

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terkait

Penyusunan skripsi ini didasari dari beberapa referensi penelitian yang dulu sudah pernah ada kemudian penulis jadikan inspirasi yang hasilnya berhubungan tentang analisis sentimen. Berikut ini ialah penelitian yang diperoleh penulis sebagai bentuk referensi :

Mathoril Hudha (2022) Berjudul “ANALISIS SENTIMEN PENGGUNA YOUTUBE TERHADAP TAYANGAN #MATANAJWAMENANTITERAWAN DENGAN METODE NAÏVE BAYES CLASSIFIER”. Penelitian ini mengumpulkan komentar dari video youtube yang menjadi fokus penelitian, yaitu wawancara oleh Najwa Shihab di acara Mata Najwa dengan kursi kosong yang mewakili Menteri Kesehatan Terawan Agus Putranto. Hasil penelitian menunjukkan bahwa sentimen netral mendominasi, dengan 1232 komentar netral, 90 komentar negatif, dan 78 komentar positif.

Penelitian ini juga menjelaskan metodologi penelitian yang digunakan, termasuk pengumpulan dan analisis data. Data dikumpulkan dengan mengambil komentar dari video youtube menggunakan *library* Selenium dalam bahasa pemrograman Python. Data kemudian diproses melalui tahap *preprocessing* teks, termasuk pembersihan, penurunan huruf besar, tokenisasi, penghapusan kata penghubung, dan *stemming*. Algoritma *Naive Bayes Classifier* digunakan untuk klasifikasi teks, yang menunjukkan hasil yang baik dalam hal akurasi. Penelitian ini

berhasil mengumpulkan 1400 komentar dari video youtube dan mencapai akurasi 90,36% dalam analisis sentimen.

Dalam penelitian ini, juga disebutkan beberapa sumber yang dikutip, termasuk penelitian dan artikel jurnal yang terkait dengan topik-topik seperti perbandingan sistem pers di Indonesia, bias algoritma dalam rekomendasi youtube, analisis sentimen di media sosial, dan aplikasi algoritma *Naive Bayes* dalam berbagai *domain*, juga menyebutkan penggunaan teknik *oversampling* dan *undersampling* untuk mengatasi ketidakseimbangan dataset dalam *data mining*.

Rizki Wahyudi & Gilang Kusumawardhana (2021) yang berjudul “ANALISIS SENTIMEN PADA REVIEW APLIKASI GRAB DI GOOGLE PLAY STORE MENGGUNAKAN SUPPORT VECTOR MACHINE”. Suatu aplikasi biasanya menggunakan review aplikasi sebagai bentuk suatu reputasi dari aplikasi tersebut, khususnya pada sebuah Google Play Store misalnya. Aplikasi Grab merupakan salah satu aplikasi yang sekarang ini banyak digunakan dan disorot karena kegunaannya dalam mengatasi masalah pengguna yaitu gojek yang dapat dipesan melalui aplikasi. Pada penelitian ini akan dianalisis tentang pengguna yang sering mereview sebuah aplikasi Grab pada Google Play Store.

Analisis sentimen untuk penelitian ini menggunakan metode *Support Vector Machine* (SVM), penggunaan metode ini berguna nantinya agar hasil pengklasifikasian teks *review* menjadi lebih baik juga akurat sehingga dapat dijadikan acuan untuk memperbaiki kelemahan aplikasi Grab tersebut. Tahap penelitiannya dimulai dari pengumpulan data, tahap *preprocessing*, tahap pembagian data, hasil dan pembahasan, perbandingan validasi, dan tahap

klasifikasi. Dalam evaluasi akan ada 1.000 hasil review pengguna yang dipakai pada penelitian ini. Hasil dari metode *Support Vector Machine* pada penelitian ini adalah hasil akurasi sebesar 85,54%, menghasilkan prediksi hasil dari *review* positif sebesar 59 *review* tentang apa yang sering diulas mengenai “ovo”, dan bagian dari *review* negatif dengan 675 *review*, yang sering muncul kata seperti “saldo_ovo” dan “tolong_perbaiki”, dari kata tersebut banyak pengguna sering mengeluhkan masalah saldo ovo.

Giovani & Ardiansyah (2020) yang berjudul “ANALISIS SENTIMEN APLIKASI RUANG GURU DI TWITTER MENGGUNAKAN ALGORITMA KLASIFIKASI”. Saat ini telah banyak pembelajaran yang dilakukan secara *online* dengan sistem komputer yang mampu memberikan kemudahan bagi para pelajar untuk mendapatkan akses pendidikan yang sama. Akses pendidikan yang berbasis elektronik ini biasa disebut *e-learning*, salah satu aplikasi yang kini marak dan telah banyak digunakan para pelajar dalam membantu sistem pembelajaran adalah aplikasi Ruang Guru.

Pada penelitian ini akan menghasilkan suatu analisis sentimen yang mengarah pada aplikasi ini berjalan dengan tepat atau tidak melalui pengambilan komentar pengguna di media sosial twitter, dengan data yang telah dilakukan *data cleaning*, diraih sentimen positif sebanyak 338 *tweet* dan juga sebanyak 175 *tweet* mengenai sentimen negatifnya. Metode yang digunakan adalah algoritma *Naïve Bayes*(NB), *Support Vector Machine* (SVM), *K-Nearest Neighbour* (K-NN), dan *feature selection*. Hasil nilai akurasi pengujian 78,55% dan untuk *Area Under Curve* 0,853. Disimpulkan bahwa pada hasil pengujian menggunakan SVM (PSO) jika

dibandingkan dengan NB, SVM, K-NN, NB (PSO), SVM (PSO), dan K-NN (PSO) akan memperlihatkannya lebih unggul nilai akurasi dan kemampuannya.

