

WEAK SUPERVISION DENGAN PENDEKATAN LABELING FUNCTION UNTUK ANALISIS SENTIMEN PADA TWITTER

Hastari Utama^{1, a)}, Erna Daniati^{2, b)}, Ahlihi Masruro^{3, c)}

^{1, 3)}Program Studi Teknik Informatika, ²⁾Program Studi Sistem Informasi,

^{a, c)}Universitas Amikom Yogyakarta, Jl. Ring Road Utara, Condong Catur, Depok, Sleman, 55283, Indonesia

^{b)} Universitas Nusantara PGRI Kediri, Mojoroto Gang 1, Mojoroto, 64112, Kota Kediri, Indonesia

Author Emails

Corresponding author: ^{a)} utama@amikom.ac.id

^{b)} ernadaniati@unpkediri.com, ^{c)} ahlihi@amikom.ac.id

Abstract. *The use of social media today has increased rapidly. One type of social media used is Twitter. This social media has billions of users from all over the world. So, in a short time the tweet data they posted was in storage. Each user is also limited to the number of characters in sending tweets. However, this collection of tweets on social media has a varied theme context. This can contain emotional sentiments such as happy, sad, joyful, sorrow, and so on. The various types of tweet data provided have great potential for analysis, especially for profit-based companies. This can include customer habits, product trends, stock indices, and so on. One type of analysis carried out is sentiment analysis. This is useful for classifying existing tweet opinions. This opinion can be positive, negative, or neutral. The results of this analysis are very necessary for companies to understand the trends occurring in this era. One solution offered in this research is to use Weak Supervision. However, there are challenges that occur with this method. This is a lower level of accuracy when compared to manual labeling. In this research, automatic labeling was carried out with weak supervision. Apart from that, a labeling function and Regex Pattern approach was used to carry out labeling automatically. It is hoped that the labeled dataset will produce a model with a level of accuracy close to that of manual labeling, especially if it can outperform manual methods. In addition, the expected contribution in this research is an effort to shorten labeling time rather than doing it manually. The research summary is no more than 500 words containing the research background, proposed research objectives and methods, targeted outcomes, and a description of the research contribution. Use the Abstract style in this section with one paragraph.*

Keywords :

Twitter, Weak Supervision, Labeling Function, Naive Bayes.

Abstraksi. Penggunaan sosial media saat ini telah meningkat dengan pesat. Salah satu jenis media sosial yang digunakan adalah Twitter. Media sosial ini memiliki miliaran pengguna dari seluruh dunia. Jadi, dalam waktu yang singkat data tweet yang mereka posting telah ada pada penyimpanannya. Setiap pengguna juga dibatasi jumlah karakter dalam melakukan pengiriman tweetnya. Namun, kumpulan tweet pada media sosial ini memiliki konteks tema yang bervariasi. Hal ini dapat memuat sentimen emosional seperti senang, sedih, gembira, duka, dan sebagainya. Berbagai jenis data tweet yang disediakan ini sangat berpotensi untuk dianalisis terutama bagi perusahaan berbasis profit. Hal ini dapat memuat kebiasaan pelanggan, tren produk, indeks saham, dan sebagainya. Salah satu jenis analisis yang dilakukan adalah sentiment analysis. Hal ini berguna untuk mengklasifikasikan opini tweet yang ada. Opini ini dapat berupa positif, negatif, atau netral. Hasil analisis ini sangat diperlukan perusahaan untuk mengetahui tren yang terjadi pada era ini. Salah satu solusi yang ditawarkan pada penelitian ini adalah dengan menggunakan Weak Supervision. Namun, ada tantangan yang terjadi pada metode ini. Hal ini adalah kurangnya tingkat akurasi jika dibandingkan dengan pelabelan secara manual. Pada penelitian ini dilakukan pelabelan otomatis dengan weak supervision. Selain itu, dilakukan pendekatan labeling function dan Regex Pattern dalam melakukan pelabelan secara otomatis. Hal ini diharapkan dataset yang dilabeli akan menghasilkan model dengan tingkat akurasi mendekati pelabelan secara manual, lebih lagi jika dapat mengungguli metode manual. Selain itu, kontribusi yang diharapkan dalam penelitian ini adalah usaha untuk mempersingkat waktu pelabelan daripada dilakukan

secara manual. Ringkasan penelitian tidak lebih dari 500 kata yang berisi latar belakang penelitian, tujuan dan metode penelitian yang diusulkan, luaran yang ditargetkan, serta uraian kontribusi penelitian. Gunakan style *Abstract* pada bagian ini dengan satu paragraf.

Kata Kunci : Twitter, Weak Supervision, Labelling Function, Naïve Bayes.

