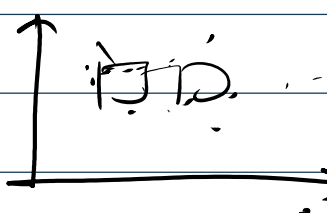


Data Preprocessing for Machine Learning ✓

Topics :

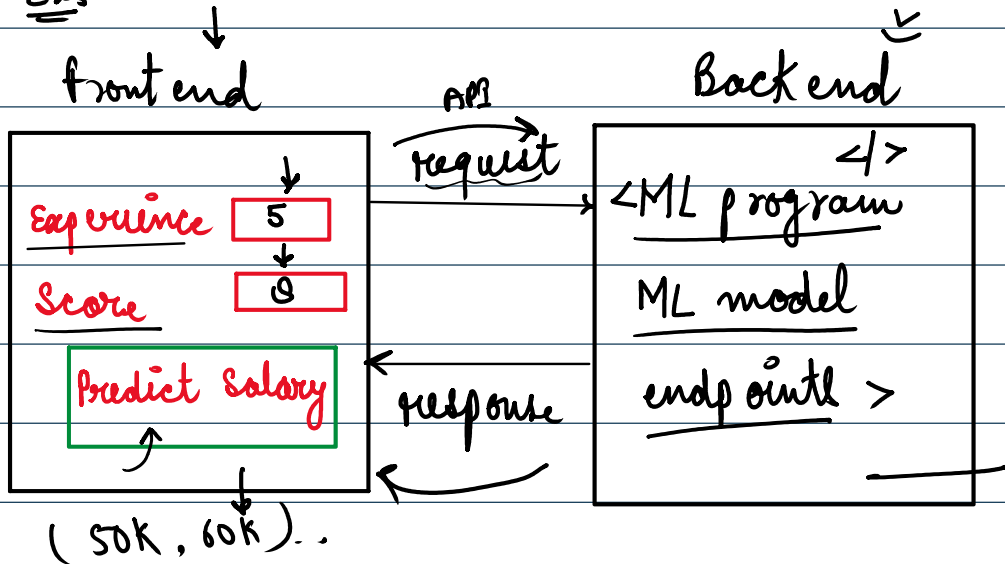
- ✓ 1. Collecting the Dataset ✓ ⇒ Kaggle
 - ✓ 2. Understanding the Data Structure and Features
 - ✓ 3. Cleaning the Data
a. Handling missing values ✓
b. Removing duplicates ✓
 4. Exploratory Data Analysis (EDA) ⇒ Explore Data
a. Visualizing distributions ⇒
b. Correlation analysis ⇒
But 
- clean data
↑
ML model

→ Bwt ⇒ weight of child at birth time ⇒ Target (variable)

- **Case** : Id number
- **Bwt** : Birthweight, in ounces (a unit of weight of one sixteenth of a pound (approximately 28 grams))
⇒ 1 ounce ⇒ ≈ 28 grams
- **Gestation** : Length of gestation, in days (the process or period of developing inside the womb between conception and birth) 9 month ↓ ≈ 280 days
- **Parity** : Binary indicator for a first pregnancy (0 = first pregnancy)
- **Age** : Mother's age in years
- **Height** : Mother's height in inches 50, 60, 40

- **Age** : Mother's age in years
- **Height** : Mother's height in inches $\Rightarrow 50, 60, 40$
- **Weight** : Mother's weight in pounds
- **Smoke** : Binary indicator for whether the mother smokes

Ex:-



Input = { Exp : 0 , Score : 0 } \Rightarrow request \Rightarrow o/p = { Prediction : 50000 }

JSON

Frontend

Backend

Module:-

ML APIs

sect ① { Dataset
Data Preprocessing }

② { Building ML model , saving in .pkl format }

③ { ML Apps, loading model, make predictions }

Input \Rightarrow Postman

O/p \Rightarrow { prediction: ... }

missing values

↓

x	y	...
5	10	...
10	15	...
—	—	...
—	7	...
0	—	...

not available
is yr
missing
value

NaN

Duplicate values

↓

x	y	...
5	10	...
10	15	...
5	10	...
5	10	...
...

→ (original value)

Duplicate
(values)