**ML Basics**

**¯\\_(ツ)\_/¯**

* AI - perform human like tasks and simulate human behaviour
* ML - subset of AI that tries to solve a specific problem and make predictions using data
* DS - (it uses ML) field that attempts to find patterns and draw insights from data

## Types of machine learning algorithms

1. **Supervised learning**

* uses **LABLED INPUT** (data) to train models. (inputs has a corresponding output labels)
* Training the model with labeled data to make predictions on unseen data.

1. **Unsupervised learning**

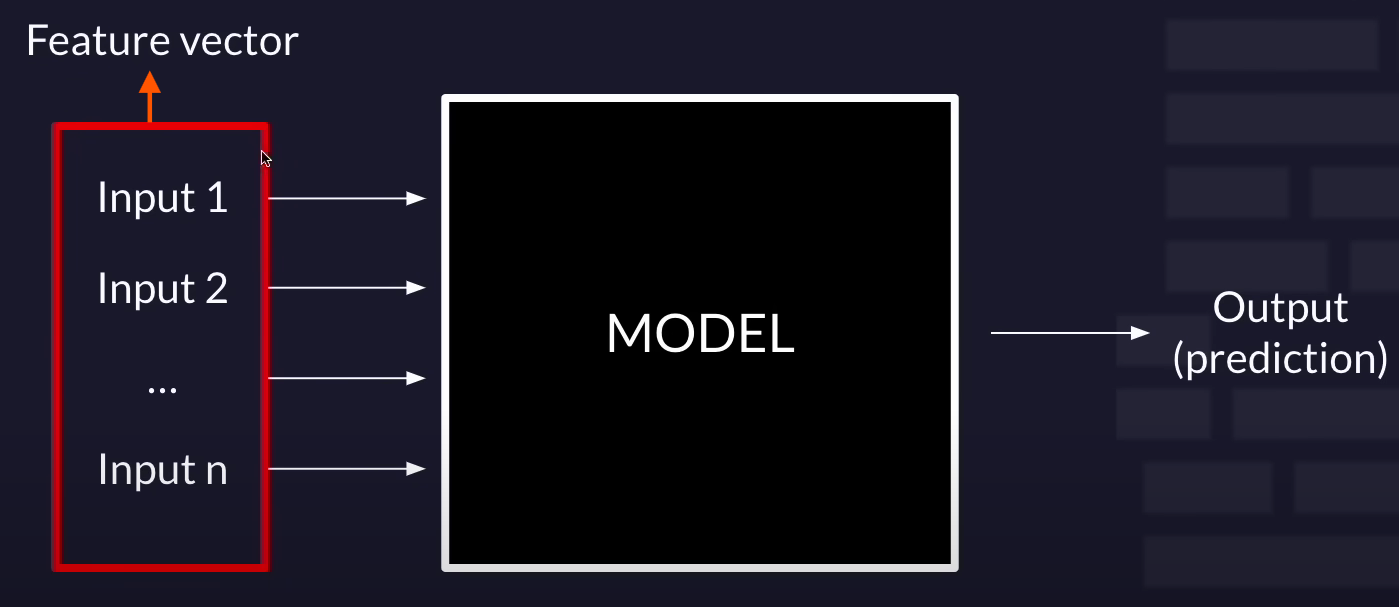
* uses **UNLABELED DATA** to learn about patterns in data (don't know cats n dogs but knows the common clusters)
* Model learns from unlabeled data, discovering patterns and structures without explicit guidance.
* it can reveal hidden classifications (**latent variables**) that were not initially apparent.

1. **Reinforcement learning**

* learning in interactive env based on **REWARDS AND PENALTIES** (like training a dog)

1. **Recommender systems**

# Supervised learning



## FEATURES = Individual measurable properties or characteristics of the data used for making predictions in a ML model.

* **Features and Feature Vectors**: Inputs to the model, structured as feature vectors.
* Features can be
  + **Qualitative** (categorical) Gender, nationality, etc.
  + **Ordinal** (inherent order) Age groups, ratings, etc.
  + or **Quantitative** (discrete or continuous) numerical data. Length, temperature, count of items, etc.

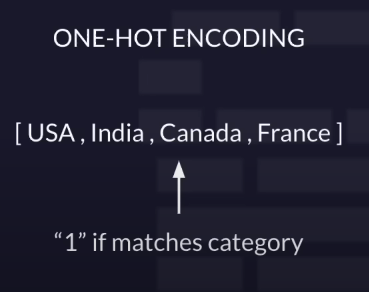
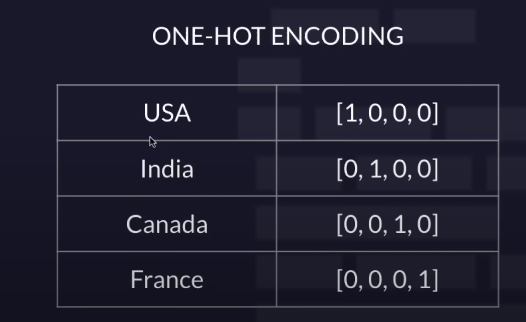
**Encoding:** Qualitative and ordinal features may require encoding for use in machine learning models.

* **One-Hot Encoding:** Commonly used for nominal (categorical) features.
* **Numerical Encoding:** Assigning numerical values to ordinal features based on their order.
* **Preprocessing:** Before feeding into the model, features often undergo preprocessing steps such as normalization or scaling to ensure uniformity and enhance model performance.

**Feature Vector** is a structured representation of data consisting of multiple features organized into a single vector or array

**QUALITATIVE** - categorical data. (finite no. of categories or groups)

1. **NOMINAL DATA** - no inherent order
   * for nominal data to feed using **ONE-HOT-ENCODING** *(is a technique in ML that turns categorical data, like colors (red, green, blue), into numerical data for machines to understand.)*

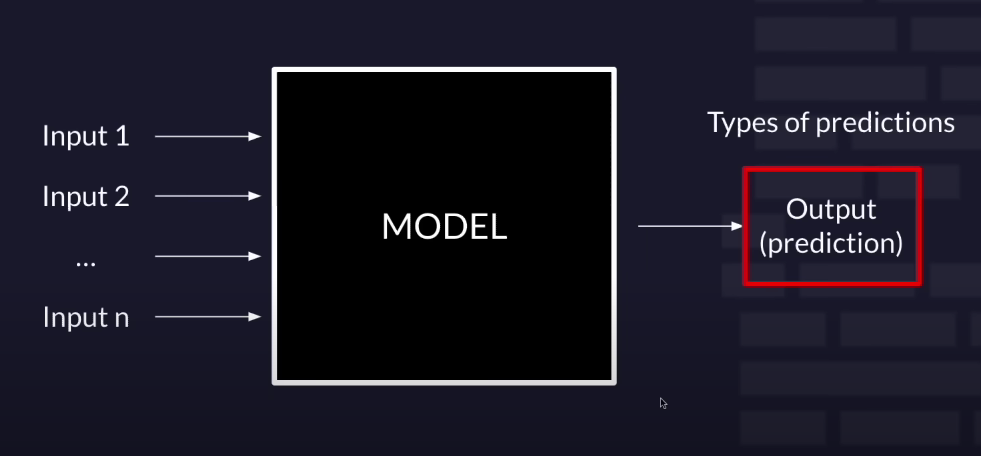
 

**ONE-HOT-ENCODING** - if matches some categories make that 1 if doesn't make 0

1. **ORDINAL DATA** - inherent order
   * like ranking 1,2,3,4,5 from bad to good

