungan dari akurasi model *MultinomialNB* adalah:

$$accuracy = \frac{(5046+8897)}{(5046+8897+143+825)} = 0.935081483$$

Classification accuracy merupakan matrik evaluasi yang mengukur seberapa akurat suatu model dalam memprediksi label kelas atau target pada data pengujian. Matrik ini dihitung dengan cara membagi jumlah prediksi yang benar (*TP* dan *TN*) dengan jumlah total sampel pada data pengujian. Dalam hal ini, *classification accuracy* yang dihasilkan oleh model *MultinomialNB* sebesar 0.9351 atau sekitar 93.51%, yang artinya model *MultinomialNB* mampu memprediksi dengan benar sekitar 93.51% dari total sampel pada data pengujian.

Namun, meskipun *classification accuracy* dapat memberikan gambaran yang baik tentang kinerja model secara keseluruhan, namun matrik ini tidak memberikan informasi tentang kinerja model pada setiap kelas. Oleh karena itu, perlu digunakan matrik evaluasi lain seperti *precision*, *recall*, dan *F1-score* untuk mengevaluasi kinerja model pada masing-masing kelas secara lebih rinci dan terperinci.

- *Precision*

  Nilai *Precision* dari model *MultinomialNB* dapat dilihat pada gambar 4.26 berikut:

```
# print precision score

precision = TP / float(TP + FP)

print('Precision : {0:0.4f}'.format(precision))
```

```
Precision : 0.9724
```

Gambar 4.26 Nilai *Precision Model MultinomialNB*

Hasil dari nilai *Precision MultinomialNB* dapat dihitung dengan persamaan 2.9.

Di mana:

$$TP = 5046$$

$$FP = 143$$

Hasil perhitungan dari nilai *Precision MultinomialNB* adalah:

$$Precision = \frac{5046}{(5046+143)} = 0.972441704$$

Precision adalah matrik evaluasi yang mengukur seberapa banyak prediksi positif yang benar dari total prediksi positif. Dalam hal ini, *precision* yang dihasilkan oleh model *MultinomialNB* sebesar 0.9724 atau sekitar 97.24%, yang artinya sekitar 97.24% dari total prediksi positif yang dilakukan oleh model *MultinomialNB* adalah benar.

- *Recall*

  Nilai *Recall* dari model *MultinomialNB* dapat dilihat pada gambar 4.27 berikut:

```
recall = TP / float(TP + FN)

print('Recall or Sensitivity : {0:0.4f}'.format(recall))
```

```
Recall or Sensitivity : 0.8595
```

Gambar 4.27 Nilai *Recall Model MultinomialNB*

Hasil dari nilai *Recall MultinomialNB* dapat dihitung dengan persamaan 2.10.

Di mana:

$$TP = 5046$$

$$FN = 825$$

Hasil perhitungan dari nilai *Recall MultinomialNB* adalah:

$$Recall = \frac{5046}{(5046+825)} = 0.859478794$$

Recall atau *Sensitivity* adalah matrik evaluasi yang mengukur seberapa banyak prediksi positif yang benar dari total data yang sebenarnya positif. Dalam hal ini, *recall* yang dihasilkan oleh model *MultinomialNB* sebesar 0.8595 atau sekitar 85.95%, yang artinya sekitar 85.95% dari total data yang sebenarnya positif dapat diprediksi dengan benar oleh model *MultinomialNB*.

- *F1-Score*

  Nilai *f1-score* dari model *MultinomialNB* dapat dilihat pada gambar 4.28 berikut:

```
f1_score = 2*(precision * recall)/(precision + recall)

print('F1-Score : {0:0.4f}'.format(f1_score))
```

```
F1-Score : 0.9125
```

Gambar 4.28 Nilai *F1-Score Model MultinomialNB*

Hasil dari nilai *f1-score* model *MultinomialNB* dapat dihitung dengan persamaan 4.2 berikut:

$$f1\ score = \frac{2 * (precision\ *\ recall)}{(precision\ +\ recall)} \qquad 4.2$$

Di mana:

$$precision\ =\ 0.972441704$$
$$recall\ =\ 0.859478794$$

Hasil perhitungan dari nilai *F1-score MultinomialNB* adalah:

$$F1 -\ score = \frac{2 * (0.972441704 *\ 0.859478794)}{(0.972441704 + 0.859478794)}$$