PENDAHULUAN

Penggunaan sosial media saat ini telah meningkat dengan pesat. Salah satu jenis media sosial yang digunakan adalah Twitter [1]. Media sosial ini memiliki miliaran pengguna dari seluruh dunia. Jadi, dalam waktu yang singkat data tweet yang mereka posting telah ada pada penyimpanannya. Setiap pengguna juga dibatasi jumlah karakter dalam melakukan pengiriman tweetnya. Namun, kumpulan tweet pada media sosial ini memiliki konteks tema yang bervariasi. Hal ini dapat memuat sentimen emosional seperti senang, sedih, gembira, duka, dan sebagainya. Selain itu, penggunaannya juga berbagai macam jenis. Mulai dari anak muda sampai orang dewasa baik pelajar maupun pemerintahan juga menggunakannya. Twitter juga memfalsifikasi fitur untuk dapat mengakses kumpulan data tweetnya melalui API. Platform ini tidak hanya menjadi alat komunikasi sosial yang penting tetapi juga menyediakan peluang yang besar bagi perusahaan untuk terlibat dengan pelanggan, membangun merek, dan mengikuti perkembangan industri. Salah satu perkembangan utama adalah pemanfaatan Twitter sebagai saluran komunikasi langsung antara perusahaan dan pelanggan. Banyak perusahaan menggunakan akun Twitter untuk merespons pertanyaan, menyampaikan informasi produk atau layanan, dan mengatasi keluhan pelanggan secara langsung. Hal ini memungkinkan perusahaan untuk menciptakan hubungan yang lebih dekat dengan pelanggan dan meningkatkan kepuasan mereka.

Berbagai jenis data tweet yang disediakan ini sangat berpotensi untuk dianalisis terutama bagi perusahaan berbasis profit. Hal ini dapat memuat kebiasaan pelanggan, tren produk, indeks saham, dan sebagainya. Twitter memiliki potensi besar untuk dianalisis oleh perusahaan dalam berbagai aspek. Pertama-tama, platform ini menyediakan akses ke data real-time mengenai tren, opini, dan percakapan publik. Dengan menganalisis pola-pola tersebut, perusahaan dapat memahami sentimen konsumen terhadap produk atau layanan mereka. Selain itu, Twitter juga merupakan sumber informasi yang berharga untuk memantau kegiatan pesaing dan tren industri. Analisis data Twitter dapat membantu perusahaan mengidentifikasi peluang baru, menilai efektivitas kampanye pemasaran, dan merespon dengan cepat terhadap perubahan dalam lingkungan bisnis. Selain itu, Twitter juga dapat digunakan sebagai alat untuk membangun citra merek dan terlibat dalam interaksi langsung dengan pelanggan. Dengan memahami pola perilaku pengguna di Twitter, perusahaan dapat mengoptimalkan strategi komunikasi mereka dan meningkatkan keterlibatan dengan audiens. Dengan demikian, potensi analisis Twitter bagi perusahaan sangat signifikan dalam mendukung pengambilan keputusan strategis dan meningkatkan daya saing di pasar. Perusahaan dapat meningkatkan keuntungannya setelah mengambil keputusan berbasis data analisis ini. Oleh karena itu, Twitter menjadi sangat penting untuk semua kalangan entitas.

Salah satu jenis analisis yang dilakukan adalah analisis sentimen. Analisis sentimen adalah proses evaluasi dan interpretasi opini, perasaan, atau sikap yang terkandung dalam teks, seperti ulasan produk, komentar media sosial, atau artikel berita. Tujuan utama dari analisis sentimen adalah untuk memahami bagaimana orang merespons suatu topik atau produk, apakah respon tersebut bersifat positif, negatif, atau netral. Beberapa langkah umum dalam analisis sentimen melibatkan pengumpulan data teks, pemrosesan teks, dan pemberian label sentimen. Hasil analisis ini sangat diperlukan perusahaan untuk mengetahui tren yang terjadi pada era ini [2]. Namun, terdapat kendala dalam melakukan analisis ini. Data yang dianalisis dalam jumlah yang masif. Selain itu, setiap baris data perlu dilakukan pelabelan. Pelabelan data ini biasanya dilakukan dengan cara yang manual. Apalagi data yang dianalisis mencapai 1 juta baris data. Hal ini tentu perlu waktu yang lama untuk dilakukan pelabelan. Hal ini juga ditambah dengan waktu proses yang lama saat dilakukan pelatihan data. Hal ini menjadi permasalahan utama yang diangkat dalam penelitian ini.

Salah satu solusi yang ditawarkan pada penelitian ini adalah dengan menggunakan Weak Supervision [3]. Metode ini diawali dengan pelabelan dari data sampel yang diambil dari dataset yang ada. Kemudian, dilakukan pemodelan dari data sampel tersebut. Model yang didapatkan kemudian digunakan untuk memprediksi label dari dataset lainnya yang belum terlabeli. Hal ini akan menggantikan peran manusia dalam melabeli data yang ada. Namun, ada tantangan yang terjadi pada metode ini. Hal ini adalah kurangnya tingkat akurasi jika dibandingkan dengan pelabelan secara manual.

Pada penelitian ini dilakukan pelabelan otomatis dengan weak supervision. Selain itu, dilakukan pendekatan labeling function dan Regex Pattern dalam melakukan pelabelan secara otomatis. Hal ini diharapkan dataset yang dilabeli akan menghasilkan model dengan tingkat akurasi mendekati pelabelan secara manual, lebih lagi jika dapat mengungguli metode manual. Selain itu, kontribusi yang diharapkan dalam penelitian ini adalah usaha untuk mempersingkat waktu pelabelan daripada dilakukan secara manual.

