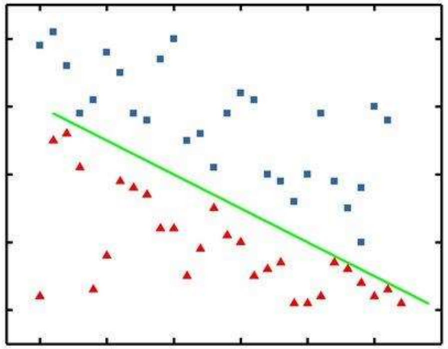
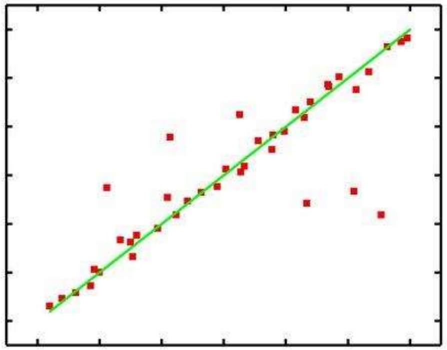
MACHINE LEARNING ALGORITHMS

Regression

- Regression Models are a class of supervised learning algorithms used for predicting a continuous outcome (target variable) based on one or more independent variables.
- Linear regression assumes a linear relationship between the input features and outcome.
- logistic regression is used for binary classification tasks. It models the probability that an instance belongs to a particular class using the logistic function.
- Non-linear regression models capture non-linear relationships between input features and the outcome.



