



Azure ML Classic Studio


Predicting Movie Ratings using Regression Model in AzureML Classic Studio.

This model (Pipeline) trains a linear regressor to predict a movie ratings prediction based on technical features. Because you're trying to answer the question "How much?" this is called a regression problem. However, you can apply the same fundamental steps in this example to tackle any type of machine learning problem whether it be regression, classification, clustering, and so on.

Gallery Link:

Movie Ratings prediction [19K45A04C6]

Movie Ratings prediction using pre-available dataset and training the model using Linear Regression. Tags: Linear Regression, Azure ML, Microsoft

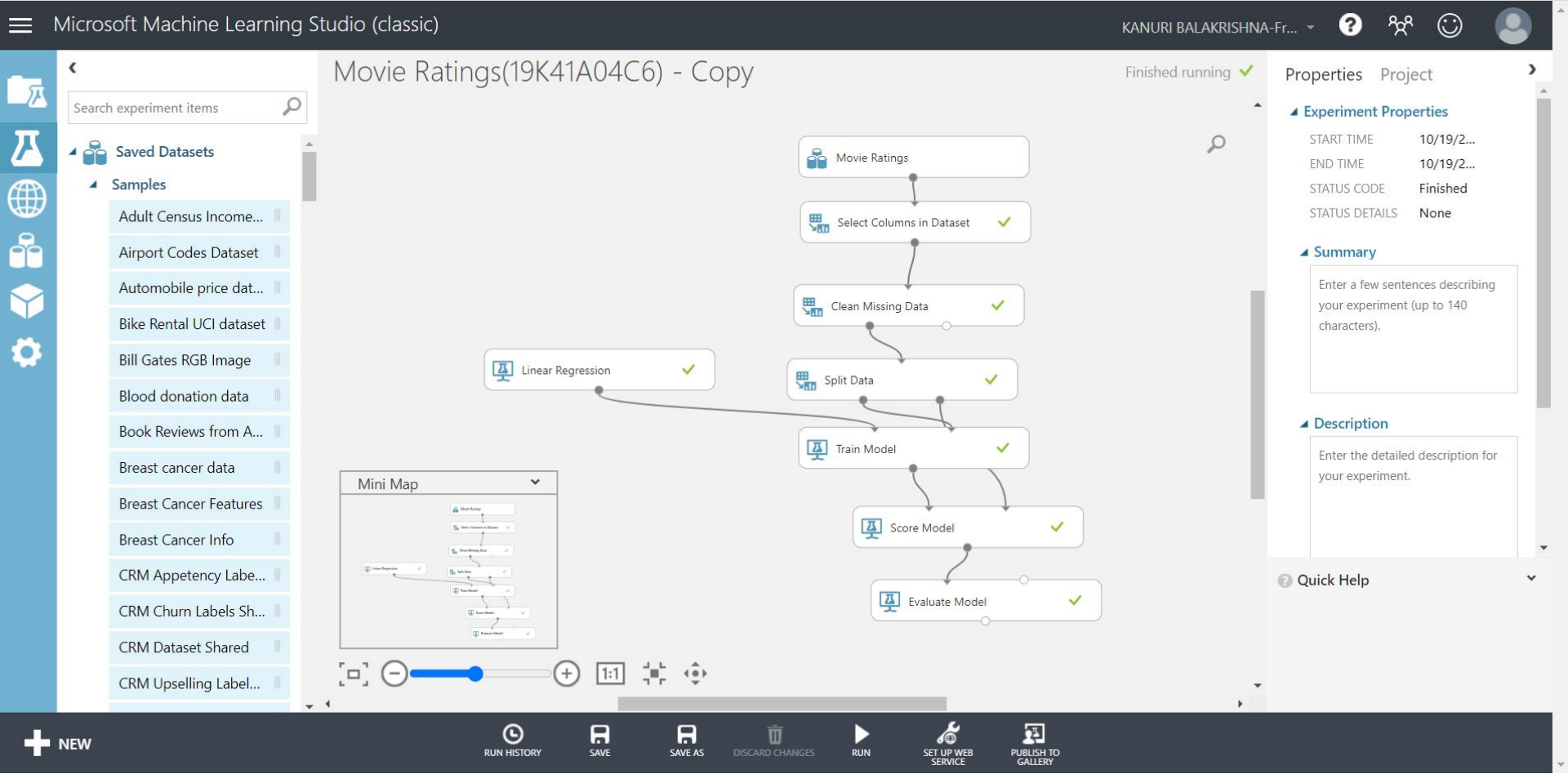
 <https://studio.azureml.net/Home/ViewWorkspaceCached/e8b0f8845b8f48479d3f0f39dddb0eec#Workspaces/Experiments/Experiment/e8b0f8845b8f48479d3f0f39dddb0eec.f-id.006c4274a73049678c3034a6bf2181dd/ViewExperiment>