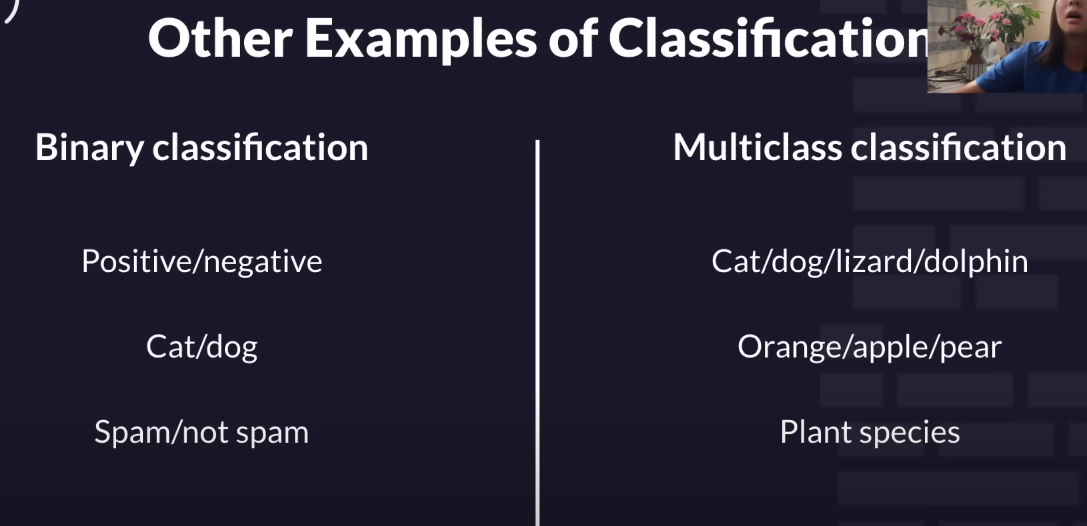
**QUANTITATIVE** - numerical value data (e.g. real no., temperature, length)

# Types of predictions



# Supervised learning Tasks

1. **Classification:** Predict discrete classes (e.g., multi-class or binary).
   1. **Multi-class Classification**: categorizing data multiple distinct classes or categories. three or more classes. *It's similar to sorting objects into multiple buckets.*
   2. **Binary Classification:** Predicting between two mutually exclusive classes, either positive or negative result. *Example - Classifying emails as spam or not spam.*

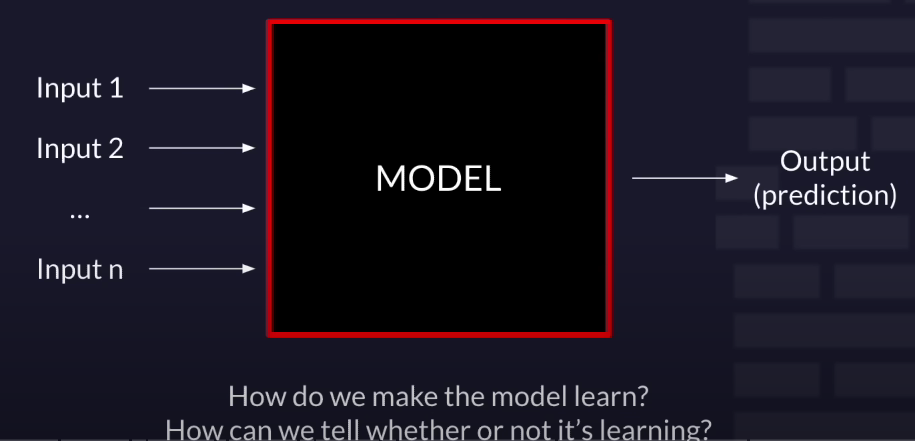


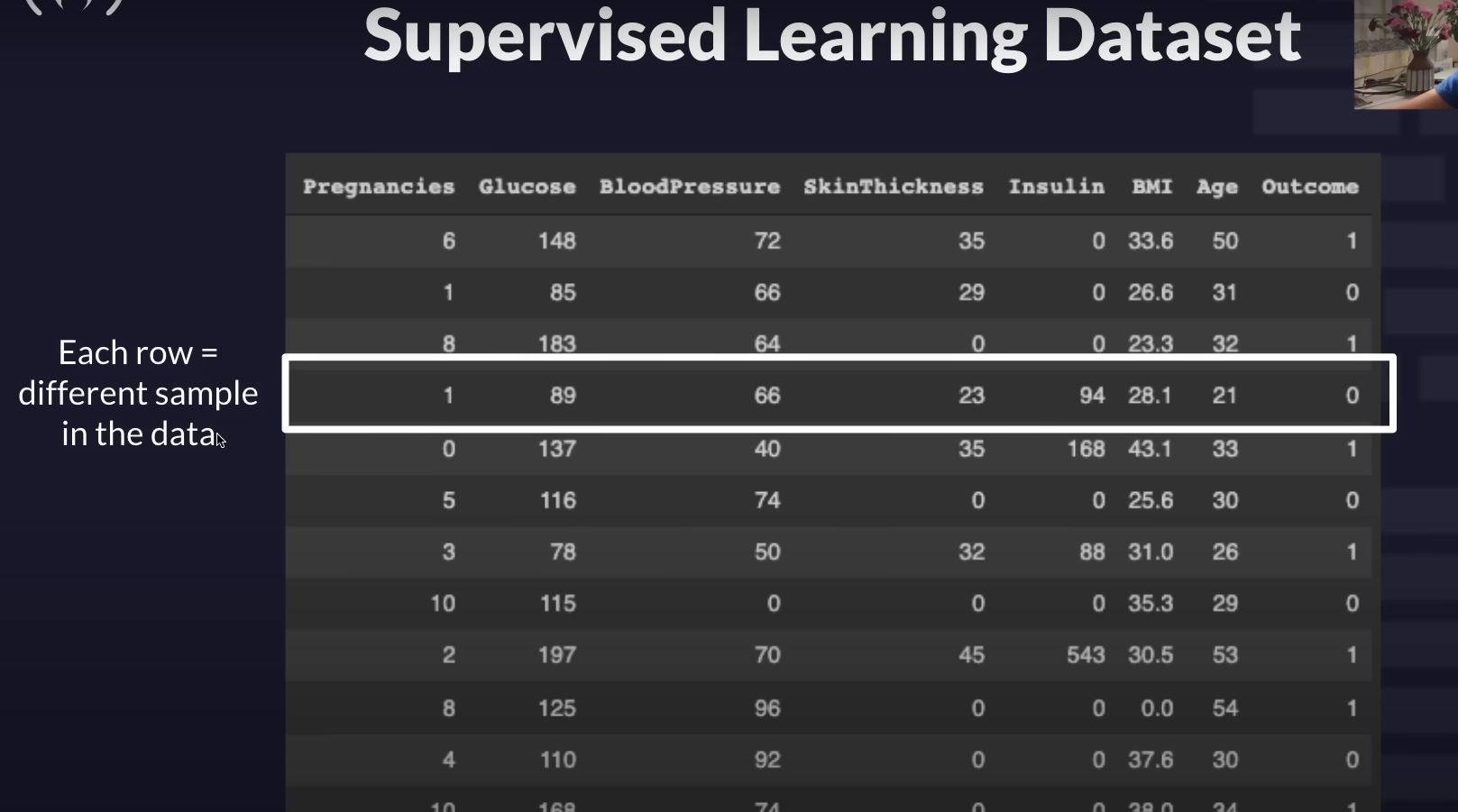
1. **Regression:** **predicting continuous numerical values or quantities** based on input features.

