

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pendidikan merupakan salah satu mekanisme utama untuk membekali individu dengan pengetahuan, keterampilan, dan kompetensi yang diperlukan dalam kehidupan kerja. Agar investasi pendidikan memberikan manfaat optimal, kompetensi yang dibentuk melalui kurikulum perlu memiliki keterhubungan dengan kebutuhan pekerjaan dan perkembangan pasar kerja. Keterhubungan tersebut dikenal melalui prinsip *link and match*, yaitu upaya menyelaraskan proses pendidikan dengan jenis kompetensi yang dibutuhkan dunia kerja. Ketika penyelarasan tidak tercapai, lulusan dapat memasuki pekerjaan yang tidak sesuai dengan jurusan atau bidang keahliannya. Dalam literatur ketenagakerjaan, kondisi ini disebut *horizontal mismatch*, yaitu ketidakcocokan antara latar belakang bidang pendidikan atau keahlian dengan bidang pekerjaan yang dilakukan. *Mismatch* dapat menyebabkan kompetensi hasil pendidikan tidak termanfaatkan secara optimal, sementara kebutuhan kompetensi pekerjaan juga tidak sepenuhnya terpenuhi. Kondisi tersebut dapat menimbulkan inefisiensi bagi individu, organisasi, dan sistem pasar kerja [1], [2]. Konsep ini berkaitan dengan *skills mismatch*, yaitu keadaan ketika keterampilan pekerja tidak sesuai dengan keterampilan yang dibutuhkan pekerjaan sehingga dapat memengaruhi produktivitas, kepuasan kerja, dan kualitas pemanfaatan tenaga kerja [3], [4], [5].

Secara makro, isu *mismatch* perlu ditempatkan dalam kondisi pasar kerja Indonesia terkini. Badan Pusat Statistik mencatat bahwa pada Februari 2026 jumlah angkatan kerja mencapai 154,91 juta orang, penduduk bekerja 147,67 juta orang, Tingkat Partisipasi Angkatan Kerja 70,56%, dan Tingkat Pengangguran Terbuka 4,68%. Rata-rata upah buruh tercatat Rp3,29 juta per bulan [6]. Besarnya jumlah penduduk yang masuk dan bertahan di pasar kerja menunjukkan bahwa kualitas kecocokan antara pendidikan, keterampilan, dan pekerjaan sama pentingnya dengan penyerapan tenaga kerja secara kuantitatif.

Data tersebut tidak digunakan untuk menyatakan bahwa seluruh pengangguran disebabkan *mismatch*. Data BPS dipakai sebagai konteks skala pasar kerja: pada jumlah angkatan kerja yang sangat besar, ketidaktepatan pemanfaatan kompetensi berpotensi memengaruhi banyak individu, organisasi, dan kebijakan pendidikan. Penelitian ini kemudian melengkapi statistik resmi dengan pembacaan opini publik digital mengenai pengalaman kerja tidak sesuai jurusan.

Tabel 1.1 Ringkasan indikator pasar kerja Indonesia Februari 2026

Indikator	Nilai	Kegunaan dalam penelitian
Angkatan kerja	154,91 juta orang	Menunjukkan skala populasi yang memasuki pasar kerja.
Penduduk bekerja	147,67 juta orang	Menunjukkan besarnya populasi yang mengalami proses pencocokan kompetensi dan pekerjaan.
Tingkat Partisipasi Angkatan Kerja	70,56%	Memberikan konteks keterlibatan penduduk usia kerja dalam pasar kerja.
Tingkat Pengangguran Terbuka	4,68%	Digunakan sebagai konteks kondisi pasar kerja, bukan ukuran mismatch.
Rata-rata upah buruh	Rp3,29 juta per bulan	Memberikan konteks hasil ekonomi pekerja secara agregat.

Sumber: [6].

Tabel 1.1 menempatkan fenomena ketidaksesuaian pendidikan-pekerjaan dalam skala pasar kerja nasional. Indikator tersebut tidak digunakan untuk mengestimasi prevalensi *mismatch*, tetapi untuk menunjukkan bahwa kualitas pencocokan kompetensi dan pekerjaan perlu dibaca bersama besarnya partisipasi dan penyerapan tenaga kerja. Analisis sentimen pada penelitian ini berfungsi sebagai pelengkap untuk membaca narasi pengalaman publik digital, bukan sebagai pengganti statistik resmi.

Pada kelompok pemuda, urgensi *mismatch* lebih kuat karena terjadi pada tahap awal pembentukan lintasan karier. Kajian BPS mengenai *mismatch*

pendidikan-pekerjaan pemuda menunjukkan bahwa pengangguran pemuda berada sekitar dua kali tingkat nasional, lebih dari sepertiga pekerja muda mengalami ketidaksesuaian vertikal, dan *mismatch* dapat berkaitan dengan penalti upah serta inefisiensi ekonomi [7]. Temuan tersebut menunjukkan bahwa transisi pendidikan-ke-kerja juga merupakan kecocokan kompetensi dan posisi kerja.

Konteks pemuda penting bagi bonus demografi. Ketika pendidikan dan pekerjaan tidak selaras, investasi pendidikan berisiko tidak dimanfaatkan secara optimal, pekerja membutuhkan adaptasi atau *reskilling* tambahan, dan pemberi kerja menanggung biaya penyesuaian kompetensi. Oleh karena itu, opini publik mengenai kerja lintas jurusan relevan dianalisis sebagai pelengkap indikator statistik pasar kerja.

