# CAPSTONE PROJECT

# MATERNAL HEALTH PROGRESS

## (A GLOBAL PERSPECTIVE ON SDG 3.1 PROGRESS)

**Presented By:**
Koyya Rohith Kiran – JNTUH College of Engineering Hyderabad – Computer Science and Engineering, specialized in Artificial Intelligence and Machine Learning
AICTE Student ID: STU682e06ebcd2fb1747846891
Internship ID: INTERNSHIP_1748937226683eaa0a58abc

edunet foundation

# OUTLINE

- Problem Statement
- Proposed System/Solution
- System Development Approach
- Algorithm & Deployment
- Result (Output Image)
- Conclusion
- Future Scope
- References

# PROBLEM STATEMENT

The Sustainable Development Goal 3.1 aims to reduce the global maternal mortality ratio to less than 70 per 100,000 live births by 2030. Monitoring progress towards this goal requires analyzing country-wise data on maternal mortality and associated health indicators such as antenatal care coverage, births attended by skilled personnel, adolescent birth rates, and healthcare expenditures. Despite global efforts, maternal health outcomes vary drastically between regions and income groups, raising the need for data-driven insights into the factors influencing maternal health

# PROPOSED SOLUTION

- **Data Collection:**
  Collected 2017 maternal health data from AI Kosh. The goal is to predict DataValue (maternal health indicator) to support SDG 3.1 – reducing maternal mortality.
- **Features Used:**
  AreaID, AreaName, TimePeriod, Source, Sector, Subsector, Goal, Target, Indicator, Unit, SubgroupDimension, Subgroup, SubgroupOrder.
  **Target:** DataValue
- **Preprocessing:**
  AutoAI handled missing values, encoded categorical features, and performed feature engineering.
- **Model Building:**
  Used IBM Watsonx AutoAI to train multiple pipelines with automated feature engineering, two rounds of hyperparameter tuning (HPO-1, HPO-2), and batch processing.
- **Best Model:**
  Batched Tree Ensemble Regressor (Snap Random Forest Regressor) with RMSE of 127.111 was selected.
- **Deployment:**
  Model deployed in IBM Watsonx Deployment Space, accessible via Watson Machine Learning API for real-time predictions.

# SYSTEM APPROACH

The system adopts a modular, end-to-end pipeline that begins with **country-wise maternal health data collection from AI Kosh**, followed by **automated preprocessing**—including **data cleaning, encoding, and feature engineering**. IBM **Watsonx AutoAI** is then used to train and evaluate multiple regression models using **cross-validation and RMSE as the performance metric**. Through two rounds of **hyperparameter optimization (HPO-1 and HPO-2)** and **batched ensemble training**, the best-performing model, the **Batched Tree Ensemble Regressor (Snap Random Forest Regressor)**, is selected. The final model is **deployed in IBM Watsonx Deployment Space**, enabling **real-time predictions via REST API**. This robust pipeline supports **data-driven decision-making** aligned with **SDG 3.1 (reducing maternal mortality)**.

# ALGORITHM & DEPLOYMENT

- **Algorithm Selection:**
  The **Batched Tree Ensemble Regressor** (Snap Random Forest Regressor) was selected by IBM AutoAI as the best-performing model for this regression task. It combines the power of ensemble learning with batch processing to improve training efficiency and accuracy.
- **Data Input:**
  The input features used were: AreaID, AreaName, TimePeriod, Source, Sector, Subsector, Goal, Target, Indicator, Unit, SubgroupDimension, Subgroup, and SubgroupOrder. The target variable was **DataValue**.
- **Training Process:**
  AutoAI automated the entire ML workflow, including **feature engineering (FE)**, **first and second rounds of hyperparameter optimization (HPO-1 and HPO-2)**, and **batched model training**. It utilized cross-validation to ensure robust model performance.
- **Prediction Process:**
  The final pipeline (Pipeline 9) achieved a **Root Mean Squared Error (RMSE) of 127.111**, indicating strong predictive performance. This model is capable of making real-time predictions via REST API after deployment.
- **Deployment:**
  The final model was **successfully deployed in IBM Watsonx Deployment Space**, enabling real-time predictions through a **REST API**. It is now ready for integration into applications or dashboards for continuous usage.

# RESULT

The **Batched Tree Ensemble Regressor (Snap Random Forest Regressor)** achieved a **Root Mean Squared Error (RMSE) of 127.111**, indicating strong predictive performance in estimating **DataValue** from the **AI Kosh dataset** using various socio-economic and health-related indicators.