Examples:

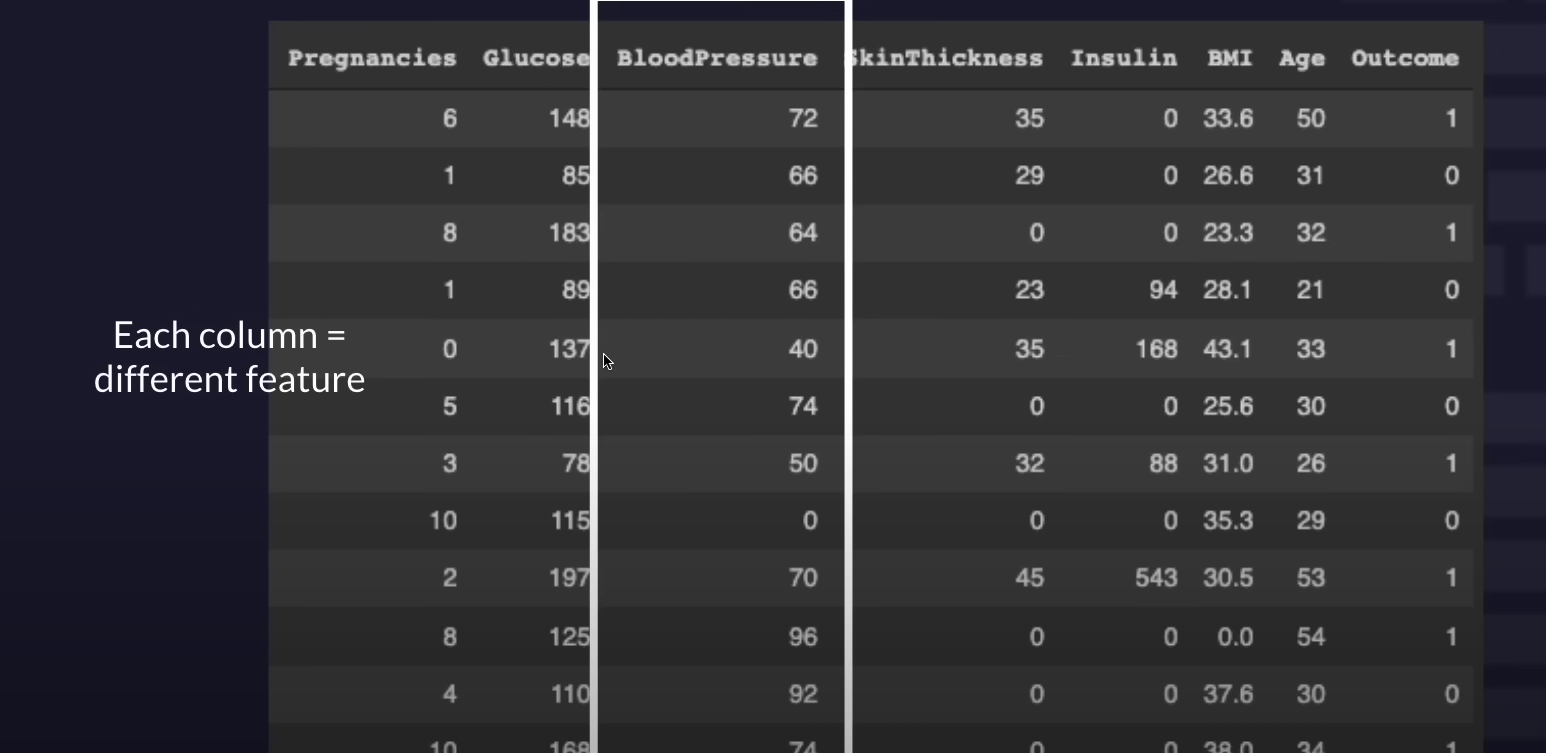
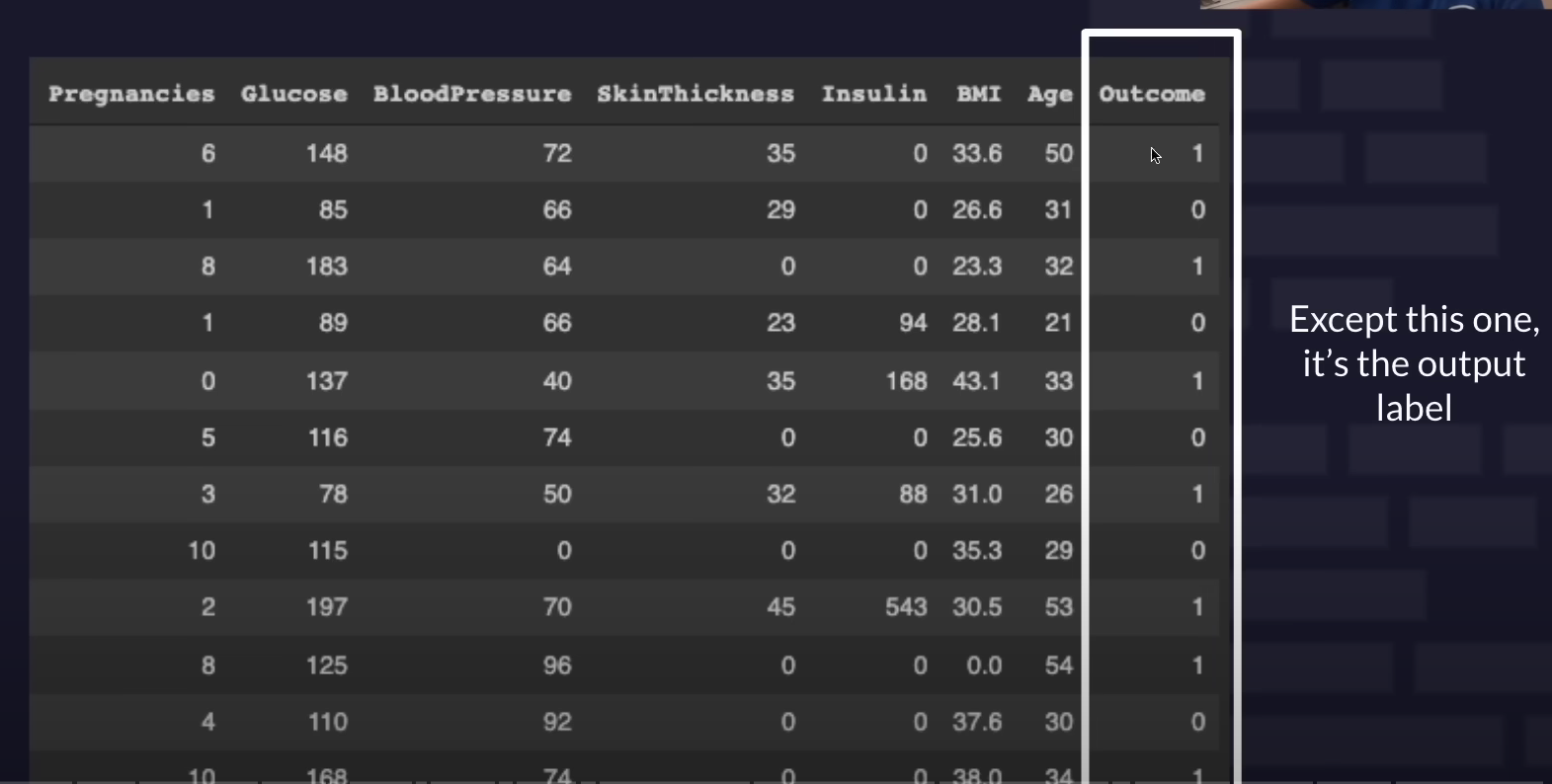
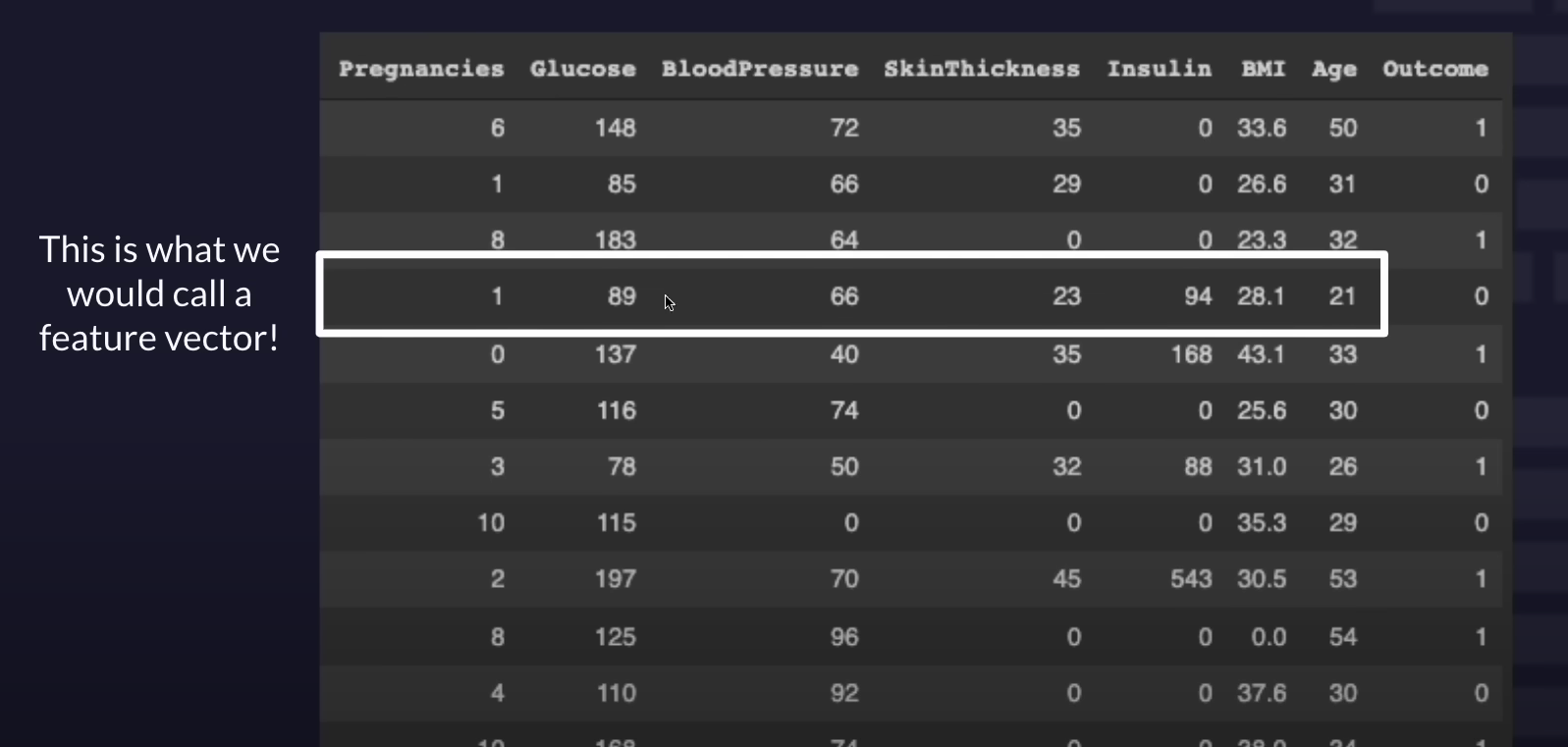
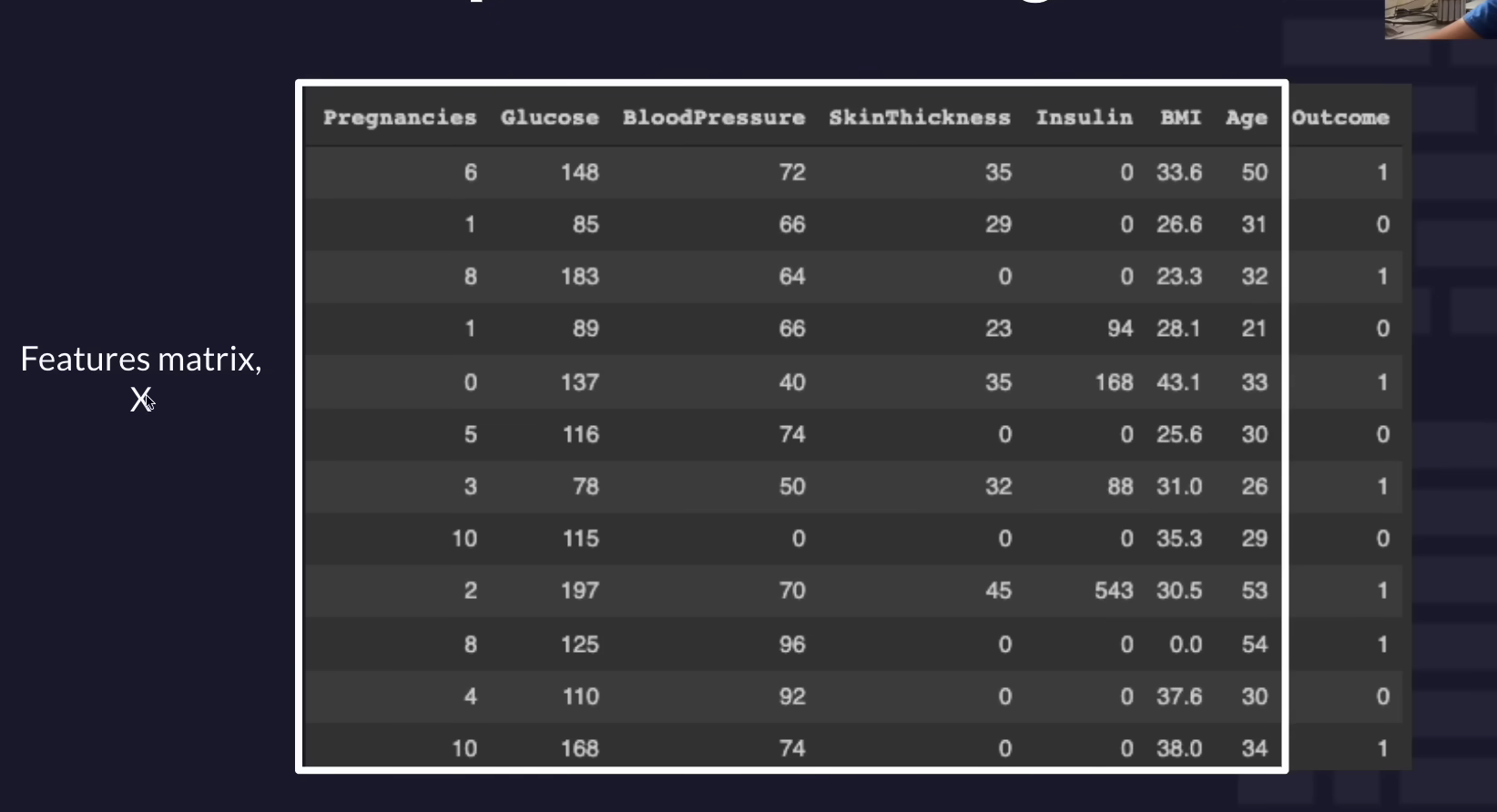
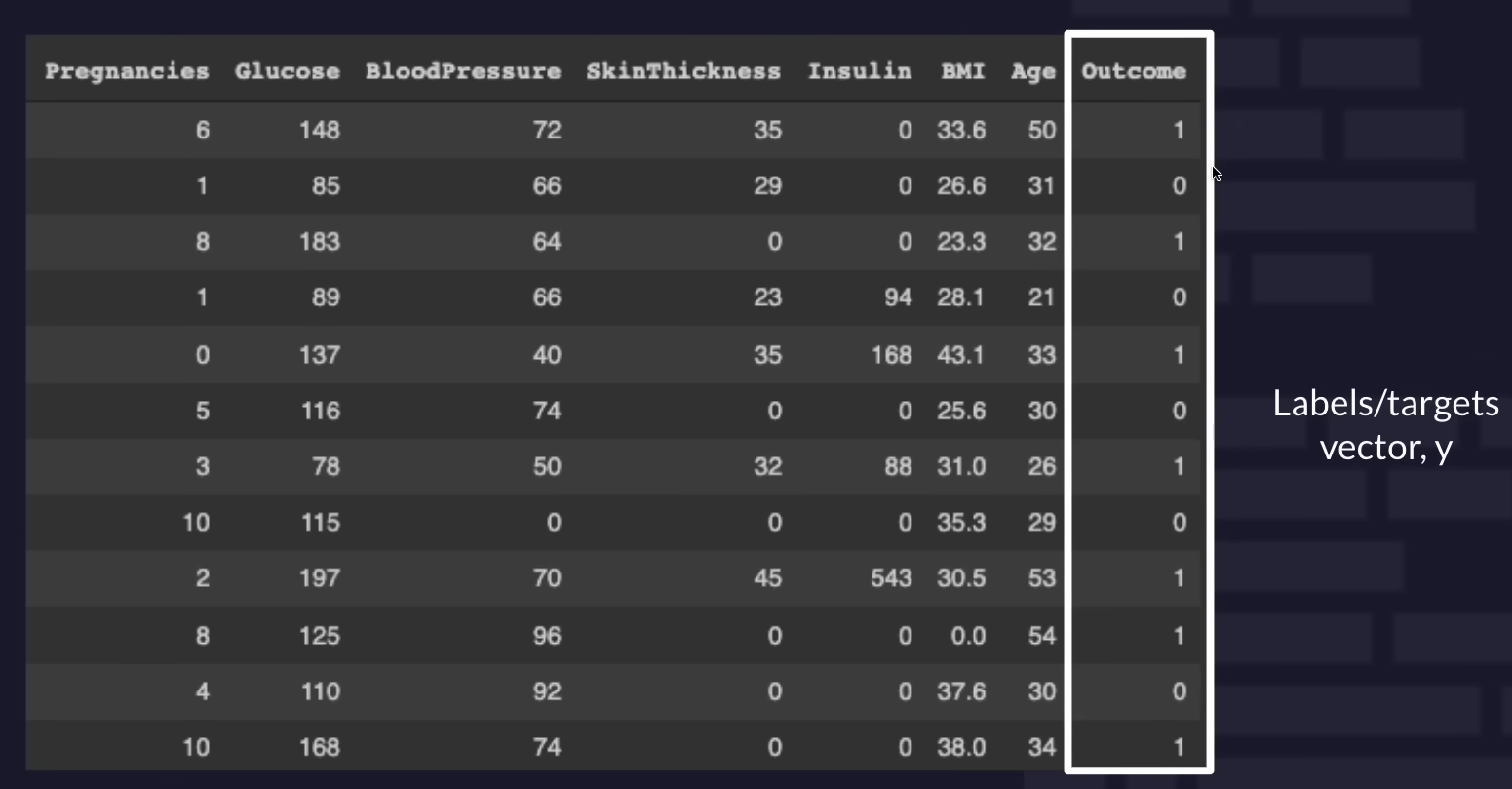
* Predicting house prices based on features like area, location, etc.
* Forecasting stock prices based on historical data.
* Estimating temperature based on weather parameters