[navigate to the link to see the Workflow and you can download the project as well.](#)

Machine Learning Project Workflow

1. Import Data
2. Explore Data (Missing values, outliers)
3. Preprocess data (Missing value imputation, outlier treatment, normalization)
4. Model Selection
5. Model Training
6. Model Testing
7. Model Deployment
8. Score Models

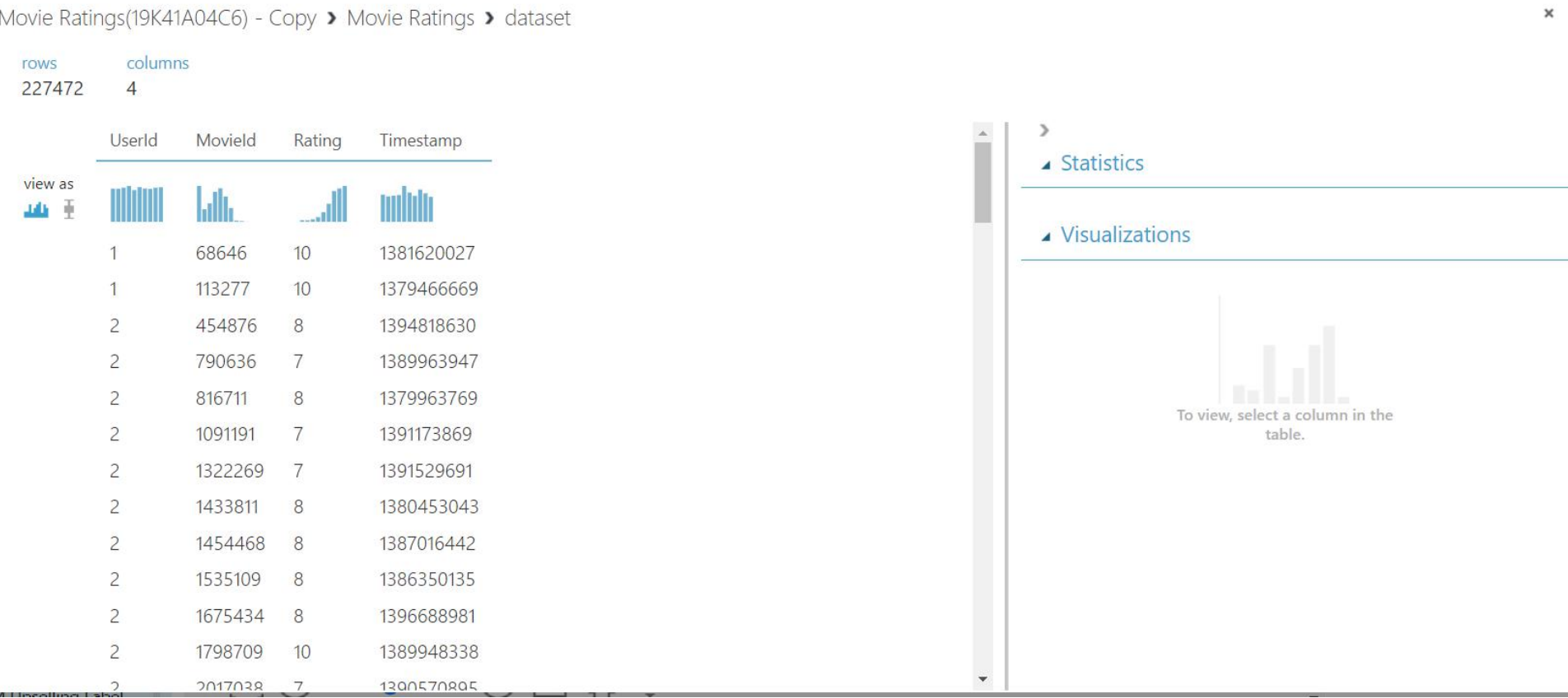
Workflow



Project Workflow

Import Data:

- importing the RAW dataset which is in CSV format.
- the dataset is pre-available in the Azure ML Classic Studio.



