



7 Steps

Machine Learning Problem Formulation

Component

1. Business Problem Statement
2. Define Unit Analysis
3. Define Variable
4. Define Label
5. Action
6. Success Criteria
7. Threshold





Business Problem Statement

Business Case Component #1

Business Problem Statement

- State a predictive problem to improve client business value.
- Know the impact of predictive result.
- This is should be in design challenge

How might we (goals) so that (impact/outcome)?

Example:

How to target promotion or offer to potential customer based on previous customer?





Define Unit Analysis

Business Case Component #2

Define Unit Analysis

1. Define unit analysis from the data (unit that represents a row / what does every row mean)
2. The row should be consistent throughout dataset
3. It's derived from business problem statement, vice versa

Example:

Business problem -> how much demand on the **several periods ahead**, so we can restock properly?

Unit Analysis -> Time unit (week/month/other)





Define Variable

Business Case Component #3

Define Variable

- Define variable (column) that will be needed for predict the label
- The variable should be consistent in measurement (Example: Currency is Rupiah or Dollar?)
- The variable should be consistent in specifications (Example: NPWP or KTP?)
- Variable should be associated with unit analysis, vice versa

Example:

What kind of product that customer might to buy?

Variable : Product pair, Order ID, Frequency





Define Label

Business Case Component #4

Define Label

- Define Label that will be predicted from the variable
- The label came after variable, because it's predictive
- Label should be associated with unit analysis, vice versa

Example:

How might we choose Supplier/Vendor so that we get fair price for new supplier?

Variable: Lead time (aggregate), cost (aggregate), services, insurance policy, profile

Label: Average price





Action

Business Case Component #5

Action

- What should user do when they get the prediction result
- Should be multiple alternative option of action depend on prediction result
- It's derived from business problem statement, vice versa
- It should be associated with the label, vice versa

Example:

How much the right amount to do restocking?

Label: Restocking needs

Action: Do restocking when product stock amount less than safety stock





Success Criteria

Business Case Component #6

Success Criteria

- Should be either minimize or maximize
- It's should be derived by statistical criteria
- The statistical criteria should be linearly correlated with business impact
- If represent by money, human resources or time, it's better
- It should be associated with label, vice versa

Example:

When does the right time to do restocking for particular product and particular distribution center?

Success Criteria: Minimize understock & overstock phenomenon





Threshold

Business Case Component #7

Threshold

- Either rank (relative / absolute) or certain score
- Threshold should be set for achieve success criteria
- Should be adjust by business condition
- Threshold should be associated with unit analysis
- It should be associated with Label, vice versa

Example:

How much demand on the several periods ahead, so we can restock properly?

Success Criteria: Minimize the occurrence of understock and overstock phenomenon

Threshold: Stock amount less than next certain time unit demand



Relation Checker Table

| | Business Problem Statement | Define Unit Analysis | Define Variable | Define Label | Action | Success Criteria | Threshold |
|----------------------------|----------------------------|----------------------|-----------------|--------------|--------|------------------|-----------|
| Business Problem Statement | | | | | | | |
| Define Unit Analysis | | | | | | | |
| Define Variable | | | | | | | |
| Define Label | | | | | | | |
| Action | | | | | | | |
| Success Criteria | | | | | | | |
| Threshold | | | | | | | |