Duei Putri, Gigih Forda, & Wahyu Eko (2022) berjudul “ANALISIS SENTIMEN KINERJA DEWAN PERWAKILAN RAKYAT (DPR) PADA TWITTER MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER”. Dewan Perwakilan Rakyat terkait kinerjanya yang selalu menjadi topik utama Masyarakat Indonesia. DPR sebenarnya sebagai lembaga negara dalam lingkungannya memiliki kekuasaan pembentuk Undang-Undang yang memiliki tanggung jawab untuk mewakilkan rakyat. Tanggung jawab yang diemban DPR kini memberikan reaksi bahwa kinerja mereka tidak nyata kepada rakyat. Opini yang ada telah banyak tersebar dalam media sosial yang rakyat gunakan untuk layanan komunikasi, layanan tersebut dipilih berdasarkan kepopulerannya salah satunya adalah Twitter. Pada penelitian ini python digunakan sebagai Bahasa pemrograman yang digunakan, terdapat *library* dan *framework* pada penelitian ini, yaitu Tweepy, NLTK, Sastawi, *Scikit-learn*. Algoritma yang digunakan pada penelitian ini adalah *naïve bayes* yang merupakan prediksi kejadian di masa depan yang didasarkan masa sebelumnya. Data yang digunakan sebanyak 1546 *tweet* yang menghasilkan 95 *tweet* positif dengan tambahan popularitas 75%, lalu *tweet* netral didapat 79%, dan *tweet* negatif 758 popularitasnya sebesar 82%. Sesuai penelitian ini hasil analisis dapat digunakan untuk memprediksi sebuah dataset baru tanpa harus melakukan *labelling* terlebih dahulu.

Mas Rizqi & Wiyli (2021) Dengan judul penelitiannya “ANALISIS SENTIMEN CUSTOMER REVIEW APLIKASI RUANG GURU DENGAN METODE BERT (BIDIRECTIONAL ENCODER REPRESENTATIONS FROM

TRANSFORMERS), pada penelitian tersebut dipilihnya google play store sebagai media untuk pengambilan Analisa sentimen pada aplikasi ruang guru, penulis menggunakan *review* aplikasi ruang guru dikarenakan sekarang ini masa dimana pembelajaran konvensional akan digantikan dengan pembelajaran daring sehingga penelitian ini memiliki kesesuaian dengan apa yang terjadi pada masa penelitiannya. Data yang berhasil dilakukan *scraping* berjumlah 5437 *records*. Pengambilan data berdasarkan hasil *review* pada google play store untuk aplikasi ruang guru yang mendapatkan banyak komentar keluhan, mengenai berbayarnya video, dan mengenai aplikasi yang mengalami *crash*.

Pada penelitian akan menggunakan 2 acuan, yaitu secara segi *Customer review* oleh Alifia dan pada algoritmanya ialah BERT oleh Saad Abdul. Melalui hasil model *pre-trained* BERT telah diperoleh nilai *f1 score*, yaitu 98,9% lalu proporsi bagian data latih juga data uji 70:30. Pada nilai akurasi yang didapat 99%, dengan presisinya adalah 64,13% lalu *recall* yang didapat 60,51%. Sehingga hasilnya nilai kriteria presentase sentimen berbobot 99% dan dinyatakan sebagai nilai valid berdasarkan 10 kali *epoch* dengan rata-rata peningkatan konsisten.

2.2 Landasan Teori

Pada bagian landasan teori pembahasan lebih difokuskan mengenai *text mining*, *data mining*, analisis sentimen, pembagian sentimen, guru di Indonesia, youtube sebagai media, konten yang disajikan dalam youtube, kanal Youtube Satu Persen, kanal Youtube Berita Satu, Instant Data Scraper, Python, Google Colab, Microsoft Excel, Metode *Naïve Bayes*, *Confusion Matriks*, dan *Word cloud*

2.2.1 *Text Mining*

Text mining adalah suatu bentuk proses ekstraksi berdasarkan pola dari data yang tidak terstruktur yang akan diambil pola-pola data juga tren dan mengenai ekstraksi pengetahuan yang berpotensi berdasarkan data teks. Informasinya berupa kecenderungan dalam bentuk pola statistik. Menjadikan masukan untuk penambahan teks merupakan data yang tidak terstruktur, misalnya dokumen, PDF, kutipan teks yang tak terstruktur, kemudian untuk masukan penambahan data ialah data yang termasuk terstruktur (Bijaksana, 2020).

2.2.2 *Data Mining*

Data mining adalah suatu proses dari ekstraksi informasi juga pola yang dapat berguna dalam *Big Data*. Pada *data mining* dapat diketahui sebagai proses *knowledge discovery*, sebuah penambahan pengetahuan yang didapat dari data, dan bentuk ekstraksi pengetahuan data. *Data mining* juga merupakan proses untuk menemukan suatu korelasi baru yang bermanfaat berdasarkan polanya dengan sejumlah repositori data yang jumlahnya besar, serta menggunakan teknologinya untuk pengenalan pola, seperti teknik matematika. *Data mining* pada tahun 1996 tujuannya digunakan dalam proses menganalisis suatu industri dengan strateginya untuk memecahkan masalah dari segi bisnis (Mas Rizqi & Wiyli, 2021).

Data mining didefinisikan sebagai proses yang menggunakan satu atau lebih cara pembelajaran komputer untuk melakukan analisis dan ekstraksi pengetahuan secara otomatis. Biasanya pola-pola dalam *data mining* bisa diketahui oleh suatu

perangkat tertentu yang nantinya memberikan Analisa data yang bermanfaat dan dapat dipelajari dengan perangkat pendukung keputusan lainnya (Darwis, 2022).