# Model Evaluation

****

****

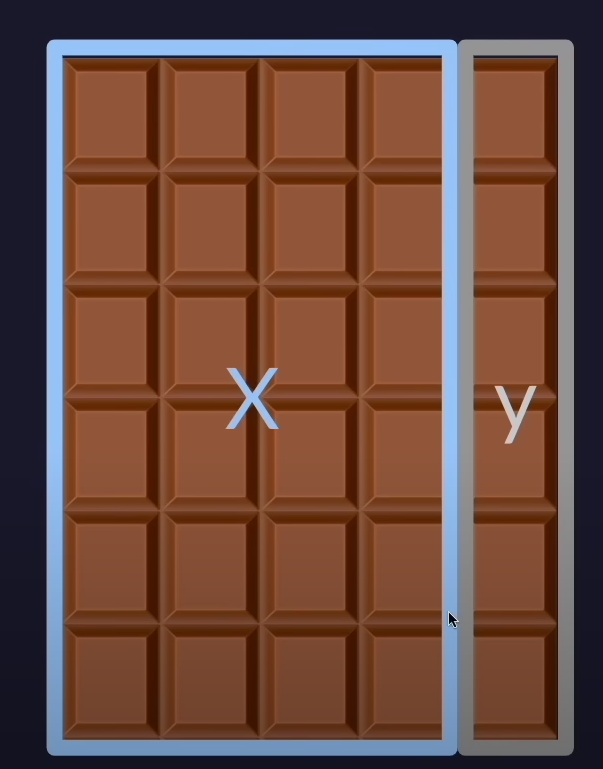
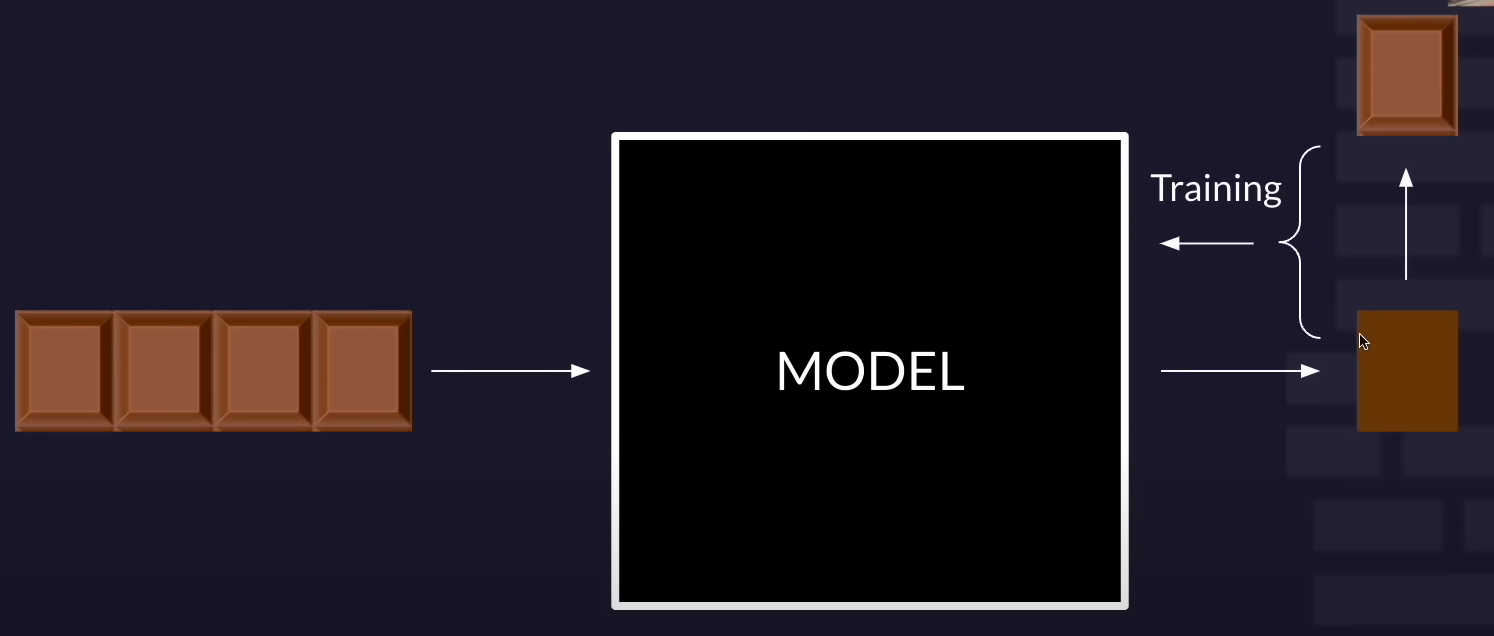
**Each row represents 1 person / 1 sample**

