KOMPUTASI STATISTIK: FROM DATA TO INSIGHT

YULIAGNIS TRANSVER WIJAYA



Daftar Unit Kompetensi List of Units of Competency

: Data Analyst Okupasi : Data Analyst

	Kode Unit Kompetensi	Judul Unit Kompetensi Title of Competency Unit
No	Cade of Competency Unit	Mengidentifikasi Kebutuhan Pengelolaan Data untuk Proses Bisnis
1	J.62DMS00.001.1	Identifying Data Management Records
	J.62DMI00.004.1	Mengumpulkan Data
2	J.62DM100.004.2	Collecting data
	J.62DMI00.005.1	Menelaah Data
3		Analyzing Data
	J.62DMI00.006.1	Memvalidasi Data
4		Validating Data
	Consumation of the Consumation o	Menentukan Objek Data
5	J.62DMI00.007.1	Defining Data Objects
		Membuat Business Intelligence
6	J.62DMS00.015.1	Creating Business Intelligence
		Creding business mensions
7	M.70BDA00.013.1	Menyusun Laporan Hasil Analisis
		Preparing Analysis Results Report

LANGKAH KERJA

- Identifikasi kebutuhan data
- 2. Mengambil data
- 3. Mengintegrasikan data
- 4. Menelaah data
- Memvalidasi data
- 6. Menentukan objek data
- 7. Membuat business intelligence
- 8. Menyusun laporan hasil analisis

1. IDENTIFIKASI KEBUTUHAN DATA

Bagian ini terkait dengan business understanding yaitu tentang

- apa masalah yang akan dicari solusinya
- apa tujuan proyek/penelitian
- apa saja data yang dibutuhkan
- darimana data diperoleh (sumber pengumpulan data)
- pemahaman tentang jenis data yang dikumpulkan apakah data kontinyu atau diskrit, time series atau cross section, numeric atau alphanumeric (string/text)
- pemahaman tentang proses pengolahan dan analisis data, kebutuhan data juga harus disesuaikan dengan rencana proses pengolahan dan analisis data

2. MENGAMBIL DATA

Metode dan tools dalam pengambilan data dapat dilakukan melalui:

- Observasi
 - Tools yang digunakan bisa berupa kamera, perekam suara, counter, dan berbagai alat sensor
- Wawancara (tatap muka langsung, bisa lewat telpon atau online)
 Tools yang digunakan bisa kuesioner, perekam suara, online form, tablet, laptop, dll.
- Kuesioner, bisa berupa kuesioner fisik ataupun berbasis online seperti Google Form, SurveyMonkey, atau bisa mengembangkan aplikasi sendiri.
- Dokumentasi
 - Dokumentasi bisa dilakukan secara manual dari dokumen fisik ataupun mengambil dokumen dari internet misal web scrapping. Tools yang digunakan bisa berupa scanner, OCR, web scraper.

3. MENGINTEGRITASKAN DATA

- Integritas data merupakan konsep yang mengacu pada data yang utuh dan akurat, yang mencakup aspek perlindungan data dari kerusakan, perubahan yang tidak sah, atau akses yang tidak diizinkan.
- Beberapa ukuran yang dapat menggambarkan Integritas data diantaranya:
 - Akurasi data, yang menjelaskan seberapa akurat data tersebut.
 - Kelengkapan data, yaitu berapa banyak data yang disimpan dalam kumpulan data.
 - Keamanan data, yaitu menjaga keamanan data dari akses ilegal.
 - Tata kelola data, yaitu memastikan data memenuhi kebutuhan organisasi.
 - Validitas data: Memeriksa kesalahan untuk memastikan bahwa data valid.

4. MENELAAH DATA

- Identifikasi tipe data
 - Skala pengukuran: nominal, ordinal, interval, rasio
 - Data numerik dan data kategorik
 - Pemeriksaan missing value(s), konsistensi data, dan outlier(s)
- Pada data tugas, terdapat 10 variabel, yaitu Serial Number, List Year, Date Recorded, Town (nominal), Address (nominal), Assessed Value (rasio), Sale Amount (rasio), Sales Ratio (rasio), Property Type (nominal), Residential Type (nominal)
 - Terdapat missing value(s) pada data Assessed Value, Sale Amount, Property Type, dan Residential Type