2.2.3 Analisis Sentimen

Dimasa sekarang ini perkembangan dari analisis telah beragam bentuknya dikarenakan kemajuan zaman yang sangatlah berperan besar bagi cara kita dalam melakukan sebuah penelitian. Analisis sentimen merupakan pengubah data yang bentuknya berupa teks menjadi informasi yang nantinya dapat bernilai positif, netral, dan negatif. Para pengguna internet terbiasa untuk menggunakan media sosial sebagai sarana untuk memberikan suatu penilaian pribadinya yang dapat diartikan, sebuah analisis sentimen. Opini publik yang mengarah pada perasaan yang mereka utarakan inilah yang menjadikannya sentimen (Arham et al., 2022).

Dalam hal ini teks dimaksudkan lisan ataupun tulisan yang memiliki makna dan bersifat umum memiliki keterikatan dengan dunia nyata, seperti hamparan Bahasa. Sedangkan untuk informasi sendiri ialah yang berbentuk teks baik didasarkan pada literatur, buku, maupun dari internet. Dapat dikatakan bahwa pengambilan suatu informasi dalam teks disebut sebagai *text mining* pada bidang analisis sentimen, bidang analisis sentimen merupakan penggambaran berupa pendapat, opini yang dipaparkan menjadi teks. Bentuk penganalisisan yang dibagi memudahkan kita untuk menilai permasalahan tersebut sesuai dengan kategorinya, yaitu positif, netral, dan negatif (Samrin & Akbar, 2023).

Penelitian tentang analisis sentimen semakin banyak bermunculan seiring dengan berkembangnya teknologi komunikasi. Kemunculan yang banyak itulah

membuat opini yang tersebar di media sosial menjadi media yang efektif untuk menggiring sentimen dan menyebarkan opini yang bisa saja dimanfaatkan oleh pihak-pihak tertentu. Kegunaan dari Analisis Sentimen adalah untuk mendapatkan informasi yang bermanfaat pada data yang tidak terstruktur (Giovani, 2020) penjelasan bahwa pada analisis sentimen akan dibagi menjadi 3, yaitu :

a. Sentimen Positif

Sentimen positif merupakan penggambaran dari pendapat seseorang yang bersifat baik, seperti memberi dukungan, memberi pujian, memberi kepuasan kepada topik masalah yang sedang dibahas.

b. Sentimen Negatif

Sentimen negatif merupakan bentuk opini atas suatu masalah atau topik yang bersifat tidak mendukung dan tidak puas sehingga cenderung tidak memihak pada apa yang disampaikan atau dibicarakan.

c. Sentimen Netral

Sentimen netral merupakan pendapat dari seseorang yang tidak memiliki kecenderungan untuk memihak pihak manapun baik itu positif maupun negatif sehingga opini ini dapat memberikan hasil sebagai penengah dari permasalahan.

2.2.4 Guru

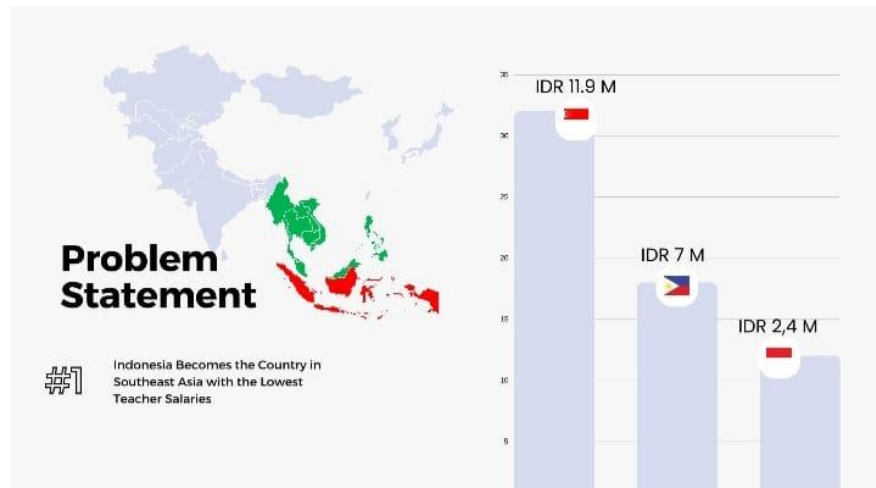
Demi memajukan bangsa, dalam Pembangunan Pendidikan bangsa Indonesia telah ditegaskan bahwa sangatlah penting peran dari seorang guru. Pada persoalan ini, nilai kompetensi dan juga profesionalisme seorang guru sangat penting dalam mengembangkan efektifitas bagi kemajuan nasional. Kualitas guru

menjadi unsur penting dalam pendidikan yang berkualitas. Oleh karena itu, diperlukan upaya untuk meningkatkan mutu tenaga pendidik.

Penugasan materi untuk guru, menunjukkan bahwa masih banyak guru yang penguasaan materinya kurang baik dan masih banyak yang mengalami permasalahan, seperti mengajar diluar bidang keahliannya, hal tersebut mengakibatkan Indonesia menjadi kekurangan orang yang berkompeten. Terdapat empat karakteristik dari guru profesional, menurut Gunawan & Imam, 2023: Dasar kemampuan profesionalnya, mampu untuk menjaga sikap dan sifatnya dalam menguasai setiap materi yang akan dia ajarkan.

- a. Kemampuan upaya profesionalnya, dapat memberikan penentuan atas pembelajaran para siswa juga siswi.
- b. Mampu mengorganisir waktu dengan baik.
- c. Imbalan secara profesional yang memberikan rasa kepuasan.

Berkaitan dengan imbalan secara professional, guru juga seharusnya mendapatkan kesejahteraan untuk hidup layak baik secara materi maupun sosial. Namun, kenyataannya saat ini kesejahteraan guru di Indonesia masih berkategori rendah. Bahkan alokasi anggaran untuk Pendidikan di Indonesia masih dibawah standar UNESCO (*United Nations Educational, Scientific and Cultural Organization*), yaitu hanya 20% dari APBN (Anggaran Pendapatan Belanja Negara). Dapat dilihat dari sebuah data di LinkedIn News Melissa Cantor, mengenai *The Lowest Teacher Salaries* dibawah ini,



Gambar 2.1 Gaji Guru Terendah

Berdasarkan gambar, Indonesia menjadi negara di Asia Tenggara dengan gaji guru terendah. Tak hanya itu bahkan ketimpangan sering terjadi antara standarisasi gaji dan tunjangan guru pada guru negeri dan swasta baik itu perdesaan maupun perkotaan.