**Github Link:** https://github.com/rohithkiran15/maternal_health_progress

**Deployment Space Public Endpoint:** https://au-syd.ml.cloud.ibm.com/ml/v4/deployments/1b092751-4d35-44ca-b2dd-6b96d965da5a/predictions?version=2021-05-01

# CONCLUSION

- The project successfully leveraged IBM Watsonx AutoAI to automate data preprocessing, model training, and selection, resulting in a high-performing predictive model. It demonstrates the effective use of AI to analyze public health data and support data-driven decision-making aligned with SDG 3.1 goals.

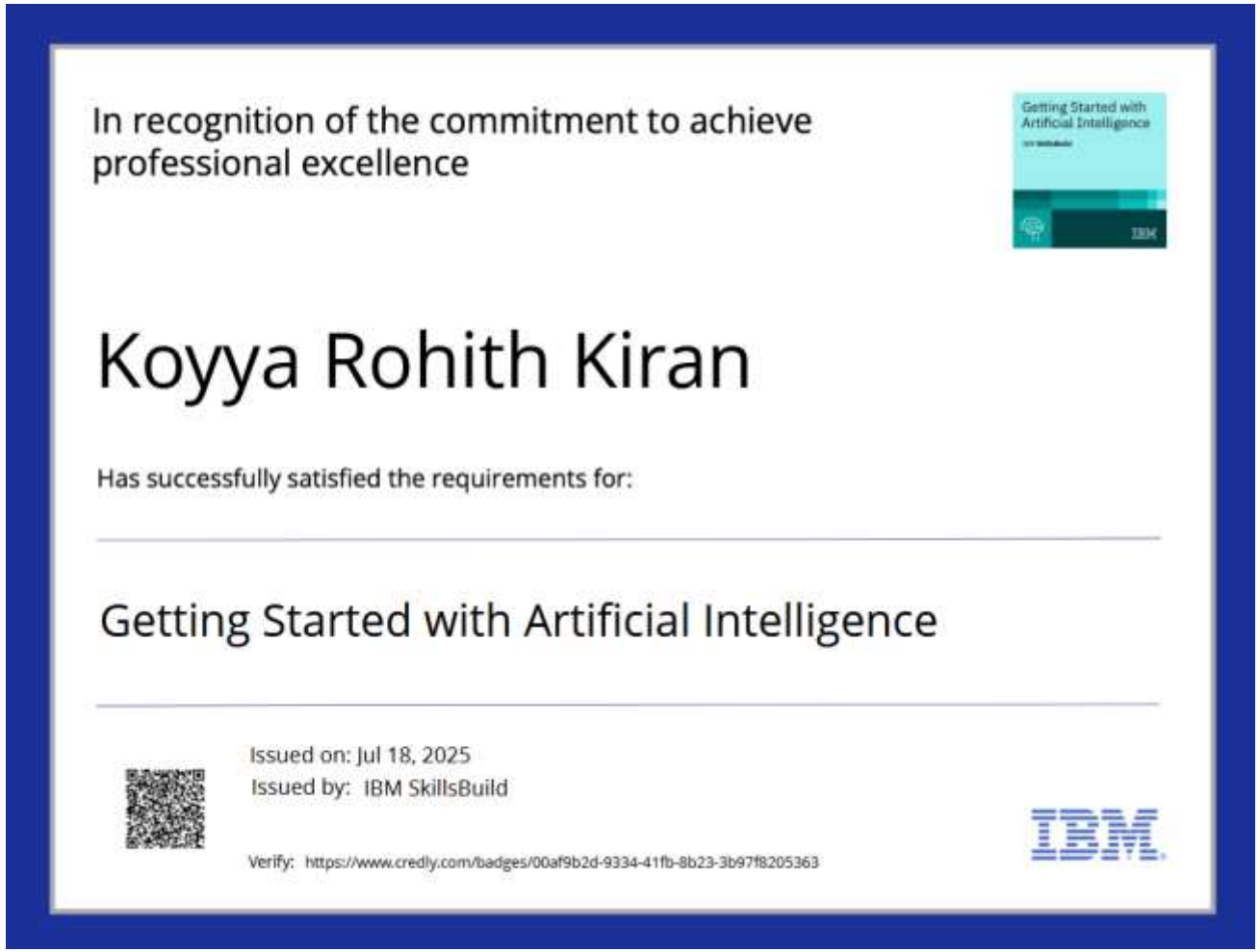
# FUTURE SCOPE

- Integrate more recent or real-time data to improve prediction relevance.
- Explore more advanced models like ensemble methods or neural networks.
- Deploy the system as a full-stack application for end-users like policymakers or researchers.
- Incorporate visual dashboards for interactive data insights.

