

Modul

PENGENALAN ILMU DATA DALAM JABATAN FUNGSIONAL ANALIS DATA ILMIAH

Pelatihan Jabatan Fungsional
Analisis Data Ilmiah



Penanggung Jawab:

1. Edy Giri Racman Putra, Ph.D.
2. Nining Setyowati Dwi Andayani, S.E., M.M.
3. Raden Arthur Ario Lelono, Ph.D.
4. Alpha Fadila Juliana Rahman, S.Pd., M.Pd.

Tim Penyusun Modul:

1. Cahyo Tri Anggoro
2. Ariani Indrawati
3. Ruben Silitonga

Diterbitkan oleh:

Direktorat Pengembangan Kompetensi - BRIN
Gedung BJ Habibie – Jl. MH. Thamrin No. 8
Jakarta

Diterbitkan pertama kali tahun 2022

KATA PENGANTAR

Pengembangan kompetensi Aparatur Sipil Negara (ASN) khususnya Pegawai Negeri Sipil (PNS) dalam mengembangkan karier jabatan fungsionalnya menjadi suatu tuntutan sehingga mampu menjalankan tugas dan fungsinya dengan sebaik – baiknya agar mampu memberikan kontribusi yang nyata terhadap bangsa dan Negara Kesatuan Republik Indonesia (NKRI).

Badan Riset dan Inovasi Nasional (BRIN) sebagai Instansi pembina 11 (sebelas) jabatan fungsional yang meliputi peneliti, perekayasa, pengembangan teknologi nuklir, analis perkebunrayaan, analis pemanfaatan iptek, analis data ilmiah, kurator koleksi hayati, penata penerbitan ilmiah, teknisi perkebunrayaan, teknisi penelitian dan perekayasaan, dan pranata nuklir. BRIN melalui kedeputian Sumber Daya Manusia Ilmu Pengetahuan dan Teknologi (SDMI) bertanggung jawab dalam penyelenggaraan pengembangan kompetensi 11 (sebelas) jabatan fungsional tersebut.

Kedeputian SDMI – BRIN melalui Direktorat Pengembangan Kompetensi sebagai penyelenggara pengembangan kompetensi jabatan fungsional bertanggung jawab dalam menyiapkan kebutuhan tersebut baik berupa pengelolaan pembelajaran, fasilitator, modul, bahan ajar dan sebagainya, yang merujuk kepada regulasi peraturan BRIN nomor 28 tahun 2022 tentang pedoman pelatihan pembentukan jabatan fungsional peneliti, peraturan BRIN nomor 29 tahun 2022 tentang pedoman pelatihan jabatan fungsional kurator koleksi hayati, peraturan BRIN nomor 30 tahun 2022 tentang pedoman pelatihan jabatan fungsional analis pemanfaatan iptek, peraturan BRIN nomor 31 tahun 2022 tentang pedoman pelatihan jabatan fungsional analis data ilmiah, peraturan BRIN nomor 32 tahun 2022 tentang pedoman pelatihan jabatan fungsional analis perkebunrayaan, peraturan BRIN nomor 33 tahun 2022 tentang pedoman pelatihan jabatan fungsional teknisi perkebunrayaan, dan peraturan BRIN nomor 34 tahun 2022 tentang pedoman pelatihan jabatan fungsional penata penerbitan ilmiah.

Kami mengucapkan syukur ke hadirat Tuhan Yang Maha Esa, atas berkat rahmat-Nya, modul pelatihan jabatan fungsional analis data ilmiah yang berjudul “PENGENALAN ILMU DATA DALAM JABATAN FUNGSIONAL ANALIS DATA

ILMIAH” dapat diselesaikan tepat waktu. Modul ini digunakan dalam pelatihan jabatan fungsional yang dibina oleh BRIN yang diselenggarakan oleh Kedepujian SDMI - BRIN melalui Direktorat Pengembangan Kompetensi. Kami berharap modul ini dapat memberikan manfaat dan kontribusi dalam meningkatkan dan mengembangkan kompetensi jabatan fungsional yang dibina BRIN.

Jakarta, Desember 2022

Plt. Deputi Sumber Daya Manusia
Ilmu Pengetahuan dan Teknologi
Badan Riset dan Inovasi Nasional

(Tanda tangan)

Edy Giri Rachman Putra, Ph.D.

DAFTAR ISI

KATA PENGANTAR	ii
DAFTAR ISI	iii
DAFTAR GAMBAR	v
DAFTAR TABEL	vi
PENDAHULUAN	1
A. Deskripsi Singkat	1
B. Alokasi Waktu	1
C. Tujuan Pembelajaran	1
D. Pokok Bahasan dan Subpokok Bahasan	2
MATERI POKOK 1: KONSEP DASAR PENGOLAHAN DATA SECARA ILMIAH	3
A. KONSEP DATA	3
B. KONSEP PENGOLAHAN DATA SECARA ILMIAH	4
C. Judul Submateripokok	4
D. Rangkuman	5
E. Evaluasi	6
MATERI POKOK 2: SIKLUS DATA	7
A. Perencanaan	7
B. Pengumpulan	7
C. Pengolahan	
D. Analisis	7
E. Penggunaan	
F. Manajemen data	
G. Penghapusan data	
H. Rangkuman	8
I. Evaluasi	8
MATERI POKOK 3: Track Kegiatan Analis Data Ilmiah	9
A. Data Analis	9
B. Data Engineer	9
C. Data Management	9

D. Rangkuman	10
E. Evaluasi	10
MATERI POKOK 4: Kegiatan / Tugas Analisis Data Ilmiah	11
A. Jabaran Kegiatan / Tugas Analisis Data Ilmiah	11
B. Judul Submateri pokok	11
C. Judul Submateri pokok	11
D. Rangkuman	12
E. Latihan	12
MATERI POKOK 5: Analisis hasil kerja Analisis Data Ilmiah	13
A. Jabaran hasil kerja / output dari masing - masing kegiatan ADI	13
B. Judul Submateri pokok	13
C. Judul Submateri pokok	13
D. Rangkuman	14
E. Evaluasi	14
DAFTAR PUSTAKA	15
LAMPIRAN-LAMPIRAN	16
Lampiran 1. Judul lampiran	17

DAFTAR TABEL

Tabel 1. xxxxxxxxxxxxxxxxxxxx

1

PENDAHULUAN

A. Deskripsi Singkat

Mata pelatihan ini memberikan penjelasan kepada peserta tentang pengenalan ilmu data secara umum dan khusus pada berbagai kegiatan Jabatan Fungsional Analis Data Ilmiah

B. Alokasi Waktu

Alokasi waktu disesuaikan dengan metode pelatihan. Untuk Klasikal, alokasi waktunya adalah 4 Jam Pembelajaran (JP) @ 45 menit; Untuk Bauran (*Blended Learning*) alokasinya adalah 6 JP (asinkronus: 3 JP dan sinkronous: 3JP) dan untuk Pelatihan Jarak Jauh (*Distance Learning*) alokasinya juga sebanyak 6 JP (asinkronus: 3 JP dan sinkronous: 3 JP)

C. Tujuan Pembelajaran

1. Hasil Belajar

Peserta mampu mengetahui ruang lingkup kegiatan Analis Data Ilmiah serta memetakan tugas di instansinya ke dalam unsur/subunsur/kegiatan Jabatan Fungsional Analis Data Ilmiah.

2. Indikator Hasil Belajar

Setelah selesai pembelajaran diharapkan peserta mampu:

- a. menjelaskan konsep data;
- b. mengetahui siklus hidup data;
- c. memetakan tugasnya ke ruang lingkup kegiatan Analis Data Ilmiah;
- d. mengetahui bentuk hasil kerja/output yang perlu dipersiapkan untuk penilaian; dan
- e. mengetahui karakteristik pengolahan data yang tercakup dalam Jabatan Fungsional Analis Data Ilmiah.

D. Materi Pokok

Mata pelatihan ini terdiri dari 5 (lima) Materi Pokok, yaitu:

1. Konsep dasar pengolahan data secara ilmiah

- b. Konsep data
 - c. Konsep pengolahan data secara ilmiah
2. Siklus data
- a. Perencanaan
 - b. Pengumpulan
 - c. Pengolahan
 - d. Analisis
 - e. Penggunaan
 - f. Manajemen data
 - g. Penghapusan
3. Kegiatan/tugas Analisis Data Ilmiah
- Jabaran kegiatan/ tugas Analisis Data Ilmiah
4. Analisis hasil kerja Analisis Data Ilmiah
- Jabaran hasil kerja/ output dari masing-masing kegiatan Analisis Data Ilmiah
5. Track kegiatan Analisis Data Ilmiah
- a. Data Analyst
 - b. Data Engineer
 - c. Data Management

MATERI POKOK 1:

KONSEP DASAR PENGOLAHAN DATA SECARA ILMIAH

Indikator Hasil Belajar:

Peserta mampu:

Menjelaskan pengertian data

Menguraikan konsep pengolahan data secara ilmiah

A. Konsep Data

Saat ini data merupakan hal yang tidak terpisahkan dalam kehidupan manusia sebab hampir pada seluruh aspek kehidupan, kehadiran data dapat membantu manusia dalam mempermudah manusia dalam melakukan pemecahan masalah ataupun melakukan pengambilan keputusan. Dalam konteks organisasi modern, saat ini level pimpinan dituntut untuk mampu mengambil keputusan berbasiskan data. Bahkan dalam beberapa kesempatan, data dianggap sebagai salah satu mata uang baru atau bahkan disetarakan dengan sumber daya energi seperti minyak bumi. Semakin strategisnya posisi data di era digital seperti sekarang ini, kemampuan untuk melakukan manajemen data dan analisis data juga semakin penting.

