
CAPSTONE PROJECT

PREDICTIVE MODELING OF ACCESS TO IMPROVED DRINKING WATER IN INDIA

Presented By:
**Mukesh K.A-Karan Arts & Science College-B.sc (Computer
Science)**

OUTLINE

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

PROBLEM STATEMENT

- The Challenge: Access to safe and improved sources of drinking water remains a critical issue in India, especially in rural and underdeveloped regions. Despite ongoing efforts under the Sustainable Development Goals (SDGs), inequalities persist in water accessibility across states and socio-economic groups. This project aims to analyze data from the 78th Round of the Multiple Indicator Survey (MIS) to assess the percentage of the population with access to improved drinking water sources. It will also explore related indicators such as use of clean cooking fuel and migration trends. By identifying patterns and disparities, the study will generate actionable insights to support evidence-based policymaking. The ultimate goal is to help ensure equitable access to clean water and contribute to India's progress on SDG targets.

PROPOSED SOLUTION

Data Collection

- Dataset: NSS 78th Round – Multiple Indicator Survey
- File Used: nss Items data.csv
- Features: State, Sector, Gender, Age Group, Indicator
- Target: Value (percentage of access to water or related indicators)

Data Preprocessing

- Handled missing values and formatted categorical fields
- Applied label encoding and one-hot encoding to non-numeric features
- Standardized the dataset for training

Machine Learning Model

- Used **IBM Watsonx.ai studio** to automate algorithm selection
- AutoAI trained multiple models and selected XGB Regressor as the best performer
- Trained model predicts the Value column based on input features
- Metrics used: MAE, R^2 Score

Deployment

- Deployed the model as an online REST API on IBM Cloud (watsonx.ai Studio)
- Deployment type: wml-hybrid_0.1
- Allows real-time prediction for new inputs (a state/sector/gender combination)

Evaluation

- AutoAI evaluated all model pipelines
- XGB Regressor had the best results with high R^2 and low MAE
- Visual and tabular outputs were used to verify accuracy

SYSTEM APPROACH

This section outlines the strategy and environment used to build and deploy the predictive system for improved drinking water access in India using IBM Cloud.

- **System Requirements**

- Processor: Intel® Core™ i3-1005G1 CPU @ 1.20GHz
- Installed RAM: 16 GB
- Storage: 512 GB NVMe SSD (Toshiba THNSN5512GPUK)
- Graphics: Intel® UHD Graphics (128 MB)
- System Type: 64-bit operating system, x64-based processor
- Operating System: Windows 11
- Input Support: No pen or touch input
- Browser: Google Chrome (used to access IBM Cloud)
- Internet: Stable broadband connection for cloud-based model training and deployment

- **Libraries / Technologies Used**

- (All processing handled automatically by IBM AutoAI on Watsonx.ai Studio)
- IBM Cloud – Watsonx.ai Studio
- AutoAI (IBM's automated ML engine)
- XGBoost Regressor (selected by AutoAI as the best model)
- IBM Watson Machine Learning Runtime (wml-hybrid_0.1)
- No manual coding required – all steps performed using IBM's graphical AutoAI interface

ALGORITHM & DEPLOYMENT

Algorithm Selection:

- The algorithm used was selected automatically by **IBM Watsonx.ai's studio tool**. After comparing multiple models internally, It chose **XGBoost Regressor** as the best-performing model for predicting the percentage of population with access to improved drinking water. The selection was based on metrics like **R² score** and **Mean Absolute Error (MAE)**, using the structure of the provided NSS dataset.

Data Input:

The following features from the dataset were used as input variables:

- State
- Sector (Rural/Urban/All)
- Gender
- Age Group
- Indicator
- The **target variable** was:
- **Value** – representing the percentage of people with access to improved drinking water or other indicators

ALGORITHM & DEPLOYMENT

Training Process:

- Dataset: nss Items data.csv uploaded to IBM Watsonx.ai Studio
- AutoAI automatically performed:
- Data preprocessing
- Encoding of categorical values
- Model selection and comparison
- AutoAI internally used cross-validation to ensure stable results across the dataset
- Best model pipeline (XGBoost Regressor) was finalized and ready for deployment

Prediction Process

- Once trained, the XGB Regressor model was deployed on **IBM Cloud** as a **REST API endpoint**
- Users can send new data points (e.g., state, gender, indicator) to the endpoint for prediction
- The model returns the **predicted percentage (Value)** based on learned patterns from the dataset

RESULT

IBM Cloud Search resources and products... Catalog Manage MUKESH K A's Account

Create About

Type Service
Provider IBM
Last updated 05/06/2025
Category AI / Machine Learning
Compliance HIPAA Enabled IAM-enabled
Location Sydney (au-syd) Frankfurt (eu-de) London (eu-gb) Tokyo (jp-tok) Dallas (us-south) Toronto (ca-tor)

Select a location
London (eu-gb)

Select a pricing plan
Prices shown are for country or location: [United States](#)

Plan	Features and capabilities	Pricing
Lite	1 authorized user 10 capacity unit-hours monthly limit Environment = # of capacity units required per hour • 1 vCPU + 4 GB RAM = 0.5 • 2 vCPU + 8 GB RAM = 1 • 4 vCPU + 16 GB RAM = 2 • Decision Optimization + Watson NLP = Environment + 5 • Synthetic Data Generator, 2 vCPU + 8 GB RAM = 7 (requires watsonx.ai Runtime)	Free

Summary

watsonx.ai Studio Free

Location: London (eu-gb)
Plan: Lite
Service name: watsonx.ai Studio-2a
Resource group: Default

☒ I have read and agree to the following license agreements:
[Terms](#)