**** ****    

Note: Y, target vector is not always the last coll

Expect this column

it is output

Think DATA BASE as choklet bar (I know spelling 🍫 is 😵)  
**** 



comparing

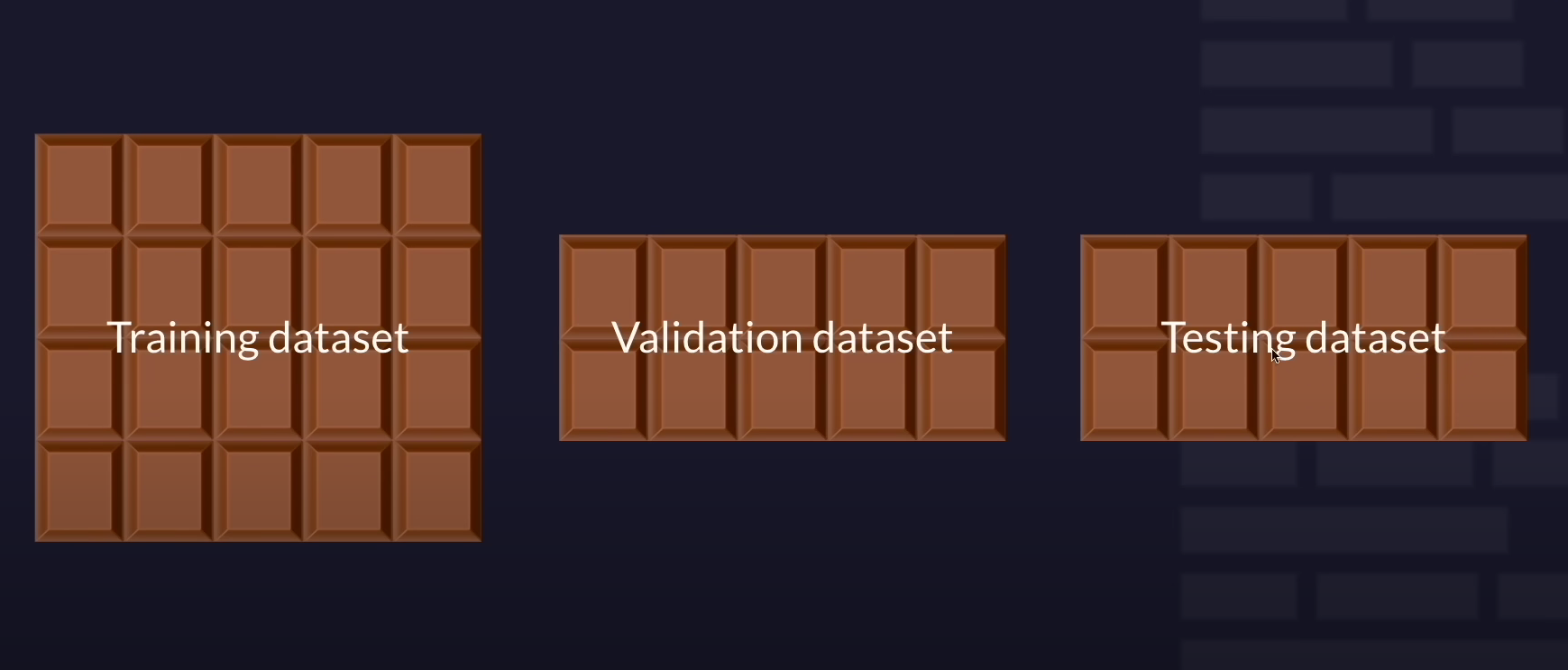
prediction

Model have seen same data again and again

**Can our model handle new Data?**

**For that we -**

* **Training Data:** Used to train the model by adjusting its parameters based on labeled examples.
* **Validation Data:** Assesses model performance during or after training, ensuring it can generalize to unseen data.
* **Testing Data:** Final evaluation on completely unseen data, providing an unbiased assessment of model performance.

