Preparing Data for Gradient Boosting

Mike West
MACHINE LEARNING ENGINEER

Module Overview

The AI hierarchy

Artificial neural networks

Supervised vs. unsupervised learning

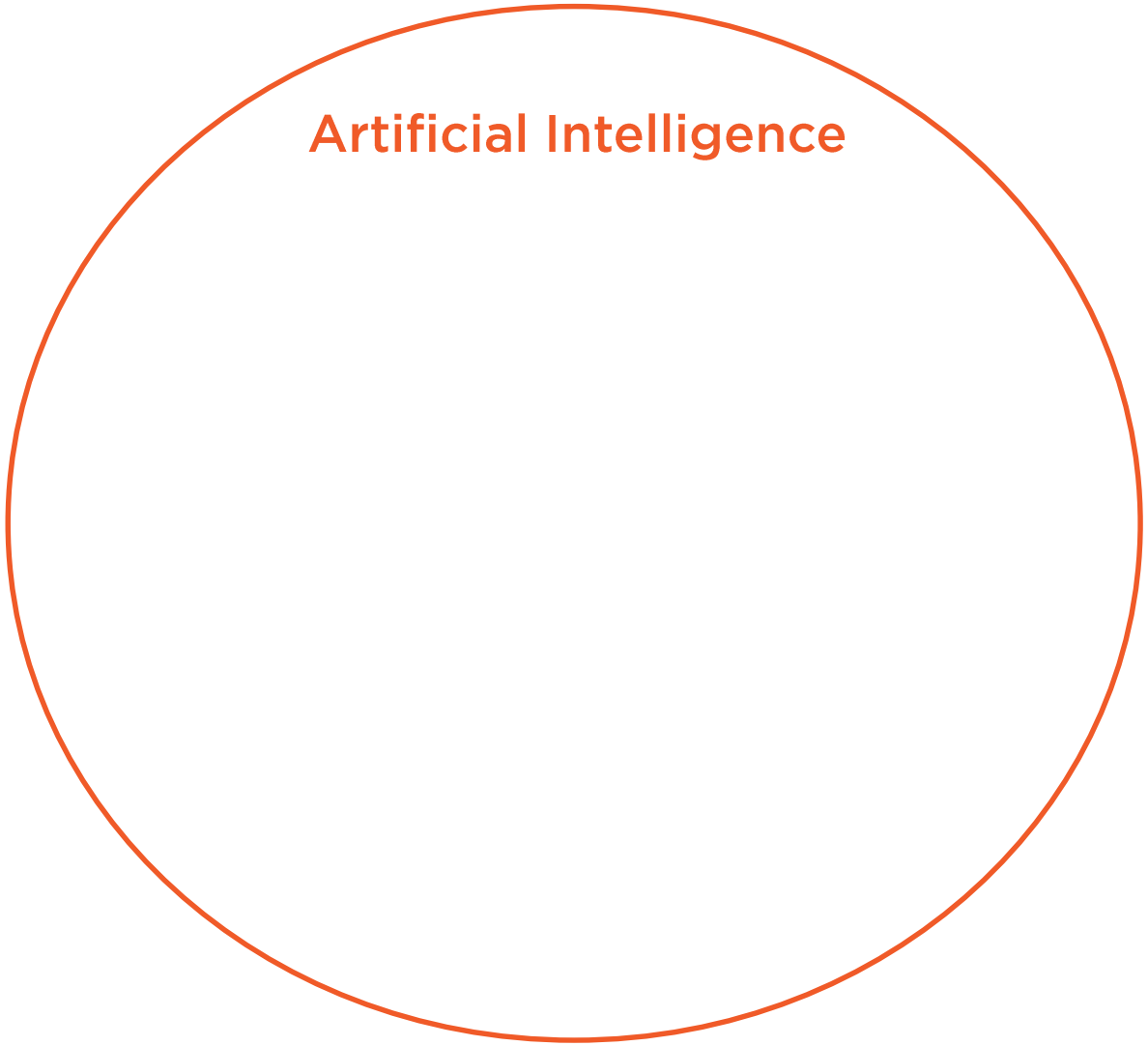
Applied machine learning

The importance of the array

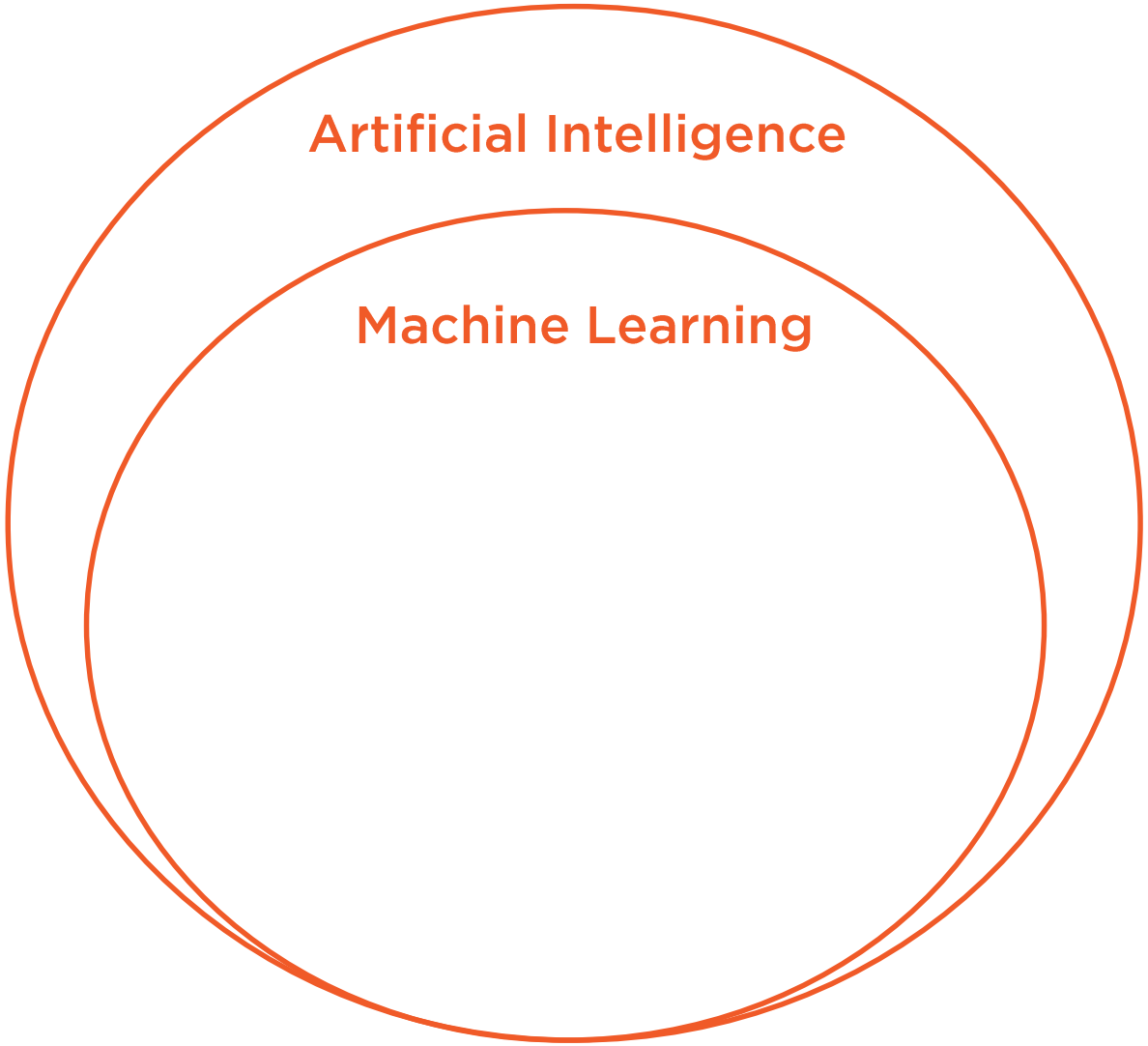
Machine learning process

