Supervised Learning - Regression

Mhd. Arsya Fikri | 191402066

Apa yang akan kita bahas?

What is Supervised Learning?

Data Preparation

Modelling (Regression)

Model Evaluation

.

.

01

Supervised Learning - Regression

What is it?

What is Supervised Learning?

There are 3 major groups of machine learning types:

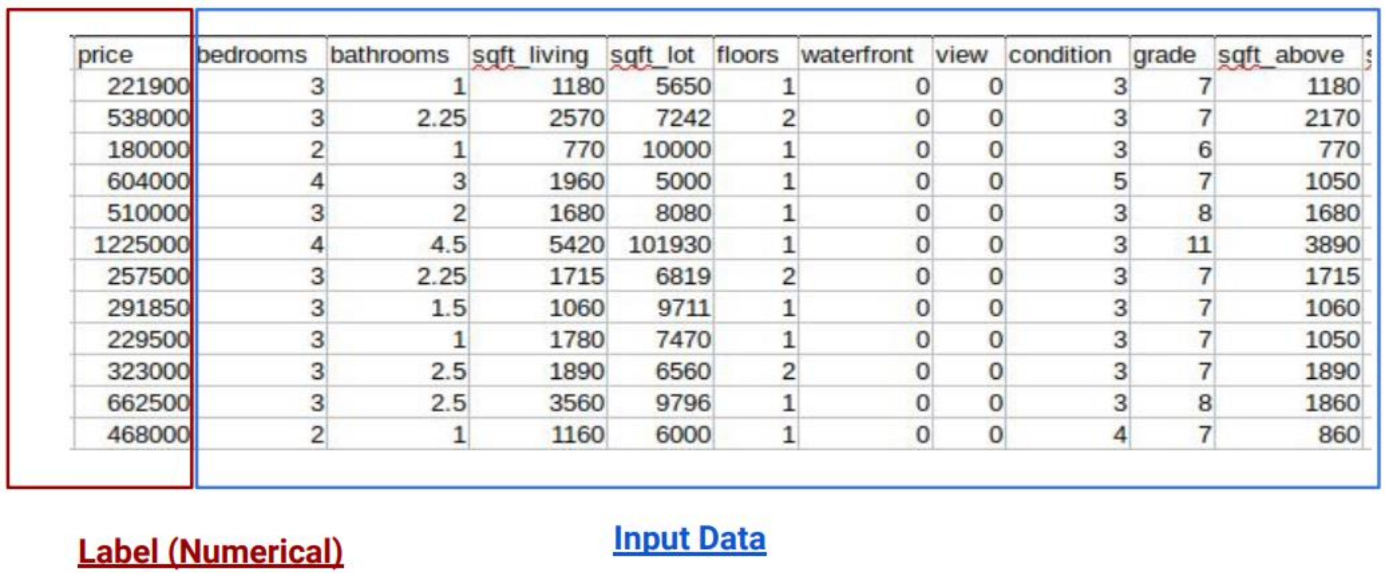
- **Supervised Learning**: Input data is called training data and has a known label or result (such as spam/not-spam or a stock price at a time).
- **Unsupervised Learning**: Input data is not labeled and does not have a known result.
- **Reinforcement Learning**: A special type of Machine Learning where the model learns from each action taken. The model is rewarded for any correct decision made and penalized for any wrong decision.

What is Supervised Learning?