TINJAUAN PUSTAKA

Penghapusan teks pemandangan adalah tugas menghapus teks dari gambar pemandangan alam, yang telah mendapatkan perhatian dalam beberapa tahun terakhir. Motivasi utama adalah untuk menyembunyikan informasi pribadi seperti nomor plat, dan papan nama rumah yang dapat muncul dalam gambar [4]. Pada penelitian ini, kami mengusulkan metode untuk menghapus teks adegan yang mendekati masalah sebagai tugas pengecatan umum dengan melibatkan weak supervision. Berbeda dengan metode sebelumnya, yang membutuhkan pasangan gambar asli yang berisi teks dan gambar dari mana teks telah dihapus, metode ini tidak memerlukan pasangan gambar yang sesuai untuk pelatihan. Hal ini menggunakan detektor teks adegan yang dilatih secara terpisah dan jaringan inpainting. Detektor teks adegan memprediksi peta segmentasi instance teks yang kemudian digunakan sebagai topeng untuk jaringan inpainting. Jaringan untuk inpainting, dilatih pada dataset gambar skala besar, mengisi daerah yang disamarkan dalam gambar input dan menghasilkan gambar akhir di mana teks asli tidak lagi ada. Hasilnya menunjukkan bahwa metode kami berhasil menghapus teks dan mengisi lubang yang dibuat untuk menghasilkan gambar yang tampak alami.

Clickbait adalah tantangan yang sulit dipahami dengan prevalensi media sosial seperti Facebook dan Twitter yang menyesatkan pembaca saat mengklik berita utama [5]. Data beranotasi yang terbatas menyulitkan untuk merancang sistem identifikasi clickbait yang akurat. Peneliti mengatasi masalah ini dengan menggunakan arsitektur berbasis pembelajaran mendalam dengan pengetahuan eksternal yang dilatih di pos dan deskripsi media sosial. Model ELMO dan BERT yang telah dilatih sebelumnya memperoleh fitur kontekstual tingkat kalimat sebagai pengetahuan; terlebih lagi, lapisan LSTM membantu menonjolkan fitur kontekstual tingkat kata. Pelatihan dilakukan pada eksperimen yang berbeda (model dengan EMLO, model dengan BERT) dengan teknik regularisasi yang berbeda seperti dropout, early stop, dan finetuning. Meneruskan sistem identifikasi tweet clickbait kontekstual (FCCTI) dengan penyempurnaan BERT dan model dengan ELMO menggunakan penyematanan pra-pelatihan sarung tangan adalah model terbaik dan mencapai akurasi identifikasi clickbait sebesar 0,847, meningkat dari baseline sebelumnya untuk penelitian ini.

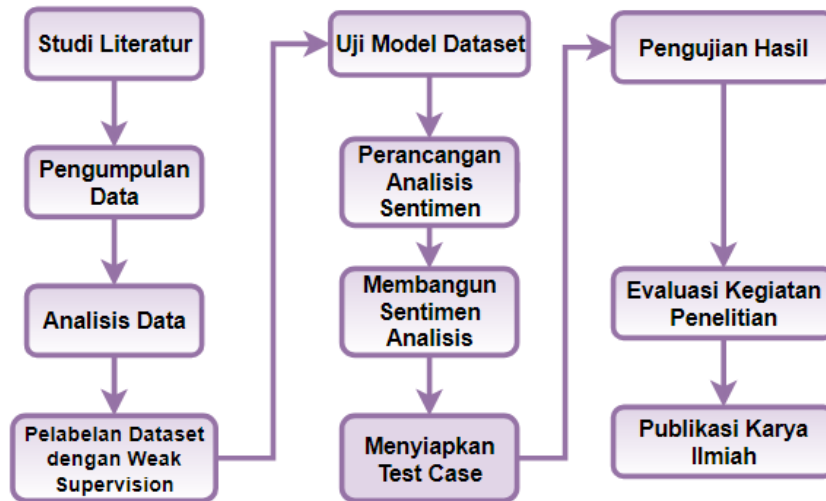
Pengembang menggunakan pencarian untuk berbagai tugas seperti menemukan kode, dokumentasi, informasi debug, dll. Secara khusus, pencarian web banyak digunakan oleh pengembang untuk menemukan contoh kode dan potongan selama proses pengkodean [6]. Baru-baru ini, pencarian kode berbasis bahasa alami telah menjadi area penelitian yang aktif. Namun, kurangnya kumpulan data skala besar di dunia nyata adalah hambatan yang signifikan. Dalam karya ini, kami mengusulkan pendekatan berbasis pengawasan yang lemah untuk mendeteksi maksud pencarian kode dalam kueri pencarian untuk bahasa pemrograman C# dan Java. Kami mengevaluasi pendekatan terhadap beberapa baseline pada dataset dunia nyata yang terdiri dari lebih dari 1 juta kueri yang ditambang dari mesin pencari web Bing dan menunjukkan bahwa model berbasis CNN dapat mencapai akurasi masing-masing 77% dan 76% untuk C# dan Java. Selain itu, kami juga merilis Search4Code, kumpulan data dunia nyata pertama dari kueri pencarian kode yang ditambang dari mesin pencari web Bing. Kami berharap kumpulan data ini akan membantu penelitian di masa mendatang tentang pencarian kode.

