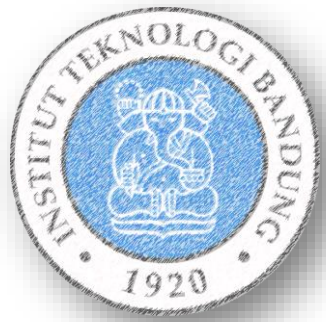


# Tugas Besar 2: Analisis Data

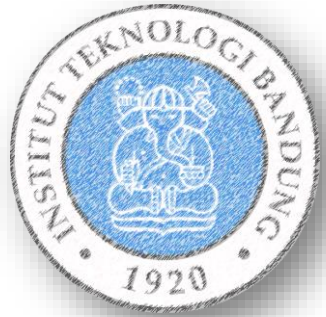
Tim Penyusun Materi Pengenalan Komputasi  
Institut Teknologi Bandung © 2021





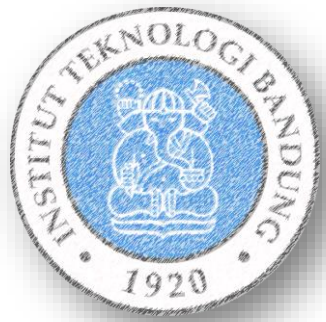
# Tujuan

- Mahasiswa mampu mengenali persoalan komputasi dalam persoalan yang diberikan, menggunakan dekomposisi dan abstraksi dalam pemecahan persoalan, serta menghasilkan artifak komputasi yang relevan
- Mahasiswa mampu berkolaborasi dalam kelompok pemecahan persoalan komputasi.
- Mahasiswa mampu berkomunikasi dengan berbagai pihak dalam rangka mengekspresikan dan bertukar ide mengenai pemecahan persoalan komputasi.



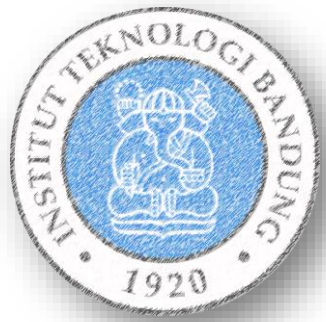
# Deskripsi Tugas (1)

- Carilah **1 dataset** dalam bentuk **tabel** di **internet** atau dari **sumber-sumber lain**.
  - File data yang dianjurkan dalam format XLS atau XLSX atau CSV atau TXT. Tidak disarankan dalam format PDF.
- Jenis topik dataset bebas. **Setiap kelompok harus menggunakan dataset yang berbeda → update siapa cepat dia dapat melalui WAG**
- Setiap tabel data harus memenuhi syarat berikut:
  - Minimum terdiri atas 5 atribut dan minimum terdiri atas 60 baris
  - Harus mengandung atribut kategorikal dan atribut kuantitatif (numerik)
- Setidaknya salah satu tabel harus memenuhi syarat berikut:
  - Salah satu atribut harus merupakan data waktu (*time-series*)



# Deskripsi Tugas (2)

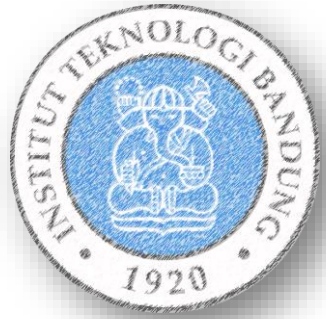
- Contoh-contoh lokasi untuk mendapatkan dataset di internet:
  - UCI Machine Learning Repositories: <https://archive.ics.uci.edu/ml/datasets.html>
  - Data Market: <https://datamarket.com/data/list/?q=provider%3Atsdl>
  - BPS: <https://www.bps.go.id/>
  - Satu Data Indonesia: <https://data.go.id/>
  - Kaggle: [Find Open Datasets and Machine Learning Projects | Kaggle](#)
- Anda dapat mencari dari sumber-sumber lain di internet, atau sumber-sumber di luar internet.
  - Jangan lupa mencantumkan referensi sumber data yang Anda gunakan pada laporan tugas.
- Anda diminta untuk melakukan beberapa analisis deskriptif (*descriptive analysis*) dan membuat laporan terkait dataset tersebut. Berikut hal-hal yang harus dijelaskan dalam laporan.