$$=\ 0.912477396$$

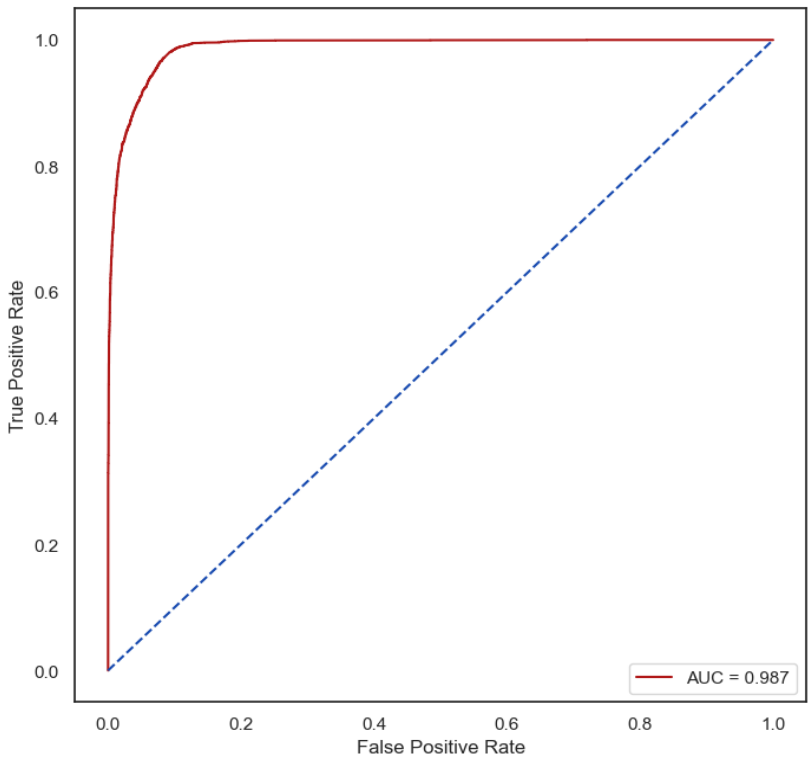
F1-score adalah matrik evaluasi yang menggabungkan *precision* dan *recall* untuk memberikan ukuran yang lebih umum tentang kinerja model. *F1-score* dihitung dengan mengambil nilai harmonic mean dari *precision* dan *recall*. Dalam hal ini, *F1-score* yang dihasilkan oleh model *MultinomialNB* sebesar 0.9125 atau sekitar 91.25%, yang artinya model *MultinomialNB* memiliki keseimbangan yang baik antara *precision* dan *recall*.

4. *Validation Model Classification*

   Validation model classification dilakukan dengan cara berikut:

   a. Kurva *ROC*

      Evaluasi hasil prediksi dapat divisualisasikan dengan kurva *ROC* dengan berfokus pada nilai *TPR* (*True Positive Rate*) dan nilai *FPR* (*Values Positive Rate*) dari satu titik. Gambar 4.29 berikut adalah hasil evaluasi model menggunakan kurva *ROC*:



Gambar 4.29 Kurva *ROC*

Kurva *ROC* adalah salah satu matrik evaluasi model klasifikasi yang digunakan untuk mengukur seberapa baik model dapat membedakan antara kelas positif dan negatif dengan memvariasikan *threshold* atau nilai ambang.

Berikut ini merupakan cara untuk menghitung nilai *True Positive Rate (TPR)* dan *Values Positive Rate (FPR)*:

$$TPR = \frac{TP}{(TP + FN)} \quad 4.3$$

$$FPR = \frac{FP}{(FP + TN)} \quad 4.4$$

Gambar 4.30 berikut merupakan hasil dari perhitungan nilai *TPR*

```
true_positive_rate = TP / float(TP + FN)

print('True Positive Rate : {0:0.4f}'.format(true_positive_rate))
```

```
True Positive Rate : 0.8595
```

Gambar 4.30 Nilai *TPR Model MultinomialNB*

Di mana:

$$TP = 5046$$

$$FN = 825$$

Maka hasil perhitungan dari nilai *TPR* model *MultinomialNB* menggunakan persamaan 4.3 di atas adalah:

$$TPR = \frac{5046}{(5046 + 825)} = 0.859478794$$

Sedangkan untuk nilai *FPR* model *MultinomialNB* dapat dilihat seperti pada gambar 4.31 berikut:

```
false_positive_rate = FP / float(FP + TN)

print('False Positive Rate : {0:0.4f}'.format(false_positive_rate))
```

```
False Positive Rate : 0.0158
```

Gambar 4.31 Nilai *FPR Model MultinomialNB*

Di mana:

$$TN = 8897$$

$$FP = 143$$

Maka hasil perhitungan dari nilai *FPR* model *MultinomialNB* dengan menggunakan persamaan 4.4 di atas adalah:

$$FPR = \frac{143}{(143 + 8897)} = 0.015818584$$

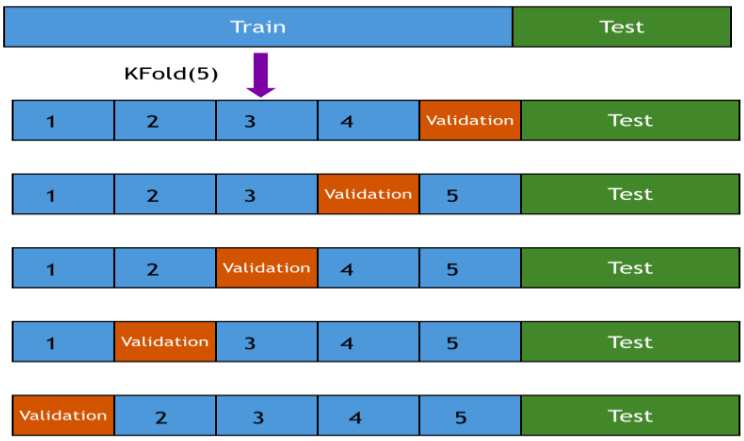
Kurva *ROC* menggambarkan nilai *True Positive Rate* (*TPR*) adalah persentase dari positif sebenarnya yang diklasifikasikan dengan benar sebagai positif, yaitu dokumen pornografi yang diklasifikasikan dengan benar sebagai pornografi / YA. Dalam kasus ini, TPR memiliki nilai 0.8595 atau sekitar 86%. *True Positive Rate (TPR)* didapatkan dengan membagi jumlah True Positive (TP) dengan jumlah True Positive ditambah jumlah *Values* Negative (FN).