Menggunakan data dari sensor detektor lingkaran untuk deteksi insiden lalu lintas hampir real-time di jalan raya sangat penting untuk mencegah kemacetan lalu lintas besar [7]. Sementara metode pembelajaran mesin yang diawasi baru-baru ini menawarkan solusi untuk deteksi insiden dengan memanfaatkan data insiden berlabel manusia, tingkat alarm palsu seringkali terlalu tinggi untuk digunakan dalam praktik. Secara khusus, inkonsistensi dalam pelabelan manusia atas insiden secara signifikan mempengaruhi kinerja model pembelajaran terawasi. Untuk itu, kami fokus pada pendekatan data-sentris untuk meningkatkan akurasi dan mengurangi tingkat alarm palsu deteksi insiden lalu lintas di jalan raya. Kami mengembangkan alur kerja pembelajaran terawasi yang lemah untuk menghasilkan label pelatihan berkualitas tinggi untuk data insiden tanpa label kebenaran dasar, dan kami menggunakan label yang dihasilkan tersebut dalam persiapan pembelajaran terawasi untuk deteksi akhir. Pendekatan ini terdiri dari tiga tahap. Pertama, kami memperkenalkan pipa prapemrosesan dan kurasi data yang memproses data sensor lalu lintas untuk menghasilkan data pelatihan berkualitas tinggi melalui pemanfaatan fungsi pelabelan, yang dapat berupa aturan heuristik yang terkait dengan domain atau sederhana. Kedua, kami mengevaluasi data pelatihan yang dihasilkan oleh

pengawasan yang lemah menggunakan tiga model pembelajaran yang diawasi—hutan acak, k-tetangga terdekat, dan ansambel mesin vektor pendukung—dan pengklasifikasi memori jangka pendek yang panjang. Hasil penelitian menunjukkan bahwa akurasi semua model meningkat secara signifikan setelah menggunakan data pelatihan yang dihasilkan oleh pengawasan yang lemah. Ketiga, kami mengembangkan pendekatan deteksi insiden real-time online yang memanfaatkan ansambel model dan kuantifikasi ketidakpastian saat mendeteksi insiden. Secara keseluruhan, kami menunjukkan bahwa alur kerja pembelajaran terawasi lemah yang kami usulkan mencapai tingkat deteksi insiden yang tinggi (0,90) dan tingkat alarm palsu yang rendah (0,08).

METODE PENELITIAN

Pada Gambar 1 ditunjukkan tahapan penelitian. Tahapan ini diawali dengan studi literatur yang dilakukan dengan tema weak supervision. Hal ini dilanjutkan dengan tahap pengumpulan data. Pengumpulan data ini dilakukan dengan mengunduh data tweet dari website kaggle.com mengenai dataset tweet penerbangan di Amerika Serikat tahun 2015. Dataset ini terdiri dari 14.641 baris data. Sumber dari dataset ini berasal dari alamat url <https://www.kaggle.com/crowdfLOWER/twitter-airline-sentiment>. Dataset ini perlu dianalisis untuk ditentukan atribut yang akan digunakan dalam pembersihan data. Selanjutnya, pada tahap pelabelan data menggunakan metode Weak Supervision. Metode spesifik yang digunakan adalah fungsi pelabelan. Hal ini digunakan sebagai solusi yang ditawarkan untuk pelabelan dengan waktu yang lebih cepat daripada dilakukan secara manual. Dataset yang telah dilabeli selanjutnya dilakukan pelatihan data untuk mendapatkan model yang digunakan untuk membangun Sentiment Analysis Engine (SAE). Hal ini perlu dilakukan evaluasi dalam bentuk pengujian hasil. Evaluasi ini memuat 4 parameter uji yaitu akurasi, recall, precision, dan f1 score.



Gambar 1. Alur Penelitian

Pada Gambar 2 ditunjukkan penerapan weak supervision pada sentiment analysis. Hal ini terdiri dari label matrix yang merupakan matriks dari label yang dikumpulkan. Selanjutnya terdapat Learning Function yang merupakan sumber pengawasan yang lemah umumnya mencakup basis pengetahuan eksternal, pola, kamus, dan bahkan heuristik khusus domain. Dalam konteks klasifikasi maksud pencarian kode, kami memanfaatkan pengklasifikasi sub-maksud rekayasa perangkat lunak [6]. Langkah awal pada alur ini dimulai dengan tokenizing. Tokenizing pada teks adalah proses memecah teks menjadi unit-unit kecil yang disebut token. Token adalah unit dasar dalam pemrosesan bahasa alami dan dapat berupa kata, subkata, frasa, atau karakter, tergantung pada tingkat kompleksitas yang diinginkan. Tujuan utama dari tokenizing adalah mengubah teks menjadi format yang dapat diolah lebih lanjut oleh model atau algoritma pemrosesan bahasa alami.

