

IMPLEMENTASI BUSINESS INTELLIGENCE UNTUK VISUALISASI DATA INDUSTRI PESERTA PELATIHAN KABUPATEN BEKASI

Shafa Aulia Nadhira, Apriade Voutama

Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang
Jl. HS. Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361., Indonesia
shafanadhira90@gmail.com

ABSTRAK

Salah satu kawasan industri terbesar di Indonesia, Kabupaten Bekasi, sangat penting bagi pertumbuhan sektor industri dan pengembangan sumber daya manusia. Partisipasi industri dalam program tersebut berfungsi sebagai indikator komitmen untuk meningkatkan kompetensi tenaga kerja. Tujuan penelitian ini adalah untuk menggunakan *business intelligence* untuk menganalisis kinerja dan pertumbuhan pelatihan industri di Kabupaten Bekasi antara tahun 2023 dan 2025. Metodologi yang digunakan melibatkan analisis *dataset* open data BekasiKab.go.id, yang menyediakan informasi tentang industri, wilayah, dan durasi pelatihan. Data kemudian diproses menggunakan *Python* dan divisualisasikan menggunakan Looker Studio dalam format *dashboard* interaktif. Hasil analisis data 199 perusahaan industri menunjukkan bahwa persentase partisipasi tertinggi berasal dari wilayah Cikarang Barat, yaitu sebesar 26,1% dari total (52 industri). Pada tahun 2024, terdapat sekitar 75 industri (37,7%), pada tahun 2025, terdapat sekitar 71 industri (35,7%), dan pada tahun 2023, terdapat sekitar 53 industri (26,6%). *Dashboard* interaktif ini telah terbukti efektif dalam memberikan pemangku kepentingan gambaran kuantitatif yang tepat untuk mengoptimalkan sebaran program pelatihan di masa depan.

Kata kunci : *Business Intelligence, Visualisasi Data, Industri, Dashboard, Looker Studio*

1. PENDAHULUAN

Industri sangat bergantung pada sumber daya manusia (SDM) untuk mencapai target organisasi, terutama di tengah tingginya kebutuhan akan tenaga kerja yang adaptif untuk transformasi digital. Untuk meningkatkan kualitas sumber daya manusia, khususnya di Kabupaten Bekasi, program pendidikan dan pelatihan berbasis industri diintegrasikan secara sistematis. Program ini sangat penting untuk mengoptimalkan kemampuan individu, memenuhi kebutuhan operasional, dan memastikan bahwa tenaga kerja tetap bersaing di masa depan [1].

Kabupaten Bekasi, sebagai salah satu pusat perindustrian terbesar di Indonesia, menjadi lokasi strategis yang mendorong pertumbuhan sektor perindustrian. Sejalan dengan SDGs 9 yang menekankan pembangunan infrastruktur yang tangguh, industri yang inklusif dan berkelanjutan, dan dorongan inovasi [2]. Oleh karena itu, data industri peserta pelatihan di Kabupaten Bekasi harus dianalisis menggunakan suatu *platform* yang dapat membantu pengumpulan, analisis dan visualisasi data untuk mengetahui perkembangan partisipasi industri dari waktu ke waktu [3].

Business Intelligence (BI) adalah salah satu alat analitik yang dapat membantu pengelola data dalam proses pengumpulan, pengolahan, analisis, dan visualisasi data guna mendukung pengambilan keputusan [3]. Dalam penelitian ini, BI digunakan untuk mengolah data industri peserta pelatihan di Kabupaten Bekasi untuk memberikan gambaran tentang sebaran dan perkembangan partisipasi industri. Selain itu, BI memungkinkan pembuatan

laporan dan visualisasi dalam bentuk *dashboard* yang lebih mudah dipahami [4].

Dengan menggabungkan konsep *Business Intelligence* (BI) dengan Looker Studio dan *Python*, penelitian ini melacak kinerja dan kemajuan pelatihan industri di Kabupaten Bekasi [5]. *Dataset* penelitian ini bersumber dari portal Open Data Kabupaten Bekasi (<https://opendata.bekasikab.go.id/>). *Python* adalah instrumen utama dalam *data wrangling* karena kemampuan untuk mengotomatisasi pembersihan format data yang tidak konsisten dengan cepat, mudah digunakan, dan berulang. Sementara itu, Looker Studio yang berbasis *cloud* memungkinkan pemangku kepentingan mengakses *dashboard* interaktif melalui peramban *web* yang dimana ini berbeda dengan perangkat lunak BI komersial yang membutuhkan instalasi atau lisensi [6]. Metode integratif ini memungkinkan penelitian untuk memberikan gambaran menyeluruh tentang tren partisipasi dan sebaran wilayah pelatihan yang dapat digunakan sebagai referensi strategis yang kuat untuk pengambilan keputusan.

2. TINJAUAN PUSTAKA

2.1. Sektor Industri dan Pengembangan Sumber Daya Manusia

Sektor industri memainkan peran penting sebagai katalisator utama pertumbuhan ekonomi, baik di tingkat regional maupun nasional, karena entitas bisnis terorganisir yang menggabungkan aktivitas ekonomi untuk menghasilkan barang atau jasa di tempat tertentu [2]. Namun, peningkatan kualitas modal manusia sering kali berkorelasi negatif dengan pertumbuhan fisik industri. Oleh karena itu,

diperlukan tindakan strategis untuk meningkatkan kompetensi tenaga kerja melalui penggabungan program pendidikan dan pelatihan yang direncanakan. Metode ini menjadi prasyarat mutlak untuk meningkatkan produktivitas dan daya saing industri, sekaligus memastikan ketersediaan lapangan kerja didukung oleh kemampuan karyawan yang memadai untuk menyesuaikan diri dengan perubahan kebutuhan pasar [3]. Oleh karena itu, tingkat partisipasi industri dalam berbagai program pelatihan dapat dianggap sebagai indikator penting yang menunjukkan sejauh mana sektor swasta terlibat dalam pengembangan sumber daya manusia yang berkelanjutan [4].