# Deskripsi Tugas (3)

## Deskripsi Data dan File

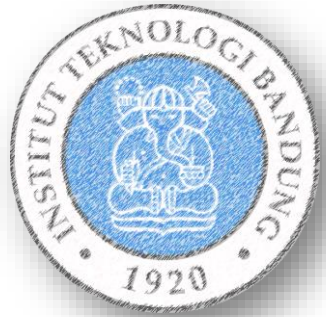
- Jelaskan **deskripsi** data tersebut: data tersebut mengenai apa, informasi/pengetahuan apa yang ingin diketahui terkait data tersebut.
- Jelaskan pula **format** data tersebut (csv atau xlsx atau txt atau format data lain), data tersebut **diambil dari mana** dan bagaimana **dimensinya** (ukuran: berapa banyak kolom dan baris), serta **ukuran file** data.
  - Tuliskan dengan menggunakan bahasa pemrograman/*spreadsheet tool* yang Anda gunakan bagaimana melakukan *loading* data dan bagaimana mengetahui berbagai informasi yang terkait dengan data.



# Deskripsi Tugas (4)

## Karakteristik Data

- Jelaskan **atribut/kolom** apa saja yang ada di dalamnya, apa **makna** tiap atribut mengenai apa, dan jelaskan jenisnya (kategorikal atau kuantitatif, jika kategorikal, apakah nominal, *binary*, dst.).
- Jelaskan juga **karakteristik** data untuk tiap atribut, misalnya terdiri atas nilai apa saja (pada atribut kategorikal), atau berapa *range* nilai datanya (pada atribut kuantitatif), berapa persen data yang kosong, dll.
  - Untuk setiap karakteristik data yang dibuat, tuliskan bahasa pemrograman/*spreadsheet tool* untuk mendapatkannya, misalnya dengan melakukan sorting, mencari nilai minimum/maksimum, dll.

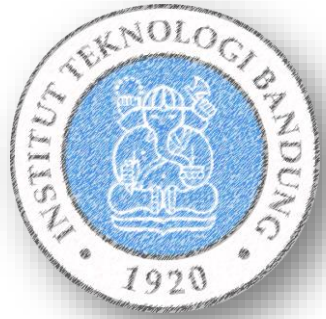


# Deskripsi Tugas (5)

## Statistik

- Tunjukkan **sampel data** dan berikan penjelasan pada sampel data tersebut. Sampel data dapat berupa: beberapa data pada baris pertama, beberapa data yang terbesar atau terkecil (lakukan sort terlebih dahulu), sampel data pada setiap kolom, dll.
- Buatlah **statistik** pada untuk setiap atribut data, minimum terdiri atas:
  - Rata-rata dan standar deviasi
  - *Percentile* (10%, 25%, 50%, 75%, 90%)
  - Ekstremum (nilai maksimum dan minimum)
  - Distribusi frekuensi nilai pada data (jika mungkin)Untuk setiap statistik yang dibuat, tuliskan dengan bahasa pemrograman/spreadsheet tool bagaimana cara mendapatkannya.
- Berikan penjelasan **informasi/pengetahuan** apa yang bisa didapatkan dari statistik tersebut.



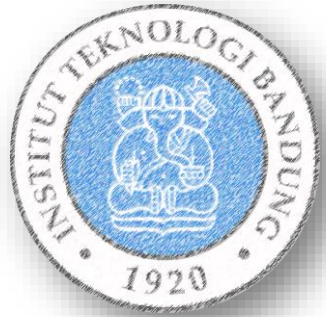


# Deskripsi Tugas (6)

## Visualisasi

- Buatlah **visualisasi** dalam bentuk grafik/*chart* masing-masing minimum 2 buah, untuk setiap kategori berikut:
  - Perbandingan kategori
  - Penampilan perubahan terhadap waktu
  - Penampilan hierarki dan hubungan keseluruhan-bagian
  - *Plotting relationships*
- Untuk tiap visualisasi yang dibuat:
  - Lengkapi visualisasi dengan informasi penting, yaitu: label sumbu x, label sumbu y, legenda, dan judul visualisasi. **Bonus:** Jelaskan penggunaan warna yang dipilih.
  - Jelaskan ***insight*** apa yang bisa Anda dapatkan dari visualisasi yang Anda tampilkan.Untuk setiap visualisasi yang dibuat, tuliskan dengan bahasa pemrograman/*spreadsheet tool* bagaimana cara mendapatkannya
- Bentuk-bentuk grafik/*chart* yang digunakan dapat menggunakan yang dicontohkan di kelas, dapat pula Anda perluas dengan menggunakan grafik lain.

