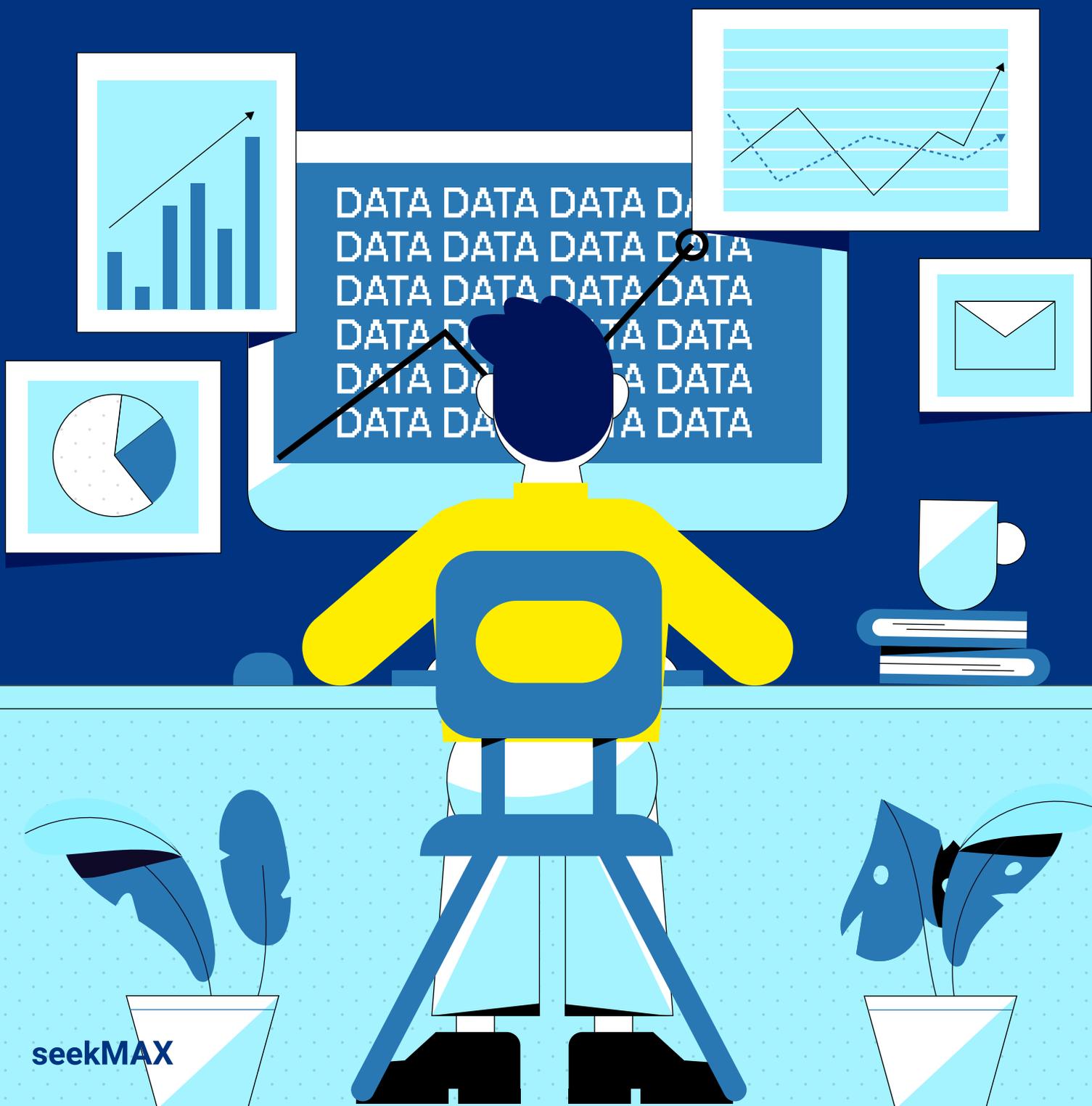


# PELATIHAN DASAR DATA SCIENCE

“Data Analyst vs Data Scientist, Apa Bedanya?”

- oleh Yohanes Christianto



# Kata Pengantar



“Data Analyst vs Data Scientist, pilih yang mana ya?”

Pertanyaan itu pasti sering terlintas di pikiran ketika kamu akan memulai karir di bidang data. Kedua profesi tersebut memang punya tujuan yang sama, yaitu menggali informasi sebanyak banyaknya dari data yang ada untuk menciptakan sebuah *insight*. *Insight* yang telah didapat tersebut akan digunakan untuk menyelesaikan masalah yang tentunya berdampak baik dalam pengembangan suatu bisnis. Nah dari situlah banyak orang yang mengira kalau keduanya merupakan profesi yang sama, hanya penyebutannya saja yang beda. Ditambah lagi di zaman era digital ini, *data analyst*, *data scientist*, dan *data engineer* merupakan profesi yang sedang ramai dibicarakan. Ah daripada makin pusing, yuk kita bahas satu persatu antara *data analyst*, *data scientist*, dan *data engineer*!

Oiya, kita belum kenal ya sama Yohanes Christianto. Yohanes Christianto ini adalah seorang *data consultant*. *Data consultant* itu sendiri memiliki ruang kerja dan tanggung jawab yang lebih luas, yaitu dimulai dari mendesain dari *data engineer* sampai ke ranah *data analyst* dan *data science*. Sebelum menjadi *data consultant*, Yohanes Christianto mengawali karirnya sebagai *software engineering* yang kemudian beralih menjadi *data engineer*. Pada profesi inilah Yohanes Christian mengenal dunia *data science*.