2.2.5 Youtube

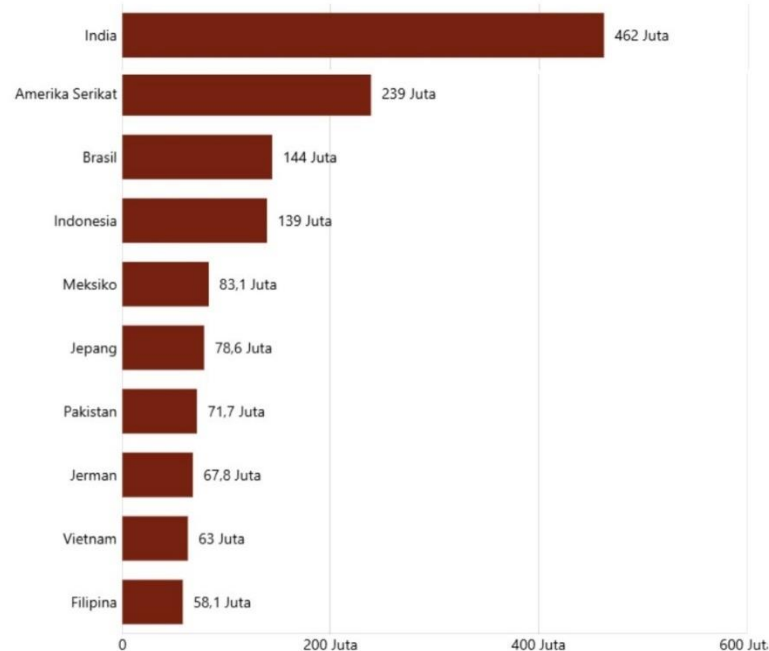


Gambar 2.2 Youtube

Youtube merupakan salah satu layanan yang dapat memberikan gambaran visual dan audio yang juga dapat memberikan kemudahan bagi pengguna yang ingin berkomunikasi. saat ini televisi telah memberikan sarana tontonan ulang dengan pemanfaatan *platform* yang sedang naik daun ini yaitu youtube. Kini

semakin majunya suatu zaman dan sarana berselancar internet, tercatat terdapat 95% pengguna internet adalah pengguna youtube. Setiap hari pengguna youtube menonton video dengan lebih dari satu miliar, bahkan untuk pengguna baru akun youtube mencapai 2 miliar jam setiap harinya.

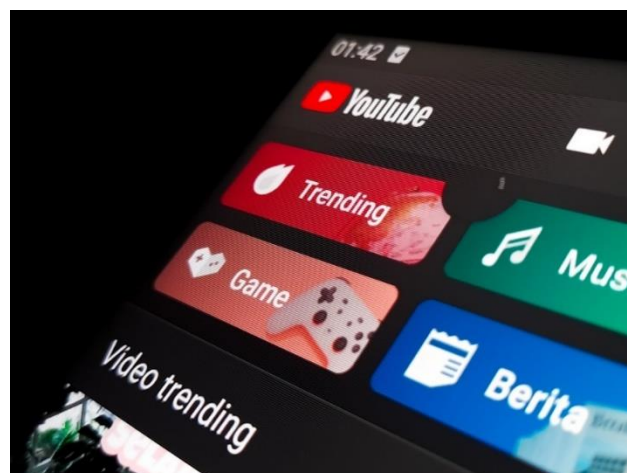
Tak hanya pertelevisian yang menggunakan youtube sebagai medianya namun ada pula yang dinamakan *youtuber*, berarti konten *creator* youtube yang mana terus membagikan tema-tema menarik sebagai bentuk tontonan yang memberikan banyak warna dalam videonya. Ada yang membagikan video tentang *tips* dan *trick*, video edukasi anak, video tentang *fashion*, video memasak, dan lain-lainnya. Mengenai isu sosial yang jarang diperdebatkan bersama juga dapat tayang di youtube. Alasan kemudahan akses informasi inilah yang membuat youtube semakin diminati hampir semua kalangan (Hudha, 2022).



Gambar 2.3 Survei Pengguna Youtube

Berdasarkan survei dari *agency* di Amerika yang dilakukan oleh *we are social* pada tahun 2023, Indonesia menduduki posisi 4 dengan pengguna YouTube terbanyak di dunia yang mencapai hingga 139 juta per Oktober 2023. Diikuti oleh Meksiko dengan 83,1 juta pengguna, Jepang 78,6 juta pengguna, Pakistan 71,7 juta pengguna, Jerman 67,8 juta pengguna, Vietnam 63 juta pengguna, dan Filipina sebanyak 58,1 juta pengguna.

2.2.6 Konten Youtube

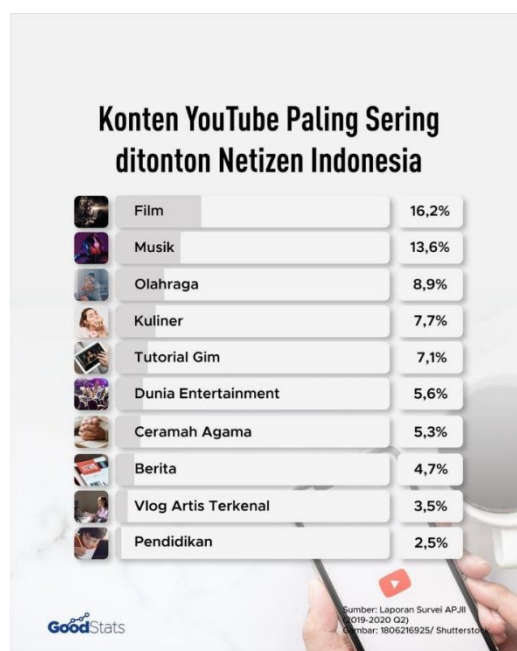


Gambar 2.4 Konten Youtube

Semakin berkembangnya teknologi komunikasi, *smartphone* yang biasa Masyarakat Indonesia gunakan sebagai alat komunikasi penghubung antar daerah kini juga berkembang menjadi media untuk menyalurkan kegiatan literasi digital. Fokus dari kegiatan literasi digital merupakan pemberdayaan pada anak muda lewat internet dengan menyebarkan suatu hal yang bersifat positif di media digital. Youtube merupakan salah satu media yang popularitasnya baik dan sangat digemari