Secara definisi, data merupakan kumpulan fakta dalam berbagai bentuk dan format mentah seperti angka, karakter, image, video, rekaman suara, atau simbol (DAMA : 2021). Perkembangan beragam format data semakin membuat kegiatan analisis data ilmiah menjadi semakin kompleks. Demikian pula dengan kegiatan manajemen data yang menuntut untuk ketersediaan akses terhadap data dimasa yang akan datang. Format data dalam konteks manajemen data terbagi menjadi beberapa bentuk sebagaimana tertuang dalam tabel berikut :

Tipe	Format yang disarankan	Format yang tidak disarankan
------	------------------------	------------------------------

	Text documents	<ul style="list-style-type: none"> • PDF/A (.pdf) • ODT (.odt) 	<ul style="list-style-type: none"> • Microsoft Word (.doc) • Office Open XML (.docx) • Rich Text File (.rtf) • PDF other than PDF/A (.pdf)
	Plain Text	Unicode text (.txt)	Non Unicode text
	Markup Language	XML HTML Related files : .css, xslt, .js, .es	SGML (.sgml) Markdown (.md)
	Programming languages	MATLAB NetCDF Text Fabric	
	Spreadsheets	ODS (.ods) CSV (.csv)	<ul style="list-style-type: none"> • Microsoft Excel (.xls) • Office Open XML Workbook (.xlsx) • PDF/A (.pdf)
	Databases	<ul style="list-style-type: none"> • SQL (.sql) • SIARD (.siard) • CSV (.csv) 	<ul style="list-style-type: none"> • Microsoft Access (.mdb, .accdb) • dBase (.dbf) • HDF5 (.hdf5, .he5, .h5)
	Statistical data	<ul style="list-style-type: none"> • SPSS (.dat/.sps) • STATA (.dat/.DO) • R • 	<ul style="list-style-type: none"> • SPSS Portable (.por) • SPSS (.sav) • STATA (.dta) • SAS (.7dat; .sd2; .tpt)
	Raster images	<ul style="list-style-type: none"> • JPEG (.jpg, .jpeg) • TIFF (.tif, .tiff) • PNG (.png) • JPEG 2000 (.jp2) • DICOM (.dcm) 	<ul style="list-style-type: none"> •
	Vector Images	<ul style="list-style-type: none"> • SVG (.svg) 	<ul style="list-style-type: none"> • Adobe Illustrator (.ai) • EPS (.eps) • WMF/EMF (.wmf, .emf) • CDR (.cdr)

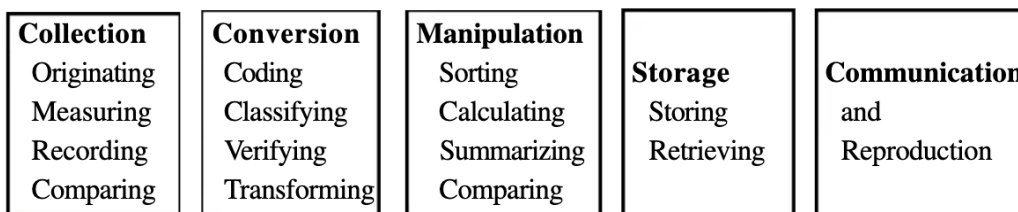
Audio	<ul style="list-style-type: none"> • BWF (.bwf) • MXF (.mxf) • Matroska (.mka) • FLAC (.flac) • OPUS 	<ul style="list-style-type: none"> • WAVE (.wav) • MP3 (.mp3) • AAC (.aac, .m4a) • AIFF (.aif, .aiff) • OGG (.ogg)
Video	<ul style="list-style-type: none"> • MXF (.mxf) • Matroska (.mkv) • 	<ul style="list-style-type: none"> • MPEG-4 (.mp4, .m4a, .m4v, ...) • MPEG-2 (.mpg, .mpeg, .m2v, .mpg2, ...) • AVI • QuickTime (.mov, .qt)
Computer Aided Design (CAD)	<ul style="list-style-type: none"> • AutoCAD DXF version R12 (ASCII) (.dxf) • SVG (.svg) 	<ul style="list-style-type: none"> • Autocad DXF, version different than R12 (ASCII) (.dxf) • DWG (.dwg) • DGN (.dgn)
Geographical Information Systems (GIS)	<ul style="list-style-type: none"> • GML (.gml) • MIF/MID (.mif/.mid) • GeoJSON (.json) 	<ul style="list-style-type: none"> • Esri Shapefiles (.shp & related files) • MapInfo (.tab & related files) • KML (.kml, .kmz) • Esri Geodatabase (.gdb) • Project files / Workspaces (.mxd, .wor, .qgs)
Georeferenced images	<ul style="list-style-type: none"> • GeoTIFF (.tif, .tiff) 	<ul style="list-style-type: none"> • TIFF World File (.tfw & .tif, possibly with additional files) • JPEG World File (.jgw & .jpg, possibly with additional files) • ERDAS IMAGINE File Format (.img)
Raster GIS	<ul style="list-style-type: none"> • ASCII GRID (.asc, .txt) • 	<ul style="list-style-type: none"> • Esri GRID (.grd & bijbehorende bestanden) • Surfer Grid (.grd; .srf) • ERDAS IMAGINE File Format (.img)
3D	<ul style="list-style-type: none"> • WaveFront Object (.obj) • Polygon file format (.ply) • X3D (.x3d) 	<ul style="list-style-type: none"> • Autodesk FBX (.fbx) • Blender (.blend) • 3D PDF (.pdf)

		<ul style="list-style-type: none"> • COLLADA (.dae) 	
	RDF	<ul style="list-style-type: none"> • RDF/XML (.rdf) • Trig (.trig) • Turtle (.ttl) • NTriples (.nt) • JSON-LD 	<ul style="list-style-type: none"> •
	Computer Assisted Qualitative Data Analysis (CAQDAS)	REFI-QDA (Qualitative Data Analysis)	<ul style="list-style-type: none"> • ATLAS.TI copy bundle • NVivo project file

B. Konsep Pengolahan Data Secara Ilmiah

Setiap organisasi, terlepas dari ukuran atau tujuannya, pasti memiliki atau menghasilkan banyak data seperti data transaksi penjualan produk, data personalia, data rantai pasok, data konsumen, dan masih banyak lainnya. Data tersebut perlu diolah untuk menghasilkan informasi yang bermanfaat. Pengolahan data merupakan proses untuk menerjemahkan data mentah menjadi bentuk yang lebih berguna untuk mendapatkan wawasan dari data tersebut. Proses ini biasanya dilakukan oleh tim *data scientists* dan *data engineers*. Dalam submateri ini, kita akan membahas mengenai konsep pengolahan data.

Seperti telah dibahas sebelumnya, pengolahan data terdiri dari kegiatan-kegiatan yang diperlukan untuk mengubah data menjadi sebuah informasi dengan berbagai alat bantu, termasuk alat manual seperti pensil dan kertas, alat mekanis seperti lemari arsip, serta alat elektro mekanis seperti mesin tik, kalkulator, dan komputer. Saat ini proses pengolahan data umumnya dilakukan dengan memanfaatkan komputer sebagai alat bantu. Terlepas dari jenis peralatan yang digunakan, kegiatan pengolahan data dapat dikelompokkan dalam lima kategori besar seperti yang ditunjukkan pada gambar berikut.



Sumber:

Gambar 1. Aktivitas pengolahan data

a. Pengumpulan Data

Data dapat berasal dari berbagai sumber dan dikumpulkan dengan berbagai metode, seperti data temperatur air yang didapatkan dari sensor, data tentang kepuasan pelanggan dari sebuah kuesioner, data karakter tanaman dari observasi lapangan, data pegawai dari basis data kepegawaian sebuah perusahaan, dan masih banyak serta beragam contoh lainnya. Data mungkin awalnya direkam pada dokumen sumber kertas dan kemudian diubah menjadi bentuk yang dapat dibaca dan diproses oleh mesin. Proses tersebut merupakan proses pengumpulan atau pemerolehan data.

b. Konversi Data

Setelah data dikumpulkan, data tersebut akan diubah dari bentuk aslinya dari dokumen sumber menjadi bentuk yang lebih sesuai untuk diproses. Seperti pada gambar x. proses konversi ini terdiri dari beberapa kegiatan berdasarkan jenis data yang akan diolah. Pertama, data akan dikodifikasikan dengan menetapkan kode identifikasi. Kode tersebut dapat berupa angka, huruf, karakter khusus, atau kombinasi semuanya, misalnya, sebuah barang diberi kode "F-01-123". Kedua adalah pengklasifikasian data dengan mengelompokkan data yang memiliki kesamaan karakteristik ke dalam kategori atau kelompok yang serupa, misalnya, pengelompokan data karya tulis ilmiah sesuai dengan bidang keilmuan. Ketiga adalah verifikasi data, kegiatan ini sangat penting untuk memastikan kelengkapan dan akurasi data. Keempat adalah proses transformasi dimana pada proses data akan diubah ke dalam bentuk lain yang lebih sesuai, misalnya data gambar akan ditransformasikan menjadi matrik untuk kemudian diolah lebih lanjut.

c. Pengolahan Data

Setelah data dikumpulkan dan ditransformasi, data siap untuk diolah lebih lanjut untuk menghasilkan sebuah informasi. Secara umum pengolahan data terdiri dari kegiatan-kegiatan berikut, tergantung pada jenis dan bentuk data:

- **Sorting**

Pada umumnya kita akan lebih mudah bekerja dengan data yang telah disusun dalam urutan yang logis, bahkan terkadang hasil pengurutan juga akan menghasilkan sebuah informasi. Data dapat disusun dalam urutan

naik ataupun menurun disesuaikan dengan kebutuhan. Sebagai contoh pada tabel x, terdapat data jumlah kerja karyawan dalam seminggu yang kemudian diurutkan dari yang paling besar hingga paling kecil untuk mengetahui siapa karyawan dengan jumlah jam kerja terbanyak.

Tabel 1. Contoh proses pengurutan

Nama	Jam Kerja
Adi	40
Budi	38
Didi	36
Edi	35

- *Calculating*

Pada kegiatan ini akan dilakukan pengolahan data secara aritmatika, item data yang direkam dapat ditambahkan, dikurangi, dikalikan, atau dibagi satu sama lain untuk membuat data baru seperti ditunjukkan pada tabel x. Misalnya, dalam menghitung gaji karyawan, jam kerja dikalikan dengan upah per jam menghasilkan gaji kotor lalu dikurangi dengan pemotongan pajak sebesar 10% dari gaji kotor maka akan didapatkan total gaji bersih.

