



Masayu Leylia Khodra
(masayu@informatika.org)

KK IF – Teknik Informatika – STEI ITB

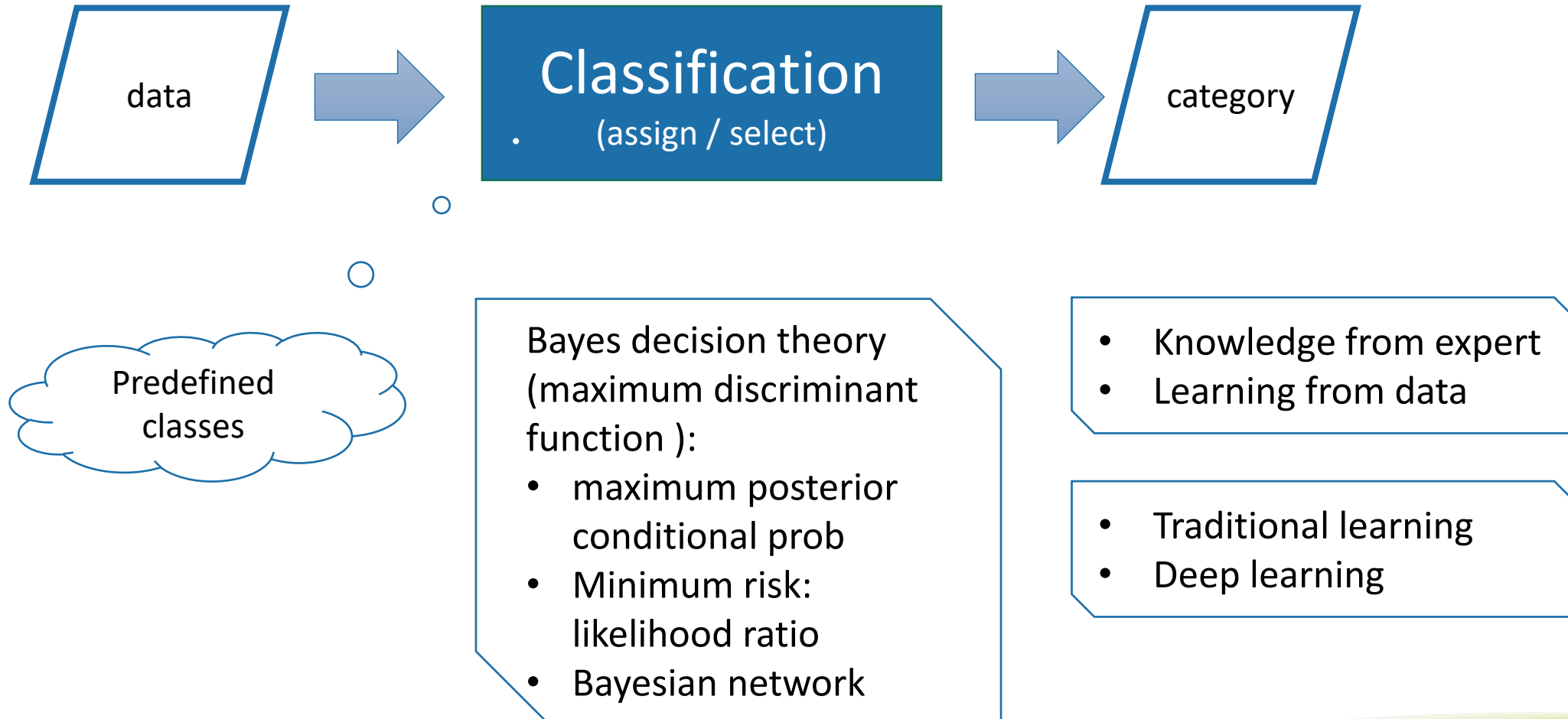
Modul 5: Classification

01 Classification Task Type

Pengenalan Pola
(*Pattern Recognition*)



Classification: Overview



Classification Task Type

Problem: size of predefined classes

Output: number of classes

Output: class or probabilities

Binary Classification
(size=2)

Single-label Classification
(output=1)

Hard Classification
(class)

Multiclass Classification
(size>2)

Multilabel Classification
(output>1)

Soft / Fuzzy Classification
(class conditional probabilities)



Examples



Fraud detection:
binary, hard/fuzzy



Fish classification:
multiclass, hard/fuzzy

Di stasiun KA Kiaracondong (jln Kiaracondong antara kebaktian s d kantor Polisi kebon jayanti) ada gepeng anak usia sekolah ngelem, bicaranya kasar tdk karuan, suka ganggu penumpang wanita! Trm ksh!