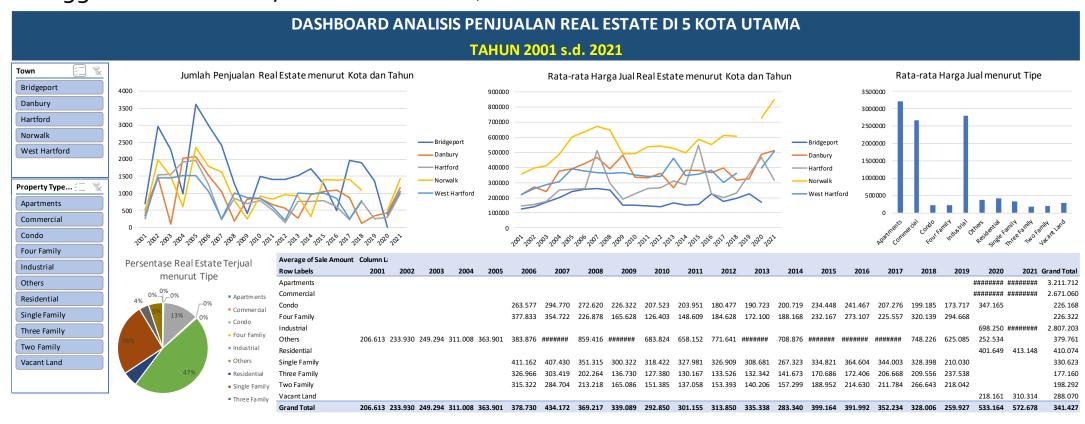
5. MEMVALIDASI DATA

- Proses validasi data melibatkan pemeriksaan tentang:
 - Ketepatan Data (Accuracy), yaitu memeriksa apakah data yang dimasukkan atau tercatat dalam sistem benar dan sesuai dengan fakta atau kejadian yang sebenarnya.
 - Konsistensi Data, yaitu menjamin bahwa data konsisten di seluruh sistem atau dataset.
 - Kelengkapan Data, yaitu memeriksa apakah semua data yang diperlukan sudah lengkap.
 - Keunikan Data, yaitu memastikan bahwa setiap entitas atau observasi dalam dataset bersifat unik dan tidak ada duplikasi.
 - Format Data, yaitu memeriksa apakah data memiliki format yang benar, termasuk jenis data yang sesuai (misalnya, tanggal seharusnya bertipe date, dan bukan string).

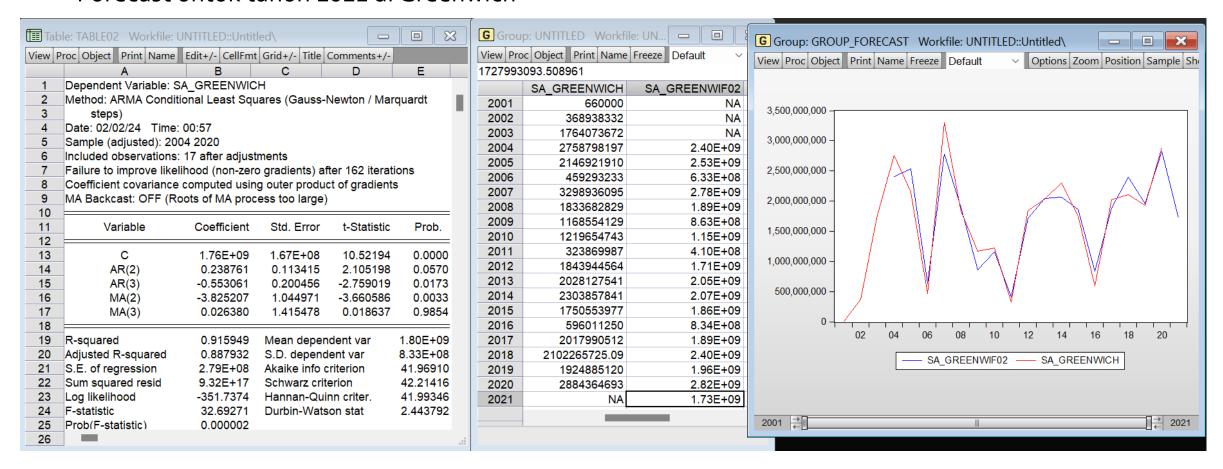
6. MENENTUKAN OBJEK DATA

- Kriteria pemilihan data mempertimbangkan
 - tujuan utama penelitian
 - relevansi dengan topik penelitian
 - kualitas data
 - ketersediaan data
 - cakupan wilayah
 - rentang waktu data