Create

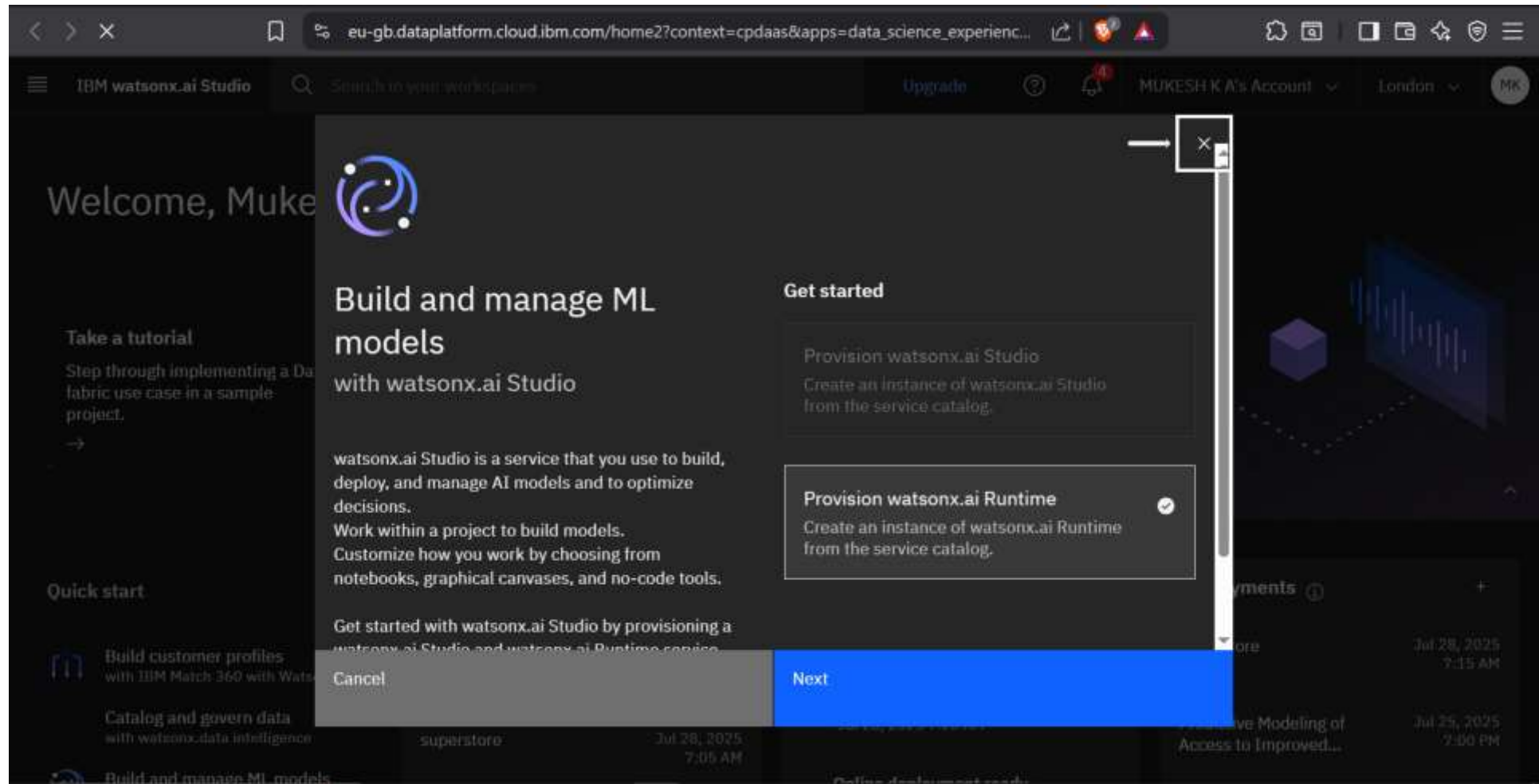
Add to estimate

RESULT

The screenshot displays the IBM Cloud console interface. At the top, the navigation bar includes the IBM Cloud logo, a search bar, and links for Catalog, Manage, and the user's account (MUKESH K A's Account). Below the navigation bar, the breadcrumb "Resource list /" is visible. The main header area shows the resource name "watsonx.ai Studio-2a" with a green status icon, an "Add tags" link, and buttons for "Details" and "Actions".

The left sidebar contains a "Manage" section with a "Plan" option. The main content area features a large card for "Studio in Cloud Pak for Data and watsonx". This card includes a Watsonx icon, a descriptive paragraph about building and deploying machine learning models, and a "Launch in" button with a dropdown arrow. To the right of the text is a 3D architectural diagram illustrating the stack: "IBM Cloud Base cloud infrastructure" at the bottom, "IBM Cloud Pak for Data, watsonx Unifying platforms" in the middle, and "IBM watsonx.ai Studio in Cloud Pak for Data and watsonx" at the top.

RESULT



RESULT

The screenshot shows the IBM Watsonx.ai Runtime creation interface. The main heading is "watsonx.ai Runtime" with sub-links for "Author: IBM", "Date of last update: Jul 23, 2025", "Docs", and "API Docs". There are two tabs: "Create" (active) and "About".

Select a region

Select a region

London

Pricing plan

Displayed prices do not include tax. Monthly prices shown are for country or region: United States

Plan	Features	Pricing
Lite	Service instance Instance includes: <ul style="list-style-type: none">• 20 capacity unit-hours (CUH) per month	Free

Summary

watsonx.ai Runtime

Region: London
Plan: Lite
Service name: watsonx.ai Runtime-xw
Resource group: Default

Create

View terms

Cancel

RESULT

The screenshot displays the IBM watsonx.ai Studio interface. At the top, there's a navigation bar with the IBM logo, a search bar, and user account information (MUKESH K A's Account, London). The main content area is titled 'Cloud Object Storage' and includes a 'Create' button. Below this, a 'Pricing plan' section is visible, showing a table of plans. The table has three columns: Plan, Features, and Pricing. The 'One-Rate' plan is highlighted. The 'Lite(deprecated)' plan is also shown, with a 'Free' pricing. The 'Standard' plan is listed at the bottom. A sidebar on the right contains a 'Summary' section with details about the service, including Region, Plan, Service name, and Resource group. At the bottom right, there are buttons for 'Create', 'View terms', and 'Cancel'.

IBM watsonx.ai Studio

Search in your workspaces

Upgrade

MUKESH K A's Account

London

Services catalog /

Cloud Object Storage

Author: IBM • Date of last update: Apr 15, 2025 • Docs • API Docs

Create About

Pricing plan

Displayed prices do not include tax. Monthly prices shown are for country or region: United States

Plan	Features	Pricing
One-Rate	One-Rate Plan is a Pay-as-You-Go option with a single, flat monthly rate (\$/GB) that includes storage, API operations, retrieval, and outbound bandwidth—making it ideal for high-activity workloads with frequent access and data transfer, such as analytics, media, and web apps. The plan includes built-in allowances that scale with stored capacity and offers automatic volume discounts as usage grows.	
Lite(deprecated)	<p>Lite plan instance is free to use for Storage capacity up to 25 GB per month. Lite plan instance is used for trial, and can be easily upgraded to Standard plan for unlimited scalability and full functionality.</p> <p>None</p> <p>Lite plan services are deleted after 30 days of inactivity.</p>	Free
Standard	<p>Standard Plan is a flexible Pay-as-You-Go option with no minimum fee—ideal for workloads with large storage needs but low or infrequent access and outbound traffic. It includes a Free Tier with 5GB of Smart Tier storage for 12 months. Charges are based on actual usage, with separate billing for storage, outbound bandwidth, API operations, and data retrieval. Multiple storage classes help you optimize costs based on how often data is accessed.</p> <p>Free Tier allowance:</p> <ul style="list-style-type: none">Storage up to 5GB/monthUp to 2000 Class A requests/monthUp to 20,000 Class B requests/monthUp to 10GB/month of data retrievalUp to 5GB/month of egress <p>Applies to aggregate total across all smart tier buckets in your account</p>	

Summary

Cloud Object Storage

Region: Global

Plan: Lite(deprecated)

Service name: Cloud Object Storage-rg

Resource group: Default

Create

View terms

Cancel

RESULT

The screenshot shows the 'Create a project' page in IBM Watsonx AI Studio. The interface is dark-themed. At the top, there's a navigation bar with the IBM Watsonx AI Studio logo, a search bar, and user account information. The main heading is 'Create a project' with a subtext 'Start with a new, blank project or select from where to import an existing project.' On the left, a sidebar shows options: '+ New', 'Local file', and 'Sample'. The main content area is titled 'Define details' and contains several form fields: 'Name' (filled with 'Predictive modeling of Access to Improved Drinking Water in India'), 'Description (optional)' (filled with a paragraph about the 78th Round of the Multiple Indicator Survey), 'Tags (optional)' (with an 'Add tags' button), 'Storage' (filled with 'Cloud Object Storage-rg'), and 'Advanced settings' (with a dropdown arrow). At the bottom right, there are 'Cancel' and 'Create' buttons.