Topik ketidaksesuaian jurusan pendidikan dan pekerjaan dipilih karena memiliki dasar empiris yang terukur, dampak yang luas, serta masih menyisakan ruang penelitian dari sisi persepsi publik. Dari sisi empiris, penelitian Yonanda dan Usman (2023) menggunakan data Sakernas Februari 2022 dan menemukan bahwa 33,50% pekerja lulusan pendidikan tinggi di Indonesia mengalami *horizontal mismatch*. Penelitian tersebut juga menunjukkan bahwa karakteristik seperti rumpun bidang studi, keikutsertaan pelatihan, dan pengalaman bekerja sebelum lulus berkaitan dengan status *horizontal mismatch* [1]. Temuan ini memperlihatkan bahwa *mismatch* berkaitan dengan faktor pendidikan, pengalaman, dan kesiapan kerja.

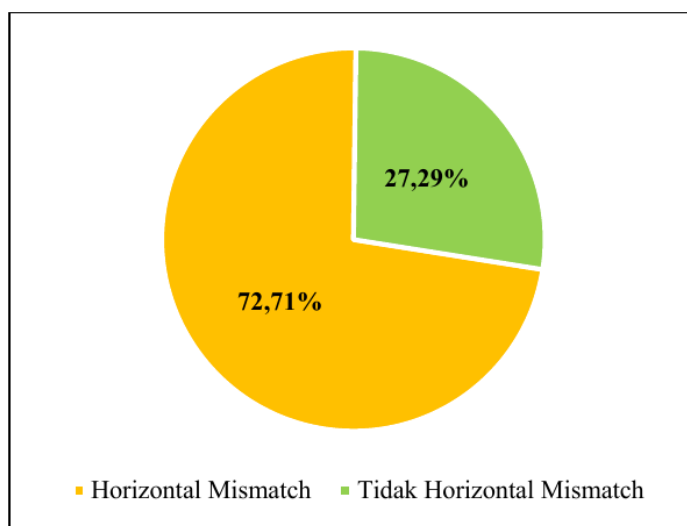
Agama	56,73%	43,27%
Humaniora	57,55%	42,45%
Sosial	70,04%	29,96%
Alam	55,94%	44,06%
Formal	61,60%	38,40%
Karakteristik	<i>Match</i>	<i>Horizontal Mismatch</i>
Keikutsertaan Pelatihan		
Pernah	73,47%	26,53%
Tidak Pernah	58,81%	41,19%
Pengalaman Bekerja Sebelum Lulus		
Bekerja Sebelum Lulus	71,65%	28,35%
Bekerja Setelah Lulus	63,75%	36,25%
Total	66,50%	33,50%

Gambar 1.1 Status horizontal mismatch pekerja lulusan pendidikan tinggi

Sumber: [1]

Berdasarkan Gambar 1.1, data asli Yonanda dan Usman menunjukkan bahwa 66,50% pekerja lulusan pendidikan tinggi berada pada kondisi *match*, sedangkan 33,50% lainnya mengalami *horizontal mismatch*. Angka ini berarti sekitar satu dari tiga pekerja lulusan pendidikan tinggi bekerja pada bidang yang tidak sesuai dengan latar belakang pendidikannya. Gambar tersebut juga memperlihatkan bahwa karakteristik seperti rumpun bidang studi, keikutsertaan pelatihan, serta pengalaman bekerja sebelum lulus memiliki perbedaan persentase *match* dan *mismatch*. Dengan demikian, data ini memberi gambaran bahwa *mismatch* berkaitan dengan karakteristik pendidikan dan pengalaman individu.

Bukti empiris lain terlihat pada lulusan Sekolah Menengah Kejuruan (SMK). Setiyana dan Oktora (2024) menggunakan data Sakernas Agustus 2022 dengan amatan 42.430 pekerja lulusan SMK dan menemukan bahwa 72,71% pekerja lulusan SMK mengalami *horizontal mismatch*, sedangkan 27,29% tidak mengalami *horizontal mismatch* [8]. Angka tersebut menunjukkan adanya ketidaksinkronan yang besar antara kompetensi vokasional yang dirancang untuk menyiapkan lulusan memasuki dunia kerja dengan realitas bidang pekerjaan yang dijalani. Kondisi ini memperkuat alasan mengapa *mismatch* perlu diteliti, karena pendidikan vokasi idealnya menjadi jalur yang paling dekat dengan kebutuhan dunia kerja.

