**Project Summary: Startup Profit Prediction**

**🔍 About the Data**

The dataset contains financial information from 50 startups, including:

* R&D Spend
* Administration Cost
* Marketing Spend
* Profit (Target variable)

The goal is to build a machine learning model that can predict a company's profit based on the other three inputs.

**🧪 Model Training**

To solve this regression problem, we trained two machine learning models:

* **Random Forest Regressor** – a bagging-based ensemble model.
* **Gradient Boosting Regressor** – a boosting-based ensemble model.

Before training, we performed the following preprocessing steps:

* Handled any missing values.
* Split the data into training and testing sets (80/20 split).
* Applied standard scaling to normalize the input features using StandardScaler.

**📈 Model Evaluation**

We evaluated both models using standard regression metrics:

* **R² Score** (explains how well the model fits the data)
* **Mean Absolute Error (MAE)**
* **Mean Squared Error (MSE)**

| **Metric** | **Random Forest** | **Gradient Boosting** |
| --- | --- | --- |
| **R² Score** | 0.910 | 0.890 |
| **MAE** | 6,437.50 | 8,664.90 |
| **MSE** | 72,625,008.62 | 88,674,767.81 |

**✅ Best Model Selection**

After comparing both models:

The **Random Forest Regressor** performed better on all evaluation metrics:

* Higher R² score → Better overall fit.
* Lower MAE and MSE → Smaller prediction errors.

Therefore, Random Forest was selected as the final model for deployment.

**💡 Why Random Forest?**

* It generalizes well, especially on small datasets like this one.
* Less sensitive to outliers and overfitting.
* Easier to interpret and debug than Gradient Boosting in small-scale projects.