Recall $y \in \left\{0, 1\right\} = Y$
We had a true system (not a model but sometimes be "the model.")

$$y = t(z_1, z_2, z_3)$$

where z_1 = has sufficient funds, z_2 = unforeseen emergency and z_3 = criminal intentions

Problem: $\left\{z_1, z_2, z_3\right\}$ is unobserved (impossible to obtain). What to do?
Next best thing: Try to define and collect information "related" to $\left\{z_1, z_2, z_3\right\}$

Thus use what you have (or what is easily available).
Let's pretend we got the resources to "define and collect."

- x_1: salary - measured by average salary
- x_2: previous loan repayment - did they ever miss previous loan payment? $\in [0, 1]$
- x_3: historical criminal record - previous crime type?
  $\left\{\text{no crime, infraction, misdemeanor, felony}\right\}$

Definition 0.1. Process Assessment: use as much as you got and whatever is cheaply available

Example: use age.

Let's say we have x_1, x_2, x_3. The idea is $\left\{z_1, z_2, z_3\right\}$ which contains some info in $\left\{x_1, x_2, x_3\right\}$.

Let $\vec{x} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}$ where the LHS is an observation/ record/ object/ input/ independent variable and the RHS is features/ attributes/ characteristics/ regressors/ covariances/ predictors.
Note that $\dim \vec{x} = p$ or d.

$$\vec{x} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \in X$$

where X is the covariance space.
Spaces: $x_1 \in \mathbb{R}$, $x_2 \in [0, 1]$ - binary ordinary variable, x_3 - categorized variables with 4 "levels"

Two Ideas:
First to do: code is numerical, such as $x_3 \in \begin{bmatrix} 0 & 1 & 2 & 3 \end{bmatrix}$ - this should only be done if predictor is "ordinal."
Next to do: Take x_3 and turn it into x_{3a} (binary no crime), x_{3b} - binary infraction, x_{3c}, x_{3d}.
This increases p from 3 to 6 - more variables to think about.
So, it is impossible to get $\left\{z_1, z_2, z_3\right\}$ but we do have $\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}$
GOAL: Do the best we can to explain y by creating a model f, the approximation - the best

relationship we can get. Does $y = f(x_1, x_2, x_3)$? No.
In fact,

$$y \approx f(x_1, x_2, x_3)$$
$$y = f(x_1, x_2, x_3) + \delta$$

where $\delta = t(\vec{z}) - f(\vec{x})$, which comes from ignorance.
How do we get f? First note there is no analytical solution.
Example: $h(x) = x^2$. Find $\min h$.
Instead, use an "empirical solution." An example of this is using data to learn from data.

Definition 0.2. Supervised Learning: uses historical examples of record and their responses

In this case, it requires 3 ingredients:

$$\mathcal{D} := \left\{ \langle \vec{x}_1, y_1 \rangle, \langle \vec{x}_2, y_2 \rangle, \langle \vec{x}_3, y_3 \rangle \right\}$$

where $\vec{x}_1$ is Bill's characteristics and y_1 is whether or not he paid back loan, $\vec{x}_2$ is Jill's, etc.
Let

$$X = \begin{bmatrix} \vec{x}_1 \\ \vec{x}_2 \\ \vdots \\ \vec{x}_n \end{bmatrix} \in X^n, \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \in Y^n$$

where $\dim(x) = n \cdot p$ and $\dim(\vec{y}) = n$.