Supervised Learning

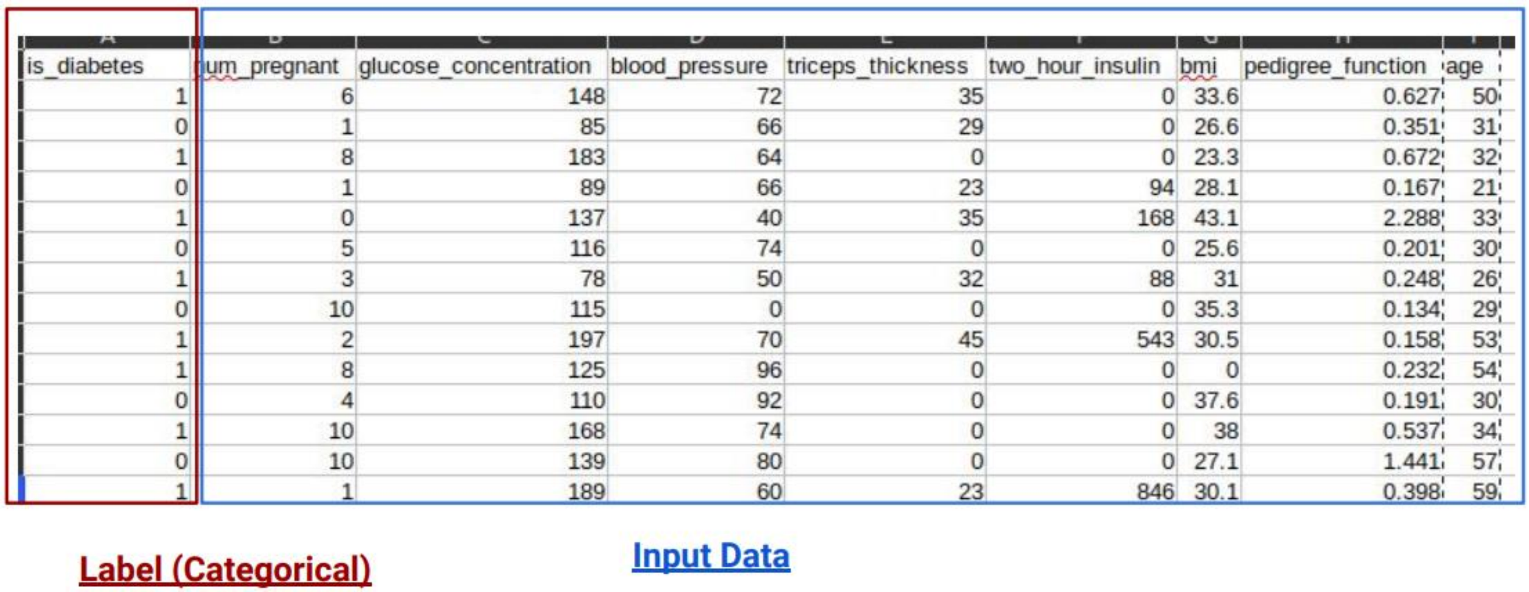
- Making predictions with a rule/often called as a model
- Has input data and labels

What is Supervised Learning?



price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	condition	grade	sqft_above
221900	3	1	1180	5650	1	0	0	3	7	1180
538000	3	2.25	2570	7242	2	0	0	3	7	2170
180000	2	1	770	10000	1	0	0	3	6	770
604000	4	3	1960	5000	1	0	0	5	7	1050
510000	3	2	1680	8080	1	0	0	3	8	1680
1225000	4	4.5	5420	101930	1	0	0	3	11	3890
257500	3	2.25	1715	6819	2	0	0	3	7	1715
291850	3	1.5	1060	9711	1	0	0	3	7	1060
229500	3	1	1780	7470	1	0	0	3	7	1050
323000	3	2.5	1890	6560	2	0	0	3	7	1890
662500	3	2.5	3560	9796	1	0	0	3	8	1860
468000	2	1	1160	6000	1	0	0	4	7	860

What is Supervised Learning?

