Insurance Charges Prediction Project Report

1. Problem Statement

The client wants a system to predict medical insurance charges based on age, sex, BMI, children and smoker status.

Goal: Build a machine-learning model to accurately estimate insurance costs.

---

2. Dataset Information

Detail	Value
Total Rows	1338
Total Columns	6
Column Names	age, sex, bmi, children, smoker, charges

---

3. Data Preprocessing

- **Convert text to numbers:**
    - sex → male = 1, female = 0
    - smoker → yes = 1, no = 0
- **Check for missing data:** No missing values.
- **Feature scaling (optional):** Standard Scaler applied to the models

---

4. Models Used

The following models were tested:

- Multiple Linear Regression
- Support Vector Machine (SVM)
- Decision Tree
- Random Forest

---

5. Model Performance (R^2 Score)

Model	**R^2 Score**
Decision Tree	0.6927
Multiple Linear Regression	0.7895
Random Forest	0.8566
Support Vector Regression (SVR)	-0.0834

---

6. Final Model and Reason

Final Model: Random Forest Regressor

Reason:

- Has the **highest R^2 score (0.8566)**
- Handles **different types of features** (numbers and categories)
- Reduces overfitting by using **many trees instead of one**
- Gives **stable and accurate predictions** for insurance charges