2.1. Visualisasi Data

Dalam dunia bisnis, visualisasi data adalah alat penting yang digunakan untuk mengubah data yang kompleks menjadi informasi yang lebih mudah dipahami melalui sistem visual manusia. Tujuan utamanya adalah menyampaikan pesan secara efektif dan jelas menggunakan representasi grafis seperti diagram dan grafik, yang dapat menunjukkan korelasi antarvariabel dan memberikan penjelasan menyeluruh tentang proses [7]. Berbagai komponen visual ini dapat disesuaikan baik secara fungsional maupun estetika untuk menyajikan hubungan data yang lebih padat, efektif, dan tepat sasaran sesuai dengan jenis informasi yang ingin disampaikan, berkat fleksibilitas format yang tinggi [5].

2.2. Business Intelligence (BI)

Business Intelligence (BI) adalah kombinasi teknik, aplikasi, dan metode untuk meningkatkan keunggulan kompetitif perusahaan melalui pengumpulan, analisis, dan presentasi data bisnis [8]. Secara operasional, BI berfungsi sebagai sistem pendukung keputusan dengan kemampuan untuk mengubah data mentah menjadi informasi yang berguna melalui pengelolaan data yang berfokus pada kinerja organisasi. BI juga membantu perusahaan menemukan peluang baru dan menangani berbagai jenis informasi untuk mendukung keberhasilan strategis dalam jangka panjang [9].

2.3. Dashboard Interaktif

Dashboard interaktif membantu menyajikan data secara real-time dan meningkatkan efisiensi analisis melalui visualisasi mudah dipahami dan wawasan menyeluruh [5]. Karena fiturnya yang dinamis, interaktif, dan antarmuka pengguna yang ramah pengguna, instrumen ini sangat penting bagi berbagai industri untuk memantau kinerja dan membantu pengambilan Keputusan [9]. Untuk memenuhi kebutuhan ini, berbagai *platform* yang dapat disesuaikan seperti Looker Studio, Power BI, Tableau, dan StatPlanet tersedia.

2.4. Python

Python dengan filosofi yang berpusat pada kode yang mudah dibaca, adalah salah satu bahasa *multiparadigm* yang paling populer yang dibuat oleh *compiler* dan *interpreter* tingkat tinggi yang dirancang untuk berbagai tujuan [10]. Ada banyak *library*, *vendor*, *API third party*, dan ekosistem yang didukung oleh *Python* yang dikembangkan oleh pihak ketiga untuk menangani pengolahan dan visualisasi data, serta kemampuan untuk melakukan berbagai tugas, seperti melakukan visualisasi data dan proses pembersihan dan pengolahan data sebelum dianalisis [11].

2.5. Looker Studio

Looker Studio merupakan instrumen BI yang dirancang untuk pengumpulan, analisis, hingga visualisasi data secara komprehensif. *Platform* ini memfasilitasi pembuatan laporan mendalam serta penyajian data *real-time* yang bersumber dari berbagai kanal guna meningkatkan akurasi analisis. Dengan pemanfaatan *Looker Modeling Language* (LookML), pengguna dapat mendefinisikan dimensi serta metrik secara lebih terstruktur. Selain itu, fitur analisis *ad hoc* pada *platform* ini memungkinkan eksekusi *query* tanpa memerlukan penulisan kode SQL yang kompleks. Bagi skala bisnis besar, *dashboard* yang dikembangkan melalui Looker Studio sangat krusial karena mampu mengintegrasikan data lokal dengan layanan Google Cloud secara kohesif [9].

2.6. Penelitian Sebelumnya

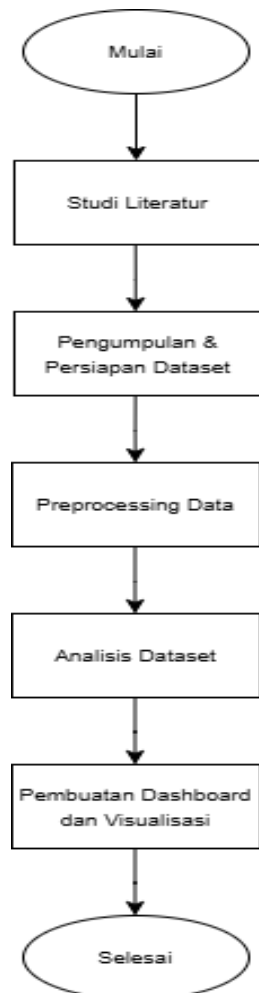
Analisis dan pengambilan keputusan telah banyak dibantu oleh visualisasi data menggunakan BI, salah satunya oleh [6] yang mengumpulkan informasi tentang bencana banjir yang terjadi di Indonesia. Penelitian ini mengolah data sekunder BNPB periode 2020–2025 melalui tahapan *data wrangling*, yang mencakup *gathering*, *assessing*, *cleaning*, dan *exporting*, sebelum disajikan dalam *dashboard* interaktif. Hasilnya menunjukkan bahwa visualisasi efektif dalam menemukan tren dan wilayah dengan frekuensi bencana tertinggi. Ini membantu pemangku kepentingan membuat kesimpulan yang lebih akurat.