Tabel 1. Contoh proses kalkulasi

Nama	Jam Kerja	Upah per jam	Gaji Kotor	Pajak (10%)	Gaji Bersih
Adi	40	85000	13600000	1360000	12240000
Budi	38	70000	10640000	1064000	9576000
Didi	36	75000	10800000	1080000	9720000
Edi	35	100000	14000000	1400000	12600000

- *Summarizing*

Pada kegiatan ini data akan diringkas menjadi bentuk yang lebih bermanfaat dan sederhana. Dari hal ini kita dapat menarik kesimpulan dari data yang telah diolah sebelumnya dan mendapatkan informasi yang berguna. Misal, jika data dalam bentuk angka maka kita dapat meringkasnya dengan menghitung total data dalam klasifikasi atau dengan memilih data strategis dari sekumpulan data. Sebagai contoh, seperti pada tabel x disajikan kesimpulan penjualan setiap jenis mobil pada seluruh cabang.

Tabel 1. Contoh proses *summarizing*

Cabang	Jenis Mobil	Jumlah Item Terjual
Jakarta	SUV	40
	MPV	50
	Sedan	34
Bogor	SUV	60
	MPV	38
	Sedan	21
Cibinong	SUV	36
	MPV	31
	Sedan	15
Serpong	SUV	35
	MPV	18
	Sedan	49

- *Comparing*

Membandingkan data dapat juga berarti melakukan evaluasi sehubungan dengan beberapa ukuran yang diketahui. Sebagai contoh, manajer perusahaan otomotif membandingkan data penjualan mobil pada setiap cabang atau membandingkan jenis mobil apa yang paling diminati oleh konsumen.

d. *Pengelolaan Hasil Keluaran*

Setelah data diolah, hasil keluaran tersebut akan dikelola untuk memastikan data tersimpan dengan baik dan dapat diakses untuk dapat digunakan kembali. Hasil keluaran dari setiap proses sebelumnya sangat penting untuk disimpan secara terorganisir agar dapat diakses dan digunakan di masa mendatang apabila dibutuhkan. Mekanisme penyimpanan untuk sistem pengolahan data dapat berupa lemari arsip dalam sistem manual, dan perangkat elektronik seperti USB, Hard Disk, atau Cloud dalam sistem berbasis komputer.

e. *Penyampaian Hasil Keluaran*

Berikutnya hasil keluaran dan informasi yang didapatkan dari proses pengolahan data tersebut perlu disampaikan kepada para pemangku kepentingan atau ke sistem pemrosesan data lain. Penyampaian dapat dilakukan secara langsung melalui komunikasi tatap muka maupun melalui media lainnya seperti email, poster, dokumen, maupun lainnya.

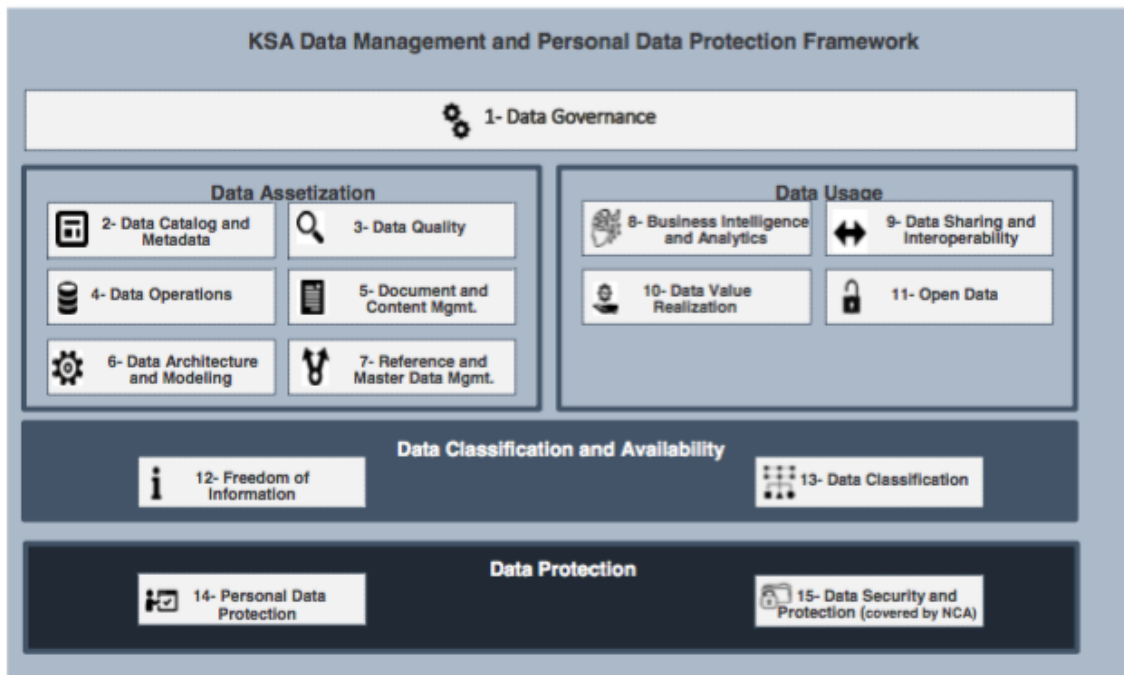
f. Manajemen Data

Dalam konteks jabatan fungsional analisis data ilmiah, manajemen data merupakan salah satu tahapan dalam siklus analisis data ilmiah. Manajemen data adalah aktifitas yang berupaya untuk memastikan data dapat digunakan secara tepat oleh pihak yang memiliki otoritas mulai dari tahap perencanaan sampai dengan evaluasi. DAMA sebuah organisasi yang melakukan pengelolaan data mendefinisikan data manajemen sebagai berikut :

“The process of developing and executing plans, policies, initiatives, and practices to enable entities to manage and govern their data and achieve the aspired value, with data considered an organizational asset” (National Data Management Office (NDMO), 2021)

Definisi tersebut menggambarkan data manajemen sebagai suatu pengembangan dan pelaksanaan perencanaan, kebijakan, inisiatif, dan praktek yang memungkinkan setiap pihak dalam organisasi untuk menata kelola data secara tepat. Adapun jika digambarkan dalam bentuk framework, maka manajemen data dapat digambarkan sebagaimana ilustrasi berikut :

Figure 1: KSA Data Management and Personal Data Protection Framework



DAMA (2021)

1. Data Governance: Data Governance merupakan suatu mekanisme untuk menentukan otoritas dan kendali suatu data melalui perencanaan dan implementasi kebijakan praktik manajemen data yang meliputi sumber daya manusia, proses, dan teknologi untuk menghasilkan penanganan asset data organisasi secara tepat dan sesuai dengan prinsip perlindungan data pribadi. [SEP]
2. Data Catalog and Metadata: Data katalog dan metadata memungkinkan individu dalam organisasi untuk memiliki akses yang efektif. Akses terhadap metadata didukung melalui metadata yang terintegrasi. Data katalog berperan sebagai alat untuk menjadi referensi utama bagi dalam melengkapi metadata. [SEP]
3. Data Quality: kualitas data memiliki focus pada peningkatan kualitas dari asset data yang dimiliki organisasi. Kualitas data memastikan bahwa data yang tersedia untuk kebutuhan pengguna telah memenuhi standard kualitas tertentu yang menjadi harapan pengguna.
4. Data Operations: The Data Operations memiliki tujuan sebagai desain, implemementasi dan dukungan terhadap tempat penyimpanan data untuk memaksimalkan nilai dari suatu data pada suatu siklus data sejak mulai data dibuat sampai dengan data dimusnahkan. [SEP]

5. Document and Content Management: Proses ini meliputi proses penciptaan, penyimpanan, akses dan penggunaan dokumen dan konten yang tersimpan diluar dari database relasional.
6. Data Architecture and Modelling: Arsitektur data dan modeling berfokus pada pendirian struktur data formal dan aliran data yang memungkinkan pemrosesan data dari hulu ke hilir yang dalam suatu entitas maupun antar entitas.
7. Reference and Master Data Management: referensi data dan master manajemen data memungkinkan untuk menghubungkan data penting kepada file tunggal utama, menyediakan poin umum untuk Reference atas seluruh data penting. [1-1] [SEP]
8. Business Intelligence and Analytics: Business Intelligence and Analytics berfokus pada analisis terhadap data yang dimiliki organisasi untuk dapat melakukan ekstraksi insight dan membuat kesimpulan terhadap informasi yang belum pernah terungkap
9. Data Sharing and Interoperability: berbagi data dan interoperabilitas meliputi pengumpulan data dari berbagai sumber dan integrasi berbagai solusi untuk mendorong komunikasi internal dan eksternal secara harmonis antara berbagai komponen teknologi informasi. Berbagi data dan interoperabilitas juga mencakup proses berbagi data yang memungkinkan terjadinya proses pertukaran data yang terstandar dan terorganisir antar entitas. [1-1] [SEP]
10. Data Value Realization: Data Value Realization mencakup evaluasi atas asset data yang berpotensi untuk dapat digunakan untuk menghasilkan pendapatan ataupun menurunkan biaya operasional suatu organisasi.
11. Open Data: Open data berfokus pada upaya untuk menyiapkan data yang dapat dimanfaatkan oleh public untuk kepentingan peningkatan transparansi, mengakselerasi inovasi, dan mendorong pertumbuhan ekonomi.
12. Freedom of Information: Freedom of Information berfokus pada penyediaan akses public terhadap informasi yang dimiliki pemerintah.
13. Data Classification : Data Classification meliputi kategorisasi data sehingga data dapat digunakan dan dilindungi secara efisien. Klasifikasi data dibuat berdasarkan penilaian dampak yang berpotensi terjadi karena kesalahan penanganan ataupun akses oleh pihak yang tidak memiliki otoritas akses.
14. Personal Data Protection: Personal Data Protection merupakan upaya perlindungan identitas subyek dan penutupan informasi pribadi yang memiliki

resiko untuk disalahgunakan.

15. Data Security and Protection: Data Security and Protection merupakan proses, manusia dan teknologi yang didesain untuk memproteksi entitas data termasuk namun tidak terbatas pada otorisasi akses terhadap data, menghindari kebocoran data, dan penjagaan terhadap pihak yang tidak berkepentingan untuk membuka data.