anak muda karena kontennya dikemas berbentuk video, yang tidak hanya bisa ditonton melalui jaringan televisi namun bisa juga melalui aplikasi *mobile*.

Menurut pendapat dari *Team leader Youtube Creator Patnership regional Asia Tenggara dan Australia* menjelaskan bahwa Indonesia memiliki pangsa pasar ekonomi digital yang dapat terus berkembang ke arah yang lebih baik. Namun, dapat dilihat dari jenis konten yang ada di Youtube, Masyarakat lebih tertarik pada konten hiburan. Potensi konten edukatif di Indonesia ternyata belum optimal. Maka dari alasan tersebut pihak youtube di masa mendatang ingin memperbanyak kreator konten, komunitas *youtuber* yang dapat membuat konten edukatif untuk menarik para pengguna youtube di Indonesia (Farah Zhafira, 2021).



Gambar 2.5 Konten Youtube Netizen

Berdasarkan survei diatas, yang dilakukan oleh Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) pada tahun 2023, konten youtube yang paling populer di Indonesia adalah film dengan persentase 16,2%. Hal ini menunjukkan bahwa masyarakat Indonesia menyukai konten hiburan yang dapat dinikmati secara visual.

gambaran dari minat dan ketertarikan masyarakat Indonesia terhadap youtube ini dapat menjadi referensi bagi para kreator konten youtube untuk membuat konten yang menarik dan sesuai dengan minat penonton (Ahmadi, 2021).

2.2.7 Channel Satu Persen



Gambar 2.6 Channel Satu Persen

Satu persen merupakan *channel* youtube yang mengulas banyak hal seputar persoalan yang sedang dibicarakan maupun isu sosial yang jarang dijabarkan. Pandangannya terhadap setiap persoalan dimaknai dari sang pendirinya yaitu Ifandi Khainur Rahim yang merupakan seorang mahasiswa lulusan psikologi Universitas Indonesia pada akhir Desember 2018 membagikan kontennya mengenai pendidikan, *self-development*, kesehatan mental, hingga filsafat. Kemudian dia menyadari potensi dari *channel* youtube Satu Persen pada pertengahan tahun 2019 bersama rekan-rekannya. Satu Persen diseriuskan menjadi sebuah *startup* oleh Evan dan Rizky sebagai rekan dari Ifandi.

Pembuatan *channel* ditujukan agar orang-orang dapat berpikir dan berdiskusi bersama mengenai tema yang akan dibahas secara terbuka di dalam media youtube

Satu Persen. Awal mula berkembangnya Satu Persen pun dahulu masih terbilang biasa, editan dari videonya sangatlah sederhana namun caranya dalam menyampaikan tema memang bagus hingga dapat menjangkau semua kalangan karena penggunaan bahasanya yang mudah dimengerti.

Perkembangan Satu Persen juga merambah pada layanannya yang Bernama, Satu *Creative Agency* dan *training skill* dasar psikologis. Satu persen juga berkolaborasi dengan Sehat Jiwa yang melatarbelakangi persoalan kesehatan jiwa yang belum dapat teratasi dengan baik di Indonesia. Melalui kontribusinya pada *startup* pendidikan *channel* youtube Satu Persen mengantarkan Ifandi masuk ke dalam jajaran *Forbes Under 30*. Satu Persen hingga kini terus mengeluarkan konten bermanfaat yang mengedepankan misinya untuk membuat semua orang memiliki *identity aware, active problem solver, dan growth mindset*.

Channel Satu Persen dipilih karena kemampuannya dalam menyajikan konten yang relevan dan aktual, terutama terkait isu-isu penting seperti peran dan kesejahteraan guru di Indonesia. Pada salah satu kontennya yang berjudul “Guru di Indonesia Gak Dihargai! Ini Alasannya” yang dibagikan tanggal 1 Agustus 2023 membahas mengenai, fenomena guru digaji rendah, kenapa guru digaji rendah dibandingkan profesi lain, guru: *supply* banyak tapi kualifikasi rendah, nilai kompetensi guru Indonesia, sangatlah relevan dengan permasalahan nyata yang sering dibicarakan dalam masyarakat.

Interaksi dari video konten youtube Satu Persen menarik perhatian dengan 5,1 ribu like lalu 132.346 kali tontonan dan dengan *subscriber* yang mencapai 3,25 juta menandakan bahwa *channel* ini memiliki *audiens* yang termasuk luas dan

berpengaruh. Para penonton banyak yang terdorong untuk berdebat dan berdiskusi mengenai konten ini. Hal ini menandakan kalau *channel* youtube Satu Persen tidak hanya menyampaikan informasi namun juga mendorong penonton untuk berpikir kritis.