Sedangkan nilai *Values Positive Rate* (*FPR*) adalah persentase dari negatif sebenarnya yang salah diklasifikasikan sebagai positif. Dalam kasus ini, *FPR* memiliki nilai 0.0158 atau sekitar 2%. *Values* Positive Rate (FPR) didapatkan dengan membagi jumlah *Values* Positive (FP) dengan jumlah *Values* Positive ditambah jumlah True Negative (TN).

ROC (Receiver Operating Characteristic) Curve adalah suatu grafik yang memperlihatkan performa dari suatu model klasifikasi pada semua threshold. Semakin ke kiri atas (mendekati titik koordinat [0,1]) kurva *ROC*, maka semakin baik performa model. Pada kasus ini, *ROC AUC* memiliki nilai 0.9873, yang dapat dikatakan sangat tinggi sehingga model dapat dikatakan sangat baik dalam melakukan klasifikasi. Berdasarkan pada bab sebelumnya yaitu pada tabel 2.2 dapat disimpulkan bahwa model klasifikasi yang dibuat dengan menggunakan model *MultinomialNB* dapat dikelompokan dalam kategori *"Excellent Classification"*.

b. *Cross Validation*

Ilustrasi 5 *fold* dapat dilihat pada gambar 4.32 berikut:



Gambar 4.32 Ilustarsi 5 *Fold*

5-fold cross-validation adalah salah satu teknik validasi silang yang umum digunakan untuk menguji performa model *machine learning*. Dalam *5-fold cross-validation*, data dibagi menjadi lima bagian yang sama besar. Pada setiap iterasi, salah satu bagian dipilih sebagai data uji dan empat bagian lainnya digunakan sebagai data pelatihan. Proses ini dilakukan sebanyak lima kali, dengan setiap bagian yang berbeda digunakan sebagai data uji pada setiap iterasi.Secara rinci, berikut adalah langkah-langkah untuk melakukan *5-fold cross-validation*:

- Membagi data menjadi lima bagian yang sama besar.
- Pilih salah satu bagian sebagai data uji dan gunakan empat bagian lainnya sebagai data pelatihan.
- Pelajari model pada data pelatihan dan evaluasi pada data uji.
- Ulangi langkah 2 dan 3 sebanyak empat kali, menggunakan bagian yang berbeda sebagai data uji pada setiap iterasi.

- Hitung rata-rata akurasi atau matrik evaluasi lainnya dari lima iterasi untuk mendapatkan evaluasi performa model yang lebih akurat.

Pada dasarnya, *5-fold cross-validation* memungkinkan untuk mengevaluasi model pada seluruh *dataset* dengan menggunakan setiap data sebagai data uji sekali. Hal ini membantu menghindari masalah *overfitting* atau *underfitting*, serta memberikan gambaran yang lebih baik tentang kemampuan model dalam melakukan generalisasi pada data yang belum pernah dilihat sebelumnya.

Hasil *cross validation* dapat dilihat pada tabel 4.17 berikut:

Tabel 4.17 *Cross Validation*

K	Accuracy	Precision	Recall	F1-Score
1	0.945889354	0.923221757	0.972669165	0.947300633
2	0.946103825	0.919726198	0.977513228	0.94773966
3	0.94665491	0.917731959	0.981261023	0.948433838
4	0.947977516	0.926011318	0.973771214	0.949290933
5	0.943899482	0.920634921	0.971567115	0.94541555
Mean				
0.946105017		0.921465231	0.975356349	0.947636123

Hasil *5-fold cross-validation* yang diberikan adalah matrik evaluasi untuk model *machine learning* yang dinilai pada 5 subset data yang berbeda secara acak. Setiap subset data akan dijadikan sebagai data uji secara bergiliran dan model akan

dilatih pada empat subset data lainnya. Kemudian, model akan diuji pada subset data yang dipilih dan hasilnya akan digunakan untuk menghitung matrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score*.

Akurasi (*Accuracy*) adalah matrik evaluasi yang mengukur seberapa sering model memberikan prediksi yang benar dari seluruh prediksi yang dilakukan. Hasil akurasi dari 5-fold cross-validation adalah rata-rata dari akurasi pada setiap subset data, yaitu 0.946.

Presisi (*Precision*) adalah matrik evaluasi yang mengukur seberapa akurat model dalam memprediksi kelas positif. Hasil presisi dari *5-fold cross-validation* adalah rata-rata dari presisi pada setiap subset data, yaitu 0.921.

Recall adalah matrik evaluasi yang mengukur seberapa baik model dalam mengklasifikasikan data kelas positif secara benar. Hasil *recall* dari *5-fold cross-validation* adalah rata-rata dari *recall* pada setiap subset data, yaitu 0.975.

