Project Development Phase
Model Performance Test

Date	10 February 2025
Team ID	LTVIP2025TMID60817
Project Name	Sustainable smart city assistant using IBM granite LLM
Maximum Marks	10 Marks

Model Performance Testing:

🔍 Model Performance Testing Template

S.No.	Parameter	Values (to fill)	Screenshot
1	**Metrics**	**Regression Model**: MAE – ____ , MSE – ____ , RMSE – ____ , R² score – ____ **Classification Model**: Confusion Matrix – ____ , Accuracy Score – ____ , Classification Report – ____	
2	**Tune the Model**	**Hyperparameter Tuning** – (e.g., grid search parameters: learning rate, batch size, #trees) **Validation Method** – (e.g., k-fold CV, hold-out validation with __% test split)	

✅ Adapting for Sustainable Smart City Assistant (IBM Granite LLM)

Since this assistant uses IBM Granite LLM, a foundation model tuned for sustainability and smart urban management tasks, ensure that the **"Values"** section includes:

- **RAG-augmented tasks** (e.g., retrieving real-time energy consumption data, local waste metrics).
- **Text generation metrics** if using Granite for synthesizing plans or reports.

Suggested Additions:

- **Relevance Score** (for RAG outputs) – e.g., precision@k, recall@k.
- **Perplexity or BLEU Score** (if comparing generated responses to ground truth).
- **User Satisfaction** – mean ratings from citizen or policymaker feedback (~1–5).

🛠️ Example of Filled-Out Template

S.No.	Parameter	Values
1	**Metrics**	**Regression (e.g., energy use prediction):** MAE – 3.2 MWh, MSE – 18.4, RMSE – 4.29 MWh, R² – 0.87 **Classification (e.g., service request triage):** Confusion Matrix – [[45, 5],[8, 42]] , Accuracy – 0.885 , Classification Report – (precision/recall/F1 for each class)
2	**Tune the Model**	**Hyperparameter Tuning:** grid search over learning_rate {1e-3, 1e-4}, batch_size {16,32}, #epochs {5,10} **Validation:** 5-fold cross-validation on temporal split; 80/20 train/test hold-out for final eval

---

ℹ️ Why Granite LLM?

IBM Granite (e.g., 3.0–3.2) — especially its Instruct and Guardian variants — offers:

- **Instruction-tuned, enterprise-ready performance**, ideal for generating sustainability plans or handling citizen queries ibm.com+4github.com+4newsroom.ibm.com+4forbes.com+1reddit.com+1eetimes.com+11ibm.com+11forbes.com+11.
- **Robust RAG capabilities**, enabling retrieval of real-time or external data when crafting assistant responses .
- **Reasoning support** via chain-of-thought models (Granite 3.2) to explain decisions or plans transparently ibm.com+2reddit.com+2reddit.com+2.