(a) Logistic Regression

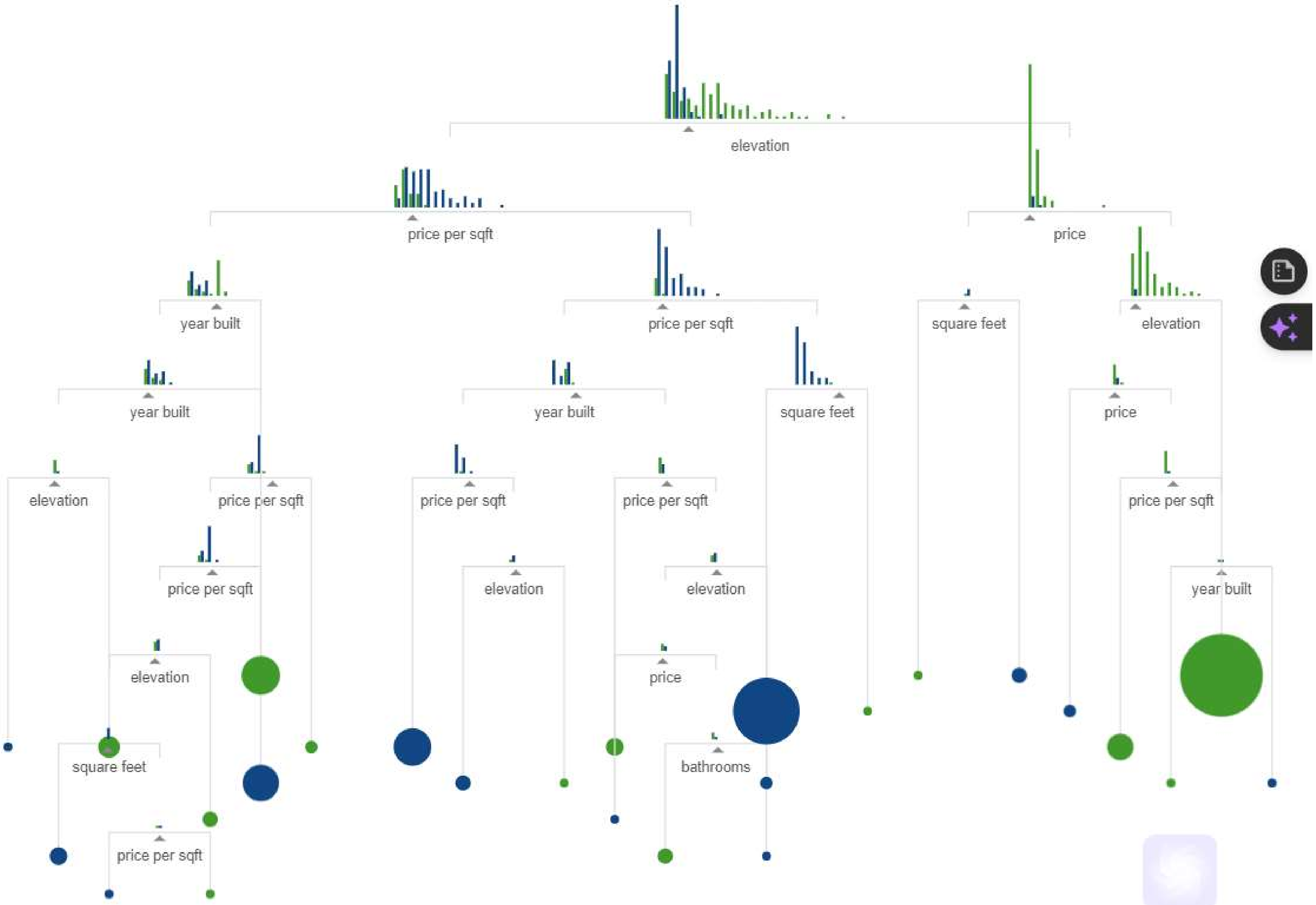


(b) Linear Regression

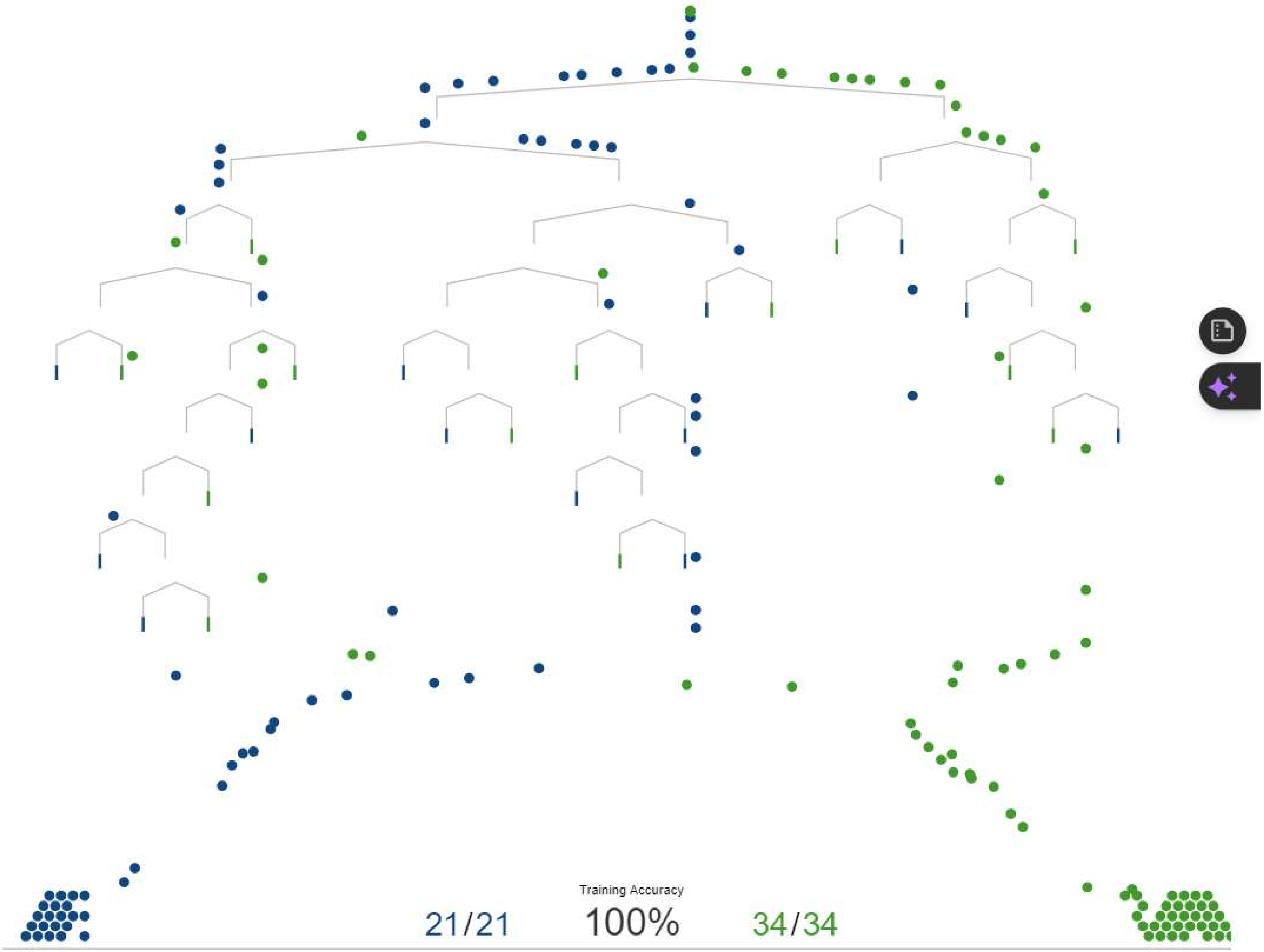
Decision Trees

- **Decision Trees** used for both classification and regression tasks.
- algorithm builds a tree-like structure by recursively splitting the dataset based on the features that provide the best separation of classes or the best prediction of the outcome/target variable.
- Model:
  - Tree-like model with nodes (decision points), branches (possible outcomes), and leaves (final predictions).
  - Root node: Represents the entire dataset.
  - Internal nodes: Contain decision rules based on feature values.
  - Leaf nodes: Represent class labels (for classification) or continuous values (for regression).
- Issues:
  - Overfitting, and Sensitive to feature scaling: Normalization often required

Decision Trees



Decision Trees



Random Forests

- **Random Forests** used for both classification and regression tasks.
- builds multiple decision trees during training and merges their predictions to improve accuracy and reduce overfitting.
- Model:
  - Collection of decision trees, each tree is trained independently on a subset of data.
  - At each node of a decision tree, a random subset of features is considered for splitting.
  - Each tree is trained on a random sample drawn from the original data (bootstrap sample)
  - Final decision by majority vote of the trees, robust and avoid overfitting.

k-Nearest Neighbors (k-NN)

- **k-NN model** - used for both classification and regression tasks.
- Model:
  - finding the k closest training examples to a new input, and making a prediction based on the labels or values of those k neighbors.
  - Good for dynamic datasets, as there is no model training phase. Memorizes the entire dataset, not suitable for large datasets.
  - Depends on distance metrics (closeness) and k (closest neighbors).