F1-score adalah matrik evaluasi yang menggabungkan presisi dan recall menjadi satu nilai. Hasil *F1-score* dari *5-fold cross-validation* adalah rata-rata dari *F1-score* pada setiap subset data, yaitu 0.948.

Dari hasil *5-fold cross-validation* ini, dapat disimpulkan bahwa model *machine learning* yang dievaluasi memiliki kinerja yang cukup baik dalam mengklasifikasikan data dengan nilai akurasi sebesar 0.946 dan *F1-score* sebesar 0.948. Namun, model perlu ditingkatkan lagi kinerjanya dalam memprediksi kelas positif dengan nilai presisi sebesar 0.921 dan nilai *recall* sebesar 0.975.

4.4 Hasil *Model Evaluation*

4.4.1 Analisa Pendapat Ahli Bahasa Indonesia

Setelah mengetahui performa dari model klasifikasi yang telah dibuat langkah selanjutnya adalah membandingkan dan mengevaluasi performa model berdasarkan pendapat ahli Bahasa Indonesia. Sampel data yang digunakan untuk mengevaluasi model dapat dilihat pada lampiran yaitu "Tabel Hasil Evaluasi Berdasarkan Pendapat Ahli Bahasa Indonesia", hasil dari analisanya dapat dilihat pada tabel 4.18 berikut:

Tabel 4.18 Hasil Analisa Pendapat Ahli Bahasa Indonesia

ID	Analisa	Hasil Klasifikasi
001	*Link*, judul dan isi cerita mengandung unsur pornografi dan tidak layak dibaca atau dikonsumsi oleh usia 18 tahun kebawah (pornografi).	Pornografi
002	*Link*, judul dan isi cerita mengandung unsur pornografi.	Pornografi
003	Cerita tersebut atau teks 3 (003) hanya boleh dibaca oleh orang dewasa yang sudah sangat matang dalam pemikiran (mengandung unsur pornografi)	Pornogarfi
004	*Link*, judul dan isi cerita mengandung unsur pornografi.	pornografi
005	*Link*, judul dan isi cerita mengandung unsur pornografi.	Pornografi

ID	Analisa	Hasil Klasifikasi
006	Isi cerita berisi pengetahuan dan dapat menumbuhkan jiwa nasionalisme dan cinta terhadap tanah air (non-pornografi).	Non-Pornografi
007	Teks tersebut adalah teks berita dan dikategorikan berita politik (non-pornografi).	Non-Pornografi
008	*Link*, judul dan isi bersifat pemberitahuan atau informasi (non-pornografi).	Non-Pornografi
009	Isi cerita atau teks tersebut (009) berisi pengetahuan atau edukasi (non-pornografi).	Non-Pornografi
010	Isi cerita berisi pengetahuan dan informasi yang bertujuan untuk mengajak seseorang atau orang lain (non-pornografi).	Non-Pornografi

Berdasarkan hasil analisa dari ahli bahasa pada tabel 4.18 di atas dapat disimpulkan bahwa ahli bahasa mengklasifikasikan lima dokumen ke dalam kategori pornografi dan lima dokumen lainnya ke dalam kategori non-pornografi, lima dokumen pertama dengan ID 001, 002, 003, 004 dan 005 diklasifikasikan sebagai *website* atau *blog* yang mengandung unsur pornografi di dalam *website* tersebut sedangkan untuk lima dokumen berikutnya dengan ID 006, 007,008, 009 dan 010 diklasifikasikan sebagai *website* atau *blog* yang tidak mengandung unsur pornografi. Berdasarkan penjelasan di atas maka hasil analisa dari klasifikasi yang telah dilakukan oleh ahli bahasa maka akan diuji dengan menggunakan sistem klasifikasi yang telah dibuat menggunakan model *MultinomialNB.*

4.4.2 Perbandingan Hasil Pendapat Ahli Bahasa Dengan Model Klasifikasi *Multinomial*

Setelah diperolehnya hasil analisa data dari konten *website* berdasarkan pendapat dari ahli Bahasa Indonesia, maka dibuatnya sistem untuk melakukan klasifikasi dengan tujuan untuk membandingkan apakah model yang dibuat mampu mengklasifikasikan teks ke dalam kelas pornografi atau kelas non-pornografi.

Sistem klasifikasi teks bekerja dengan memproses dan mengelompokkan teks ke dalam kategori atau label berdasarkan informasi yang terkandung dalam teks tersebut. Sistem ini memanfaatkan algoritma *Machine Learning* dan *Natural Language Processing (NLP)* untuk mengidentifikasi pola dan karakteristik pada teks, dan kemudian menggunakan informasi tersebut untuk menentukan kategori atau label yang tepat.