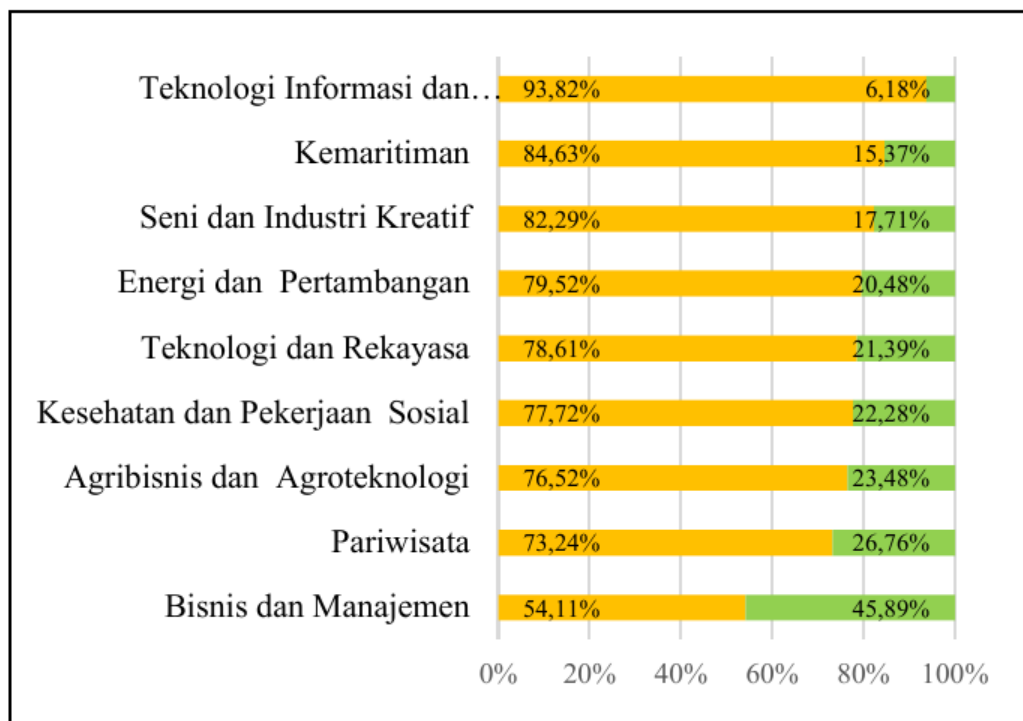


Gambar 1.2 Persentase tenaga kerja lulusan SMK berdasarkan status horizontal mismatch

Sumber: [8]

Gambar 1.2 memperlihatkan bahwa 72,71% pekerja lulusan SMK mengalami *horizontal mismatch*, sedangkan hanya 27,29% yang tidak mengalami *horizontal mismatch*. Proporsi tersebut menunjukkan bahwa masalah ketidaksesuaian bidang pendidikan dan pekerjaan pada lulusan vokasi jauh lebih dominan dibandingkan kondisi yang sesuai. SMK secara konseptual dirancang untuk menyiapkan lulusan yang siap kerja, sehingga tingginya *mismatch* pada kelompok ini mengindikasikan adanya masalah link and match antara pendidikan vokasi dan kebutuhan dunia kerja.

Studi Setiyana dan Oktora juga menunjukkan bahwa *mismatch* tidak tersebar secara merata antarbidang keahlian. Bidang Teknologi Informasi dan Komunikasi memiliki proporsi *horizontal mismatch* tertinggi, yaitu 93,82%, diikuti Kemaritiman sebesar 84,63%, Seni dan Industri Kreatif sebesar 82,29%, Energi dan Pertambangan sebesar 79,52%, serta Teknologi dan Rekayasa sebesar 78,61% [8]. Data ini penting karena memperlihatkan bahwa mismatch berkaitan antara bidang keahlian tertentu dan struktur kesempatan kerja yang tersedia.



Gambar 1.3 Persentase status horizontal mismatch lulusan SMK berdasarkan bidang keahlian

Sumber: [8]

Berdasarkan Gambar 1.3, tingkat *horizontal mismatch* lulusan SMK berbeda antarbidang keahlian. Bidang Teknologi Informasi dan Komunikasi mencatat *mismatch* tertinggi sebesar 93,82%, diikuti Kemaritiman sebesar 84,63%, Seni dan Industri Kreatif sebesar 82,29%, Energi dan Pertambangan sebesar 79,52%, Teknologi dan Rekayasa sebesar 78,61%, Kesehatan dan Pekerjaan Sosial sebesar 77,72%, Agribisnis dan Agroteknologi sebesar 76,52%, Pariwisata sebesar 73,24%, serta Bisnis dan Manajemen sebesar 54,11%. Perbedaan ini menunjukkan bahwa *mismatch* tidak terjadi secara seragam, melainkan dipengaruhi oleh karakter bidang keahlian dan peluang kerja yang tersedia. Oleh karena itu, penelitian ini perlu melihat *mismatch* sebagai isu yang memiliki variasi pengalaman dan persepsi publik, bukan hanya sebagai satu angka agregat.

Tabel 1.2 Perbandingan data asli horizontal mismatch

Sumber data	Kelompok pekerja	Temuan utama	Makna bagi penelitian
Yonanda & Usman (2023) [1]	Pekerja lulusan pendidikan tinggi, Sakernas Februari 2022	33,50% mengalami <i>horizontal mismatch</i> dan 66,50% <i>match</i>	<i>Mismatch</i> juga terjadi pada kelompok terdidik sehingga bukan hanya persoalan lulusan tertentu.
Setiyana & Oktora (2024) [8]	42.430 pekerja lulusan SMK, Sakernas Agustus 2022	72,71% mengalami <i>horizontal mismatch</i> dan 27,29% tidak <i>horizontal mismatch</i>	<i>Mismatch</i> pada lulusan vokasi lebih tinggi sehingga menunjukkan persoalan <i>link and match</i> yang kuat.
Setiyana & Oktora (2024) [8]	Lulusan SMK menurut bidang keahlian	TIK memiliki <i>horizontal mismatch</i> tertinggi sebesar 93,82%	Bidang keahlian tertentu menghadapi ketidaksesuaian yang lebih tajam sehingga

			isu <i>mismatch</i> bersifat struktural.
--	--	--	--

Tabel 1.2 menegaskan perbandingan utama antara dua kelompok pendidikan yang sering menjadi perhatian dalam isu *mismatch*. Pada lulusan pendidikan tinggi, *horizontal mismatch* sebesar 33,50% menunjukkan bahwa masalah ini tetap signifikan pada kelompok terdidik. Pada lulusan SMK, angka 72,71% menunjukkan persoalan yang lebih tajam pada jalur vokasi. Sementara itu, data bidang keahlian SMK menunjukkan bahwa *mismatch* dapat mencapai 93,82% pada bidang tertentu. Perbandingan ini memperkuat alasan pemilihan topik karena *mismatch* terbukti memiliki skala berbeda antarjenjang dan bidang pendidikan, sehingga layak diteliti dari perspektif opini publik.