# Daftar Isi

<b>Bab 1: Profesi Apa Saja yang Ada Dalam Big Data?</b> .....	<b>1</b>
1.1 Data Scientist	
1.2 Data Analyst	
1.3 Data Engineer	
<b>Bab 2: Bedanya Data Scientist dan Data Analyst</b> .....	<b>8</b>
2.1 Tujuan dan Lingkup Kerja	
2.2 Skill dan Tools	
2.3 Career Path	
<b>Bab 3: Menjadi Data Engineer</b> .....	<b>12</b>
3.1 Tujuan dan Lingkup Kerja	
3.2 Skill dan Tools	
3.3 Career Path	
<b>Daftar Pustaka</b> .....	<b>17</b>

Bab 1:

# Profesi Apa Saja yang Ada dalam Big Data?



Dunia *big data* tengah menjadi topik hangat yang sedang ramai dibicarakan karena penerapannya dapat ditemukan di berbagai industri. Lantas, apa sih *big data* itu? *Big data* adalah kumpulan data yang begitu kompleks sehingga diperlukan pemrosesan yang panjang meliputi pencarian dan pengambilan data, analisis data, kurasi data, berbagi dan mentransfer data, penyimpanan data, visualisasi data hingga pembaruan dan privasi data (Dianta et al., 2022). Nah keliatan kan bagaimana kompleksnya dalam mengolah data? Maka dari itu terbentuklah profesi *data analyst*, *data scientist*, dan *data engineer*. Namun, secara praktiknya dalam dunia industri, *data analyst*, *data scientist*, dan *data engineer* sering kali disamakan bahkan sering tertukar karena ketiga profesi itu memang berperan dalam *big data*. Lalu kira-kira bagaimana ya cara membedakannya? Yaudah yuk kita bahas satu per satu.

## 1.1 Data Scientist



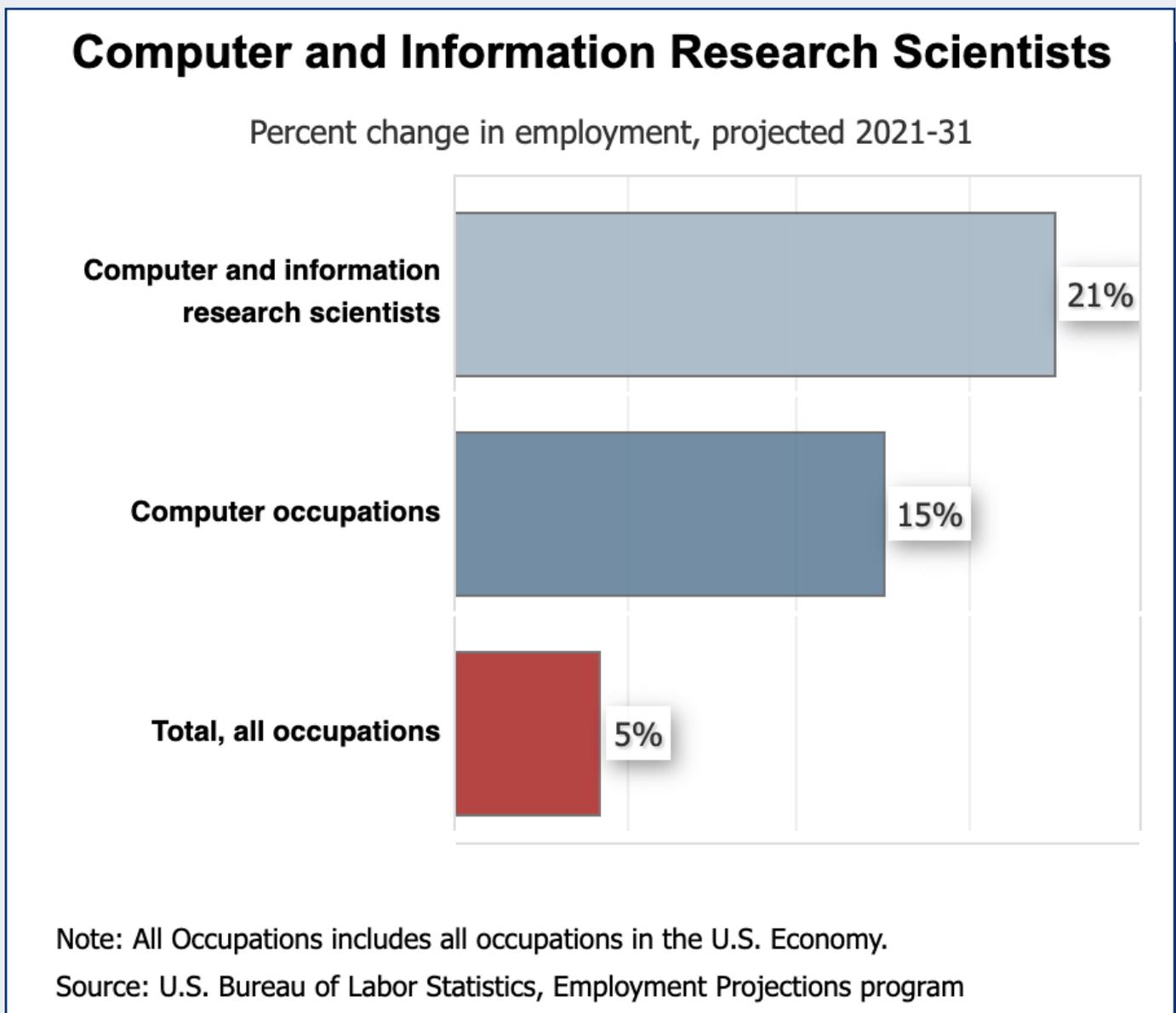
Ada kabar baik nih buat kamu yang tertarik menjadi *data scientist*. **Saat ini pekerjaan sebagai *data scientist* sangat dibutuhkan oleh banyak perusahaan sedangkan di Indonesia belum banyak yang berprofesi sebagai *data scientist*** (Hairani & Amrullah, 2020; Ismail & Zainal Abidin, 2016).

*U.S. Bureau of Labor Statistics, Employment Projections* program pun menyatakan bahwa diperkirakan pekerjaan pada bidang *data scientist* ini akan mengalami peningkatan sebanyak 21% dalam jangka waktu 2021 hingga 2031, lebih cepat bila dibandingkan untuk semua pekerjaan (U.S. BUREAU OF LABOR STATISTICS, 2023). Rata-rata sebanyak 3.300 lapangan pekerjaan tersedia setiap tahunnya selama satu dekade. Tertarik untuk mencoba menjadi seorang *data scientist*?