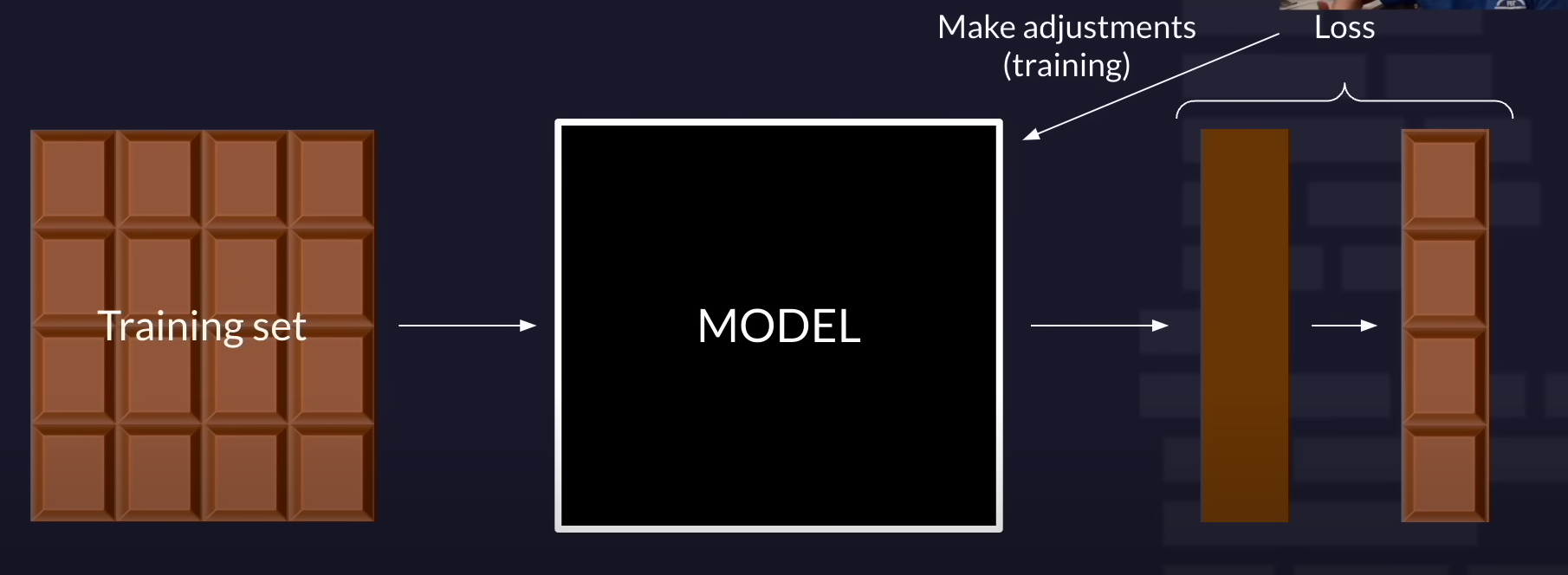


20% or 10% 🍫

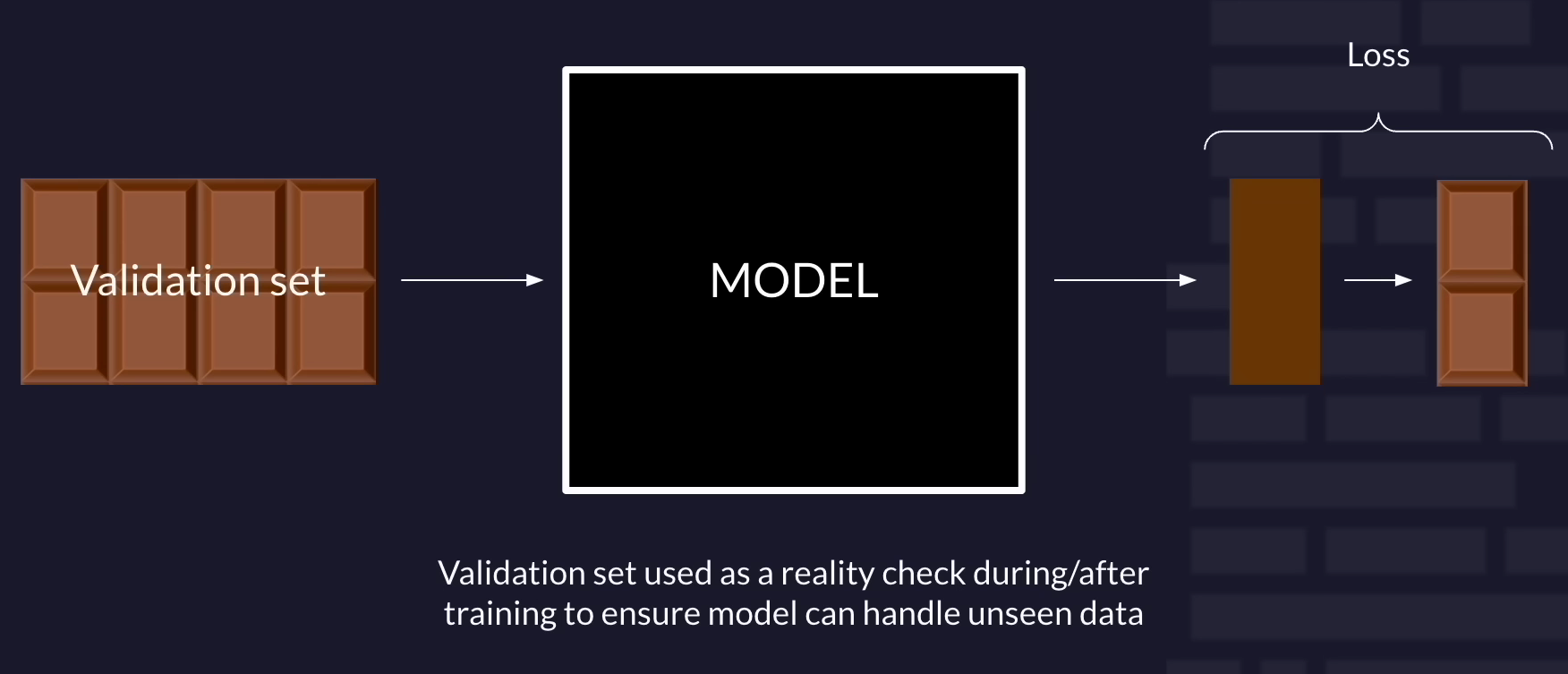
20% or 10% 🍫

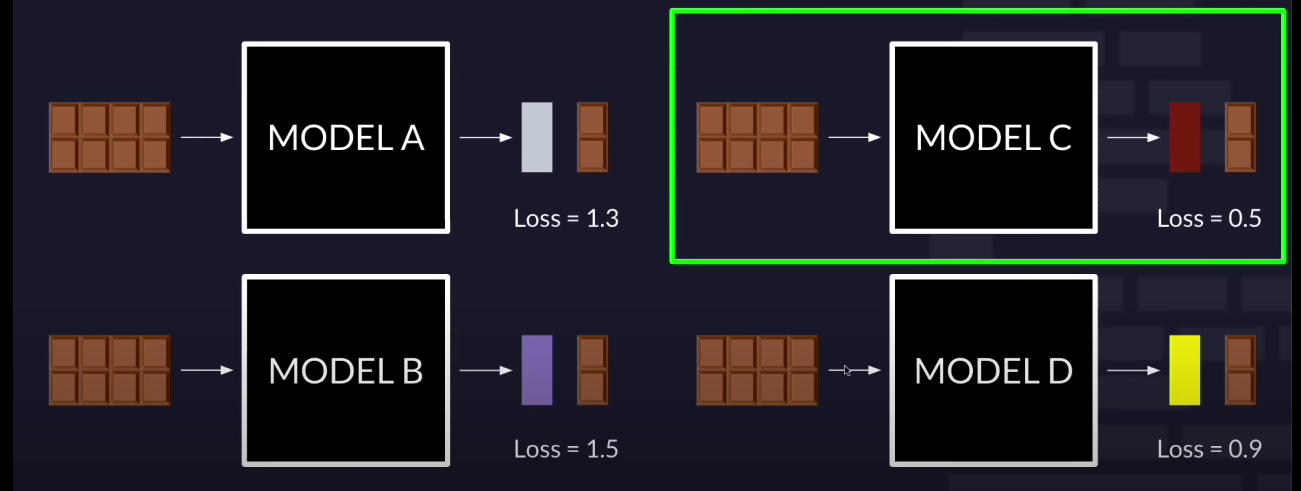
60 % or 80%🍫

This is the Full Data Set --🍫 slice



This loss never gets to the Training foAmodel

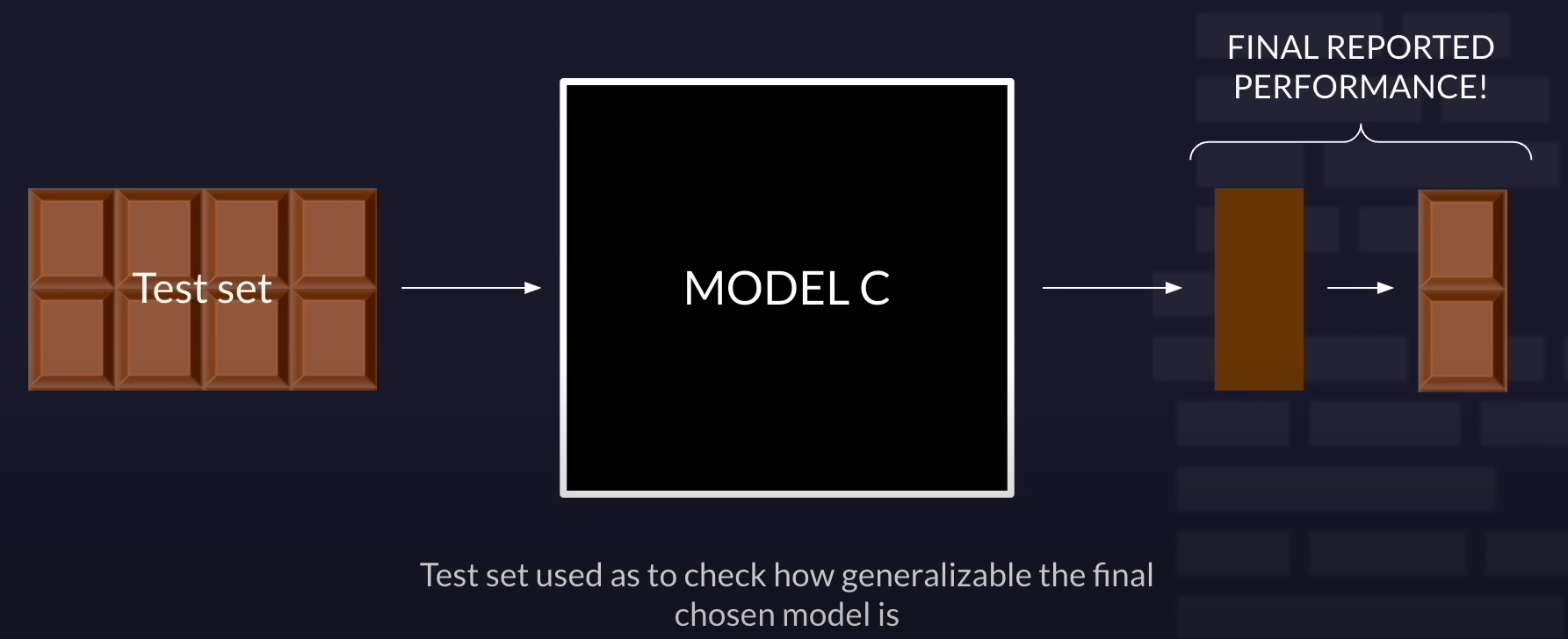




Performs best

**Now, run TEST SET threw MODEL C**

So, we can see how this model perform on data that it never seen before during this Process