Studi tersebut menunjukkan bahwa *dashboard* interaktif dapat membantu memahami data yang kompleks. Oleh karena itu, untuk memvisualisasikan data industri peserta pelatihan di Kabupaten Bekasi, penelitian ini menggunakan metode serupa menggunakan *Python* dan Looker Studio. Metode ini dapat digunakan untuk menganalisis data tentang sebaran industri dan tingkat partisipasi pelatihan secara lebih sistematis.

3. METODE PENELITIAN

Metode deskriptif kuantitatif digunakan dalam penelitian ini untuk mengumpulkan informasi penting dan menyiapkan *dataset* dengan menggunakan

berbagai sumber referensi dan menganalisis data yang relevan. Data industri peserta pelatihan di Kabupaten Bekasi selama periode tahun 2023–2025 dipilih melalui teknik *sampling purposive* selama proses pengumpulan data. Tahapan penelitian ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian

3.1. Studi Literatur

Proses penelitian diawali dengan pelaksanaan studi literatur guna menelaah berbagai referensi terdahulu yang relevan dengan topik kajian. Melalui pengkajian dan analisis mendalam terhadap literatur terkait, peneliti diharapkan memperoleh paradigma baru serta wawasan komprehensif yang berfungsi sebagai landasan teoretis dan acuan dalam menentukan langkah-langkah tindakan selanjutnya. Pendekatan sistematis ini krusial untuk memastikan bahwa penelitian memiliki landasan yang kuat dan terarah.

3.2. Pengumpulan & Persiapan Dataset

Siklus analisis melibatkan pengumpulan data, yang menuntut relevansi, akurasi, dan komprehensivitas untuk memastikan bahwa hasil penelitian valid [12]. Untuk penelitian ini, teknik *crawling* digunakan untuk mengumpulkan data pada

platform resmi <https://opendata.bekasikab.go.id/>, yang membatasi cakupan pada tahun 2023 hingga 2025. Pemilihan rentang waktu ini didasarkan pada data riil bahwa itu memberikan data paling baru yang mungkin untuk mencerminkan kondisi pelatihan industri nyata saat ini dan variabelnya lebih konsisten dibandingkan tahun-tahun sebelumnya. Untuk memastikan bahwa setiap parameter dapat dibandingkan secara objektif selama proses analisis, konsistensi variabel ini sangat penting. Tahap preparasi data adalah yang paling penting karena data dari sumber heterogen seringkali memiliki struktur yang tidak terorganisir atau dikategorikan sebagai *dirty dataset*. Proses ini dilakukan secara menyeluruh untuk meningkatkan kualitas data awal sehingga dapat diproses menjadi informasi yang akurat dan dapat diandalkan bagi para pemangku kepentingan [13].

3.3. Preprocessing Data

Pra-pemrosesan data berarti membersihkan dan mengubah data mentah menjadi format yang sesuai untuk analisis. Untuk meningkatkan manfaat dan kualitas data, penanganan nilai yang kosong, dan penghapusan duplikat diperlukan [14]. Tahap *praprocessing* melibatkan data yang telah diperoleh sebelumnya. Pada tahap ini, eksplorasi data dilakukan untuk mengolah data dari kondisi awal yang tidak memberikan informasi yang jelas sehingga dapat diuraikan secara jelas sehingga pengguna dapat memahaminya [15].

Sebagai bagian dari proses ini, data diproses dengan lebih efektif dan akurat melalui pembersihan, pengisian nilai yang kosong, transformasi, dan penyesuaian tipe data.

3.4. Analisis Dataset

Peneliti melakukan proses analisis *dataset* menggunakan pendekatan BI dengan memanfaatkan *Python* dan Looker Studio. *Dataset* tersebut diolah melalui serangkaian prosedur hingga menjadi bentuk visualisasi data berupa *dashboard* interaktif yang diharapkan dapat digunakan sebagai acuan dalam proses pengambilan keputusan. Proses analisis dimulai dengan memuat data industri peserta pelatihan ke dalam *Python* untuk dilakukan pembersihan dan penyesuaian data sebelum divisualisasikan menggunakan Looker Studio [13].

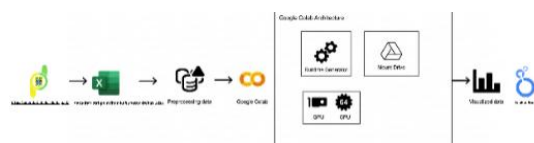
3.5. Pembuatan Dashboard dan Visualisasi

Dataset industri peserta pelatihan divisualisasikan untuk memudahkan pembaca memahaminya. Grafik batang, diagram garis, dan tabel digunakan untuk menunjukkan sebaran dan perkembangan partisipasi industri. Untuk membuat data lebih terorganisir dan mudah dipahami oleh pengguna, langkah selanjutnya adalah membuat *dashboard* interaktif menggunakan Looker Studio. *Dashboard* interaktif adalah alat visual yang memungkinkan eksplorasi data secara dinamis

dengan menampilkan data dalam format yang mudah digunakan dan dapat disesuaikan. *Dashboard* ini, dengan penyajian yang lebih sistematis, membantu proses pengambilan keputusan dengan memberikan informasi yang relevan dan informatif [9].