Apakah kamu ingin menjadi seorang *data scientist*? Sudah tahu belum “apa itu *data scientist*” dan “apa aja sih kerjaan *data scientist*”?

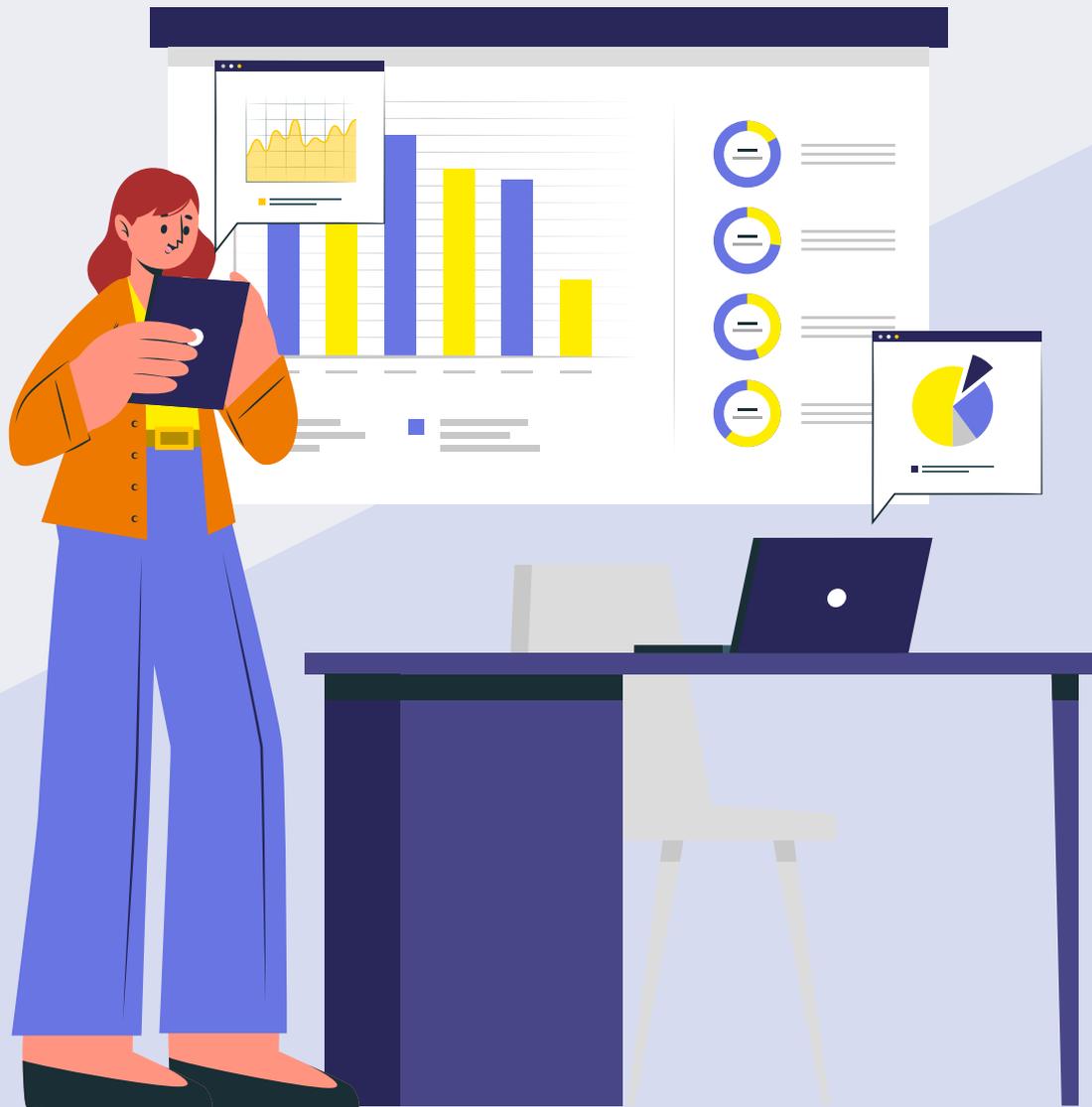
*Data scientist* itu seseorang yang bekerja untuk **membersihkan, merapikan, dan memproses *big data*** (Dinov, 2018; Hartatik et al., 2023) dari sebuah perusahaan yang tentunya akan sering bertemu dengan bahasa pemrograman seperti SQL, Python, dan R (Aleryani, 2020). Dengan demikian, seorang *data scientist* harus menguasai tiga pilar ilmu utama yaitu matematika, statistik dan komputer pemrograman (Kuncoro, 2020). Selain itu, ketika kamu menjadi seorang *data scientist*, kamu juga harus memiliki keahlian di bidang komunikasi untuk menyajikan hasil data yang nantinya menghasilkan sebuah *insight*. *Insight* tersebut harus dapat diterima dan dimengerti oleh semua divisi.