Selanjutnya, proses dilanjutkan pada stop word removal. Stop word removal adalah tahap dalam pra-pemrosesan teks yang melibatkan penghapusan kata-kata penghenti atau stop words dari suatu dokumen atau kalimat. Stop words adalah kata-kata umum seperti "dan", "atau", "di", yang umumnya tidak membawa makna khusus dalam analisis teks

dan cenderung sering muncul dalam dokumen. Dengan menghapus stop words, tujuan utamanya adalah mengurangi dimensi data dan meningkatkan efisiensi analisis teks. Penghapusan stop words dapat membantu dalam fokus pada kata-kata yang lebih informatif dan relevan, yang pada gilirannya dapat meningkatkan kualitas ekstraksi fitur atau analisis sentimen. Namun, perlu dicatat bahwa dalam beberapa konteks atau tugas analisis teks, beberapa stop words mungkin memiliki arti khusus dan sebaiknya tidak dihapus. Meskipun demikian, stop word removal tetap menjadi teknik yang umum digunakan dalam pra-pemrosesan data teks untuk memperbaiki kinerja model dan meningkatkan interpretabilitas hasil analisis.

Stemming adalah proses dalam pra-pemrosesan teks yang bertujuan untuk menghapus awalan atau akhiran kata sehingga hanya menyisakan akar kata atau kata dasarnya. Tujuan dari stemming adalah untuk mengurangi variasi kata ke bentuk dasarnya, sehingga kata-kata yang serupa dapat dianggap sebagai satu entitas. Hal ini membantu meningkatkan konsistensi dan mengurangi dimensi data dalam analisis teks. Contoh dari proses stemming adalah mengubah kata-kata seperti "berlari", "berlarian", dan "lari" menjadi bentuk dasarnya, yaitu "lar". Dengan cara ini, kata-kata yang semantisnya serupa dapat dianggap sebagai satu kelompok, memudahkan analisis teks dan ekstraksi pola. Stemming sering digunakan dalam aplikasi pemrosesan bahasa alami (NLP) dan sistem temu kembali informasi. Meskipun stemming dapat membantu mengurangi variasi kata, perlu diingat bahwa hasil stemming tidak selalu menghasilkan kata yang benar dalam bahasa yang dimaksud, karena beberapa kata hasil stemming mungkin tidak bermakna atau tidak digunakan dalam bahasa sehari-hari.

Dalam konteks Weak Supervision, beberapa konsep utama melibatkan Label Matrix, Learning Function, Generative Model, dan Discriminative Model.

1. Label Matrix

Label matrix adalah representasi dari label atau anotasi yang diberikan pada data pelatihan. Namun, dalam konteks Weak Supervision, label pada label matrix mungkin tidak sepenuhnya akurat atau pasti. Label dapat berasal dari berbagai sumber, seperti label heuristik, noisy, atau sebagian terlewat.

2. Learning Function

Learning function merujuk pada fungsi atau model pembelajaran yang digunakan untuk menghubungkan input data dengan output label. Fungsi pembelajaran ini dapat berasal dari berbagai jenis model pembelajaran mesin, termasuk model berbasis statistik, jaringan saraf, atau metode pembelajaran mesin lainnya. Dalam konteks Weak Supervision, learning function harus dapat menangani ketidakpastian atau kebisingan dalam label data.

3. Generative Model

Generative model adalah tipe model yang mencoba memahami dan memodelkan distribusi probabilitas dari data. Dalam konteks Weak Supervision, generative model dapat digunakan untuk memodelkan hubungan antara data dan label serta memahami ketidakpastian dalam label. Generative model dapat menghasilkan label yang sesuai dengan distribusi probabilitas yang ditemukan.