Movie Ratings Prediction RAW dataset (CSV format)

Explore Data

- this basically includes data visualization to search for any missing values in the Dataset.
- if any missing values are found, then they needs to be cleaned.
- selecting the required columns and clean the data using the Clean Missing Value module (Just Drag n' Drop)

Microsoft Machine Learning Studio (classic) interface showing a workflow for data exploration and cleaning. The workflow includes modules: Movie Ratings, Select Columns in Dataset, Clean Missing Data, Split Data, Train Model, Score Model, and Evaluate Model. The 'Clean Missing Data' module is highlighted, and the 'Properties' pane on the right shows the 'Selected columns' as 'UserId, MovieId, Rating, Times'.

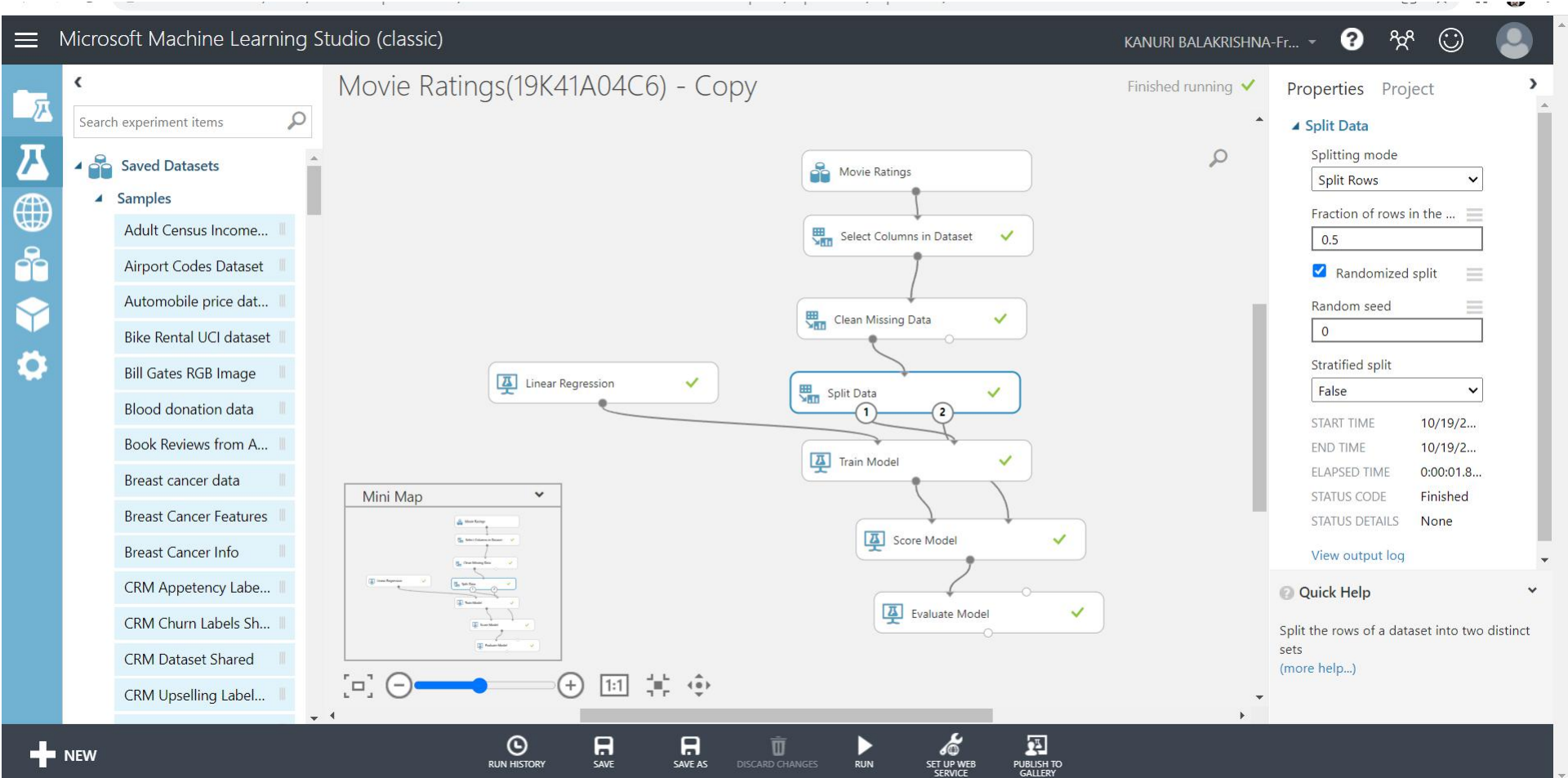
as the normalized loss has 41 missing values in the Dataset, those missing values are to be cleaned.