4. HASIL DAN PEMBAHASAN

Alur pengolahan data hingga visualisasi *dashboard* digambarkan pada Gambar 2. Tahap *import, preprocessing* dengan *Python*, dan visualisasi melalui *Looker Studio* adalah semua komponennya. Sebelum diintegrasikan, proses ini membersihkan data, menyesuaikan formatnya, dan menangani nilai yang tidak ada. Untuk mempermudah analisis dan pengambilan keputusan, distribusi peserta berdasarkan wilayah dan kategori dapat disajikan secara interaktif dengan metode ini [13].



Gambar 2. Prosedur Ekstraksi Dataset

4.1. Persiapan Dataset

Penelitian ini menggunakan *dataset* dari opendata.bekasikab.go.id dalam format *.xlsx* yang memuat data industri peserta program pendidikan dan pelatihan. *Dataset* mencakup atribut nama kabupaten, pelaku usaha atau kegiatan industri, tahun pelaksanaan, kecamatan, desa, serta zona kawasan industri pada periode 2023–2025.

4.2. Preprocessing Data

Untuk memastikan *dataset* bersih dan siap untuk digunakan, fase *preprocessing* dilakukan melalui mekanisme *data wrangling* (*gathering, assessing, cleaning, dan exporting*). Standardisasi ini sangat penting untuk menjamin integritas dan akurasi hasil analisis pada tahap berikutnya [6].

4.3. Gathering Data

Pada Untuk mempermudah pengolahan, *dataset* diambil dari portal Open Data Kabupaten Bekasi dalam format *.xlsx* dan kemudian dikonversi ke *CSV*. Selanjutnya, fungsi `read_csv()` digunakan untuk memuat data ke lingkungan *Python* menggunakan *library* *Pandas*.

```

[0] # Gathering (load data)
✓ 0d import pandas as pd
file_path = '/content/dt_industri_yang_mengikuti_pendidikan_pelatihan.csv'
df = pd.read_csv(file_path)

[1] # Menampilkan 5 data teratas
✓ 0d print("In Data Awal:")
print(df.head())
  
```

Gambar 3. Gathering Data

Dari kode tersebut, akan menghasilkan *output* seperti:

```

***
Data Awal:
   id nomor  kode_kabupaten nama_kabupaten \
0  1      1      3216        BEKASI
1  2      2      3216        BEKASI
2  3      3      3216        BEKASI
3  4      4      3216        BEKASI
4  5      5      3216        BEKASI

   data_pelaku_usaha_dan_atau_kegiatan  tahun  satuan_data  kecamatan \
0  PT. INDOFOOD SUKSES MAKUR  2023  PERUSAHAAN  CIKARANG BARAT
1  PT. HERO SUPERMARKET TBK  2023  PERUSAHAAN  CIKARANG BARAT
2  PT. BUNORI VINA INDONESIA  2023  PERUSAHAAN  CIKARANG PUSAT
3  CV. SALUYU GALIH MULYA  2023  PERUSAHAAN  CIKARANG UTARA
4  CV. SUMBER UTAMA BERKAH  2023  PERUSAHAAN  CIKARANG UTARA

   desa  zona_kawasan
0  SUKADANAU  LUAR KAWASAN
1  SUKADANAU  LUAR KAWASAN
2  PASTIRANTI  GTIC
3  MEKARIUKTI  NON KAWASAN
4  MEKARIUKTI  NON KAWASAN
  
```

Gambar 4. Output Gathering Data

Pada gambar di atas ditampilkan 5 baris pertama *dataset* peserta pelatihan industri di Kabupaten Bekasi. *Dataset* tersebut memuat beberapa atribut, yaitu kode kabupaten, nama kabupaten, data pelaku usaha atau kegiatan, tahun, satuan data, kecamatan, desa, serta zona kawasan.

4.4. Assessing Data

Proses *Assessing Data* pada penelitian ini adalah sebagai berikut:

- Proses *Assessing Data* dilakukan menggunakan fungsi `df.info()` untuk melihat informasi kolom dan tipe *dataset*

```

[6] # Assessing Data
✓ 0d # Melihat informasi kolom dan tipe data
print("Informasi Data")
print(df.info())

***
Informasi Data
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 199 entries, 0 to 198
Data columns (total 10 columns):
 #   Column                                     Non-Null Count  Dtype
---  ---
 0   id                                         199 non-null    int64
 1   nomor                                     199 non-null    int64
 2   kode_kabupaten                           199 non-null    int64
 3   nama_kabupaten                           199 non-null    object
 4   data_pelaku_usaha_dan_atau_kegiatan      199 non-null    object
 5   tahun                                     199 non-null    int64
 6   satuan_data                              199 non-null    object
 7   kecamatan                               196 non-null    object
 8   desa                                     196 non-null    object
 9   zona_kawasan                             196 non-null    object
dtypes: int64(4), object(6)
memory usage: 15.7+ KB
None
  
```

Gambar 5. Struktur DataFrame

Hasil pada Gambar 5 menunjukkan bahwa *dataset* terdiri dari 199 baris dan 10 atribut dengan tipe data *object* dan *int64*. Informasi pada bagian *Non-Null Count* menunjukkan bahwa terdapat beberapa atribut yang memiliki jumlah data kurang dari 199, sehingga mengindikasikan adanya *missing value* pada atribut tersebut.