4. Discriminative Model

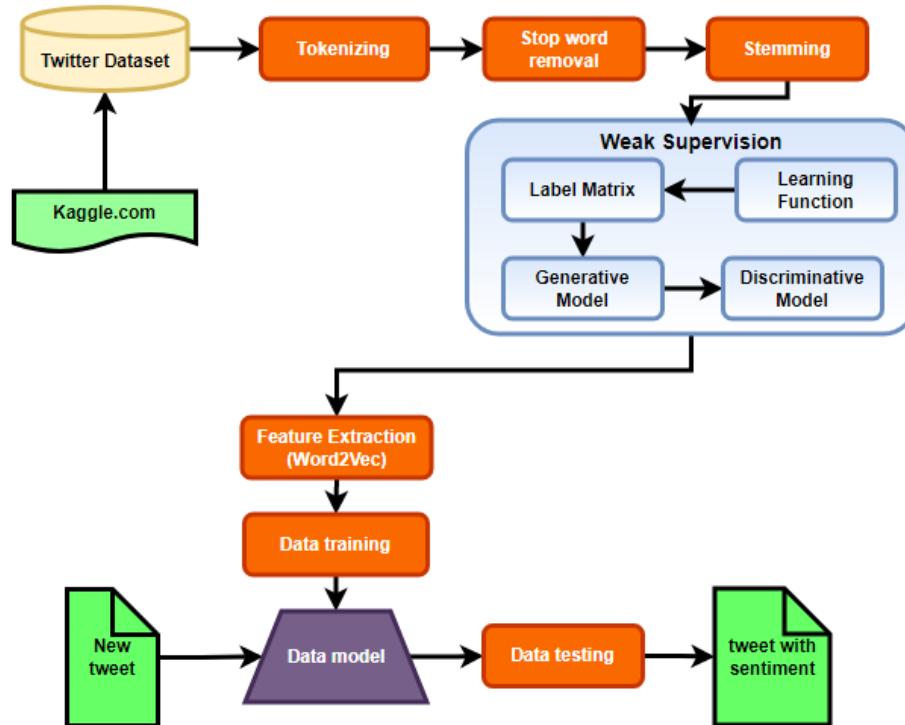
Sebaliknya, discriminative model fokus pada memodelkan batas keputusan langsung antara kelas atau label. Dalam konteks Weak Supervision, discriminative model dapat digunakan untuk mengidentifikasi pola-pola yang membedakan antara kelas atau label, walaupun label mungkin noisy atau tidak pasti.

Dalam aplikasi Weak Supervision, label matrix sering kali tidak dapat diandalkan sepenuhnya, dan learning function harus mampu menangani ketidakpastian, kebisingan, atau keterbatasan dalam data pelatihan. Generative model dan discriminative model adalah dua pendekatan berbeda untuk mengatasi masalah ini, di mana generative model cenderung lebih fokus pada pemahaman distribusi data dan ketidakpastian, sementara discriminative model lebih fokus pada memodelkan keputusan langsung antara kelas atau label.

Weak supervision adalah pendekatan dalam pembelajaran mesin di mana label data yang digunakan untuk melatih model tidak selalu bersifat sepenuhnya akurat atau lengkap. Pendekatan ini memanfaatkan label data yang diberikan secara heuristik, noisy, atau hanya sebagian. Berikut adalah tahap-tahap yang biasanya dilalui dalam weak supervision:

1. Pengumpulan Data dengan Label Heuristik (Heuristic Labeling)

Pada tahap ini, label data diberikan secara heuristik atau berdasarkan aturan-aturan sederhana. Ini bisa dilakukan secara otomatis atau manual. Contohnya, dalam analisis sentimen, dapat digunakan kata-kata kunci tertentu untuk memberikan label positif atau negatif.

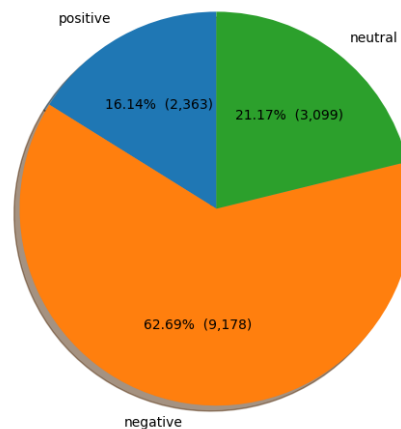


Gambar 2. Arsitektur Analisis Sentimen

2. Labeling Noisy atau Ambigu (Noisy or Ambiguous Labeling).
 Label data yang dihasilkan melalui heuristik mungkin noisy atau ambigu. Tahap ini melibatkan pemahaman bahwa label-label ini mungkin tidak sepenuhnya akurat atau pasti. Pengolahan lebih lanjut mungkin diperlukan untuk mengatasi ketidakpastian dan kebisingan dalam label.
 3. Pemrosesan Data (Data Processing).
 Data dengan label heuristik atau noisy kemudian diproses untuk mempersiapkannya agar dapat digunakan dalam pelatihan model. Ini mungkin melibatkan langkah-langkah seperti pembersihan data, tokenisasi, dan ekstraksi fitur.
 4. Pembuatan Model dengan Label Lemah (Model Training with Weak Labels).
 Model mesin dilatih menggunakan data yang diberi label heuristik atau noisy. Algoritma pembelajaran mesin diadaptasi untuk mengatasi ketidakpastian dalam label data. Model ini mencoba untuk menemukan pola atau representasi yang dapat mewakili informasi yang bermanfaat meskipun ketidakpastian dalam label.
 5. Validasi dan Pemantauan Model (Model Validation and Monitoring).
 Model yang dihasilkan kemudian dievaluasi dan divalidasi menggunakan data yang memiliki label yang lebih andal. Pemantauan terus-menerus diperlukan untuk memastikan bahwa model berkinerja dengan baik dan mampu mengatasi ketidakpastian dalam label data.
 6. Iterasi dan Peningkatan (Iteration and Improvement).
 Proses ini sering kali melibatkan iterasi berulang. Model dapat diperbarui atau disesuaikan dengan menggunakan informasi tambahan atau dengan meningkatkan heuristik atau aturan yang digunakan dalam pemberian label.
- Weak supervision memberikan fleksibilitas dalam mengatasi keterbatasan dalam label data, yang seringkali sulit atau mahal untuk diperoleh secara akurat. Meskipun demikian, pendekatan ini membutuhkan perhatian khusus terhadap penanganan ketidakpastian dan kebisingan dalam data untuk memastikan kualitas dan keandalan model yang dihasilkan.

