

Structured vs Unstructured Data

Structured Data

- Organized in predefined format
- Easy to query and analyze
- Efficient storage
- Standard analytical tools

Examples:

Tables, databases, spreadsheets

VS

Unstructured Data

- No predefined structure
- Rich information content
- Complex processing required
- Advanced analytical methods

Examples:

Text, images, audio, video



Semi-Structured Data

Hybrid format with some organization (JSON, XML)

80-90%

of enterprise data is unstructured and growing rapidly

Structured Data Examples

Classification Example

Classification with PCA Features

Age	Income	Credit Score	Education	Approved
25	45K	680	Bachelor	No
35	75K	720	Master	Yes
42	90K	780	PhD	Yes
28	52K	650	Bachelor	No
50	110K	800	Master	Yes

PC1	PC2	PC3	Gender	Churn
2.4	-1.2	0.8	M	No
-0.5	3.1	-1.5	F	Yes
1.8	0.3	2.2	M	No
-2.1	2.7	-0.9	F	Yes
3.2	-0.8	1.4	M	No

Regression Example

Size (sq ft)	Bedrooms	Age (years)	Distance (km)	Price (\$K)
1200	2	10	5.2	320
1800	3	5	3.1	485
2400	4	2	1.8	650
950	1	15	8.5	245
2100	3	8	4.3	560

Regression with PCA Features

PC1	PC2	PC3	PC4	Sales (\$M)
12.5	-3.2	1.8	0.5	2.4
8.7	4.1	-2.3	1.2	3.8
15.3	-1.5	3.4	-0.8	5.2
6.2	2.8	-1.1	0.9	1.9
11.8	0.3	2.7	-0.4	4.5