# Lab 1: Building Machine Learning models in a No-code environment

# BigML Diabetes Prediction: Classification & Regression Exercise

This document contains exercises for performing both classification and regression analysis using the Diabetes dataset in BigML. You will learn how to predict whether a patient has diabetes (classification) and how to estimate glucose levels (regression).

# Part 1: Binary Classification for Diabetes Prediction

## Objective

This exercise guides you through using BigML to perform a binary classification task to predict whether a patient has diabetes using the Diabetes dataset.

### Step 1: Upload the Diabetes Dataset

1. Download the [Diabetes dataset (diabetes.csv)](https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database) from Kaggle.  
2. Log in to BigML (<https://bigml.com/> ).

3. Go to Dashboard.   
4. Go to Sources and click 'Upload a local file'.  
5. Upload the diabetes.csv file.  
6. Click 'Create' to process the dataset.

### Step 2: Create a Dataset

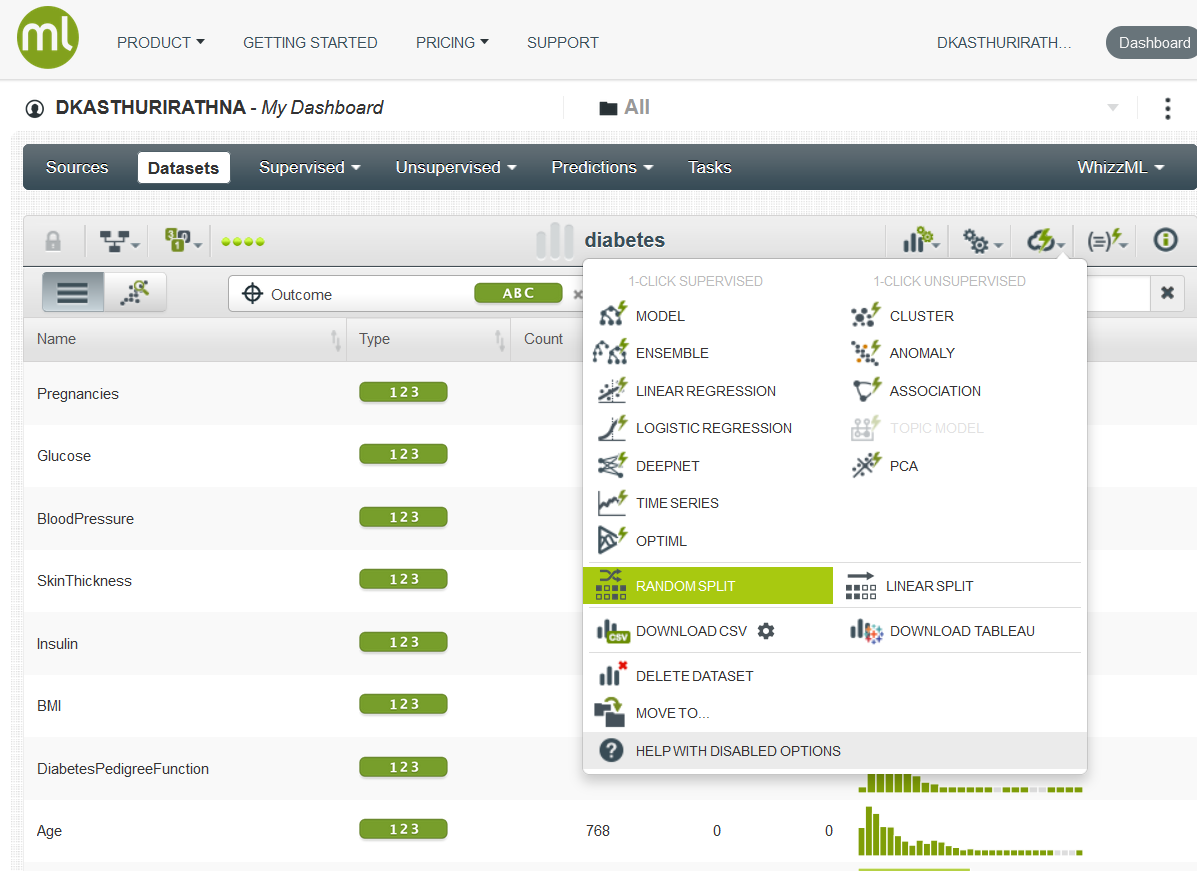
1. Click '1-Click Dataset' to generate a structured dataset.  
2. BigML will process the dataset and display field information.