2.2.8 Youtube Berita Satu Mengenai Guru Digaji Sangat Rendah



Gambar 2.7 Channel Berita Satu

Youtube *channel* Berita Satu berjudul “Mengapa Guru di Indonesia Digaji Sangat Rendah?” yang dibagikan pada 1 Juni 2023 menyampaikan mengenai tantangan kapitalisme dan kemajuan teknologi, para guru semakin dibebani oleh tuntutan profesionalisme yang tinggi, sementara kompensasi yang mereka terima tidak sebanding. Isu ini diperburuk dengan perubahan RUU Sisdiknas yang dirancang oleh Kemendikbud pada tahun 2022, menimbulkan kontroversi karena terkesan buru-buru dan tidak mencantumkan tunjangan profesi guru secara eksplisit. Pada versi RUU Sisdiknas (Sistem Pendidikan Nasional) bulan April

2022, pasal 127 ayat 1-10 masih mengatur mengenai tunjangan untuk profesi guru, tunjangan khusus bagi guru daerah terpencil, dan tunjangan kehormatan dosen. Namun pada bulan Agustus 2022, pasal tersebut diubah menjadi pasal 105, yang hanya menyebutkan bahwa pendidik berhak mendapat upah, jaminan sosial, dan penghargaan sesuai prestasi kerja tanpa menyebutkan secara spesifik tentang tunjangan profesi guru.

Kontroversi ini diperparah dengan kurangnya partisipasi organisasi PGRI (Persatuan Guru Republik Indonesia) dalam proses pembentukan RUU ini. Menurut Ketua Litbang PB PGRI, Sumardiansyah, mereka hanya diberikan waktu kurang dari 10 menit untuk menyampaikan pendapat dan hanya diberikan naskah RUU secara mendadak, sehingga minim ruang dalam membangun suatu dialektika.

Menghilangkan pasal yang mengatur tunjang guru profesi dalam RUU Sisdiknas telah merendahkan martabat guru, karena peraturan tersebut membuat kesejahteraan guru menjadi standar minimum, bahkan di bawah minimum. Jika tunjang profesi dihapuskan maka guru hanya akan mengandalkan gaji pokok, karena tidak semua guru mendapatkan tunjangan khusus dan tunjang fungsional yang jumlahnya tidak signifikan. Selain itu, kenyataannya tunjangan kinerja juga tidak dapat dinikmati oleh semua guru, karena tergantung pada kekuatan APBD (Anggaran Pendapatan Belanja Daerah) daerah masing – masing (Dhobith, 2024).

2.2.9 Instant Data Scraper



instant Data Scraper CRX 1.0.8 for Chrome

Gambar 2.8 Instant Data Scraper

Tools Instant Data Scraper menggunakan AI (*Artificial Intelligence*) digunakan untuk memprediksi data yang tepat dan menyimpannya ke dalam file Excel. Penggunaan alat ini tidak memerlukan skrip pada situs web. Saat menggunakannya terdapat pengaturan waktu *delay* sehingga proses data yang diambil efisien (Setiyani, 2020). Instant Data Scraper termasuk ekstensi *browser* yang memudahkan kita untuk mengekstrak data dari situs web secara gratis dan memudahkan analisis lebih lanjut. Setelah ekstensi mendeteksi data pada halaman terdapat pilihan tombol *download* CSV atau *export* agar data tersimpan di komputer sesuai dengan format yang diinginkan.

Instant Data Scraper dapat digunakan dalam berbagai situasi dan industri yang memerlukan proses ekstraksi data dari situs web. Dalam analisis bisnis Instant Data Scraper berguna untuk mengambil data untuk menganalisis tren dan pembuatan laporan bisnis.

2.2.10 Python



Gambar 2.9 Python

Python merupakan bahasa pemrograman tingkat tinggi yang mendukung pemrograman berorientasi objek. Python memiliki perbedaan dengan bahasa pemrograman lain yaitu dalam penulisan sintaks yang relatif mudah dan sederhana. Hal ini menjadikan Python sebagai bahasa pemrograman yang populer untuk berbagai keperluan, termasuk analisis data (Duei Putri, 2022).

Kecepatan Python menjadikannya bahasa pemrograman yang cocok untuk analisis data yang melibatkan data besar. Kemudahan penggunaan Python menjadikannya bahasa pemrograman yang cocok untuk pemula. Fleksibilitas Python memungkinkannya untuk digunakan untuk berbagai jenis analisis data. Komunitas pengguna Python yang besar menyediakan dukungan dan sumber daya yang berlimpah (Sandi & Utami, 2023).

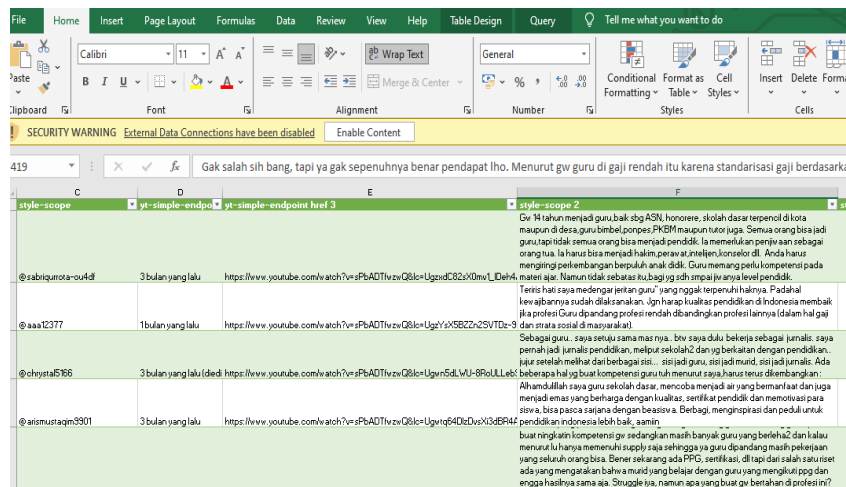
2.2.11 Google Colab



Gambar 2.10 Google Colab

Google Colab merupakan layanan gratis dari google untuk menulis dan menjalankan Bahasa Pemrograman Python langsung pada *browser*, biasa disebut *online- based service*. Akses gratis dari Google Colab hanyalah memerlukan akun google untuk memulai membuka *notebook* Colab. Layanan ini berguna untuk pembelajaran mesin ataupun analisis data. Kelebihan dari Google Colab ini adalah hasil *code* dapat langsung disimpan ke dalam Google Drive sehingga memudahkan proses penyimpanan dan pengaksesan kembali. Google Colab ini sangat cocok untuk analisis data dan visualisasi data karena penggunaannya yang fleksibel dengan memberikan akses gratis komputasi GPU dan TPU, sangat penting untuk pengerjaan tugas yang memerlukan komputasi tinggi.