Gambar 1. Persentase Kebutuhan Pekerjaan Sebagai Data Scientist



Sumber: U.S. BUREAU OF LABOR STATISTICS, 2023

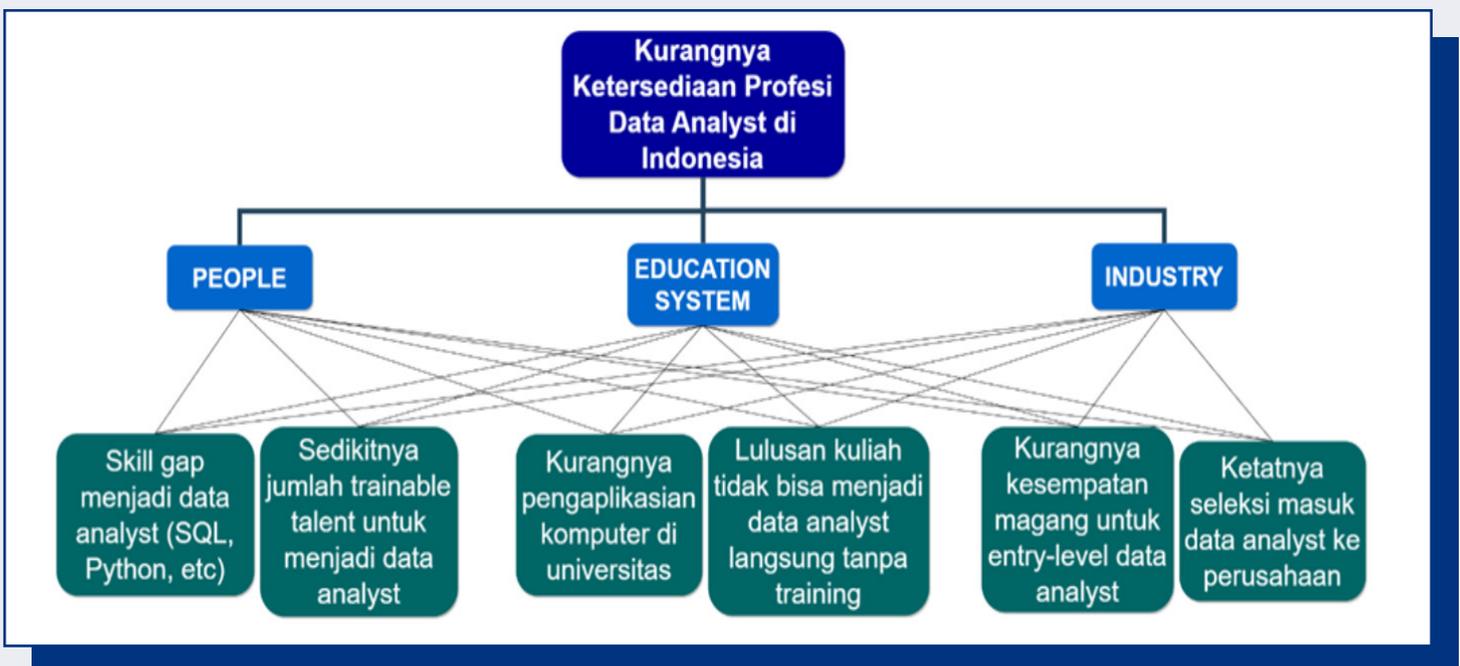
## 1.2 Data Analyst



Selanjutnya adalah profesi yang “mirip” dengan *data scientist* yaitu *data analyst*. Seorang *data analyst* itu bekerja untuk **membersihkan, menganalisis dan memvisualisasikan data** (Setiawan, 2021). Tugasnya dimulai dari memperoleh, membersihkan dan menganalisis informasi perkembangan bisnis untuk masa depan dari berbagai sudut dan aspek. Hal-hal lain yang harus dilakukan oleh seorang *data analyst* adalah harus memvisualisasikan data yang menghubungkan dengan divisi pemasaran, divisi penjualan, divisi teknis dan strategi bisnis (Cheng et al., 2023). Dengan demikian, seorang *data analyst* harus menguasai teknik dan alat yang digunakan seperti program ms. excel, tableau dan SQL untuk membuat grafik dan infografis dalam memvisualisasikan data (Ismail & Zainal Abidin, 2016).

Ada kabar baik nih buat kamu yang tertarik menjadi *data analyst*. Orang yang **berprofesi sebagai *data analyst* masih terbilang langka sehingga SDM yang mahir di bidang tersebut sangat dibutuhkan di Indonesia**. Kurangnya minat terhadap profesi tersebut disebabkan oleh faktor *people* (sumber daya manusia), *education system* (sistem pendidikan di Indonesia) dan industri (Susanto & Kurniawan, 2023). Berikut adalah gambarannya:

Gambar 2. Faktor yang Penyebab Kurangnya Ketersediaan Profesi Data Analyst di Indonesia

