

---

# CAPSTONE PROJECT

## Tracking Maternal Health Progress Toward SDG 3.1: A Global Data Analysis

**Presented By:**

**Student Name- Mohd Sheeban Khan**

**College Name- Invertis University**

**Department- Computer Applications**

# OUTLINE

- Problem Statement
- Proposed System/Solution
- System Development Approach
- Algorithm & Deployment
- Result
- 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

- This project proposes a machine learning-driven approach to analyze global maternal health data using IBM Cloud Lite services.


Key components include:

- Preprocessing country-wise datasets on maternal health.
- Using supervised learning models to predict health indicators.
- Providing visual insights and performance metrics.
- Hosting model deployment using IBM Watson Studio & Cloud services.
-  GitHub Repository: <https://github.com/Spoofy23/IBM-SkillsBuild-Internship-on-AI-Cloud-Technologies-Project/tree/main>

# System Approach

- **Data Source:** Publicly available WHO & UN maternal health datasets
- **Cleaning:** Handled nulls, outliers, standardized column names
- **Pivoting:** For subgroup-based indicator analysis
- **Modelling:** Linear Regression using Scikit-Learn
- **Tools Used:**
  - IBM Watson Studio
  - IBM Cloud Lite
  - Jupyter Notebook
  - Python (Pandas, Scikit-learn)

# Algorithm & Deployment

- **Algorithm Selection: Linear Regression**
  - In our case, it was used to predict DataValue, such as antenatal care coverage or births attended by skilled personnel.
- **Data Input (Target Variable): DataValue**
  - Represents the numeric value of maternal health indicators such as:
    - Antenatal care coverage
    - Births attended by skilled health personnel
    - Adolescent birth rate
- **Input Features:**
  - TimePeriod (e.g., 2015-16)
  - Indicator (e.g., Antenatal care coverage)
  - Subgroup (e.g., Urban, Rural)
  -  These features were one-hot encoded to convert categorical values into numerical form for machine learning.
- **Training Process:**
  - Used train-test split (80/20) to divide data for training and testing.
  - Achieved a very low Mean Squared Error (MSE), indicating accurate predictions.
- **Prediction Process:**
  - The trained Linear Regression algorithm predicts future maternal health indicator values based on encoded historical data inputs like TimePeriod, Indicator, and Subgroup, without requiring real-time data during prediction using model which was trained and saved within IBM Watson Studio.

# Result

Tracking\_Maternal\_Health\_Progress\_Mark2

✔ Deployed

Online

API reference

Test

Enter input data

Text

JSON

Enter data manually or use a CSV file to populate the spreadsheet. Max file size is 50 MB.