### Step 3: Set the Target Variable

1. Go to the Datasets tab and open the dataset and navigate to the 'Fields' tab.  
2. Click on the 'Outcome' field.  
3. In the settings panel, set 'Objective = Yes' (marks it as the target variable 🎯).

### Step 4: Split the Dataset

1. Click 'Split' on the dataset page.  
2. Choose 'Random Split' and set:  
 - 80% for training.  
 - 20% for testing.  
3. Click 'Create'.



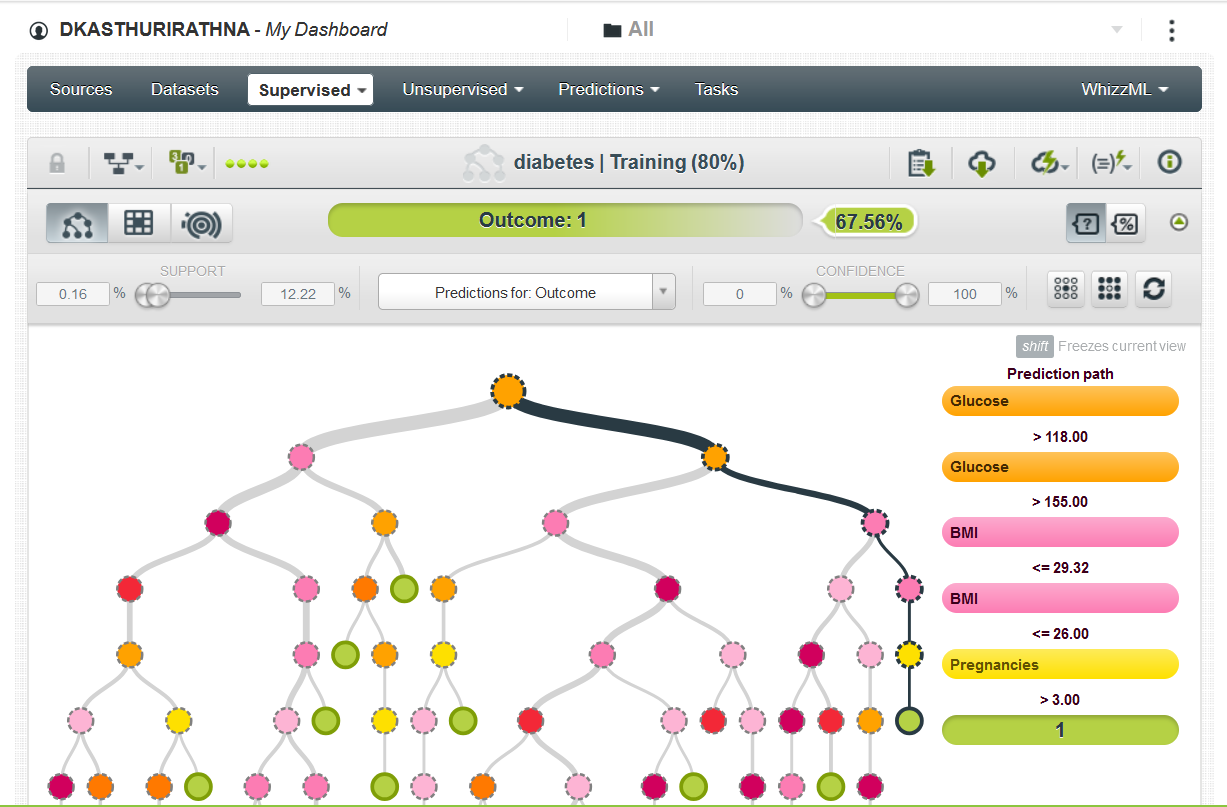
### Step 5: Train a Model

1. Open the training dataset.  
2. Under ‘1-Click supervised’ click 'Model' to train a Decision Tree model.

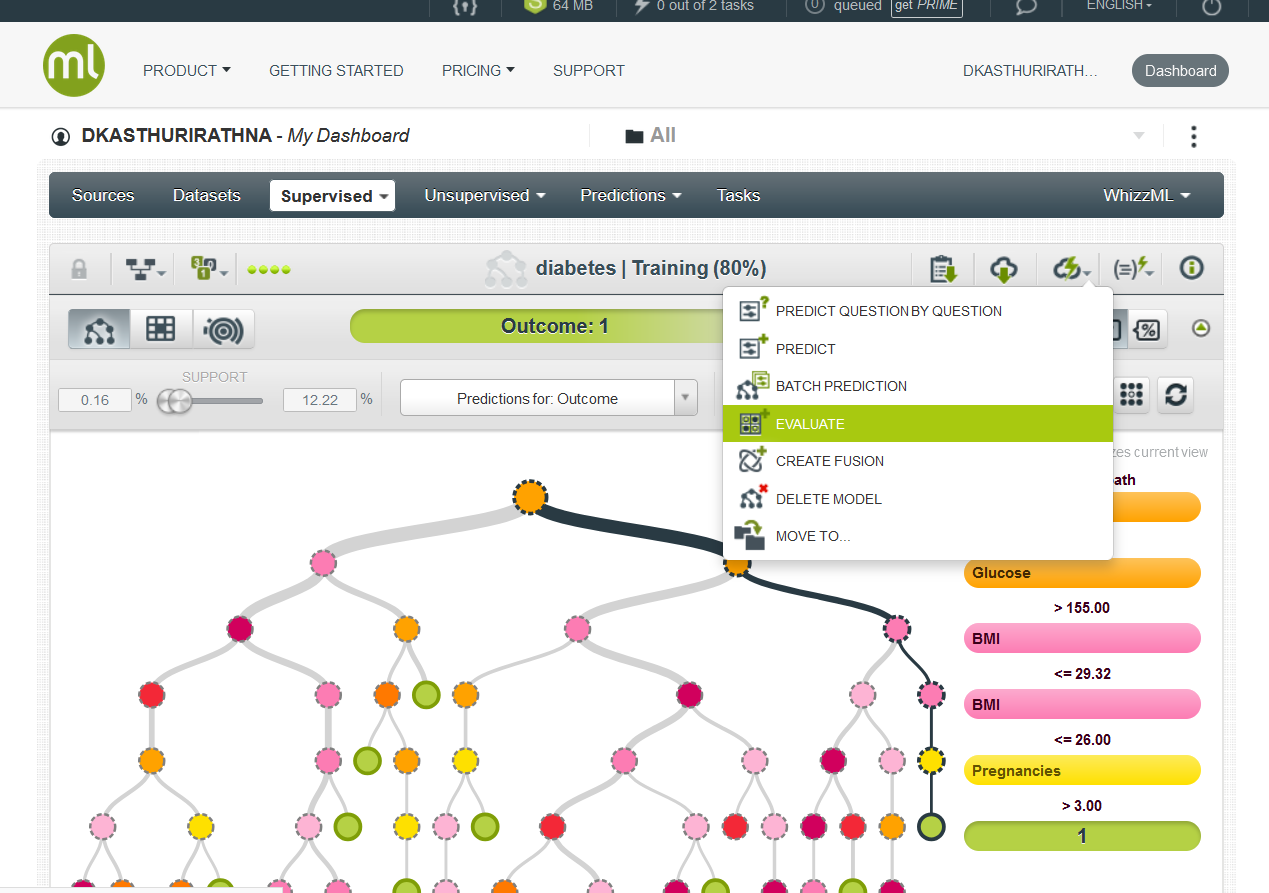