Microsoft Machine Learning Studio (classic) interface showing a workflow for data cleaning. The workflow includes modules: Movie Ratings, Select Columns in Dataset, Clean Missing Data, Split Data, Train Model, Score Model, and Evaluate Model. The 'Clean Missing Data' module is highlighted, and the 'Properties' pane on the right shows the 'Selected columns' as 'All columns'.

Data Cleaning

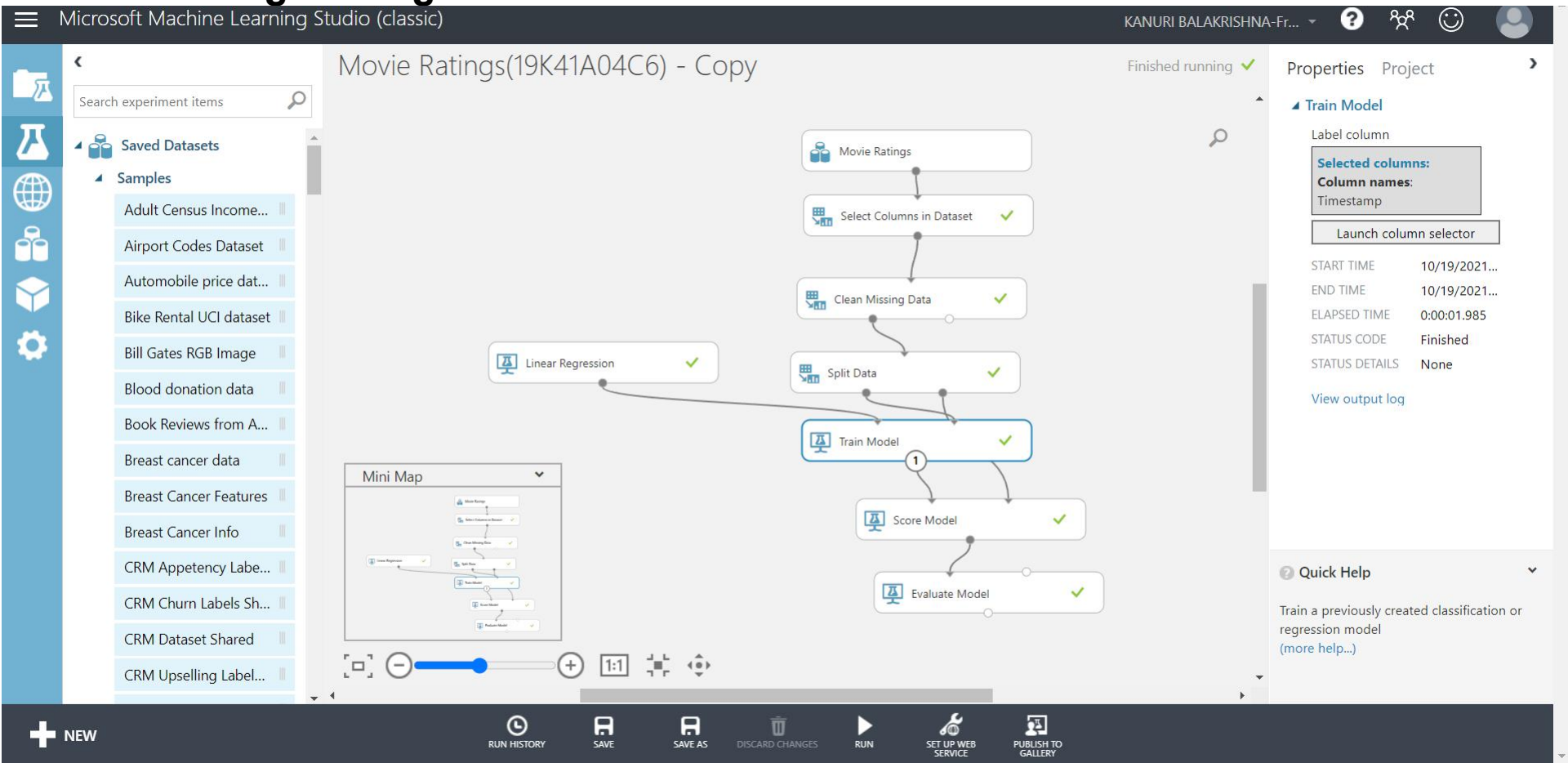
Split Data

- Use the Split Data module to randomly divide the input data so that the training dataset contains 70% of the original data and the testing dataset contains 30% of the original data.