- Memeriksa jumlah *missing value* disetiap kolom *DataFrame*

```

[7] # Cek Missing Values (Data yang Kosong)
✓ 0d print("\n[1] Jumlah Missing Values per Kolom:")
print(df.isna().sum())
  
```

Gambar 6. Cek Missing Value

Gambar 6 menunjukkan kode yang menunjukkan proses teknis untuk memverifikasi dan mengidentifikasi akumulasi nilai yang hilang (*missing value*) pada setiap kolom *DataFrame*. Implementasi ini mengidentifikasi kolom yang

mengandung data kosong dan kuantitasnya, dengan rincian luaran (*output*) sebagai berikut:

```
***
[1] Jumlah Missing Values per Kolom:
id          0
nomor       0
kode_kabupaten 0
nama_kabupaten 0
data_pelaku_usaha_dan_atau_kegiatan 0
tahun       0
satuan_data 0
kecamatan   3
desa        3
zona_kawasan 3
dtype: int64
```

Gambar 7. Output Missing Value

Gambar 7 menunjukkan hasil perhitungan *missing value* tiap atribut. Terdapat 3 pada kecamatan, 3 pada desa, dan 3 pada zona kawasan.

c. Periksa Duplikat

untuk menghindari perhitungan ganda, menjaga kualitas dan keakuratan data dengan fungsi seperti gambar dibawah ini:

```
[8]
✓ 0d # Cek Data Duplikat
print("\nJumlah Duplikat:", df.duplicated().sum())

***
Jumlah Duplikat: 0
```

Gambar 8. Cek Duplikat

Pada gambar diatas, terlihat bahwa tidak adanya jumlah *duplikat* pada *dataset*. Yang artinya tidak adanya *data* ganda pada *dataset* tersebut.

d. Menampilkan ringkasan statistik pada kolom numerik dalam *dataset*

```
[10]
✓ 0d # Cek Statistik Deskriptif (Untuk kolom angka)
# Ini membantu melihat anomali
print("\nRingkasan Statistik:")
print(df.describe())

***
Ringkasan Statistik:
           id      nomor  kode_kabupaten  tahun
count  199.000000  199.000000         199.0  199.000000
mean    100.000000  100.000000         3216.0  2024.090452
std      57.590508   57.590508           0.0    0.786155
min       1.000000   1.000000         3216.0  2023.000000
25%      50.500000   50.500000         3216.0  2023.000000
50%     100.000000  100.000000         3216.0  2024.000000
75%     149.500000  149.500000         3216.0  2025.000000
max     199.000000  199.000000         3216.0  2025.000000
```

Gambar 9. Cek Statistik Deskriptif

Tujuan dari tahap ini, yang menggunakan fungsi *describe()*, adalah untuk memeriksa distribusi dan kemungkinan anomali pada kolom numerik. Hasil *statistik deskriptif* menunjukkan bahwa kolom *tahun* memiliki jumlah data (*count*) sebanyak 199, yang menunjukkan bahwa seluruh entitas memiliki nilai yang lengkap. Tidak ditemukan nilai anomali untuk atribut tersebut dari tahun 2023 hingga 2025.

e. Cek Keunikan Nilai (*Inconsistency*)

```
[10]
✓ 0d # Cek Keunikan Nilai (Inconsistency)
# Membantu melihat apakah ada nama kecamatan yang typo
print("\nVariasi Nilai di Kolom Kecamatan:")
print(df['kecamatan'].unique())

Variasi Nilai di Kolom Kecamatan:
['CEKARANG BARAT' 'CEKARANG PUSAT' 'CEKARANG UTARA' 'CEKARANG SELATAN'
 'CIBITUNG' 'CIBITUNG' 'TARUMAJAYA' 'SERANG BARU' 'TAMBIN SELATAN'
 'CEKARANG' 'CIBITUNG' 'BOJONGMANGU' 'KEJONGMANGU' 'BABELAN'
 'CEKARANG' 'CIBITUNG' 'KALIJAYA' 'CEKARANG BARAT' 'nan'
 'CEKARANG SELATAN' 'CIBITUNG' 'CEKARANG UTARA' 'BABELAN'
 'CEKARANG TIMUR' 'BABELAN' 'BANDUNG' 'CEKARANG SELATAN'
 'CEKARANG PUSAT' 'TAMBIN SELATAN']
```

Gambar 10. Cek Nilai Unik

Inkonsistensi penulisan seperti kesalahan ketik, spasi berlebihan, dan karakter khusus ditemukan dengan berhasil melalui pemeriksaan menggunakan fungsi *unique()* pada kolom kecamatan. Inkonsistensi ini kemudian diperbaiki pada tahap perbaikan data.

4.5. Cleaning Data

Pada tahap *cleaning data*, dilakukan proses pembersihan *dataset* untuk memperbaiki permasalahan yang ditemukan pada tahap *assessing data*. Proses ini meliputi penanganan *missing value*, penghapusan spasi berlebih, serta perbaikan inkonsistensi penulisan pada beberapa atribut agar data menjadi lebih konsisten dan siap digunakan pada tahap analisis.

a. Menangani Missing Value

Berdasarkan hasil pengecekan sebelumnya, terdapat nilai kosong pada atribut kecamatan, desa, dan zona_kawasan. Nilai kosong tersebut kemudian diisi dengan nilai "Unknown" agar tidak mengganggu proses analisis data. Kode yang digunakan adalah sebagai berikut:

```
[12]
✓ 0d # Mengisi missing value
df['kecamatan'] = df['kecamatan'].fillna('Unknown')
df['desa'] = df['desa'].fillna('Unknown')
df['zona_kawasan'] = df['zona_kawasan'].fillna('Unknown')
```