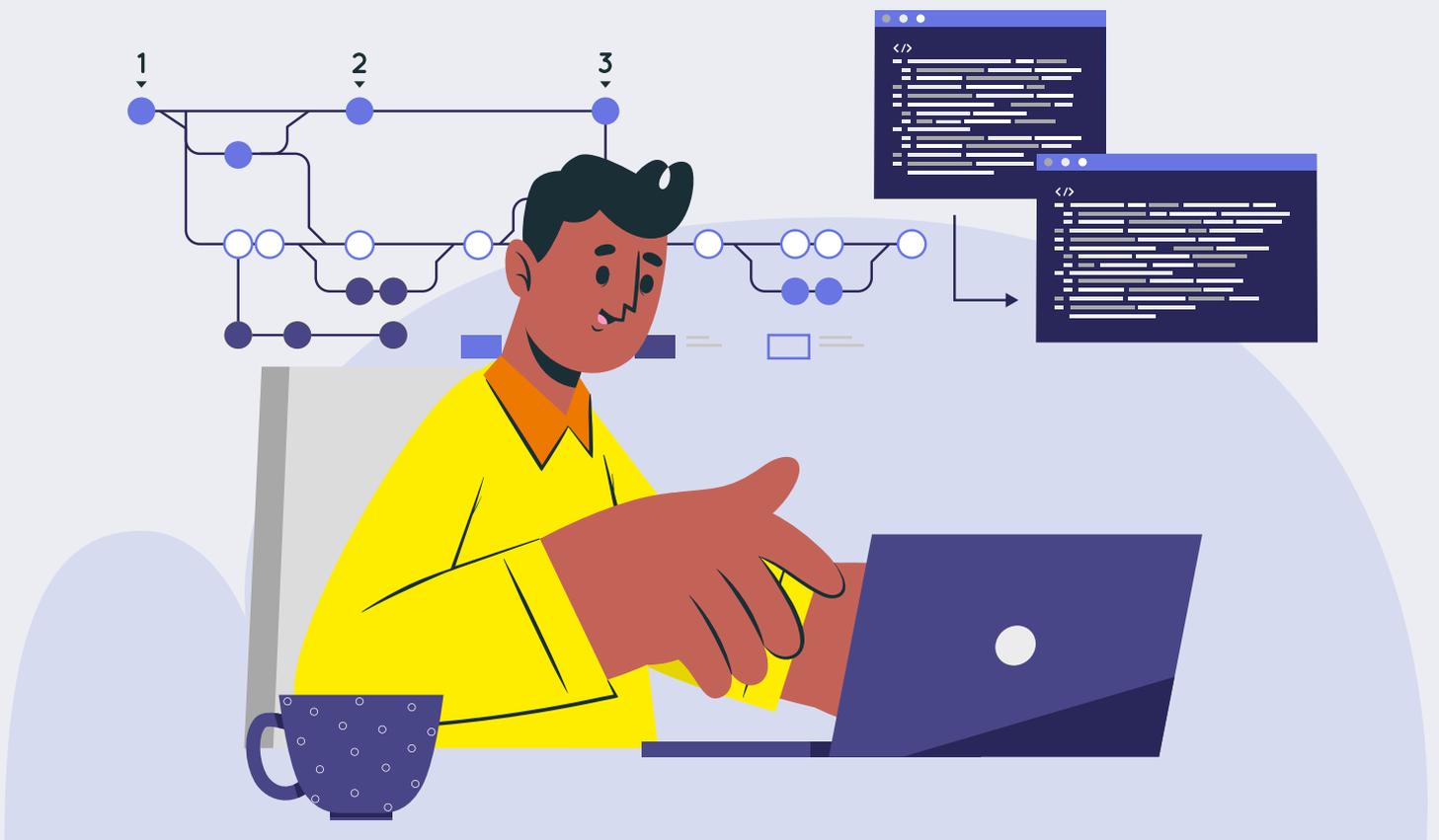


Sumber: **Susanto & Kurniawan, 2023**

## 1.3 Data Engineer

Dari topik-topik sebelumnya, "*data engineer*" di-mention terus ya. Kamu tahu ga sih apa itu *data engineer*? Jadi, seorang *data engineer* adalah orang yang mengembangkan serta membuat desain arsitektur sebuah manajemen data dan memonitor infrastrukturnya di dalam sebuah perusahaan. Singkatnya, seorang *data engineer* **mewadahi alur perputaran data** untuk perusahaan dengan skala besar (Suharmato, 2019). Selain itu, seorang *data engineer* harus memastikan bahwa data dapat dikumpulkan dan diakses secara efisien dari sumbernya, dibersihkan,

dan diproses seperlunya. Tujuannya adalah untuk mengoptimalkan sistem pertukaran data antara *data scientist* dan *data analyst*. Seorang *data engineer* harus memiliki pengetahuan tentang pemrograman, *big data*, dan tentunya matematika. Selanjutnya, *tools* yang akan sering kamu gunakan ketika menjadi seorang *data engineer* salah satunya adalah Python. Selain itu, kamu juga harus menguasai *database*, dan SQL. Selain itu sebisa mungkin kamu harus memahami alat ETL, skrip shell, dan pemrograman dasar juga (Abeykoon et al., 2020).



Meskipun ada beberapa perbedaan antara *data engineer*, *data scientist* dan *data analyst* namun ketiganya masih saling keterkaitan satu sama lain. **Data analyst dan data scientist tidak dapat saling bertukar data tanpa data engineer. Sementara itu data engineer juga tidak akan bekerja tanpa adanya data scientist dan data analyst. Saat ini lowongan pekerjaan untuk ketiga profesi tersebut masih sangat terbuka lebar khususnya di Indonesia. Ditambah lagi banyak sektor bisnis seperti bidang perbankan, kesehatan, pendidikan, transportasi, media, perjalanan hingga logistik pasti membutuhkan ketiga profesi tersebut** (Alonso-Fernández et al., 2019).

Bab 2:

# Perbedaan Data Scientist dan Data Analyst



Sudah tahu kan peran serta keterkaitan antara profesi *data engineer*, *data scientist* dan *data analyst*? Nah seperti yang sudah kita singgung pada awal pembahasan, banyak orang yang berpikir kalau *data scientist* dan *data analyst* adalah suatu profesi yang sama. Sekilas keduanya memang tampak mirip karena sama-sama bersinggungan dengan data lalu menganalisisnya. Namun nyatanya, *data scientist* memiliki tanggung jawab yang lebih banyak dibanding *data analyst*.

Sederhananya, *data analyst* bekerja dengan data yang sudah terstruktur dengan tujuan yang lebih *tangible* sedangkan *data scientist* harus memecahkan hal yang bersifat *intangible* dengan data mentah yang belum tentu terstruktur. Gimana? Sudah ada bayangan? Nah biar lebih jelas, yuk kita bahas perbedaannya dari berbagai aspek!

## 2.1 Tujuan & Lingkup Pekerjaan

Pertama yang akan kita bahas adalah dari aspek tujuan dan lingkup pekerjaan dari seorang *data scientist* dan *data analyst*. Berikut ini adalah tabel perbedaan antara *data scientist* dan *data analyst* dari aspek tujuan dan lingkup pekerjaan:

**Tabel 1. Perbedaan *Data Scientist* dan *Data Analyst* Dari Aspek Tujuan dan Lingkup Pekerjaan**

DATA ANALYST	TINJAUAN	DATA SCIENTIST
Pemeliharaan data	Tujuan	Analisis data menggunakan <i>machine learning</i> dan <i>deep learning</i>
Mengumpulkan data dan <i>data processing</i>	Peran	Mengintegrasikan data
Mengoptimalkan efisiensi dan kualitas statistik data		Ikut serta dalam perencanaan analisis data
Analisis statistik dan interpretasi data	Tanggung jawab profesi	Mengembangkan pemodelan atau strategi
Merepresentasi data melalui laporan dan visualisasi data	<i>Output</i>	Menjembatani antara <i>stakeholder</i> dan <i>customers</i>

Sumber: Baihaqi, 2022

## 2.2 Skill dan Tools

Perbedaan selanjutnya adalah dari *skill* dan *tools* yang harus dimiliki oleh seorang *data scientist* dan *data analyst*. Yuk kita lihat tabel di bawah ini!

