Machine Learning Model Evaluation Report

1. Modeling Problem

We explored two types of machine learning problems:

Regression Problems

- **Model 1**: Predicting Canada's *Per Capita Income* over the years.
- **Model 2**: Predicting *Maximum Temperature* in Lahore using historical weather data.

Classification Problems

- **Model 3**: Classifying iris flowers using **KMeans (unsupervised)**.
- **Model 4**: Predicting iris flower species using **Random Forest Classifier**.
- **Model 5**: Predicting iris flower species using **Support Vector Machine (SVM)**.

2. Models Trained

Regression Models

Model	Dataset	Technique	Preprocessing	Tuning
Linear Regression	Canada Per Capita Income	`LinearRegression()`	None	No
Linear Regression with Pipeline	Lahore Weather Data	`Pipeline + FunctionTransformer`	Date converted to ordinal	No

Classification Models

Model	Dataset	Technique	Preprocessing	Tuning

KMeans Clustering	Iris Dataset	`KMeans(n_clusters=3)`	`MinMaxScaler`	No
Random Forest Classifier	Iris Dataset	`RandomForestClassifier()`	None	No
Support Vector Classifier	Iris Dataset	`SVC(C=10)`	None	GridSearchCV, RandomizedSearchCV

3. Evaluation Metrics

Regression Models Evaluation

Model	R^2 Score	Additional Notes
Canada Income Prediction	High R^2 (visually linear)	Predicted 2025 income successfully
Lahore Temperature Prediction	`model.score()` used	Custom pipeline used with ordinal date

Recommended Improvements:

- Evaluate using **MAE**, **MSE**, and **R^2** explicitly.
- Add cross-validation or time-series split for better evaluation.

Classification Models Evaluation

Model	Accuracy	Precision	Recall	F1-Score
KMeans + Label Mapping	Reasonable	Used mode-based mapping	Good for unsupervised baseline	Decent but not optimal
Random Forest	~90%	High	High	High
SVM	~97%	High	High	High

Hyperparameter Tuning:

- **SVM**: Tuned using `GridSearchCV` and `RandomizedSearchCV` with parameters `C` and `kernel`.

4. Visualizations Included

- Scatter plots for income and temperature trends.
- Cluster visualizations for KMeans using petal features.
- Heatmaps for confusion matrices of classification models.
- Elbow curve for optimal K in KMeans.

5. Key Observations & Improvements

Observations

- Linear models work well on clean, linear trends (e.g., income, temperature).
- Random Forest outperforms SVM slightly on Iris dataset without tuning.
- KMeans is good as an unsupervised learning example but less accurate than supervised methods.

Improvements

- Add error metrics (RMSE, MAE) for regression.
- Add cross-validation scores for classification robustness.
- For temperature prediction, consider seasonality trends or time-series models.
- Incorporate more features (e.g., humidity, wind) for temperature prediction.
- Use classification reports and ROC curves for deeper classification insight.