Data Splitting

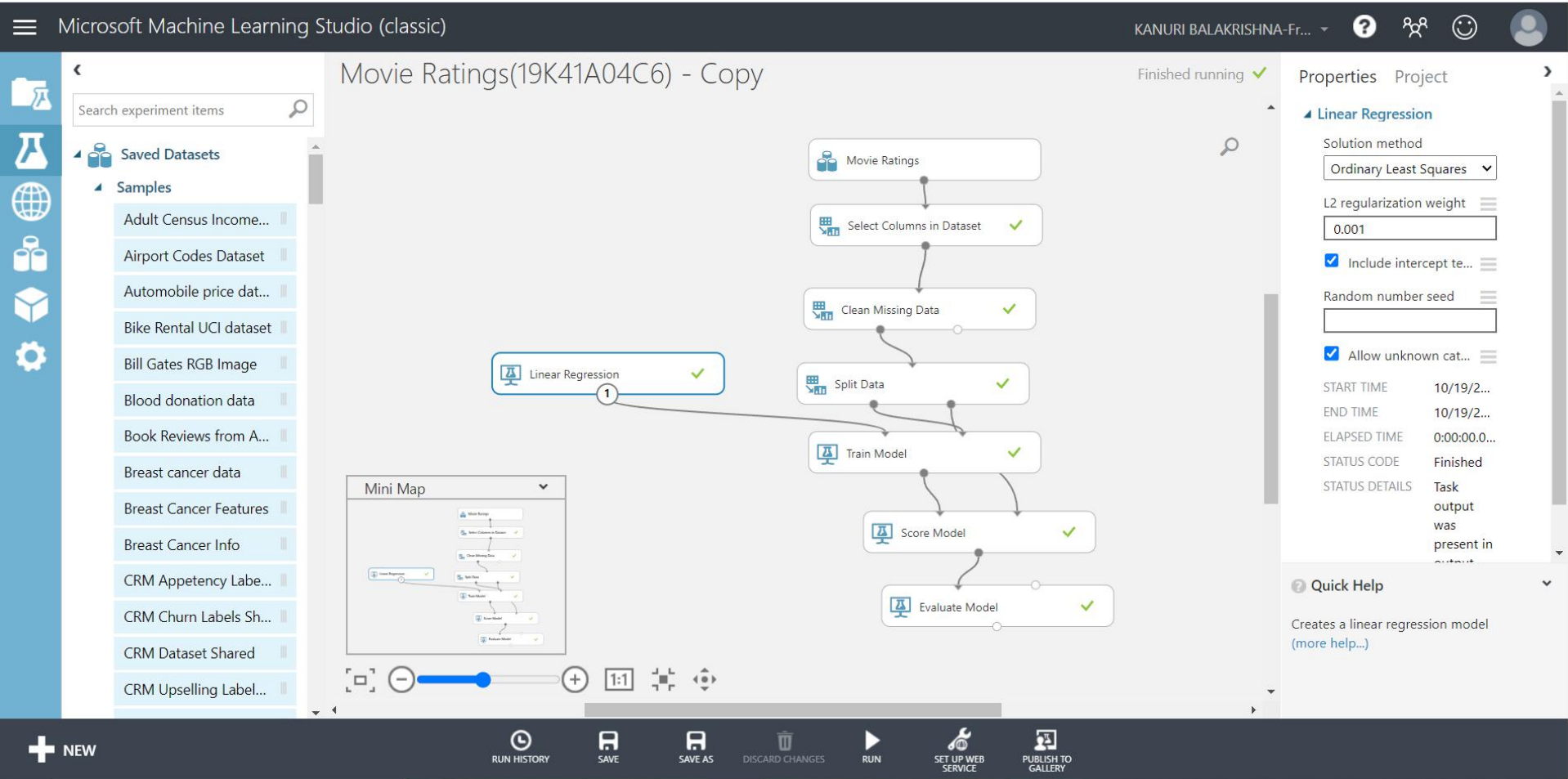
Model Training and Algorithm



Model Training

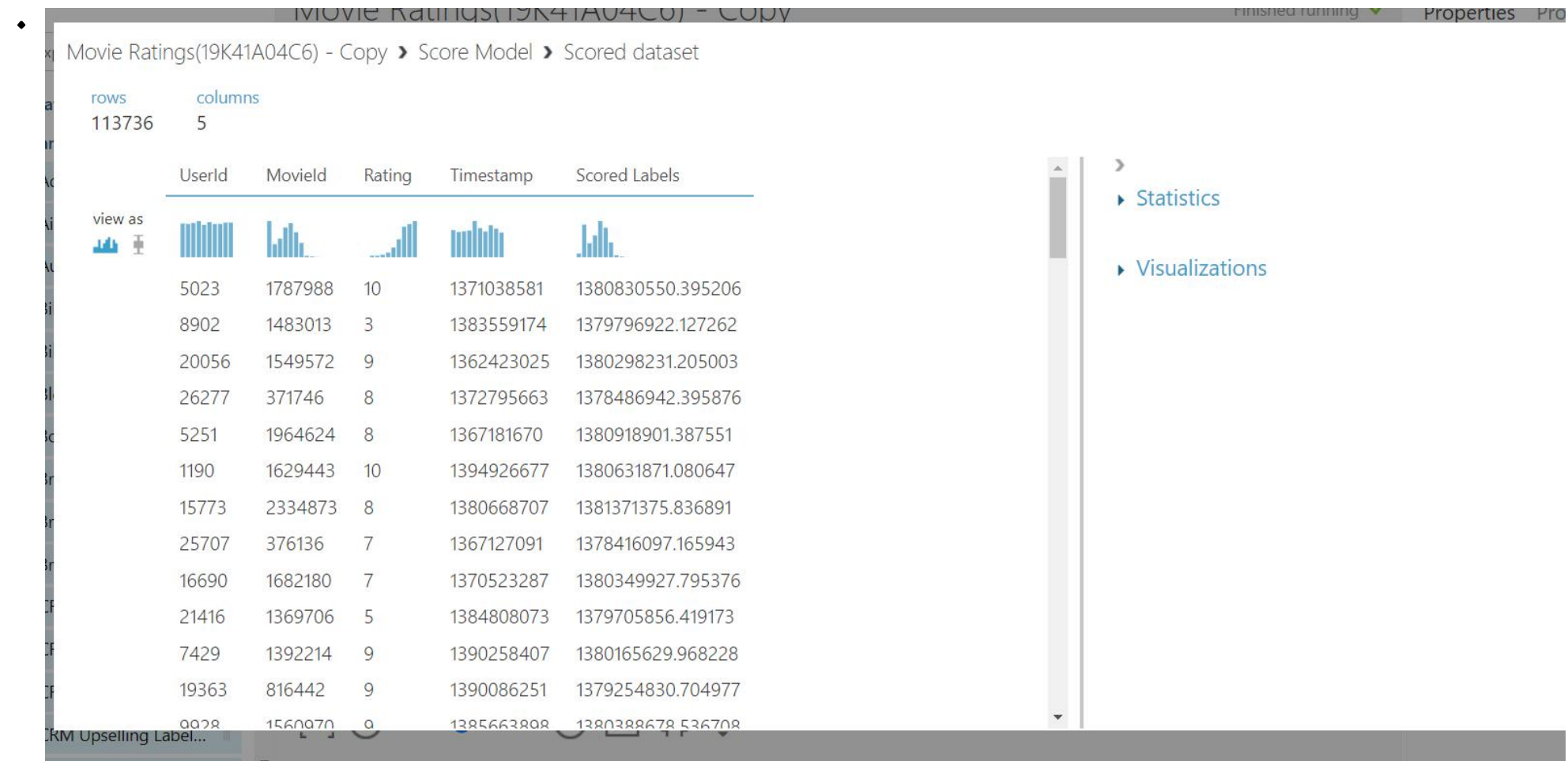
Using Linear Regression to train the model

- Since the goal of this sample is to predict movie ratings prediction, and because the label column (ratings) is continuous data, We use linear regression for this pipeline.