Clean and model the Titanic dataset

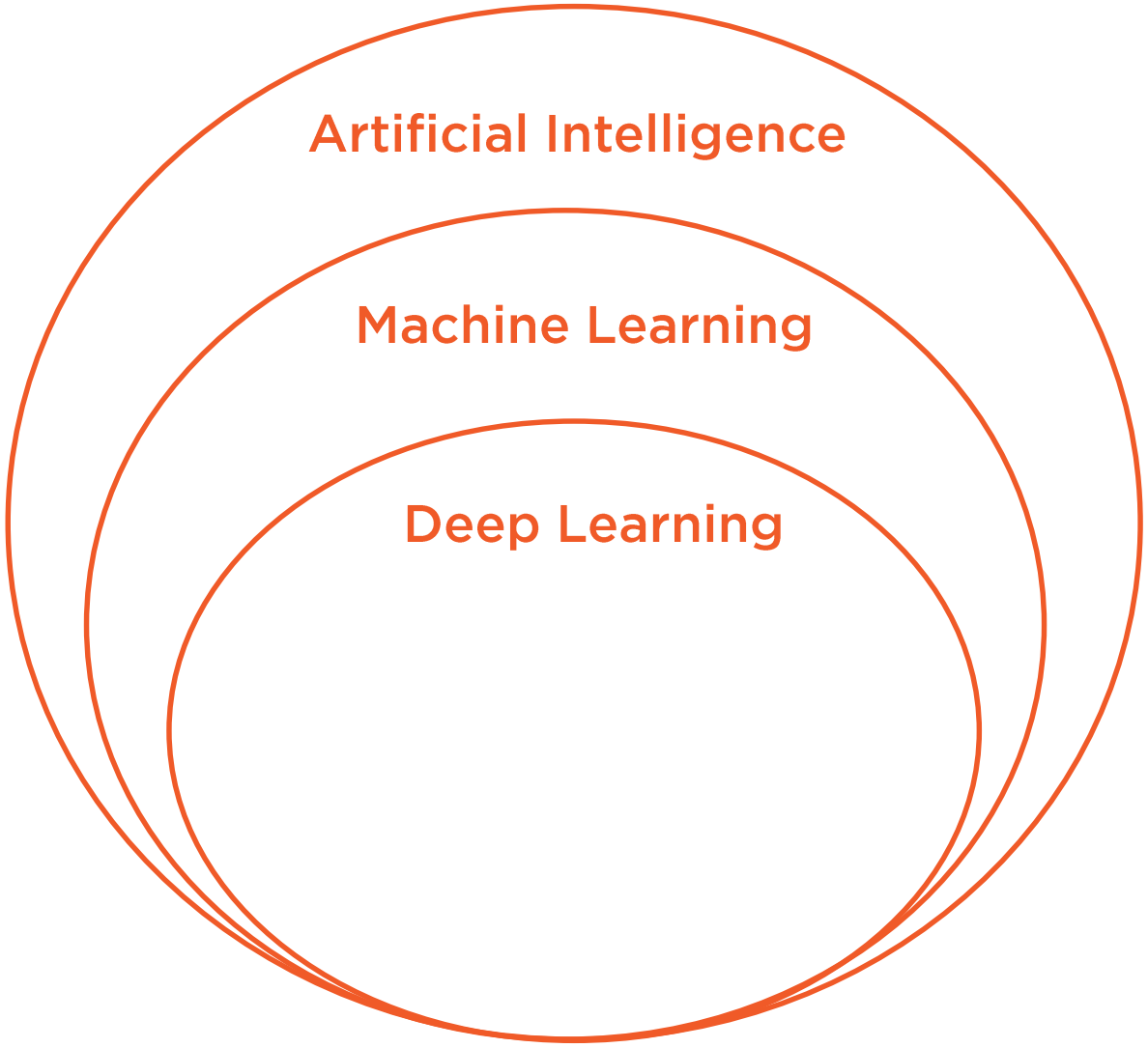
The AI Hierarchy



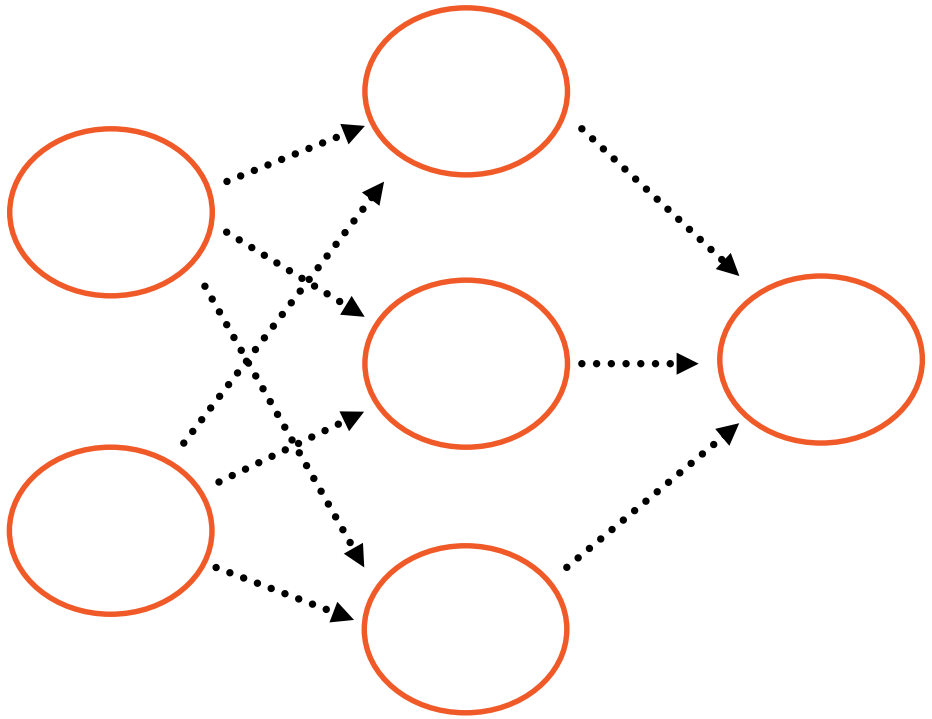
The AI Hierarchy



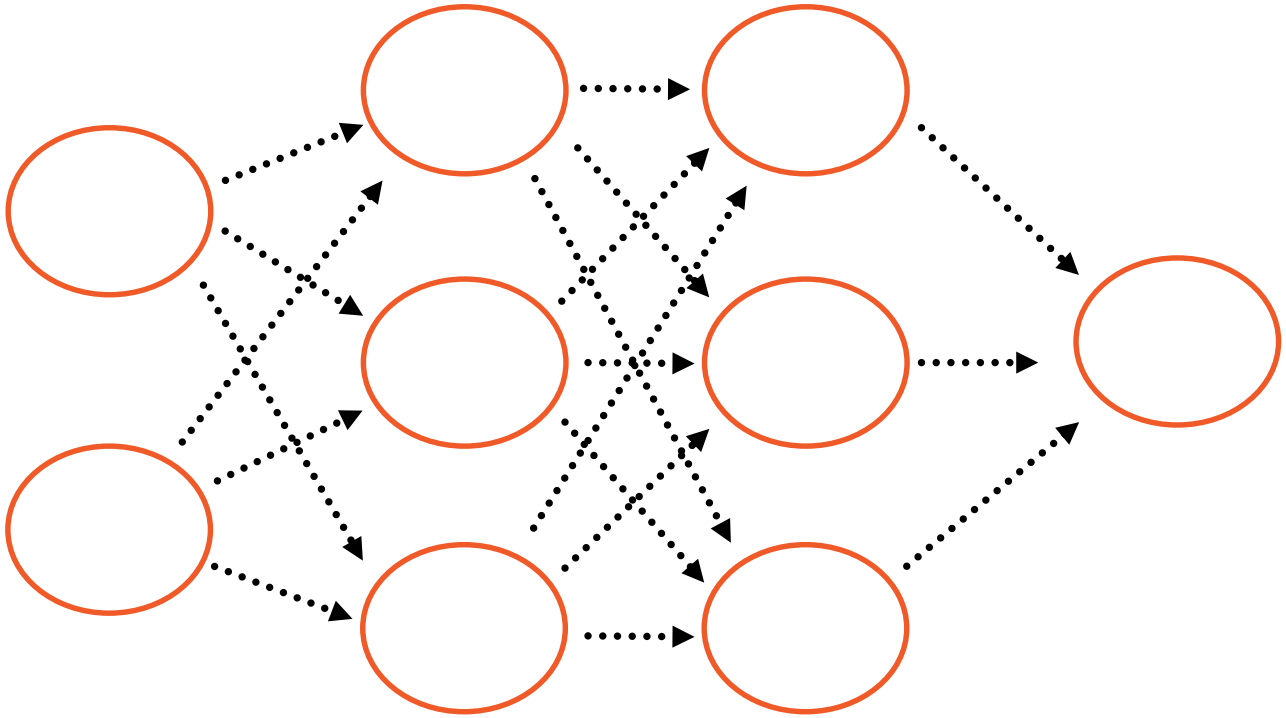
The AI Hierarchy



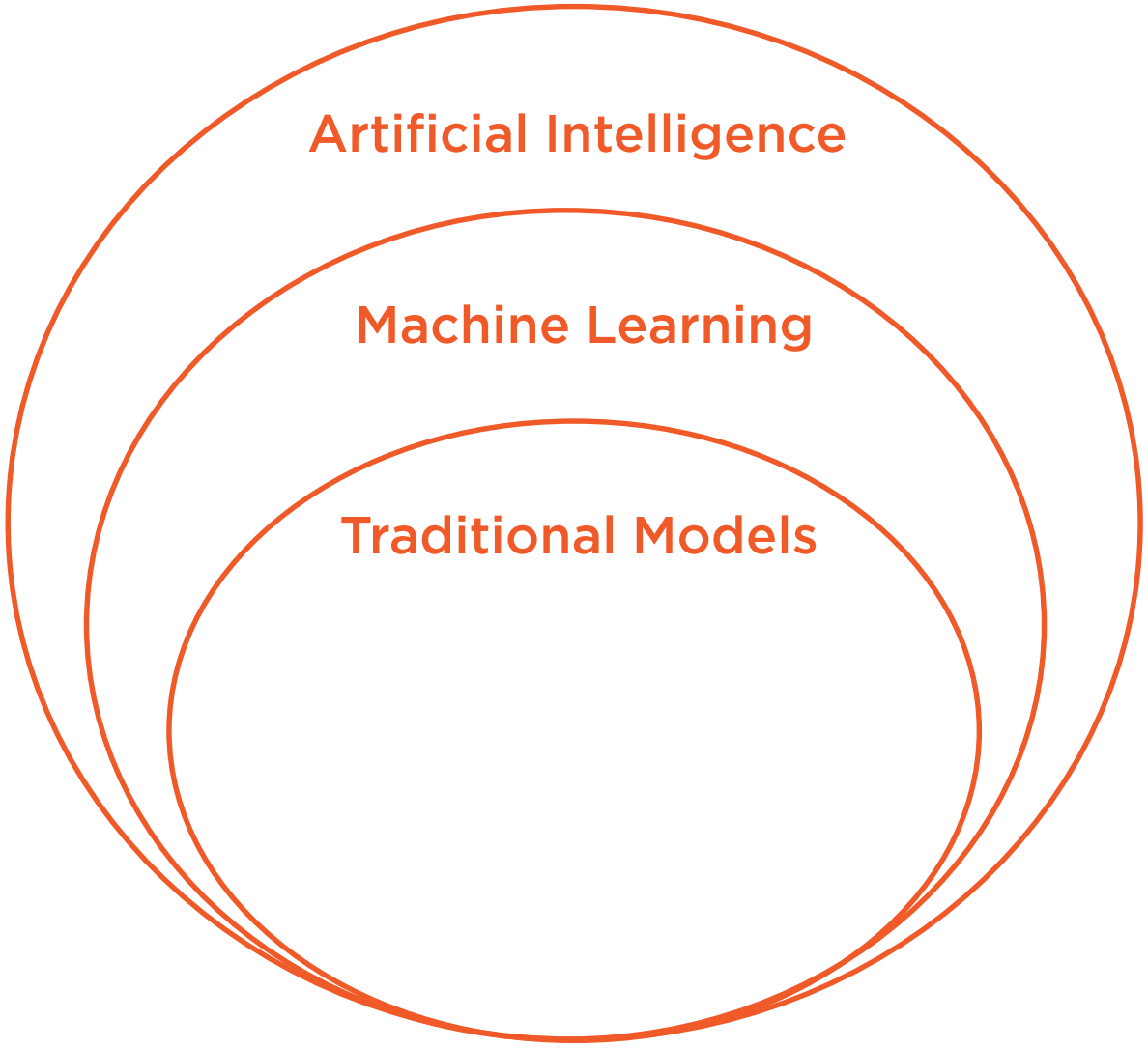
Artificial Neural Network



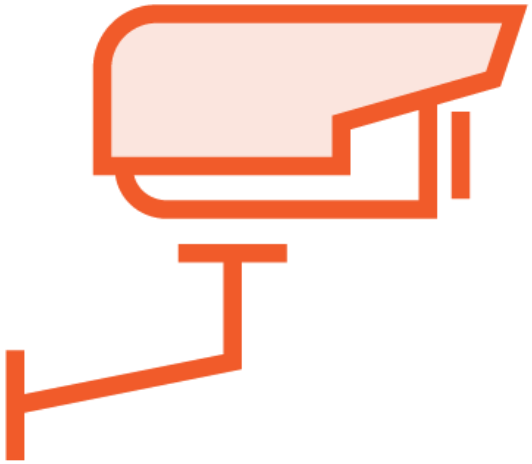