Gambar 11. Mengisi Missing Value

```
[10]
✓ 0d # Mengecek kembali missing value
print("Jumlah missing value setelah dibersihkan")
df.isnull().sum()

Jumlah missing value setelah dibersihkan
id          0
nomor       0
kode_kabupaten 0
nama_kabupaten 0
data_pelaku_usaha_dan_atau_kegiatan 0
tahun       0
satuan_data 0
kecamatan   0
desa        0
zona_kawasan 0
dtype: int64
```

Gambar 12. Jumlah Missing Value setelah di Cleaning

Gambar di atas menunjukkan proses *cleaning data* untuk menangani *missing value* pada atribut kecamatan, desa, dan zona kawasan. Nilai kosong pada atribut tersebut diisi dengan nilai "Unknown"

agar data tetap lengkap dan dapat digunakan pada tahap analisis.

b. Menghapus Spasi Berlebih

Beberapa data kategori memiliki spasi tambahan di awal atau akhir teks, sehingga perlu dibersihkan agar tidak dianggap sebagai kategori yang berbeda.

```
[13] ✓ Od
# Menghapus spasi berlebih
df['kecamatan'] = df['kecamatan'].str.strip()
df['desa'] = df['desa'].str.strip()
df['zona_kawasan'] = df['zona_kawasan'].str.strip()

[25] ✓ Od
# Mengecek kembali spasi yang berlebihan & typo pada kecamatan
print(df['kecamatan'].unique())

['CIKARANG BARAT' 'CIKARANG PUSAT' 'CIKARANG UTARA' 'CIKARANG SELATAN'
'CIBITUNG' 'CIKARANG TIMUR' 'TARUMAJAYA' 'SERANG BARU' 'TAMBUN SELATAN'
'CIBARUSAH' 'BOJONGMANGU' 'KEDUNGWARINGIN' 'BABELAN' 'CIKARANG\x00 BARAT'
'KALIJAYA' 'Unknown' 'BANDUNG']
```

Gambar 13. Menghapus Spasi Berlebih

Selanjutnya dilakukan pembersihan spasi berlebih pada beberapa atribut bertipe teks menggunakan fungsi `str.strip()`. Langkah ini bertujuan untuk memastikan konsistensi penulisan data sehingga tidak terjadi perbedaan kategori yang disebabkan oleh spasi tambahan.

c. Memperbaiki kesalahan penulisan (Typo)

```
[22] ✓ Od
# Memperbaiki typo pada kecamatan
df['kecamatan'] = df['kecamatan'].replace({
    'CIKARANG BARAT': 'CIKARANG BARAT',
    'CIKARANG SELATAN': 'CIKARANG SELATAN',
    'BABELAB': 'BABELAN'})

[25] ✓ Od
# Mengecek kembali spasi yang berlebihan & typo pada kecamatan
print(df['kecamatan'].unique())

['CIKARANG BARAT' 'CIKARANG PUSAT' 'CIKARANG UTARA' 'CIKARANG SELATAN'
'CIBITUNG' 'CIKARANG TIMUR' 'TARUMAJAYA' 'SERANG BARU' 'TAMBUN SELATAN'
'CIBARUSAH' 'BOJONGMANGU' 'KEDUNGWARINGIN' 'BABELAN' 'CIKARANG\x00 BARAT'
'KALIJAYA' 'Unknown' 'BANDUNG']
```

Gambar 14. Memperbaiki Kesalahan Penulisan

Dari tahap *assessing data* ditemukan beberapa kesalahan penulisan pada atribut kecamatan, seperti:

- CIKARANG BARAT → CIKARANG BARAT
- CIKARNG SELATAN → CIKARANG SELATAN
- BABELAB → BABELAN

Oleh karena itu dilakukan proses koreksi nilai kategori untuk memastikan keseragaman penulisan *nama* kecamatan dalam *dataset*.

d. Memperbaiki karakter spasi khusus

```
[14] ✓ Od
# Memperbaiki karakter spasi khusus
# Mengganti non-breaking space menjadi spasi biasa
df['kecamatan'] = df['kecamatan'].str.replace('\xa0', ' ', regex=True)
df['kecamatan'] = df['kecamatan'].str.strip()

[17] ✓ Od
# Mengecek kembali karakter spasi khusus
print(df['kecamatan'].unique())

['CIKARANG BARAT' 'CIKARANG PUSAT' 'CIKARANG UTARA' 'CIKARANG SELATAN'
'CIBITUNG' 'CIKARANG TIMUR' 'TARUMAJAYA' 'SERANG BARU' 'TAMBUN SELATAN'
'CIBARUSAH' 'BOJONGMANGU' 'KEDUNGWARINGIN' 'BABELAN' 'CIKARANG BARAT'
'KALIJAYA' 'Unknown' 'BANDUNG']
```

Gambar 15. Memperbaiki karakter spasi khusus

Gambar di atas menunjukkan proses *cleaning data* untuk memperbaiki karakter spasi khusus (`\xa0`) dengan menggantinya menjadi spasi biasa menggunakan fungsi `str.replace()` agar data lebih konsisten.

4.6. Integrasi Data (Exporting)

Tahap *exporting* diimplementasikan untuk melakukan retensi atau penyimpanan terhadap *dataset* yang telah melewati serangkaian prosedur *gathering*, *assessing*, hingga *cleaning*. Hasil pemrosesan data tersebut kemudian diekspor ke dalam format berkas baru guna memfasilitasi konektivitas dengan *platform* Looker Studio. Langkah ini krusial untuk memastikan bahwa data yang telah tervalidasi dapat ditransformasikan ke dalam representasi visual yang komprehensif.