Label: “Dinas Sosial (Dinsos) Kota Bandung” and “Satuan Polisi Pamong Praja (Satpol PP) Kota Bandung”

Complaint disposition:
multilabel, hard/fuzzy

Image sources: <https://thenounproject.com/term/fraud-protection/2102650/>

Duda dkk, 2001

Winata, G. I., & Khodra, M. L. (2015, August). Handling imbalanced dataset in multi-label text categorization using Bagging and Adaptive Boosting. In *2015 International Conference on Electrical Engineering and Informatics (ICEEI)* (pp. 500-505). IEEE.



Discussion 1

- Determine classification task type from Part-of-speech (POS) tagger and Named Entity Recognition (NER). Explain your answers.

Stanford CoreNLP 4.0.0 (updated 2020-04-16)

— Text to annotate —

American faces prison over bad hotel review in Thailand

— Annotations —

parts-of-speech ✕

named entities ✕

Part-of-Speech:

	JJ	NNS	NN	IN	JJ	NN	NN	IN	NNP
1	American	faces	prison	over	bad	hotel	review	in	Thailand

Named Entity Recognition:

	NATIONALITY								COUNTRY
1	American	faces	prison	over	bad	hotel	review	in	Thailand



Discussion 2

Determine classification task type from this object detection demo. Explain your answers.



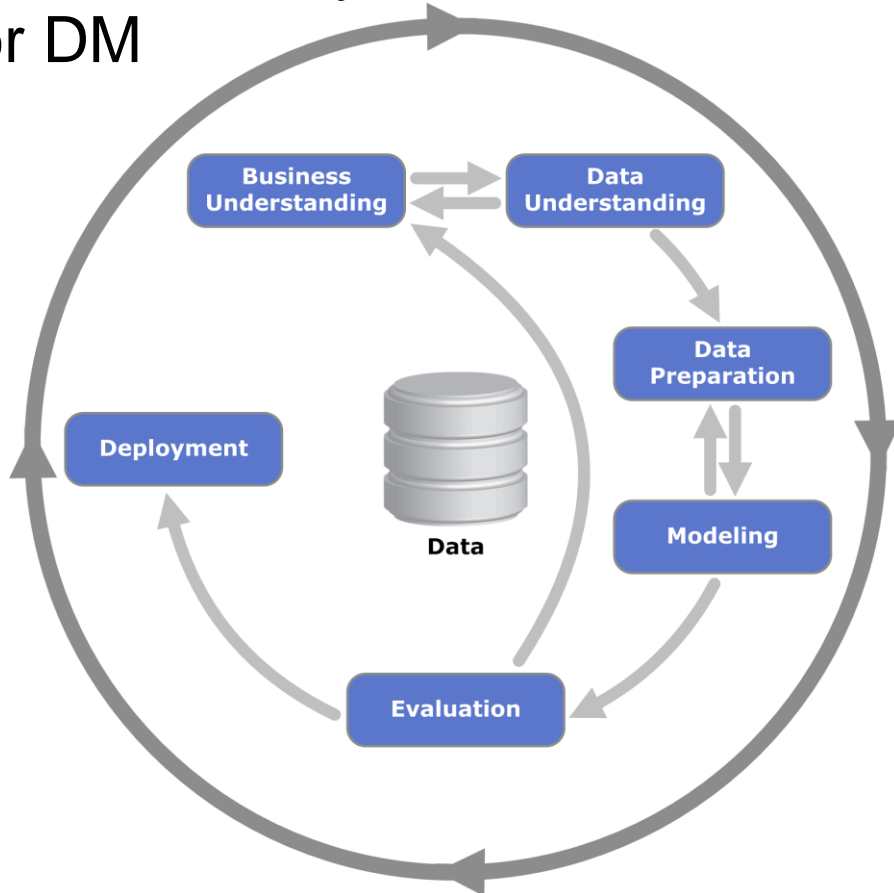
<https://demo.sensifai.com/>

General	
PREDICTED CONCEPT	PROBABILITY
Dish	91.22 %
Meal	89.99 %
Food	89.95 %
Cuisine	71.75 %
Lunch	63.68 %
Breakfast	35.41 %
Produce	35.26 %
Restaurant	15.81 %
Meat	13.57 %
Vegetable	13.31 %