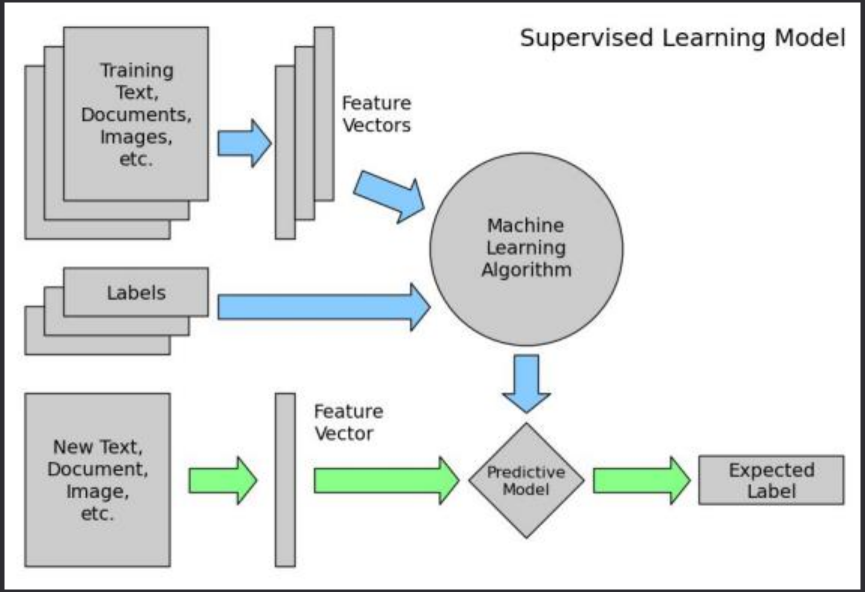


is_diabetes	num_pregnant	glucose_concentration	blood_pressure	triceps_thickness	two_hour_insulin	bmi	pedigree_function	age
1	6	148	72	35	0	33.6	0.627	50
0	1	85	66	29	0	26.6	0.351	31
1	8	183	64	0	0	23.3	0.672	32
0	1	89	66	23	94	28.1	0.167	21
1	0	137	40	35	168	43.1	2.288	33
0	5	116	74	0	0	25.6	0.201	30
1	3	78	50	32	88	31	0.248	26
0	10	115	0	0	0	35.3	0.134	29
1	2	197	70	45	543	30.5	0.158	53
1	8	125	96	0	0	0	0.232	54
0	4	110	92	0	0	37.6	0.191	30
1	10	168	74	0	0	38	0.537	34
0	10	139	80	0	0	27.1	1.441	57
1	1	189	60	23	846	30.1	0.398	59

Label (Categorical)

Input Data

How Supervised Learning Works?


Supervised Learning Model
Training Text, Documents, Images, etc.
Feature Vectors
Machine Learning Algorithm
Labels
New Text, Document, Image, etc.
Feature Vector
Predictive Model
Expected Label

02

Supervised Learning - Regression

Data Preparation

Data Preparation

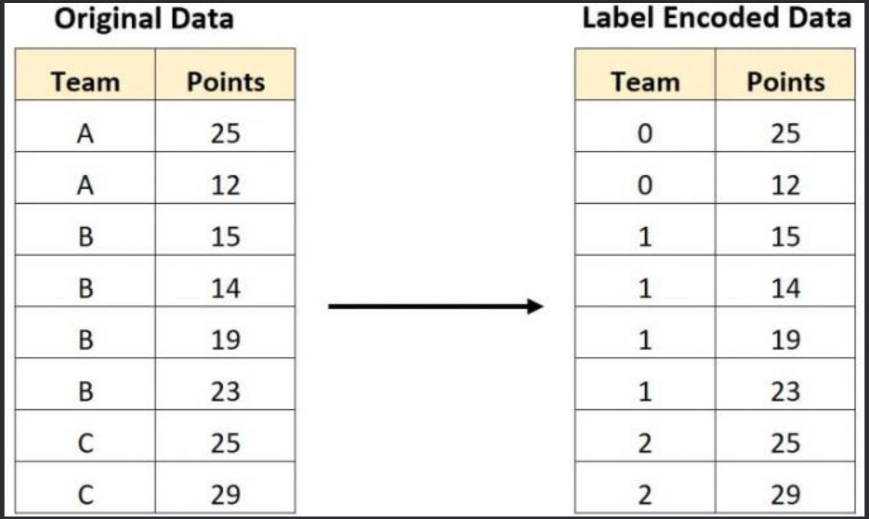
- **Encoding**. Representing every single piece of data in a way that a computer can understand (the name literally means "convert to computer code").

- **Train Test Split**. The train-test split is a data preprocessing technique for evaluating the performance of a machine learning algorithm. The procedure involves taking a dataset and dividing it into two subsets.
  - The first subset is used to fit the model and is referred to as the training dataset.

  - The second subset is not used to train the model; instead, the input element of the dataset is provided to the model, then predictions are made and compared to the expected values. This second dataset is referred to as the test dataset.