Selain itu, kajian kebijakan dan perencanaan pembangunan juga menegaskan bahwa *mismatch* merupakan celah antara pendidikan dan kebutuhan pasar kerja. Putranto dkk. (2024) menjelaskan bahwa penyelarasan pendidikan dan jenis pekerjaan diharapkan menjadi solusi yang menguntungkan sisi penawaran dan permintaan tenaga kerja, tetapi Indonesia masih menghadapi *vertical mismatch* yang signifikan terutama pada kelompok muda [9]. LPEM juga menyoroti adanya kesenjangan keterampilan dan peran pendidikan/pelatihan vokasi sebagai salah satu instrumen untuk mengurangi *mismatch* [10]. Dengan demikian, penyebab *mismatch* tidak dapat dilepaskan dari hubungan antara kurikulum, layanan karier, pelatihan, pendidikan vokasi, dan kebutuhan industri.

Secara sebab-akibat, *mismatch* dapat dipahami sebagai hasil dari beberapa faktor yang saling berkaitan. Pertama, faktor individu dan pendidikan, seperti pilihan bidang studi, rumpun ilmu, pengalaman kerja sebelum lulus, pelatihan, dan akses informasi karier [1]. Kedua, faktor kelembagaan dan *link and match*, yaitu belum optimalnya penyelarasan antara kurikulum pendidikan, pendidikan vokasi, dan kebutuhan pasar kerja [9], [10]. Ketiga, faktor dinamika pasar kerja, yaitu terbatasnya kesempatan kerja yang benar-benar sesuai kompetensi, tekanan pengangguran, dan fleksibilitas pasar kerja yang mendorong seseorang menerima pekerjaan di luar bidang pendidikannya [7], [8]. Keempat, faktor wilayah dan

struktur ekonomi yang membuat peluang bekerja sesuai bidang pendidikan berbeda antarwilayah dan sektor pekerjaan [11].

Pemetaan penyebab *mismatch* dalam penelitian ini dibagi menjadi empat kelompok, yaitu faktor individu, faktor lembaga pendidikan, faktor organisasi/industri, dan faktor pasar kerja. Faktor individu mencakup pilihan jurusan, kesiapan skill, pengalaman magang atau kerja sebelum lulus, serta akses informasi karier. Faktor lembaga pendidikan mencakup keselarasan kurikulum, layanan *career center*, dan kemitraan industri. Faktor organisasi mencakup kebutuhan rekrutmen yang menekankan *skill* tertentu, biaya pelatihan ulang, dan fleksibilitas penempatan kerja. Faktor pasar kerja mencakup ketersediaan lowongan yang tidak selalu sejalan dengan bidang studi, tekanan pengangguran, dan perubahan kebutuhan kompetensi. Pemetaan ini digunakan agar pembahasan *mismatch* menunjukkan hubungan antara individu, organisasi, dan struktur pasar kerja.

Tabel 1.3 Level, penyebab, dan pemetaan dampak

Level	Penyebab/indikasi	Dampak yang dipetakan
Individu	Jurusan dan pekerjaan tidak linear, <i>skill</i> tidak terpakai, pengalaman kerja terbatas	Kebingungan karier, kepuasan kerja menurun, kebutuhan <i>reskilling</i> , potensi penalti upah
Organisasi/industri	Kebutuhan kompetensi tidak selalu sama dengan lulusan yang tersedia	Biaya adaptasi dan pelatihan meningkat, produktivitas awal belum optimal, rekrutmen makin berbasis <i>skill</i>
Pasar kerja	Kesempatan kerja sesuai bidang terbatas dan perubahan kebutuhan <i>skill</i>	Inefisiensi alokasi tenaga kerja, transisi pendidikan ke kerja lebih panjang, bonus demografi kurang optimal
Ruang digital/X	Pengalaman <i>mismatch</i> diungkapkan sebagai keluhan, pertanyaan, atau cerita adaptasi	X relevan sebagai pelengkap data resmi karena menangkap narasi aktual

Akibat dari faktor-faktor tersebut muncul pada beberapa level. Pada level individu, *mismatch* dapat menyebabkan *underutilization of skills*, ketidakpuasan karier, penurunan kepercayaan diri, serta potensi penalti upah atau pendapatan [12], [13]. Pada level organisasi, *mismatch* dapat meningkatkan kebutuhan adaptasi, reskilling, biaya pelatihan tambahan, dan risiko produktivitas yang belum optimal [10]. Pada level makro, *mismatch* dapat menghambat kualitas transisi pendidikan ke kerja, menimbulkan inefisiensi pasar kerja, dan mengurangi optimalisasi bonus demografi, terutama ketika dialami oleh pekerja muda [7], [9]. Temuan lintas konteks juga menunjukkan bahwa vertical dan horizontal mismatch dapat berkaitan dengan kecocokan keterampilan, penalti upah, dan kualitas pemanfaatan tenaga kerja [13], [14].

Mismatch tidak selalu dipahami publik melalui istilah teknis *horizontal mismatch* atau *vertical mismatch*. Dalam percakapan sehari-hari, fenomena ini sering muncul melalui narasi seperti “salah jurusan”, “kerja tidak sesuai jurusan”, “skill tidak terpakai”, atau “akhirnya kerja di luar bidang”. Studi pendidikan kejuruan tentang *students with the wrong major* menunjukkan bahwa ketidaksesuaian jurusan dapat berkaitan dengan konflik akademik, tekanan psikologis, penurunan kepercayaan diri, dan perasaan inferior [15]. Meskipun konteksnya berada pada tahap pendidikan, temuan tersebut relevan sebagai hulu dari pengalaman mismatch ketika individu memasuki dunia kerja.