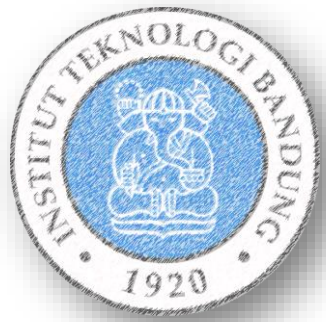




# Deskripsi Tugas (7)

## Korelasi

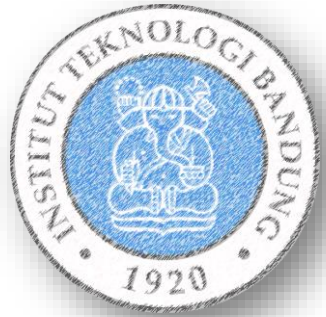
- Tunjukkan dan jelaskan **korelasi** antar semua atribut kuantitatif yang ada dalam tabel.
- Jelaskan dalam bentuk **angka** dan perjelas dengan menggunakan **visualisasi** dalam bentuk grafik yang terkait.
- Tuliskan dengan bahasa pemrograman/*spreadsheet tool* bagaimana cara mendapatkan angka korelasi dan membangkitkan visualisasi dalam bentuk grafik.



# Deskripsi Tugas (8)

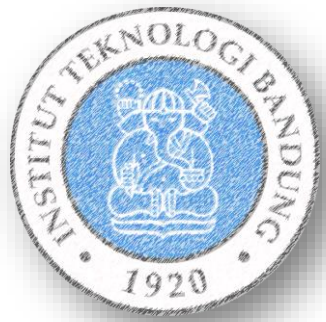
## Data cleansing - Bonus

- **Bonus:** Jika data Anda dalam keadaan “kotor”, yaitu mengandung terlalu banyak data kosong atau data yang salah (misalnya: data umur seharusnya  $> 0$ , tetapi pada data terdapat angka negatif), buatlah pembahasan khusus tentang:
  - Deskripsi tentang tingkat kekotoran data, misalnya: pada atribut yang mana, berapa persen data yang kotor, dll.
  - Bagaimana Anda mengatasinya:
    - Jika dibiarkan apa adanya, jelaskan apa alasannya.
    - Jika “dibersihkan”, jelaskan apa alasannya dan bagaimana melakukannya: apakah diubah nilainya, atau dihapus baris yang mengandung data kotor, dll.
- Tuliskan dengan bahasa pemrograman/spreadsheet tool bagaimana untuk mengecek “kekotoran data” dan bagaimana melakukan pembersihan data (jika dilakukan).



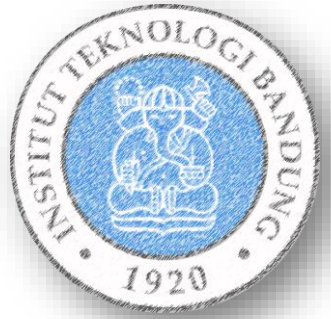
# Petunjuk Pengerjaan

- Deliverable: **laporan** dan **video presentasi**
- Tugas dikerjakan berkelompok dengan @ kelompok 4 s.d. 5 orang, sama dengan kelompok pada tugas sebelumnya.
- Jadwal Kegiatan:
  - Minggu 13: publikasi tugas
  - Masa pengerjaan: minggu 13 s.d. minggu 14 akhir (1,5 pekan)
  - Pengumpulan laporan dan video presentasi: Minggu ke-14 akhir pada tgl **28 November 2021 pukul 23.59**



# Format Laporan

- Laporan berisi hal-hal yang dijelaskan dalam deskripsi tugas.
- Format laporan:
  - Ditulis dalam format kertas A4
  - Font formal (Times New Roman, Calibri, atau Arial) dengan ukuran min. 10 dan maksimum 10 dan spasi teks = 1. Setiap halaman laporan harus memiliki minimum nomor halaman (boleh menambahkan informasi lain pada header dan footer dokumen).
  - Berikan sampul laporan yang minimum mengandung informasi berikut: kode dan nama mata kuliah, daftar NIM dan nama anggota kelompok, judul tugas, bulan dan tahun pemberian tugas.
  - Tuliskan pula bagaimana pembagian kerja dalam kelompok.
- Laporan dikumpulkan dalam bentuk softcopy
- **Link pengumpulan laporan softcopy: [edunex.itb.ac.id](http://edunex.itb.ac.id)**



# Presentasi

- Setiap kelompok diwajibkan membuat presentasi dalam bentuk video berdurasi 10-15 menit (setiap anggota kelompok ikut presentasi).
  - Video diunggah di situs penyedia layanan video atau dalam bentuk file di cloud. Link video dituliskan dalam laporan yang dikirimkan ke [edunex.itb.ac.id](http://edunex.itb.ac.id)