### Step 6: Evaluate the Model

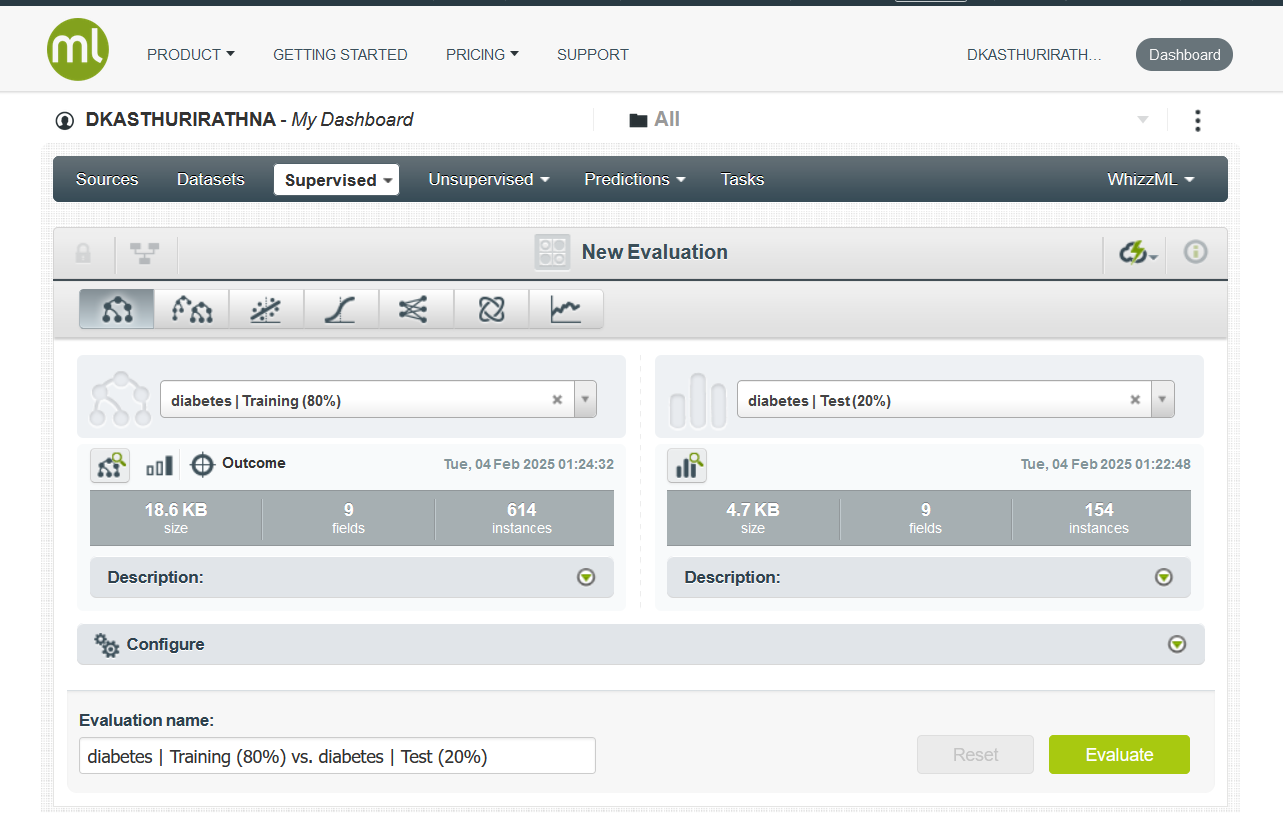
1. Open the model details page (Supervised->Models).

2. Click the path to visualize the model structure. (Note: After clicking a leaf node, click the root node to get the full view). Navigate a path till a leaf node to extract a rule.