IBM watsonx.ai Studio

Search in your workspaces

Upgrade ? 🔔 MUKESH K A's Account London

Create a project

Start with a new, blank project or select from where to import an existing project.

- + New
 - Local file
 - Sample

Define details

Name

Predictive modeling of Access to Improved Drinking Water in India

Description (optional)

This project uses data from the 78th Round of the Multiple Indicator Survey to predict the percentage of the population with access to improved drinking water across Indian states. By analyzing features like region, gender, sector, and related indicators, the model identifies patterns and disparities to support data-driven policies for equitable water access and SDG progress.

Tags (optional)

Add tags

Add tags to make projects easier to find. To add tags, separate them with commas and press Enter.

Storage

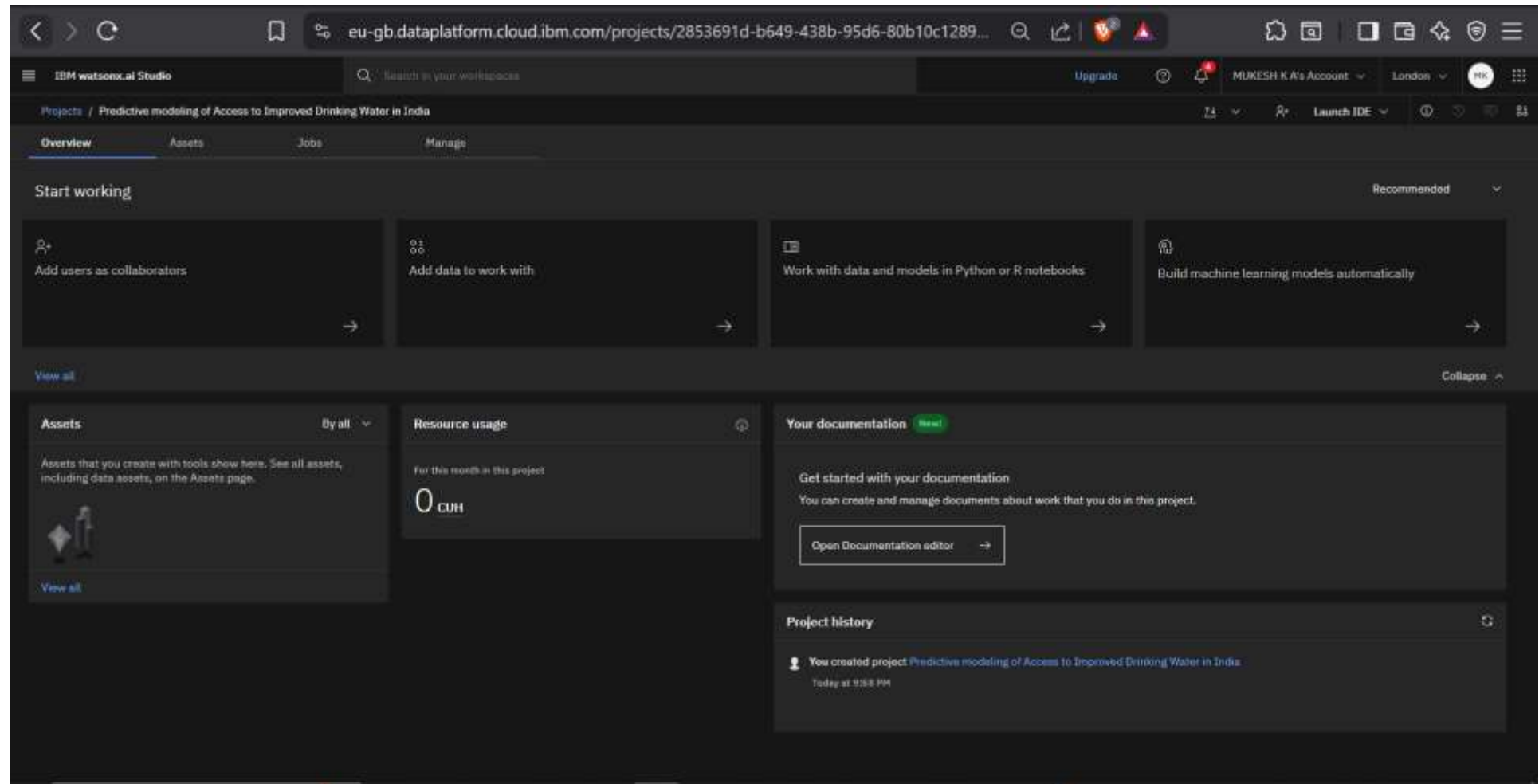
Cloud Object Storage-rg

Project includes integration with [Cloud Object Storage](#) for storing project assets.

Advanced settings

Cancel Create

RESULT



RESULT

The screenshot displays the IBM watsonx.ai Studio web interface. The browser address bar shows the URL: `eu-gb.dataplatform.cloud.ibm.com/projects/2853691d-b649-438b-95d6-80b10c1289...`. The top navigation bar includes the 'IBM watsonx.ai Studio' logo, a search bar, an 'Upgrade' button, and user account information for 'MUKESH K A's Account' in the 'London' region. The main content area is titled 'Projects / Predictive modeling of Access to Improved Drinking Water in India'. Below this, there are tabs for 'Overview', 'Assets', 'Jobs', and 'Manage', with 'Manage' being the active tab. On the left sidebar, under the 'Project' section, the 'General' tab is selected, showing options for 'Access control', 'Environments', 'Resource usage', and 'Services & integrations'. The 'Tools' section shows 'Pipeline'. The main panel displays the 'General' settings for the project, including a 'Details' section with the project name, description, and tags, and a 'Storage' section showing 'Storage used' as 0 Bytes and the bucket name 'predictivemodelingofaccesstoimpro-donotdelete-pr-d2hqc6ad23vjmw'. A 'Manage in IBM Cloud' link is also present.

IBM watsonx.ai Studio

Search in your workspaces

Upgrade

MUKESH K A's Account

London

Projects / Predictive modeling of Access to Improved Drinking Water in India

Overview Assets Jobs Manage

Project

General

Access control

Environments

Resource usage

Services & integrations

Tools

Pipeline

General

Details

Name

Predictive modeling of Access to Improved Drinking Water in India

Description

This project uses data from the 78th Round of the Multiple Indicator Survey to predict the percentage of the population with access to improved drinking water across Indian states. By analyzing features like region, gender, sector, and related indicators, the model identifies patterns and disparities to support data-driven policies for equitable water access and SDG progress.

Tags

Add tags to make projects easier to find.

Project ID

2853691d-b649-438b-95d6-80b10c1289c2

Storage

Storage used

0 Bytes

Bucket

predictivemodelingofaccesstoimpro-donotdelete-pr-d2hqc6ad23vjmw

Manage in IBM Cloud

RESULT

The screenshot shows the 'Associate service' dialog in IBM watsonx.ai Studio. The dialog prompts the user to 'Choose an existing or add a new service to associate with your project.' It features filters for 'Default' and 'Locations', a search bar for 'Find services', and a table of available services. The table has columns for Name, Type, Plan, Location, Status, and Group. One service, 'watsonx.ai Runtime-xw', is listed with a status of 'Not associated'. At the bottom, there are 'Cancel' and 'Associate' buttons.