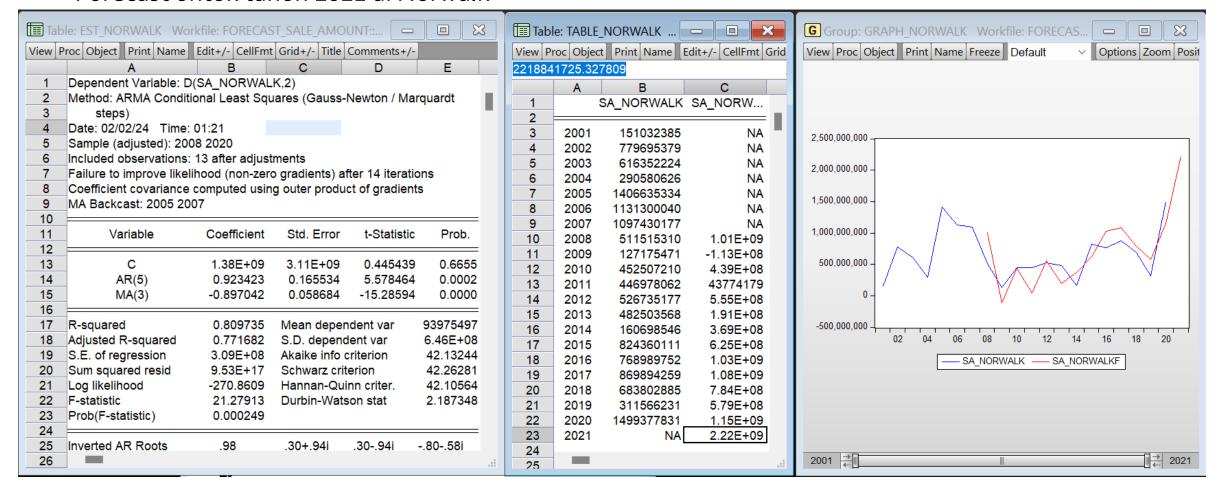
 Membuat rancangan visualisasi, dapat terdiri atas grafik dan tabel (dapat dibuat secara sederhana menggunakan software, contoh: Ms-Excel)



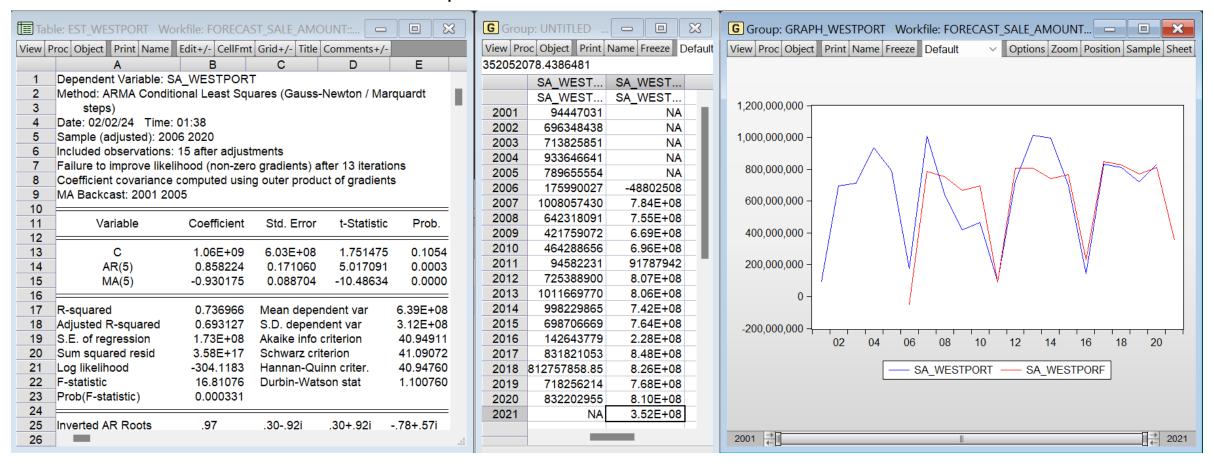
Forecast untuk tahun 2021 di Greenwich.



