**Project Title: company sales prediction by using Random forest**

**Project Duration:** 1 Days  
**Date Completed:** April 21, 2025  
**Author:** *sagar khese*

**1. project overview :**

To predict company sales performance based on various feature like income ,advertising , location , education and demographic .this classification help the business in targeted marketing and strategy planning

**2. Dataset Overview**

* **Source file:** company data.csv
* **Shape:** 400 rows × 11 columns
* **Features:** location, Age, urban ,population, education
* **Target Variable:** sales\_cat(high = 1, average =0 , low = 2)

**3. Solution Architecture**

* **Raw data -- > data processing -- > Feature encoding (**location, Age, urban ,population, education) **-- > Model Training (RM) -- > Model Evaluation (accuracy score , ,correlation map) -- > Model saving**

**4. Methodology**

| **Step** | **Reason** |
| --- | --- |
| **Data Collection** | Loading csv file from local library |
| **Data processing** | Convert sales numerical columns into categorical cols and Encoded categorical variables using label encoder. |
| **Exploratory Data Analysis (EDA)** | Visualized categorical feature distributions and correlation heatmaps , bar graph and pariplot |
| **Feature Engineering** | Separated features x and target y |
| **Modeling** | Random forest classifier with entropy criterion and training model |
| **Evaluation** | Accuracy score for training data |
| **Deployment Prep** | Saved the model using Joblib for future use |

**5. Time Taken**

| **Task** | **Time Spent** |
| --- | --- |
| Data Cleaning & EDA | 1 hour |
| Model Building | 1.30 minutes |
| Visualization | 30 minutes |
| Documentation | 1 hour |
| **Total Time** | **4 hours** |

**6. Challenges Faced**

* Encoding multi-categorical data while preserving interpretability.
* Risk of overfitting due to evaluation on training data.
* Understanding which features most affect sales category.

**7. Complexity**

* **Complexity:** medium
* **Involves multiple preprocessing steps**
* Needs care in evaluation to avoid misleading accuracy
* **Skills Required:** Python, Pandas, Seaborn, Scikit-learn, Data Visualization