Name	Type	Plan	Location	Status	Group
<input checked="" type="checkbox"/> watsonx.ai Runtime-xw ⓘ	watsonx.ai Runtime	Lite	London	Not associated	Default

RESULT

The screenshot displays the IBM watsonx.ai Studio web interface. The browser address bar shows the URL: eu-gb.dataplatform.cloud.ibm.com/projects/2853691d-b649-438b-95d6-80b10c... The page title is 'IBM watsonx.ai Studio'. The main heading is 'Build machine learning models automatically'. Below this, it says 'Define the details to create an AutoAI experiment asset and open it in the AutoAI tool.' The form is divided into two main sections: 'Define details' and 'Define configuration'. In the 'Define details' section, the 'Name' field is filled with 'Predictive modeling of Access to Improved Drinking Water in India'. The 'Description (optional)' field contains the text: 'This project uses data from the 78th Round of the Multiple Indicator Survey to predict the percentage of the population with access to improved drinking water across Indian states. By analyzing features like region, gender, sector, and related indicators, the model identifies patterns and disparities to...'. The 'Tags (optional)' field is empty. In the 'Define configuration' section, the 'watsonx.ai Runtime service instance' is set to 'watsonx.ai Runtime-xw'. The 'Environment definition' is set to 'Large: 8 CPU and 32 GB RAM'. A note below states: 'This environment definition consumes 20 capacity units per hour for training. For details, see watsonx.ai Runtime plans.' At the bottom of the form, there are three buttons: 'Cancel', 'Back', and 'Create'.

IBM watsonx.ai Studio

eu-gb.dataplatform.cloud.ibm.com/projects/2853691d-b649-438b-95d6-80b10c...

Projects / Predictive modeling of Access to Improved Drinking Water in India

Build machine learning models automatically

Define the details to create an AutoAI experiment asset and open it in the AutoAI tool.

Define details

Name

Predictive modeling of Access to Improved Drinking Water in India

Description (optional)

This project uses data from the 78th Round of the Multiple Indicator Survey to predict the percentage of the population with access to improved drinking water across Indian states. By analyzing features like region, gender, sector, and related indicators, the model identifies patterns and disparities to...

Tags (optional)

Add tags to make assets easier to find.

Start typing to add tags

Define configuration

watsonx.ai Runtime service instance

watsonx.ai Runtime-xw

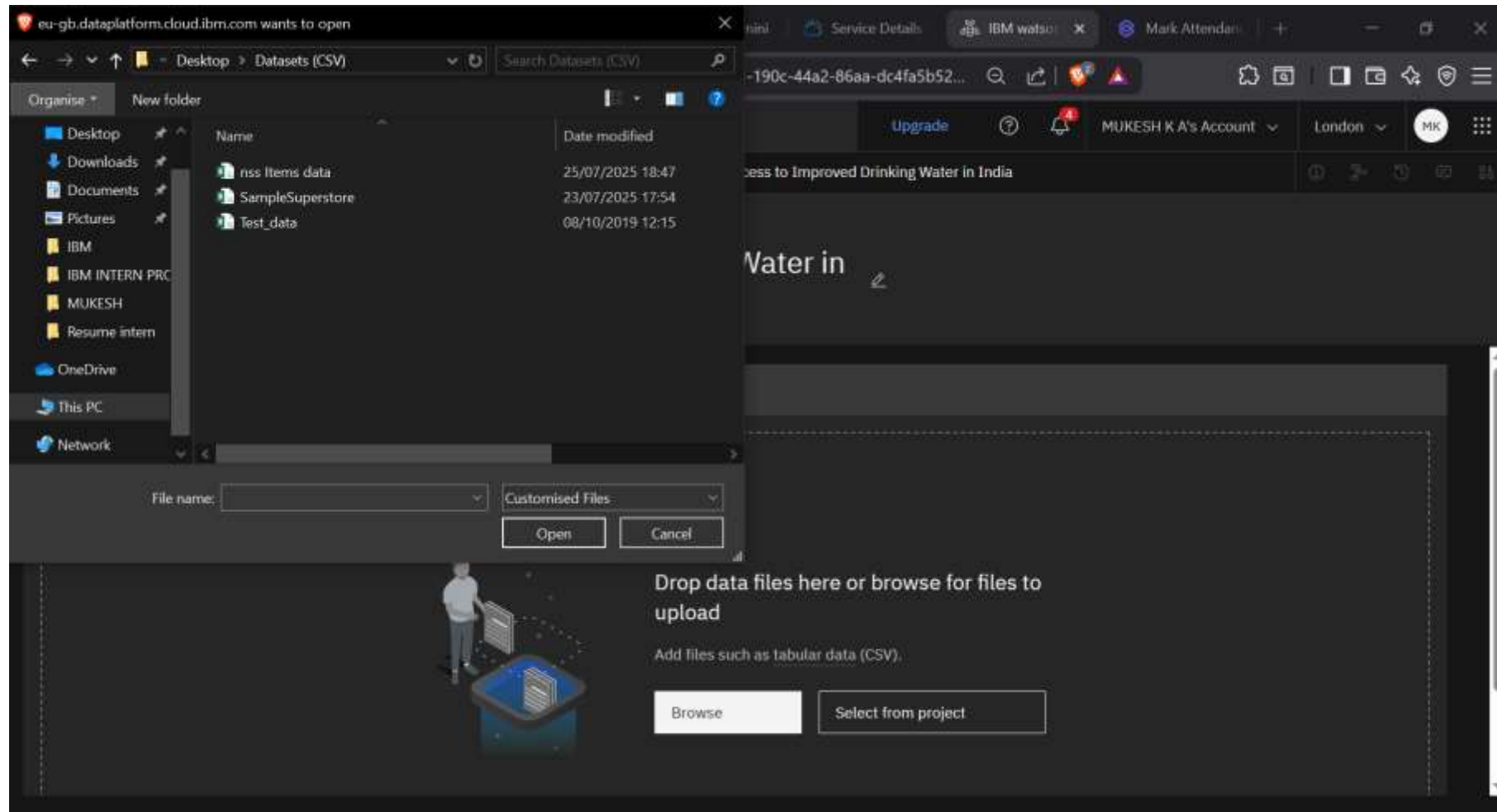
Environment definition ⓘ

Large: 8 CPU and 32 GB RAM

This environment definition consumes 20 capacity units per hour for training. For details, see watsonx.ai Runtime plans.

Cancel Back Create

RESULT



RESULT

The screenshot displays the IBM Watsonx AI Studio web interface. The browser address bar shows the URL: `eu-gb.dataplatform.cloud.ibm.com/ml/auto-ml/6c207cf3-190c-44a2-86aa-dc4fa5b52...`. The interface includes a top navigation bar with the IBM Watsonx AI Studio logo, a search bar, an 'Upgrade' button, and user account information for 'MUKESH K A's Account' in 'London'. The main header area shows the project path: 'Projects / Predictive modeling of Access to Improved Drinking Water in India / Predictive modeling of Access to Improved Drinking Water in India'. Below this, the title 'Predictive modeling of Access to Improved Drinking Water in India' is displayed, with an 'Autosaved: 22:04:26' timestamp. The interface is divided into two main panels. The left panel, titled 'Add data source', contains a dashed box with the text 'Add files such as tabular data (CSV)'. Inside this box are two buttons: 'Browse' and 'Select from project'. Below the dashed box, a file named 'nss Items data.csv' is listed with a size of '122.16 KB' and 'Columns: 6'. The right panel, titled 'Configure details', features a section 'Create a time series analysis?' with a line graph icon. The text below this section states: 'Enable this option to predict future activity over a specified date/time range. Data must be structured and sequential. [Learn more](#)'. To the right of this text are two buttons: 'Yes' and 'No'.

