

A. Penjelasan Dataset

Dataset pada kasus Pemilihan Presiden Amerika Serikat 2016 memiliki beberapa file, yaitu `primary_results.csv`, `county_facts.csv`, `county_facts_dictionary.csv`. Pada `primary_results.csv`, terdapat beberapa atribut yaitu:

- `state`: negara bagian pemilihan
- `state_abbreviation`: singkatan negara bagian
- `county`: daerah pemilihan
- `fips`: kode FIPS daerah
- `party`: partai (demokrat atau republik)
- `candidate`: nama kandidat presiden
- `votes`: jumlah suara
- `fraction_votes`: persentase jumlah suara

Kemudian file `county_facts.csv` berisi data demografik tiap daerah yang bersumber dari sensus. Sedangkan `county_facts.csv` berisi penjelasan kolom-kolom pada `county_facts.csv`.

B. Analisis

Untuk menganalisis Pemilihan Presiden Amerika Serikat 2016, saya menggunakan bahasa R dan IDE RStudio. Saya juga menggunakan asumsi bahwa setiap lulusan SMA akan melanjutkan ke bangku kuliah. Untuk library yang digunakan, saya hanya menggunakan `readr` untuk mengimport file csv. Berikut analisis untuk membuktikan dan memecahkan persoalan yang diberikan:

- Menggabungkan data pada file `primary_results.csv` dan `county_facts.csv` berdasarkan FIPS-nya. Data yang digabungkan tersebut disimpan dalam "data".
- Mengambil data dari "data" yang kandidatnya adalah Donald Trump dan menyimpannya pada `datasubset`.
- Membuat dataframe yang berisi rata-rata tiap demografik dan menyimpannya pada `rataan`.

Untuk kasus pertama, yaitu membuktikan bahwa mayoritas orang Amerika Serikat yang mendukung Donald Trump untuk menjadi calon presiden dari partai Republik berasal dari kalangan masyarakat yang tidak mengenyam bangku kuliah, saya mencari persentase masyarakat yang tidak mengenyam bangku kuliah dan mendukung Donald Trump terhadap total semua dukungan untuk Trump. Didapatkan bahwa persentasinya adalah 0.9999855. Sehingga dapat disimpulkan bahwa memang benar mayoritas orang Amerika Serikat yang mendukung Donald Trump untuk menjadi calon presiden dari partai Republik berasal dari kalangan masyarakat yang tidak mengenyam bangku kuliah.

Untuk kasus kedua, yaitu mendapatkan satu faktor dominan lain yang memengaruhi kemenangan Trump di pemilihan kandidat calon presiden, saya melihat dari rata-rata pada `rataan` tadi. Setelah diurutkan berdasarkan rataannya, dapat dilihat bahwa rata-rata paling besar adalah pada `MAN450207`, yaitu `Manufacturers shipments, 2007 ($1,000)`. Sehingga dapat disimpulkan bahwa mayoritas pendukung Trump adalah masyarakat dari kalangan produsen pengiriman.