DATA MINING PROJECT

Predicting the critical temperature of the superconductors.

By Group 4

Apr 3,2025

BACKGROUND

Superconductivity is the ability of certain materials to conduct electricity with zero electrical resistance at temperatures above the boiling point of liquid nitrogen, which was unexpectedly discovered in copper oxide (cuprate) materials in 1987. The critical temperature for superconductors is the temperature at which the electrical resistivity of metal drops to zero.

PROJECT GOAL

Develop a machine learning model to predict the critical temperature (Tc) of superconductors by Implementing and comparing two models – Linear Regression and a Neural Network – based on material properties.

DATA SET

- The dataset contains superconductors' material properties and corresponding critical temperature (Tc) values.

Data Set Size:

- Total Samples: 21,263 superconductors
- Dataset Characteristics - Multivariate
- Number of Attributes: 81 predictors + 1 target (Tc)

Data Source:

https://archive.ics.uci.edu/ml/datasets/Superconductivty+Data

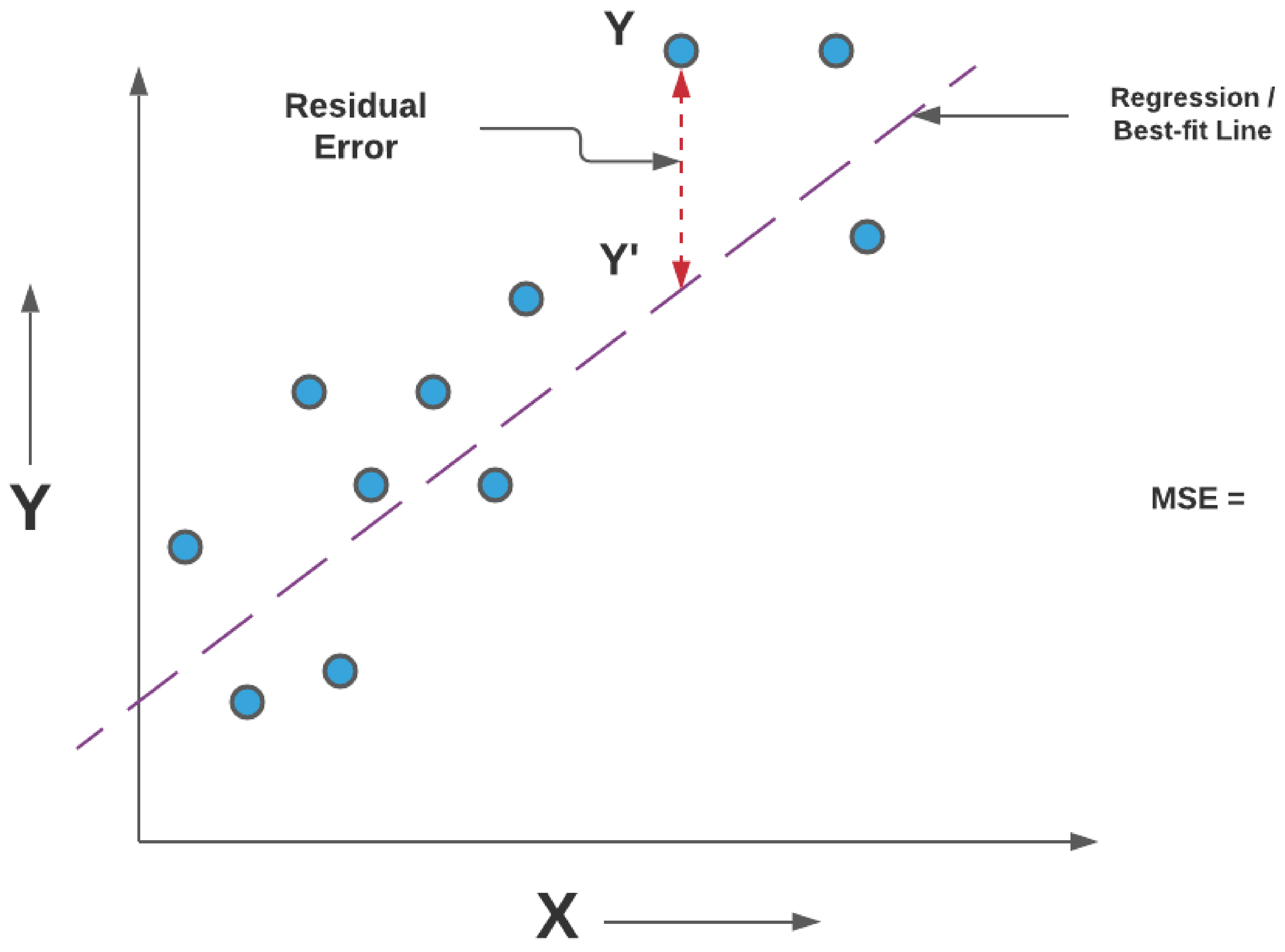
DATA PREPROCESSING

- Split the dataset into features and targets.
- Scales the features using StandardScaler to ensure that they have zero mean and unit variance.
- Splits the data into training and testing sets, with 80% for training and 20% for testing.