HASIL DAN PEMBAHASAN

Pada Gambar 3 terdapat kumpulan data Twitter mengenai penerbangan menunjukkan bahwa mayoritas tweet, sebanyak 62.69%, dapat dikategorikan sebagai netral, yang mencerminkan adanya persepsi umum yang tidak memiliki sentimen positif atau negatif yang dominan terhadap topik tersebut. Di sisi lain, tweet dengan sentimen negatif mencapai 21.17%, menunjukkan bahwa sebagian warganet memiliki pandangan kritis atau merasa tidak puas terhadap aspek-aspek tertentu terkait penerbangan. Sementara itu, tweet dengan sentimen positif sebanyak 16.14% menunjukkan bahwa sebagian masyarakat merasa senang atau puas dengan pengalaman penerbangan. Analisis sentimen ini dapat memberikan wawasan berharga kepada perusahaan penerbangan atau peneliti untuk memahami opini dan respon masyarakat terhadap layanan penerbangan serta memungkinkan mereka untuk meningkatkan kualitas layanan atau merespons masalah yang mungkin muncul.



Gambar 3. Spesifikasi data

Data yang diberikan mencakup informasi tentang penggunaan Labelling Function menggunakan Classifier Naïve Bayes, dengan evaluasi terhadap Label Model menggunakan sejumlah metrik pada set data pengembangan (dev). Metrik evaluasi tersebut mencakup akurasi (accuracy), recall, precision, dan f1 score. Pertama-tama, Labelling Function menggunakan Classifier Naïve Bayes memiliki performa yang dievaluasi dengan baik pada set data pengembangan, dengan tingkat akurasi sebesar 0.65. Akurasi mengukur sejauh mana model dapat mengklasifikasikan data dengan benar secara keseluruhan. Kemudian, evaluasi dilanjutkan dengan melihat recall, yang mencapai nilai sebesar 0.75. Recall mengukur kemampuan model untuk mengidentifikasi keseluruhan instans dari suatu kelas. Nilai yang tinggi pada recall menunjukkan bahwa model cenderung dapat menangkap sebagian besar instance positif yang sebenarnya. Selanjutnya, precision, yang memiliki nilai 0.55, mengukur sejauh mana instans yang diklasifikasikan sebagai positif oleh model benar-benar positif. Precision yang relatif rendah mungkin menunjukkan adanya kecenderungan model untuk memberikan label positif pada instans yang sebenarnya negatif. Terakhir, f1 score, dengan nilai 0.63, adalah gabungan dari recall dan precision. F1 score memberikan gambaran keseluruhan tentang keseimbangan antara kemampuan model untuk mengidentifikasi positif dan menghindari memberikan label positif secara keliru. Secara keseluruhan, hasil evaluasi terhadap Label Model menunjukkan kinerja yang memadai, dengan beberapa catatan terkait precision yang dapat menjadi fokus perbaikan atau peningkatan model untuk tugas klasifikasi yang lebih baik. Berikut ini adalah deskripsi dari hasil uji model Weak Supervision yang didapatkan:

Metode : Labelling Function
 Classifier : Naïve Bayes
 Accuracy : 0.65

Recall : 0.75
 Precision : 0.5454545454545454
 f1 score : 0.631578947368421

Label Model yang didapatkan sebelumnya digunakan untuk melakukan pelabelan pada keseluruhan data latih. Setelah data latih mendapat label, hal ini perlu dilakukan evaluasi untuk mengukur kinerja dari label model tersebut. Data yang diberikan mencakup hasil evaluasi dari penerapan Label Model menggunakan Classifier Naïve Bayes pada keseluruhan dataset. Evaluasi tersebut mencakup metrik akurasi (accuracy), recall, precision, dan f1 score.

Pertama-tama, akurasi dari model tersebut mencapai nilai sebesar 0.56. Akurasi mengukur sejauh mana model dapat mengklasifikasikan keseluruhan dataset dengan benar. Nilai akurasi yang relatif rendah dapat menunjukkan bahwa model mungkin mengalami kesulitan dalam melakukan klasifikasi yang tepat. Selanjutnya, recall memiliki nilai sebesar 0.71. Recall mengukur kemampuan model untuk mengidentifikasi keseluruhan instans dari suatu kelas tertentu. Nilai recall yang tinggi menandakan bahwa model cenderung dapat menangkap sebagian besar instance positif yang sebenarnya. Kemudian, precision dari model ini mencapai 0.44. Precision mengukur sejauh mana instans yang diklasifikasikan sebagai positif oleh model benar-benar positif. Nilai precision yang relatif rendah mungkin menunjukkan kecenderungan model memberikan label positif pada instans yang sebenarnya negatif. Terakhir, f1 score, dengan nilai 0.54, adalah gabungan dari recall dan precision. F1 score memberikan gambaran keseluruhan tentang keseimbangan antara kemampuan model untuk mengidentifikasi positif dan menghindari memberikan label positif secara keliru. Secara keseluruhan, hasil evaluasi menunjukkan kinerja yang dapat ditingkatkan, terutama dalam hal akurasi dan precision. Peningkatan model mungkin diperlukan untuk meningkatkan kemampuannya dalam mengklasifikasikan dataset secara lebih tepat. Berikut ini adalah deskripsi dari evaluasi hasil pelabelan Weak Supervision pada dataset yang ada:

Classifier	: Naïve Bayes
Accuracy	: 0.5638148667601683
Recall	: 0.7076923076923077
Precision	: 0.43914081145584727
f1 score	: 0.5419734904270987

KESIMPULAN

Weak supervision memberikan fleksibilitas dalam mengatasi keterbatasan dalam label data, yang seringkali sulit atau mahal untuk diperoleh secara akurat. Meskipun demikian, pendekatan ini membutuhkan perhatian khusus terhadap penanganan ketidakpastian dan kebisingan dalam data untuk memastikan kualitas dan keandalan model yang dihasilkan. Labeling Functions (LFs) adalah aturan atau fungsi yang menghasilkan label untuk data tanpa memerlukan annotator manusia. LFs bisa berupa aturan heuristik, pencocokan pola, atau metode lainnya untuk memberikan label. Namun, LFs mungkin menghasilkan label yang noisy atau tidak sepenuhnya akurat. Labeling function pada weak supervision memiliki keunggulan dalam meningkatkan skalabilitas dan fleksibilitas dalam pengumpulan label data. Keunggulan pertama terletak pada kemampuan labeling function untuk dengan cepat memberikan label pada data tanpa perlu keterlibatan annotator manusia secara manual. Dengan demikian, proses tersebut menjadi lebih efisien, terutama ketika data yang besar dan kompleks harus diatasi. Keunggulan kedua adalah fleksibilitas, di mana labeling function dapat memanfaatkan berbagai sumber informasi atau aturan heuristik yang mencerminkan pengetahuan domain ahli. Fleksibilitas ini memungkinkan integrasi pengetahuan yang mendalam ke dalam proses pembuatan label, yang mungkin sulit dicapai dalam situasi pengumpulan label manual. Namun, labeling function juga memiliki beberapa kelemahan yang perlu diperhatikan. Salah satu kelemahan utama adalah potensi noisy atau tidak konsistennya label yang dihasilkan. Labeling function dapat tergantung pada aturan atau sumber informasi yang tidak selalu akurat, sehingga label yang dihasilkan mungkin kurang dapat diandalkan. Selain itu, kurangnya penanganan ketidakpastian pada tingkat keakuratan label juga menjadi kelemahan. Hal ini membuat sulit untuk mengevaluasi sejauh mana kita dapat mempercayai label yang diberikan oleh labeling function. Oleh karena itu, perlu hati-hati dalam mengelola kelemahan ini dan mempertimbangkan solusi atau metode tambahan untuk meningkatkan keakuratan dan konsistensi labeling function dalam konteks weak supervision.

DAFTAR PUSTAKA

- [1] M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowledge-Based Syst.*, vol. 226, p. 107134, 2021, doi: 10.1016/j.knosys.2021.107134.
- [2] S. Sivakumar and R. Rajalakshmi, "Analysis of sentiment on movie reviews using word embedding self-attentive LSTM," *Int. J. Ambient Comput. Intell.*, vol. 12, no. 2, pp. 33–52, 2021, doi: 10.4018/IJACI.2021040103.
- [3] L. M. Chen, B. X. Xiu, and Z. Y. Ding, "Multiple weak supervision for short text classification," *Appl. Intell.*, 2022, doi: 10.1007/s10489-021-02958-3.
- [4] J. Zdenek and H. Nakayama, "Erasing scene text with weak supervision," *Proc. - 2020 IEEE Winter Conf. Appl. Comput. Vision, WACV 2020*, pp. 2227–2235, 2020, doi: 10.1109/WACV45572.2020.9093544.
- [5] R. K. Mundotiya and N. Yadav, "Forward context-aware clickbait tweet identification system," *Int. J. Ambient Comput. Intell.*, vol. 12, no. 2, pp. 21–32, 2021, doi: 10.4018/IJACI.2021040102.
- [6] N. Rao, C. Bansal, and J. Guan, "Search4Code: Code search intent classification using weak supervision," *Proc. - 2021 IEEE/ACM 18th Int. Conf. Min. Softw. Repos. MSR 2021*, pp. 575–579, 2021, doi: 10.1109/MSR52588.2021.00077.
- [7] Y. Sun, T. Mallick, P. Balaprakash, and J. Macfarlane, "A data-centric weak supervised learning for highway traffic incident detection," *Accid. Anal. Prev.*, vol. 176, no. January, p. 106779, 2022, doi: 10.1016/j.aap.2022.106779.