Kondisi ideal (*das sollen*) dalam sistem pendidikan dan ketenagakerjaan adalah ketika pendidikan membekali kompetensi yang selaras dengan kebutuhan pekerjaan, individu memiliki informasi karier yang memadai untuk memilih jurusan dan keterampilan secara sadar, serta pasar kerja mampu menyerap tenaga kerja sesuai kompetensinya. Kondisi nyata (*das sein*) menunjukkan bahwa mismatch masih tinggi pada pekerja muda, lulusan pendidikan tinggi, dan lulusan SMK [1], [7], [8]. Perbandingan antara *das sollen* dan *das sein* tersebut menunjukkan *gap* utama penelitian, yaitu investasi pendidikan belum sepenuhnya terkonversi menjadi kecocokan kerja yang produktif dan sejahtera.

Tabel 1.4 Pemetaan permasalahan, bukti, dan dampak ketidaksesuaian jurusan dengan pekerjaan

No	Permasalahan	Data/fakta utama	Dampak	Pihak terdampak
1	<i>Horizontal mismatch</i> pada lulusan pendidikan tinggi	33,50% pekerja lulusan pendidikan tinggi mengalami <i>horizontal mismatch</i> [1]	Kompetensi tidak termanfaatkan optimal dan jalur karier dapat melenceng dari bidang studi	<i>Fresh graduate</i> , pekerja terdidik, HR
2	<i>Horizontal mismatch</i> pada lulusan SMK	72,71% pekerja lulusan SMK mengalami <i>horizontal mismatch</i> [8]	Pendidikan vokasi belum sepenuhnya terserap sesuai bidang keahlian	Lulusan SMK, sekolah vokasi, industri
3	<i>Mismatch</i> pemuda dan bonus demografi	Lebih dari sepertiga pekerja muda mengalami <i>mismatch</i> tingkat pendidikan dan berkaitan dengan <i>wage penalties</i> [7]	Risiko inefisiensi ekonomi dan bonus demografi tidak optimal	Pemuda, pemerintah, keluarga
4	Kondisi pasar kerja Indonesia 2026	TPT Februari 2026 sebesar 4,68%, angkatan kerja 154,91 juta; penduduk bekerja 147,67 juta, rata-rata upah buruh Rp3,29 juta [6]	Menunjukkan skala transisi pendidikan-kerja dan kebutuhan melihat kualitas kecocokan kompetensi	Pekerja, pencari kerja, institusi pendidikan, pemerintah

5	Pengangguran dan tekanan transisi kerja	Isu pengangguran lulusan SMK menjadi perhatian kebijakan [16]	Pencari kerja dapat menerima pekerjaan di luar bidang sebagai strategi bertahan	Pencari kerja, lulusan, pembuat kebijakan
6	Percakapan publik di ruang digital	DataReportal mencatat 212 juta pengguna internet, 143 juta identitas pengguna media sosial, dan 25,2 juta audiens iklan <i>X</i> di Indonesia [17]	Keluhan dan pengalaman <i>mismatch</i> menyebar cepat sebagai opini publik	Masyarakat, institusi pendidikan, pembuat kebijakan

Sumber: [1], [7], [8], [9], [10], [16], [17]

Tabel 1.4 menjelaskan hubungan antara data, permasalahan, dampak, dan pihak yang terdampak. Data pada tabel tersebut memperlihatkan bahwa *mismatch* juga memiliki konsekuensi praktis bagi lulusan, institusi pendidikan, industri, dan pemerintah. Dengan memetakan permasalahan berdasarkan bukti dan dampaknya, penelitian ini memiliki dasar yang lebih jelas untuk menunjukkan bahwa analisis sentimen di *X* dapat menjadi pelengkap data resmi dalam memahami bagaimana warganet memaknai ketidaksesuaian jurusan pendidikan dan pekerjaan.

Tabel 1.4 memperjelas bahwa *mismatch* bukan hanya persoalan preferensi karier individual. Permasalahan ini memiliki bukti statistik, berdampak pada individu dan organisasi, serta memiliki implikasi sosial-ekonomi pada level makro. Tabel tersebut juga menjelaskan mengapa media sosial *X* relevan digunakan sebagai sumber data. Ketika pengalaman *mismatch* dibahas melalui narasi publik, data media sosial dapat menjadi pelengkap data survei resmi karena mampu menangkap keluhan, normalisasi, stigma, maupun pengalaman langsung yang muncul secara lebih aktual.

Media sosial *X* dipilih sebagai sumber data karena platform ini menyediakan ruang percakapan publik yang cepat, terbuka, dan aktual. DataReportal melaporkan besarnya ekosistem digital Indonesia, dengan 212 juta pengguna internet, 143 juta identitas pengguna media sosial, dan estimasi 25,2 juta jangkauan audiens iklan *X* [17]. Angka tersebut menunjukkan adanya basis percakapan yang cukup besar untuk memetakan opini publik mengenai isu pendidikan, pekerjaan, karier, dan mismatch secara sosial-komputasional.

Dalam penelitian berbasis *X*, pendekatan *text analytics* dan *sentiment analysis* telah banyak digunakan untuk membaca respons publik terhadap isu sosial, kebijakan, dan layanan. Beberapa studi analisis sentimen pada media sosial *X/Twitter* menggunakan alur *crawling*, *preprocessing*, representasi teks, dan pengujian beberapa algoritma *machine learning* seperti *Support Vector Machine*, *Naïve Bayes*, dan *Logistic Regression* [18], [19], [20]. Penelitian lain juga menunjukkan bahwa perbandingan algoritma tersebut dapat dievaluasi secara terukur menggunakan metrik seperti *accuracy*, *precision*, *recall*, dan *F1-score* [21], [22]. Dengan demikian, pemilihan SVM, *Naïve Bayes*, dan *Logistic Regression* pada penelitian ini memiliki dasar metodologis yang sesuai dengan praktik penelitian analisis sentimen berbasis teks.