Tujuan dari pembuatan sistem klasifikasi teks ini adalah untuk menguji kinerja dari model klasifikasi yang telah dibuat apakah mampu mengklasfikasikan konten *website* yang belum pernah dilihat sebelumnya, tabel 4.19 berikut adalah hasil pengujian dari sistem klasifikasi yang telah dibuat:

Tabel 4.19 Hasil Perbandingan Analisa Ahli Bahasa Dengan Sistem

ID	Klasifikasi Berdasarkan Ahli Bahasa	Klasifikasi Berdasarkan Sistem Model *MultinomialNB*
001	Pornografi	Pornografi
002	Pornografi	Pornografi
003	Pornografi	Pornografi
004	Pornografi	Pornografi

ID	Klasifikasi Berdasarkan Ahli Bahasa	Klasifikasi Berdasarkan Sistem Model *MultinomialNB*
005	Pornografi	Pornografi
006	Non-Pornografi	Non-Pornografi
007	Non-Pornografi	Non-Pornografi
008	Non-Pornografi	Non-Pornografi
009	Non-Pornografi	Non-Pornografi
010	Non-Pornografi	Non-Pornografi

Dalam konteks analisis klasifikasi teks, sistem klasifikasi yang efektif dan memiliki tingkat akurasi yang baik mengindikasikan bahwa model klasifikasi mampu mengenali pola dalam data dan dapat memprediksi label kelas dengan akurat.

Dapat dilihat pada tabel 4.19 di atas dapat dijelaskan bahwa berdasarkan analisa ahli bahasa ID 001, 002, 003, 004 dan 005 diklasifikasikan sebagai konten *website* yang mengandung unsur pornografi yang kemudian diuji menggunakan sistem klasifikasi model *MultinomialNB* yang telah dibuat didapatkannya hasil bahwa hasil klasifikasi yang dilakukan oleh sistem mengklasifikasikan hal yang sama dimana ID 001, 002, 003, 004 dan 005 juga diklasifikasikan sebagai konten *website* yang mengandung unsur pornografi.

Sama halnya dengan ID 006, 007, 008, 009 dan 010 baik dari hasil analisa ahli bahasa dan klasifikasi berdasarkan sistem klasifikasi yang dibuat mengklasifikasikan dokumen sebagai konten *website* yang tidak mengandung unsur pornografi. Oleh karena itu dapat disimpulkan bahwa model klasifikasi menggunakan algoritma *Multinomial Naïve Bayes* mampu mengklasifikasikan konten *website* ke dalam kategori yang mengandung unsur pornografi dan yang tidak mengandung unsur pornografi secara akurat.

4.5 Operation

Seperti yang telah dijelaskan pada sub bab sebelumnya bahwa pada tahap *operation* dalam *machine learning life cycle,* model *machine learning* yang telah dikembangkan dan diuji diimplementasikan ke dalam produksi dan mulai digunakan untuk memproses data secara *real-time*. Tahap ini melibatkan implementasi model, *monitoring* kinerja, perawatan model, dan skalabilitas sistem. Tujuan dari tahap operasi adalah memastikan model *machine learning* yang beroperasi dalam lingkungan produksi selalu dapat berfungsi dengan baik dan menghasilkan hasil yang akurat.

Namun dalam penelitian ini dimana peneliti hanya berfokus pada pembuatan model klasifikasinya saja maka untuk tahap operation atau pengembangan ini dapat dijelaskan sebagai berikut:

4.5.1 *Environment*

Environment dari model klasifikasi *machine learning* yang dibuat dalam penelitian ini yaitu dengan menggunakan *Anconda*, dimana *Anaconda* adalah lingkungan *open-source* yang mencakup banyak pustaka (*library*) untuk *data science* dan *machine learning*, yang dapat membantu dalam mengelola proyek *machine learning* secara efisien. *Anaconda* memiliki fitur standar seperti manajemen lingkungan *virtual*, pengaturan paket, dan instalasi *software* dengan mudah.

Sementara itu, *Jupyter Notebook* adalah aplikasi lingkungan pengembangan yang digunakan untuk membuat dan berbagi kode dalam bentuk *notebook*, yang membantu peneliti untuk mempertahankan dan mengeksekusi kode secara interaktif, memvisualisasikan data, dan menjelaskan hasil. *Jupyter Notebook* bekerja pada lingkungan *anaconda*, sehingga sangat memudahkan untuk membuat dan mengelola lingkungan *virtual* untuk pembuatan model klasifikasi *machine learning*.

Dengan demikian, *environment* yang digunakan mencangkup lingkungan *virtual* yang telah dibangun di dalam *Anaconda* serta

lingkungan pengembangan interaktif menggunakan *Jupyter Notebook*. *Environment* atau lingkungan ini memungkinkan peneliti dalam melakukan berbagai eksperimen *machine learning* pada suatu lingkungan yang terisolasi dan bersih, serta mempermudah proses mempertahankan dan mengeksplorasi hasil eksperimen dengan menggunakan *Jupyter Notebook*. Dengan tujuan untuk mengembangkan model machine learning dengan menggunakan metode *Multinomial Naïve Bayes.*

4.5.2 *Deployment*

Tahap *deployment* dalam penelitian ini dilakukan dengan cara membuat sistem klasifikasi berdasarkan ekstraksi fitur dan model yang telah dibuat. Berdasarkan cara yang telah diimplementasikan dalam penelitian ini untuk menyediakan hasil klasifikasi model *Machine Learning* dapat dianggap sebagai *deployment*. Walaupun pada dasarnya *deployment* biasanya terkait dengan mengintegrasikan model *Machine Learning* ke dalam *software* atau aplikasi, namun dalam penelitian ini, peneliti sudah menyediakan fitur dan model terpisah dari aplikasi atau *software*.

