

Dokumentasi Praktek ETL menggunakan KNIME

- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
 - Proses membaca data dari dua sumber yang berbeda
 - Proses Modeling
- Evaluation
- Deployment

Business Understanding

Data Understanding

- Super bowl adalah acara tahunan American Football yang menentukan juara dari National Football League (NFL). Pertandingan puncak pada satu musim NFL dan menghasilkan kesimpulan dari NFL playoffs. Pertandingan ini dihelat di salah satu kota amerika, ditentukan empat tahun sebelum penyelenggaraan. mulai dari januari 1971, pemenang dari AFC akan menghadapi pemenang dari NFC dan kedua tim akan diadu pada laga puncak NFL Playoffs.
- dataset ini berisi para finalis dari superbowl, mulai dari tahun 1967 sampai 2020.
- dataset ini 54 row dan mempunyai 10 coloumns
 - Date : Tanggal dilaksanakan superbowl final
 - SB : Superbowl Tittle, superbowl final pertama kali dilaksanakan tahun 1967, setiap tahunnya memasang tittle dengan menggunakan huruf romawi , superbowl terakhir adalah LIV(54).
 - Winner : nama tim yang keluar sebagai pemenang
 - Winner pts : points yang diraih pada tim pemenang pada laga final
 - Loser : nama tim yang mengalami kekalahan
 - Loser pts : points yang diraih oleh tim yang mengalami kekalahan
 - MVP : nama most valuable player pada superbowl
 - Stadium : Lokasi stadium dilaksanakannya partai puncak
 - City : Lokasi kota dilaksanakan partai puncak
 - State : negara bagian amerika tempat dihelatnya pertandingan puncak
- Source dataset : <https://www.kaggle.com/timoboz/superbowl-history-1967-2020>

Data Preparation

- memisahkan dengan cara membuat dataframe pada data asli, dan memilih kolom yang akan dimasukkan ke file 1 dan 2



```
import pandas as pd
data = pd.read_csv("superbowl-asli.csv")
data.dropna(inplace = True)
data_sb_asli = pd.DataFrame(data)
def select_columns(data_sb_asli, column_names):
    new_frame = data_sb_asli.loc[:, column_names]
    return new_frame
selected_columns = ['Date', 'SB', 'Winner']
new = select_columns(data_sb_asli, selected_columns)
new.to_csv('superbowl-1.csv', index=False)
selected_columns = ['Winner Pts', 'Loser', 'Loser Pts', 'MVP', 'Stadium', 'City', 'State']
new_2 = select_columns(data_sb_asli, selected_columns)
new_2.to_csv('superbowl-2.csv', index=False)
```



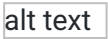


 code: [split](#)

Modeling

Proses membaca data dari dua sumber yang berbeda

Proses membaca dari MYSQL

- yang pertama membaca data dari mysql, dengan menggunakan mysql connector nodes dari knime

- data di mysql seperti dibawah

- Merapikan row yang ada di mysql dengan menggunakan syntax ini

```
DELETE FROM `table 1` WHERE `COL 1`='Date'
```
- hasil menjalankan syntax di atas

- melakukan konfigurasi disesuaikan dengan mysql yang ada di phpmyadmin, mulai dari database, port dari localhost dan username.

- db table selector untuk mengambil Koneksi DB sebagai input dan memungkinkan untuk memilih tabel atau tampilan dari dalam database yang terhubung.

- db reader untuk Mengeksekusi kueri input dalam database dan mengambil hasilnya ke dalam tabel data KNIME.

- hasil akhir dari pembacaan database yang terhubung dari mysql


Proses membaca dari csv

- memasang csv reader untuk membaca file csv
- melakukan konfigurasi , menentukan path file dimana csv disimpan
- hasil dari csv reader

Proses Modeling

- menggunakan joiner node, dengan mensambungkan node dari database yang mengolah mysql dan csv
- melakukan configure dengan memilih kolom yang urutan nya sama , bisa dibilang foreign key agar datanya berhasil di append
- dan hasilnya seperti ini
- nama coloumn di sesuaikan agar mudah dimengerti
- menggunakan coloumn filter untuk menentukan coloumn mana yang mau ditampilkan
- hasil akhir

Evaluation

- dari kedua data yang ada di mysql dan csv telah berhasil di join
- hasil join
- data asli
- dari kedua data di atas, berhasil karena dari hasil join sama dengan data asli ketika sebelum dipisah

Deployment

simpan ke csv

- data yang pertama akan disimpan ke csv dengan menggunakan csv writer
- memilih penempatan dan konfigurasi lain nya

- data berhasil tersimpan

simpan ke db

- data akan disimpan ke database menggunakan db writer
- memilih konfigurasi lain nya, seperti nama dan db yang dituju
- berhasil tersimpan