Tren tersebut menjelaskan mengapa penelitian ini perlu dilakukan sekarang. Pertama, BPS secara terbaru menyoroti *mismatch* pemuda sebagai isu yang berkaitan dengan bonus demografi, *wage penalties*, dan inefisiensi ekonomi [7]. Kedua, studi empiris menunjukkan angka *horizontal mismatch* yang cukup tinggi pada lulusan pendidikan tinggi dan lulusan SMK [1], [8]. Ketiga, percakapan publik di *X* bergerak cepat sehingga pengalaman dan sentimen yang muncul hari ini dapat berubah atau hilang jika tidak segera terdokumentasi [17]. Dengan kata lain, urgensi penelitian terletak pada kebutuhan untuk menangkap persepsi publik secara aktual sebagai pelengkap data resmi dan studi statistik.

Berdasarkan literatur dan konteks tersebut, *research gap* penelitian ini terletak pada empat hal. Pertama, *gap* objek material, yaitu penelitian *mismatch* di Indonesia lebih banyak menggunakan data survei seperti Sakernas dan berfokus pada pengukuran determinan *mismatch* [1], [8], [9], [11], sedangkan penelitian

yang menempatkan percakapan publik sebagai objek utama masih terbatas. Kedua, *gap* konteks, yaitu data survei memiliki siklus rilis tertentu dan tidak selalu menangkap dinamika opini publik secara *real-time*, sementara *X* menyediakan data percakapan aktual terkait pengalaman kerja dan pendidikan [17]. Ketiga, *gap* metode, yaitu studi *mismatch* umumnya menggunakan pendekatan ekonometrika atau regresi, sedangkan studi analisis sentimen lebih banyak diterapkan pada isu lain seperti layanan publik, aplikasi, produk, kesehatan, atau isu sosial-politik [18], [19], [20], [21], [22]. Keempat, *gap* evaluasi komparatif, yaitu perbandingan SVM, *Naïve Bayes*, dan *Logistic Regression* memang telah digunakan pada topik lain seperti ulasan *marketplace*, perbandingan model sentimen, dan ulasan *LinkedIn* [23], [24], [25], tetapi belum banyak digunakan secara spesifik pada isu ketidaksesuaian jurusan pendidikan dan pekerjaan berbasis data *X* berbahasa Indonesia.

Gap antara penelitian *mismatch* berbasis survei dan penelitian berbasis media sosial terletak pada jenis bukti yang dihasilkan. Survei ketenagakerjaan dapat menunjukkan prevalensi, determinan, dan dampak *mismatch* secara statistik, sedangkan data *X* dapat memperlihatkan bagaimana pengalaman tersebut dikomentari, dinormalisasi, dikeluhkan, atau diperdebatkan oleh warganet dalam percakapan sehari-hari. Oleh karena itu, penelitian ini tidak menggantikan data resmi, tetapi melengkapi data resmi dengan pembacaan opini publik digital pada dataset yang dikumpulkan.

Tabel 1.5 Pemetaan *gap* penelitian

Jenis <i>gap</i>	Kondisi penelitian terdahulu	Kebutuhan penelitian ini
Gap objek material	Studi <i>mismatch</i> lebih banyak berbasis survei tenaga kerja dan data resmi [1], [8], [9], [11]	Menganalisis percakapan publik tentang ketidaksesuaian jurusan pendidikan dan pekerjaan di <i>X</i> .
Gap konteks	Data survei kuat secara statistik, tetapi tidak selalu menangkap opini <i>real-time</i> [17]	Menggunakan data <i>X</i> untuk membaca pengalaman dan sentimen publik secara aktual.

Gap metode	Analisis <i>mismatch</i> umumnya memakai ekonometrika, sedangkan analisis sentimen banyak diterapkan pada isu lain [18], [19], [20], [21], [22]	Menerapkan analisis sentimen pada domain <i>mismatch</i> jurusan dengan pekerjaan.
Gap evaluasi komparatif	SVM, <i>Naïve Bayes</i> , dan <i>Logistic Regression</i> banyak dibandingkan pada topik lain [23], [24], [25]	Membandingkan ketiga algoritma pada data <i>X</i> berbahasa Indonesia tentang <i>mismatch</i> .

Tabel 1.5 memperjelas posisi penelitian ini di antara penelitian terdahulu. Studi sebelumnya sudah banyak membahas *mismatch* melalui survei tenaga kerja, pendekatan statistik, maupun kajian kebijakan, sedangkan studi analisis sentimen lebih banyak digunakan untuk isu sosial, layanan, atau topik publik lain. Celah yang ditampilkan dalam tabel tersebut menunjukkan bahwa penelitian ini diperlukan karena menggabungkan isu ketenagakerjaan dengan analisis opini publik berbasis media sosial. Dengan demikian, penelitian ini tidak mengulang penelitian terdahulu, tetapi mengisi ruang yang belum banyak dikaji, yaitu persepsi warganet terhadap ketidaksesuaian jurusan pendidikan dan pekerjaan di Indonesia.