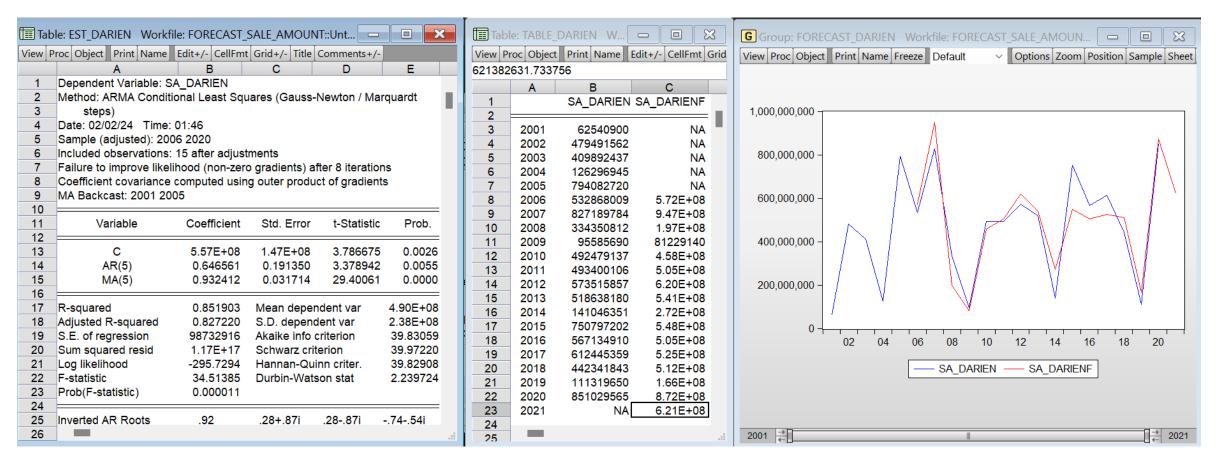
Forecast untuk tahun 2021 di Norwalk



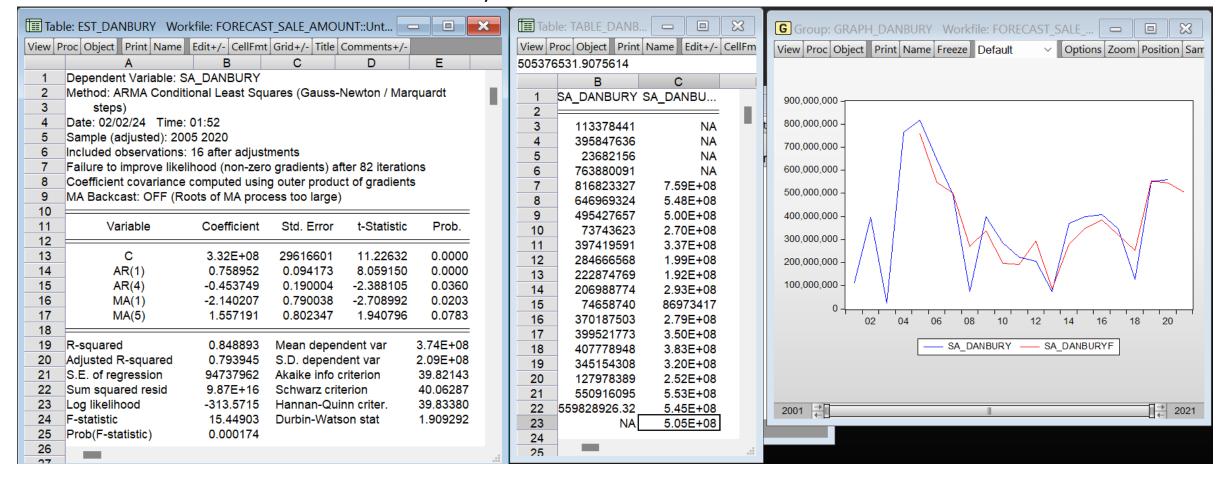
Forecast untuk tahun 2021 di Westport



Forecast untuk tahun 2021 di Darien



Forecast untuk tahun 2021 di Danbury



KUIS-PRAKTIKUM:

"Analisis Penjualan Properti di Amerika Serikat Tahun 2001-2021"

Sumber data: https://s.stis.ac.id/data-properti-us

- Data diperoleh/sudah disiapkan oleh stakeholder. Data tersebut merupakan bentuk terstruktur, namun perlu preprocessing. memuat tipe: numerik, kategorik, date, string.
- 2. Download file di https://s.stis.ac.id/data-properti-us