RESULT

The screenshot displays the IBM Watsonx.ai Studio interface. At the top, the header includes the IBM Watsonx.ai Studio logo, a search bar, and user account information (MUKESH K A's Account). The main title of the experiment is "Predictive modeling of Access to Improved Drinking Water in India". Below the title, the interface is divided into two main sections. On the left, there is a file selection area with buttons for "Browse" and "Select from project". A file named "nss Items data.csv" is listed with a size of 122.16 KB and 6 columns. On the right, the "What do you want to predict?" section shows the "Prediction column" set to "Value". Below this, the "Prediction type" is set to "Regression", and the "Optimized for" metric is "RMSE & run time". A large green checkmark is visible in the background. At the bottom right, there is a blue "Run experiment" button. The interface also shows "Autosaved: 22:04:26" and "CUH remaining: 10.73 CUH".

IBM watsonx.ai Studio

Search in your workspaces

Upgrade

MUKESH K A's Account

London

MK

Projects / Predictive modeling of Access to Improved Drinking Water in India / Predictive modeling of Access to Improved Drinking Water in India

Configure AutoAI experiment

Predictive modeling of Access to Improved Drinking Water in India

Autosaved: 22:04:26

Browse

Select from project

nss Items data.csv

Size: 122.16 KB | Columns: 6

What do you want to predict?