Tabel 2. Perbedaan *Data Scientist* dan *Data Analyst* dari Aspek *Skill* dan *Tools*-nya

DATA ANALYST	SKILL & TOOL	DATA SCIENTIST
Excel	Kemampuan dasar	<i>machine learning</i> dan <i>deep learning</i>
SQL	Keterampilan pemrograman	SQL, Pig, R, Scala, Python, Java, JavaScript
Tableau, polymer search, akkio	Database Management	MySQL, SQL Server, Oracle
Matematika dan Statistika	Bidang Ilmu yang harus dikuasai	Matematika, Statistika, Ilmu Komputer
SAS	Tools	Apache Spark, IBM SPSS, Matlab, PyTorch, SAS

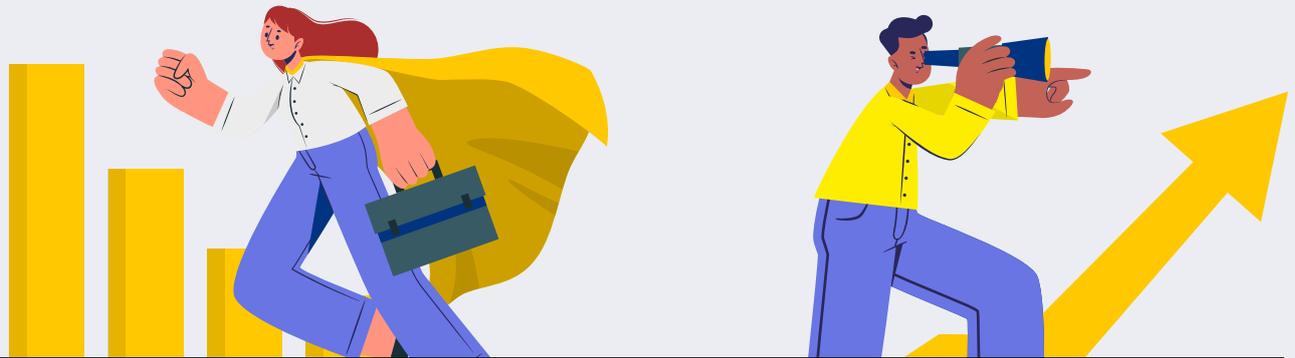


Sumber: Baihaqi, 2022

## 2.3 Career Path

Nah sekarang sudah mulai jelas kan tentang seluk-beluknya *data scientist* dan *data analyst*? Ada lagi nih jenjang karirnya untuk profesi *data scientist* dan *data analyst* biar makin jelas masa dunia kerjamu. Awalnya kamu pasti memulainya pada **entry level** sebagai junior *data scientist* dan *data analyst*. Selanjutnya pada **mid-level**, kamu mulai menduduki posisi *data scientist* dan *data analyst*. Lama-lama karirmu akan meningkat ke **senior level** menjadi senior *data scientist* dan *data analyst*, yang kemudian kamu akan menjadi **manager** atau sebagai *chief data scientist* dan *data analyst*. Nah point tambahan disini, **kamu bisa bergeser posisi dari *data scientist* ke *data analyst* ataupun sebaliknya**. Berikut ini adalah tabel tahapan karirmu sebagai *data scientist* dan *data analyst*:

Tabel 3. Levelling Career Data Scientist dan Data Analyst



DATA ANALYST	LEVELING CAREER	DATA SCIENTIST
Junior Data Analyst	Entry Level	Junior Data Scientist
Data Analyst	Mid-Level	Data Scientist
Senior Data Analyst	Senior Level	Senior Data Scientist
Chief Data Analyst	Manager	Chief Data Scientist

Sumber: Baihaqi, 2022

Bab 3:

# Menjadi Data Engineer

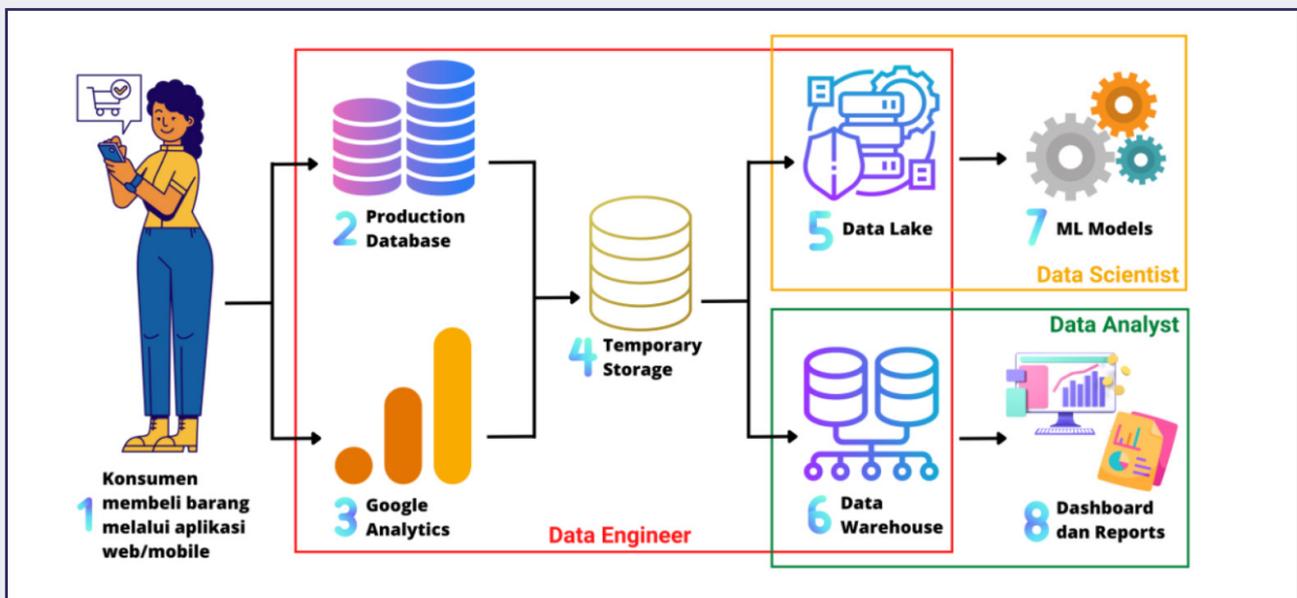


*Data analyst, data scientist, dan data engineer* merupakan profesi yang sangat berperan dalam dunia *data science*. Seperti yang telah dijelaskan sebelumnya bahwa *data analyst* dan *data scientist* tidak dapat saling bertukar data tanpa *data engineer*. Sementara itu *data engineer* juga tidak akan bekerja tanpa adanya *data scientist* dan *data analyst*. *Data engineer* di sini sangat berperan karena mereka menjadi wadah hasil kerja *data scientist* dan *data analyst*.

## 3.1 Tujuan & Lingkup Pekerjaan

Secara umum, pengertian dari *data engineer* itu sendiri merupakan seorang yang mengatur dan mengelola proses arsitektur data pada sebuah perusahaan, di mana arsitektur tersebut berupa *database, pipeline* atau *warehouse*. Dalam mengatur arsitektur data, seorang *data engineer* berurusan dengan volume data yang jumlahnya dan ukurannya cukup besar. Kebutuhan data yang semakin besar tidak mungkin ditampung dengan hanya bermodalkan infrastruktur yang konvensional saja.

Gambar 3. Ilustrasi Hubungan Kerja Data Analyst, Data Scientist, dan Data Engineer



Sumber: Amira, 2023

Nah terus, apa saja yang perlu dipersiapkan kalau kita akan memulai karir kita sebagai *data engineer*? Yuk kita dalami tugas dan tanggung jawab seorang *data engineer*!

- Memastikan penyimpanan dan pengumpulan data telah sesuai dengan persyaratan bisnis dan standar industri yang ada
- Melakukan integrasi *software* manajemen data yang telah ada dalam perusahaan
- Melakukan riset untuk melihat kesempatan jika ada akuisisi data dalam bisnis
- Membuat secara *custom* suatu komponen *software* dengan berbagai bahasa pemrograman dan *tools* untuk menggabungkan sistem yang berbeda. Di sisi lain, seorang *data engineer* akan mengembangkan struktur dan infrastruktur data dalam bisnis.
- Melakukan penyimpanan dan memproses data agar aman. *Data engineer* harus siap menjadi garda terdepan dalam menghadapi serangan *cyber*.

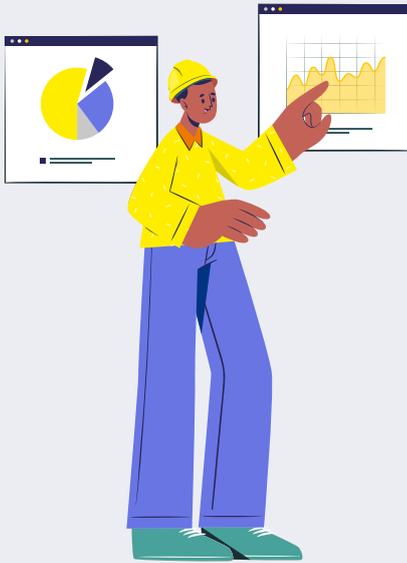
## 3.2 Skill dan Tools

Nah selanjutnya, *skill* dan *tools* apa sih yang dibutuhkan kalau berminat menjadi *data engineer*? Jadi *skill* yang dibutuhkan dibagi menjadi dua, ada *skill* teknis dan *skill* non-teknis. Kalau *skill* teknis yang dibutuhkan itu kamu harus mahir di bidang matematika dan statistika. Lalu kamu juga harus bisa memvisualisasikan data, paham dengan struktur data, bahasa pemrograman, *database system*, *data pipeline*, *data mining*, dan *machine learning*.

Kalau *skill* non-teknisnya gimana? Jadi kamu harus mempunyai kemampuan komunikasi (verbal dan tulis), kemampuan koordinasi dan kolaborasi, harus berpikir kritis, dan harus cepat beradaptasi. Selanjutnya, kamu harus bisa mempresentasikan hasil kerja kamu agar orang selain divisimu mengerti. Terakhir seorang *data engineer* harus *business acumen*.

Apa sih *business acumen* itu? *Business acumen* merupakan kemampuan untuk memahami dan menghadapi berbagai situasi bisnis, semua risiko dan peluangnya, dengan tajam dan akurat untuk mendapatkan hasil yang menguntungkan.

Tabel 4. Skill yang Dibutuhkan untuk Seorang Data Engineer



SKILL TEKNIS	SKILL NON-TEKNIS
Matematika dan statistika	Kemampuan komunikasi (verbal dan tulis)
Visualisasi data	Kemampuan koordinasi dan kolaborasi
Struktur data	Critical thinking
Bahasa pemrograman	Presentasi
Database systems	Business acumen
Data pipeline	Adaptability
Data mining	
Machine learning	

Sumber: Alizanovic, 2023

Dalam menjalankan tanggung jawab dan menerapkan *skill-skill* tersebut, seorang *data engineer* membutuhkan beberapa *tools* seperti berikut, beserta contoh *platform* yang biasa digunakan:

Tabel 5. Tools yang Digunakan Untuk Data Engineer



FUNGSI	CONTOH PLATFORM
ETL tools	Talend, IBM Datastage, Pentaho, dan lain sebagainya
Tools untuk platform komersial terkait teknologi big data	IBM, Oracle, dan lain sebagainya
Tools dan frameworks berbasis Hadoop	HDFS, MapReduce, HIVE, dan lain sebagainya
Database management system tools	MySQL atau Microsoft SQL Server
Real-time processing frameworks tools	Apache Spark, Beam, atau Kafka
Tools untuk data mining dan modeling	Rapid Miner, Weka, dan KNIME

Sumber: Alizanovic, 2023

### 3.3 Jenjang Karir

Kalau jenjang karirnya gimana ya? Mirip juga dong sama *data analyst* dan *data scientist*? Yap, betul! Jenjang karir buat kamu pemula sebagai *data engineer*, di posisi *entry level* di mana kamu menjadi seorang *junior data engineer*. Selanjutnya di *mid level*, kamu sudah dapat dikatakan sebagai *data engineer officer*. Setelah itu, kamu akan memasuki jenjang *senior level* di mana kamu akan menjadi *senior data engineer* dan kalau kamu sudah *expert* banget nih, kamu bisa menjadi *chief data engineer* atau setara dengan *manager*.

Tabel 6. Jenjang Karir Untuk Seorang *Data Engineer*

FUNGSI	CONTOH PLATFORM
ETL tools	Talend, IBM Datastage, Pentaho, dan lain sebagainya
Tools untuk platform komersial terkait teknologi big data	IBM, Oracle, dan lain sebagainya
Tools dan frameworks berbasis Hadoop	HDFS, MapReduce, HIVE, dan lain sebagainya
Database management system tools	MySQL atau Microsoft SQL Server
Real-time processing frameworks tools	Apache Spark, Beam, atau Kafka
Tools untuk data mining dan modeling	Rapid Miner, Weka, dan KNIME

Sumber: Baihaqi, 2022

# Daftar Pustaka



- Abeykoon, V., Perera, N., Widanage, C., Kamburugamuve, S., Kanewala, T. A., Maithree, H., Wickramasinghe, P., Uyar, A., & Fox, G. (2020). Data Engineering for HPC with Python. *Proceedings of PYHPC 2020: 9th Workshop on Python for High-Performance and Scientific Computing, Held in Conjunction with SC 2020: The International Conference for High Performance Computing, Networking, Storage and Analysis*, 13–21.  
<https://doi.org/10.1109/PyHPC51966.2020.00007>
- Aleryani, A. (2020). A data analysis perspective by the Business Analyst and Data Scientist. *International Journal of Scientific and Research Publications (IJSRP)*, 10(9), 234–243.  
<https://doi.org/10.29322/ijsrp.10.09.2020.p10525>
- Alizanovic, V. (2023, March 10). *Big Data Engineer: Definisi, Tugas, Skill, dan Jenjang Karier*. Pacmann.io.  
<https://pacmann.io/blog/big-data-engineer#skill-yang-wajib-dimiliki-big-data-engineer>
- Amira, Q. H. (2023, February 6). *Bagaimana perbedaan posisi Data Engineer, Data Scientist, dan Data Analyst pada praktiknya?* LinkedIn.  
<https://www.linkedin.com/pulse/bagaimana-perbedaan-posisi-data-engineer-scientist-dan-/?originalSubdomain=id>
- Baihaqi, F. (2022, December 13). *Perbedaan Karir Data Analyst, Data Engineer dan dan Data Scientist*. Kelas.Com.  
<https://kelas.work/blogs/perbedaan-karir-data-analyst,-data-engineer-dan-data-scientist>
- Cheng, L., Li, X., & Bing, L. (2023). Is GPT-4 a Good Data Analyst? *ArXiv*.  
<http://arxiv.org/abs/2305.15038>
- Dianta, I. A., Aqham, A. A., & Setiawan, D. (2022). Penerapan Big Data Untuk Mengatur Sistem Analisis Data. *Jurnal Ilmiah Teknik Mesin, Elektro Dan Komputer*, 2(1).
- Dinov, I. D. (2018). *Data Science and Predictive Analytics*. Springer International Publishing AG part of Springer Nature.  
<https://doi.org/https://doi.org/10.1007/978-3-319-72347-1>

- Hairani, H., & Amrullah, A. Z. (2020). Pelatihan Pengenalan Data Science untuk Meningkatkan Kemampuan dalam Pengolahan Data. *Jurnal Abdidas*, 1(3), 95–99. <https://doi.org/10.31004/abdidas.v1i3.31>
- Hartatik, Kwintiana, B., Nengsih, T. A., Bradja, A., & Harto, B. (2023). *DATA SCIENCE FOR BUSINESS: Pengantar & Penerapan Berbagai Sektor*. PT. Sonpedia Publishing Indonesia.
- Ismail, N. A., & Zainal Abidin, W. (2016). Data Scientist Skills. *IOSR Journal of Mobile Computing & Application*, 03(04), 52–61. <https://doi.org/10.9790/0050-03045261>
- Kuncoro, B. A. (2020). *Pengenalan Prinsip Data Science untuk Pemula*. Mripat Publisher.
- Setiawan, I. (2021). Perbedaan Data Engineer, Data Scientist dan Data Analyst. *Jurnal Jurnal Kajian Pendidikan FKIP Universitas Dwijendra*, 12(2). <http://ejournal.undwi.ac.id/index.php/widyaaccarya/index>
- Suharmato, P. (2019). *Artificial Intelligence-Driven Innovation In Smart City*. 978–623. <https://doi.org/10.30998/simponi.v0i0.370>
- Susanto, A. N. A. P., & Kurniawan, G. I. (2023). Analisis Terbatasnya Peminatan Profesi Data Analyst Di Indonesia Berdasarkan Pendekatan Analytical Hierarchy Process (AHP). *Journal of Information System, Applied, Management, Accounting and Research*, 7(1), 217–224. <https://doi.org/10.52362/jisamar.v7i1.1042>
- U.S. BUREAU OF LABOR STATISTICS. (2023, February 6). *Computer and Information Research Scientists*.