2.2.12 Microsoft Excel



Gambar 2.11 Microsoft Excel

Microsoft Excel adalah aplikasi perangkat lunak lembar kerja spreadsheet yang kuat yang dikembangkan oleh Microsoft untuk Windows, macOS, Android, dan iOS. Ini adalah alat penting untuk analisis data, pemodelan keuangan, penganggaran, dan membuat presentasi. Microsoft Excel membantu pengguna di seluruh dunia mengatur data, menganalisis keuangan, dan membuat presentasi. Dengan antarmuka yang intuitif, rumus yang beragam, dan opsi visualisasi data yang canggih

2.2.13 Metode *Naïve Bayes*

Metode *Naïve Bayes* merupakan sebuah metode yang membantu dalam mengklasifikasi, dengan perhitungan probabilitas yang menggunakan statistik dengan memperkirakan kehadiran dan kekurangan suatu kelas (Ahmadi, 2021). Metode ini memberikan kejelasan dari peluang tentang kejadian masa depan atau mampu memprediksi masa depan. Pemilihan metode ini juga didasari karena mulai banyak yang menggunakannya. Begitu pula alasannya didasari oleh simpelnya

metode ini dengan hasil yang cepat dan ketepatannya yang tinggi. Berikut bentuk umumnya :

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

Keterangan :

C = sebuah *Class* spesifik

X = *vector* inputan atau atribut

$P(C|H)$ = kemungkinan dari *Class* yang didasari oleh inputan X yang telah diketahui atau *posterior probability*.

$P(X|C)$ = kemungkinan setiap input X didasari kondisi *Class* atau *likelihood*.

$P(C)$ = kemuninan dari *Class* yang dicari atau *prior probability* terhadap semua data.

$P(X)$ = kemungkinan keseluruhan data suatu inputnya.

2.2.14 Confusion Matrix

		PREDICTED	
		Positive	Negative
ACTUAL	Positive	TRUE POSITIVE	FALSE NEGATIVE
	Negative	FALSE POSITIVE	TRUE NEGATIVE

Gambar 2.12 Confusion Matriks

Confusion matrix adalah suatu alat yang digunakan untuk melakukan pengukuran mengenai seberapa baik *classifier* yang berfungsi untuk mengenali tuple dari bentuk kelas yang berbeda. Menggunakan *confusion matrix* hasil yang didapat akan berdasarkan pada penghitungan *accuracy*, *recall*, *precision*, dan *f1-score* (Farah Zhafira, 2021).

- a. *Accuracy* adalah bagian metrik yang menunjukkan persentase data yang akan diklasifikasikan benar oleh model. Pengukuran ini merupakan salah satu metrik evaluasi untuk mengukur kinerja keseluruhan seberapa baik model dapat membuat prediksi benar.

$$Accuracy = \frac{TP_{Negatif} + TP_{Netral} + TP_{positif}}{Total} \times 100\%$$

$TP_{Negatif}$ = Jumlah *True Positive* untuk sentimen negatif.

TP_{Netral} = Jumlah *True Positive* untuk sentimen netral.

$TP_{positif}$ = Jumlah *True Positive* untuk sentimen positif.

- b. *Precision* adalah pengukur dari seberapa tepat prediksi positif yang telah dibuat model. Jadi presisi menunjukkan proporsi dari data yang diprediksi positif dan sudah benar positif. . Pengukuran ini sangatlah penting pada situasi dimana *false positives* (prediksi positif yang salah) harus diminimalkan.

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

$TP(\text{true positives})$ = jumlah kasus diprediksi positif, benar-benar positif.

$FP(\text{false positives})$ = jumlah kasus yang diprediksi positif, sebenarnya negatif.

- c. *Recall* adalah mengukur lengkapnya model untuk mengidentifikasi semua data positif yang sebenarnya. Pengukuran ini sangatlah penting pada situasi dimana *false negatives* (kasus positif yang tidak terdeteksi) harus diminimalkan.

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

$TP(\text{true positives})$ = jumlah kasus diprediksi positif, benar-benar positif.

$FN(\text{false negatives})$ = jumlah kasus diprediksi negatif, tetapi sebenarnya positif.

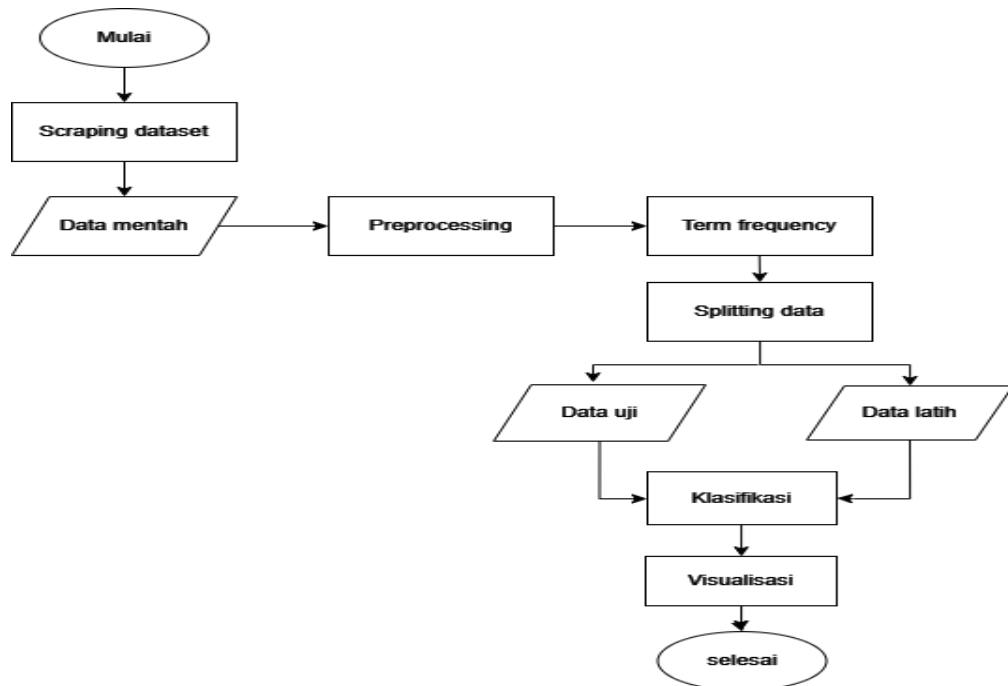
- d. *F-1 Score* adalah penggabungan dari *precision* dan *recall* yang akan memberikan gambaran keseluruhan kinerja model klasifikasi. Kegunaannya sebagai penyeimbang antara kedua nilai saat terjadi suatu distribusi kelas yang tidak merata.