Dalam melaksanakan kegiatan manajemen data, terdapat prinsip yang diacu secara global yang dikenal sebagai prinsip FAIR. Prinsip FAIR digunakan sebagai acuan untuk kemudahan bagi publik untuk menggunakan kembali (reuse) suatu dataset untuk berbagai keperluan lanjutan.

Prinsip FAIR sendiri merupakan akronim dari :

F : Findability, yakni prinsip kemudahan suatu data untuk dapat ditemukan oleh mesin pencari otomatis maupun metode pencarian manual pada suatu sistem database. Untuk mendukung kemudahan ini, kelengkapan metadata merupakan unsur utama yang menjadi titik kritis prinsip Findability. Selain untuk kemudahan pencarian kembali, kelengkapan metadata digunakan agar data secara mudah oleh pengguna karena metadata mampu mendeskripsikan gambaran utuh terkait data mulai dari tahapan penciptaan, pengolahan, sampai dengan penggunaan kembali oleh pihak lain.

A : Accesibility, yakni suatu prinsip yang mencerminkan suatu dataset untuk bisa dengan mudah diakses oleh publik. Untuk dapat memenuhi prinsip ini, suatu dataset wajib untuk dilengkapi dengan metadata yang memiliki standard yang mampu tertelusur oleh dengan baik oleh mesin pencari melalui protokol komunikasi terstandard.

I : Interoperable, suatu prinsip yang mampu dioperasionalisasikan pada berbagai macam platform pengolahan data yang umum atau lazim digunakan oleh analis data. Untuk memenuhi standard ini, suatu dataset harus tersimpan dalam suatu format yang umum sebagaimana termuat dalam tabel pada BAB pengenalan konsep data dibagian awal modul ini.

R : Reusable, suatu prinsip yang mengacu pada kondisi suatu dataset yang dapat digunakan kembali oleh pihak lain. Untuk dapat memenuhi prinsip ini, suatu dataset sedapat mungkin memiliki akses terbuka dengan lisensi data terbuka seperti creative commons sehingga pihak lain yang ingin menggunakan data dapat dengan mudah untuk memanfaatkan kembali suatu dataset dengan terlebih dahulu mengetahui ketentuan penggunaan kembali suatu dataset.

C. Rangkuman

Data merupakan kumpulan fakta dalam berbagai bentuk dan format mentah seperti angka, karakter, image, video, rekaman suara, atau symbol. Kehadiran data dapat membantu dan mempermudah manusia dalam melakukan pemecahan masalah ataupun melakukan pengambilan keputusan. Perkembangan beragam format data membuat kegiatan analisis data ilmiah menjadi semakin kompleks, demikian pula dengan kegiatan manajemen data yang menuntut ketersediaan akses terhadap data.

Data perlu diolah untuk menghasilkan informasi yang bermanfaat. Pengolahan data merupakan proses menerjemahkan data mentah menjadi bentuk yang lebih berguna untuk mendapatkan wawasan dari data tersebut. Proses ini biasanya dilakukan oleh tim *data scientists* dan *data engineers*. Kegiatan pengolahan data meliputi; pengumpulan data, konversi data, pengolahan data, pengelolaan hasil keluaran, penyamaan hasil keluaran dan manajemen data.

D. Evaluasi

1. Jelaskan apa yang dimaksud dengan data!
2. Jelaskan mengapa pengolahan data perlu!
3. Uraikan tahapan proses pengolahan data secara ilmiah!

MATERI POKOK 2: SIKLUS PENGOLAHAN DATA

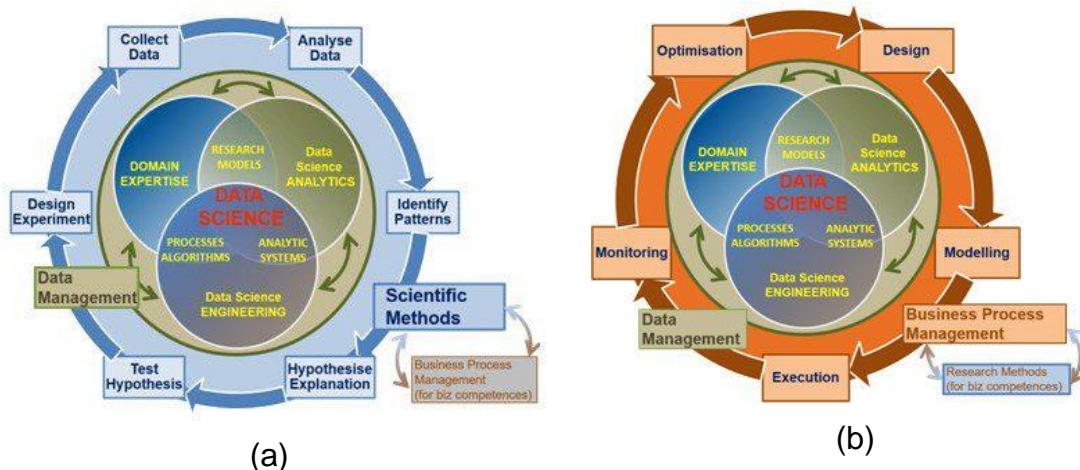
Indikator Hasil Belajar:

Peserta mampu:

Menguraikan secara ringkas siklus pengolahan data

Mempraktikkan siklus pengolahan data

Seperti telah dijelaskan secara singkat pada bab sebelumnya, siklus pengolahan data terdiri dari serangkaian langkah dimana data mentah (*input*) dimasukkan ke dalam sistem untuk menghasilkan wawasan yang dapat ditindaklanjuti (*output*). Setiap langkah memiliki urutan tertentu yang terus berulang secara siklik seperti diilustrasikan pada gambar (a), hasil keluaran siklus pengolahan data pertama dapat disimpan dan dimasukkan sebagai input untuk siklus berikutnya. Pada bab ini kita akan membahas siklus data secara lebih detail yang terdiri dari tujuh tahapan umum yaitu perencanaan manajemen data, pengumpulan data, pengolahan data, analisis data, interpretasi data, dan manajemen data.



Sumber: Edison

Gambar 1. Hubungan antara kelompok kompetensi ilmu data untuk profesi /profil (a) berorientasi umum atau penelitian dan (b) berorientasi bisnis.

A. Perencanaan dan Perancangan

Membuat perencanaan manajemen data merupakan langkah awal dari sebuah proyek dan sangat penting. Perencanaan yang baik akan menghemat banyak waktu, uang, dan energi. Banyak lembaga penyandang dana penelitian mensyaratkan dokumen rencana pengelolaan data untuk diajukan sebagai bagian dari aplikasi penelitian. Pada tahap ini kita banyak hal yang perlu dirancang, antara lain mengidentifikasi kebutuhan data yang akan dikumpulkan atau digunakan untuk menjawab pertanyaan penelitian, bagaimana data akan dikumpulkan, bagaimana data akan diolah dan dianalisis, hingga bagaimana penyimpanan dan pengaturan akses data untuk digunakan kembali.

Secara umum, perencanaan manajemen data meliputi beberapa elemen berikut:

- Tipe data: Data apa saja yang Anda butuhkan atau hasilkan? Apa sumber data Anda? Dalam format apa data Anda? Bagaimana metode Anda dalam mengumpulkan data? Berapa banyak data yang akan Anda kumpulkan atau hasilkan?
- Detail kontekstual (metadata): Bagaimana Anda akan mendokumentasikan dan menjelaskan data Anda?
- Penyimpanan, pencadangan, dan keamanan: Bagaimana dan dimana Anda akan menyimpan dan mengamankan data Anda?
- Ketentuan untuk perlindungan dan privasi: Masalah privasi dan kerahasiaan apa yang harus Anda tangani? Apa peraturan dan kebijakan terkait?
- Kebijakan untuk digunakan kembali: Bagaimana peneliti lain dapat menggunakan data Anda?
- Akses dan berbagi: Bagaimana Anda menyediakan akses data kepada peneliti lain? Bagaimana peneliti lain dapat menemukan data Anda?
- Arsip dan penyediaan akses: Bagaimana rencana Anda untuk menyimpan data dan menyediakan akses jangka panjang?
- Peran dan pengawasan rencana: Siapa yang akan bertanggung jawab atas aspek manajemen data di seluruh proyek, dan sumber daya apa yang diperlukan untuk implementasi?

B. Pengumpulan Data

Secara umum pengumpulan data merupakan proses pemerolehan data dengan berbagai teknik diantaranya, observasi, survei, wawancara baik individual maupun fokus grup, monitor sensor, eksperimen, dan lainnya. Data mentah yang dikumpulkan dapat memiliki berbagai jenis format data, seperti tabular, teks, gambar, suara, dan lain-lain. Data mentah yang dikumpulkan memiliki dampak besar pada validitas dari output yang dihasilkan. Oleh karena itu, penting dilakukan rencana pengumpulan data mulai dari analisis kebutuhan data, analisis sumber data yang jelas dan akurat, serta metode pengumpulan data.

1. Analisis Kebutuhan Data

Hal pertama yang perlu dilakukan pada tahap ini adalah analisis kebutuhan data. Analisis kebutuhan data adalah bagian yang sangat penting dari proses pengolahan dan analisis data. Analisis kebutuhan data serta aspek-aspek yang terdapat didalamnya perlu dipahami dengan baik. Secara umum, proses analisis kebutuhan data menggunakan pendekatan *top-down* yang menekankan pada kebutuhan bisnis atau tujuan penelitian sehingga data yang nanti akan dikumpulkan relevan.

Secara umum, analisis kebutuhan data dilakukan berdasarkan apa yang akan dilakukan dan bagaimana cara melakukannya untuk mencapai tujuan tertentu. Pada tahap ini kita akan mengidentifikasi, menganalisis, mendefinisikan, dan mendokumentasikan terkait kebutuhan data, sumber data, menilai rentang waktu serta sumber daya yang dibutuhkan sesuai dengan tujuan tertentu.

2. Pengumpulan Data

Pada tahap ini, kita akan melaksanakan proses pengumpulan data sesuai dengan perencanaan yang telah disusun sebelumnya. Terdapat beberapa metode pengumpulan data yang umum digunakan, antara lain:

- Wawancara dan *Focus Group Discussion*, dengan melakukan diskusi secara langsung dengan subjek tentang topik atau masalah tertentu. Melalui metode ini, kita dapat mengumpulkan data kualitatif dan kuantitatif

berdasarkan tanggapan dari audiens atas pertanyaan-pertanyaan yang diajukan.