4.7. Analisis Data

Pada tahap ini, teknik BI digunakan untuk menganalisis sebaran dan dinamika industri peserta pelatihan di Kabupaten Bekasi. Fokus analisis adalah tren partisipasi tahunan dan pemetaan konsentrasi industri berdasarkan kecamatan dan zona wilayah. Dengan melihat visualisasi ini, pemangku kepentingan dapat melihat pola persebaran dan perbandingan data secara langsung tanpa mengolah data asli. Ini memudahkan pengambilan keputusan tentang pengembangan program pelatihan yang lebih sesuai dengan tujuan.

4.8. Dashboard dan Visualisasi

Dashboard ini menggunakan *Pivot Table*, *Line Chart*, dan *Bar Chart* untuk menyajikan kategori, sebaran wilayah, dan tren industri per tahun. Visualisasi ini membuat lebih mudah untuk melacak perubahan jumlah industri dari waktu ke waktu dengan cara yang informatif dan mudah dipahami.



Gambar 16. Dashboard Industri Peserta Pelatihan Kabupaten Bekasi

4.9. Kartu Ringkasan Dashboard

Total Industri	Total Kecamatan	Total Zona
199	17	18

Gambar 17. Kartu Ringkasan Dashboard

Visualisasi kartu ringkasan menampilkan informasi umum mengenai data industri yang digunakan dalam *dashboard*. Pada *dashboard* ini, kartu ringkasan menunjukkan total industri sebanyak 199, total kecamatan sebanyak 17, dan total zona sebanyak 18. Informasi ini berfungsi sebagai gambaran awal mengenai cakupan data yang dianalisis sebelum pengguna melihat visualisasi yang lebih rinci pada grafik dan tabel.

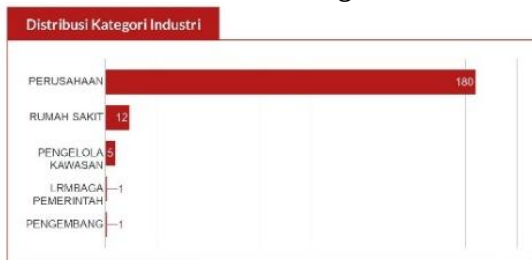
4.10. Visualisasi Data Perbandingan Jumlah Industri Berdasarkan Kecamatan



Gambar 18. Perbandingan Jumlah Industri Berdasarkan Kecamatan

Menurut peta wilayah, ada konsentrasi industri yang tidak merata di Kabupaten Bekasi, dengan mayoritas terkonsentrasi di daerah *Cikarang*. Kecamatan Cikarang Barat memiliki industri terbanyak dengan 52 unit, disusul Cikarang Selatan dengan 38 unit, dan Cikarang Utara dengan 30. Di sisi lain, kecamatan seperti Tambun Selatan, Cikarang Pusat, hingga Cibitung, dan Serang Baru memiliki intensitas industri yang lebih rendah.

4.11. Visualisasi Distribusi Kategori Industri



Gambar 19. Distribusi Kategori Industri

Visualisasi distribusi kategori industri disajikan dalam bentuk *bar chart horizontal* untuk menunjukkan jumlah industri berdasarkan kategori. Pada *dashboard* terlihat bahwa kategori perusahaan merupakan kategori yang paling dominan, yaitu sebanyak 180 data, jauh lebih tinggi dibandingkan kategori lain seperti rumah sakit, pengelola kawasan, lembaga pemerintah, dan pengembang. Hal ini menunjukkan bahwa data industri peserta pelatihan di Kabupaten Bekasi didominasi oleh entitas perusahaan.

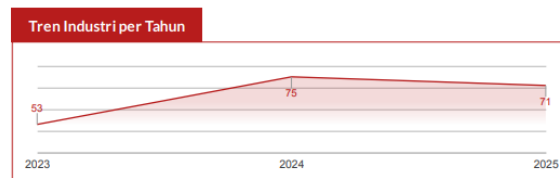
4.12. Visualisasi Jumlah Industri per Zona Kawasan



Gambar 20. Jumlah Industri per Zona Kawasan

Jika kita melihat jumlah industri per zona kawasan, kita melihat bahwa kategori luar kawasan lebih banyak dibandingkan dengan zona *lainnya*. Hasil menunjukkan bahwa peserta pelatihan cukup banyak tersebar di luar zona industri formal, bukan hanya di kawasan formal seperti MM2100, Jababeka, dan Lippo.

4.13. Visualisasi Tren Industri per Tahun



Gambar 21. Tren Industri per Tahun

Jumlah industri meningkat dari 53 pada 2023 menjadi 75 pada 2024, dan turun sedikit menjadi 71 pada 2025. Secara keseluruhan, grafik ini *menunjukkan* pola pertumbuhan yang signifikan pada tahun kedua, sementara tingkat partisipasi tetap tinggi pada tahun terakhir.

4.14. Visualisasi Tabel Ringkasan Industri

Kecamatan	Tahun / Total Industri		
	2024	2025	2023
CIKARANG BARAT	14	21	17
CIKARANG SELATAN	14	7	17
CIKARANG UTARA	15	8	7
TAMBUN SELATAN	10	11	1
CIKARANG PUSAT	6	3	5

Gambar 22. Tabel Ringkasan Industri

Untuk melengkapi grafik utama, tabel ringkasan menyajikan rincian angka per kecamatan dan tahun (2023–2025). Melalui tabel ini, pengguna dapat membandingkan data secara menyeluruh dan spesifik, terutama di daerah dengan banyak industri seperti Cikarang dan Tambun Selatan.