Deep Learning Network



The AI Hierarchy

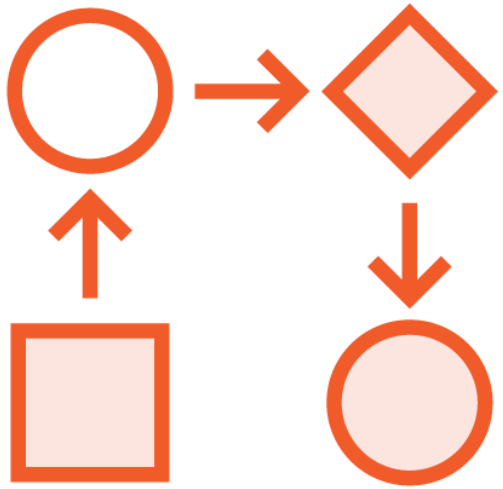


Two Types of Machine Learning



Supervised

The models look for patterns in an existing dataset



Unsupervised

The models find patterns and structure on unlabeled data

Applied Machine Learning

The majority of real world modeling is supervised learning

Supervised learning always starts with sourcing data

Data wrangling is critical to model performance

Two-dimensional Array

	0	1	2	3
0	1	Haws, Zech	male	22
1	2	West, Mike	male	51

Two-dimensional Array

	0	1	2	3
0	1	0,1	male	22
1	2	West, Mike	male	51

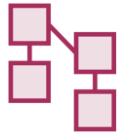
Two-dimensional Array

	0	1	2	3
0	1	0,1	male	22
1	2	1,1	male	51

The Machine Learning Process

Supervised machine learning starts with sourcing data

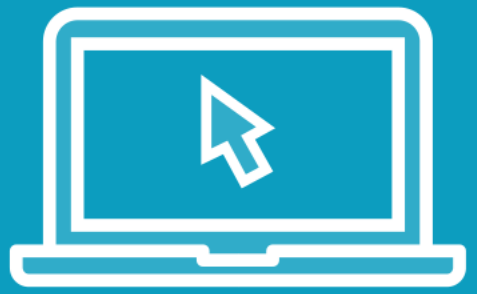
Data wrangling is massaging the sourced data



Modeling is looking for patterns in the wrangled dataset

The final step is putting the model in production

Demo



Import libraries

Load Titanic dataset

Wrangle the Titanic dataset

Train XGBoost model on dataset

Make predictions

Summary

- Defined AI hierarchy
- The deep learning network
- Supervised machine learning
- Unsupervised learning
- The array
- Machine learning process
- XGBoost model on Titanic dataset