- Survei, dengan menyebarkan kuesioner baik fisik maupun digital kepada target peserta. Metode ini memberikan gambaran tentang pengalaman, pengetahuan, pendapat, dan harapan dari para peserta tentang suatu hal.
- Observasi, dengan melakukan pengamatan langsung tentang suatu fenomena. Secara umum, melalui observasi kita dapat menemukan realita dan dinamika suatu keadaan yang tidak dapat ditemukan dengan metode lainnya.
- Eksperimen, dalam kasus tertentu perlu dilakukan eksperimen terlebih dahulu untuk mendapatkan data yang dibutuhkan. Misalnya, dalam kegiatan penelitian efisiensi suatu obat, para peneliti perlu melakukan uji klinis terhadap para subjek untuk mengukur dampak obat pada kelompok tertentu.
- Analisis konten dan dokumen, terkadang kita dapat mengumpulkan banyak data melalui konten dan dokumen yang tersedia dari sumber tertentu atau kegiatan yang telah selesai.

C. Persiapan Data

Akurasi informasi dari hasil pemrosesan data akan sangat bergantung pada kualitas data. Data mentah seringkali tidak lengkap, terdapat duplikasi, memiliki format yang tidak konsisten antar data, dan terdapat data anomali. Oleh karena itu, perlu dilakukan proses persiapan data sebelum ke tahap selanjutnya. Persiapan data merupakan proses pemilahan atau penyaringan data mentah. Tujuan dari langkah ini untuk menilai tingkat kelengkapan dan akurasi data, memperbaiki kesalahan dan melengkapi nilai yang hilang, pemberian label (anotasi), serta menghilangkan inkonsistensi pada data sehingga dapat digunakan untuk menyusun informasi berkualitas tinggi. Berikut merupakan beberapa kegiatan yang dilakukan pada tahap persiapan data:

1. Data Cleaning

Seperti disebutkan sebelumnya, dalam melakukan pengolahan data seringkali kita jumpai masalah kualitas data seperti nilai yang hilang (*missing value*), duplikasi,

inkonsistensi, dan data anomali. Permasalahan kualitas data tersebut perlu ditangani dengan baik dan dapat diselesaikan dengan beberapa teknik, antara lain:

- *Missing value*, terjadi ketika terdapat nilai hilang atau tidak lengkap pada dataset. masalah ini dapat langsung ditangani dengan kembali mengumpulkan lebih banyak data dan informasi, namun jika tidak memungkinkan, kita dapat menggunakan pengetahuan domain atau algoritma yang ada untuk mengisi nilai yang hilang berdasarkan korelasi antar atribut.
- Duplikasi, terjadi ketika terdapat lebih dari satu data yang sama persis. Kita dapat mengindeks data berdasarkan atribut tertentu untuk mengetahui duplikasi dan menghapusnya.
- Inkonsistensi, terjadi ketika nilai atau format pada dataset tidak konsisten.
- *Outlier*, terjadi jika terdapat nilai yang berbeda atau anomali dari data mayoritas. Jika dibutuhkan kita dapat melakukan koreksi dengan beberapa teknik pada data tersebut atau dalam beberapa kasus data anomali tetap harus disertakan.

2. **Data Integration**

Data yang dikumpulkan dapat berasal dari berbagai sumber. Pada tahap ini akan dilakukan penggabungan data dari berbagai sumber tersebut menjadi satu kesatuan. Dalam penggabungan data tersebut kita perlu memastikan seluruh atribut data mempunyai format yang sama.

3. **Data transformation**

Setelah data terintegrasi menjadi satu, langkah berikutnya merupakan *data transformation*. Dalam tahapan ini, kita dapat mengubah struktur data, format data, atau nilai data sehingga menghasilkan dataset dalam bentuk yang lebih sesuai untuk proses analisis data. Berikut merupakan beberapa tahapan yang biasa dilakukan pada saat *data transformation*:

- *Aggregation*, dengan menggabungkan seluruh data dalam format yang seragam
- *Normalization*, dengan mengubah nilai pada data ke dalam skala tertentu

- *Discreditization*, dengan mengelompokkan data ke dalam interval yang lebih kecil. Misalnya, terdapat
- *Concept hierarchy generation*, dengan menambahkan hirarki baru pada dataset. Misalnya, dalam terdapat label burung, ayam, dan bebek maka kita dapat menambahkan label unggas sebagai hirarki baru untuk menggabungkan ketiganya.

4. **Data Selection**

Pada tahapan ini, data yang telah ditransformasi akan dipilah berdasarkan variabel atau fitur yang paling penting dalam meraih tujuan, maka kita dapat berfokus dan mengambil sampel pada data yang paling relevan. Terdapat tiga teknik yang dapat diterapkan dalam melakukan *data selection*, yaitu:

- *Attribute selection*, dengan mengkombinasikan fitur atau label, sebagai contoh terdapat label pasien dan jenis kelamin “pria/wanita” maka kita dapat menggabungkan label tersebut menjadi pasien pria/pasien wanita.
- *Numerosity selection*, dengan merepresentasikan data sebagai sebuah model seperti model regresi.
- *Dimensionality reduction*, dengan melakukan pengurangan jumlah fitur yang berulang pada dataset.

D. **Pemrosesan dan Analisis Data**

Pada tahap ini, data akan diproses untuk mengekstrak informasi dan dapat dianalisis lebih lanjut melalui proses pengorganisasian, pengindeksan, dan pemrosesan data. Informasi disini berarti hasil identifikasi hubungan dan pola penting dari data yang dapat membantu memecahkan permasalahan atau mencapai tujuan tertentu. Dalam sejarahnya, kemampuan dan efisiensi pemrosesan data secara manual bertahap digantikan oleh mesin dan komputer.

Banyak metode yang dapat digunakan pada langkah ini untuk menghasilkan keluaran yang diinginkan, misalnya, pemrosesan data pengenalan jenis tanaman memerlukan teknik klasifikasi, pemrosesan data iklim memerlukan analisis deret waktu, dan lain sebagainya. Metode pemrosesan data mengacu pada teknik dan algoritma yang digunakan untuk mengekstraksi informasi dari data. Berikut merupakan beberapa metode yang sering digunakan.

- Klasifikasi, adalah metode yang digunakan mengelompokkan data yang memiliki kesamaan karakteristik ke dalam kategori atau kelompok yang serupa, biasanya data yang digunakan adalah data kategorikal. Pengklasifikasian membutuhkan data latih untuk dapat membentuk sebuah model klasifikasi. Data dilatih menggunakan data yang telah diberi label oleh para ahli, sehingga data berikutnya yang belum dikategorikan dapat dikategorikan secara otomatis berdasarkan model pembelajaran mesin dari data latih. Sebagai contoh, identifikasi jenis cabai, sistem memerlukan data latih berupa karakteristik dan gambar dari setiap jenis cabai untuk dapat mengenali jenis cabai secara otomatis.
- Regresi, adalah metode untuk mempelajari hubungan antara suatu variabel terikat dengan variabel bebas lainnya, biasanya digunakan untuk data numerik. Hubungan tersebut dapat digunakan untuk memprediksi hasil di masa mendatang. Sebagai contoh, dampak polusi udara terhadap prediksi sinar matahari.
- *Clustering*, adalah metode untuk menemukan kelompok data yang berbeda berdasarkan karakteristiknya. Berbeda dengan klasifikasi yang membutuhkan data latih, *clustering* merupakan jenis pembelajaran tanpa pengawasan (*unsupervised learning*) dan dapat digunakan baik untuk data kualitatif maupun kuantitatif. Sebagai contoh, analisis minat dan bidang kepakaran para peneliti berdasarkan karya tulis ilmiah yang dihasilkan.
- *Association Rule Mining*, adalah metode untuk menemukan hubungan antar variabel yang sering terjadi bersamaan dalam dataset. Sebagai contoh,
- Analisis deret waktu, adalah sekumpulan metode untuk mendeteksi trend dan pola dari data *time series*. Data time series merupakan kumpulan data yang diindeks dengan waktu, data jenis ini dapat digunakan diberbagai domain. Sebagai contoh,

E. Penyampaian Hasil Analisis

Pada tahapan ini hasil dari tahap pemrosesan data akan ditransmisikan dan ditampilkan kepada pengguna dalam bentuk lain yang lebih mudah dipahami seperti grafik, tabel, file vektor, audio, video, dokumen, dll. Informasi baru tersebut

yang kemudian akan disampaikan dan dianalisis lebih lanjut untuk mendapatkan sebuah wawasan bagi pemangku kepentingan. Penyampaian dapat dilakukan secara langsung melalui komunikasi tatap muka maupun melalui media lainnya seperti email, poster, dokumen, maupun lainnya.

F. Manajemen Data

Tahap terakhir dari siklus pengolahan data adalah manajemen data yang mencakup penyimpanan, pengaturan akses, ketentuan untuk perlindungan, kebijakan penggunaan kembali, dll. Pada tahap ini, data dan hasil keluaran proses pengolahan data disimpan dalam format yang dapat diakses secara cepat kapanpun dibutuhkan, dan juga memungkinkan untuk digunakan sebagai input dalam siklus pengolahan data lainnya.

G. Rangkuman

Siklus pengolahan data terdiri dari serangkaian langkah dimana data mentah (*input*) dimasukkan ke dalam sistem untuk menghasilkan wawasan yang dapat ditindaklanjuti (*output*). Setiap langkah memiliki urutan tertentu yang terus berulang secara siklus. Secara detail, siklus pengolahan data terdiri dari 6 (enam) tahapan umum yaitu, (i) perencanaan dan perancangan, (ii) pengumpulan data, (iii) pengolahan data, (iv) pemrosesan dan analisis data, (v) penyampaian hasil analisis/interpretasi data, dan (vi) manajemen data.

Perencanaan merupakan awal tahapan yang sangat penting antara lain karena mengidentifikasi kebutuhan data yang akan dikumpulkan atau digunakan untuk menjawab pertanyaan penelitian, bagaimana data akan dikumpulkan, bagaimana data akan diolah dan dianalisis, hingga bagaimana penyimpanan dan pengaturan akses data untuk digunakan kembali.