Dengan menyediakan *output* dalam format yang dapat digunakan langsung untuk mengklasifikasikan dokumen teks dari *website* atau *blog*, walaupun untuk menggunakan sistem klasifikasi tersebut membutuhkan *software* atau perangkat lunak yang mendukung bahasa pemrograman *python*, peneliti memungkinkan pengguna untuk mengklasifikasikan dokumen dengan mudah tanpa harus mengembangkan aplikasi khusus untuk itu. Tentu saja, cara ini tersedia untuk digunakan dan dapat disebarkan untuk pengguna yang membutuhkan untuk mengklasifikasikan konten teks dari *website* atau *blog*.

Maka, secara keseluruhan, berdasarkan apa yang telah dilakukan oleh peneliti dapat dianggap sebagai *deployment*. Peneliti menyediakan fitur dan model yang dihasilkan oleh proyek *machine learning* menggunakan algoritma *Naïve Bayes* dengan metode *Multinomial Naïve Bayes*, sehingga memungkinkan pengguna untuk mengklasifikasikan dokumen dengan

cepat dan mudah tanpa harus mengembangkan aplikasi atau *website* yang rumit.

Berikut adalah sistem klasifikasi yang telah dibuat dengan menggunakan model *Multinomial Naïve Bayes* dengan menyimpan ektraksi fitur, model algoritma *multinomial naïve bayes* dan membuat sistem klasifikasi:

```
joblib.dump(tfidf, 'tfidf.joblib')
```

```
['tfidf.joblib']
```

Gambar 4.33 Simpan Ekstraksi Fitur

Berdasarkan *script* seperti pada gambar 4.33 di atas dapat dijelaskan sebagai berikut. Dimana, joblib.dump(tfidf, 'tfidf.joblib') digunakan untuk menyimpan objek *tfidf* ke dalam *file* tfidf.joblib menggunakan *library joblib*.

'*tfidf'* adalah fitur yang telah dibuat sebelumnya dengan menggunakan *library scikit-learn* untuk menghitung skor *TFIDF (Term Frequency-Inverse Document Frequency)* dari suatu *dataset* teks. Skor *TFIDF* digunakan untuk mengevaluasi seberapa penting sebuah kata dalam dokumen, dengan mempertimbangkan frekuensi kemunculan kata tersebut di dokumen dan frekuensi kemunculan kata tersebut di seluruh dokumen dalam *dataset*.

Dalam konteks ini, joblib.dump() digunakan untuk menyimpan objek *tfidf* agar dapat digunakan kembali di kemudian hari tanpa perlu menghitung ulang skor *TFIDF*-nya dari awal. Dengan menggunakan joblib, objek *tfidf* dapat disimpan secara efisien di dalam *file* tfidf.joblib dan diambil kembali menggunakan joblib.load() ketika diperlukan.

```
# Menyimpan model
joblib.dump(classifier, 'model_mnb.pkl')
```

```
['model_mnb.pkl']
```

Gambar 4.34 Simpan Model MultinomialNB

Berdasarkan *script* seperti pada gambar 4.34 di atas dapat dijelaskan sebagai berikut. Dimana, joblib.dump(classifier, 'model_mnb.pkl') digunakan untuk menyimpan model klasifikasi yang telah dilatih menggunakan algoritma *Multinomial Naïve Bayes* ke dalam *file* model_mnb.pkl menggunakan *library joblib*.

'classifier' adalah objek model klasifikasi yang telah dilatih sebelumnya dengan menggunakan *library scikit-learn* untuk melakukan klasifikasi teks pada *dataset* tertentu. Algoritma *Multinomial Naïve Bayes* adalah salah satu metode klasifikasi yang umum digunakan dalam pemrosesan teks, khususnya untuk tugas klasifikasi dokumen.

Dalam konteks ini, joblib.dump() digunakan untuk menyimpan objek model klasifikasi classifier agar dapat digunakan kembali di kemudian hari tanpa perlu melatih model klasifikasi dari awal. Dengan menggunakan *joblib*, model klasifikasi dapat disimpan secara efisien di dalam file model_mnb.pkl dan diambil kembali menggunakan joblib.load() ketika diperlukan. Hal ini sangat berguna dalam aplikasi praktis di mana model klasifikasi perlu disimpan dan diambil kembali untuk digunakan dalam konteks tertentu.

```python
# Fungsi untuk melakukan prediksi label teks
def get_class(text):
    # Mengubah teks menjadi vektor fitur
    text_vector = tfidf.transform(text).toarray()

    # Melakukan prediksi menggunakan model klasifikasi
    classification = classifier.predict(text_vector)

    # Menampilkan hasil prediksi
    if classification [0] > 0.5:
        print("Klasifikasi: Pornografi")
    else:
        print("Klasifikasi: Non-Pornografi")
```

Gambar 4.35 Sistem Klasifikasi

Berdasarkan *script* seperti pada gambar 4.35 di atas mendefinisikan sebuah fungsi bernama get_class yang digunakan untuk melakukan prediksi label (klasifikasi) teks apakah termasuk kategori pornografi atau tidak. Berikut adalah penjelasan mengenai bagaimana fungsi tersebut bekerja:

1. Fungsi ‘get_class’ menerima sebuah argumen text yang berisi teks yang akan diklasifikasikan.
2. Fungsi pertama-tama mengubah teks tersebut menjadi vektor fitur menggunakan tfidf.transform(text).toarray(). ‘tfidf’ di sini adalah objek *TFIDF* yang telah dibuat sebelumnya dan disimpan ke dalam file menggunakan *joblib*.
3. Setelah vektor fitur teks didapatkan, fungsi melakukan prediksi klasifikasi menggunakan model klasifikasi yang telah dilatih sebelumnya dan disimpan ke dalam file menggunakan *joblib*. Prediksi klasifikasi dilakukan dengan memanggil classifier.predict(text_vector).
4. Hasil prediksi klasifikasi kemudian diperiksa apakah nilai prediksi lebih besar dari 0,5 atau tidak. Jika nilai prediksi lebih besar dari 0,5, fungsi akan mencetak "Klasifikasi: Pornografi". Jika tidak, fungsi akan mencetak "Klasifikasi: Non-Pornografi".

5. Fungsi ‘get_class’ mengembalikan output berupa teks hasil prediksi klasifikasi. Namun, karena dalam *script* di atas tidak menggunakan return, maka output dari fungsi ini hanya berupa hasil cetakan di console saja.

```
# Teks website yang akan diklasifikasikan
text = input("Input Text Here: ")

# Menampilkan teks input
print(text)

# Melakukan prediksi label teks
get_class([text])
```

Gambar 4.36 Menampilkan Output Klasifikasi

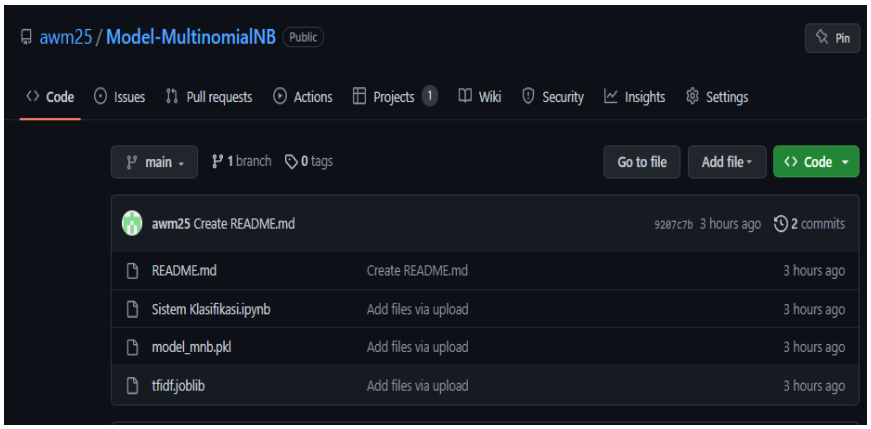
Script di atas melakukan hal berikut:

1. Memasukkan teks melalui fungsi input() dan menyimpannya ke dalam variabel text.
2. Menampilkan teks yang dimasukkan oleh pengguna ke dalam console menggunakan print(text).
3. Memanggil fungsi get_class() dengan argumen berupa teks yang dimasukkan oleh pengguna yang telah dikemas dalam list ([text]). Fungsi ini akan melakukan prediksi label teks apakah termasuk kategori pornografi atau tidak. Hasil prediksi kemudian akan ditampilkan ke dalam console menggunakan perintah print() di dalam fungsi get_class().

Dengan cara ini, *script* di atas dapat digunakan untuk memasukkan teks apa pun dan melakukan prediksi klasifikasi teks tersebut apakah termasuk kategori pornografi atau tidak.

Berdasarkan penjelasan di atas maka peneliti telah menyiapkan model klasifikasi *Multinomial Naïve Bayes*, ekstraksi fitur serta *script* di atas sebagai sistem klasifikasi yang telah di upload pada github dengan link

sebagai berikut https://github.com/awm25/Model-MultinomialNB.git di dalam link tersebut terdapat tiga file seperti pada gambar berikut:



Gambar 4.37 File Link Github

Sehingga apabila pengguna ingin menggunakan model yang telah dibuat tersebut dapat diambil atau diunduh melalui github dan dijalankan menggunakan *tools*/*software* yang mendukung bahasa pemrograman *python*.

Namun ada beberapa hal yang perlu diperhatikan apabila pengguna ingin menggunakan sistem klasifikasi tersebut agar dapat memperhatikan lingkungan atau *environment* yang digunakan, hal ini dikarenakan dalam pengembangan model klasifikasi yang dibuat tersebut masih dalam ruang lingkup *internal* sehingga untuk menggunakan sistem klasifikasi tersebut disarankan dengan menggunakan lingkungan atau *environment* dari *software* sebagai berikut:

- *Anaconda versi 3.1.0*
- *Python versi 3.9.13*
- *Jupyter Notebook versi 6.4.12*
- *Scikit-Learn versi 1.2.0*

Hal yang mendasari agar menghindari menghindari penggunaa *environment* yang berbeda ini adalah lingkungan *internal* yang berbeda dapat memiliki konfigurasi sistem operasi, bahasa pemrograman, *library*,

atau *dependency* yang berbeda yang dapat mempengaruhi cara ekstraksi fitur dan model klasifikasi dijalankan dan jika versi perangkat lunak yang digunakan pada lingkungan *internal* yang berbeda berbeda dengan versi yang digunakan saat model dan ekstraksi fitur disimpan, maka mungkin ada perbedaan dalam cara model dijalankan.

