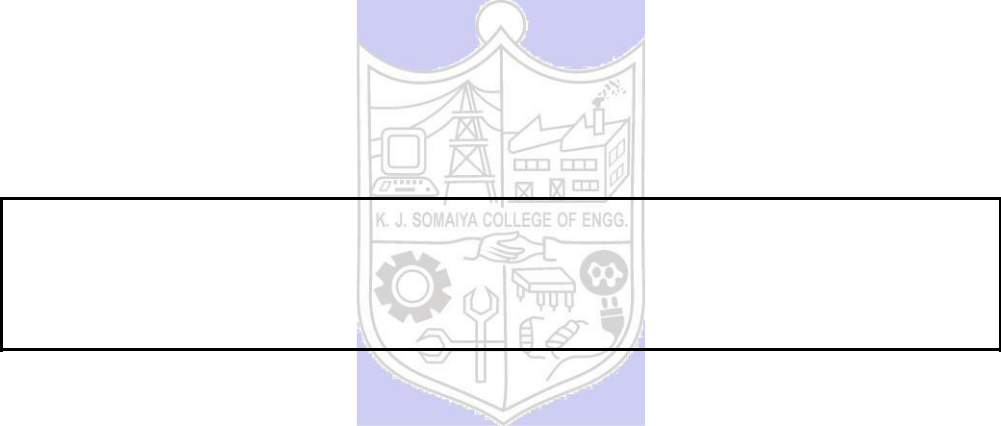
KJSCE/IT/B.Tech/SEM-VI/CC/2020-2021



**Experiment No. 07**

**Title: Big Data analytics using Microsoft Azure (PaaS)**

(Autonomous College Affiliated to University of Mumbai)

KJSCE/IT/B.Tech/SEM-VI/CC/2020-2021

**Batch:A4** **Roll No.:1914078** **Experiment No.:07**

**Aim: Case study and implementation of Big Data analytics using Microsoft Azure (PaaS)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Resources needed: Microsoft Azure Machine learning Studio Account**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

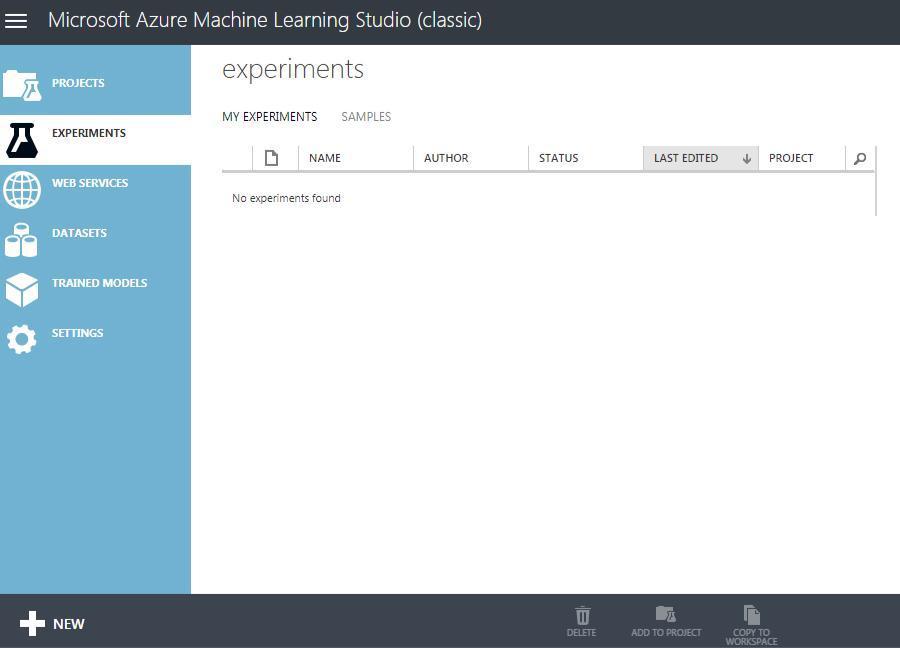
**Describe the following points with respect to the business under consideration:**

1. Problem faced by the business
2. Approach/ Methodology followed by the business
3. Skill Sets, infrastructure and other impact on the business during implementation
4. Similar approaches followed by other businesses
5. **Follow Machine learning studio for developing an application**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Procedure:**