Data Preparation: Encoding

- **Label Encoding**



Original Data

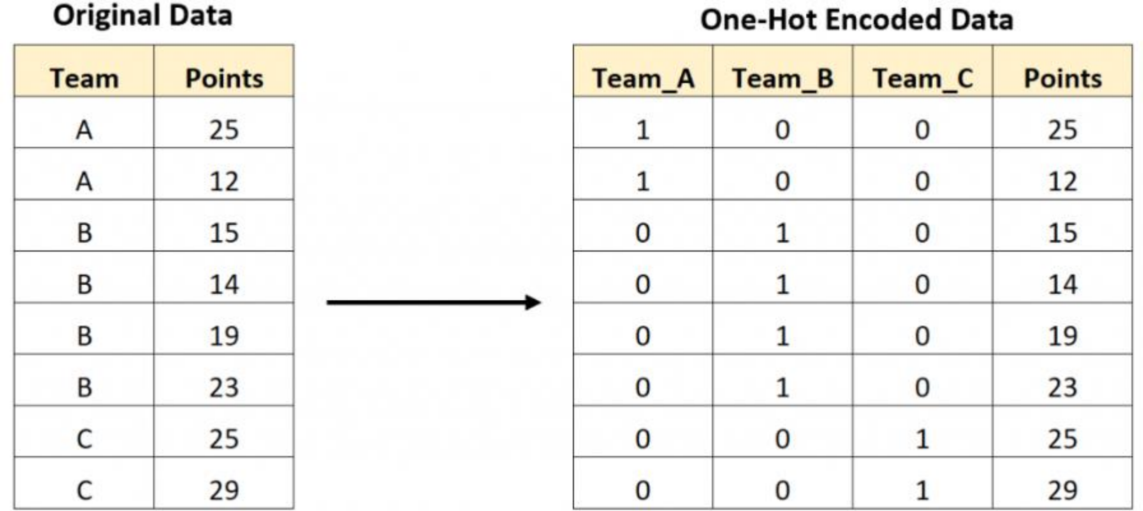
Team	Points
A	25
A	12
B	15
B	14
B	19
B	23
C	25
C	29

Label Encoded Data

Team	Points
0	25
0	12
1	15
1	14
1	19
1	23
2	25
2	29

Data Preparation: Encoding

- **One-Hot Encoding**



Original Data

Team	Points
A	25
A	12
B	15
B	14
B	19
B	23
C	25
C	29

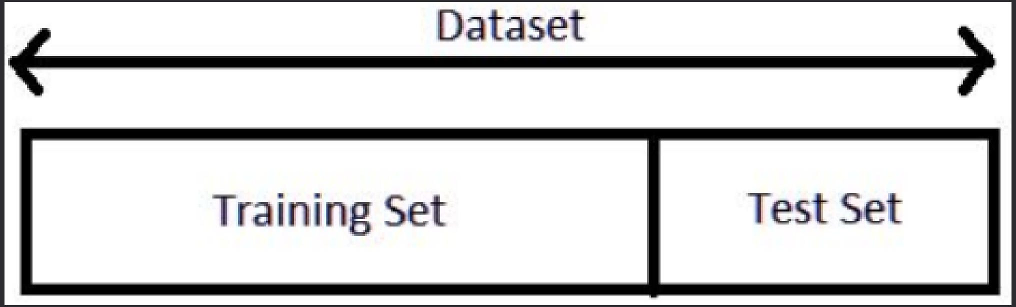
→

One-Hot Encoded Data

Team_A	Team_B	Team_C	Points
1	0	0	25
1	0	0	12
0	1	0	15
0	1	0	14
0	1	0	19
0	1	0	23
0	0	1	25
0	0	1	29

Data Preparation: Train Test Split

- **Train Dataset**: Used to fit the machine learning model.
- **Test Dataset**: Used to evaluate the fit machine learning model.



03

Supervised Learning - Regression

Modelling (Regression)

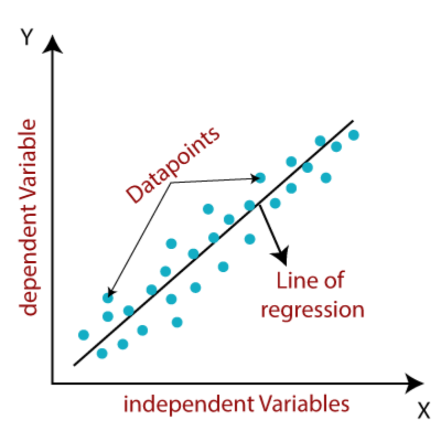
Modelling

- The process of modeling means training a machine learning algorithm to predict the labels from the features, tuning it for the business need, and validating it on holdout data.
- There are so many algorithm we can use as a model.