Prediction column

Value

Prediction column: Value

CUH remaining: 10.73 CUH

PREDICTION TYPE

Regression

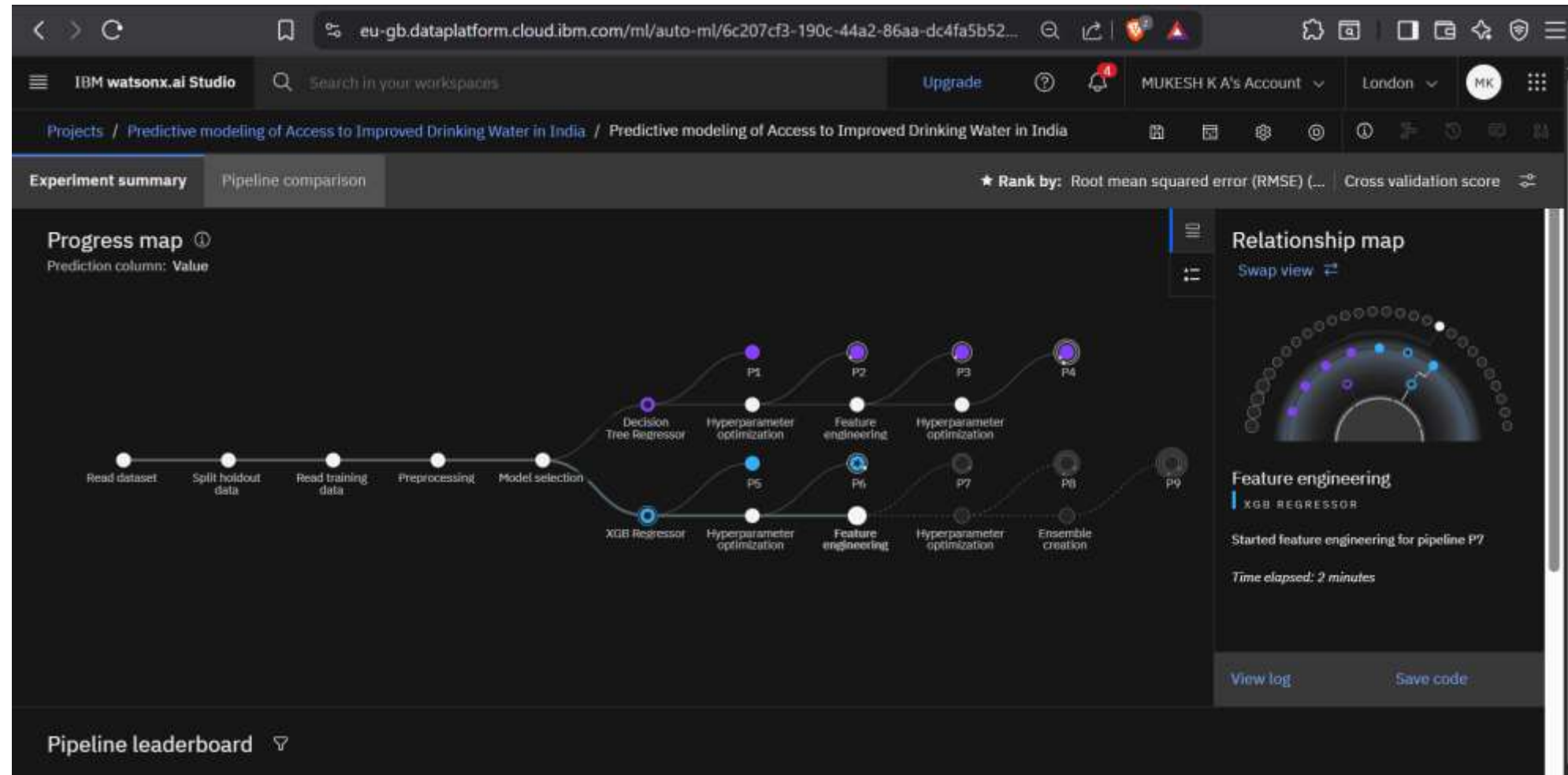
OPTIMIZED FOR

RMSE & run time

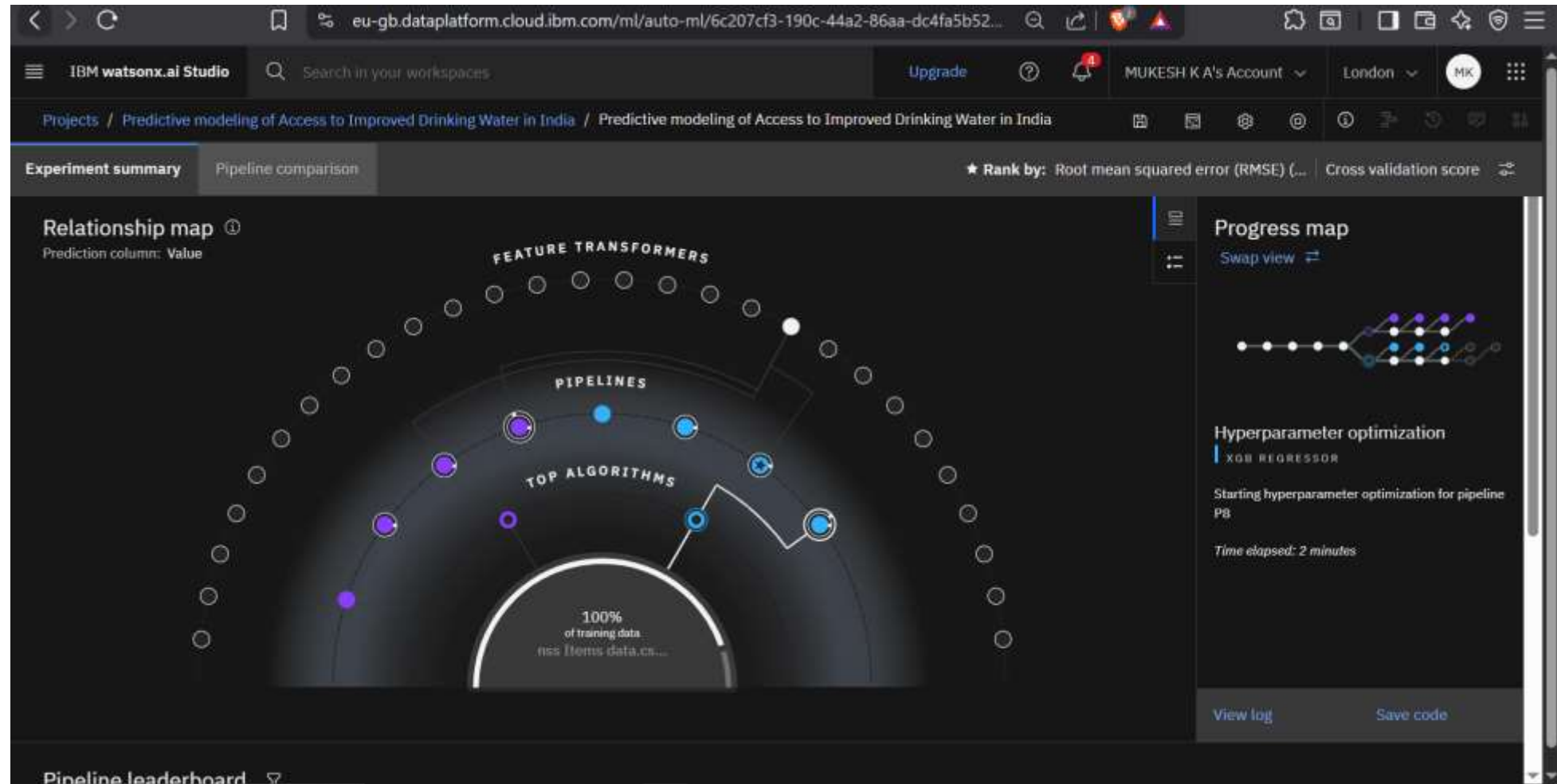
Experiment settings

Run experiment

RESULT



RESULT



RESULT

IBM watsonx.ai Studio

Search in your workspaces

Upgrade

MUKESH K A's Account

London

MK

Projects / Predictive modeling of Access to Improved Drinking Water in India / Predictive modeling of Access to Improved Drinking Water in India

Experiment summary Pipeline comparison

★ Rank by: Root mean squared error (RMSE) (...) Cross validation score

nss Items data.cs...

View log Save code

Pipeline leaderboard

	Rank	Name	Algorithm	Specialization	RMSE (Optimized) Cross Validation	Enhancements	Build time
★	1	Pipeline 8	XGB Regressor		4.827	HPO-1 FE HPO-2	00:00:33
	2	Pipeline 7	XGB Regressor		4.827	HPO-1 FE	00:00:25
	3	Pipeline 6	XGB Regressor		4.860	HPO-1	00:00:06
	4	Pipeline 5	XGB Regressor		4.860	None	00:00:01

RESULT

The screenshot shows the IBM watsonx.ai Studio interface. At the top, the browser address bar displays the URL: eu-gb.dataplatform.cloud.ibm.com/ml/auto-ml/6c207cf3-190c-44a2-86aa-dc4fa5b52... The studio header includes the 'IBM watsonx.ai Studio' logo, a search bar, an 'Upgrade' button, and user information for 'MUKESH K A's Account' in 'London'. The breadcrumb trail indicates the current project: 'Projects / Predictive modeling of Access to Improved Drinking Water in India / Predictive modeling of Access to Improved Drinking Water in India'.

The 'Save as' dialog is open, featuring a left sidebar with 'Experiments' (selected), 'Pipelines', and a star icon. The main area is divided into two sections:

- Select asset type:** Contains two options: 'Model' (selected with a checkmark) and 'Notebook'. The 'Model' description reads: 'Create a watsonx.ai Runtime model asset that you can test with new data, deploy to generate predictions, and trace lineage activity.' The 'Notebook' description reads: 'Create a notebook if you want to view the code that created this model pipeline or interact with with the model programatically.'
- Define details:** Contains three fields:
 - Name:** A text box containing 'P8 - XGB Regressor: Predictive modeling of Access to Improved Drinking W'.
 - Description (optional):** A text box containing 'Model description'.
 - Tags:** A section with the instruction 'Add tags to make assets easier to find.' and a text box labeled 'Add a tag' with a plus icon.

At the bottom right of the dialog are two buttons: 'Cancel' and 'Create'.

RESULT

The screenshot displays the IBM watsonx.ai Studio interface. The top navigation bar includes the logo, a search bar, an 'Upgrade' button, a help icon, a notification bell with a red '4', the user's account 'MUKESH K A's Account', the location 'London', and a profile icon 'MK'. The breadcrumb trail shows the path: Projects / ... / P8 - XGB Regressor: Predictive modeling of Access to Improved Drinking Water in India.

The main content area is divided into two panels. The left panel, titled 'Input (1)', contains a table with the following data:

Column	Type
Age Group	other
Gender	other
Indicator	other
Sector	other
State	other

The right panel, titled 'About this asset', provides details about the model asset. It includes fields for Name, Description, Asset Details, Tags, Last modified, and Created on. The Asset Details section shows the Type as 'wml-hybrid_0.1' and the Model ID as '6f83dd5a-7d7a-47...'. The Software specification is 'hybrid_0.1' and the Hybrid pipeline software specifications are 'autoai-kb_rt24.1-py3.11'.

About this asset

Name
P8 - XGB Regressor: Predictive modeling of Access to Improved Drinking Water in India

Description
No description provided.

Asset Details
Type: wml-hybrid_0.1
Model ID: 6f83dd5a-7d7a-47...
Software specification: hybrid_0.1
Hybrid pipeline software specifications: autoai-kb_rt24.1-py3.11

Tags
Add tags to make assets easier to find.

Last modified
23 seconds ago by MUKESH K A

Created on

RESULT

The screenshot shows the 'Promote to space' dialog box in IBM Watsonx.ai Studio. The dialog is titled 'Promote to space' and includes the instruction: 'Promote the asset to a deployment space to deploy the asset or to support a deployment.'

Target deployment space

Predictive modeling of Access to Improved Drinkin... x v

Why don't I see all of my spaces? ⓘ

☐ Go to the model in the space after promoting it

Description (Optional)

Description of assets

Selected assets (1)

Name	Format	Version	Status
PB - XGB Regressor: Predictive mo...	Model	Cur... v	Queued

Promoting an asset promotes dependent assets as well. For example, promoting a model also promotes the associated software specification and package extensions. You will see all promoted assets in the target space.