Dengan demikian, kontribusi penelitian ini berada pada tiga aspek. Pertama, penelitian ini memperluas kajian mismatch dari pengukuran statistik tenaga kerja menuju pemahaman opini publik. Kedua, penelitian ini menghadirkan bukti berbasis data *X* yang aktual untuk membaca pengalaman warganet mengenai ketidaksesuaian jurusan pendidikan dan pekerjaan. Ketiga, penelitian ini menyediakan evaluasi komparatif model *machine learning* klasik yang relatif ringan secara komputasi, yaitu *SVM*, *Naïve Bayes*, dan *Logistic Regression*, sebagai *baseline* analisis sentimen bahasa Indonesia pada domain sosial-ketenagakerjaan.

Penelitian ini diarahkan untuk mendukung pengambilan keputusan pada beberapa level. Pada level individu, hasil penelitian dapat menjadi bahan pertimbangan bagi mahasiswa, calon mahasiswa, *fresh graduate*, dan pencari kerja dalam memilih jurusan, membangun keterampilan tambahan, dan merancang strategi karier. Pada level institusi pendidikan, hasil penelitian dapat menjadi masukan bagi perguruan tinggi, SMK, dan *unit career center* untuk mengevaluasi

kurikulum, memperkuat layanan bimbingan karier, serta memperluas kemitraan industri. Pada level organisasi dan kebijakan, hasil penelitian dapat mendukung kebutuhan *reskilling*, strategi rekrutmen, serta penyusunan intervensi untuk menekan *mismatch* [7], [9], [10].

Penelitian ini dirancang melalui alur pengumpulan data dari *X* menggunakan kata kunci terkait ketidaksesuaian jurusan pendidikan dan pekerjaan, *quality control data*, pra-pemrosesan teks bahasa Indonesia, *ablation preprocessing*, pelabelan sentimen berbasis *lexicon domain-specific*, ekstraksi fitur menggunakan TF-IDF, analisis *class imbalance*, pemodelan dengan SVM, *Naïve Bayes*, dan *Logistic Regression*, serta evaluasi menggunakan *accuracy*, *balanced accuracy*, *precision*, *recall*, *macro F1-score*, *weighted F1-score*, MCC, *Cohen's Kappa*, *confusion matrix*, dan *error analysis*. *Output* penelitian berupa distribusi sentimen, pengaruh *preprocessing* terhadap performa model, karakteristik kesalahan klasifikasi, serta *prototype/local deployment dashboard* sebagai media reporting hasil penelitian.

Selain menjelaskan urgensi fenomena *mismatch*, penelitian ini perlu ditegaskan sebagai penelitian *Natural Language Processing* karena objek data yang dianalisis berbentuk teks tidak terstruktur dari media sosial *X*. Unggahan warganet memuat variasi bahasa Indonesia informal, *slang*, singkatan, negasi, *intensifier*, *hashtag*, *mention*, dan frasa domain seperti “salah jurusan”, “tidak sesuai jurusan”, “jurusan tidak terpakai”, serta “banting setir karier”. Karakteristik tersebut membuat teks tidak dapat langsung dianalisis tanpa tahapan pembersihan, normalisasi, representasi numerik, dan pemodelan klasifikasi.

Dengan demikian, Bab I menegaskan kontribusi Sistem Informasi melalui *text mining* dan *NLP*. Tahapan *preprocessing*, *lexicon-based labeling*, representasi TF-IDF, perbandingan SVM, *Naïve Bayes*, dan *Logistic Regression*, serta analisis kesalahan digunakan untuk mengubah opini publik berbasis teks menjadi informasi sentimen yang terukur. Revisi ini juga menempatkan *preprocessing* sebagai bagian yang diuji melalui *ablation study*.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang, rumusan masalah penelitian ini adalah:

- 1) Bagaimana karakteristik, cakupan, dan kualitas *dataset* unggahan publik berbahasa Indonesia pada media sosial *X* yang membahas ketidaksesuaian jurusan pendidikan dan pekerjaan?
- 2) Bagaimana tahapan pengumpulan, penggabungan *batch scraping*, deduplikasi, pembersihan, *preprocessing*, pelabelan sentimen, ekstraksi fitur, pembagian data, dan penanganan ketidakseimbangan kelas diterapkan untuk menghasilkan *dataset* siap model?
- 3) Bagaimana perbandingan kinerja *Support Vector Machine*, *Naïve Bayes*, dan *Logistic Regression* dalam mengklasifikasikan sentimen positif, negatif, dan netral, serta algoritma mana yang menghasilkan kinerja dan kemampuan generalisasi terbaik?

1.3 Batasan Masalah

Agar penelitian terfokus dan terukur, batasan masalah ditetapkan sebagai berikut:

- 1) Sumber data dibatasi pada unggahan publik media sosial *X* yang tersedia pada *file* hasil *scraping* penelitian.
- 2) Unit analisis utama adalah teks unggahan. Gambar, audio, video, dan isi tautan eksternal tidak dianalisis.
- 3) Data penelitian dibatasi pada unggahan berbahasa Indonesia, variasi *slang* dan singkatan dinormalisasi apabila tersedia pada kamus.
- 4) Topik dibatasi pada ketidaksesuaian jurusan pendidikan dan pekerjaan berdasarkan *query*, *theme_hint*, dan *quality filter* relevansi.
- 5) Periode *dataset* yang lolos *quality filter* adalah 2 Januari 2020 sampai 10 Mei 2026.
- 6) Umur dan lokasi pengguna tidak dianalisis karena metadata tersebut tidak tersedia secara lengkap dan valid untuk menyimpulkan karakteristik demografis atau wilayah pengguna.
- 7) Target klasifikasi terdiri atas tiga kelas: negatif, netral, dan positif. Label final dibentuk menggunakan *lexicon domain-specific* beserta aturan negasi dan *intensifier*.
- 8) Representasi teks dibatasi pada TF-IDF word *unigram-bigram* dan word *unigram-trigram*.