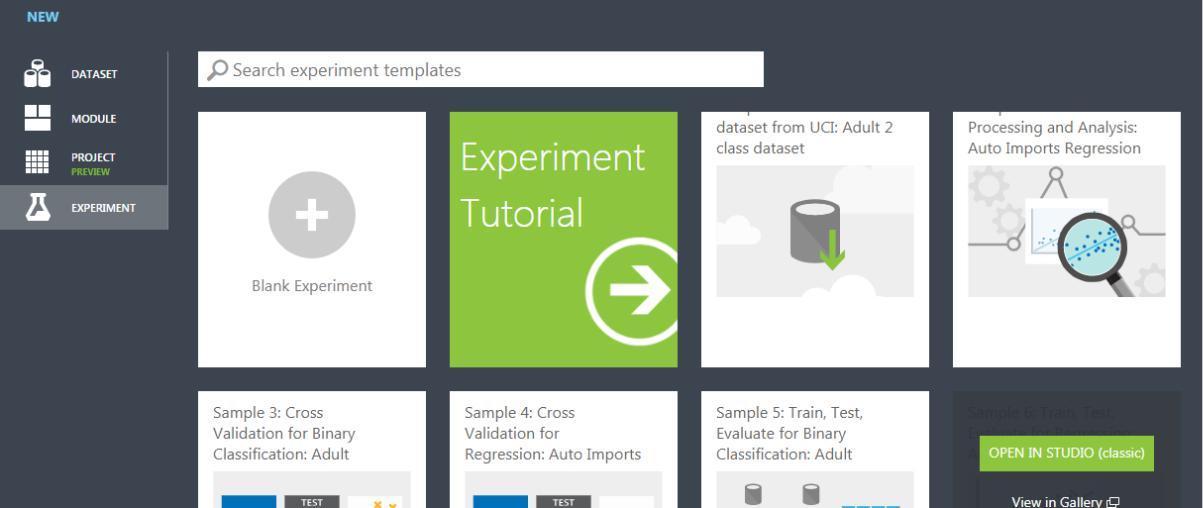
**Step 1: Open Azure Machine learning Studio Account:** [https://studio.azureml.net](https://studio.azureml.net/)



(Autonomous College Affiliated to University of Mumbai)

KJSCE/IT/B.Tech/SEM-VI/CC/2020-2021

**Step 2: Click on New and explore Experiment Tutorial and samples**



**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Results: (Program / Steps with screenshots):**

**Description of the dataset**

The dataset contains information about residential buildings situated in **Ames, Iowa.**

The gist of the problem is predicting the sale price of the house which is based on many factors. Appropriately, the dataset also has numerous columns which represent these factors. The dataset also contains null or missing values which pose a problem to predicting prices using a machine learning model.

1. **Problem faced by the business:**

There are innumerable factors that affect housing prices. It’s very volatile and changes frequently. Some of these factors are:

* 1. **Economic factors:** 
     1. **Growth in the Economy:** With a growing economy and rise in wages, people are likely to spend more on houses.
     2. **Unemployment:** Fear of unemployment can stop people from investing in real estate, possibly causing a drop in prices.
     3. **Interest rates**
     4. **Customer Trust**
     5. **Mortgage Availability**
  2. **Demographic Factors:**
     1. Life expectancy increased for the elderly
     2. Divorce rates rise
     3. As per now, children leave their homes in early years itself
     4. Increase in Marriages
     5. They are dreaming to be more Independent.

1. **Approach/ Methodology followed by the business:**

**There are a few ways in which a buyer can predict house prices in today’s market:**

* 1. **A comparative market analysis:**  A buyer can ask a local real estate agent for a CMA or a comparative market analysis. A CMA provides an agent’s evaluation of a property to estimate/predict its value.
  2. **FHFA House Price Index Calculator**
  3. **Hiring a professional appraiser**
  4. **Evaluating comparable properties**

1. **Skill Sets, infrastructure and other impact on the business during implementation:**

* A software engineer that has experience in machine learning will prove to be very helpful. Any client can use this model to predict house prices on their own without having to avail professional help, saving money.