- Identifikasi Atribut Data: Number, Year, Date Record, Town, Address, Assessed Value, Sale Amount, Sales Ratio, Property Type, Residential Type, Non Use Code, Assessor Remarks, OPM remarks, Location.
- 4. Jumlah Record Data: 997.214

1. Identifikasi Kebutuhan Data

Instruksi Kerja:

- 1. Tentukan asset data organisasi dan aspek pengelolaannya
- 2. Lakukan identifikasi tipe model operasional data

- Business understanding (diketahui masalah, tujuan yang ingin dicapai, rancangan Solusi dan tahapan untuk mencapai tujuan)
- 2. Memahami dan mengidentifikasi tipe data
- 3. Diperlukan Preprocessing:
 - a. Missing value -> imputasi
 - b. Data aneh -> perbaiki/hapus
 - c. Outlier (jika ada) -> hapus/trim

2. Mengambil Data

Instruksi Kerja:

1. Tentukan Tools pengambilan data sesuai tujuan teknis data analyst

- 1. Mengunduh data yang akan digunakan
- 2. Jika data yang diunduh berupa file csv, data dibuka menggunakan ms. excel
- 3. Buka file melalui menu Get data → Import file from csv

3. Mengintegritaskan Data

Instruksi Kerja:

1. Integritas data diperiksa sesuai tujuan teknis data analyst

- 1. Memeriksa kelayakan data untuk dianalisis
 - a. Penanggung jawab (kolom remarks)
 - b. Pastikan datanya bisa dipertanggungjawabkan (dari sumber yang dipercaya)

4. Menelaah Data

Instruksi Kerja:

- 1. Lakukan analisis tipe dan relasi data
- 2. Tentukan Karakteristik data yang terkumpul lalu sajikan dengan visualisasi grafik

- Tipe data (nominal, numerik, ordinal)
- Cleaning data
- Relasi data: pola hubungan antar variabel
- 4. Visualisasi data dalam bentuk grafik

5. Memvalidasi Data

Instruksi Kerja:

- 1. Lakukan pengecekan kelengkapan data
- 2. Buatlah rekomendasi kelengkapan data

- 1. Pengecekan kelengkapan (banyak noise? Missing value? Series tidak lengkap?)
- Apa yang dilakukan untuk mengatasi data yang belum lengkap? (minta dilengkapi? Atau mengambil data sejenis dari sumber terpercaya agar dapat dilakukan untuk analisis selanjutnya)

6. Menentukan Objek Data

Instruksi Kerja:

- 1. Tentukan kriteria dan teknik pemilihan data
- 2. Tentukan attributes (columns) dan records (row) data

- 1. Apakah perlu seleksi kolom tertentu? (pakai semua atau beberapa kolom tertentu?) untuk mencapai tujuan
- 2. Kolom mana saja yang mau diolah (misal kolom 1-5), record nya (misal ambil tahun 2001 ke atas)

7. Membuat Business Intelligence

Instruksi Kerja:

- 1. Membuat dashboard ataupun analisis Business Intelligence
- 2. Buat analisis disertai visualisasi hubungan keterkaitan antara sale amount, type property, tahun terdaftar (list year) dengan lokasi (kota)
- 3. Lakukan analisis trend penjualan rumah tahun 2001 s.d. 2020 berdasarkan harga dan type rumah di 5 kota utama
- 4. Lakukan prediksi (forecast) penjualan properti (sale amount) untuk tahun 2021 di 5 kota utama

Kegiatan:

- 1. Membuat rancangan visualisasi
- 2. Visualisasi dan analisis: sale amount, type property, tahun terdaftar (list year) dengan lokasi (kota)
- 3. Analisis time series (bikin grafik, analisis trend, forecast 5 kota utama)
- 4. Time series forecast untuk 2021 di 5 kota utama (pakai model ARIMA univariat)

Catatan:

5 kota utama dipilih berdasarkan kelengkapan datanya (dari tahun 2001-2020) dan data record penjualan terbanyak.

Kota terpilih: Bridgeport, Danbury, Norwalk, Hartford, dan West Hartford

8. Menyusun Laporan Hasil Analisis

Instruksi Kerja:

1. Susun dokumentasi laporan diterapkan sesuai dengan panduan organisasi.

Kegiatan:

1. Buat laporan untuk organisasi kita (penyiapan data, pengolahan, dll.)

THANKS