3. Click 'Evaluate'.

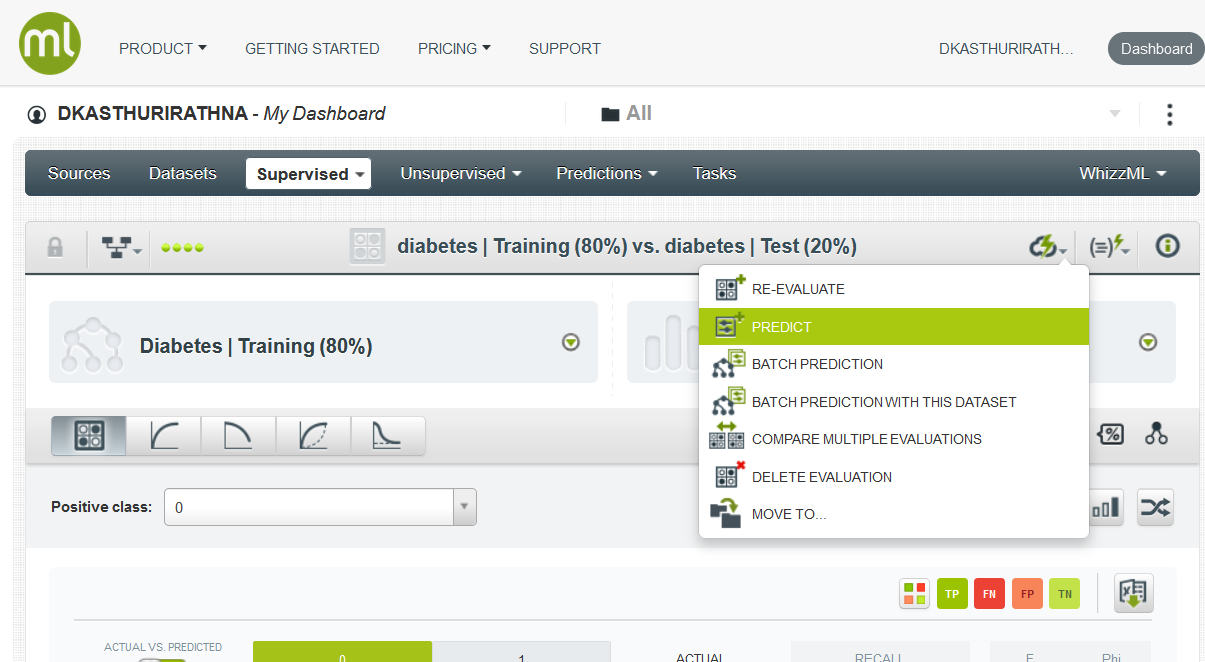




4. Select the testing dataset.  
5. Review performance metrics such as:  
 - Accuracy  
 - Precision  
 - Recall  
 - F1-Score  
 - Confusion Matrix.

### Step 7: Make Predictions

1. Click 'Predict' in the model details page.



2. Enter input values (e.g., Glucose, BMI, Age, etc.).  
3. Click 'Predict' → BigML will return either 1 (Diabetes) or 0 (No Diabetes).

# Part 2: Regression Analysis for Glucose Level Prediction

## Objective

This exercise guides you through performing regression analysis in BigML to predict Glucose levels using the Diabetes dataset.

### Step 1: Set the Target Variable to Glucose

1. Open the dataset and navigate to the 'Fields' tab.  
2. Click on the 'Glucose' field.  
3. In the settings panel, set 'Objective = Yes' (marks it as the target variable 🎯).

### Step 2: Split the Dataset

1. Click 'Split' on the dataset page.  
2. Choose 'Random Split' and set:  
 - 80% for training.  
 - 20% for testing.  
3. Click 'Create'.

### Step 3: Train a Model

1. Open the training dataset.  
2. Under ‘1-Click supervised’ click 'Model' to train a Decision Tree model.

### Step 4: Evaluate the Model

1. Open the model details page.  
2. Click 'Evaluate'.  
3. Select the testing dataset.  
4. Review performance metrics such as:  
 - Root Mean Squared Error (RMSE)  
 - R² Score  
 - Error distribution.  
5. Click the path to visualize the regression model structure.

### Step 5: Make Predictions

1. Click 'Predict' in the model details page.  
2. Enter input values (e.g., Pregnancies, BMI, Age, etc.).  
3. Click 'Predict' → BigML will return the predicted Glucose level.

## Submission

Upload the Results pages of both exercises as screenshots in a single zip file. On the screenshots the username (on top left) should be your username. The zip file name should be your registration number.

## Summary

✔ You successfully performed both classification (Diabetes Prediction) and regression (Glucose Level Prediction) in BigML.  
✔ You trained, evaluated, and visualized both models.  
✔ You uploaded screenshots as proof of completion.

## Optional Challenges

* Try Feature Engineering: Create new features from the dataset.
* Explore other low-code/no-code environments like <https://h2o.ai/> ,Google cloud API ([Cloud APIs | Google Cloud](https://cloud.google.com/apis)) and [Orange Data Mining](https://orangedatamining.com/).
* Experiment with Ensemble Models for improved accuracy.
* Explore Advanced Evaluation Metrics such as Lift Charts.
* Export these models as web services to client apps.

## References

BigML tutorial videos

<https://www.youtube.com/watch?v=6xbNpILmQYo&list=PL1bKyu9GtNYHAk0PUojkLYZzASoYVcsTQ>