5. KESIMPULAN DAN SARAN

Dengan menggunakan *Python* dan *Looker Studio*, implementasi BI berhasil menampilkan 199 data industri dari 17 kecamatan dan 18 zona kawasan di Kabupaten Bekasi. Hasil menunjukkan bahwa *dataset* terdiri dari 199 data industri, tersebar di 17 kecamatan dan 18 zona wilayah. Kategori perusahaan adalah yang paling banyak digunakan. Karena lokasinya sebagai pusat industri manufaktur, Kecamatan Cikarang Barat memiliki jumlah industri peserta pelatihan tertinggi, mencapai 26,1% (52 industri), sedangkan kategori zona kawasan industri terbesar berada di peringkat Luar Kawasan. Selain itu, puncak partisipasi terjadi pada 75 industri pada tahun 2024 (37,7%), meningkat signifikan dari 53 industri pada tahun 2023 (26,6%). Pada tahun 2025, partisipasi turun sedikit menjadi 71 industri (35,7%). Secara keseluruhan, *dashboard* interaktif ini berfungsi dengan baik untuk memberikan analisis singkat tentang penyebaran dan dinamika partisipasi dalam pelatihan industri.

Memperpanjang rentang waktu data, meningkatkan variasi visualisasi, dan meningkatkan fitur filter untuk analisis yang lebih menyeluruh adalah saran untuk penelitian selanjutnya. Selain itu, jika *dataset* tambahan dari sumber resmi tersedia, penelitian dapat dikembangkan untuk mengevaluasi efektivitas program pelatihan industri secara lebih mendalam.

DAFTAR PUSTAKA

- [1] K. Sinaga and P. Axido, "PENGARUH PENDIDIKAN DAN PELATIHAN TERHADAP KINERJA KARYAWAN PADA PT . KAWASAN INDUSTRI ABSTRAK : ABSTRACT :," vol. 1, pp. 116–131, 2021.
- [2] M. Ringan, K. U. E. Bangkit, and D. A. N. Bolu, "Faculty of Economics Riau University, Pekanbaru, Indonesia," 2017.
- [3] L. Yulia, "Fenomena Kesenjangan Kesempatan Kerja (Gap Employment) antara struktur peluang Kerja Industri dengan Kesiapan Sumber Daya Manusia di Kabupaten Karawang," vol. 5, no. 1, pp. 4465–4475, 2026.
- [4] A. J. Ahmad, L. Hakim, and N. Mustari, "Strategi Pengembangan Sumber Daya Manusia Sektor Pariwisata Melalui Partisipasi Masyarakat," vol. 7, pp. 252–261, 2024.
- [5] A. F. Ariani, K. Aulia, L. Ode, and A. Arafat, "PENGEMBANGAN DASHBOARD INTERAKTIF MENGGUNAKAN LOOKER STUDIO UNTUK VISUALISASI DAN PREDIKSI HARGA KOMODITAS CABE DI JAWA TIMUR," vol. 8, no. 4, pp. 8067–8074, 2024.
- [6] N. E. Nurhayati, U. Panca, S. Bekasi, K. Bekasi, L. Studio, and L. Studio, "IMPLEMENTASI BUSINESS INTELLIGENCE DALAM VISUALISASI DATA," no. 2, pp. 54–60, 2025.
- [7] A. Medika, M. Microsoft, and P. Bi, "(2) 1,2)," vol. 8, no. 2, pp. 52–56, 2023.
- [8] S. Nugraha, "Implementasi business intelligence untuk visualisasi data milik thrift shop," 2022.
- [9] J. S. Informasi and U. Tarumanagara, "PERANCANGAN DASHBOARD UNTUK VISUALISASI DATA PENJUALAN PT . KZP".
- [10] J. Blank and K. Deb, "Pymoo : Multi-Objective Optimization in Python," 2020, doi: 10.1109/ACCESS.2020.2990567.
- [11] D. A. R. Eaton, "Toytrees : A minimalist tree visualization and manipulation library for Python," vol. 2020, no. September 2019, pp. 187–191, 2020, doi: 10.1111/2041-210X.13313.
- [12] N. Saepul *et al.*, "PENGEMBANGAN SISTEM VISUALISASI INTERAKTIF DATA STATISTIK KABUPATEN BINTAN PADA PLATFORM " BINTAN IN HAND " UNTUK MENINGKATKAN PEMAHAMAN," vol. 16, no. Januari, pp. 7–13, 2025.
- [13] C. City *et al.*, "Implementasi Business Intelligence untuk Visualisasi Laju Indeks Pembangunan Manusia Kota Cirebon Menggunakan Google Collab," vol. 21, no. 1, pp. 33–45, 2024, doi: 10.30595/sainteks.v21i1.21356.
- [14] M. D. Feters, L. A. Curry, and J. W. Creswell, "Achieving Integration in Mixed Methods Designs — Principles and Practices," pp. 2134–2156, doi: 10.1111/1475-6773.12117.
- [15] A. Agung, A. Daniswara, and I. K. D. Nuryana, "Data Preprocessing Pola Pada Penilaian Mahasiswa Program Profesi Guru," vol. 05, pp. 97–100, 2023.