EVALUATION METRICES

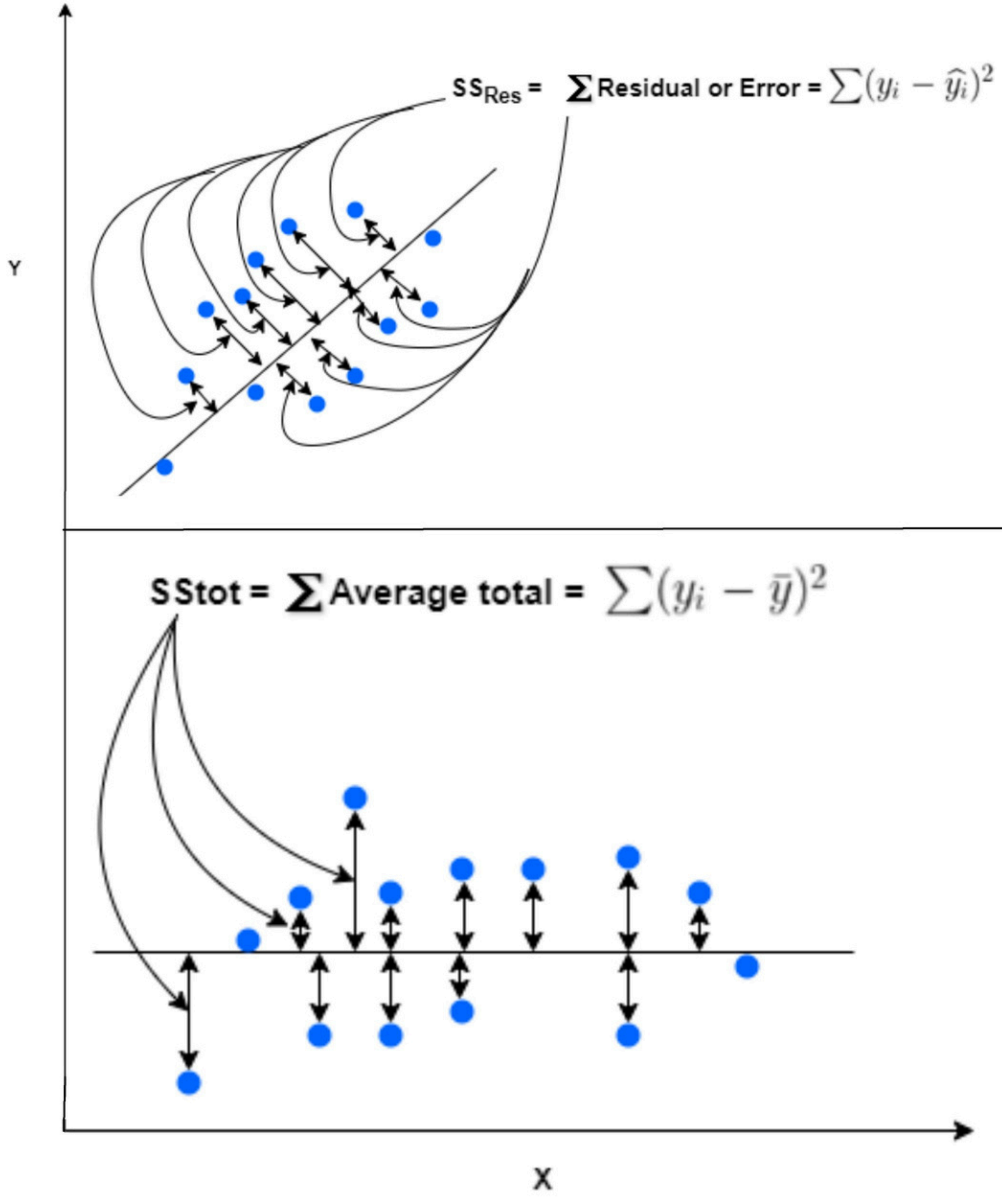
- Mean Squared Error(MSE)
- R-Squared Error(R^2 Error)

MEAN SQUARED ERROR



$$MSE = \frac{\text{Sum } (Y - Y')^2}{N}$$

R^2 ERROR



$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum(y_i - \widehat{y_i})^2}{\sum(y_i - \bar{y})^2}$$

LINEAR REGRESSION

- We trained a **Linear Regression model** using the training data (X_train, y_train). After that we uses that trained model to make predictions on both the training data and the test data (X_test). The model's performance is evaluated using two metrics: **Mean Squared Error (MSE)** and **R-squared (R^2)**, calculated for both training and test sets.