Modelling

- **Linear Regression**

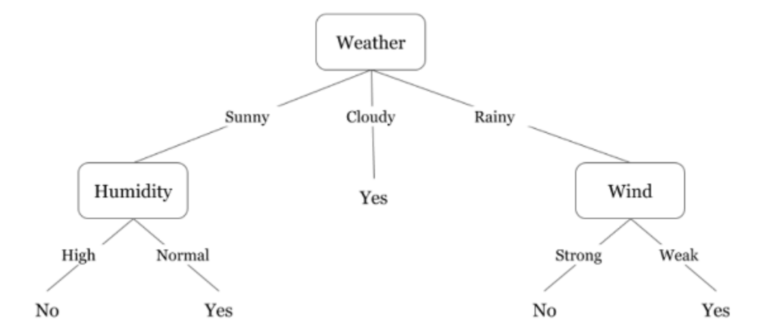
Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (X) variables, hence called as linear regression.



Modelling

- **Decision Tree**

A decision tree is a tree-like structure that represents a series of decisions and their possible consequences. It is used in machine learning for classification and regression tasks. An example of a decision tree is a flowchart that helps a person decide what to wear based on the weather conditions.

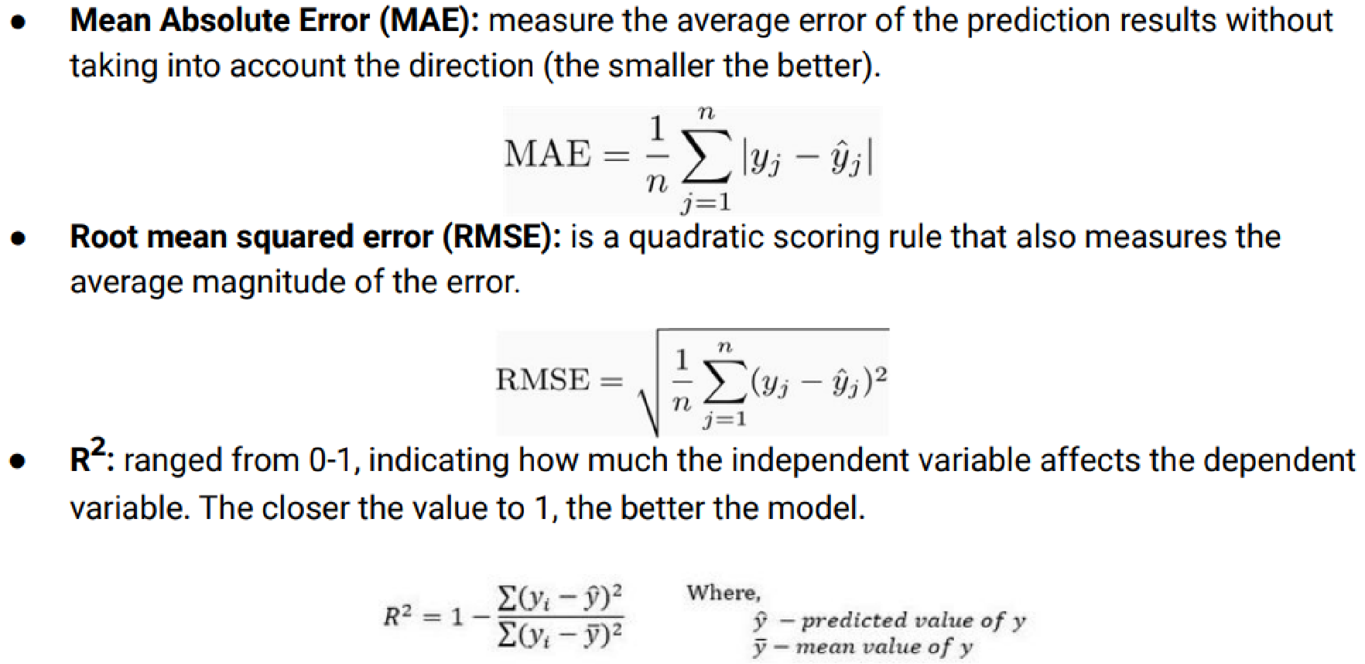


04

Supervised Learning - Regression

Model Evaluation (Regression)

Model Evaluation



- **Mean Absolute Error (MAE):** measure the average error of the prediction results without taking into account the direction (the smaller the better).

$$\mathrm{MAE} = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j|$$

- **Root mean squared error (RMSE):** is a quadratic scoring rule that also measures the average magnitude of the error.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

- **R^2:** ranged from 0-1, indicating how much the independent variable affects the dependent variable. The closer the value to 1, the better the model.