# Metrics of performance

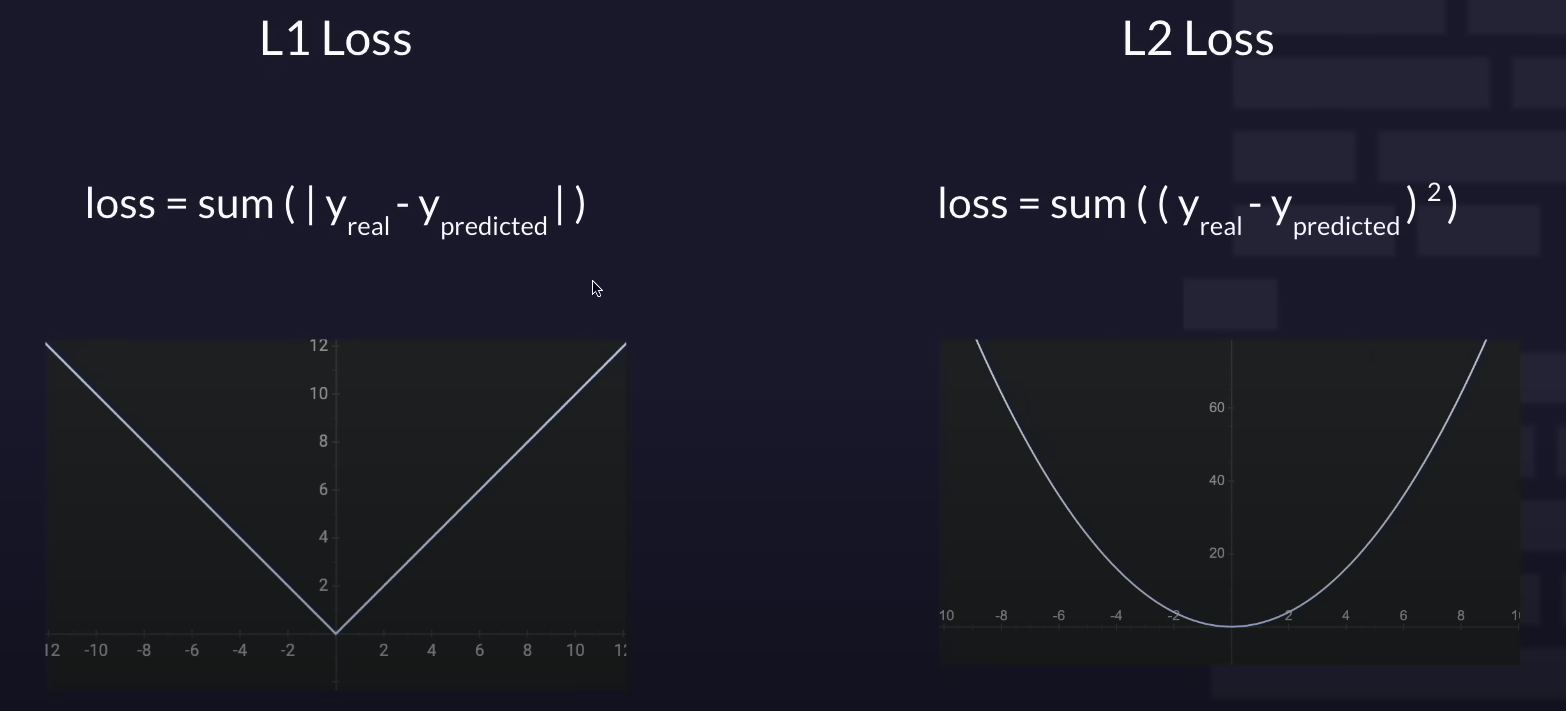
# LOSS

**Loss Function:** Measures the disparity between predicted and actual values.

1. **Types**:

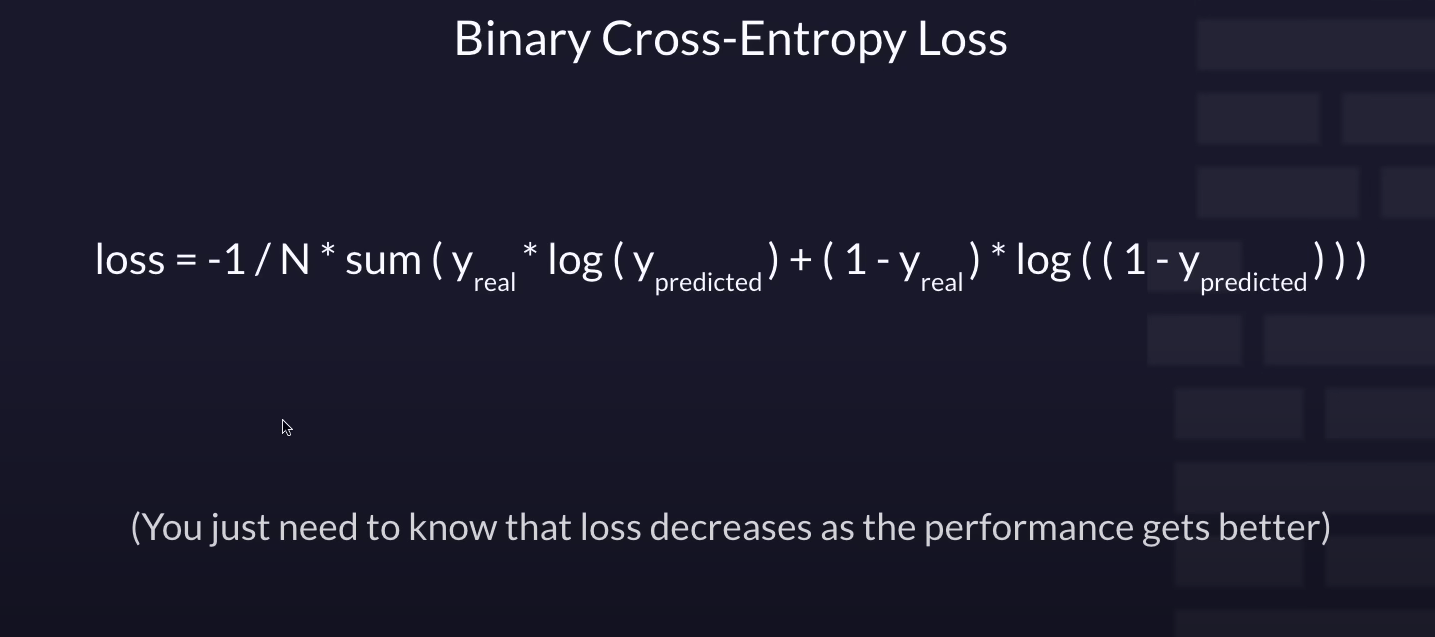
* **L1 Loss**: Absolute difference between predicted and actual values.
* **L2 Loss**: Quadratic difference between predicted and actual values.
* **Binary Cross Entropy Loss (log loss)**: Specific to binary classification tasks.

1. **Purpose**: Guides model training by quantifying prediction errors.



**further from 0 more loss + or -**

**further from 0 more loss + or -**



THE F🦆 ------ I DON’T KNOW WHAT THE FUCK IS THIS

😵😵😵😵

------end------

**Unsupervised learning and Neuron network** Not included.