- **Linear Regression Performance**:
  Train MSE: *310.2180*, Train R^2: *0.7368*
  Test MSE: *302.0075*, Test R^2: *0.7376*

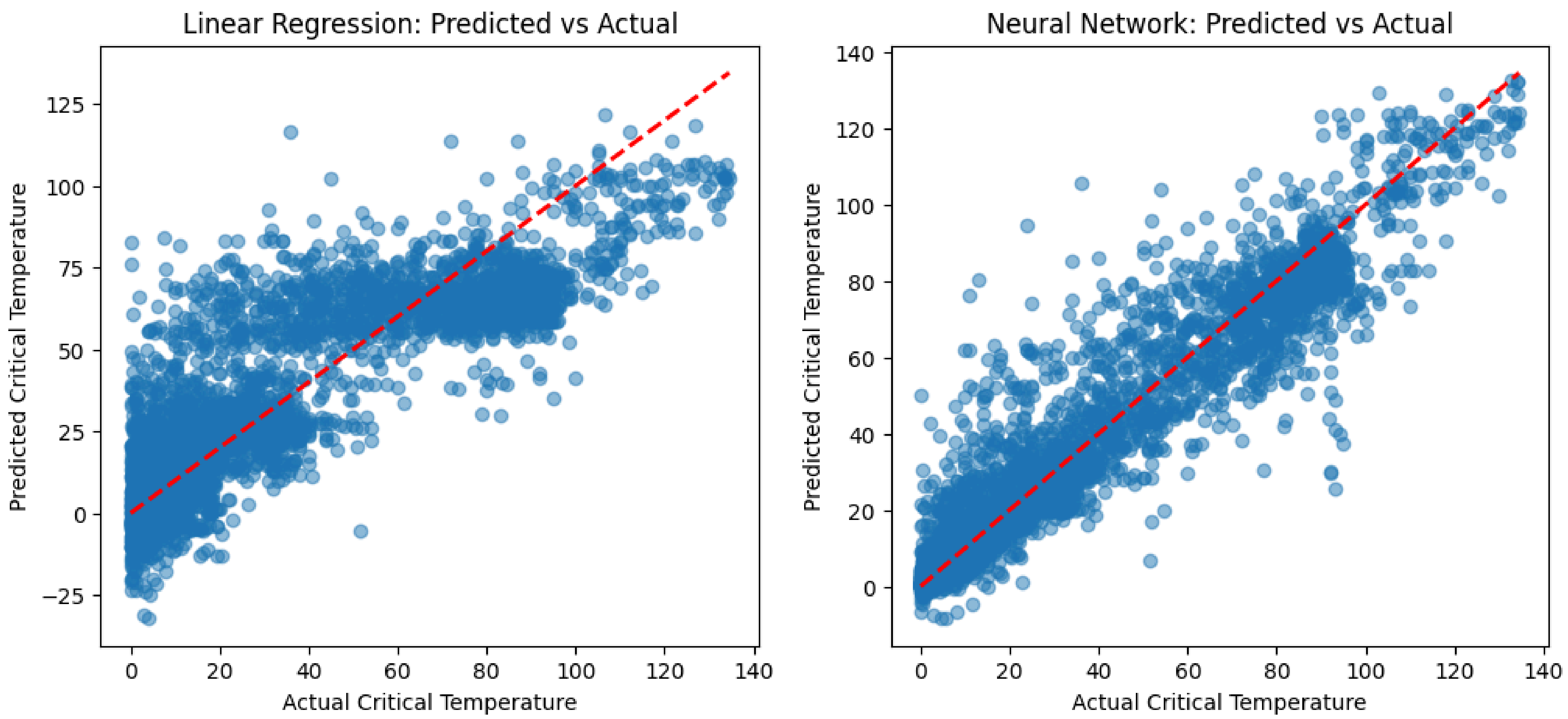
ANN

- We used a neural network regression model (**MLPRegressor**) to predict a target variable. The model has two hidden layers (100 and 50 neurons) and uses the **sigmoid** activation function.After training, the model is used to predict values for both the training and test datasets. The performance of the model is then evaluated using **Mean Squared Error (MSE)** and **R-squared (R^2)** metrics, which are printed for both training and test sets.

- **Neural Network Performance**:
  Train MSE: *88.5137*, Train R^2: *0.9249*
  Test MSE: *107.4809*, Test R^2: *0.9066*

MODEL COMPARISON

- <u>**Model Comparison**</u>:
  Linear Regression - Test MSE: *302.0075*, Test R^2: *0.7376*
  Neural Network - Test MSE: *107.4809*, Test R^2: *0.9066*



THANK YOU