Download CSV template ⬇

Browse local files ↗

Search in space ↗

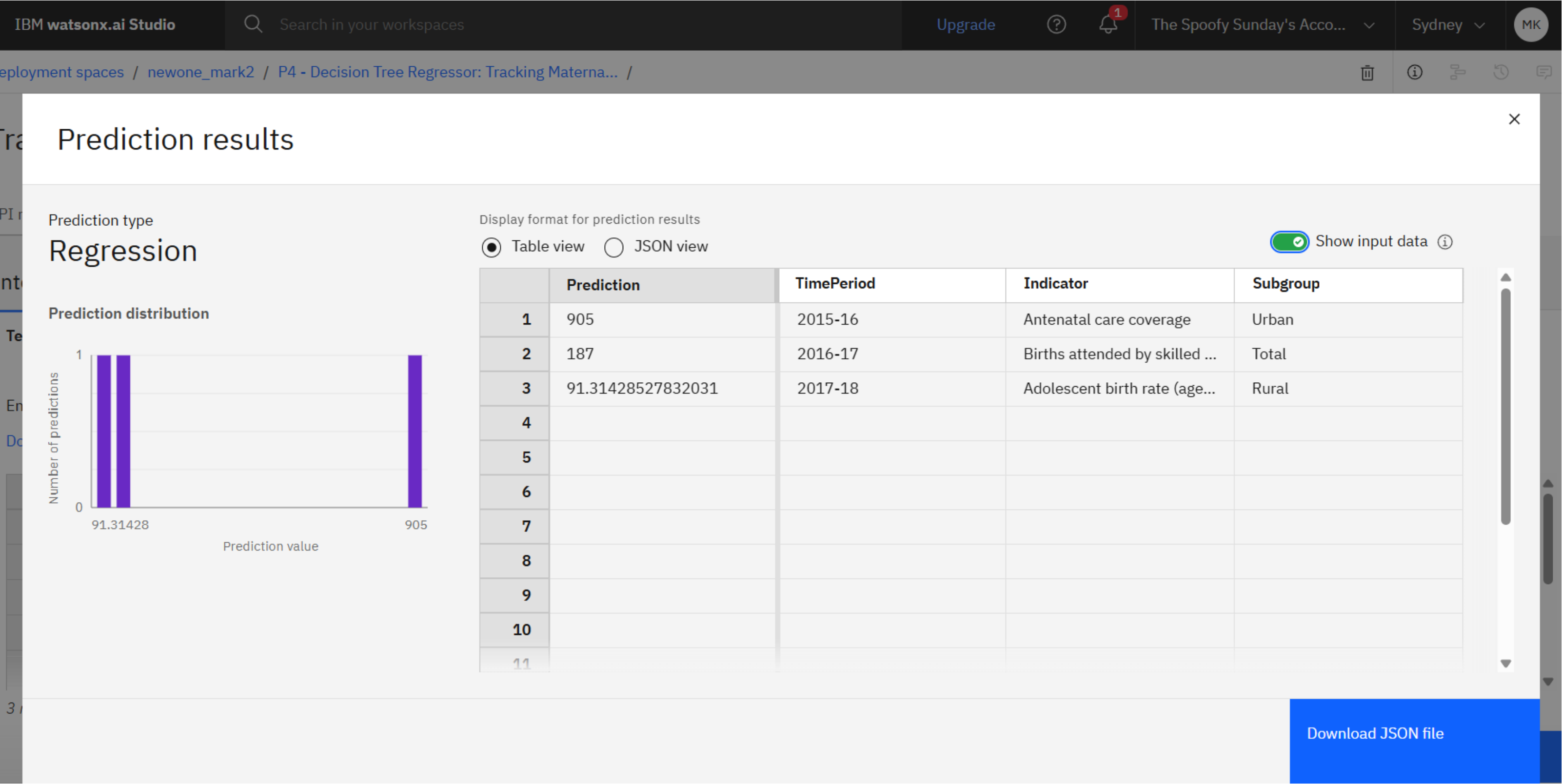
Clear all ×

	TimePeriod (other)	Indicator (other)	Subgroup (other)
1	2015-16	Antenatal care coverage	Urban
2	2016-17	Births attended by skilled personnel	Total
3	2017-18	Adolescent birth rate (aged 15-19)	Rural
4			
5			

3 rows, 3 columns

Predict

# Result





---

# Conclusion

- The project successfully showcases the use of IBM Cloud in addressing a real-world health goal (SDG 3.1).
- It demonstrates how machine learning models can uncover key trends in maternal health.
- While some indicators showed high variance, the project establishes a strong foundation for scalable analysis.

# Future scope

- Integrate real-time WHO API data
- Use more advanced models like XGBoost or LSTM
- Develop an interactive dashboard using IBM Cloud Functions
- Collaborate with health organizations for deeper validation
- Extend model to SDG 3.2 and other healthcare goals

---

# References

- WHO Global Health Observatory
- United Nations SDG 3.1 Tracker
- Scikit-learn Documentation
- IBM Cloud Lite & Watson Studio Docs
- Dataset: Maternal Mortality Indicators (country-wise)

# IBM Certifications

credly certificate( getting started with AI)



# IBM Certifications

credly certificate( Journey to Cloud)



# IBM Certifications

credly certificate( RAG Lab)

7/24/25, 10:07 PM

Completion Certificate | SkillsBuild

IBM SkillsBuild

Completion Certificate



This certificate is presented to

Mohd Sheeban Khan

for the completion of

**Lab: Retrieval Augmented Generation with LangChain**

(ALM-COURSE\_3824998)

According to the Adobe Learning Manager system of record

Completion date: 24 Jul 2025 (GMT)

Learning hours: 20 mins

**THANK YOU**