H. Evaluasi

1. Coba uraikan secara ringkas urutan tahapan pengolahan data !
2. Case Study. Pengajar memberi tugas case study untuk mempraktekkan proses pengolahan data.

MATERI POKOK 3:

Track Kegiatan Analisis Data Ilmiah

Indikator Hasil Belajar:

Peserta mampu:

Menjelaskan pentingnya Jabfung Analisis Data Ilmiah

Menjelaskan perbedaan ruang lingkup kegiatan Data Analisis, Data Engineer, dan Data Management

Penyusunan jabatan fungsional analisis data ilmiah disusun dari beberapa spesialisasi atau track berdasarkan peran dan tanggung jawab dari masing – masing tahapan analisis data ilmiah. Dalam penentuan track jabatan fungsional analisis data ilmiah, acuan yang digunakan adalah *Education for Data Intensive Science to Open New Science Frontiers (EDISON)*. EDISON sendiri merupakan project yang didanai oleh Horizon 2020 yang bertujuan untuk menyediakan framework yang dapat digunakan sebagai panduan oleh para pemangku kepentingan pada berbagai bidang yang secara intensif memanfaatkan data dalam berbagai proses bisnisnya (Demchenko, 2017).

Pada perguruan tinggi, EDISON Framework dapat dimanfaatkan untuk mendesain kurikulum dan mata kuliah yang hendak diajarkan. Pada sektor industri atau bisnis, EDISON Framework dapat digunakan untuk menentukan seperangkat kompetensi dan keterampilan yang spesifik sesuai dengan domain dan karakteristik organisasi pengguna.

Lebih lanjut, EDISON Framework juga menjadi salah satu rujukan dalam penentuan track jabatan fungsional analisis data ilmiah sebagai upaya untuk menentukan peran dan tanggung jawab bagi pemangku jabatan fungsional yang selaras dengan visi dan misi lembaga atau organisasi dimana pemangku jabatan tersebut duduk. Salah satu keluaran spesifik EDISON Framework yakni Data Science Profesional Profiles (DSPP) yang merupakan pengembangan lebih lanjut dari European Skills Competency Qualification and Occupation (ESCO) atau

dalam konteks Indonesia dapat disetarakan dengan Standar Kompetensi Kerja National Indonesia (SKKNI).

DSPP didefinisikan sebagai perpanjangan dari Keterampilan, Kompetensi, Kualifikasi, dan Pekerjaan Eropa (ESCO) menggunakan kelompok klasifikasi teratas ESCO. Definisi DSPP memberikan instrumen penting untuk mendefinisikan efektif struktur organisasi dan peran yang terkait dengan posisi Ilmu Data dan dapat juga digunakan untuk membangun individu jalur karir dan kompetensi yang sesuai dan transfer keterampilan antara organisasi dan sektor.

Berikut merupakan Edison Data Science framework dan interelasinya dengan DSPP yang menjadi rujukan dalam pengembangan Jabatan fungsional analis data ilmiah :

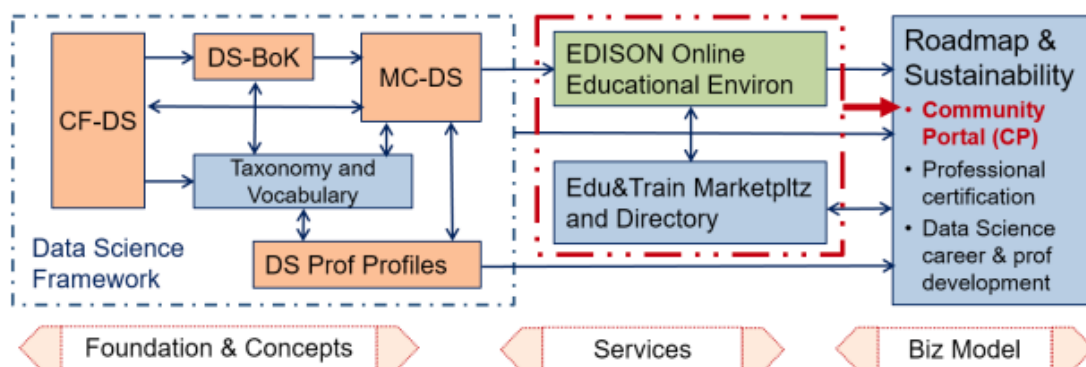


Figure 1 EDISON Data Science Framework components.

Bagan di atas menggambarkan bagaimana interelasi antara Profil Professional Data Science (DSPP) dengan Data Science Competency Framework (CF-DS) beserta dengan Data Science Body of Knowledge serta Data Science – Model Curriculum (MC-DS). Data Science Competency Framework (CF-DS) yang merupakan pondasi atau basis dari keseluruhan pengembangan Edison framework yang meliputi kompetensi inti yang dibutuhkan untuk keberhasilan kerja seorang analis data ilmiah di berbagai sektor industri dan lingkungan kerja melalui serangkaian tahapan jenjang karir. DSPP menjadi instrument penting untuk menentukan struktur organisasi yang efektif dan peran terkait posisi jabatan

ilmuwan data (data scientist) yang kemudian diadopsi pada jabatan fungsional analis data ilmiah. Framework DSPP juga dapat digunakan sebagai acuan dalam membangun jenjang karir individu seseorang serta bagaimana kemudian hubungannya dalam membangun kompetensi tersebut untuk dapat mencapai jenjang karir yang diharapkan.

Jabatan fungsional analis data ilmiah pada lembaga pemerintahan merupakan jabatan fungsional yang mempunyai ruang lingkup, tugas tanggung jawab dan wewenang untuk melaksanakan analisis data ilmiah (Peraturan Menteri Pemberdayaan Aparatur Negara dan Reformasi Birokrasi Nomor 86 Tahun 2020). Jabatan fungsional ini merupakan respon atas kebutuhan profesional pada bidang analisis data pada berbagai lembaga pemerintahan seiring dengan semakin intensifnya pemanfaatan data dalam berbagai pengambilan keputusan pada pemerintahan. Jabatan fungsional ini merupakan adopsi dari Ilmuwan data atau data scientist yang merupakan profesional di bidang data. data scientist sendiri memiliki beragam definisi dan belum ada satu definisi yang cukup solid yang mampu menggambarkan seperti apa data scientist itu sendiri. Namun demikian, untuk menggambarkan data scientist ada satu definisi yang cukup komprehensif yang diambil dari NIST SP1500-1 document yakni :

“A Data Scientist is a practitioner who has sufficient knowledge in the overlapping regimes of expertise in business needs, domain knowledge, analytical skills, and programming and systems engineering expertise to manage the end-to-end scientific method process through each stage in the big data lifecycle, till the delivery of an expected scientific and business value to science or industry.”

Definisi tersebut menggambarkan ilmuwan data adalah seseorang yang memiliki cukup pengetahuan pada lintas disiplin yakni kebutuhan bisnis, dan domain pada ilmu pengetahuan tertentu, kemampuan analisis serta kepakaran pada bidang programming dan sistem engineering. Dengan kompleksnya kemampuan yang dimiliki oleh seorang ilmuwan data, maka peran seorang ilmuwan data dapat berada pada berbagai lini di suatu organisasi atau suatu proyek seperti data engineer, data analis, data arsitek atau data steward (Demchenko, 2017). DSPP sebagai framework yang menggambarkan profil profesional seorang ilmuwan data mengelompokkan grup kompetensi yang kemudian diadopsi pada jabatan

fungsiional Analisis data ilmiah sebagai berikut :

1. Data Science Analytics (Termasuk analisis statistik, machine learning, data mining, business analytics)
2. Data Science Engineering (Termasuk Software dan application engineering, data warehousing, big data infrastructure and tools)
3. Data Management and Governance (Termasuk data stewardship, curation and preservation)

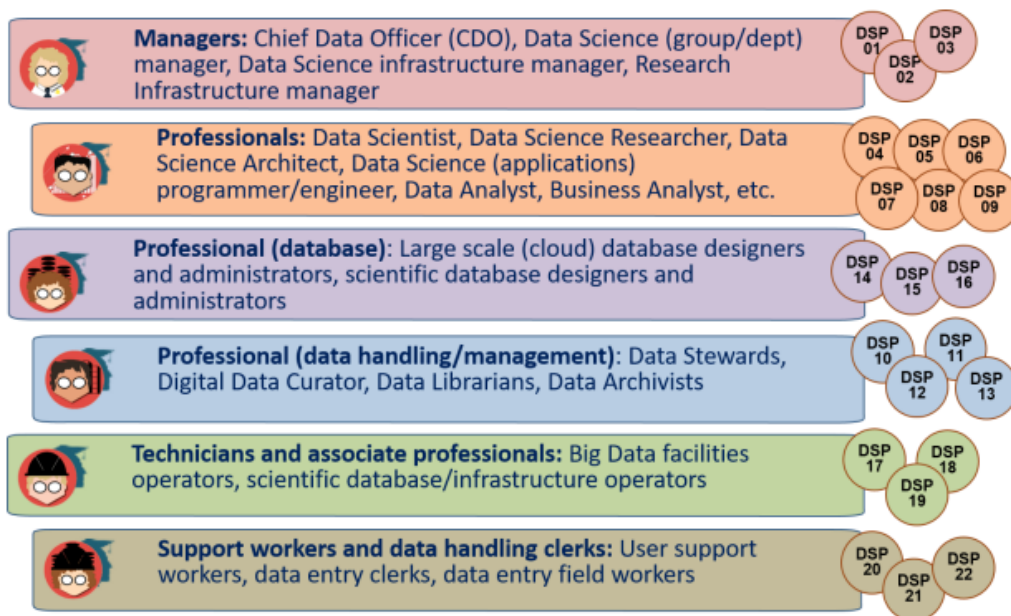


Figure 4. Data Science Professional profiles and their grouping by the proposed new professional groups compliant with the ESCO taxonomy.