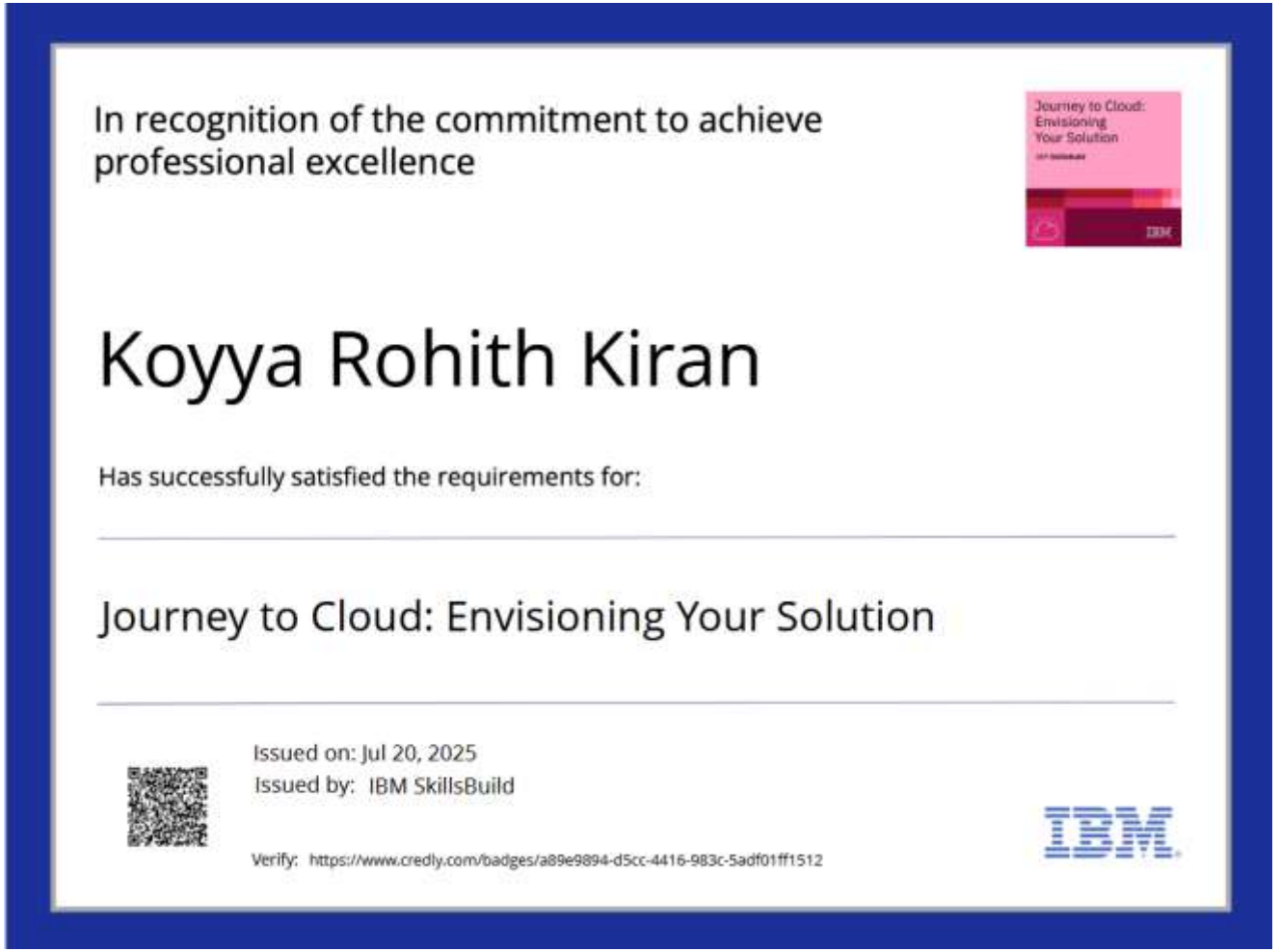
# REFERENCES


- AI Kosh dataset link: https://www.data.gov.in/resource/sustainable-development-goals-national-indicator-framework-version-31-2021
- IBM Watsonx.ai.studio: Documentation for IBM watsonx as a Service — Docs | IBM watsonx
- IBM Skills Build: https://skills.yourlearning.ibm.com/?lang=en&ngo-id=0302&source=login&strategy=google

# IBM CERTIFICATIONS



# IBM CERTIFICATIONS



# IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate



This certificate is presented to

Koyya Rohith Kiran

for the completion of

**Lab: Retrieval Augmented Generation with LangChain**

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

**Completion date:** 20 Jul 2025 (GMT) **Learning hours:** 20 mins

edunet foundation

# THANK YOU

edunet foundation