

## 1) Problem statement

Predict a customer's medical insurance **charges** (continuous target) from available demographic/health attributes.

---

## 2) Dataset basic info

- File used: /mnt/data/insurance\_pre.csv
  - Total rows: **1338**
  - Total columns: **6**
  - Columns found: age, sex, bmi, children, smoker, charges
  - Missing values: **none**
- 

## 3) Pre-processing

- Detected categorical columns: sex, smoker (and any other object cols).
  - Converted categorical variables to dummy/one-hot variables via `pd.get_dummies(..., drop_first=True)`.
  - Standardized numeric features (StandardScaler) prior to model training.
  - Train/test split: 70% train / 30% test (`random_state=0`).
- 

## 4) Machine Learning Algorithms:

### (i) Multiple Linear Regression

**Best  $R^2$  Score = 0.7894790349867009**

### (ii) Support Vector Machine

S. No	C	Linear R <sup>2</sup>	Rbf R <sup>2</sup>	Poly R <sup>2</sup>	Sigmoid R <sup>2</sup>
1	10	- 0.00161763 2488647213 8	- 0.08196910396 420853	- 0.093116155328 48516	- 0.09078319814 614
2	100	0.54328181 96692804	- 0.12480367775 039669	- 0.099761723336 66167	- 0.11814554828 411405
3	1000	0.63403693 1263208	- 0.11749092439 183229	- 0.055505937517 909665	- 1.66590813155 33064
4	3000	0.75908903 67741108	- 0.09621285083 097098	0.048928963896 86563	- 12.0190481058 5073
5	5000	0.76489381 50145156	- 0.07310107353 590722	0.146223786594 05622	- 31.5682820719 9063

#### Best R<sup>2</sup> Score:

Linear kernel at C = 5000: **R<sup>2</sup> = 0.7649**

#### (iii) Decision Tree

Sl.No	Criterion	Splitter	R <sup>2</sup> Value
1	squared_error	best	0.682050774394739
2	squared_error	random	0.7194937388580465
3	friedman_mse	best	0.6939741374705168
4	friedman_mse	random	0.7890426027556149
5	absolute_error	best	0.6686785651066409
6	absolute_error	random	0.7092140339803142
7	poisson	best	0.7167036501017701
8	poisson	random	0.718530210384976

**Best R<sup>2</sup> Score:** friedman\_mse with random splitter: **R<sup>2</sup> = 0.7890**

#### (iv)Random Forest

Sl.No	n_estimators	criterion	random_state	R <sup>2</sup> Value
1	100	squared_error	0	0.8538307913484513
2	50	squared_error	0	0.8498329315421834
3	10	squared_error	0	0.83303041340085
4	100	absolute_error	0	0.8520093621081837
5	50	absolute_error	0	0.8526655993519747
6	10	absolute_error	0	0.835063555313752
7	100	friedman_mse	0	0.8540518935149612
8	50	friedman_mse	0	0.8500716139332296
9	10	friedman_mse	0	0.8331662678473348
10	100	poisson	0	0.8526334258892607
11	50	poisson	0	0.8491075958392151
12	10	poisson	0	0.8313991040134341

**Best R<sup>2</sup> Score:** friedman\_mse with 100 estimators: **R<sup>2</sup> = 0.8541**

---

#### 5)Best Mode:

**Final Model:** Random Forest Regressor

**Justification:** It achieved the highest R<sup>2</sup> score (0.8541) using friedman\_mse with 100 estimators, showing superior accuracy, stability, and generalization compared to other models.