Sumber gambar : (Demchenko, 2017)

A. Data Analisis

Sebagai salah satu track dalam jabatan fungsiional analisis data ilmiah, data analisis memiliki peran dan tanggung jawab untuk membuat analisis baik menggunakan teknis analisis statistik, analisis prediksi, pada data yang tersedia untuk menemukan insight baru dan mengungkapkan model atau hubungan baru. Kompetensi yang perlu dimiliki oleh seorang yang menduduki jabatan fungsiional analisis data ilmiah dengan pilihan track data analisis antara lain sebagai berikut :

“Importance of using different statistics and data analytics

techniques for complex datasets, with stress on statistical techniques; development of the specialised tools and visualisation were ranked with less importance (Demchenko, 2017)”

Adapun ruang lingkup pekerjaan seorang yang memilih track data analisis antara lain : business understanding, analisis, modeling, visualisasi, interpretasi sampai dengan menyampaikan insight yang ditemukan kepada penggunanya.

B. Data Engineer

Data engineer merupakan salah satu track yang dapat dipilih oleh seseorang yang menduduki jabatan fungsional analisis data ilmiah dengan ruang lingkup pekerjaan melakukan perancangan, pengembangan, aplikasi analisis dan pengumpulan data, perekayasa fitur, serta penyimpanan data.

Adapun kompetensi yang harus dimiliki oleh seseorang yang memilih pada track Data Engineer antara lain :

- Kemampuan perekayasa aplikasi untuk pengumpulan dan analisis data
- Mengembangkan dan mengimplementasikan storage dan processing untuk big data seperti (Data lakes, hadoop, Hbase, Cassandra, MongoDB)

C. Data Management

Data management merupakan salah satu track dalam jabatan fungsional analisis data ilmiah yang memiliki peran dan tanggung jawab dalam mendesain dan mengembangka kebijakan tata kelola data, mengimplementasikan dan mengevaluasi kebijakan yang telah di laksanakan serta melakukan seleksi dan penilaian terhadap suatu dataset untuk kepentingan preservasi serta untuk kepentingan komersial jika dimungkinkan.

Adapun kompetensi yang dimiliki oleh seorang yang memilih track data management antara lain :

- Mengembangkan dan menerapkan strategi untuk pengumpulan,

pengolahan, kurasi serta penentuan aksesibilitas dat.

- Mengembangkan dan menerapkan standar metadata sesuai dengan kebutuhan bidang / industry.
- Mengimplementasikan lisensi dan perlindungan kekayaan intelektual terhadap suatu dataset.

Adapun bila dituangkan dalam suatu matriks, maka perbandingan kompetensi antara masing – masing track jabatan fungsional sebagai berikut :

Table 3.1. Competences definition for different Data Science competence groups

Data Analytics (DSDA)	Data Science Engineering (DSENG)	Data Management (DSDM)	Research Methods and Project Management (DSRM)	Domain related Competences (DSDK): Applied to Business Analytics (DSBA)
DSDA Use appropriate data analytics and statistical techniques on available data to discover new relations and deliver insights into research problem or organizational processes and support decision-making.	DSENG Use engineering principles and modern computer technologies to research, design, implement new data analytics applications; develop experiments, processes, instruments, systems, infrastructures to support data handling during the whole data lifecycle.	DSDM Develop and implement data management strategy for data collection, storage, preservation, and availability for further processing.	DSRM Create new understandings and capabilities by using the scientific method (hypothesis, test/artefact, evaluation) or similar engineering methods to discover new approaches to create new knowledge and achieve research or organisational goals	DSDK Use domain knowledge (scientific or business) to develop relevant data analytics applications; adopt general Data Science methods to domain specific data types and presentations, data and process models, organisational roles and relations
DSDA01 Effectively use variety of data analytics techniques, such as Machine Learning (including supervised, unsupervised, semi-supervised learning), Data Mining, Prescriptive and Predictive Analytics, for complex data analysis through the whole data lifecycle	DSENG01 Use engineering principles (general and software) to research, design, develop and implement new instruments and applications for data collection, storage, analysis and visualisation	DSDM01 Develop and implement data strategy, in particular, in a form of data management policy and Data Management Plan (DMP)	DSRM01 Create new understandings by using the research methods (including hypothesis, artefact/experiment, evaluation) or similar engineering research and development methods	DSBA01 Analyse information needs, assess existing data and suggest/identify new data required for specific business context to achieve organizational goal, including using social network and open data sources

D. Rangkuman

Jabatan fungsional analis data ilmiah merupakan jabatan yang mempunyai ruang lingkup, tugas tanggung jawab dan wewenang untuk melaksanakan analisis data ilmiah. Urgensi pembentukan jabatan ini disebabkan adanya kebutuhan professional pada bidang analisis data pada berbagai lembaga pemerintahan seiring dengan semakin intensifnya pemanfaatan data dalam pengambilan keputusan pada pemerintahan.

Jabatan Fungsional Analis Data Ilmiah dikelompokkan dalam 3 (tiga) Kompetensi, yakni (i) Data Analis (ii) Data Engineer. dan (iii) Data Management.

E. Evaluasi

1. Jelaskan alasan mengapa Jabatan Fungsional Analis Data Ilmiah semakin penting belakangan ini !
2. Jelaskan perbedaan ruang lingkup kegiatan utama antara Data Analis, Data Engineer, dan Data Management !

MATERI POKOK 4:

KEGIATAN / TUGAS ANALIS DATA ILMIAH

Indikator Hasil Belajar:

Peserta mampu:

Menjelaskan garis besar kegiatan/tugas Analis Data Ilmiah

Melaksanakan tugas-tugas Analis Data Ilmiah

A. MELAKUKAN PERENCANAAN AKUISISI DATA

Pemanfaatan *big data* di berbagai sektor saat ini meningkat pesat seiring semakin tingginya kebutuhan pengambilan keputusan berbasis data. Salah satu tahapan yang dilakukan untuk pemanfaatan *big data* adalah bagaimana mekanisme dan metode untuk melakukan akuisisi data yang tersedia di berbagai platform digital. Perencanaan akuisisi data adalah suatu kegiatan yang mencoba untuk menyusun rencana akuisisi data untuk memenuhi kebutuhan strategis organisasi dari berbagai sumber dengan menggunakan metode tertentu. Rencana akuisisi data dibuat sebelum melakukan akuisisi data dengan memuat beberapa unsur paling tidak rencana metode, waktu, tools, serta biaya (Peraturan Kepala LIPI nomor 7 Tahun.....)

B. MELAKUKAN AKUISISI DATA

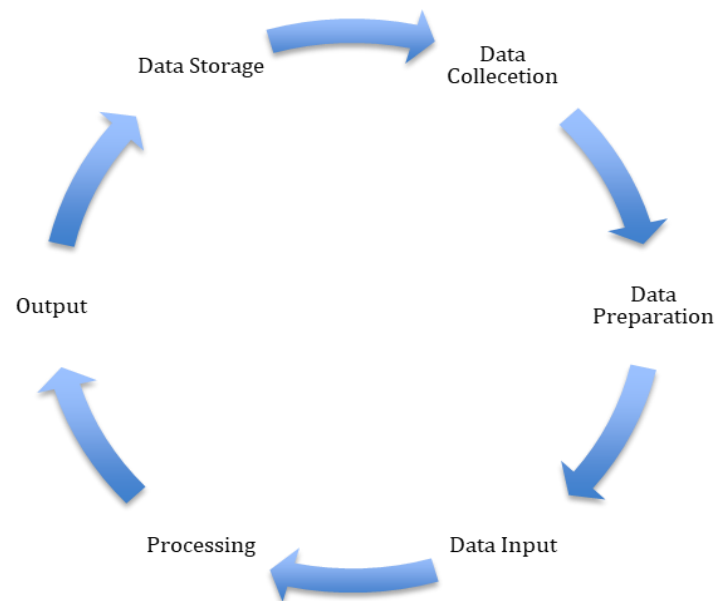
Akuisisi data merupakan suatu aktivitas pengambilan data dari berbagai sumber dan metode dengan tetap memperhatikan lisensi dan ketentuan perundang – undangan yang berlaku. Akuisisi data mencakup aktivitas pembuatan query ke beragam database.

C. MELAKUKAN PEMROSESAN DATA

Pemrosesan data merupakan suatu kegiatan pengalihbentukan suatu data mentah menjadi data yang dapat digunakan untuk keperluan lebih lanjut. Menurut OECD, pemrosesan data merupakan :

“The operation performed on data in order to derive new information according to a given set of rules.”

Pemrosesan data mencakup aktivitas pengumpulan, tabulasi, manipulasi, dan persiapan output data dan metadata. Jika digambarkan dalam diagram, maka pemrosesan data sebagai berikut :



1. Data collection merupakan tahapan pertama dalam kegiatan pemrosesan data yang bersumber dari beragam sumber data. tahapan ini merupakan tahapan yang cukup kritis yang menentukan kualitas keluaran data yang akan diproses. Seleksi terhadap sumber data yang terpercaya menjamin kualitas data yang akan diproses untuk keperluan analisis lebih lanjut.
2. Pada tahapan data preparation, data yang tidak lengkap, tidak memiliki nilai akan di eliminasi. Tahapan ini bertujuan untuk memastikan data sudah bersih dan siap untuk tahapan pemrosesan.
3. Data input merupakan tahapan memasukan data kedalam sistem atau aplikasi pemrosesan. Pada umumnya proses input data membutuhkan intervensi sumber daya manusia.
4. Pemrosesan data. pada tahapan ini data yang sudah melalui tahapan pra pemrosesan data untuk memastikan kualitas data di rubah bentuknya untuk bisa memberikan nilai tambah ataupun untuk kemudahan analisis lanjutan sehingga data menjadi lebih memiliki makna.

5. Output atau keluaran data dapat terdiri dari beragam format dan bentuk seperti text, audio, video, imaget. Keluaran data dari hasil pemrosesan membuat data dapat dioperasionalkan pada sistem operasi yang lain dengan lebih mudah serta dipahami.
6. Data storage merupakan tahapan terakhir dari kegiatan pemrosesan data yakni kegiatan untuk menyimpan data hasil pemrosesan kedalam tempat penyimpanan yang aman. Data storage menjamin aksesibilitas data dimasa yang akan datang dengan melakukan penyimpanan sesuai dengan format yang terstandar untuk kegiatan preservasi data.