Cancel Promote

RESULT

The screenshot displays the IBM Watsonx AI Studio interface. The main heading is "Promote to space" with a subtext: "Promote the asset to a deployment space to deploy the asset or to support a deployment." Below this, a green checkmark indicates "Promotion completed." A table titled "Selected assets (1)" shows the promoted asset:

Name	Format	Version	Status
P8 - XGB Regressor: Predictive modeling of Acc...	Model	Current	Promoted

Below the table, a note states: "Promoting an asset promotes dependent assets as well. For example, promoting a model also promotes the associated software specification and package extensions. You will see all promoted assets in the target space." A blue "Close" button is located at the bottom right of the modal.

A success notification popup is visible in the top right corner, stating: "Success: Successfully promoted P8 - XGB Regressor: Predictive modeling of Access to Improved Drinking Water in India to the deployment space. Go to the [deployment space](#) to prepare the assets for deployment. Timestamp 10:13:13 PM".

RESULT

The screenshot displays the IBM watsonx.ai Studio interface. At the top, the header includes the logo, a search bar, an 'Upgrade' button, a help icon, a notification bell, and user account information for 'MUKESH K A's Account' in the 'London' region. Below the header, the main title 'Predictive modeling of Access to Improved Drinking Water in India' is centered. A navigation bar contains tabs for 'Overview', 'Assets' (which is selected), 'Deployments', 'Jobs', and 'Manage'. On the left side, there is a sidebar with a search bar 'Find assets', a list of '1 asset' under 'All assets', and 'Asset types' including 'Models' with a count of 1. The main content area shows a table titled 'All assets' with two columns: 'Name' and 'Last modified'. The table contains one entry: 'PB - XGB Regressor: Predictive modeling of Access to Improved Drinking Water in India', which is a 'Machine learning model from AutoAI' created '21 seconds ago' by 'MUKESH K A (You)'. At the bottom, there is a pagination bar showing 'Items per page: 20', '1-1 of 1 items', and '1 of 1 pages'.

IBM watsonx.ai Studio

Search in your workspaces

Upgrade

MUKESH K A's Account

London

Deployment spaces /

Predictive modeling of Access to Improved Drinking Water in India

Overview Assets Deployments Jobs Manage

Find assets

Import assets

New asset

1 asset

All assets

Asset types

Models

Name	Last modified
PB - XGB Regressor: Predictive modeling of Access to Improved Drinking Water in India Machine learning model from AutoAI	21 seconds ago MUKESH K A (You)

Items per page: 20 1-1 of 1 items 1 of 1 pages

RESULT

The screenshot shows the 'Create a deployment' dialog box in the IBM Watson AI Studio interface. The dialog is titled 'Create a deployment' and has a close button (X) in the top right corner. It is divided into two main sections: 'Define details' and 'Name'. In the 'Define details' section, there is a label 'Associated asset' with a dropdown menu showing 'PB - XGB Regressor: Predictive modeling of Access to Improved Drinking Water in India'. Below this, there is a 'Deployment type' section with two options: 'Online' and 'Batch'. The 'Online' option is selected, indicated by a checkmark and a description: 'Run the model on data in real-time, as data is received by a web service.' The 'Batch' option is described as: 'Run the model against data as a batch process.' In the 'Name' section, there is a text input field containing the text 'Predictive modeling of Access to Improved Drinking Water in India'. At the bottom of the dialog, there are two buttons: 'Cancel' and 'Create'.

eu-gb.dataplatform.cloud.ibm.com/ml-runtime/models/88bc4273-9b9c-485a-80d0-6...

IBM watsonx.ai Studio

Deployment spaces / ... / PB - XGB Regressor: Predictive modeling of Access to Improved Drinking Water in India

Create a deployment

Define details

Associated asset
PB - XGB Regressor: Predictive modeling of Access to Improved Drinking Water in India

Deployment type

Online ☒

Run the model on data in real-time, as data is received by a web service.

Batch

Run the model against data as a batch process.

Name

Predictive modeling of Access to Improved Drinking Water in India

Cancel Create

RESULT

The screenshot displays the IBM watsonx.ai Studio interface. The browser address bar shows the URL: eu-gb.dataplatform.cloud.ibm.com/ml-runtime/models/88bc4273-9b9c-485a-80d0-6... The page title is "Deployment spaces / ... / PB - XGB Regressor: Predictive modeling of Access to Improved Drinking Water in India". The main content area is divided into two tabs: "Deployments" (active) and "Model details". The "Deployments" tab shows a table with one deployment item:

Name	Type	Status	Tags	Last modified
Predictive modeling of Access to Improved Drinking Water in India	Online	Initializing		6 seconds ago MUKESH K A (You)

At the top right of the table is a "New deployment" button. The right sidebar contains information about the asset:

- About this asset**
- Name:** PB - XGB Regressor: Predictive modeling of Access to Improved Drinking Water in India
- Description:** No description provided.
- Asset Details:**
 - Type: wml-hybrid_0.1
 - Model ID: 88bc4273-9b9c-48...
 - Software specification: hybrid_0.1
 - Hybrid pipeline software specifications: autoai-kb_rt24.1-py3.11
- Tags:** Add tags to make assets easier to find.
- Source asset details:**
- Last modified:**

At the bottom left, it says "Items per page: 20" and "1-1 of 1 items". At the bottom right, it says "1 of 1 pages".

RESULT

The screenshot displays the IBM watsonx.ai Studio interface. The top navigation bar includes the logo, a search bar, an 'Upgrade' button, a help icon, a notification bell, the user account 'MUKESH K A's Account', the location 'London', and a profile icon 'MK'. The breadcrumb trail shows the path: 'Deployment spa...' / 'Predictive modeling of Access to Improved Drinking Water in...' / 'P8 - XGB Regressor: Predictive modeling of Access to Improved Drinking Water L...'. The main content area is titled 'Predictive modeling of Access to Improved Drinking Water in India' and indicates the model is 'Deployed' and 'Online'. Below the title, there are tabs for 'API reference' and 'Test'. The 'API reference' tab is active, showing 'Endpoints for scoring' with a help icon. It lists a 'Private endpoint' and a 'Public endpoint', both with their respective URLs and a 'Bearer <token>' field set to 'IAM'. A link to 'Learn more about the 2021-05-01 version query parameter' is provided. Below the endpoints, there is a 'Code snippets' section with tabs for 'cURL', 'Java', 'JavaScript', 'Python', and 'Scala'. The 'cURL' tab is active, showing a note: '# NOTE: you must set \$API_KEY below using information retrieved from your IBM Cloud account (https://eu-gb.dataplatform.cloud)'. On the right side, a sidebar titled 'About this deployment' provides details about the deployment, including the name, description, deployment ID, serving name, software specification, and tags.

IBM watsonx.ai Studio

Search in your workspaces

Upgrade

MUKESH K A's Account

London

MK

Deployment spa... / Predictive modeling of Access to Improved Drinking Water in... / P8 - XGB Regressor: Predictive modeling of Access to Improved Drinking Water L...

Predictive modeling of Access to Improved Drinking Water in India

Deployed Online

API reference Test

Endpoints for scoring

Private endpoint

Public endpoint

Learn more about the 2021-05-01 version query parameter

Code snippets

cURL Java JavaScript Python Scala

NOTE: you must set \$API_KEY below using information retrieved from your IBM Cloud account (https://eu-gb.dataplatform.cloud)

About this deployment

Name

Predictive modeling of Access to Improved Drinking Water in India

Description

No description provided.

Deployment Details

Deployment ID: 6db7e34d-1af5-45...