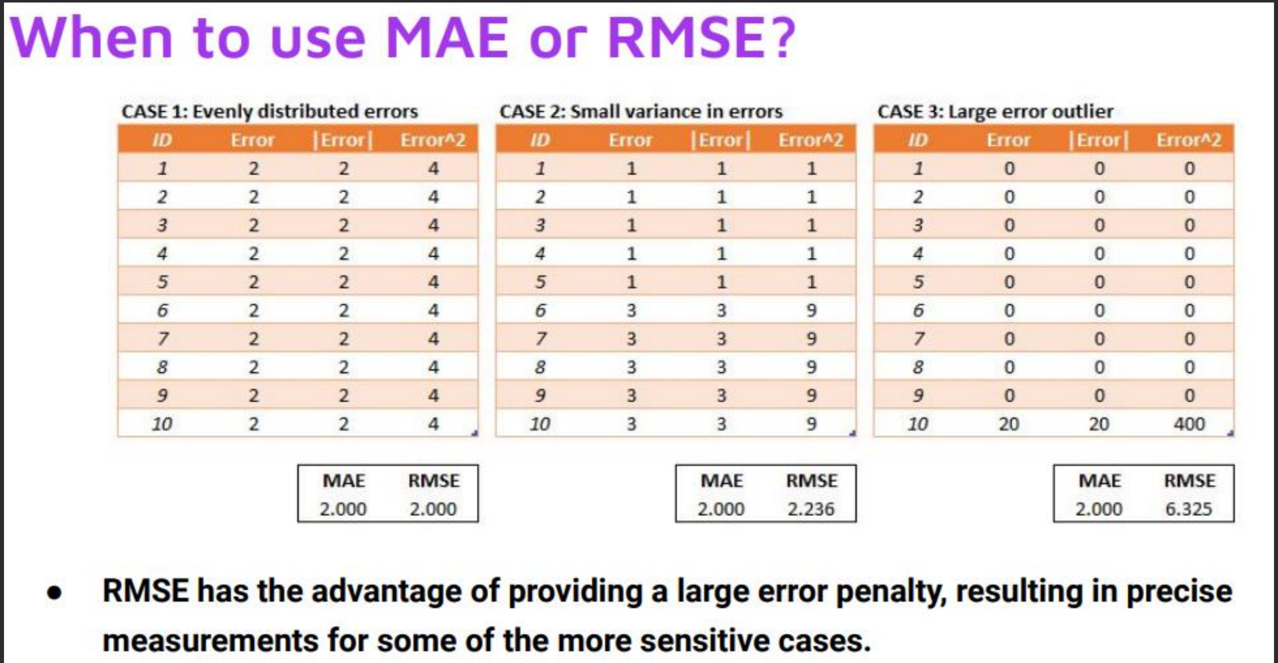
$$R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2}$$

Where,

$\hat{y}$ − *predicted value of y*

$\bar{y}$ − *mean value of y*

Model Evaluation



When to use MAE or RMSE?

CASE 1: Evenly distributed errors

ID	Error	|Error|	Error^2
1	2	2	4
2	2	2	4
3	2	2	4
4	2	2	4
5	2	2	4
6	2	2	4
7	2	2	4
8	2	2	4
9	2	2	4
10	2	2	4

MAE	RMSE
2.000	2.000

CASE 2: Small variance in errors

ID	Error	|Error|	Error^2
1	1	1	1
2	1	1	1
3	1	1	1
4	1	1	1
5	1	1	1
6	3	3	9
7	3	3	9
8	3	3	9
9	3	3	9
10	3	3	9

MAE	RMSE
2.000	2.236

CASE 3: Large error outlier

ID	Error	|Error|	Error^2
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	0
5	0	0	0
6	0	0	0
7	0	0	0
8	0	0	0
9	0	0	0
10	20	20	400

MAE	RMSE
2.000	6.325

- **RMSE has the advantage of providing a large error penalty, resulting in precise measurements for some of the more sensitive cases.**

Thanks!

Do you have any questions?

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**

r/cats

Posted by u/WifeKnowsImOnHere • 5h

Duško conquering his fear of the bathtub with his binky