- 9) Algoritma yang dibandingkan adalah varian SVM linear, *Naïve Bayes*, dan *Logistic Regression* dengan *tuning* parameter.
- 10) Evaluasi menggunakan *accuracy*, *balanced accuracy*, *precision*, *recall*, *macro F1*, *weighted F1*, *MCC*, *Cohen's Kappa*, *confusion matrix*, *cross-validation*, gap generalisasi, *multiclass capability check*, *error analysis*, dan *quality gate* internal.

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

Tujuan penelitian ini adalah:

- 1) Mendeskripsikan karakteristik, cakupan, distribusi, dan kualitas *dataset* unggahan publik X berbahasa Indonesia yang membahas ketidaksesuaian jurusan pendidikan dan pekerjaan.
- 2) Menjelaskan dan mengimplementasikan *pipeline* pengumpulan, penggabungan *batch*, deduplikasi, pembersihan, *preprocessing*, pelabelan, TF-IDF, *stratified split*, dan penanganan *class imbalance* hingga terbentuk *dataset* siap model.
- 3) Membandingkan SVM, *Naïve Bayes*, dan *Logistic Regression* melalui *tuning*, *cross-validation*, evaluasi multiclass, generalisasi, dan *quality gate* untuk menentukan model final yang paling dapat dipertanggungjawabkan.

1.4.2 Manfaat Penelitian

- 1) Manfaat Teoretis:
 - a) Memberikan kontribusi pada pengembangan kajian ketenagakerjaan dan *mismatch* melalui pendekatan analitik media sosial, sehingga memperkaya perspektif yang selama ini dominan berbasis survei/ekonometrika.
 - b) Menjadi referensi empiris mengenai penerapan *baseline machine learning* (SVM, *Naïve Bayes*, *Logistic Regression*) untuk analisis sentimen bahasa Indonesia pada domain isu sosial-ketenagakerjaan
 - c) Menjadi rujukan mengenai penerapan *lexicon-based labeling* dan analisis frasa dominan untuk memahami isu sosial-ketenagakerjaan berbasis teks media sosial berbahasa Indonesia.
- 2) Manfaat Praktis:

- a) Bagi pemerintah/pembuat kebijakan: menyediakan masukan berbasis data percakapan publik sebagai pelengkap data resmi untuk merancang intervensi karier, pelatihan, dan penyelarasan pendidikan dengan pasar kerja [7], [9].
- b) Bagi perguruan tinggi/SMK/lembaga pendidikan: menjadi bahan evaluasi isu *mismatch* yang muncul di publik sehingga dapat memperkuat kurikulum, layanan bimbingan karier, dan kemitraan industri.
- c) Bagi dunia industri/HR: membantu memahami persepsi dan keluhan publik terkait *mismatch* serta kebutuhan *reskilling/upskilling* yang sering muncul dalam wacana.
- d) Bagi mahasiswa/pencari kerja: meningkatkan kesadaran terhadap pola pengalaman *mismatch* serta mendorong pengambilan keputusan karier yang lebih berbasis informasi.
- e) Bagi peneliti selanjutnya: menyediakan rujukan alur penelitian, skema pengolahan data, dan hasil evaluasi model sebagai baseline untuk riset lanjutan (misalnya pengembangan model yang lebih advanced atau analisis topik/aspek).

1.5 Sistematika Penulisan

Penulisan skripsi ini disusun secara sistematis ke dalam lima bab sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi latar belakang penelitian yang menjelaskan fenomena dan permasalahan ketidaksesuaian jurusan pendidikan dengan pekerjaan yang tercermin dari data ketenagakerjaan serta percakapan publik di media sosial X. Selain itu, bab ini memuat rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan skripsi sebagai arah dan batas ruang lingkup penelitian.

BAB II LANDASAN TEORI

Bab ini memaparkan teori-teori dan konsep yang menjadi dasar penelitian, meliputi konsep *education-employment mismatch* (*horizontal* dan *vertical mismatch*), konsep analisis sentimen dan *text*

mining pada media sosial, tahapan *Natural Language Processing* (NLP) untuk teks berbahasa Indonesia, teknik representasi teks seperti TF-IDF, serta algoritma klasifikasi yang digunakan yaitu *Support Vector Machine* (SVM), *Naïve Bayes*, dan *Logistic Regression*. Bab ini juga menyajikan penelitian terdahulu yang relevan sebagai landasan pembandingan dan penguatan posisi penelitian.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan metode dan tahapan penelitian yang digunakan, mencakup gambaran umum objek penelitian, metode penelitian, teknik pengumpulan data (data primer hasil *scraping X*), proses pelabelan data berbasis *lexicon domain-specific*, pra-pemrosesan teks, ekstraksi fitur, pembangunan model klasifikasi, analisis frasa dominan, serta metode evaluasi performa model (*confusion matrix*, *accuracy*, *precision*, *recall*, dan *F1-score*).

BAB IV ANALISIS DAN HASIL PENELITIAN

Bab ini memuat hasil implementasi dan analisis penelitian, meliputi deskripsi dataset, hasil *quality control*, *preprocessing*, distribusi sentimen, hasil pelabelan berbasis *lexicon*, hasil pelatihan dan pengujian model SVM, *Naïve Bayes*, dan *Logistic Regression*, perbandingan performa berdasarkan metrik evaluasi, serta interpretasi isu atau frasa dominan yang muncul pada setiap kelas sentimen.

BAB V SIMPULAN DAN SARAN

Bab ini berisi simpulan penelitian berdasarkan hasil analisis dan pembahasan, serta saran yang ditujukan bagi pihak terkait maupun rekomendasi untuk penelitian selanjutnya.

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A