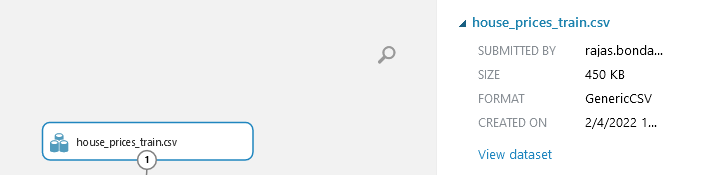
1. **Similar approaches followed by other businesses**

A regressor model will suit best for this task as we’ve to predict the value of one numerical target variable using a set of predetermined independent variables. Similar models can be used for

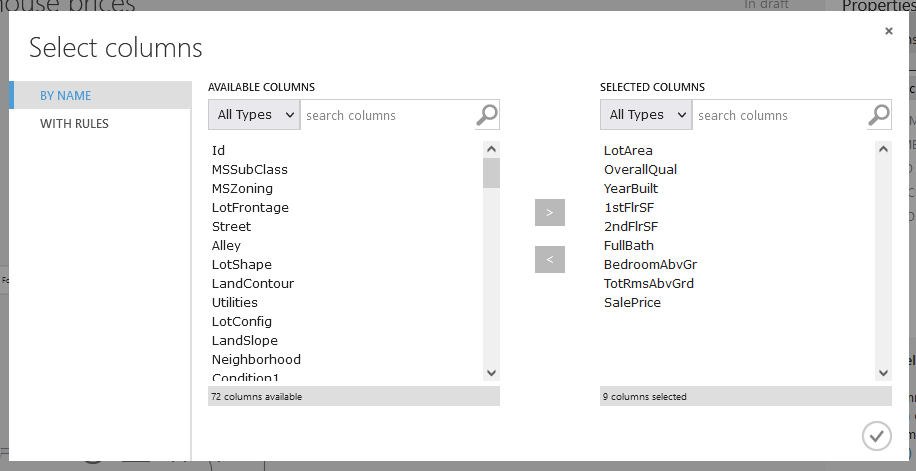
1. **Follow Machine learning studio for developing an application**

[Link to the dataset](https://www.kaggle.com/alexisbcook/missing-values/data?select=train.csv)

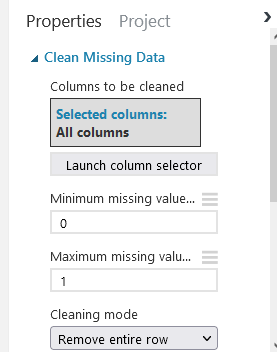
1. Uploading and choosing dataset



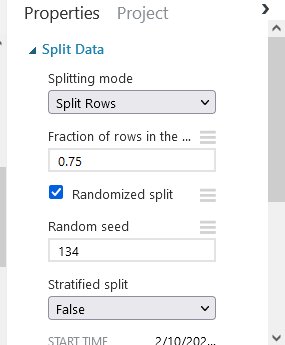
1. Selecting columns in dataset
   1. The dataset has over 72 columns. Thus only,
      1. LotArea: area of the house
      2. OverallQual: Quality of the house 1-10
      3. YearBuilt
      4. 1stFlrSF: Area of 1st floor in square feet
      5. 2ndFlrSF: Area of 2nd floor in square feet
      6. FullBath: Full bathrooms
      7. BedroomsAbvGrd: No. of bedrooms above ground
      8. TotRmsAbvGrd: No. of rooms above ground
      9. SalePrice: Target variable