D. MELAKUKAN PELESTARIAN DATA

Pelestarian data merupakan upaya untuk menjamin ketersediaan akses untuk jangka panjang. Dalam tahapan manajemen data, pelestarian data merupakan tahapan paling akhir setelah data digunakan untuk keperluan lebih lanjut. Dalam melakukan pelestarian data, seorang analis data melakukan penilaian terhadap suatu data apakah data tersebut layak untuk dilestarikan atau masuk dalam fase berikutnya yakni pemusnahan data.

Secara umum preservasi data terbagi menjadi tiga tahapan sebagai berikut (Pulselli et al., 2021) :

1. Preservasi jangka pendek terhadap data yang masa pakainya sudah dapat diprediksi namun tidak dapat dipastikan aksesibilitasnya di masa yang akan datang
2. Preservasi jangka menengah yakni akses berkelanjutan terhadap data digital melampaui perubahan teknologi dengan batas waktu tertentu
3. Preservasi jangka panjang yakni akses permanen atas data digital

Akses jangka panjang terhadap data merupakan tujuan utama dari preservasi data. ancaman terhadap usangnya teknologi menjadi salah satu ancaman yang harus mampu dimitigasi dan diprediksi oleh seorang analis data ilmiah. Untuk mencapai tujuan tersebut, seorang analis data ilmiah perlu memiliki kerangka kerja untuk tindakan preservasi data. berikut ini merupakan salah satu model kerangka kerja yang lazim digunakan untuk melakukan preservasi data yang diadopsi dari CIFOR, lembaga penelitian bergerak di bidang sosial ekonomi kehutanan.

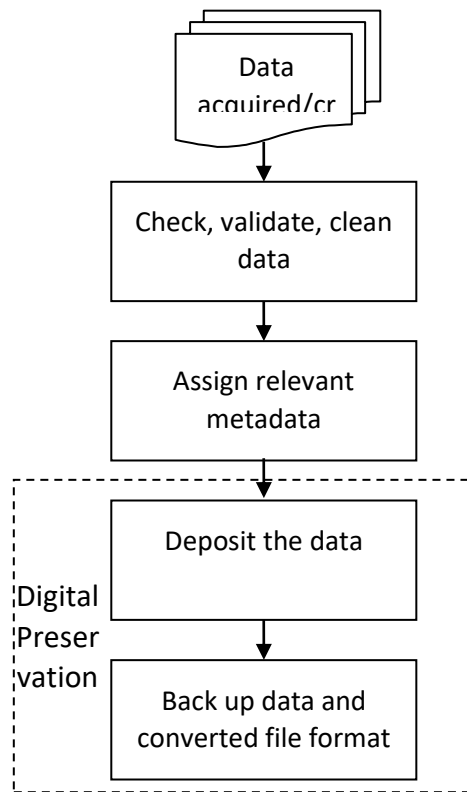


Figure 2. Digital preservation in RDM practice
(Figure modified from CIFOR, 2013)

E. Rangkuman

Jabatan Fungsional Analisis Data Ilmiah secara umum melakukan kegiatan atau tugas-tugas yang dapat dikelompokkan dalam 4 (empat) kegiatan besar, yakni: (i) Melakukan perencanaan akuisisi data, (ii) Melakukan akuisisi data, (iii) Melakukan pemrosesan data, dan (iv) Melakukan pelestarian data.

Pemanfaatan *big data* di berbagai sektor saat ini meningkat pesat. Salah satu tahapan yang perlu dilakukan untuk pemanfaatan *big data* adalah dengan menerapkan mekanisme dan metode yang tepat untuk melakukan akuisisi data yang tersedia di berbagai platform digital.

F. Evaluasi

1. Jelaskan apa yang dimaksud dengan big data dan mengapa perencanaan akuisisi data sangat diperlukan!
2. Jelaskan tahapan atau siklus kegiatan dalam pemrosesan data !

MATERI POKOK 5:

ANALISIS HASIL KERJA ANALIS DATA ILMIAH

Indikator Hasil Belajar:

Peserta mampu:

Mempraktikkan.....

Mempraktikkan.....

Mempraktikkan.....

A. Kesesuaian Rencana Kerja Tahunan

Dalam kegiatan analisis data ilmiah, salah satu keluaran yang akan dihasilkan oleh seorang analis data ilmiah adalah merupakan solusi berbasis data. Untuk mencapai tujuan tersebut, seorang analis data ilmiah membuat suatu rencana pekerjaan dalam satu rentang waktu yang dikenal dengan rencana kerja tahunan. Dalam suatu rencana kerja termuat rencana penggunaan sumber daya yang terdiri dari :

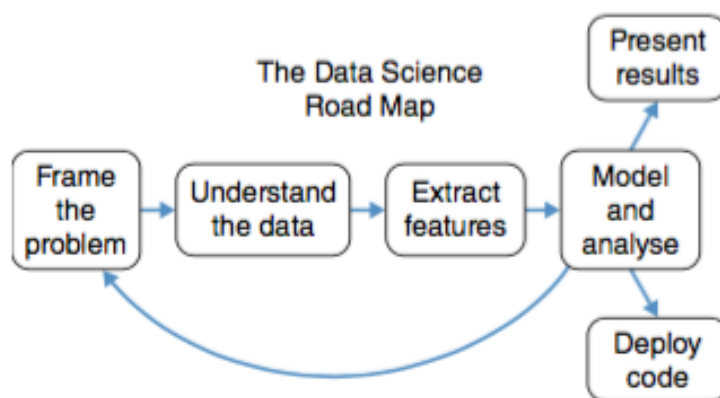
1. Sumber daya manusia
2. Sumber daya teknologi
3. Anggaran
4. Jadwal Pelaksanaan (Timeline)

Dalam melakukan analisis hasil kerja analis data ilmiah, seorang analis data ilmiah perlu merujuk kembali dokumen rencana kerja tahunan sebagai panduan untuk menilai apakah tujuan dari kegiatan analisis data ilmiah telah sesuai dengan tujuan awal pelaksanaan kegiatan analisis data ilmiah. Penilaian dilakukan terhadap kualitas, ketepatan waktu, efektifitas dan efisiensi penggunaan sumber daya.

Berikut salah satu contoh matriks penilaian terhadap hasil kerja analisis data ilmiah :

Tahapan Kegiatan	Kebutuhan SDM	Tools	Anggaran	Rentang Waktu
Pengumpulan data & Data Warehousing	2 Orang Data Engineer	Server	Rp. 3.500.000	35 hours

Rencana kerja tahunan merupakan tahapan perencanaan yang digunakan sebagai panduan untuk memulai kegiatan bagi seorang analis data ilmiah. Kegiatan analisis data ilmiah merupakan upaya untuk menemukan solusi dari suatu permasalahan dengan memanfaatkan big data sehingga solusi ataupun keputusan yang dihasilkan bisa menjadi lebih presisi sesuai dengan konteks permasalahan. Secara umum, roadmap untuk kegiatan analisis data ilmiah tergambar dalam bagan sebagai berikut :



sumber : field cady (2017)

analisis hasil pekerjaan analisis data ilmiah dilakukan terhadap setiap tahapan dalam roadmap tersebut diatas. Roadmap tersebut menunjukkan proses berulang mulai dari pemahaman konteks permasalahan, pemahaman terhadap data yang telah diakuisisi, serta melakukan ekstraksi fitur, serta melakukan modeling dan analisis data dan kemudian melakukan proses pengulangan sampai dengan menemukan insight baru dalam kegiatan analisis data ilmiah (Cady, 2017).

B. Akuisisi data penelitian

Saat melakukan perancangan akuisisi data penelitian, analisis yang dilakukan terhadap hasil kerja akuisisi data penelitian dilakukan dengan menyusun seperangkat pertanyaan umum :

1. seberapa besar dataset?
2. Apakah sudah data set sebelumnya?
3. Apakah data tersebut cukup representatif?
4. Apakah dataset tersebut akan memiliki outliers atau memiliki sumber noise yang luar biasa?
5. Seberapa banyak isian data yang masih kosong?

Dari pertanyaan tersebut kemudian dilakukan penilaian terhadap dataset yang telah berhasil diakuisisi dan disimpan dalam data warehouse.

C. Analisis Data

Setelah data dikumpulkan oleh data engineers dan disimpan pada data warehouse, tahapan selanjutnya adalah melakukan analisis dan pemodelan data. Sebagai contoh, pada saat kita ingin membuat solusi untuk user interface, meskipun kita memiliki akses terhadap corpus data pada twitter, google, facebook, dan aplikasi lainnya, namun apakah kita bisa melakukan analisis untuk menyimpulkan perilaku manusia dalam berhubungan dengan interface suatu aplikasi. Oleh karena itu, analisis data diperlukan untuk mengetahui secara komprehensif terkait perilaku pengguna aplikasi (Schutt & Cathy O'Neil, 2013).

D. Rangkuman

Berisi uraian..... dst.....
se.....
.....
.....
.....
.....
.....
.....
.....

E. Evaluasi

1. Isi dengan evaluasi/latihan/pertanyaan/dsb.
2. Bisa berupa Pertanyaan konsep, proses dan evaluasi, Studi Kasus, dan upa Bahan diskusi lainnya

DAFTAR PUSTAKA

Cady, F. (2017). The Data Science Handbook. In *The Data Science Handbook*.

<https://doi.org/10.1002/9781119092919>

Demchenko, Y. (2017). *EDISON Data Science Framework: Part 4 . Data Science Professional Profiles (DSPP)*. 37.

National Data Management Office (NDMO). (2021). *National Data Management Office Data Management and Personal Data Protection Standards*. January, 1–173.

Pulselli, R. M., Broersma, S., Martin, C. L., Keeffe, G., Bastianoni, S., & van den Dobbelsteen, A. (2021). Future city visions. The energy transition towards carbon-neutrality: lessons learned from the case of Roeselare, Belgium. *Renewable and Sustainable Energy Reviews*, 137. <https://doi.org/10.1016/j.rser.2020.110612>

Schutt, R., & Cathy O'Neil. (2013). Doing data science. In *O'Reilly* (Vol. 51, Issue 12). <https://doi.org/10.5860/choice.51-6803>

LAMPIRAN-LAMPIRAN

Lampiran 1. Judul lampiran.....