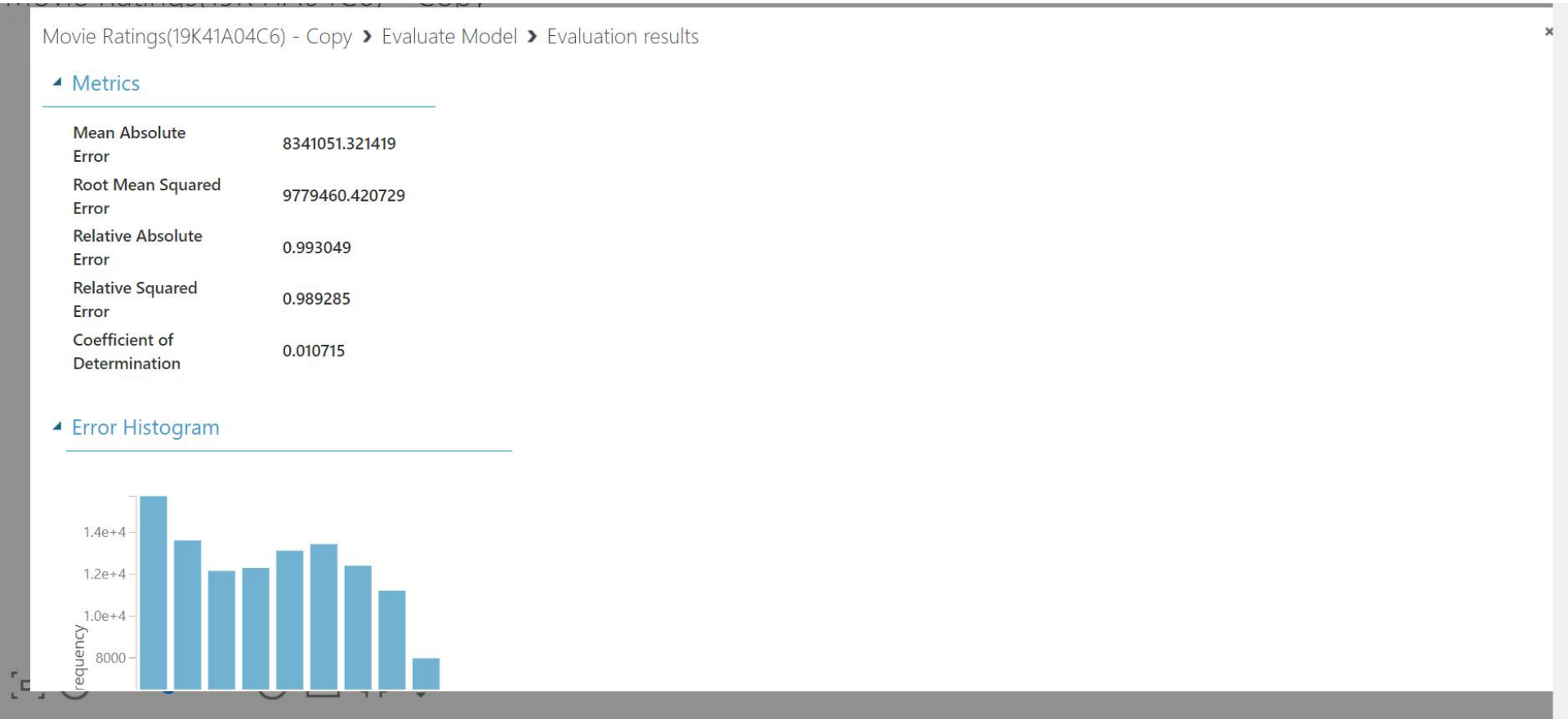
Linear Regression

Score Model and Evaluate Model



After the model is trained, we can use the Score Model module to generate predicted results.

Score Labels



- After the score model, we can use the Evaluate Model module to generate predicted results.

Evaluate Labels