$$F1\ Score = 2 \times \frac{presisi \times recall}{presisi + recall}$$

presisi = hasil dari mengukur semua data positif yang sebenarnya.

recall = hasil dari seberapa tepat prediksi positif.

2.3 Kerangka Pemikiran



Gambar 2.13 Kerangka Pemikiran

2.3.1 Scraping Data Komentar

Scraping ini merupakan tahap yang akan digunakan untuk mengambil data dengan pengumpulan suatu data dari *database*, sebelum data tersebut diolah atau dianalisis. Pada pengumpulan yang dilakukan adalah mengunduh data dari komentar pada youtube (Ole, 2023).

2.3.2 Data Mentah

Data mentah merupakan data yang belum mengalami pengolahan sehingga masih asli (Diki Hendriyanto, 2022). Pada saat pengunduhan data nantinya data

yang diunduh akan berasal dari komentar youtube Satu Persen yang berfokus pada salah satu konten mereka tentang guru.

2.3.3 *Preprocessing*

Tahap *Preprocessing* adalah tahap pembangunan data yang bertujuan untuk mengurangi atribut yang tidak penting dalam proses klasifikasi data. Data yang digunakan pada proses ini adalah data mentah yang masih kotor. Proses dalam *preprocessing* meliputi *cleaning*, *labelling*, *case folding*, *filtering*, *tokenizing*, dan *stemming*.

- a. *Cleaning* adalah Tahap pembersihan, misalnya karakter selain huruf jadi para prosesnya akan dihapus URL dari setiap komentar. Jadi tujuan untuk *cleaning* ini adalah membersihkan suatu komentar.
- b. *Labelling* merupakan tahap penentuan label dari setiap komentar dalam dokumen. Tujuannya memberikan bobot pengelompokkan polaritas sehingga menghasilkan sentimen positif, sentimen negatif, dan sentimen netral (Mujahidin, 2022).
- c. *Case Folding* merupakan tahap pengubahan huruf yang terjadi misalkan terdapat huruf yang diubah menjadi huruf kecil semua. Langkahnya dengan memeriksa setiap karakter lalu nantinya akan ditemukan huruf kapital yang langsung dijadikan huruf kecil.
- d. *Filtering* adalah setiap kata akan muncul pada kategorinya dan pada kata - kata yang tidak bermakna untuk pengkategorian dokumen, nantinya kata-kata tersebut akan dibuang.

- e. *Tokenizing* adalah Tahap pemrosesan untuk memisahkan kalimat menjadi beberapa bagian kata yang biasa disebut dengan *token*. Tahap ini dilakukan agar mempermudah pada bagian proses selanjutnya (Setiyani, 2020).
- f. *Stemming* merupakan pengubahan dari kata berimbuhan menjadi kata dasar. bertujuan untuk menghemat waktu pengklasifikasian (Ahmadi, 2021).

2.3.4 *Term Frequency*

Term Frequency merupakan bentuk dari proses perhitungan mengenai jumlah term pada suatu dokumen. Nantinya akan terjadi pembobotan atas setiap kata yang akan muncul di sebuah dokumen (Diki Hendriyanto, 2022).

2.3.5 *Splitting data*

Splitting data merupakan pembagian data pada dokumen menjadi dua bagian, yaitu dinamakan set pelatihan dan set pengujian. Tujuannya agar pengembangan model analisis sentimen dapat bekerja dengan baik dalam memprediksi sentimen data.

2.3.6 *Data Latih*

Data latih merupakan data yang digunakan dalam terbentuknya suatu *classifier*. Data ini berguna untuk mempermudah pembelajaran sistem juga memahami tentang *machine learning* (Hudha, 2022).

2.3.7 Data Uji

Pada tahap ini data uji merupakan evaluasi mengenai uji hasil dari pembelajaran tentang algoritma serta pembelajaran mesin (Darwis, 2021). Data testing ini akan mengukur sejauh mana keberhasilan dari pengklasifikasian.

2.2.8 Visualisasi *Word Cloud*

Visualisasi *word cloud* merupakan visualisasi yang menampilkan data teks ke dalam bentuk visual sehingga mudah dibaca dan dipahami. Hasil dari *word cloud* dapat memberikan *insight* dengan tampilannya yang menunjukkan setiap kata dari data set muncul memenuhi *word cloud*, kata tersebut merepresentasikan frekuensi kata yang paling sering muncul. Ukuran kata yang berbeda pada *word cloud* mengartikan kata tersebut menjadi topik yang paling sering dibicarakan. Adanya visualisasi *word cloud* memudahkan proses filter pada sentiment yang kurang relevan sehingga mempersingkat waktu (Hutabarat, 2024).