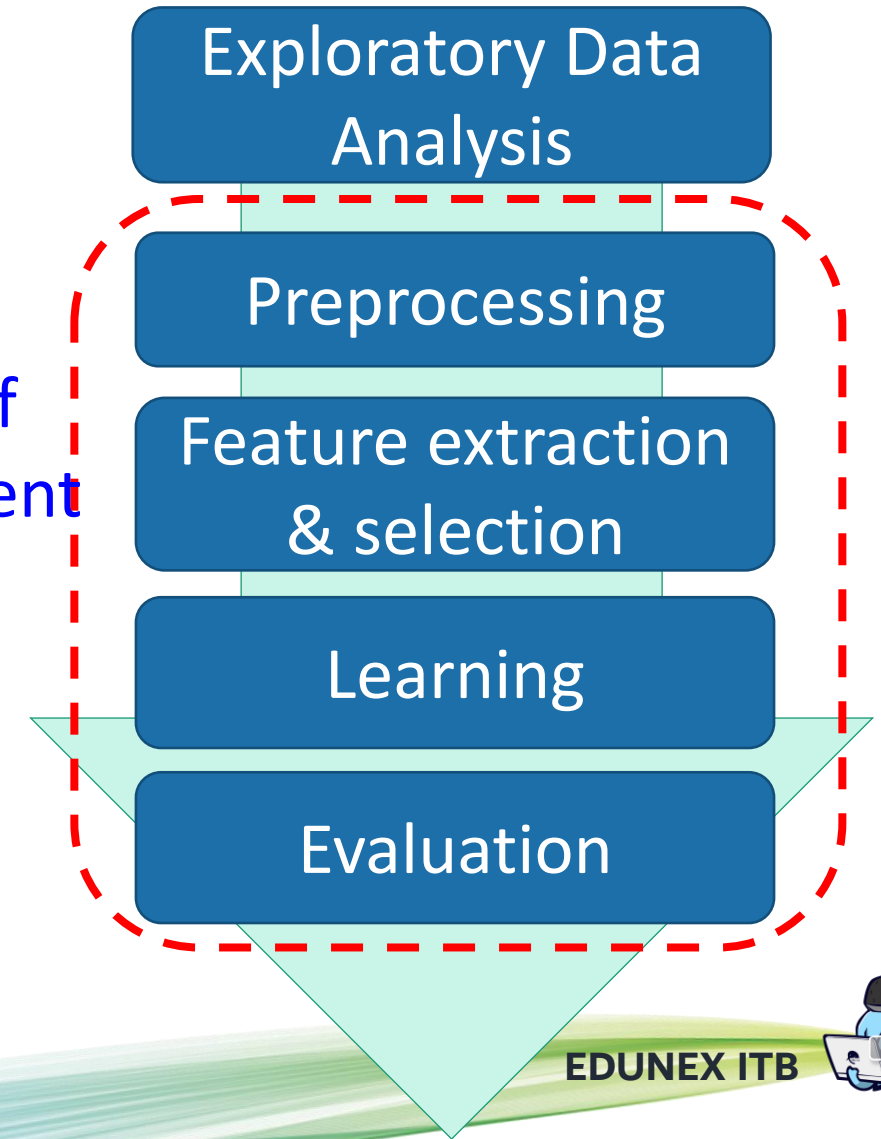


Building Classification Model from Data

Cross-Industry Standard Process for DM



Design of Experiment



Summary

Classification
task:

binary vs multiclass,
single- vs multi-label,
hard vs soft

Building
classification
model from data

Design of Experiments