Selanjutnya untuk mengoperasikan sistem klasifikasi yang telah dibuat tersebut dapat dilakukan dengan cara berikut:

a. Pastikan telah menginstal semua *library* dan *package* yang dibutuhkan, yaitu *Scikit-Learn* dan *joblib*.
b. Unduh dan simpan *file* "sistem klasifikasi.ipynb", "model_mnb.pkl", dan "tfidf.joblib" pada direktori yang sama.
c. Jalankan aplikasi *Jupyter Notebook* pada *Anaconda Navigator*
d. Pada halaman *Jupyter Notebook*, navigasikan ke direktori yang berisi file yang sudah di unduh tersebut dan buka *file* "sistem klasifikasi.ipynb".
e. Setelah file terbuka, kemudian mengeksekusi seluruh kode yang terdapat didalamnya dengan menekan tombol *"Run"* atau dengan menekan tombol *"Shift + Enter"* pada setiap sel.
f. Pastikan bahwa *file* "model_mnb.pkl" dan "tfidf.joblib" terletak pada direktori yang sama dengan *file* "sistem klasifikasi.ipynb". Jika *file* tidak ditemukan, pastikan untuk menyesuaikan path-nya di dalam kode.

4.5.3 *Monitoring* atau *updating*

Dalam penelitian ini, *monitoring* dan *updating* pada model Machine Learning yang telah dibuat dapat dilakukan dengan memantau kinerja model secara teratur dan melakukan pembaruan terkait faktor-faktor yang mempengaruhi hasil klasifikasi.

Untuk memantau kinerja model, dapat dilakukan dengan menggunakan beberapa matrik evaluasi, seperti akurasi, presisi, *recall*, dan *F1-score*. Pemantauan dilakukan secara teratur untuk

memastikan bahwa model tidak mengalami *overfitting* atau *underfitting*, dan dapat menghasilkan prediksi yang akurat pada data baru.

Sementara itu, pembaruan dapat dilakukan dengan mengumpulkan data baru yang mencerminkan populasi aktual dan mengembangkan model baru yang berdasarkan pada data tersebut. Dalam penelitian ini, untuk kedepannya dapat memutakhirkan model yang sudah ada dengan data baru yang dihasilkan. Setelah model diperbaharui, maka dapat menguji model pada data baru ini, dan melihat apakah hasil klasifikasinya sudah membaik.

Terlepas dari itu, jika model tidak memerlukan pembaruan berdasarkan pada faktor-faktor yang mempengaruhi hasil klasifikasi, maka model dapat terus digunakan tanpa perlu melakukan *updating*. Namun, pemantauan secara konstan, baik pada performa model maupun perubahan kondisi eksternal, tetap penting untuk memastikan bahwa hasil klasifikasi yang dihasilkan tetap sesuai dengan tujuan awal dari proyek *Machine Learning*.

4.6 Kelemahan Sistem

Berikut adalah beberapa kelemahan sistem yang dibuat:

1. *Dataset* yang tidak seimbang dapat menjadi kelemahan dalam pengembangan sistem. Meskipun kelemahan ini tidak mempengaruhi kinerja model secara signifikan dalam penelitian ini, namun dapat menghasilkan prediksi yang tidak akurat dan bias. Yang dimaksud dengan ketidakseimbangan *dataset* adalah dalam penelitian ini terdapat dua label kelas yaitu kelas pornografi (YA) dan kelas non-pornografi (TIDAK) dimana kelas pornografi memiliki ukuran atau jumlah dataset sebanyak 17.297 ribu data label kelas atau 35% dan untuk label kelas non-pornografi sebanyak 32.405 ribu data label kelas atau 65%. Oleh karena itu, penelitian selanjutnya dapat mempertimbangkan teknik *sampling* atau

weighting pada dataset untuk memastikan keseimbangan antara kelas positif dan negatif.

2. Penggunaan nilai parameter *alpha 0.1* pada model *MultinomialNB* yang tidak seimbang pada bagian *validation curve* dapat mengakibatkan *overfitting* atau *underfitting* pada model seperti pada gambar 4.21 *Validation Curve*.
3. Nilai *recall* kelas positif yang hanya sebesar 85.95% dapat menjadi kelemahan dalam pengembangan sistem. Hal ini menunjukkan bahwa model cenderung memprediksi *false negative* pada kelas positif, sehingga dapat menghasilkan prediksi yang tidak akurat. Hal ini disebabkan oleh ketidakseimbangan label kelas pada dataset seperti yang disebutkan pada poin sebelumnya.
4. Dalam mengklasifikasikan dokumen atau konten *website* dan *blog* dilakukan dengan cara konvensional karena dalam penelitian ini hanya berfokus pada pembuatan model klasifikasi dan tidak menerapkan atau mengimplementasikannya ke dalam bentuk aplikasi atau *website*.