Serving name:

No serving name.

Software specification:

hybrid_0.1

Hybrid pipeline software specifications:

autoai-kb_rt24.1-py3.11

Copies:

1

Tags

Add tags to make assets easier to find.

Associated asset

P8 - XGB Regressor: Predictive modeli

RESULT

IBM watsonx.ai Studio

Search in your workspaces

Upgrade

?

5

MUKESH K A's Account

London

MK

Deployment spaces / Predictive modeling of Access to Improved Drinking Water in India / P8 - XGB Regressor: Predictive modeling of Access to Improved Drinking Water in India /

Predictive modeling of Access to Improved Drinking Water in India 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

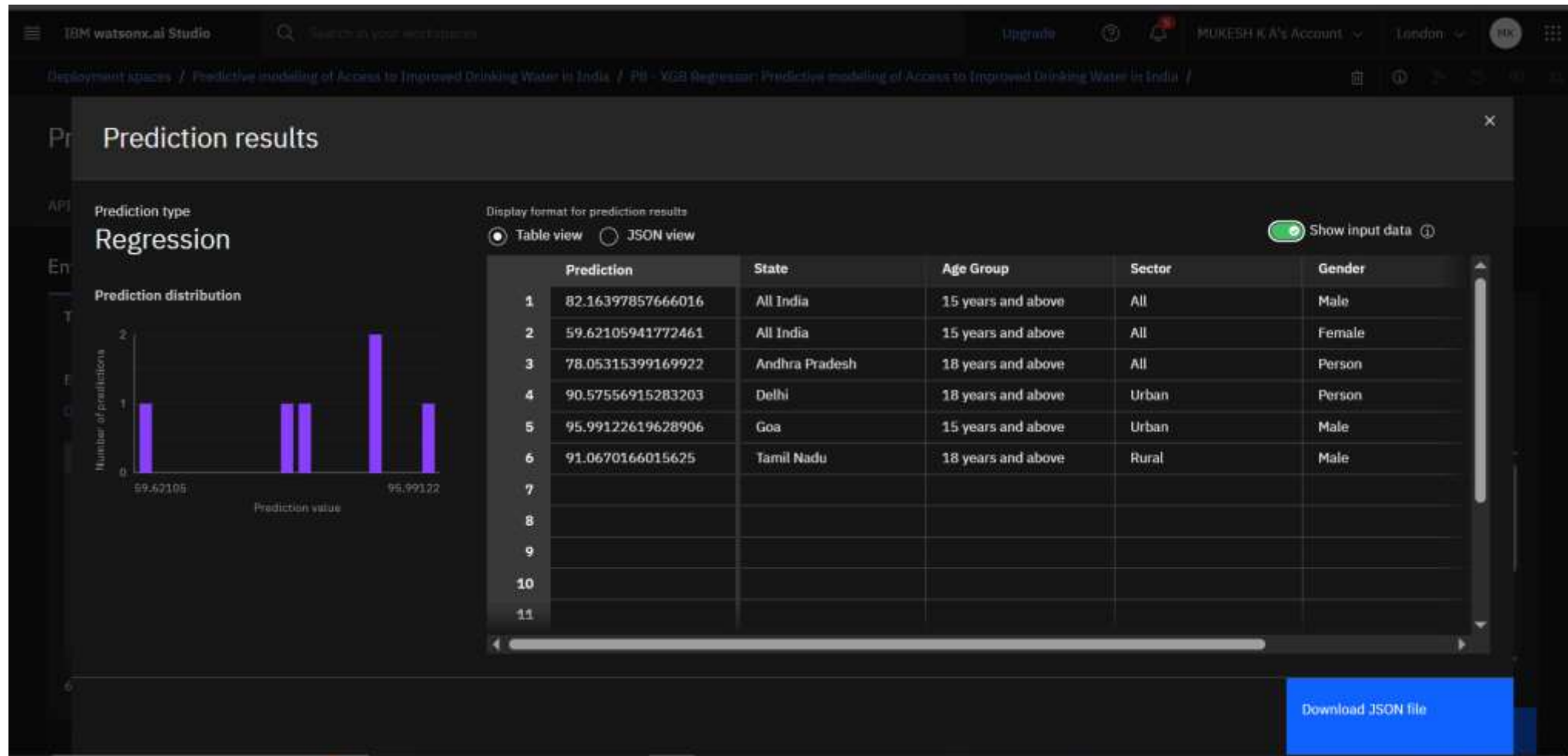
Clear all

	State (other)	Age Group (other)	Sector (other)	Gender (other)	Indicator (other)
1	All India	15 years and above	All	Male	Percentage of Persons Who Used MoE
2	All India	15 years and above	All	Female	Percentage of Persons Who Used MoE
3	Andhra Pradesh	18 years and above	All	Person	Percentage of Persons Who Used MoE
4	Delhi	18 years and above	Urban	Person	Percentage of Persons Who Used MoE
5	Goa	15 years and above	Urban	Male	Percentage of Persons Who Used MoE
6	Tamil Nadu	18 years and above	Rural	Male	Percentage of Persons Who Used MoE

6 rows, 5 columns

Predict

RESULT



CONCLUSION

- The system successfully predicted the percentage of the population with access to improved drinking water using IBM Cloud's AutoAI. The XGBoost Regressor model achieved high accuracy, demonstrating the effectiveness of machine learning in analyzing large-scale public datasets.
- The solution provides valuable insights into regional and demographic disparities in water access, supporting evidence-based decisions to advance India's Sustainable Development Goals (SDG 6).

Challenges Faced

- Data preprocessing inside AutoAI offered limited manual control
- Some indicators were text-heavy and required standardization
- Deployment initialization took time due to cloud processing dependencies

Potential Improvements

- Include more features (e.g., income level, education) if available
- Visualize state-level data using maps for better clarity
- Add interactive filtering to explore different indicator categories

FUTURE SCOPE

Incorporate Additional Data Sources

- Include data on income level, education, sanitation, and health access
- Use government datasets from NFHS or Census of India for richer modeling

Expand Geographic Coverage

- Apply the model to district-level or village-level data for hyper-local insights
- Compare predictions across states and union territories

Improve Model Performance

- Fine-tune hyperparameters manually if AutoAI customization is enabled
- Experiment with ensemble models or deep learning if supported in future environments

FUTURE SCOPE

Build Interactive Dashboard

- Use Streamlit or Power BI to create dashboards for policymakers
- Enable filters by region, gender, or indicator type for real-time exploration

Integrate Emerging Technologies

- Use Edge AI for local health monitoring in rural areas
- Explore geospatial ML models and time-series forecasting for predictive planning

REFERENCES

- **IBM Watsonx.ai Studio Documentation**
IBM Cloud official documentation for AutoAI and deployment
<https://dataplatform.cloud.ibm.com>
- **AI Kosh Dataset Link –**
https://aikosh.indiaai.gov.in/web/datasets/details/improved_source_of_drinking_water_multiple_indicator_survey_78th_round.html
- **Scikit-learn Documentation – Regression Models**
https://scikit-learn.org/stable/supervised_learning.html
- **XGBoost Documentation – Extreme Gradient Boosting**
<https://xgboost.readthedocs.io>

IBM CERTIFICATIONS



IBM CERTIFICATIONS



IBM CERTIFICATIONS





THANK YOU