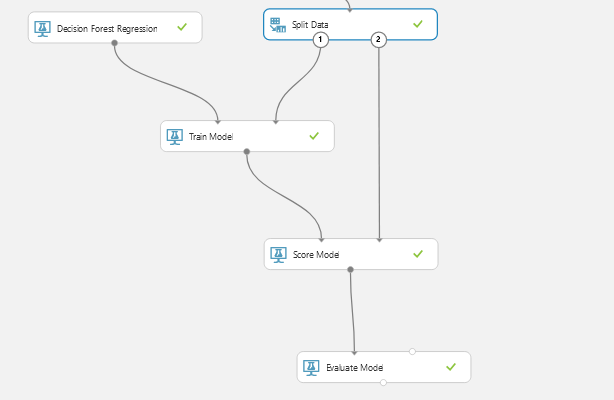
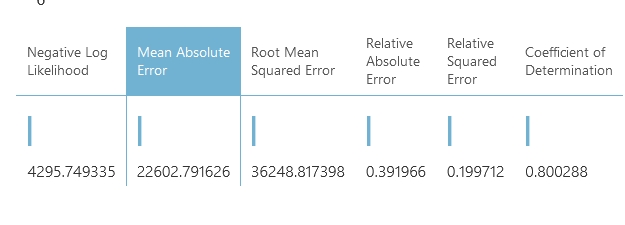
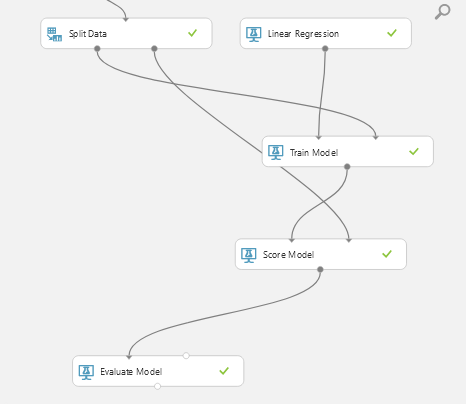
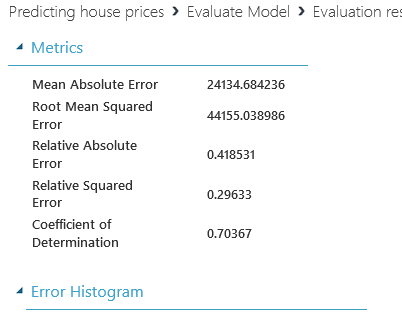
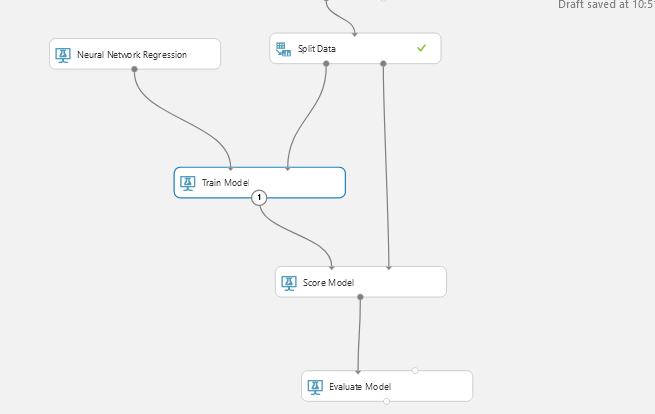
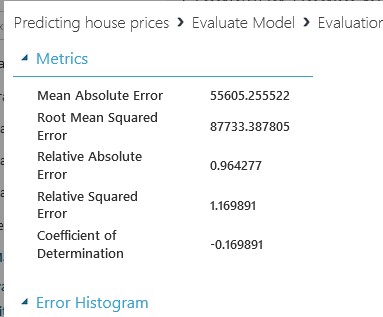


1. Clean Missing Data: Removing rows with missing data



1. Split Data: Splitting data 75%-25%



1. Training regression models
   1. Decision Forest Regressor
      1. 
      2. Score:
         1. 
   2. Linear Regression Model
      1. 
      2. Score
         1. \
   3. Neural Network Regression
      1. 
      2. Score:
         1. 

Out of the above chosen models; Decision Forest Regression, Linear Regression and Neural network regression model; Decision Forest Regression had the least Mean absolute error.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Questions:**

Discuss the tangible and intangible benefits the business has observed after the implementation.



This model will help in predicting sale prices of a house using the chosen variables. Of course, variables can be added or removed to make the model more precise for future use. It can be used realtors to see how each of the variables affects the sale price. Even builders can use this model to efficiently make a house that will have a high Sale Price.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Outcomes:**

Study the Evolution of Cloud Computing and its models

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Conclusion: (Conclusion to be based on the objectives and outcomes achieved)**

During the course of this experiment, we performed big data analytics using Microsoft Azure Machine Learning Studio(PaaS). It was used to train several models for predicting house prices.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

(Autonomous College Affiliated to University of Mumbai)

KJSCE/IT/B.Tech/SEM-VI/CC/2020-2021

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

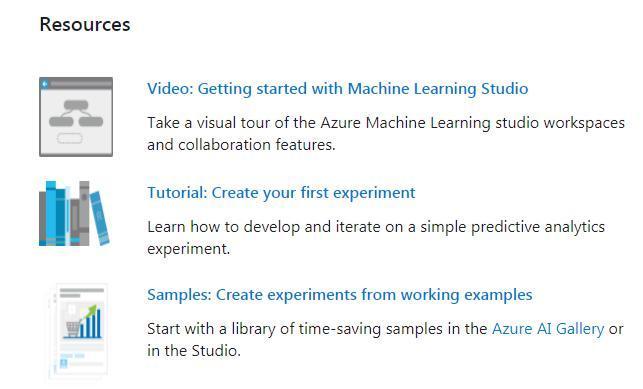
**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Grade: AA / AB / BB / BC / CC / CD /DD**

**Signature of faculty in-charge with date**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**References:**



(Autonomous College Affiliated to University of Mumbai)